DISSERTATION

Titel der Dissertation

Similarity, Cross-linguistic Influence and Preferences in Non-Native Vowel Perception

An experimental cross-language comparison of German vowel identification by non-native listeners

verfasst von

Mag. Nadja Kerschhofer-Puhalo

angestrebter akademischer Grad

Doktorin der Philosophie (Dr. phil.)

Wien, 2014

Studienkennzahl lt. Studienblatt:	A 092 328
Dissertationsgebiet lt. Studienblatt:	Doktoratsstudium Allgemeine/Angewandte Sprachwissenschaft
betreut von:	emer. o. Univ.-Prof. Mag. Dr. Wolfgang U. Dressler

Similarity, Cross-linguistic Influence and Preferences in Non-Native Vowel Perception

Contents

Acknowledgments

I gratefully acknowledge the help, encouragement and advice of many people who supported me during the work on this dissertation.

First, I want to express my gratitude to my advisor, Wolfgang U. Dressler not only for his overseeing this project over the past years, but also for introducing me to the true diversity of linguistic research. His high standards provided me with a model for my own scientific work.

I am endebted to Rudolf de Cillia for his continuous support and encouragement throughout the various stages of this project. His ready ear always made a difference.

For invaluable comments on various parts and versions of this work I owe thanks to Katarzyna Dziubalska-Kołaczyk, David Goldstein, Hans Christian Luschützky, Melanie Malzahn and Sylvia Moosmüller. All remaining faults are of course mine.

For their precious advice, intellectual and practical support and friendship in different stages of this study I want to thank my colleagues at the Department of Linguistics of the University of Vienna and my colleagues at the Acoustics Research Institute of the Austrian Academy of Science.

Special thanks goes to Friedrich Neubarth for his assistance in converting my survey data into a searchable format.

Thanks also to the more than 200 persons participating in the perceptual experiments and to all those who helped me find them, especially to Katalin Árkossy, Elisabeth Herden, Sigrun Inmann-Trojer, Ute Karlavaris-Bremer and Grete Kernegger.

For their help with the manuscript I would like to thank Axel Brom and Laura Wang.

My personal thanks goes to Werner Mayer for his cordial support and enthusiasm in our current project.

This dissertation was supported in part by a dissertation grant F-98-G by the University of Vienna that encouraged me to undertake the large-scale data collection for this project. Part of the data acquisition was also supported by a pre-dissertation travel grant of the University of Vienna.

Most importantly, I want to thank my family for their constant support, love and encouragement, for all their wise words and critical questions. I would never have completed this project without them. Thank you!

Abstract

This dissertation presents an empirical study on German vowel identification by non-native listeners. The central tenet of the study is that the acquisition of second language phonetics and phonology is conditioned by three major forces: (1) *language-specific* phonological and phonetic patterns and *cross-linguistic influence* mainly from the learners' first language L1 on the target language L2, (2) *intra-lingual similarity* of L2 sounds as perceived by L2 learners, and (3) the influence of *universal preferences* and *biases* for particular sound patterns.

Similarity is one of the central concepts in current models of the acquisition of second language phonology. The operationalization of this construct, however, has so far not received sufficient attention. This study aims to integrate contributions from second language acquisition research, empirical studies in acoustic and articulatory phonetics and theoretical approaches from cognitive psychology for a better understanding of the diverse phonetic and cognitive aspects of similarity and its role in non-native vowel perception.

The aim of the experimental part is to investigate the perception of German vowel contrasts by L2 learners from diverse language backgrounds with smaller or larger vowel inventories. Methods of descriptive statistics and higher statistical procedures such as Hierarchical Clustering and Multidimensional Scaling are used to make inferences about inter-category distances and perceived similarity, confusion probability and the direction of perceptual substitutions.

The German vowel system is described in terms of five major classes according to the location of narrowing in the vocal tract (constriction location, cf. Wood 1979): pharyngeal /ɑ ɑ:/, uvular /ɔ o:/, velar /ʊ u:/, mid-palatal /ɛ ɛ: e: œ ø:/ and pre-palatal /ɪ i: ʏ y:/ vowels are further differentiated by mandibular aperture, rounding, constriction degree and phonemic length. The articulatory description of German vowel categories is supplemented with an acoustic analysis of the input stimuli from the perception experiment.

The empirical basis of the study is a corpus of vowel confusion data collected in an identification experiment performed with 173 non-native listeners from ten L1 sub-samples and 18 German natives (control group). Starting with a contrastive analysis for each of the ten L2 learners' native languages, the empirical material is then evaluated on a language-specific, a vowel-specific and a learner-related level. In a cross-language comparison, typical patterns of wrong identifications (*difficulties*), category confusion and direction of perceptual substitutions (*preferences*) are discussed. Moreover, *perceptual vowel maps*, i.e. spatial

representation of the high-dimensional perceptual vowel space, are derived for each of the ten language sub-samples by means of Multidimensional Scaling (MDS).

Confusion matrices and perceptual vowel maps – spatial MDS representations of the L2 listeners' perceptual vowel space – summarize the results of the identification task for each of the language sub-samples and reveal language-specific as well as more common cross-linguistic patterns of the *difficulties* associated with specific contrasts, typical patterns of *perceptual confusion* and *preferences* for specific response categories. Moreover, the data reveal several *asymmetries* in the direction of perceptual substitution processes (perceptual *bias*). Within a given vowel class, some input categories cause more difficulties than others and some qualities attract more responses than others. While *a*-vowels, /iː/, /uː/, /ɛ/ and /ɔ/ are particularly stable, most difficulties are observed with contrasts of front rounded /yː ʏ øː œ/ vowels, /eː/ and /ɛː/. Several sub-samples show a strong tendency of /iː/, /uː/ and /yː/ to attract responses for /eː/-, /oː/- and /øː/-stimuli, respectively, whereas perceptual substitutions in the opposite direction are rather uncommon. For front rounded qualities, the MDS maps of the listeners' perceptual space reveal language-specific "similarity clusters" with respect to substitutions of front rounded vowels with either back rounded or front non-rounded categories.

The study contributes to the modelling of similarity in second language acquisition in three ways: First, it demonstrates the necessity to distinguish *inter*-language similarity from *intra*-language similarity and phonetic distance and similarity from psychological similarity. Second, it demonstrates empirically the non-linear and language-specific relation between articulatory or acoustic properties of vowel stimuli and the L2 listeners' responses. Finally, it provides empirically grounded ways of operationalizing *perceived similarity* between L2 vowel categories in terms of similarity scores and distances in spatial MDS representations.

The study demonstrates that perceived similarity between German vowel categories cannot be predicted directly from physical properties of the acoustic signal but is determined by the listener's attentional tuning to specific dimensions of the perceptual vowel space and by language-specific and more general physical and cognitive *biases* associated with stimuli as well as responses.

Perceived similarity s_{ij} between vowel categories of the target language is modelled as the result of the complex interaction of phonetic proximity p_{ij}, stimuli biases b_i and response biases b_j ($s_{ij} = p_{ij} * b_i * b_j$). *Biases* vary according to characteristics of the acoustic signal, the set of stimuli and response categories and the listeners' language experience (in L1, L2, Ln), L2 proficiency and their individual conception of the target language vowel system.

Deutsche Zusammenfassung – German Abstract

Gegenstand dieser Dissertation ist die theoretische und empirische Beschreibung von Faktoren, die die Identifikation von Vokalen und den Erwerb phonetischer und phonologischer Strukturen der Zielsprache bestimmen: (1) sprachspezifische phonologische und phonetische Strukturen und wechselseitiger sprachlicher Einfluss (*cross-linguistic influence*) zwischen der Erstsprache der Lernenden (L1) und der Zielsprache (L2), (2) *intra-linguale Ähnlichkeitsbeziehungen* zwischen L2-Kategorien, wie sie von den Lernenden selbst perzipiert werden, und (3) sprachübergreifende, universale Präferenzen (*biases*) für bestimmte lautliche Strukturen.

„Ähnlichkeit" ist zentrales Element vieler Modelle der Fremdspracherwerbsforschung. Eine Operationalisierung dieses Konzepts für den Fremdspracherwerb steht aber noch aus. Die vorliegende Studie verbindet Beiträge der artikulatorischen und akustischen Phonetik, der Fremdspracherwerbsforschung und theoretische Ansätze der kognitiven Psychologie, um neue Zugänge zu Konzepten phonetischer und psychologischer Ähnlichkeit und ihrer Rolle im Fremdspracherwerb zu gewinnen.

Im empirischen Teil der Arbeit wird ein Korpus von Konfusionsdaten, d.h. Verwechslungen zwischen L2-Kategorien aus einem Vokal-Identifikationstest mit 173 Deutschlernenden zehn verschiedener Erstsprachen untersucht. Mit Methoden der deskriptiven Statistik und höheren statistischen Verfahren der Hierarchischen Clusteranalyse und der Multidimensionalen Skalierung werden Aussagen zu akustischer und perzeptueller Ähnlichkeit und Distanz, der Fehleranfälligkeit bestimmter Kategorien und der Richtung perzeptueller Substitutionsprozesse abgeleitet.

Grundlage der Datenanalyse ist eine detaillierte artikulatorische Beschreibung deutscher Vokalkategorien nach Konstriktionsort (Region der Verengung im Vokaltrakt, cf. Wood 1979) und weiteren differenzierenden Merkmalen (Kieferwinkelöffnung, Konstriktionsgrad, phonemische Länge, Rundung) sowie eine akustischen Analyse der verwendeten Input-Stimuli. Die 15 deutschen Vollvokale werden in fünf Klassen eingeteilt: pharyngale *a*-Vokale /ɑ: ɑ/, uvulare *o*-Vokale /o: ɔ/, velare *u*-Vokale /u: ʊ/ sowie palatale Vokale, die in prä-palatale /i: ɪ/ und /y: ʏ/ und mittel-palatale Qualitäten /e: ɛ/ und /ø: œ/ unterteilt werden.

Das empirische Material wird unter sprachspezifischen, vokalspezifischen und lernerspezifischen Aspekten sowie sprachübergreifend analysiert. Mittels deskriptiver Statistik werden Fehlerhäufigkeiten und Präferenzen für bestimmte Response-Kategorien sprachspezifisch und sprachübergreifend beschrieben. Die Identifikationsdaten werden in

Konfusionsmatrizen zusammengefasst und quantitativ wie auch qualitativ analysiert. Aus den Konfusionsdaten werden Werte perzipierter Ähnlichkeit und Distanz zwischen deutschen Vokalkategorien errechnet. Mittels Multidimensionaler Skalierung werden daraus räumliche Repräsentationen des perzeptuellen Vokalraums (*perceptual vowel maps*) abgeleitet, die die sprachspezifisch geprägte Wahrnehmung perzeptueller Ähnlichkeiten und Distanzen zwischen Vokalkategorien durch Lernende verschiedener Erstsprachen visualisieren.

Die Daten zeigen sprachspezifische wie auch sprachübergreifende Schwierigkeiten und Präferenzen. Während ɑ-Vokale, /i:/, /u:/, /ɛ/ und /ɔ/ generell geringe Fehlerquoten zeigen, ist die perzeptuelle Unterscheidung vorderer gerundeter Vokalen /y: ʏ ø: œ/ sowie /e:/ und /ɛ:/ deutlich schwieriger. In den meisten Sprachsamples zeigt sich eine starke Tendenz, /e:/-, /o:/- bzw. /ø:/-stimuli als /i:/, /u:/ bzw. /y:/ zu perzipieren, während umgekehrte Substitutionen deutlich seltener sind. Die MDS-Repräsentationen des perzeptuellen Vokalraums zeigen sprachspezifische Substitutionsmuster für vordere gerundete *ü*- und *ö*-Vokale, die entweder als hintere runde *u*-/*o*-Qualitäten oder als vordere ungerundete *i*-/*e*-Qualitäten wahrgenommen werden.

Die Studie leistet einen wesentlichen Beitrag zur Modellierung von *Ähnlichkeit* im Kontext fremdsprachlichen Lernens, indem sie (1) die Notwendigkeit einer grundlegenden Unterscheidung von (a) *inter*-lingualen Ähnlichkeiten zwischen Erst- und Zielsprache und *intra*-lingualen Ähnlichkeiten von Kategorien innerhalb der Zielsprache aufzeigt sowie (b) die Unterscheidung von phonetischer Ähnlichkeit vs. psychologischer Ähnlichkeit verdeutlicht, (2) die non-lineare und sprachspezifische Beziehung artikulatorisch-akustischer Parameter und der Wahrnehmung durch Deutschlernende auf Basis eines großen Samples von 173 Probanden mit zehn verschiedenen Erstsprachen empirisch belegt und (3) Ansätze zur metrischen Operationalisierung perzeptueller Ähnlichkeit und räumlicher Visualisierung des sprachspezifisch geprägten perzeptuellen Vokalraums mittels MDS präsentiert.

Die Untersuchung zeigt, dass perzeptuelle Ähnlichkeiten zwischen Vokalkategorien des Deutschen auf akustischen Eigenschaften des Lautsignals einerseits und sprachspezifischen wie auch allgemeineren physischen und kognitiven Präferenzen (*biases*) andererseits und der relativen Gewichtung (*attentional weight*) von Dimensionen im perzeptuellen Raum beruhen.

Wahrgenommene perzeptuelle Ähnlichkeit s_{ij} zwischen Kategorien i und j wird als Resultat der Interaktion von phonetischer Nähe p_{ij} und der unterschiedlichen Gewichtung von Stimuli-Biases b_i und Response-Biases b_j modelliert ($s_{ij} = p_{ij} * b_i * b_j$). Biases sind von Eigenschaften des Input-Signals, dem Set von Stimuli- und Response-Kategorien und individuellem sprachlichen Wissen, Erfahrungen und Repräsentationen des Zielsprachen-Systems bestimmt.

Table of Contents

Introduction

A crucial component of foreign language learning is the acquisition of the characteristic sound patterns of the target language. At the same time, it is precisely the acquisition of phonetic and phonological patterns of the target language that may lead to considerable difficulties and a traceable "foreign accent" even with very advanced learners. To account for a "foreign" or "non-native accent" in the target language, several explanations have been proposed: phonetic habits due to transfer from the listeners' native language (L1) into the second language (L2[1]) and "reduced neural plasticity" after a so-called "critical period" are among the most widely accepted explanations for the learners' failure to acquire a native-like accent in L2. Moreover, factors such as inadequate input, social identity or insufficient motivation have been proposed to account for L2 learners' incomplete or inadequate achievement in L2 sound production and the fact that a native-like accent is rarely achieved by adult or so-called "late learners".

Still of greater importance, however, is the *inadequate perception* of phonetic information in the L2 as already stated by Lado (1957):

> *"Much less known, and often not even suspected, may be the fact mentioned above that the speaker of one language listening to another does not actually hear the foreign language sound units – phonemes. <u>He hears his own</u>. Phonemic differences in the foreign language will be consistently missed by him if there is no similar phonemic difference in his native language."* (Lado 1957: 11[2])

Numerous studies consider an accurate *perception* of L2 sounds as a central determinant for the development of an appropriate pronunciation in L2. The accurate perception of sounds, contrasts and phonetic details in L2 is considered not only to improve native-like speech production but also to facilitate the development of phonological categories of the target language and to contribute to better word-recognition and ease in speech processing. Several studies on the interaction of perception and production support the argument that the development of L2 sound perception precedes that of production and that – if accuracy in

---

[1] In studies on foreign language acquisition, the *target language* is frequently referred to as *second language* (*L2*), which can be the first foreign language that is acquired by a learner in lifetime after his or her *first language* (*L1*) or any other language acquired after L2 (L3, L4, ..., Ln). L1 is also referred to as *native language*. To simplify, "second language" (*L2*) will be used in this study as a *cover-term* to refer to any language acquired subsequently to L1, irrespective of the total number of languages that have been acquired by the learner, the attained level of proficiency or the context of acquisition (see chapter 1).
Research on *second language acquisition* (SLA) focusses on the learners' language behaviour and language knowledge and the development of the so-called interlanguage. *Interlanguage* refers to the emergent linguistic system that learners develop in acquiring the target language as manifested in *non-native* language production and perception. This study will investigate the identification of German vowels by adult non-native learners acquiring German as L2.

[2] emphasis mine.

perception is not given – production can be native-like only by chance. Furthermore, these studies show *language-specific perception patterns* that constrain the development of mental representations for L2 sound categories. Therefore, the study of *perception* in L2 is a crucial issue for an understanding of difficulties in second language acquisition and the development of phonological categories in the target language.

Second language acquisition (SLA) researchers as well as language teachers agree that *inter-lingual differences* between L1 and L2 have a crucial influence on L2 acquisition and performance. More specifically, it is the *learners' perception* of these inter-lingual differences and their implicit and explicit assumptions that have an influence on the emerging interlanguage system. Phonetic structures, which are phonologically distinctive in L2 but not in L1, are often not correctly identified, discriminated and categorized in L2 perception and therefore cannot be adequately acquired. Moreover, sub-phonemic differences ("fine phonetic details") between L2 sounds and corresponding sounds in L1 may not be recognized sufficiently due to the listeners' perception of L2 sounds as "similar" or even "equivalent" to L1 categories. Inadequate perception due to the influence of the learner's L1 leads to "non-native" pronunciation of L2 sounds.

Experimental studies in SLA provide abundant evidence for the effects of influence from the learners' native language on the perception of the target language. Structures of the learners' native language are commonly assumed to be transferred into the target language. However, *transfer* of linguistic forms from L1 to L2 is not the only strategy of L2 learners to acquire the target language system. Rather, L2 learning does not only involve transfer but also implies a variety of other language-related cognitive processes. Processes of interaction between L1 and L2 such as comparisons of structures between or within language systems, discrimination and categorization, ratings of perceived differences and distances, judgements about similarity and typicality, strategies of preference or avoidance, reaction times, etc. are consciously or subconsciously applied. Cognitive processes involving interactions between the target language and other previously or subsequently acquired languages represented in the learners' mind will be referred to here as *cross-linguistic influence* (following Odlin 1989, 2005; Jarvis & Pavlenko 2008; Jarvis 2012).

The acquisition of a second language's *phonological system* involves the establishment of relations of functional contrast, similarities and differences between linguistic structures together with hypotheses on probability distributions and strategies of preference or avoidance for specific linguistic structures. In other words, it is not only transfer from L1 to L2 that shapes the emergent phonological system of a learner's interlanguage. Rather, the specific ways that L2 sound structures are perceived, produced and mentally represented are determined by *cross-linguistic influence* of the the native language, the target language and other languages represented in the learner's mind.

Beside a number of empirical studies that describe the phenomenon of cross-linguistic influence and language-specific perception of sounds in L2 acquisition, several theoretical models and hypotheses have been proposed to describe constraints on L2 perception as influenced by the L1 phonological system. A number of models of L2 speech perception and acquisition have been proposed such as Kuhl's Native Language Magnet Model (Kuhl 1995; Kuhl & Iverson 1995), Flege's Speech Learning Model (Flege 1995) or Best's Perceptual Assimilation Model (Best 1995; Best/Tyler 2007). These models posit cognitive processes such as "perceptual assimilation", "equivalence classification" or "category goodness" to account for the influence of L1 on the learners' success to perceive L2 sounds correctly and to develop adequate mental representations for L2 categories. The primary aim of research within these theoretical accounts is to predict the relative *difficulty* of non-native contrasts in L2 perception, production and mental representation. To account for difficulties in L2 acquisition, SLA research largely focusses on contrastive analyses comparing L1 and L2 structures and on experimental testing. Difficulties are assumed to result from inter-lingual *differences* and *similarities* between structures of L1 and L2.

"Similarity" is one of the central concepts in many studies on second language acquisition. The assumption that *similarity* between sounds in L1 and L2 limits perceptual accuracy and that perceptual accuracy limits production and mental representations of L2 sounds is a common notion in models and theories on the acquisition of second language phonology. It is, however, surprising that the operationalization of the construct "similarity" itself has, so far, not received sufficient attention in second language acquisition research.

One of the central aims of this study is the operationalization of "similarity" as a central factor in non-native vowel identification. A further aim is to demonstrate the interacting effects of similarity and cross-linguistic influence in L2 vowel perception in an experimental cross-language comparison.

The study shall provide a comprehensive approach to L2 perception by integrating contributions from *second language acquisition* research, empirical studies in articulatory and acoustic *phonetics* and theoretical concepts from *cognitive psychology* for a better understanding of the diverse phonetic and cognitive aspects of similarity and its role in non-native vowel perception.

Similarity between vowel categories will be discussed under two aspects: (1) in terms of an analysis of *articulatory* and *acoustic* characteristics of vowel signals (objectively measurable similarity) and (2) in terms of *perceived* similarity of vowel categories within the L2 system (subjective similarity as perceived by L2 learners).

In theoretical accounts, similarity commonly refers to inter-lingual relations between L1 and L2 structures, whereas *intra*-lingual relations of similarity between items *within* the L2 system are commonly disregarded. Both types of similarity relations may even hinder the development of adequate L2 categories. A central tenet of the study is that not only inter-lingual relations between L1 and L2 sounds, but also *intra-lingual similarity* relations of categories within the target language have a crucial role in the development of mental representations for L2 sound categories. Both, cross-linguistic influence in general and intra-lingual similarity relations established by the learners are considered to determine the learners' success in developing adequate L2 sound categories.

The effects of cross-linguistic influence and intra-lingual similarity of categories within the target language will be exemplified by a specific sub-set of the phonological system – the vowel inventory. The aim of the experimental part is to investigate the perception of German vowel contrasts by L2 learners from diverse language backgrounds with smaller or larger vowel inventories. German has a comparatively rich and complex system of vowel contrasts that may cause considerable difficulties in L2 acquisition particularly for learners from native languages with a smaller, less complex vowel inventory. This study will show how the perceptual identification and categorization of German vowel sounds and the learners' difficulties as well as their strategies of preference or avoidance in an experimental task are influenced not only by language-specific factors but also by learner-related and more universal aspects of human speech perception and cognition.

The experimental design of this study is largely motivated by the author's experience in teaching German as a foreign language in groups with learners from diverse language backgrounds. For many German learners, problems in perception and categorization of German vowel phonemes do not only have an effect on "foreign accent" in L2 but appear also to cause severe problems in the acquisition of orthography, lexical items and morphological forms (cf. Kerschhofer-Puhalo 2010b, 2012, in press a).

The major purpose of an empirical investigation of German vowel perception by L2 learners from diverse language backgrounds is to describe possible language-specific and/or universal patterns of *difficulties*, *category confusion* and systematic perceptual vowel *substitutions*. Integrating theoretical approaches and empirical data on L2 German vowel perception, the study's results will contribute not only to *phonological and typological theory*, but will also serve *pedagogical purposes*, by providing controlled empirical evidence for difficulties and perceptual confusion of German vowel categories which were observed among L2 learners from ten different languages. The ten languages under investigation are Albanian, Arabic, English, Farsi, Hungarian, Mandarin, Polish, Romanian, SerBoCroatian, and Turkish. The language choice was motivated by statistics on the most common native languages of students in the Austrian educational system.

In the world's languages, substantial variation in number and quality of vowel contrasts across phonological systems is observed. This *variation* concerns not only the number and type of phonemic contrasts but also sub-phonemic phonetic variation along articulatory and acoustic parameters, i.e. phonetic details in the realization of a given vowel category. While phonetic variation is considered to vary continuously along specific articulatory or acoustic parameters, vowel perception is commonly regarded as categorical. However, *categorial perception* is language-specific. In other words, no linear relationships between gradual variation along phonetic parameters and categorical perception can be assumed. Rather, category boundaries vary language-specifically and language users from diverse language backgrounds exploit different phonetic parameters to differing extent according to the structure of their native language vowel system.

Even though typological studies and cross-linguistic comparisons reveal enormous variation in number and types of vowel contrasts, system constellations and phonetic parameters for phonemic contrasts, there are some distributional patterns that are particularly common in the vowel systems of the world's languages. *Common patterns* in vowel inventories and the frequent occurrence of specific vowel qualities, e.g. a preference for /i/, /u/ and /a/ in the languages of the world, have been argued to be "*universal*" and motivated by human speech capacity in general and by characteristics of the human vocal tract and the peripheral auditory system in particularly. Sounds that are "easier" in articulation and perception are expected to occur "universally" more frequently and are also referred to as less "marked" than more complex, more marked structures that universally occur less frequently.

A large number of studies from diverse theoretical backgrounds have argued that phonetically grounded universal preferences and markedness conditions determine the distributional patterns of vowel systems in the languages of the world as well as the organization of individual vowel systems. However, a question of intense scientific debate is whether these universal forces are merely phonetic or cognitive in nature.

Universal preferences and *markedness constraints* have also been an issue in second language acquisition research. Studies in SLA show that structures and developmental sequencies in learners' interlanguages largely correspond to patterns observed in natural language systems. Universal factors resulting from major characteristics of human speech capacity are assumed to determine vowel inventories in natural languages as well as L2 learners' difficulties and their strategies in L2 language acquisition. These universal forces may even "overrule" language-specific characteristics of the learners' L1 explaining the occurrence of common patterns of difficulties and preferences in interlanguages independently of the learners' native language. More specifically, less difficulties with universally more common "unmarked" sound structures are observed in interlanguages with L2 learners from diverse language backgrounds.

Asking for the phonetic and/or cognitive bases of these universal preferences and constraints in language systems and interlanguages, however, also addresses the issue of the relation between *phonetic form* and *phonological function*. For native but particularly for non-native perception, it is important to emphasize that no linear relation can be assumed between articulatory constellations and the resulting acoustic characteristics of the speech signal on the one side and the listeners' perception of this signal on the other side. Rather, the matching of acoustic signal (form) and phonemic category (function) must be considered to vary according to the listeners' native language and his/her linguistic knowledge and experience in the target language.

A fundamental differentiation in the present study refers to the human articulatory and acoustic vowel space as distinct from the language-specific perceptual vowel space. The *perceptual vowel space* refers to the cognitive organization of a particular vowel system and the listeners' selective attention to specific phonetic parameters. The perceptual vowel space in a language user's mind may be compared to a language-specifically tuned *cognitive map*.

The current study will present a method to visualize L2 learners' cognitive maps for the L2 vowel system by the statistical procedure of Multidimensional Scaling (MDS) based on category confusion data from an identification task.

The cognitive vowel map, i.e. the psychological structure of a vowel system, may vary according to the listeners' native language, their awareness of particular phonemic contrasts in the language under consideration and their perceptual sensitivity and selective attention to specific phonetic details in the signal.

In the course of the L2 acquisition process, the cognitive map of the L2 vowel system may be reorganized by increased awareness of particular phonemic contrasts and phonetic characteristics that the learner has not been aware of before. The mental representation of the L2 vowel system may also be further refined by increased perceptual sensitivity and perceptual re-allocation to specific phonetic details that the listener has previously not been aware of. It is expected here that the listener's knowledge and experience will not only determine the extent to which L2 vowel sounds are identified correctly, but also influences his/her strategies in cases where correct identification seems at risk.

In the experimental part of this study, a category identification task is used to reveal L2 listeners' difficulties in identifying German vowel categories correctly. Wrong identifications are considered to reflect the learners' difficulties in differentiating specific vowel contrasts. The results are subject to a quantitative as well as a qualitative analysis, revealing difficulties in correct identification, category confusion and perceptual substitutions, and the learners' tendencies to prefer or neglect specific response options.

The purpose of the empirical investigation is to exemplify the interaction of language-specific and universal factors based on empirical evidence for difficulties that L2 learners of German may encounter upon perceiving phonemes of a phonological system, which is more complex than their native phonemic system. The data analysis provides a language-specific description for each of the L1 sub-samples and a cross-language comparison of patterns of perceptual confusion and perceived similarity between L2 categories.

This kind of large-scaled overview and cross-language comparison of German vowel perception by learners from diverse language systems may reveal the influence that universal phonetic and cognitive factors may have in second language vowel perception better than studies focussing only on one single language. The data analysis will not only provide evidence for phonological or typological theory but will also serve pedagogical purposes.

Unlike many other studies, which investigate only a subset of L2 phonemes, the full repertoire of German vowel phonemes is investigated in the present study. The decision to include the whole vowel inventory in the study is motivated by three arguments: (1) the dynamic relationships between all members of the target system have to be considered in order to understand the process of system organization and category formation of new or similar L2 sounds, (2) in a cross-language comparison of vowel perception the possible candidates for perceptual substitutions may vary language-specifically, e.g. for front rounded vowels both front unrounded or back rounded vowels are possible substitutes, (3) for pedagogical purposes, an empirical investigation of the complete vowel inventory is more relevant and useful than studies of single vowel qualities or a sub-set (e.g. only front vowels) of vowels.

The current study will ask for the phonetic and cognitive factors that determine (1) the listeners' *difficulties* in identifying German vowels correctly, (2) patterns of perceptual *confusion* between L2 categories, and (3) the listeners' *preferences* for specific categories as response options.

Referring to the notions of similarity and cross-linguistic influence, the study will ask for those factors that influence *perceived similarity* of vowel categories between and within language systems and for the impact of cross-linguistic influence interacting with universal preferences and constraints that determine the listeners' performance in an identification task in L2. The theoretical concepts of *markedness* and *universals* will be applied here to extend the scope and descriptive power of traditional contrastive analysis of the systems of L1, L2 and IL.

To conclude, the central tenet of this dissertation is that cross-language speech perception is conditioned by three major factors:

(1) *cross-linguistic influence* between the learners' first language L1 and the target language L2 (and other languages acquired),

(2) relationships of *intra-lingual perceived similarity* between vowel categories of L1 and L2 and between categories within L2 as established by the learner at a given stage of L2 acquisition, and

(3) cross-linguistically, by the influence of universal phonetic and phonological *preferences and markedness* conditions that determine or constrain the acquisition process and the listeners' difficulties and choices of specific response categories in an identification task.

An integrated view of these three major factors shaping interlanguage phonology enables a comprehensive account of perceptual learning in second language acquisition.

Organization of the dissertation

Chapter 1 will resume basic theoretical notions and concepts of SLA research that have been claimed to play a role in the acquisition of phonetic and phonological patterns in L2 acquisition. Moreover, the chapter will review concepts of markedness and universals and their status within second language acquisition theory.

Chapter 2 discusses fundamental aspects of human speech perception and language processing and the development of speech perception in L1 and L2, i.e. the mechanisms by which sounds are processed, acquired, internalized, and used in L1 or L2. Moreover, experimental methods in L2 perception are reviewed. The last section of chapter 2 discusses current theories of L2 perception.

Chapter 3 discusses different aspects of “similarity” in phonetics and SLA research. The chapter reviews aspects of phonetic similarity in articulatory, acoustic and perceptual terms and relates the diverse notions of similarity with the learners’ ease or success in identifying L2 sound patterns correctly. The impact of acoustic-phonetic structures on the language-specific perception and mental representation of L2 sounds is discussed in the last section of this chapter.

In chapter 4, phonetic and phonological definitions of vowels are reviewed. The discussion will focus on aspects of system size and contrast types and on the phonetic correlates of phonemic oppositions. The chapter discusses phonetic, phonological and typological aspects of vowel inventories and presents evidence for specific patterns that cross-linguistically seem to be preferred in the vowel inventories of the world’s languages. Theories on vowel inventory typology that have been proposed to account for these universal patterns and preferences such as Markedness Theory, Natural Phonology, Quantal Theory and articulatory-oriented approaches are discussed in the last section.

Chapter 5 presents a phonetic and phonological description of the German vowel system. German vowel categories will be classified in terms of articulatory regions of the vocal tract and will be described in terms of articulatory and acoustic properties. The acoustic-phonetic data presented here is based on a detailed analysis of the input stimuli from the experimental part of this study. A description of the German vowel system that is based on articulatory and acoustic characteristics offers insights into *phonetic similarities* between German vowel categories. The last section of this chapter presents an analysis of phonetic similarity of German vowels in terms of acoustic-phonetic properties and deduces predictions for difficulty and category confusion between German vowels by means of Hierarchical Clustering.

The experimental design, the input material and the test procedure of the current study are presented in chapter 6.

Chapter 7 provides a detailed discussion of procedures of data analysis and crucial aspects of data interpretation. The goals, scope and limitations of data analysis and data interpretation are described for each of the type of quantitative or qualitative analysis preocedure (for an overview of all types of analyses performed in this study, see Table 7.8).

A detailed discussion of the study’s experimental results focussing on language-specific, learner-related and category-specific aspects is presented in chapter 8, 9, 10 and 11.

Chapter 8 describes individual factors and their influence on the L2 learners' differential success in the perception experiment.

Chapter 9 provides a first overview of the major results for the full sample of 173 non-native listeners and presents a cross-language comparison on language-specific differences and commonalities in difficulties, perceptual substitution processes and category preferences. A general analysis of results for the full L2 sample will include the following aspects: (1) a quantitative analysis of (a) L2 listeners' *difficulties*, i.e. the ratio of wrong and correct identifications, and (b) the listeners' *preferences* for specific response categories in a cross-language comparison and a vowel-specific comparison, followed by (2) a qualitative analysis of (a) *confusion patterns* and perceptual substitution processes, (b) *similarity scores* for the full L2-sample, and (c) *cognitive maps* of the *perceptual vowel space* for each of the ten sub-samples of L2 listeners derived from a Multidimensional Scaling Analysis (MDS).

Chapter 10 discusses the results for each of the language sub-samples separately. For each of the participants' native languages, the analysis will offer (1) a description of *general characteristics* of the language under discussion, (2) a *phonological and phonetic description* of the vowel system, (3) a brief *contrastive outline* describing previous research and possible areas of difficulty with L2 German vowel perception, and (4) a detailed analysis of the quantitative and qualitative results of the perception experiment (as outlined above).

Chapter 11 provides a cross-language comparison of vowel-specific difficulties, preferences and directions of perceptual substitution patterns for each of the German vowel categories. The vowel-related focus of this chapter allows for a *cross-language comparison* for German vowel classes and single categories and reveals common patterns observed in several language sub-samples. Moreover, it provides insights into category-specific patterns *within* a given vowel class. These common patterns within a given vowel class refer not only to comparisons of error rates (difficulties) but also to the occurrence of perceptual substitutions (confusion patterns) and the listeners' bias to select specific response patterns (preferences and asymmetries).

Chapter 12 will take up the diverse strings of argumentation and resume the experimental results focussing on the issues of "similarity" and "bias" to account for the differential results in terms of error rate (difficulties), category confusion due to similarity, and category preferences (biases) and asymmetries in perceptual substitutions. The major aim of this last chapter is to relate the phonetic facts and empirical results with theoretical contributions from research in cognitive psychology and findings on speech perception and language learning to identify those elements that are essential to a model of *intra-lingual similarity* of L2 categories and *perceptual bias* in second language acquisition.

General conclusions, pedagogical implications and an outlook for further research are given in the final section of this study.

1 Cross-linguistic Influence, Similarity and Markedness in Language Acquisition

1.1 Fundamental concepts in second language acquisition research

Abundant evidence for influences from learners' native language (L1) on a subsequently acquired language is found in studies on the acquisition of phonetic and phonological patterns in foreign language learning. While the idea of cross-linguistic influence from the learners' native language on the acquisition of a subsequently acquired language is widely accepted, less consense is given with respect to substance and motivation of these influences by language-specific, individual or more general "universal" factors in human language that have been claimed to be relevant in language performance, first language acquisition, language variation and change as well as in foreign language learning.

This chapter will resume basic theoretical notions and concepts of second language acquisition (SLA) research that have been claimed to have a role in the acquisition of the target's language phonetic and phonological patterns. It will start with the notions of transfer and crosslinguistic influence (CLI) (section 1.1.1) and the concept of "similarity" (section 1.1.2) as crucial elements in SLA theories, followed by a discussion of the status of markedness and universals in SLA research and theoretical linguistic frameworks (section 1.1.3, 1.1.4, 1.1.5 and 1.2) and will then discuss the interaction of language-related, individual and universal factors in second language acquisition (section 1.3).

In this study, the notion *first language* or *native language* (*L1*) will refer to the *first* language acquired by a speaker/listener from a chronological perspective. The term *second language* (*L2*) will be used here as a *cover term* for *any language acquired subsequently and non-simultaneously*[3] to L1, irrespective of the attained level of proficiency, the context of acquisition, the type of received input or the total number of languages that have been acquired by the learner. *Second language acquisition* (SLA) will include here the acquisition of a third language (L3) or any other language acquired after L1 (Ln). L2 can also be referred to as *target language* as opposed to the *source language* (L1).

---

[3] Cases of simultaneous L1 acquisition will be here referred to as simultaneous bilingualism.

1.1.1 Transfer, interference and cross-linguistic influence

Speech perception and speech production in second language acquisition are largely influenced by the learners' native language (L1). Most current theories of second language acquisition assume that influences from L1 are a crucial factor in acquiring the phonological system of L2 and that phonetic and phonological *differences* between L1 and L2 have an influence on the acquisition process in perception and production.

One of the central concepts in applied linguistics, second language research and language teaching is *language transfer*. However, its status, its nature and its importance has been reassessed several times in the last decades of research on language learning.

Since the 1950s, *transfer* and *interference* are central notions in studies on language contact and foreign language acquisition (Weinreich 1953; Lado 1957) referring to the influence of earlier learned linguistic structures on the target language, whether from the native language L1 or from any other language previously or subsequently acquired language.

> *"Those instances of deviation from the norms of either language which occur in the speech of bilinguals as a result of their familiarity with more than one language, i.e. as a result of language contact, will be referred to as INTERFERENCE phenomena."* (Weinreich 1953: 1)

Weinreich (1953: 7) distinguishes (a) *mere transfer* or "borrowing" of linguistic elements, from (b) *inter-lingual identification* of forms that are considered as "equal" or "similar" by the bilingual. Referring to the acquisition of a phonemic system, Weinreich (1953: 18f) described four basic types of interference:

(1) *sound substitutions*, when a sound of L2 is identical with one from L1 but differs in pronunciation),

(2) *under-differentiation*, i.e. confusion of two L2 sounds whose counterparts are not distinguished in L1 (e.g. the confusion of German /œ/ and /ɛ/ by Polish learners of German, see Kerschhofer 1995),

(3) *over-differentiation*, i.e. phonemic distinctions from L1 are imposed on the sounds from L2, where they are not required (e.g. the use of durational information by German listeners for contrasts that differ in spectral cues such as /ɛ/ - /æ/ in American English, see Bohn & Flege 1990), and

(4) *reinterpretation of L2 distinctions*, when redundant features in L2, due to their relevant status in L1, are used to distinguish phonemes in L2.

Weinreich's definition of interference focuses on "deviations from the norm", i.e. on what has later been termed *negative transfer*. While positive transfer contributes to the learners' success in the target language, negative transfer of L1 into L2 may result in errors, overproduction, underproduction, miscomprehension and other effects that constitute the divergence of the learners' performance from the norms of the target language (Odlin 1989: 36, 167).

The differentiation of positive and negative transfer was a basic concept in the Contrastive Analysis paradigm. The *Contrastive Analysis Hypothesis* (CAH) assumes that *differences* between languages could be used to predict difficulties in language learning. In case of similar structures, positive transfer would occur while differences would result in negative transfer, i.e. interference. In other words, similar elements would be easy for L2 learners while L1-L2 differences would cause difficulties (Lado 1957: 2). A more differentiated account was presented by Stockwell, Bowen & Martin (1965a) in their hierarchy of difficulty that distinguished structural and functional correspondences in L1 and L2.

While CAH in its earlier strong version (Lado 1957) attempted to predict all errors in L2 performance by transfer from L1, in its weaker less predictive version transfer from L1 served as post-hoc explanation for interferences in non-native performance (Wardhaugh 1970).

However, a straightforward contrastive analysis of L1 and L2 cannot account for all phenomena observed in the acquisition of second language speech (cf. Dulay, Burt & Krashen 1982: 97f.). Research findings from numerous studies on the phonological acquisition of a second language sound system show much more than simple transfer from the first language. In particular, CAH provides no complete explanation for the relative *degree of difficulty* of specific sound structures and for *individual differences* in foreign accent and acquisition rate. Therefore, additional concepts have to be considered. As Broselow (1983) states,

> "*our inability to predict the occurrence and nature of many errors may well stem from inadequacies in our understanding of native speakers' competence rather than from the failure of the contrastive analysis hypothesis*". (Broselow 1983 ([1987: 292f]).

With the decline of behaviourism, structuralism and the audiolingual method and with the shift from behavioristic to cognitive approaches to learning, CAH and the concept of transfer lost its popularity. But though CAH in its strong form could not be proved empirically, until today the concept of transfer as well as notions of CAH are of persistent influence on research and teaching practice in second language acquisition (Leather & James 1991; Hansen & Zampini 2008).

Research on transfer and other phenomena in language acquisition and cognition developed in phases as roughly described in Figure 1.1 (for a detailed discussion, see Jarvis & Pavlenko 2008: 4ff; Odlin 1989: 6ff): In phase 1, transfer became a widely accepted variable in language acquisition serving as *explanans* (cf. Weinreich's (1953) four types of transfer/interference). In the 1970s, a shift of research from transfer as explanans to transfer as *explanandum* caused increasing interest not only on the identification, scope and quantity of transfer, but also on transfer itself, its sources, causes, constraints and directionalities (phase 2) remained important issues which receive ongoing scientific attention until today. Over the years, the phenomenon had attracted sufficient recognition to give rise to several theoretical models and hypotheses (phase 3), whereby empirical research became highly theory-driven. Advances in research technology (e.g. brain-based methods, see section 2.3.7) and other disciplines (e.g. neurolinguistics) caused a further shift and expansion of empirical methods (phase 4).

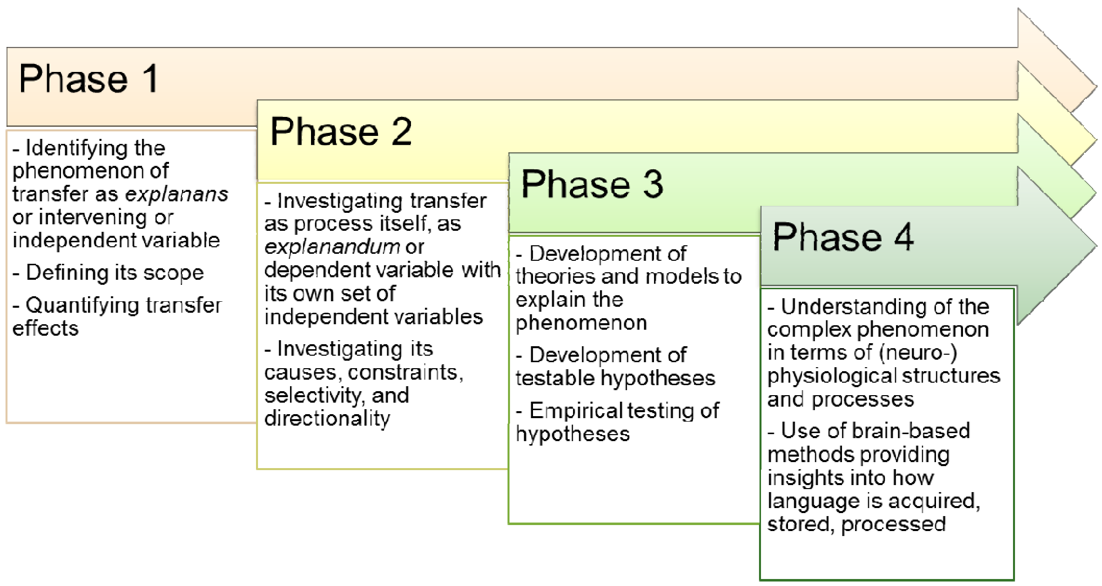

Figure 1.1: Phases of research on transfer (cf. Jarvis & Pavlenko 2008)

Several research questions as well as theoretical concepts that developed in the earlier phases of SLA research continue to be of influence until today. One of the most influential concepts is *interlanguage* (Selinker 1972, 1992).

Interlanguage (IL) (Selinker 1972) refers to emerging linguistic systems developed by language learners approximating eventually the target language system but using features from their native language L1. Selinker (1972) defines interlanguage (IL) as a

> "*separate linguistic system based on the observable output which results from a learners' attempted production of a TL norm. This linguistic system we will call 'interlanguage' (IL).*" (Selinker 1972: 214)

Interlanguages are idiosyncratically based on the learner's experiences and are shaped by several learning mechanisms like transfer from L1, transfer of training, strategies of L2 learning (e.g. simplification), strategies of L2 communication, and overgeneralization of target language patterns (Selinker 1972, 1992).

Odlin (1989: 25ff) emphasizes that an attempt to define what transfer is requires statements about *what transfer is not* (for discussion, see also Jarvis & Pavlenko 2008: 10ff). Transfer is (1) not simply a consequence of habit formation, (2) not simply interference, (3) not simply a falling back on the native language and (4) results not always from native language influence. A fully adequate definition of transfer

> "*... seems unattainable without adequate definitions of many other terms, such as* strategy*,* process*, and* simplification*. Such definitions may presuppose an account of bilingualism that accurately characterizes relations between transfer, overgeneralization, simplification, and other second language behaviors. An adequate account of bilingualism would in turn have to include an accurate neurological model of language, since, presumably, the influence of one language on another has something to do with the storage of two knowledge systems within the same brain (…). Thus, one might plausibly argue that a fully adequate definition of transfer presupposes a fully adequate definition of language.*" (Odlin 1989: 28[4])

However, in his book "*Language Transfer. Cross-linguistic influence in language learning*" as a working definition for "transfer" Odlin (1989) proposed an often cited and rather broad definition of *transfer* as

> "*the influence resulting from similarities and differences between the target language and any other language that has been previously (and perhaps imperfectly) acquired*" (Odlin 1989: 27)

This definition emphasizes two factors that will be crucial in the present study, namely the existence of

(1) *cross-linguistic differences* between the systems of the target language L2 and the native language L1 or any other so far acquired language(s),

(2) *cross-linguistic similarities* of items in the languages involved and

(3) structures revealing more than direct transfer, i.e. *other types of cross-linguistic influence* of L1, L3 or Ln on the target language.

In more recent studies, due to criticism of transfer and the CAH, the notion *cross-linguistic influence* (*CLI*) is preferred to refer to the influence of one language on another in language use (comprehension and production) and to refer to a variety of forms of language-related

---

[4] emphasis mine.

behaviour, including categorization, similarity and typicality judgements, reaction times, etc. (Odlin 1989, 2005; Kellerman 1995; Jarvis & Pavlenko 2008; Jarvis 2012).

Both terms, CLI and transfer, are often used interchangeably. However, they do not cover the same range of phenomena observed in language learning. While transfer is a very useful notion to describe the literal adoption of forms or structures or meanings from one language to the other, the concept of *cross-linguistic influence* seems to be more appropriate to refer to the learners' actual *use* of specific language forms, structures and functions (or more specifically their overuse, underuse, or even avoidance) in one language under the influence of another language (Odlin 1989; Jarvis & Pavlenko 2008; Jarvis 2012).

Inter-lingual identifications (Weinreich 1953), equivalence classifications (Flege 1987, 1995) and the relative degree of *similarity* of elements of L1 and L2 as perceived by the learners cannot be described by mere transfer from one language to the other. Rather, inter-lingual identifications and relations of similarity between linguistic items must be regarded as the result of cross-linguistic influence of two (or even more) languages involved. CLI can also manifest itself in *preferences* for specific linguistic structures instead of others (Jarvis & Pavlenko 2008: 11ff). CLI moreover covers those mental processes that determine not only the outcome of L2 use in everyday communication but also in an experimental setting.

Moreover, CLI allows to refer to the influence of any language on another in any direction (e.g. L2 > L3 or L2 > L1).

> "*Crosslinguistic influence is the often preferred term for a phenomenon more commonly known as transfer, which is the influence of one language on another, as witnessed in the language use (both comprehension and production) and other language-related behavior (e.g., categorization, gesturing, similarity and typicality judgments, reaction times) of both individuals and discourse communities. The former term is often preferred because it more naturally refers to a wider variety of crosslinguistic effects—including the overuse, underuse, and avoidance of language forms, functions, and structures in one language due to the influence of another language, as well as crosslinguistic effects at the level of conceptualization and mental processing—whereas the latter term is often interpreted as the literal transfer of a form, structure, or meaning from a person's knowledge of one language to their use of another (…)*" (Jarvis 2012: 1)

Therefore, in the current study, *cross-linguistic influence* (CLI) will be preferred to account not only for direct transfer of linguistic structures from one language to another but also to refer to the wide range of implicit and explicit assumptions and strategies in L2 performance, including mental processes that are involved when learners are establishing mental representations of vowel categories and relationships of *similarity* and *contrast* of vowel categories.

To conclude, the notion of cross-linguistic influence (CLI) is preferred in the present study because of its wider scope that refers to several forms of language behaviour, language processing and language learning. Cross-linguistic influence also includes effects at the level of c*onceptualization* and mental *processin*g that are crucial in the present study.

1.1.2 Similarity

A crucial role in predicting difficulties in language acquisition is assigned to the concept of *similarity*. The Contrastive Analysis Hypothesis postulates that a contrastive description of differences between L1 and L2 will explain difficulties in L2,

> "*The greater the difference between two systems, i.e. the more numerous the mutually exclusive forms and patterns in each, the greater is the learning problem and the potential area of interference.*" (Weinreich 1953: 1)

Structures *similar to L1* are considered to cause less problems and to be acquired more easily by an L2 learner: "*those elements that are similar to his native language will be simple for him, and those elements that are different will be difficult"* (Lado 1957: 2). Lado (1957) expected that L2 sounds similar to L1 will be easier in acquisition:

> *"In learning the sound system of a foreign language one finds sounds that are <u>physically similar</u> to those of the native language, that structure similarly to them, and that are similarly distributed. <u>"Learning"</u> of such phonemes <u>occurs by simple transfer without difficulty</u>. On the other hand, one also finds sounds that are not part of the sound system of the native language, that structure differently, or that are differently distributed. Learning of these occurs more slowly, and difficulty with them is more persistent. In fact, learning of the latter actually means learning the sounds of the language. We therefore seek to find those problems, and we will find them by the structural comparison of the two sound systems."* (Lado 1957: 12[5])

However, several studies have demonstrated that it is exactly *similarity* that causes transfer from L1 and triggers substitution processes. Sounds substituted for a specific L2 sound usually are acoustically or articulatorily similar to the original target sound. *Similar* sound structures in L2 are more difficult than sounds that are perceived as clearly *different* from L1 categories. Dissimilar sounds are expected to show less L1 influence (e.g. Flege 1988b, c, 1995; Wode 1981; Bohn 2002). In this sense, similarity may even hinder correct acquisition, when learners produce dissimilar sounds more authentically than similar sounds. This fact, as is argued by Flege (e.g. 1987, 1988c, 1995; Bohn & Flege 1990), is due to *equivalence classification*, i.e. a sort of "assimilation" of L2-structures to L1-categories. "New", i.e. non-

[5] emphasis mine.

equivalent sounds will be learned better than equivalent or "similar" sounds: For example, Flege (1987) found that French /y/ was better acquired than French /u/ by native English L2 learners of French, because /y/ has no equivalent while French /u/ is subject to an equivalence classification with English /u/.

Correct perception in L2 requires a fairly well-defined L2 sound system. However, the L2 sound system may be quite incomplete or ill-defined due to insufficient language experience or wrong category formation.

Several current models of second language acquisition and cross-linguistic speech perception that will be discussed in section 2.4 base their predictions on similarities between native and non-native speech sounds (e.g. Flege 1995; Best 1995; Best & Tyler 2007). Similarity is considered to be a key issue in predicting the learners' success in distinguishing segmental contrasts by the *perceived similarity* of L1 sounds with L2 sounds. The difficulties learners encounter in L2, it is argued, are conditioned by similarities to L1 sounds.

In terms of the Flege's (1995) Speech Learning Model (see section 2.4.2), based on the distinction of "new" vs. "similar" sounds, similarity is considered to cause *equivalence classification* that may prevent learners to establish L2 categories adequately. In terms of the Perceptual Assimilation Model (Best 1993, 1995; Best & Tyler 2007), a high degree of similarity leads to patterns of assimilation to L1 sounds and poor discrimination (see section 2.4.3).

Generally, studies dealing with L2 sound perception usually consider the *inter-lingual* similarity of the phonemic systems of L1 and L2 in perceiving, processing and mentally organizing L2 sounds, whereas intra-lingual similarity and confusion between categories within L2 receive insufficient attention. A fact that has often been neglected but is considered to be crucial in the present study is the occurrence of confusion and equivalence classifications between sounds *within* the target language system, i.e. *among categories* of L2. L2 categories, which are articulatorily or acoustically similar or show orthographic similarities, are often subject to category confusion and substitution processes. The role of these *intra-lingual similarities* and their effect on perception, pronunciation and the acquisition of morphological and lexical items in L2, however, deserves more attention in L2 research to account for difficulties and category confusion with beginners as well as advanded learners. While equivalence classifications of L2 sounds with sounds from L1 are more common with beginners, *intra*-lingual confusion between L2 categories may persist even in more advanced stages of acquisition as the empirical data of the present study will show.

1.1.3 Markedness

Though it is commonly accepted that the emerging phonological system in interlanguages is characterized to a large extent by L1 phonological structures, transfer does not seem to be the only force influencing the acquisition process but appears to interact with other factors. An important issue in L2 research concerns the observation that not all L1 structures equally lend themselves to transfer into the target language and that not all L2 structures are equally "easy" or "difficult" in L2 acquisition.

An important shift of interest in the research history of *transfer* is the search for principles that govern the *transferability* of linguistic structures instead of simply documenting what is actually transferred in L2 acquisition (see phase 2 in Figure 1.1).

Kellerman (1983) proposed two major constraints on transferability: (1) the *psychotypology* constraint that leads L2 learners to transfer structures only if they are considered to be similar in L1 and L2, and (2) the *transferability* constraint stating that structures that are perceived as marked by the L2 learners are less likely to transfer.

A related argument is presented in Andersen's (1983) principle of *transfer to somewhere* that essentially states that structures compatible with natural acquisitional principles are more likely to be transferred to the target language.

A concept closely related to the issue of transferability and learnability in L2 acquisition is the concept of *markedness* (Eckman 1977; 2008; White 1987; Major 1987a; 2008; Major & Kim 1999). Markedness has been used in many linguistic fields to account for the relative difficulty of specific sounds in language processing, language acquisition and language disorder or for their resistance to sound change. In phonological studies and studies in second language acquisition, *phonetic ease and difficulty* have often been associated with notions of (un)*markedness*, together with (dis)*preference* (Vennemann 1983), and (un)*naturalness* (as in Natural Phonology, see section 1.2.4).

In very general terms, markedness refers to patterns of occurrence and the likelihood of occurrence of linguistic structures (Major 2008: 78). The term has been used for phonological phenomena as well as for morphological, syntactic and lexical structures. However, since the term "markedness" was first proposed by Nicholas Trubetzkoy (1939) and Roman Jakobson (1941), the notion has been used in several different theoretical frameworks such as structuralism, generative phonology, Chomskyan principle-and-parameter-syntax, Optimality Theory, first and second language acquisition, creole studies and others. Thus, the notion "markedness" has developed a considerable range of diverging senses in the various theoretical frameworks. Some authors (Moravcsik & Wirth 1986; Battistella 1990, 1996; Andersen 2001) make attempts to comprise the whole set of different senses of the notion, while others use it as a rather theory-neutral everyday term. Nevertheless, some common core meaning or intuitive basic sense of the term seems to exist.

Definitions of markedness refer to several linguistic facts such as *complexity* (marked sounds are more complex, i.e. have additional features), *difficulty* (difficult sounds are dispreferred), statistical *frequency* (more frequent sounds are less marked), typological *implications* (the presence of the marked implies the presence of the unmarked in a system), and *order of acquisition* (less marked structures are acquired earlier by children in L1 acquisition).

In its original sense, markedness referred to the relation between the members of a phonological opposition, in which one member having an additional specification or feature is more marked than the other (Trubetzkoy 1939). Roman Jakobson adopted the notion of markedness to lexical and grammatical issues. He described unmarkedness as lack of specification, in other words, the difference between marked and unmarked structures appears not to be a difference between A and non-A, but between A and *indifference* between A and non-A. Jakobson (1941 [1968]) associated differences in phonetic complexity and contrastivity with asymmetries in phonological oppositions and markedness as property of sounds with phonetic difficulty. Additionally, he pointed to the fact that markedness is reflected in typological patterns as well as in order of acquisition and sound change (for a detailed historical review, see Andersen 1989; Battistella 1990, 1996). According to Jakobson (1941, 1971), unmarked structures demonstrate a “first in, last out”-behaviour: they are acquired earlier and easier in L1 acquisition and are more stable in case of language loss (e.g. in aphasia). Oppositions which are relatively rare in the languages of the world usually also are acquired later in first language acquisition.

Although, crosslinguistically, phonological structures comprise an abundant variety, some phonological structures are rare or even not present at all in the languages of the world. Greenberg’s (1966, 1978) approach, referred to as *typological markedness*, focused on implicational relationships between two members of a phonological opposition and on their cross-linguistic distribution and relative frequency of occurrence. In this approach, markedness refers to a *multidimensional correlation* of several phenomena. Unmarked structures are more *frequent* than marked ones, but frequency of occurrence is considered to be only one influential factor among others: *Implicational hierarchies* predict that the presence of a more marked sound in a language system implies the existence of a less marked sound in the same system; *neutralization* of contrasts usually results in preservation of unmarked structures; unmarked values are preferred in inventories for *phonemic differentiation* (e.g. more non-nasal than nasal vowels in a language); unmarked structures show greater *allophonic variation* than marked ones (Croft 1999; Haspelmath 2006).

From a *typological* point of view, three types of relations between languages can be described: (1) languages with none of the members of an opposition, (2) languages with both members of the opposition, and (3) languages with only one, namely the unmarked element of the opposition.

Chomsky & Halle (1968) appealed to the terms "marked" and "unmarked" to refer to phonemes and the presence or absence of a certain feature, e.g. [+/– nasal]. In Chomsky & Halle's (1968) chapter 9, the "intrinsic content" of features is considered. SPE's binary feature system (+/–) is not directly compatible with a notion of "markedness" since it allows no third value "unmarked". Therefore, Chomsky & Halle proposed new values "marked" (*m*) and "unmarked" (*u*) to refer to some notion of neutrality, naturalness or expectedness. Chomsky's principle-and-parameter approach (Chomsky 1981) refers to markedness considering that possible parameter settings are not equal and that the unmarked parameter value functions as a default case, while the marked value is chosen if evidence requires this choice. Some feature combinations like e.g. *[+nasal, -voice] have to be classified as "unnatural".

A way of restricting the possible overgeneration of "unnatural" rules, processes or constraints in formal grammars is *functional grounding* (Archangeli & Pulleyblank 1994). *Constraint*-based formalisms over ordered rules aim at a more nuanced and richer conception of markedness to explain why some elements are more "natural" than others in certain positions.

In recent works on markedness, a *listener*-oriented and perception-based view has been receiving more interest connected with constructs such as "salience", "contrast", "robustness", "enhancement", similarity and others, arguing for the grounding of phonology on speech *perception* (for a collection of papers that link phonetic research with phonological description, see Hume & Johnson 2001a).

The precise relation of phonological knowledge, markedness amd phonetic *difficulty* has been an issue of ample discussion in phonological literature. While some approaches consider markedness as based on physiological, acoustic and perceptual factors (Ohala 1981, 1983, 1993b; Andersen 2001; Blevins 2004), others regard the *knowledge* about phonetic difficulty and markedness as part of linguistic knowledge (e.g. Archangeli & Pulleyblank 1994; Hayes, Kirchner & Steriade 2004). Hayes & Steriade (2004) posit that knowledge of markedness is considered to derive from phonetic knowledge and that this knowledge leads learners to postulate independently similar constraints that act as a source of systematic similarities among grammars. Phonetic knowledge is defined by the speakers' partial understanding of the physical conditions under which speech is produced and perceived and is conceived as the source for markedness constraints (referring to OT, see section 1.2.3) as components of grammar.

In the present study, the notion of phonological markedness will refer primarily to phonetic and phonological *complexity* and its articulatory and acoustic correlates. Markedness is assumed to be *phonetically grounded* and manifested in production as well as perception, in cross-linguistic distribution in the languages of the world and *frequency* of occurrence in a specific language system such as a learner's native language or an interlanguage.

1.1.4 Universals

The notion of markedness is closely related to the theoretical construct of language *universals*, referring to general preferences for certain linguistic structures. Both constructs, markedness and universals, have their origins in the work of the Prague School of Linguistics (Trubetzkoy 1939; Jakobson 1941) but have been used in other frameworks such as *Universal Grammar* (Chomsky 1981) or in the study of *typological universals* (Greenberg 1966; 1978; Croft 1990/2003, Maddieson 1999). A recurrent issue in research on L1 and L2 acquisition refers to the question whether typologically recurring structures and implicational conditions acquisition reflect universal principles and universal or "natural" sequences in phonological acquisition and speech processing. A number of researchers agree that there are some universal forces at action. However, the nature and source of these universals is a matter of intense debate.

Despite the immense variety of linguistic structures in the world's languages there are certain core properties which languages have in common. The description of a considerable number of languages does not only reveal diversity and variation but also constraints to this variation. These constraints, demonstrating that languages do not vary infinitely, may be considered to represent linguistic universals (Croft 2003: 5).

Universals are not to be discovered or verified by observing a single language. Studies on *typological universals* infer universals directly from comparing descriptions of several different languages. Their aim is to

> "*formulate cross-linguistic and, if possible, universally valid empirical generalizations about language structure (...) which hold true for some significant universe of languages and which at the same time are capable of being refuted by actual language data*" (Greenberg 1978: vi).

While some properties seem to imply the existence of others, there are also some absolute universals that hold true for all languages, e.g. (a) all languages have consonants and vowels, or (b) all languages make a distinction between nouns and verbs. There are other properties of languages which cannot be regarded as essential for all languages, even if they have a high degree of probability and therefore represent significant *tendencies* such as e.g. (c) most languages have a vowel /i/, or (d) most languages have adjectives (Whaley 1997: 32).

An *implicational universal* has a precondition, e.g. “if a language has nasal vowels, then it also has oral vowels” (for discussion, see Croft 2003: 57). Implications are unidirectional.

One of the fundamental questions concerns the nature of universals as being innate or triggered by language experience. Referring to L2, a crucial issue in second language acquisition is the modelling of the interplay of universals, markedness and the learners’ language experience in L1 and L2 to account for the learners’ performance nad rate of L2 acquisition.

1.1.5 Difficulty

The variable success of learners in L2 pronunciation and perception is generally associated with the relative *difficulty* of specific speech sounds.

While articulatory difficulty can be regarded as a property of an individual sound (in a particular context), the perception of single speech sounds cannot be attributed to be difficult to perceive as such. Rather, perceptual difficulty has to be defined in terms of the contrasts a sound enters into, i.e. the difficulty of a specific speech sound lies in the correct categorization onto the set of contrasting categories of a given language.

Perceptual difficulty is postulated to be related with *perceptual markedness*, which depends on the *salience* of specific features and the *stability* of contrasts the sound is involved in (Lindblom 1986; Flemming 2004). The concept of perceptual markedness is used to account for the fact that some L2 vowels due to their reduced salience and distinctivity from other contrasting sounds are more difficult to categorize than other less marked, more salient and more distinctive ones. Less confusable and more stable unmarked contrasts are preferred over more confusable marked ones.

From this notion of perceptual markedness, perceptibility scales and *implicational perceptual hierarchies* can be deduced. Along an implicational hierarchy, the phonetic salience of sounds involved decreases with consequences on phonemic contrastivity, the preference and frequency of cross-language occurrence decreases and the perceptual difficulty increases, especially if the marked element is not present in L2.

Implicational hierarchies in perception can be described

- as steps down in phonetic salience
- as steps down in preference and frequency of cross-language occurrence
- as steps up in difficulty in speech perception.

However, as the discussion of phonetic facts and experimental results of the present study will show, the concept of "perceptual markedness" as well as the notion of *salience* refers to characteristics of a given sound that have to be differentiated on more than one descriptive level (for discussion, see chapter 4 and 12).

In the description of L2 acquisition, markedness and implicational conditions may contribute to predict the learners' rate and order of acquisition for specific speech sounds. Perceptually more marked sounds are less salient in contrast to similar sounds, they will occur later and less frequently in interlanguages and will therefore cause more difficulties in perception and categorization even with advanced learners, especially if a corresponding sound is not present in L1. The *similarity* of L2 sounds as perceived by an L2 learner from a given L1 is a crucial factor to determine the relative difficulty of a sound (see section 2.4). A detailed investigation of articulatory, acoustic and perceptual aspects of similarity of vowels in L2 is the major aim of the present study.

1.2 Language acquisition, markedness and universals in phonological theory

Several models and theories assume that interlanguages – just as all natural languages – are governed by universal principles. Depending on the theoretical perspective, the role of universals in second language acquisition can be interpreted in many different ways (cf. Major 2001: 84): (a) a general linguistic learning perspective would describe overgeneralization or hypercorrection as *universal learning principles*, (b) a *parameter-based view* would consider universal parameters as being set, deset, or reset, (c) Generative Phonology or Natural Phonology will describe *rules and processes* appearing at the surface and the ordering and reordering of rules and processes, (d) Markedness Theory will describe the order of acquisition of linguistic structures as reflecting their universal status of *markedness*, (e) Optimality Theory will describe the re-ranking or un-ranking of L1 *constraints* aiming at the L2-specific set and *ranking* of constraints.

The following section will briefly discuss a few theoretical approaches to phonology and their basic assumptions for universals and foreign language acquisition.

1.2.1 Language typology and universals

Typological studies are based on the observation and classification of linguistic phenomena, followed by a typological generalization of observed structures. The functional-typological approach, mostly associated with Joseph Greenberg, aims at the construction of explanations of the observed structures and deduced generalizations. Explanatory models refer to competing motivations, economy, iconicity, processing, semantic maps in conceptual space, and several others (Croft 2003: 2f). For example, the universal "all languages have vowels" can be attributed to the fact that it is impossible or at least very difficult to articulate speech without vowels.

The generative approach proposed by Noam Chomsky posited that universals are innate and determined by generative grammar and its underlying principles and properties, while more empiricist-typological approaches posit more inductive generalizations and propose explanations in terms of more general cognitive, social-interactional, processing, perceptual or other abilities. However, in both approaches, universals are considered to rest in human abilities, with probably an innate component (even if not entirely innate), and to be ultimately biological, founded in genetics with Chomskyan innate linguistic knowledge and in evolutionary theory with Greenberg (Croft 2003: 5f). Both approaches describe universals and markedness conditions to hold for all languages, and therefore, are expected to hold also for interlanguages.

1.2.2 Universal Grammar

Noam Chomsky argued that underlying unity of linguistic structures has its origin in human biology. In his view, humans are genetically equipped with some kind of "language faculty" that enables us to acquire complex linguistic structures in an astoundingly short time span in early childhood. In this approach, an innate Universal Grammar (UG) is assumed to underly the grammars of all individual languages. It is supposed to structure the process of language acquisition and its representation in the human brain. Universal Grammar is defined as a system of principles, conditions, and rules that are elements or properties of all human languages. Chomsky assumes UG as an innate language faculty, which enables a child to learn a language in spite of the qualitatively and quantitatively insufficient input it is exposed to. L1 acquisition is considered to be conditioned by interaction of L1 input and UG. Within a generativist perspective of language learning, it is assumed that L2 learning is also guided by innate principles of UG and that due to these principles all languages show similar basic structures.

The basic question for second language acquisition discussed in several works on UG is whether the innate language faculty is disposable for the acquisition of a second language. Within this framework, the lack of success in mastering the pronunciation and intonation patterns of L2 could be taken as evidence that UG does not operate in second language acquisition. The issue of the initial state of language acquisition and the role of the L1 in interaction with UG is discussed quite controversially: While some researchers assume that L2 learners in SLA have no access at all to UG (Bley-Vroman´s Fundamental Difference Hypothesis, see Bley-Vroman 1991, 2009) or that access to UG is given only partially (White´s Transfer Hypothesis, see White 1990, 2000, 2003), other researchers hypothesize full access to Universal Grammar but differ in their assumptions on whether SLA will develop without transfer (Epstein, Flynn & Martohardjono 1996) or whether transfer interacts with UG. The Full Transfer/Full Access Hypothesis (Schwartz & Sprouse 1996, 2000) proposes that L2 learners have full access (FA) to UG and that the initial state of L2 acquisition is the final state of L1 acquisition (full transfer, FT). They claim that the creation of new L2 sound representations and the adjustment of the L2 initial perception grammar must go through developmental stages which can be found in the development of L1 perception as well.

Of course, it is not necessary to assume that all similarities occurring in the languages of the world are due to innate universal principles. With regard to phonological acquisition, constraints may be universal without being innate (Lindblom 1990; Donegan 1993; Boersma 1998; Hayes 1999; Hayes & Steriade 2004). Some shared characteristics may be due to characteristics of the human speech organs while others could be explained by external factors and the environment in which language is acquired.

1.2.3 Optimality Theory

One of the more recent phonological theories incorporating universals and markedness is Optimality Theory (OT). The classical OT model was developed by Prince & Smolensky (1997, 2004; cf. also Kager 1999; Tesar & Smolensky 2000; McCarthy 2002; for a more functional approach with relevance for SLA, see Boersma 1998, 2000, 2007; Broselow, Chen & Wang1998; Hancin-Bhatt 2008). A central claim of OT is that all languages share a common set of innate, universal *constraints*. Languages differ with respect to the language-specific ranking of these constraints. Two types of constraints are distinguished: Faithfulness constraints and Markedness constraints. Faithfulness constraints match the input/underlying form with the output/fully structured surface form while markedness constraints ensure the well-formedness of the output. Possible conflicts of constraints are solved in a language-specific ordering of constraints, resulting in cross-language variation.

Children in L1 acquisition have to acquire the language-specific ranking of their first language. Higher ranked constraints exert a stronger effect on output forms than lower ranked constraints. The model entails that a number of inactive constraints still are present but not active in the language concerned. For second language acquisition, OT assumes that in the initial stages of L2 acquisition learners operate with the constraint rankings from their L1. Their task in L2 acquisition is to acquire the L2-specific constraint rankings, i.e. to perform a re-ranking of constraints. Intermediate stages are neither L1- nor L2-rankings but are the results of universal constraints. This re-ranking of constraints results in interlanguage forms that are neither L1-like nor L2-like (Hancin-Bhatt 2008). Learners will perform re-ranking as long as they perceive differences between target-language structures and their own performance.

In OT, markedness is encoded in universal hierarchies of markedness constraints. OT markedness constraints reflect empirically established cross-linguistic tendencies, the evidence for these constraints is purely typological. The description of individual languages relates language-specific facts to general explanations from language typology and about grammars of all languages (McCarthy 2002: 1). The re-ranking of constraints is expected to develop earlier with less marked than with more marked structures since acquisition proceeds from the least marked to the marked structures. For similar phenomena, re-ranking will be less frequent and the L1-ranking will be more persistent (Major 2008: 80).

1.2.4 Natural Phonology

Another linguistic framework that operates with the concepts of universals and markedness is Natural Phonology (NP), founded by Stampe (1969, 1973a, b) and further developed by Stampe (1987) and Donegan (Donegan & Stampe 1978, 1979; Donegan 1978, 1985, 1993, 1995, 2001, 2002a, b). The framework was revised and expanded to other linguistic components as morphology and text linguistics in Natural Linguistics by Dressler (1984, 1985, 1995, 1999, 2002) and applied to the study of second language acquisition (e.g. Dziubalska-Kołaczyk 1990; Dressler & Dziubalska-Kołaczyk 1994; Kerschhofer 1995, 1998a; Bogacka 2004, 2007).

NP describes language as "*natural reflection of the needs, capacities, and world of its users*" (Donegan & Stampe 1979: 127). Like OT, NP considers phonological production and perception to be based on conflicting but coexisting universal principles.

Naturalness, just as markedness, is a relatively gradual concept, not only based on frequency in the languages of the world and on intuitive plausibility, but rather in the sense of being "*cognitively simple, easily accessible (especially to children), elementary and therefore universally preferred, i. e. derivable from human nature, or unmarked/less marked*" (Dressler 1999b: 135).

Natural Linguistics is a theory based on *preferences* rather than absolute constraints (cf. Dressler 1990; Dziubalska-Kołaczyk 2001b). Preferences are derived from non-linguistic higher universal principles such as the *principle of least effort* or the *principle of cognitive economy* (Dressler 1985b, 1999; Dziubalska-Kołaczyk 2002: 28). These preferences have typological or language-specific consequences in the languages of the world, e.g. the absence of clusters in a language or the absence of front rounded vowels in a system. In this sense, preferences account for structures being more or less marked, where more preferred structure are less marked.

Conflicts between preferences such as the tendency towards perceptual clarity vs. articulatory ease may arise. These will be solved either according to universal higher-order preferences and universal principles or they may be solved locally within a given language-type or a given language (language-specifically), so that "*preferences in the use and acquisition of language become frozen in preferences of language structure*" (Dressler 1999a: 394).

In NP, every regular sound substitution such as assimilations, insertions, or deletions reflects the operation of one or more *natural phonological processes*. These processes (comparable to constraints in OT) are based on human phonetic abilities, i.e. they are phonetically motivated. Processes apply involuntarily and are "*noticeable to a speaker only when he confronts pronunciations to which it does not apply, as when he tries to imitate a foreign sound or sound sequence and finds that he has difficulty doing so.*" (Donegan 1978: 5)

Within a language, the phonemes as well as their phonetic realizations are the output of language-specific phonological processes. Prelexical phonological processes determine the phoneme inventory and the phonotactic constraints of a given language while postlexical processes form the phonetic output of phonemes. The same process may apply prelexically and postlexically.

Thus, the phonological system of a language is the output of all those phonological processes a language fails to suppress: *"The living phonology of a language, everything that determines the 'accent' of its speakers, is the collective, systematic manifestation of those natural phonological processes which the language fails to challenge."* (Donegan 1978: 1).

In first language acquisition, a child learns to inhibit some of the universal set of natural phonological processes in order to arrive at a language-specific phonology, i.e. a system that enables the listener to decode the speaker's intentions from the flow of speech and to encode the own phonological intentions.

In L2 acquisition, learners have to modify the L1-specific process inventory and find the appropriate combination of limitation and suppression of processes that are characteristic for the target language. A process that remains unsuppressed in the L1-phonology of an adult speaker represents a constraint on his ability to produce certain sounds or sound sequences. The speaker has failed to learn to pronounce its input. This explains why sounds that do not occur in a speaker's L1 are "hard to pronounce" and/or "difficult to perceive" and are subject to substitution when the adult speaker encounters them in a foreign language (Donegan 1978: 12).

According to NP, in the initial stages of second language acquisition, learners have a specific number of phonemes from their L1, while all other possible phonemes have been suppressed by fortition processes in production and perception. Fortitions *"... seem to limit the set of sounds that can be perceived by the hearer as intended-by-the-speaker."* (Donegan 1985: 26). L2 learners therefore must suppress part of their L1-fortition processes to be able to perceive new L2 sounds correctly. Otherwise they would fall through a kind of L1 filter (cf. Trubetzkoy 1939 [1958]: 47). The suppression of these fortition processes succeeds only gradually. Therefore a part of the pronunciations occurring in an interlanguage might result from perceptual substitutions caused by L1-fortition processes that limit the learner's inventory (Donegan 1985: 26).

Wrong substitutions (due to L1 processes) may occur, leading to an incorrect reconstruction of the original intention which means that L2 learners can be misled by paradigmatic and syntagmatic processes (mainly of their L1).

Processes affect speaker's perception of foreign sounds as well as their production. In L2, processes account for deviations from the L2 norm and the production of "similar" sounds from L1 as well as for misperceptions, i.e. wrong identifications of L2 sounds with other L1 or L2 sounds resulting in sound substitutions. The complex interactions of processes in production and processes in perception are particularly relevant to explain L2 speech acquisition.

In second language acquisition, in addition to conflicts between the perceptual and articulatory teleology of processes, conflicts between phonological processes from L1 and from L2 may arise. Processes from L1 may be transferred to L2, where their application is inappropriate; e.g. the process of final devoicing is transferred to English by German speaking learners, resulting in devoicing of voiced obstruents in English. The suppression of processes from L1 could be transferred to an L2, where this suppression is not applied. The latter case will cause less problems than the former one, since voiceless obstruents in final position are unmarked and “natural”.

To summarize, the acquisition of an L2 sound system is considered to depend on the status of natural phonological processes in L1 and L2. The model’s predictions for the relative degree of difficulty of vowel sounds and vowel contrasts are summarized in section 4.7.2.

process in L1	**process in L2**	**predictions for IL**
present	present	L1 process applies in L2 – positive transfer – correct structures Subtype: The process applies in L2 but under more or less restricted conditions. Learners will have to adapt their process applications to L2 conditions.
	suppressed	Negative transfer, L2 structures are more marked than in L1 – Learners have to suppress the process in L2 – relatively high degree of difficulty
suppressed	suppressed	Marked structure but positive transfer – no problems expected in L2 (however, in some cases/contexts/styles the unmarked structure might arise by process application)
	present	A natural process applies, from which unmarked structures result. Learners will find it relatively easy to neutralize restrictions from L1. Relatively few difficulties are expected in L2.
latent (“dormant”): no input structures in L1, therefore the process is not suppressed in L1, but remains latent.	no input structures in L2	No problems will be evident.
	input structures given in L2	The latent process will apply. No severe problems are expected.

Table 1.1: The status of natural processes in L1 and L2 and the relative difficulty in L2 acquisition and interlanguage

1.2.5 Functional Grounding

Functional grounding refers to the view that phonetic facts are linked with phonological grammar and that phonological entities or processes are in some way based on or determined and restricted by functional factors.

Formally grounded factors may refer to formal linguistic entities like features, constituents etc., while functionally grounded schemata refer to phonetic and psycholinguistic properties. Grounded structures reflect physical correlates of features involved. While functionally grounded constraints are considered to be induced by learners from functional properties of the ambient language, formally grounded factors are considered to be innate (Hayes 1999; Steriade 2001a, b).

In their *Functional Phonology* approach, Archangeli & Pulleyblank (1994) propose that phonological rules and representations are constrained by the interaction of formal conditions drawn from a limited universal pool and substantive conditions of a phonetically motivated nature.

Many researchers assume that functional grounding must be internal to grammar, i.e. formal entities and operations are directly constrained by functional factors (e.g. Chomsky & Halle 1968; Stampe 1973a, b; Donegan 1978; Donegan & Stampe 1979; Dressler 1990; Archangeli & Pulleyblank 1994). This sense of grammar-internal functional grounding has also been implemented in many studies in Optimality Theory (OT) (e.g. Prince & Smolensky 1993; Boersma 1998; Hayes 1999; Flemming 2002; Kirchner 1998, 2001; Steriade 2001b; 2003; Hayes, Kirchner & Steriade 2004).

Other researchers posit that there is no need to integrate functional factors into formal phonology and that it is simpler to describe functionally motivated tendencies to be outside of grammar and grounded in external basic principles of speech perception, language acquisition or language change (e.g. Ohala 1981, 1993a, b; Andersen 1981; Blevins 2004).

In OT, markedness constraints *may* reflect functionally motivated tendencies, i.e. phonological structures can be restricted by functional factors. It is necessary to state that OT does not inherently require that all constraints are based on functionally motivated tendencies, but it allows functionally motivated tendencies to exert an influence in grammars of natural languages (Smith 2008). In phonetically-driven works in OT, phonological constraints are justified in functional grounds to produce phonetically-driven formal accounts of various phonological phenomena (Hayes 1999). Phonological constraints are well motivated when they render speech easier to articulate or when they render contrasting forms easier to distinguish perceptually.

In many empirical studies investigating second language speech acquisition, a basically surface-oriented phonetic approach is predominant (e.g. Flege 1995; Best 1995; Kuhl & Iverson 1995), while more theoretical SLA studies focus on underlying structures such as features and phonemes, e.g. Hancin-Bhatt´s Feature Competition Model (Hancin-Bhatt 1994), or Brown´s Underspecification Theory and Feature Geometry (Brown 1998, 2000; Brown & Matthews 1997), claiming that phonological categories of L1 (features, phonemes) will constrain which L2 sounds will be perceived and produced correctly. In this sense, functional typology compares structures of L1 and L2 to predict or explain crosslinguistic influences, but a functional approach goes beyond the surface-orientied structural contrast of CA to identify more abstract patterns, principles, and constraints (cf. Saville-Troike 2012: 56ff).

However, especially in initial stages of SLA, learners solely have access to phonetic L2 surface structures (phones) and phonological structures from L1, while phonemic structures of L2 only have to develop. In this context, the question of the role of an innate learning system is of relevance (see section 2.2.1).

To conclude, the crucial problem for phonological description is to incorporate research results from phonetics into formal phonology, i.e. to integrate mere phonetic explanations into phonological descriptions of linguistic phenomena. In other words, the task for language learners as well as for SLA phonologists is to bridge the gap between surface phonetic structures and underlying phonological representations.

1.3 Language-related, learning-related and universal factors in second language acquistion

1.3.1 Developmental processes

A major flaw of the Contrastive Analysis Hypothesis is its prediction of L2 substitutions solely on the basis of transfer from L1. Several studies showed that there are some sound substitutions that do not fit into this concept of transfer and that other factors must be at work such as the presence of (universal) *developmental processes* in the phonological acquisition of L1 and L2. Developmental processes result in *developmental substitutions* that cannot be accounted for by simple sound transfer from L1. In L2 acquisition, these developmental processes are part of the interlanguage system but can neither be directly attributed to L1 nor to L2. Rather, they can be described as general learning strategies that assist learners to acquire the target system. Developmental processes generate interlanguage structures which are neither part of L1 nor L2 but often constitute a sort of intermediate stage of acquisition. Interestingly, learners with different language backgrounds often show the same or similar substitutions in the L2, and many of these substitutions are also found with children acquiring their first language.

Developmental substitutions which are not due to transfer from L1 clearly reveal the existence of other forces than transfer shaping interlanguage phonology. If these substitutions cannot be accounted for by direct transfer from L1, it seems reasonable to conclude that they are due to language universals and markedness constraints.

Eckman´s *Structural Conformity Hypothesis* (SCH) assumes that universal generalizations that hold for primary languages will also hold for interlanguages (Eckman 1991: 24). Assumptions of SCH are motivated by patterns in interlanguages, in which IL structures adhere to markedness principles, but are not an area of difference between L1 and L2.

The Similarity Differential Rate Hypothesis (Major & Kim 1996) incorporated the notions of similarity and dissimilarity into a hypothesis about the rate of second language phonological acquisition: The rate of acquisition for dissimilar phenomena will be faster than for similar phenomena. Although not stated explicitly in the hypothesis, effects of *markedness* as an intervening factor are accounted for in that the degree of markedness can increase or decrease the rate of acquisition (Major 2008).

1.3.2 Individual strategies and the active role of the learner

To account for the immense variability of learning rate and succes among L2 learners, another important aspect of language acquisition has to be taken into account: the active role of the language learners and their individual strategies and hypotheses concerning the phonological system of the target language.

Selinker (1972: 215ff) postulates five central processes in second language learning: (1) language transfer, (2) transfer-of-training, (3) strategies of second language learning, (4) strategies of second-language communication, and (5) overgeneralization of target language linguistic material; moreover, several other non-central processes are assumed that influence the surface from of interlanguage (IL) utterances such as spelling pronunciations, cognate pronunciations, holo-phrase learning or hypercorrection (1972: 220f). Selinker considers *strategies of L2 learning* items in IL performance that are a result of an identifiable approach of the learner to the material to be learned.

The term “strategy” has been used in different senses and extent in language learning research (cf. e.g. Stern 1975; Rubin 1975, 1987; Ellis 1986; Skehan 1989). Other notions with similar connotations such as “tactics”, “techniques” or “learning behaviours” have been used (for discussion, see e.g. Griffiths & Parr 2001; Griffiths 2004).

In a very broad sense, L2 *learning strategies* can be considered as specific actions, behaviours, steps, or techniques that learners use – often consciously – to improve their progress in apprehending, internalizing and using the L2 (Oxford 1990). Learning strategies can be defined as tools for active, self-directed involvement needed for developing L2 communicative ability (O'Malley & Chamot 1990).

Rubin (1987) names three types of language learning strategies, – (1) learning strategies, (2) communication strategies, and (3) social strategies – and further distinguishes cognitive and meta-cognitive learning strategies. While cognitive learning strategies contribute directly to language learning, such as clarification/verification, guessing/inductive inferencing, deductive reasoning, practice, memorization, and monitoring, meta-cognitive strategies are rather used to plan, organise or manage the learning process (Rubin 1987).

In the present study, dealing with the acquisition of L2 phonetics and phonology, the use of the term “strategy” will rather refer to specific cognitive operations, which are at work when learners acquire a phonological system in L2: Cross-linguistic influence in L2 acquisition (see section 1.1.1) refers to several types of such operations such as similarity judgements, overgeneralizations, simplifications, approximations, or hypercorrections that will be discussed in the language-specific analysis in chapter 10, the vowel-specific data analysis in chapter 11 and in terms of theoretical concepts in chapter 12.

Even though highly ideosyncratic, L2 learners’ individual strategies and hypotheses may be motivated by *universal preferences* for certain linguistic structures and may provide more insight into the phonological acquisition process in L2.

Individuals’ considerations and hypotheses about the auditive and perceptual *similarity* of certain speech sounds and the influence of the orthographic systems of L1 and L2 are of particular relevance for the present study. Relationships of *perceived similarity* between L2 categories are established by the individual learner and are not directly predictable from the speech signal. The operationalization of relations of perceived similarity between L2 categories is the central aim of the present study.

With respect to difficulties in L2 pronunciation, Kerschhofer (1995) showed that *articulatory approximations* are predominant in the production of German vowels by Polish learners. Some of these approximations were due to the insufficient command of articulatory gestures, featuring at the phonetic surface of the learners’ performance in L2, while other instances of wrong realizations in the learners’ L2 production were due to ill-defined phoneme categories in the learners’ cognitive representation that could be explained by the insufficient perceptual distinction of phonetic details and phonemic contrasts in L2 German. Most of these operations result from *individual considerations* and (temporary) *hypotheses* about the L2 sound system and are substantially influenced by the learners' linguistic experience in the native and the target language (and further foreign languages) as well as by the specific setting of language learning.

Another type of strategy refers to *hypercorrect* behaviour observed mainly with learners in more advanced stages of acquisition. Hypercorrection in the context of the present study are not understood in a sociolinguistic sense (cf. Labov 1966, 1972), but rather as a learner's reaction to previous errors, behaviours or hypotheses that proved to be wrong. Eckman & Iverson (2000: 213f) define hypercorrection of an L2 contrast or "*hypercontrast*" as a "*lexically inappropriate substitution of one target language phoneme for another in contradiction to the native language transfer pattern*". This phenomenon seems to constitute a kind of substitution reversal, such that the "new" L2 phoneme is used in contexts where the L1 phoneme would be appropriate (see also Janda & Auger 1992; Eckman, Iverson & Yung Song 2013). In the current study, this type of hypercorrect substitutions may also refer to substitutions, where one L2 phoneme *i*, e.g. /y:/ or <ü>, is used in contexts, where another L2 category *j* (e.g. /u:/ or <u>) would be appropriate. Such substitutions are driven by previously experienced difficulties with category *i* that learners try to avoid by overuse of *i* in contexts, where its use is inappropriate.

Hypercorrections are understood here as instantaneous, deliberate, intentional acts that are conditioned by the learner's previous experience with the target language, either in every-day life or in a classroom situation or in an experimental setting. They are symptomatic for the interlanguage systems "under construction" and for unstable mental representations of L2 categories due to insufficient language experience, misperception and misinterpretation of L2 sounds.

In the data of the present study, the phenomenon of hypercorrect perceptual substitutions is observed in several L2 sub-samples. It is particularly evident in substitutions of front non-rounded *i*- and *e*-vowels that are hypercorrectly substituted by front rounded *ü*- and *ö*-vowels or *u*- and *o*-vowels. These hypercorrect substitutions are observed e.g. with the group of native Polish learners (see section 10.9) or native English learners (see section 10.5), and is most conspicuous with Polish learners in initial stages of L2 German acquisition. In the sample of Polish beginners (and less frequently with Polish advanced learners), substitutions of *i*- and *e*-qualities with rounded qualities are observed for all front non-rounded vowels, particularly for /ɪ/ (for a comparison of beginners and advanced learners, see Table 10.58 and Table 10.59 and section 12.5). Partly, such hypercorrect responses may also be due to reactivity in an experimental setting and the participants awareness of the fact that they are tested in a research study (see section 6.5).

To summarize, the vowel confusion data from non-native listeners will show that vowel identification in L2 involves the activation of several cognitive, attentional and developmental factors that vary across individuals and different points in time and are therefore difficult to isolate even in experimental contexts.

1.3.3 Cross-linguistic influence, markedness and difficulty in L2

Several models and hypotheses were formulated regarding the relative ease or difficulty learners will show in the acquisition of specific L2 structures.

One problem with the Contrastive Analysis Hypothesis was its failure to predict degrees of difficulty or rate of acquisition or developmental sequences. The theoretical constructs of universals and markedness have proved to be useful to predict or explain the relative degree of difficulty and to describe developmental sequences in second language acquisition.

Eckman's (1977) Markedness Differential Hypothesis predicts areas of difficulty by the interaction of differences between L1 and L2 and markedness:

> *"(a) Those areas of the target language which differ from the native language and are more marked than the native language will be difficult;*
> *(b) The relative degree of difficulty of the areas of difference of target language which are more marked than the native language will correspond to the relative degree of markedness;*
> *(c) Those areas of the target language which are different from the native language, but are not more marked than the native language will not be difficult."* (Eckman 1977 [1987: 61])

Referring to the role of markedness, Eckman's (1991) *Structural Conformity Hypotheses* (SCH) postulates that interlanguages and primary languages are similar in at least one important respect: they both undergo the same set of universal generalizations. Learners will perform better on less marked structures relative to more marked structures since the "*universal generalizations that hold for primary languages hold also for Interlanguages*" (Eckman 1991: 24). SCH is particularly relevant for cases, where IL patterns do not arise from differences between L1 and L2 but adhere to markedness principles.

Eckman's (1977, 2008) *Markedness Differential Hypothesis* (MDH) incorporates the notion of typological markedness into a contrastive description of L2 phonological acquisition. The basic assumption is that in addition to a contrastive analysis of L1 and L2 the relative markedness of L2 sounds can account for differences in acquisition. Typologically marked sounds not present in L1 will cause more problems in L2 learning than unmarked sounds.

More marked forms will be more difficult relative to less marked structures. In this sense, MDH is a special case of Eckman's (1991) *Structural Conformity Hypothesis* (SCH): MDH describes cases in which universal generalizations hold for the IL and the structures for which the generalizations hold are ones in which the NL and TL differ (Eckman 2008).

The *Ontogeny Model* (Major 1987) and the *Ontogeny Phylogeny Model* (OPM, Major 2001) describe three major factors determining the development of the interlanguage: *L1*, *L2*, and *universals* (U) and their varying influence in different stages of acquisition. In general, it is stated that while transfer processes would decrease over time, universals will first increase and then decrease again. The fact that learners with a variety of language backgrounds often make the same mistakes in L2 and that these mistakes can also be observed with children acquiring their L1, is considered as evidence that these errors cannot be attributed to the influence of L1 but that they are due to universals (Major 2001: 3). While some of these universals may involve general cognitive processes (e.g. overgeneralization), others are specifically linguistic, including some core features of languages. Major's models assume access to universals including principles of UG (Universal Grammar) without explaining the extent of access to UG. Universals include the universal set of properties of the human language capacity (anatomical, functional, and processing properties typical for our species) and the resulting universal characteristics of languages. Only those universals that are not already part of L1 or L2, i.e. the "latent" universals, will first increase and then decrease (Major 2001: 83).

Major (1987a; 2001) proposed a hierarchical relationship between transfer from L1 and universals. The *Ontogeny Phylogeny Model* (OPM, Major 2001) posits explicit claims on the relationship of transfer, universals (and markedness), and similarity. OPM considers *transfer* as the major influence factor in the early stages of L2 acquisition; in later stages, transfer from L1 will decrease, L2 will increase and *universals* will increase and then decrease in more advanced stages. The relative proportions of transfer from L1 and universals depend on the status of the phenomenon under consideration as "normal" or "similar" or "marked". Compared to "normal" phenomena, L2 is considered to develop more slowly for similar and marked L2 phenomena. The influence of L1 is relatively more important for similar phenomena. With marked phenomena, universals are more important than L1. For marked structures, OPM predicts the chronological development of the interlanguage as follows: L2 increases slowly, L1 decreases and then decreases slowly, universals increase rapidly and then decrease slowly. Thus, for marked structures, the role of universals is much greater than the role of L1, compared to less-marked structures.

1.4 Conclusions

The correct *perception* of L2 phonemic contrasts is fundamental for the successful acquisition of phonological categories in the target language. As a consequence of insufficient perception, inadequate mental representations of phonological categories and misconceptions of L2 contrasts, and the internal structure categories and their phonetic properties may arise. Misperceptions of L2 sounds are conditioned by inter-lingual differences and similarities between L1 and L2 as well as by intra-lingual differences and similarities within the L2 system.

Difficulties in L2 perception or production may be due to *cross-linguistic influence (CLI)* of L1 and L2 that are not limited to mere transfer from L1 but also refer to L1 influence on conceptualization, processing and use of L2 structures including hypercorrection, overuse, underuse or avoidance of particular structures (Jarvis 2012). CLI may also refer to backlash effects on the listeners' native language and other previously of subsequently acquired languages.

Markedness has been claimed to affect developmental processes in L2 acquisition and the relative ease or difficulty of particular L2 structures. However, the construct of markedness differs considerably in different theoretical frameworks. In SLA research (cf. Eckman 1977; 2008; White 1987; Major 1987a; 2008; Major & Kim 1999), only some aspects of the wide range of meanings of markedness have been received (see section 1.1.3 and 1.1.4).

Principles of universal and/or typological markedness conditions may have an impact on success and rate of acquisition and may partly account for the fact that non-native phonemic contrasts differ in terms of *difficulty*. Moreover, it is expected that these conditions may influence the learners' *preferences* for specific categories as targets in perceptual substitutions. Therefore, research on the interaction of L1and L2 has to ask for the *phonetic grounding* of markedness patterns (see chapter 4), i.e. for the influence of intrinsic characteristics of speech sounds and the effect of physiological and cognitive properties of human language and speech capacity.

Characteristic phonetic differences between L1 and L2 sounds that are considered as "equivalent" may not be perceived accurately, particularly when L1 and L2 sounds are very similar. "*Equivalence classifications*" (Flege 1987, 1995) established between L1 sounds and "similar" L2 sounds may inhibit learners from establishing adequate mental representations for L2 sound categories, whereas "new" L2 categories will be easier established for less similar sounds (see section 1.1.1, 1.1.2 and 2.4.2).

Two types of *similarity* relationships have to be distinguished in L2 acquisition:

(1) inter-lingual similarities of phonological structures between categories of L1 and L2

(2) intra-lingual similarities between categories within the target language system.

Models of speech perception and phonological acquisition in L2 usually focus on inter-lingual similarities (see section 1.1 and 2.4), whereas *intra*-lingual relationships of categories within the L2 system have so far received insufficient attention in SLA research. Intra-lingual similarities in phonetic as well as in phonological or psychological terms must be considered as a crucial element for a model of phonological acquisition processes in L2.

Perceived intra-lingual similarities between L2 categories can account for category confusion, inadequate category formation and instability of category representations in L2. It is necessary to emphasize that it is not mere phonetic similarity but rather *psychological similarity* between categories as perceived by the learners that crucially determines the learners' interlanguage (see chapter 3). The interaction of intrinsic phonetic properties of vowels and psychological dimensions of perceived similarity between categories determining the L2 listeners' performance in a perceptual identification task is of central interest in this study.

The major aim of the present study is the theoretical and experimental discussion of *intra-lingual similarities* between L2 vowel categories and the interaction of language-specific, individual and more general or universal factors on the learners' sensitivity to L2 vowel contrasts and their mental representations.

To conclude, the *interlanguage* system, i.e. the learner's L2 system "under construction", is determined by at least three major factors:

(1) *crosslinguistic influence* of the native language and the target language,

(2) relationships of contrast and *similarity* as established by the learner, and

(3) patterns of universal and/or typological *preferences*.

2 Speech Perception in Language Processing and Language Acquisition

The previous chapter introduced fundamental notions of second language acquisition research and learning theory and provided a short review of some linguistic concepts related to first and second language acquisition. This chapter will give an outline of research on second language speech perception, it will discuss fundamental aspects of human speech processing (section 2.1) and of the development of speech perception in L1 and L2 (section 2.2). It will moreover provide an overview of common experimental methods in L2 perception studies (section 2.3) and compare the most influential models of L2 speech perception (section 2.4).

2.1 On the nature of human speech processing

2.1.1 Variation in the speech signal

The structure of natural speech signals and the relationships between their articulatory, acoustic and auditory features are extremely complex. The speech signal has three substantial characteristics: (1) it is *continuous*, i.e. the acoustic signal is a continuum which contains no single segments, (2) it is highly *variable*, i.e. there are hardly any stable phases in the signal, due to the articulators which are – just as the acoustic waves produced – constantly in motion during speech production, and (3) it is articulatorily and acoustically *complex*, i.e. the auditive interpretation of a phonetic input stimulus requires an integrative interpretation of several acoustic parameters that are associated with a particular phonetic feature.

Perceptually significant components cannot be readily isolated by filtering or by temporal segmentation of the speech string. The exact correspondence of acoustic parameters to auditory events is language-specific, context-dependent, and highly variable. The perceptual mapping of the signal onto phonemic categories is considered to be part of a person's *linguistic knowledge*, underlying the decoding process of the continuous speech signal into sound categories.

Due to the continuous movements of the articulators, local extrema (maxima and minima) in some acoustic parameters may occur in the speech signal at some points in time together with discontinuities and dislocations at other points in time. Local extrema may function as markers for specific elements. Quantal acoustic properties in the vicinity of these discontinuities or extrema help to define the features of these elements (Stevens 1997: 463). Experiments with synthetic speech signals with carefully controlled acoustic patterns allow for identifying and describing some of the acoustic cues that support the perception of particular phonetic categories based on listeners' judgments of speech sounds.

Despite the immense acoustic variability of speech signals, native listeners seem to have no problem with the correct phonetic interpretation of the acoustic input. *Acoustic variability* refers not only to differences along the temporal axis of speech production but also to the immense articulatory and acoustic heterogeneity among different realizations of the same sound intention. Even within a single phonetic context, realizations of the same phoneme may vary considerably according to speech rate, the speaker's gender and age, the speaking style, the speaker's intention, etc. In other words, the same phonological segment may be realized in several ways differing articulatorily and acoustically due to co-articulation. Speaker-specific differences are hypothesized to be reduced prior to identification of linguistic categories by processes of so-called speaker normalization (models proposed e.g. by Traunmüller 1981; Syrdal & Gopal 1986; Miller 1989; Nearey 1989).

The *segmentation* of the speech string or the articulatory continuum into smaller units is another task that has to be accomplished by the listener, but continuousness and intrinsic variability usually do not affect phonological distinctiveness for a competent listener in a particular language. Segmentation is only possible if the listener and the speaker share the same "code" to decode the message in the speech input, i.e. if the linguistic knowledge about the relevant elements in the phonological system and the phonetic features that serve to distinguish phonological contrasts is available to the speaker and the listener. Phonetic features serve to distinguish phonological contrasts in a given language. However, a differentiation in terms of distinctive features on the phonetic or auditory level of single segments is not sufficient for the interpretation of the complex speech signal. The dynamic patterns of formant movements along the time axis (e.g. CV or VC transitions) provide substantial information for the differentiation of vowels (e.g. Hillenbrand et al. 1995; Jenkins et al. 1994).

However, spoken language processing cannot be reduced to a stimulus-driven segmentation procedure. Language is processed in two different directions which are described as "bottom-up" and "top-down" processing. In *bottom-up processing*, received auditory information is converted into neural signals and phonetic feature information is extracted and further processed. Bottom-up processing of single segments usually goes along with *top-down processing* that provides information from higher-level linguistic entities such as words or phrases. In top-down processing, stored information about language and the world is used to interpret the sense conveyed by the speech string. The input signal is associated with previous knowledge and expectations about the informational content of the speaker's message which allows to compensate for insufficient or ambigue information in the input signal. Thus, top-down processing enables us to disambiguate the available speech input and to "understand" or "perceive" more information than is actually available in the acoustic signal.

2.1.2 Categorization of speech sounds

The speech signal comprises an enormous range of variability in articulatory and acoustic parameters. Changes in acoustic parameters occur as a consequence of movements of the articulators in a non-monotonic way. Although the range of variability is constrained to some degree by human psychophysical capacities, humans could not cope with the enormous amount and variety of speech sounds without the cognitive mechanism of *categorization*, i.e. without the capacity of dividing and classifying the input into classes or "categories" based on some common features as criteria for similarity.

Perceptual categorization implies (1) the ability to identify a wide range of different phones as being "the same", i.e. as belonging to the same category, irrespectively of some auditorily detectable inter-stimuli differences, and (2) the ability to distinguish exemplars of one category from those of other categories in spite of some properties they may share but which do not function as differentiating criteria.

Categories are mental constructs that link two levels of representation, a discrete level and a parametric level. Categories can be understood as *labels* associated with a probability distribution over the parametric phonetic space. To acquire these probability distributions, extensive experience with a language is necessary to elaborate and refine the relevant levels of representation (Pierrehumbert 2003: 119).

Phonological categories can be conceived as the end-products of three successive processing stages: (1) the extraction of acoustic cues, (2) the analog-to-digital transformation of acoustic cues into phonetic categories, and (3) the grouping of phonetic categories into language-specific phonological ones (Werker & Logan 1985; Samuel & Kat 1996; Serniclaes 2000). A phonological category in a given language can be defined as a density distribution over the parametric level. Learning the phonetic patterns of a language involves the learning of the language-specific probability distributions over the *parametric phonetic space*, which can be understood as a high-dimensional *cognitive map* on which a metric of proximity and similarity in terms of acoustic and articulatory properties is defined (Pierrehumbert 2003: 128). Basic assumptions of the "categorical view" implicitly guide the linguistic view on vowel identification in L2 acquisition and will be discussed in section 12.1.3 and 12.1.4.

2.1.3 Categorical vs. continuous speech perception

Investigating relations between acoustic parameters and articulatory parameters reveals regions in the vocal tract, where acoustic parameters stay relatively stable in spite of changes in articulation whereas adjacent regions exhibit discontinuity and abruptness (Stevens 1997: 463). The same type of relation can be observed between certain acoustic parameters and specific auditory responses to acoustic inputs. That is, gradual changes of acoustic parameters may correspond to abrupt changes in auditory responses.

Numerous experiments have described this phenomenon in human speech perception by testing listeners' judgments in identification tasks and discrimination tasks. In early experiments with synthetic stop consonants that varied in acoustically equal steps (e.g. Liberman et al. 1961; Abramson & Lisker 1970; Pisoni 1971), listeners showed no gradual transitions in identification in spite of the continuous quality of the input; rather, a clear and abrupt separation of categories depending on the phonemic significance of stimuli in L1 was observed. Similar experiments with synthetic vowels showed no such discontinuities between categories; the discrimination of phonetic differences between different vowel sounds proved to be generally better than between consonants, even within a phonemic category (e.g. Fry, Abramson, Eimas & Liberman 1962), whereas with consonants, acoustic differences were only perceived if the stimuli belonged to different phonetic categories (cf. Liberman et al. 1957; Pisoni 1971). These results demonstrate that phonetically different vowel stimuli are to a certain degree discriminated along an acoustic continuum even within a category.

The observed differences between vowel and consonant discrimination led researchers to distinguish two *modes of speech perception*: a categorical mode and a continuous mode of perception (Liberman 1957; Liberman et al. 1967a; Studdert-Kennedy 1973; Studdert-Kennedy, Shankweiler & Pisoni 1972; Wode 1990; for a historical review, see Liberman 1996).

In the *categorical perception mode*, the hearer can make only judgments about the "name" of a stimulus, but not about its particular sound quality. Categorical perception seems to be typical for some types of sounds, especially for plosives.

In a *continuous mode* of perception, the hearer is able to discriminate between two identical phonetic types or allophones, while in categorical perception only discrimination between sounds which are identified as belonging to different categories is possible.

For language acquisition both modes are essential. While categorical perception serves to develop basic categories, continuous perception enables hearers/speakers to acquire fine-grained details and to self-monitor their speech to adjust their speech production to the norms of the L2 speech community (Wode 1990, 1994a, b, c).

Categorical perception is typically characterized by (1) a clear category boundary, (2) a corresponding discrimination peak, and (3) the predictability of discrimination function from identification; the resistance to contextual effects and sequential dependencies could not be proved to be characteristic for categorical perception (Damper & Harnad 2000).

Continuous perception allows the detection of gradual differences between sounds within a given category. Several studies report that listeners can distinguish gradient phonetic variation within a vowel or consonant category and are able to rate phonetic variants within a native category as “good” or “poor” variants of a given category (e.g. Grieser & Kuhl 1989; Kuhl 1993, 1998; Miller 1994; Allen & Miller 2001). This has led to the assumption of “ideal” or *prototypical* representatives of a phonetic category (Kuhl 1991, 1993; Kuhl & Iverson 1995).

It is important to note that the distinction between the categorical and the continuous mode of perception mainly refers to *discrimination* abilities. However, the ability to *identify* speech sounds is equally important in speech perception.

In identification tasks with vowels, language-specific boundaries between vowel categories can be observed. In several experimental studies on categorical perception (e.g. Janota 1965; Lindner 1966; Scholes 1967; Sendlmeier 1985; Rochet 1995), adults discriminated phonetic units that crossed the “phonetic boundary” between categories but not stimuli that fell within a category. This effect showed to be language-specific.

Within the *Motor Theory of Speech Perception* (Liberman 1957; Liberman, Cooper, Shankweiler & Studdert-Kennedy 1967a, b; Mattingly & Studdert-Kennedy 1991; Liberman & Mattingly 1985), categorical perception has been assumed to result from the categorical nature of speech production. Speech perception is assumed to involve a process of gesture recovery, where gestures correspond to abstract speaking intentions of the speaker that are directly perceived by the listener/perceiver (Fowler 1986). Within this framework, the perception of consonants is generally supposed to be more categorical than that of steady-state vowels because the articulatory gestures for consonants are more discrete than the relatively continuous gestures for vowel production (Damper & Harnad 2000: 844).

Generally, in all sound dimensions more differences can be discriminated than identified correctly. While identifications along a single dimension are restricted, relative discrimination is possible between a much larger number of stimuli (Damper & Harnad 2000: 843). The ability to discriminate differences among stimuli within most auditory continua indicates that these continua are, in fact, perceived *continuously*. This holds especially for vowels.

However, vowel identification always involves an element of categorization. Massaro (1987) clearly distinguishes categorization from categorical perception:

> "*I cannot understand why categorization behavior was (and continues to be) interpreted as evidence for categorical perception. It is only natural that continuous perception should lead to sharp category boundaries along a stimulus continuum.*" (Massaro 1987: 115)

Fujisaki & Kawashima (1969; 1970) could demonstrate that categorical effects may also originate from memory limitations and the level of phonetic vs. auditory coding (cf. Pisoni 1971) and showed that differences in perception between vowels and consonants are partly due to the duration of acoustic parameters. Vowel stimuli of the same phonetic category with a very short duration of 40-50 ms (which corresponds to the duration of CV-transitions) are reported to be less well discriminated than equal longer stimuli (Sendlmeier 1985: 43).

To summarize, *categorical perception* refers to perceptual sensitivity for between- but not for within-category differences, while *continuous perception* refers to perceptual sensitivity for changes along an acoustic continuum.

Categorical perception refers to the ability to discriminate between-category but not to within-category differences along a stimulus continuum, while *continuous perception* refers to a relatively linear relationship between changes in a stimulus and changes in the perception of this stimulus.

So called *prototypes* represent very good or "typical" instances of a given phonetic category that can be clearly distinguished from other adjacent categories. Sound categories and category prototypes are *language-specific*. Speech perception in L2 is affected by the specific set of phonetic categories and prototypes of the listeners' L1.

Categorical perception of phonemes in a particular language arises from deactivating or rather reducing the initial sensitivity to phonetic distinctions irrelevant in the native language. In other words, the sensitivity to phonetic details that are not phonologically relevant is reduced during L1 acquisition. This process of "desensitization" already takes place in infancy and is induced by linguistic properties of the infant's ambient language (see section 2.1.4 and 12.5).

2.1.4 Language-specific speech perception

Due to the language-specific phonetic-acoustic correspondences that are part of the learners' linguistic knowledge acquired in L1, the same acoustic input signal may be interpreted differently by listeners from different languages as has been demonstrated already in early works on experimental psycho-phonetics (Janota 1965; Lindner 1966) and numerous subsequent studies (e.g. Scholes 1967; Rochet 1995).

Language-specific patterns of selective perception, which result in highly automatic patterns of perceptual processing, guarantee robust and efficient perception in L1, but are not easily modified by subsequent linguistic experience in L2 (Strange & Shafer 2008: 156). In L2, the ability to discriminate or to identify speech sounds is limited to a large degree by the linguistic knowledge available to the listener. As we will see in the empirical data presented in this study, the same input signal may be interpreted in very different ways by listeners with different language backgrounds, depending on the language-specific distribution of vowel categories and category boundaries in the L2 learners' L1 as well as by their individual construction of the L2 sound system to be acquired.

In foreign language acquisition, automatic selective perception patterns that are appropriate in the listeners' L1 may result in inadequate perception processes, which may delimit the correct discrimination or identification of L2 sounds and hinder the development of correct L2 sound categories and of native-like production patterns.

Especially for adults, so called "late learners", and for "naïve listeners" with no knowledge of L2 sound patterns, some phonological contrasts and phonetic details in L2 are very difficult to differentiate in perception. With some phonological structures, these difficulties may persist even after years of language experience in L2. This "perceptual accent" causes problems in L2 sound identification and word recognition; it moreover seems to be the reason for slower speech processing in L2 and reduced sensitivity for non-native phonemic contrasts under noise-conditions (Bürki-Cohen, Miller & Eimas 2001) and it helps to explain the incomplete acquisition of L2 articulatory gestures as an underlying reason for a non-native accent in L2 speech production (e.g. McAllister 1991; Mayo et al. 1997).

Language-specific perceptual and cognitive processes of sound perception develop during childhood and are induced through exposure to a child's ambient language. Research in infants' speech perception has shown that infants are born with high perceptual sensitivities for acoustic differences in the speech input, even if these differences are not relevant in the ambient language. Earlier works assumed therefore an innate universal set of phonetic feature detectors that enables infants to encode speech into linguistic units (Eimas et al. 1971). These were assumed to atrophy during development. However, it could be proved that perceptual abilities of infants are rather due to *auditory processing* than to innate linguistic structures. Instead of atrophy, it is exposure to the ambient language that changes and reorganizes the original perceptual abilities of infants (see section 2.2.1 below) in that a sort of reduction of perceptual sensitivity to within-category acoustic variation is acquired due to language exposure (Kuhl 2000; Iverson et al. 2001; Werker & Tees 2005). These perceptual changes due to early language experience in L1 can impede the acquisition of non-native phonemes by adults (Kuhl 2000). The degree of interference of the native language is related to the extent to which the phonological systems of L1 and L2 differ (Flege 1995; Best 1995; Iverson et al. 2001).

In the present study, it will be argued that the acquisition of a second language phonological system implies (1) a "re-sensitization" to specific phonetic details and continuous variation, which are relevant in the target language, and (2) an re-allocation of selective attention and attentional weight to specific phonetic parameters that are relevant in L2 (see section 2.2.1 and 12.5). An important prerequisite for the development of appropriate L2 categories and a native-like production is the learners' conscious perception of phonetic details that are relevant for phonological contrasts in L2. However, even if in discrimination tasks involving a comparison of stimuli, subjects may be able to *discriminate* L2 vowel sounds ("continuous mode"), their ability to *identify* and categorize L2 vowels correctly ("categorical mode") is expected to be less accurate, especially with universally marked vowel sounds and with distinctions that are not used in the L1 system (e.g. Frieda & Nozawa 2007). The current study will use a perceptual identification task involving forced-choice vowel categorization to demonstrate where problems in the correct identification of German vowels in L2 occur (see section 6.2).

2.1.5 Auditory, phonetic and phonemic processing in perception

Concerning speech perception, three separate processing factors or levels are commonly distinguished: (1) auditory, (2) phonetic, and (3) phonemic processing (Werker & Logan 1985; Strange 2011). However, to which extent these levels of processing are altered by language experience is quite uncertain.

It is generally assumed that *auditory processing* is unaffected by language exposure. Thus, auditory factors in speech perception are supposed to be "universal" in the sense that they do not show an effect of previous linguistic knowledge even with adults (Flege, Munro & Fox 1994; Strange 2011). This assumption is supported by the fact that adults can *discriminate* acoustic differences between non-native phonemes, even if they cannot categorize them correctly, nor would they realize them correctly in production. The ability to discriminate acoustic differences is stronger in non-linguistic than in linguistic input.

However, electro-physical studies with adults reporting the existence of language-dependent memory traces that are activated in early processing of speech (Näätänen et al. 1997; Winkler et al. 1999) support the assumption that language exposure and experience may also affect auditory processing. This evidence suggests that perceptual changes due to language experience do occur at an early phonetic or late auditory level and are prior to recognition or categorization of speech in terms of higher-level linguistic units. Phonological categorization is driven by phonetic processing and the phonemic representation of categories in a given language. Thus, language-specific low-level auditory processing interferes with higher-level linguistic representations. The relative salience of within-category and between-category acoustic differences is influenced by language-specific low-level perceptual processing (Iverson et al. 2001).

2.1.6 Phonetic traces and phonemic processing

Studies using brain-based methods (see section 2.3.7) show that speech perception in a given language is based on a set of language-specific *phonetic traces* (and their combinations in syllables and words), which develop for phonemes in the native language in the first few months of life providing recognition models or templates that are used by the central auditory system in the perception of these sounds (Näätänen 2001:7f). These traces are permanent or based on long-term memory. Exposure to a familiar sound of the native language activates the corresponding trace or recognition model, together with other mechanisms of sound-analysis. These speech sound recognition models could account for the categorical perception of phonemes if we assume that there is a (nearly) constant phoneme code carried by the trace, which is activated by different types (prototype and non-prototype) of the phoneme. The limited set of response alternatives available in this setting may explain the absence of continuous mapping of stimuli in the system (Näätänen 2001: 7f).

For a given speech sound, the individual *discrimination accuracy* is considered to depend on the informational sharpness of the *central sound representation* (CSR) on the auditory dimension involved (Näätänen 2001). CSRs underlie and correspond to perception, providing sensory information. A sound is considered to enter (if at all) into long-term memory in the form of a CSR, which represents the sound as an auditory event in time rather than as a static set of certain levels of different stimulus features. The accuracy of perception, sensory memory, recognition, and discrimination is determined by the central sound representation. The notion of *representational width* (*Rw*) of a sound represents a measure for the discriminability of a sound. The narrower the *Rw* of a sound is, the better its discriminability from a different sound/stimulus and the better the resolution ability of the system. The MMN amplitude is highly correlated with the accuracy of behavioral discrimination (Näätänen 2001: 3f).

Despite the large acoustic variation due to different voices and different word contexts, some complex acoustic invariance must exist that is shared by different phones, encoded in the phonetic trace, e.g. the formant frequencies and their ratios for vowels. Thus phonetic traces may act as recognition patterns or templates in speech perception. These phonetic traces also account for the fact that spoken language is not only perceived acoustically, just as any other complex sound, but categorically. While speech and non-speech sounds are mapped continuously into auditory perception, speech sounds additionally are pre-perceptually classified into different phonetic *categories c*orresponding to different phonemes and their combinations.

In early language acquisition, perception clearly leads production. Phonetic categorization and perceptual parsing develops before infants learn to speak and before they develop large vocabularies (Pierrehumbert 2003). Exposure to one's mother tongue causes changes in the brain in the form of language-specific memory traces and may additionally cause increased sensitivity of the afferent mechanisms for the specific type of acoustic variation present in the input (Näätänen 2001: 12f). Traces may be (partially) activated also by sounds nearly matching with phoneme-specific invariant *codes* (categorical effect) but their number is limited to the number of different phonemes an individual can recognize perceptually. Further, they may also provide the sensory information necessary for production and control of pronunciation. In learning to understand and speak a foreign language, these speech-sound traces have to be developed for phonemes, syllables and words.

2.2 Speech perception and language acquisition

2.2.1 The development of speech perception in L1 acquisition

The speech processing system of adults exhibits astounding levels of speed, accuracy and robustness in parsing complex phonetic patterns. However, the perception of sound contrasts deviating phonetically and/or phonologically from contrasts in the listeners' native language appears to be constrained systematically by the listeners' language experience in their L1. This effect becomes most apparent with adult learners in L2 and especially with "naïve listeners" but begins to emerge already in infancy and throughout early childhood (Werker & Tees 1984b).

In their first months of life, infants appear to be endued with amazing learning strategies, demonstrating pattern perception as well as statistical (probabilistic and distributional) computational skills, which are not predicted by classical theories of L1 acquisition (cf. Skinner 1957; Chomsky 1957). Neither Skinnerian reinforcement nor Chomskyan specification of innate linguistic knowledge is supported by more recent studies on the acquisition of speech in L1. Rather, infants in their first year of life seem to perceptually "map" critical aspects of their ambient language. By exposure to the ambient language, statistical properties of speech are picked up. That is, linguistic experience alters infants' perception of speech, warping perception towards the child's native language (Kuhl 2000b). With age and the particular input the infant is exposed to, these abilities begin to change and to redevelop. This universal phonetic capacity gets eventually lost and is not available any more for adults who demonstrate considerably more difficulties in non-native phonetic discrimination.

Studies with infants in their first few months of life have demonstrated that their ability to discriminate speech sounds is remarkably good at this early age even for features that are not present in the ambient language (e.g. Eimas et al. 1971; Werker & Tees 1984a, b; Best & McRoberts 2003; Kuhl et al. 2006). Moreover, young infants are able to discriminate even subtle differences "categorically" (Eimas et al. 1971; Eimas & Miller 1983; Eimas, Miller & Juszczyk 1987).

During the first year of life, considerable changes occur to infants' speech perception and production. Already at the age of 6-12 months language-general perception abilities redevelop, i.e. exposure to the ambient language shows its effect (Werker & Tees 1984a, b). Discrimination abilities for non-native contrasts deteriorate already at the age of 6-12 months (Best et al. 1988; Grieser & Kuhl 1989; Kuhl et al. 1992; Werker & Tees 1984a, b). At the end of the first year of life, native language speech perception skills show improvement (Kuhl et al. 2008). At the age of 10-12 months, the perceptual organization of speech sounds is focused almost exclusively on phonetic contrasts occurring in the ambient language. Simultaneously, the ability to discriminate other speech sounds, which are not relevant in the native language, decreases. These developmental changes in perceptual abilities during the first year of life are considered to occur due to exposure to the ambient language and are influenced or induced by the specific kind of linguistic input infants are exposed to. In other words, with increasing age, infants either lose the ability to perceive differences or learn to ignore those contrasts that are not relevant for the native language.

That is, before infants have developed a vocabulary, their perception of phonetic segments is affected by phonetic properties of the ambient language (Houston 2008). Presumably, word learning plays a substantial part in the further shaping of speech perception and the knowledge about linguistic contrastivity of speech sounds (Best et al. 2001; Houston 2008). The detection of the contrastive function of speech sounds during lexical learning thus fosters the development of language-specific phonetic and phonological categories.

Older children and adults clearly demonstrate language-specific patterns of perception so that we can assume that language-general patterns of perception given at a very young age develop into language-specific patterns. These research findings have led to the view that infants are born with a kind of universal speech discrimination system that is able to discriminate all of the world's phonetic contrasts and that adults have lost this "universal" phonetic capacity (e.g. Eimas et al. 1971), though Nittrouer (2001) warns against a too simplified interpretation of infants' speech perception (for critique see Aslin et al. 2002, for a replique see Nittrouer 2002).

Empirical studies on cross-language speech perception mostly use two kinds of populations, infants and adults, while the perception of older children is a less studied field (Walley 2007; 2008). Studies in L1 speech perception mostly finish between 12 and 24 months. Only few studies about speech perception in childhood have been undertaken. The available data suggests that the way a native language is perceived may change in the course of childhood and does not achieve adult levels before the age of 12 (Hazan & Barrett 2000; Walley & Flege 1999). While younger children (about 5 years of age) are claimed to classify speech rather on similarity relations, older listeners would use phonemic identity. Generally, children's representations appear not to be as fine-grained or "segmental" as adults' representations but appear to be more "holistic" (Walley 2008: 452f). Younger children seem to rely more on dynamic cues for phonetic decisions and seem to focus more on the recovery of syllabic structure. Phonemic category boundaries in perception tasks with younger and older children seem to depend to some extent on lexical familiarity, i.e. category identification in familiar words showed to be much better. With lexical growth, more attention to details of the acoustic signal is required (Walley 2008; Walley & Metsala 1992). Moreover, the influence of developing reading and writing skills and the establishment of phoneme-grapheme-correspondences has to be considered as influential factor in the development of L1 phonemic categories (Kerschhofer-Puhalo 2010b, 2012, in press a).

The study by Walley & Flege (1999) with 5- and 9-year old children and adults with a native and a non-native vowel continuum showed no significant effect of age in the location of phoneme boundaries. This result is consistent with other studies (e.g. Kuhl et al. 1992) that assume that the vowel space is shaped very early in infancy. Still, infants' discrimination abilities at the end of the first year of life are not adult-like (Houston 2008). Younger childrens' perception is not as fine-grained or segmental as that of older listeners. Moreover, L1 perceptual categories continue to develop in early and later childhood (Walley 1993, 2008; Baker et al. 2002). This may explain why Walley & Flege (1999) observed no marked age differences in the location of phoneme boundaries in L1 and L2 continua (for other studies about phoneme boundaries in late childhood and adolescence, see e.g. Flege & Eefting 1986; Hazan & Barrett 2000; Nittrouer & Miller 1997).

These effects have been demonstrated particularly with consonants (Eimas et al. 1971; Werker & Tees 1983, 1984a, b, 1992, 1999; Jusczyk 2000). As to the perception of vowel contrasts, developmental changes are less well documented (but see Kuhl et al. 1992; Polka & Werker 1994). While consonant discrimination with infants seems to be categorical (Eimas & Miller 1980b), the discrimination of vowels by infants differs somewhat from the discrimination of consonants. Infants, like adults, seem to be able to discriminate not only distinctive vowel phonemes but also vowel sounds of the same category that differed in their formant frequencies; thus they seem to perceive vowels in a continuous manner (Houston 2008: 418).

2.2.2 Perception and foreign accent in L2 acquisition

Several studies link problems in L2 production with difficulties in perception, assuming that the inadequate perception of L2 speech sounds explains problems learners have when learning L2 sounds and explaining problems of inaccurate production (or "accentedness") by the inadequate perception of L2 contrasts that is influenced by the learners' L1.

The *Motor Theory of Speech Perception* (Liberman 1957; Liberman, Cooper, Shankweiler & Studdert-Kennedy, 1967a, b; Mattingly & Werker 1991; Liberman & Mattingly 1985) claims that perceptual entities in general are not acoustic or auditory events but refer to articulatory gestures. A revised version of the theory by Liberman & Mattingly (1985) proposes a biologically distinct system for phonetic information, a module which is specialized to detect the intended gestures of the speaker as a basis for phonetic categories. By establishing a lawful relationship between gestures and acoustic patterns, the module causes perception of phonetic structure without translation from preliminary auditory impressions (Liberman & Mattingly 1985).

A large number of studies shwoed language-specific effects in perception. Rochet's (1995) study with a synthetic continuum of French vowels showed cross-language differences in classification along the continuum by French, English and Portuguese subjects. Vowel stimuli that were classified as /y/ by French listeners were perceived as /i/ by Portuguese and as /u/ by English participants. In production, the participants showed similar patterns as in perception. Numerous studies have followed which show that the learners' language background has a strong effect on the perception of L2 sounds (e.g. Scholes 1967; Flege 1987; Flege & Hillenbrand 1984; Best et al. 1996; Levy & Strange 2008).

To account for learners' foreign accent in L2, several explanations have been proposed. One of the most accepted explanations are inadequately established *phonetic habits*, mostly influenced from L1 (Weinreich 1953; Lado 1957), together with "reduced neural *plasticity*" due to neurological maturation (Lenneberg 1967; Penfield & Roberts 1959) to account for the incomplete or inadequate adaptation of the sensomotoric programs for L2 sound production. This view is closely associated with the notion of a *"critical period"*, a construct to explain why the acquisition of a native-like accent is extremely rare with so-called "late learners" (Lenneberg 1967; Scovel 1988, 2000). A considerable number of other explanatory factors such as inadequate input, wrong habits acquired in early stages of learning, insufficient motivation, social identity and several others have been proposed to account for learners' failure to acquire a native-like accent (Flege, Munro & MacKay 1995; Bohn 1998; Piske, MacKay & Flege 2001).

However, the *inadequate perception* of phonetic information in L2 seems to be one of the best candidates to explain many difficulties in the acquisition of L2 pronunciation patterns. Phonetic segments, which are phonologically distinctive in L2 but not in L1, are often not correctly discriminated and categorized in perception and therefore not adequately pronounced. Moreover, phonetic differences between L2 sounds and corresponding L1 sounds may not be recognized in perception due to a certain degree of similarity of L2 sounds to L2 categories, resulting in an inadequate reproduction of L2 sounds.

Several models and hypotheses describing constraints on L2 perception have appeared in the literature, all of which assume that the perception of L2 phonetic segments is influenced by the L1 phonological system. Trubetzkoy (1939 [1958]: 47) compared the L1 phonological system to a system of "sieves" filtering the relevant features of the speech input.

Current *models of perception and production in L2* such as Flege's Speech Learning Model (Flege 1995), Best's Perceptual Assimilation Model (Best 1995; Best/Tyler 2007), and Kuhl's

Native Language Magnet Model (Kuhl 1995; Kuhl & Iverson 1995) suggest that perceptual accuracy limits L2 production.

Flege (1988, 1995) hypothesized that the production of an L2 phonetic segment will typically be no more native-like than its perceptual representation and that in early stages of acquisition, L2 production will be less native-like. To achieve a native-like level in L2 production, learners have to detect the phonetic features for L2 phonemic contrasts in the acoustic signal and to establish new phonetic categories for L2 in long-term memory. Thus, the correct perception of L2 categories is considered to be a substantial prerequisite for appropriate speech production in L2. Moreover, learners have to link perceptual categories with articulatory gestures in order to acquire a native-like accent. The inadequate perception of phonetic differences between phonologically identical segments in L1 and L2 is considered to hinder the acquisition of a native-like accent in L2.

To summarize so far, the perception of sound categories is not universal but language-specific. It is influenced by the phonological system of a listener's L1. Learning a first language basically involves the development of selective perceptual processes. These are acquired during the first years of life, they apply automatically and subconsciously (Linell 1982) and have a persistent influence on every other language a human being subsequently acquires.

To conclude, three components seem to be substantial for the successful acquisition of L2 sound patterns: (1) the accurate *perceptual discrimination* of the relevant properties that distinguish L2 sounds from each other as well as from L1 sounds, (2) the *storing and structuring* of relevant information in long-term memory (phonetic *categories*), and (3) the learning of *articulatory gestures* to reproduce L2 sounds adequately (Flege 1995: 236). These components will be discussed in more details in the following sections.

2.3 Empirical research on L2 perception

Differences between the phonological systems of the listeners' native language L1 and the target language L2 may lead to different interpretations of the same acoustic signal. Sebastián-Gallés (2008:547) describes different types of "perceptual illusions" in everyday-terms such as "deafness" (listeners cannot hear differences due to ignorance of contrasting information in the signal), "mutations" (listeners change one sound into another), and "mirage" (listeners create information not present in the input signal, e.g. epenthetic vowels). Experimental designs have to account for these types of possible problems in L2 perception.

Until the late 1950s and early 1960s, relatively few empirical studies in L2 speech perception and production were carried out (with a few exceptions, e.g. Polivanov 1931; Fries 1945; Weinreich 1953; Lado 1957). Thereafter, researchers' interest in the acquisition of L2 sounds grew considerably. Critiques to the Contrastive Analysis Paradigm soon revealed that a mere structuralistic description and comparison of L1 and L2 phoneme inventories could not fully explain the difficulties in L2 production and the varying success among learners but that explanations had to consider the individual L2 learners themselves (Flege 1988b).

From the 1970s on, research interest began to shift more towards perceptual phenomena underlying the phonological acquisition of L2. Empirical studies focussed on different aspects of difficulties in sound *discrimination* and on *categorical perception* (for a historical overview, see Strange 1995b). Empirical insights into L1 acquisition, especially the finding of categorical perception of speech sounds by infants and their ability to distinguish segmental contrasts that are not present in the ambient language in their first months of life (Eimas et al. 1971) intensified researchers' interest in the perception of non-native speech perception. Subsequently, several streams of investigation with specific approaches and research methods developed.

A substantial part of research focused on adults' perceptions of L2 phonetic distinctions that do not occur in L1. Another important field of research was the role of *language experience* on perceptions of non-native phonetic distinctions. The effect of laboratory *training* on perception of non-native segmental contrasts constituted a further important issue of investigation. In the 1980s, a number of productive research programs on these issues had been established by prominent researchers like James E. Flege, Winifred Strange, Catherine Best and their colleagues. Numerous papers presented at several specialized meetings of the American Acoustical Society of America (see the Journal of the American Acoustical Society of America and the volume edited by Strange 1995a) and the proceedings of the New Sounds conferences in Europe (Leather & James 1990, 1992, 1997, 2000) presented results from a large range of experimental studies in L2 speech production and perception. The series of New Sounds conferences have been continued recently (Rauber, Watkins & Battista 2007; Dziubalska-Kołaczyk, Wrembel & Kul 2010; Cardoso & Trofimovich 2014), reflecting the growing scope of empirical methods in investigations of L2 speech production and perception.

While many studies on cross-language speech perception have dealt with segmental perception, the study of suprasegmentals in L2 has so far received less attention, although suprasegmental differences in syllable structure, temporal organization of speech, stress, accent, and tone, and differences in phonotactic constraints have substantial influence on the acquisition of a second language phonological system (more recent studies e.g. by Sereno & Wang 2007; Burnham & Mattock 2007; Aoyama & Guion 2007).

2.3.1 Experimental investigation on L2 perception

Experimental studies in speech perception use several types of tasks to gather information about cross-language speech perception that will be discussed in this section.

Behavioural studies on speech perception ask listeners to listen to stimuli with differing acoustic features such as formant frequencies, duration, fundamental frequency, stimuli under noise etc. and to give responses according to the experimental setting.

Two major types of perception tasks are particularly common to test segmental perception in second language acquisition: *discrimination* tasks and *identification t*asks (for an overview cf. Repp 1984; Strange 1995b; Flege 2003; Hansen Edwards & Zampini 2008). The listeners' responses can be the identification of the stimulus, the discrimination between two different stimuli, or a matching or comparison of one stimulus against another. Segment-based studies on perception mostly use some form of discrimination or identification tasks that will be described in the next section.

Additionally to the task type, several other design variables are available for experimental tests concerning the *input material* itself: The input signal may consist of *natural or synthetic stimuli*, of *non-word items or real words*, and of speech material produced or presented in a continuous speech context, in single words (citation form) or in isolation. The input may consist of utterances from a single speaker or from more than one speaker.

More recently, *brain-based methods* have been introduced to the field of cross-language perception to study language-specific speech perception (see section 2.3.7).

2.3.2 Variability in L2 speech perception

Several factors are expected to explain the L2 listeners differential success and behavioural output in experimental performance in second language speech perception. Variability in the output of perceptual experiments can be described on four different levels:

(1) a *language-specific level* defined by the linguistic structure of L1 and L2,

(2) a *methodological or contextual level*, where task-related situational variables like the experimental setting, the stimulus material, the task type and instructions to the listeners, and the response options offered may influence the experimental output,

(3) an *individual level* specified by person-related variables like quantity and quality of language experience, age, individual strategies and hypotheses about L2, etc. and

(4) a *speech-inherent or language-universal level* depending on underlying mechanisms in human speech production and perception.

Individual, person-related factors are determined by the listeners' language background and their language learning history. They include a person's first language and all other languages acquired, the age of learning, length of exposure and quality of input in the target language, language aptitude, motivation, and several other factors (for discussion, see chapter 8).

While person-related factors are inherent to the L2 learner and cannot be directly changed or manipulated in an experimental setting, *learner-independent factors* can be manipulated or controlled by the researcher by choice of the speech input presented in the experiment, the elicitation technique or the response alternatives. Studies investigating the perception of L2 sounds use different kinds of speech input such as natural, synthetic or synthetically manipulated natural speech or different forms of responses in the task.

2.3.3 Language experience and the choice of subjects

Many studies compared perceptions of experienced learners of L2 with unexperienced or "naïve" listeners (beginners), and some have tried to operationalize language experience in their experimental design (e.g. Piske, MacKay & Flege 2001).

The variation and complexity of factors influencing L2 speech perception is hard to describe within only a few dimensions. Best & Tyler (2007: 15f) suggest three dimensions of experience to compare *naïve listeners* with *experienced L2 learners*: (1) L1 acquisition at the onset of L2 learning, (2) ratio of usage of L1/L2 on an average daily basis, and (3) ratio of L1/L2 in the language environment. However, the operationalization and measurement of these three dimensions is difficult to control in an experimental setting. Moreover, other individual factors influencing the subjects' success in the L2 acquisition of German have to be considered, such as e.g. the number of foreign languages learned, pronunciation training and others (for discussion of the role of these factors and the effect on experimental results in the present study, see chapter 8).

2.3.4 Natural vs. synthetic stimuli

Numerous studies on cross-language speech perception use non-native minimal distinctions, that is, single phonetic feature contrasts that mostly are linguistically irrelevant in the subjects' L1 but of linguistic relevance in the target language.

Segmentation, i.e. the ability to identify "segments" in the continuous speech signal, and *normalization*, the ability to appropriately perceive sounds and words realized by different speakers in different contexts and at different rates as "the same", are two essential components of speech perception.

In an experimental setting, it can never be excluded that a modification of input (as in modified natural speech) may influence the speech processing procedure of the stimulus. The use of different experimental methods in studies on speech perception can possibly reduce this risk of creating artefacts (Sendlmeier 1985: 70f.). Both, experimental procedures which investigate human sensitivity to specific acoustic parameters in the speech signal by using synthetically manipulated stimuli, and experiments using natural units of speech, can complement each other in providing insights into mechanisms of language-specific speech processing. The use of *synthetic stimuli* allows the control of phonetic variation whereas with natural speech input subjects may use redundant acoustic features to compensate problems in discrimination and identification. Another advantage of synthetic stimuli is the control over contextual influences that could not be controlled to the same extent in natural speech due to coarticulation and speaker-specific characteristics.

Undoubtedly, experimental studies with controlled synthetic input stimuli gather important evidence for speech processing mechanisms, but they are far from natural speech perception processes with respect to the stimulus material as well as to the experimental situation, which clearly differs from every-day speech processing. Especially in experiments with learners at a non-advanced level in L2, synthetic speech stimuli may be extremely irritating and exhausting and may cause even more problems than natural stimuli would. Therefore, the current study will use natural speech stimuli to gather evidence for perceptual vowel categorization in L2.

If we wish to examine natural language processing in studies on second language perception, we may also wish to use existing words of the target language. However, in many cases, minimal pairs with real words do not exist for all possible segmental distinctions. Another problem linked with real words is that a morphological or semantic relation between stimuli in a discrimination task (prime and target) and the unequal frequency of accourrence could influence the target word processing (Meunier & Segui 2002). Therefore, the use of non-words is very common in experiments on speech perception. This study will use mono- and disyllabic non-words to explore the full set of German vowels in L2 speech perception.

2.3.5 Discrimination, identification and perceptual assimilation tasks

2.3.5.1 Discrimination tasks

Discrimination tasks test the ability to perceptually distinguish specific phonemic contrasts. In discrimination tasks, two or more stimuli are presented to the subjects who are asked to make a comparative decision on the relationship between them, i.e. to decide whether the stimuli are "the same" or "different".

There are several forms of discrimination tasks differing in number and sequence of items presented: AX, ABX, AXB or *category discrimination tasks* (CDT), and *oddity discrimination tasks* (ODT). In ODT tasks, the "odd" item, i.e. the one that differs from the other(s) in a (mostly) triad, has to be selected. The triads contain potentially confusable sounds of the target language (see discussion of possible results and advantages vs. disadvantages in Flege 2003).

Usually, results from discrimination tasks reflect the *auditory* component rather than or in addition to long-term memory representations of L2 categories. Typically, vowels are inherently more discriminable than consonants (see e.g. Stevens et al. 1969; Strange 1995b). Of course, the discriminability (related to auditory short term memory) of phones might be better than the ability to correctly identify L2 sounds (related to representations in long term memory). A "different" response could refer to phonetically non-relevant information in the input signal. L2 learners may use different phonetic cues than natives to distinguish two categories of L2, e.g. differences in length instead of formant frequencies (Flege et al. 1997; for discussion, see Flege 2003: 24). Therefore, it is appropriate to perform experiments with a control group of native listeners.

While the perception of consonants has often been described as categorical, the perception of vowels has been described as being more continuous (Fry et al. 1962) in that sensitivity within a vowel category is often higher than predicted. Still, phonemic boundary effects have been found with vowels but appear to vary with the exact experimental design.

Categorical discrimination tasks with stimuli spoken by different native speakers of L2 (e.g. Flege, Munro & Fox 1995; Gottfried 1983) encourage responses in a general rather than a token-specific mode.

Inter-stimulus interval (ISI) is one of the determinant factors: In stimuli presented with a relatively short ISI, the differentiation and distinction of phonetic differences which normally do not function distinctively in L1 phonetic sequences is easier. Thus, with shorter ISI the possibility increases that the responses are based on information in auditory short-term memory (see e.g., Werker & Logan 1985; Strange 2010).

Macmillan, Goldberg & Braida (1988) suggest that the listeners' accuracy in discrimination tasks is limited by three major types of variance: (1*) sensory variance* due to resolution limits of the auditory system, (2) *context variance* owing to the range of stimuli presented in block of trials, and (3) *trace variance* connected with the length of time a stimulus is stored in memory. The presence of phonemic boundary effects has been found to vary with these three types of variance. A high degree of context variance and increases in trace variance may promote vowel boundary effects (for discussion, see Iverson & Kuhl 2000: 877ff). Moreover, the duration of the interval between the stimuli (ISI) is expected to have an effect on whether the signal is analyzed by auditory memory traces or by higher-order categorical labels[6] (Repp et al. 1978; Gerrits & Schouten 2004). However, vowel perception is considered to be less categorical (for discussion, Gerrits & Schouten 2004).

2.3.5.2 Identification tasks

In *identification tasks*, speech stimuli are presented to the listeners, who are asked to indicate the category to which they think the stimulus belongs to by selecting one of the given response alternatives or by giving some other kind of active response (in oral or written form). Repeated instances of the same category are presented in random order, results are scored in terms of wrong/correct classification.

Identification tasks require an active *categorization* by the subjects. Instances of misidentifications reflect patterns of confusion between L2 sounds and provide insights into how the L2 learners' mental representations of L2 categories differ from that of natives (Flege 2003: 32). Frequent misidentifications of phoneme $/X/_{L2}$ as $/Y/_{L2}$ by L2 learners would suggest that the two L2 vowel categories are systematically confused with one another and are perceived as "similar" to each other and therefore cannot be differentiated sufficiently.

Response items can be either chosen from the L1 repertoire to test "assimilation" to native categories, or they may represent sound categories from L2. Response categories can be presented as orthographical labels or as key words representing the categories tested.

Identification tasks, where listeners have to choose among response categories from the *L2* repertoire, allow conclusions about the difficulties of advanced learners which may be largely due to persistent problems in the discriminability of some "similar" L2 categories. Insufficient

[6] To compensate for larger inter-stimuli intervals in discrimination tasks, subjects must rely on labels they have assigned to the input stimuli, due to the relatively short time span of auditory memory (200-300 ms) (Gerrits & Schouten 2004).

discrimination is reflected in instances of perceptual confusion of the categories involved; these may consequently be responsible for insufficient distinction in production. Difficulties in discriminating L2 sounds can then be described as a function of experience with L2 (beginners vs. advanced learners).

Identification tasks most closely mimic the process of phoneme decision in natural speech perception. However, there are some weaknesses to the method that will be discussed in section 2.3.6.

2.3.5.3 Perceptual assimilation tasks

In recent years, a growing number of studies used *perceptual assimilation tasks* or cross-language categorizations (e.g. Hentschel 1986; Strange 1995b, 2007; Strange et al. 2001; Levy 2004) that are claimed to provide direct measures of perceived L1/L2 similarity relationships (see Strange 2007b for a critique of these techniques).

In *perceptual assimilation tasks*, a subtype of identification tests, the identification of L2 stimuli with L1 response categories is required to show which vowel(s) in L2 are identified by the listeners as instances of a category of their L1. Additionally, goodness of *fit ratings* may be performed (e.g. Guion et al. 2000; Iverson et al. 2001; Best et al. 2003; Frieda & Nozawa 2007).

This experimental design provides answers to the question which L2 phones are assigned to which category of L1 and allows predictions about learning problems especially in the initial stages of foreign language learning, when "equivalence classifications" between L2 and L1 sounds (Flege 1987) are quite common. However, for advanced learners the results of perceptual assimilation tests may be less relevant as learners appear to be conscious about the "different" nature of "new" L2 sounds. Goodness of fit-ratings may reduce this effect to a certain degree but nevertheless forced choice identification tasks using L1 response categories cannot provide sufficient experimental evidence for L2 speech learning. However, assimilation tasks are currently a wide-spread experimental design in studies on L2 speech perception.

2.3.6 Testing L2 vowel perception – methodological considerations

A critical comparison of different methods to gain evidence on L2 sound perception reveals that some methodological aspects have to be considered carefully in test performance and data interpretation. All task types described above consist of two major components: (1) the speech signal (input) and (2) the listeners' behaviour (output). The most obvious methodological problem with behavioural output from identification tasks and discrimination tasks is that they always involve some kind of productive element such as selecting a category, marking it on an answering sheet or repeating it orally. The experimental design and the choice of categories and response modes has an influence on the learners' performance in the test that is difficult to control for. Especially when a larger number of categories are tested, the process of deciding and choosing from alternatives will be more difficult. This may lead to secondary errors when subjects – despite correct perceptual identification – eventually may choose a wrong response option.

Another problem of perception tasks is that the perception of the categories tested may be influenced by the test procedure itself in that during performance some effect of familiarization or even learning is induced; occasionally there are cases reported where learning of one L2 category appears to deteriorate another adjacent L2 category (e.g. Major 1987; see discussion in Flege 2003: 20-23).

Moreover, decisions in the productive phase of the task always contain a decisional element that is influenced by previous decisions and all stimuli a learner has been experienced before.

An important experimental aspect concerns the choice of *response alternatives*, their number, form and manner of presentation. Finding labels that can be used unambiguously by the subjects is not easy. L2 spelling conventions or even phonetic symbols are usually not easy to handle for the participants. Some experiments (e.g. on voiced vs. voiceless contrasts) are limited only to a choice between two response alternatives (e.g. "press button A or B"), which represent the members of the phonemic opposition of interest. In some studies, reaction times are measured to provide additional insight into the relative difficulty of L2 categories. Some studies use keywords for response instead of graphemic representations (Levy 2004; Strange 2007), but participants might be biased to prefer the more familiar word (Flege et al. 1995; Gottfried 1984).

Another serious problem with identification tasks is the possible influence of orthographic representation. In many experimental settings, answering options are presented in a written way. This might be a problem for subjects who are not sufficiently familiar with the writing system of the language tested. Particularly, orthographic systems that exhibit less transparency in phoneme-grapheme-correspondences (e.g. English or French, and to a certain degree also German) cause problems in language acquisition. In a perception task, this might cause instances of what may be called "spelling perceptions" (in analogy to "spelling pronunciations"), where subjects choose the wrong orthographic representation, misled by their L1 (e.g. native speakers of English choosing <u> when perceiving short [a]). However, even if not directly involved in the test procedure, differences in the orthographic representation of the same phoneme in L1 and L2 (or even of L3 or L4) may always have an influence the representation and categorization of phoneme categories to some extent (Bassetti 2009). Alternatively, pictures representing keywords that contain the sounds under consideration can be presented.

Strange and colleagues posit that direct assessment of perceptual similarity is possible in perceptual assimilation tasks, in which non-native segments are presented to listeners who are asked to categorize them with respect to "the most similar" native category and to give a rating on "category goodness" (e.g. Bohn & Flege 1990; Guion, Flege, Akahane-Yamada & Pruitt 2000; Strange, Bohn, Trent & Nishi 2004; Levy 2009). By this type of task it should be possible to establish cross-language perceptual similarity independent of identification or discrimination performance in order to predict L2 learning difficulties appropriately. However, perceptual assimilation might be a cognitive mechanism found with beginners, but similarity relationships between L2 categories established by more advanced learners will not directly be represented by the results of a perceptual assimilation task with category goodness ratings. Therefore, in the present study, the perceptual assimilation paradigm will not be accepted here for advanced learners. Rather, an identification task will be used where listeners are asked to identify stimuli as belonging to one of the target language categories offered in the experiment, revealing potential confusion between different categories of the target language. The interpretation of these patterns of confusion will reveal the learners' difficulties and possible similarity relationships between L2 categories as established by beginners and advanced learners of L2 German.

Interlanguages are emerging systems that are influenced by a number of internal and external factors. The development of L2 sound categories and their mapping in the perceptual space is determined not only by acoustic, phonetic and phonological features but also by the learners' actively established connections and associations with other subsystems as morphology, lexicon and orthography.

2.3.7 Brain-based methods

Apart from behavioural studies, where participants are required to give some kind of active response, *brain-based methods* have been introduced from experimental psychology and neuroscience to investigate possible unconscious differences of perception in L1 and L2 (see Strange & Shafer (2008) for an introduction and Näätänen et al. (2007) for a detailed review of basic research methods using brain-based technologies). Brain-based methods are an important tool supplementary to behavioral data to investigate language-specific speech perception.

Event-related brain potentials (ERPs), measured with electroencephalography, are used to distinguish and identify psychological and neural sub-processes by measuring brain activity during cognitive processing in perceptual tasks.

For auditive perception, most evidence has been gathered in analyzing the differences by *Mismatch Negativity* (MMN) response (e.g. Dehaene-Lambertz 1997; Näätänen 2001; Hohlfeld, Mierke & Sommer 2004; Sebastián-Gallés 2008) to test whether participants can neurologically distinguish between certain sounds. In these types of studies, auditory stimuli are presented with a rare deviant sound embedded in a repetitive sequence of the same sound type. Mismatch negativity – the amplitude of the ERP index of discrimination – occurs after the deviant sound in the sequence indicating the passive discriminability of sounds. MMN data indicate that for each sound, both speech or non-speech, neural representations develop in sensory memory that correspond to the percept of this sound in the neurophysiological substrate of auditory sensory memory. MMN results from any discriminable auditory change deviant from a repetitive "standard" stimulus. No MMN can be elicited if two phonemes are indiscriminable to the subject or learner. The MMN amplitude is larger with easier discriminations and with improved discrimination ability, e.g. after training (Näätänen 2001: 7). MMN data can reveal the accuracy of this representation for each auditory feature separately, e.g. duration, formant frequency etc.

Several studies indicate that different auditory event-related potentials (ERPs) can be obtained for native and non-native speech perception (e.g. Näätänen et al. 1997; Näätänen 2001; Winkler et al 1999; Hohlfeld, Mierke & Sommer 2004). Winkler et al. (1999) found that Hungarians could not behaviorally discriminate the Finnish /e/ - /ä/ contrast and had no MMN to this contrast.

MMN can be elicited regardless of whether the subject is paying attention to the sequence or not and even when the stimuli are not behaviorally, i.e. consciously, discriminated by the subjects. Therefore, MMN data serve as an objective measure of auditory discrimination accuracy.

2.4 Models of cross-linguistic speech perception

Studies on cross-language speech perception provide ample evidence for language-specific patterns of speech perception. Several theoretical approaches have been proposed to account for the inaccurate perception of L2 speech sounds. Current Models of L2 speech perception are all based on the premise that perception in L2 is determined by L1 (Flege 1995; Flege 2002; Kuhl & Iverson 1995; Kuhl 1998a, b; Best 1995; Best & Tyler 2007). Models of non-native speech perception suggest the existence of differences in the discrimination and categorization of certain L2 vowel contrasts and formulate theoretical predictions to account for the relative difficulty of L2 sound structures in perception.

Most approaches agree to a certain degree on the fact that specific features or properties necessary to develop accurate perceptual representations are not accessible to L2 learners and that this inaccessability is due to two main factors: One is the factor *age*, referring to the chronological age at which L2 learning began and to possible constraints in the learnability of a native-like pronunciation for L2 speech sounds due to a "critical" or "sensitive" period (cf. Lenneberg 1967; Penfield & Roberts 1959; Birdsong 1999), though this idea has been challenged in more recent works (for a critical discussion, see e.g. Bongaerts et al. 1997; Flege & MacKay 2010). Another string of argumentation refers to possible limitations in the emergence of *perceptual representations* for phonetic categories in L2 that may be restricted by the amount of *language experience* in L2 (Piske, MacKay & Flege 2001; Flege 2003). For a more detailed discussion, see section 8.7.

2.4.1 The Native Language Magnet Model (NLM)

Patricia Kuhl's *Native Language Magnet model* (*NLM)* (Kuhl 1992; 1993; Kuhl & Iverson 1995) primarily aims to describe developmental changes in auditory perception reflecting the reorganization or attuning of phonetic perception patterns in infants' acquisition of their first language. These developmental changes result in language-specific perceptual patterns. In L1 acquisition, linguistically functional contrasts of L1 continue to be well perceived while the discrimination of non-native phonetic contrasts deters gradually (but not equally for all sounds, see Polka & Bohn 1996). The underlying auditory-phonetic space is considered to be "warped" towards the language-specific phonological system of the native language (Kuhl 2000a, b).

The *Native Language Magnet* model (Kuhl 1992, 1993) suggests that speech perception is initially determined by nonlinearities in auditory perception and is altered by the acquisition of phoneme categories. Perception becomes fine-tuned to the characteristics of one's native language. According to the Native Language Magnet model, exposure to one's first language leads to the formation of language-specific *category prototypes for L1* phonemes. These category prototypes are exceptionally good exemplars of phonetic categories that act like *perceptual magnets*. Discrimination in the vicinity of a prototype will be more difficult compared to non-prototypes. Non-native categories would lack this prototype, especially for naïve listeners in L2. An acoustically similar L2 vowel would then show a “perceptual magnet effect” like a native vowel. If the L2 vowel presented is not similar to any native vowel, it should be more easily discriminated from surrounding vowels and no magnet effect will be observed.

Kuhl’s (1991) experiment with synthesized exemplars of /i/ showed that some tokens in a particular region of the vowel space received very high goodness ratings. On the basis of goodness ratings, an excellent exemplar (*prototype P*) and a poor exemplar (*non-prototype NP*) were defined and manipulated acoustically. In a subsequent discrimination task adults and infants showed more difficulties to discriminate *P* from its variants than in discriminating *NP* and variants of *NP* (Kuhl 1991; Grieser & Kuhl 1989; Kuhl et al. 1992). This effect seems to be typically human, whereas monkeys do not exhibit this perceptual magnet effect (Kuhl 1991).

Thus, the *perceptual magnet effect* refers to the occurrence of poor sensitivity near best exemplars, which is explained by a kind of “shrinking” the perceptual space near best exemplars. It is hypothesized that phoneme categories are represented in terms of prototypes, i.e. a single abstract exemplars that are representing all members of a given category (Iverson & Kuhl 2000).

Kuhl, Williams, Lacerda, Stevens & Lindblom (1992) investigated the effect of language experience on the same vowel prototypes with listeners from two different languages (English and Swedish) in the early phases of life in a study with American and Swedish 6-month old infants with the prototype English /i/ and Swedish /y/. The results showed that the perceptual magnet effect is strongly affected by exposure to a specific language. In both language groups the foreign-language vowel was treated as a non-prototype.

The NLM model describes three phases of development in infancy which characterize the transition from early universal patterns of phonetic perception to language-specific patterns of perception: In phase 1, the initial state, infants show the ability to differentiate all sounds occurring in human speech. This ability is supposed to derive from their *general auditory processing* rather than speech-specific mechanisms. In phase 2, due to infants’ sensitivity to distributional properties in the ambient language input, phonetic representations are developed. Thus, *language experience* is “warping” perception, resulting in a distortion of the original sensitivities: perceptual sensitivity is decreasing near the category modes while increasing near the boundaries between categories (Kuhl 1991; Iverson et al. 2003). Representations that are most often activated (*prototypes*) eventually function as *perceptual* *magnets*, for other members of the same category. In phase 3, this distortion of perception shows its effect, the *perceptual magnet effect*, facilitating phonetic abilities in the native language while reducing perceptual sensitivity in a foreign language. To summarize, at the end of the first year of life a dual change is supposed to happen in infants' speech perception: speech perception of non-native contrasts declines while native-language speech perception skills improve and continue to develop at least during the first ten years of life.

Kuhl, Conboy, Coffey-Corina, Padden, Rivera-Gaxiola & Nelson (2008) present an expanded version of NLM model (*NLM-e*): In phase 1, infants discriminate phonetic units universally. In this phase, phonetic abilities are relatively crude, reflecting *general acoustic constraints* and allowing for developmental improvement. Variation across phonetic contrasts is explained by the acoustic salience of a phonetic contrast (e.g. fricatives are more difficult). Further, directional asymmetries are observed, where a change in one direction results in significantly better performance than in the other direction (for vowels, see Polka & Bohn 1996, 2003). In phase 2, the core of NLM-e, *phonetic learning* takes place, caused by the distributional patterns of the ambient language and by exaggerated cues in infant-directed speech. These exaggerated cues, together with social interaction and increased attention and arousal, result in more robust and durable learning. For vowels, learning occurs earlier than for consonants (Werker & Tees 1984a; Kuhl et al. 1992; Polka & Werker 1994; Best & McRoberts 2003). Further, NLM-e posits a link between the perception of acoustic signals and speech production induced by vocal play and vocal imitation of heard input. By imitating stored perceptual representations, language-specific speech production evolves. By the end of this phase, infants' perception is altered and the detection of native cues is enhanced. In phase 3, *enhanced speech perception abilities* lead to an improvement of three independent skills which play a role in word acquisition: (a) the perception of phonotactic patterns, (b) the detection of word-like units and (c) the resolution of phonetic details in early words. By phase 4, relatively *stable neural representations* have been established (neural commitment) and new utterances do not cause shifts in the neural coding of distributional properties. From this moment on, future learning will be affected by *native language patterns*. In infancy neural networks are not completely formed and thus do not restrict learning processes, whereas with adults representations are stable and relatively unaffected by a small amount of input in a foreign language. Thus, exposure to a new language will not automatically create new neural structure as observed with infants. The underlying assumption of the model is that the degree of "plasticity" in learning phonetic structures in L2 depends on the stability of the underlying perceptual representations and on the degree of neural commitment (Kuhl et al. 2008).

2.4.2 The Speech Learning Model (SLM)

The *Speech Learning Model (SLM)* has been developed by James Flege (Flege 1988, 1992, 1995, 1999a, 2002) and colleagues. SLM is based on a wide range of perception and production studies and has been most influential on numerous empirical studies in L2 production and perception.

Flege (1995: 238f) hypothesizes that sounds in L1 and L2 are related perceptually to each other and that this relation refers to a position-sensitive, allophonic level rather than to an abstract phonemic level (as was claimed within Contrastive Analysis, e.g. Lado 1957).

A central tenet of SLM is the *"new" vs. "similar"* dichotomy and the establishment of *equivalence classifications* of L1- and L2-sounds. "New" phonetic categories can be established for L2 sounds if learners are able to discern (at least some) relevant phonetic differences to L1 sounds. If phonetic differences between sounds from L1 and L2 can be identified, a new phonetic category can be established.

The greater the perceived cross-language phonetic dissimilarity between sounds of L1 and L2 the greater the probability that these sounds will be discriminated. That is, phonetic similarity of L1 and L2 sounds is supposed to hinder the establishment of a new L2 category. Accurate perception and production in L2 is only possible if the learner develops adequate long-term representations (categories) for L2 sounds. The relative difficulty of a given L2 phoneme will depend on the *degree of similarity* of a given L2 sound to the closest native one.

Category formation for "new" L2 sounds is more likely than for "similar" sounds. If the new phonetic category for an L2 sound established by a bilingual matches with native speakers of L2, then the L2 sound can be stored and produced accurately. Eventually, L2 sound production will correspond to the phonetic category representation in long-term memory.

However, category formation in L2 may be blocked by the mechanism of "*equivalence classification*" (Flege 1988). Then, a single category will be used to process perceptually linked sounds from L1 and L2, so called "diaphones" (cf. Weinreich 1953). These diaphones will be similar in production.

The ability or likelihood of discerning phonetic differences in L2 that are non-contrastive in L1 seems to decrease with *age of learning* (AOL). The more developed the L1 system is the more it will interfere with L2 learning. As sound categories of the L1 develop and stabilize through childhood and into adolescence, they seem to become more powerful attractors of L2 sounds (Flege 2002; Walley 2007; Baker et al. 2002). Therefore, adult learners may be more likely to judge L2 sounds as members of L1 categories than children. However, age is not the only determinant factor for achievement in L2. *Language experience* and *language use* of L1 and L2 are important predictors for success in L2 speech production (e.g. Flege, Bohn & Jang 1997; Piske, MacKay & Flege 2001; Flege & MacKay 2004).

Phonetic categories established by bilinguals may differ from monolinguals' categories if bilinguals' categories refer to different phonetic features or different weighting of features in category formation or if bilinguals strive to keep representations for L1 and L2 sounds apart (Bohn & Flege 1992). By hypothesis, L2 categories of bilinguals may be deflected away from existing L1 categories to preserve phonological contrast within an individual's phonological space (Flege 1995: 242).

One quite controversial but important postulate of SLM is that the mechanisms and processes underlying successful L1 speech acquisition including category formation remain intact over the life span (Flege & MacKay 2010, 2011). These capacities include the accurate perception of feature patterns, the categorization of segments with common properties and the ability to relate perceptual to productive properties of speech.

Concerning the perception and production of L2 vowels, the SLM generates specific predictions (Flege 1995: 243ff): Phonetic differences between L1 and L2 sounds can be perceived even by adult L2 learners. The greater the distance perceived between an L2 vowel and the most similar L1 category, the greater the possibility that a new category for L2 will be established.

Depending on whether a new category has been established or not, the model predicts different production patterns for L2 vowels. If no new category has been established, no native-like production patterns can be achieved unless by chance. If a category is established, the model predicts accurate production in L2, unless learners attempt to "deflect" the target sound away from an L1 category. Thus, Flege's model of second language speech learning accounts for L2 perception and production in a very integrative way.

To summarize, SLM makes the following predictions for the acquisition of L2 phonemic contrasts: L2 phonemes may be *similar* or even identical to a native one, or they may be "*new*", i.e. dissimilar from any L1 phoneme. The insufficient perception of sound contrasts results in "equivalence classifications", i.e. L2 sounds are perceived with the sound categories of L1. "New" sounds in L2, being perceivably *different* from L1, are easier to perceive accurately than "similar" sounds. L2 phonemes are considered to be easiest to learn when they are far from existing L1 categories. These L2 phonemes are assumed to be learned without changing any of the L1 categories. Therefore, the *relative difficulty* of a sound in L2 can be defined as a function of *degree of similarity* to existing L1 categories. The greater the perceived difference between the categories of L1 and L2 the greater the likelihood of new category formation in L2.

2.4.3 The Perceptual Assimilation Model (PAM)

Best's (1994, 1995) *Perceptual Assimilation Model (PAM)* was originally developed to explain non-native speech acquisition by naïve listeners but has been extended later to PAM-L2 for the perception of L2 learners (Best & Tyler 2007). The central tenet of PAM is that incoming foreign sounds are compared and assimilated to the listeners' native inventory in one of several ways. PAM posits that non-native speech sounds are *perceptually assimilated* as more or less good exemplars of the listener's native language sound inventory. The change in perception of non-native contrasts observed with infants at 10-12 months of age is explained by infants' recognition of the articulatory gestures underlying speech. Due to listeners' native language experience, older listeners assimilate unfamiliar non-native phones perceptually to the articulatorily most similar native phoneme.

The discrimination of a non-native contrast is expected to be very good or excellent if both members are assimilated to two different native categories ("*two category formation assimilation" TC*). Discrimination is hard or less successful for contrasts in which members are perceived as equally good (or poor) tokens of a native phoneme and, therefore, are assimilated to only one native category ("*single category formation*" *SC*). If both segments are assimilated to a single L1 category, they may differ in their perceived *category goodness (CG)*, and, therefore may be better discriminated than in the single category pattern. Their "*goodness of fit*" to the native categories is crucial for the decision whether non-native phones belong to a single or to two different categories. Finally, non-native phones may fail to sufficiently match any given native phoneme and will be judged as *"uncategorized"*. A non-native phone in an L2 contrast may be perceived as a native phoneme or an uncategorized speech sound. L2 contrasts may result in an *uncategorized-categorized assimilation* with better discrimination or in an *uncategorized-uncategorized assimilation* with poor to moderate discrimination.

Concerning the performance in discrimination tasks, it is expected within PAM that TC contrasts are better discriminated than CG contrasts, and CG contrasts better than SC contrasts. Note that these predictions refer to the ability of discrimination. Perceptual identification or production is not the focus of the model's predictions.

The major emphasis of PAM in its earlier version was to describe the discrimination of non-native consonant contrasts by naïve adult listeners, i.e. by persons with little or no experience in L2. Later studies extended the model to the perception of vowels (Best, Faber & Levitt 1996; Best et al. 2003). Results of these studies can account for the "initial state" of L2 learners beginning to learn a new language but not for sound perception by more advanced learners.

Best & Tyler (2007) extended the PAM to predict patterns of speech perception by more experienced L2 learners (*PAM-L2*), discussing four possible cases of L2 minimal contrasts (Best & Tyler 2007: 28ff):

(1) if only one L2 phonological category is perceived as equivalent to an L1 category: If, at the phonetic level, one member of the L2 contrast is perceived as a good token of an L1 category, the chance for perceptual learning is relatively restricted. If, however, one L2 phone is perceived as deviant from a category in L1, that is if the learner perceives phonetic differences but no differences in its phonological function, they can be easily dissimilated at the phonetic level.

(2) if both L2 phonological categories are perceived as equivalent to only one L1 phonological category but not as equally good exemplars (cf. *category goodness assimilation contrast*): Discrimination and recognition of lexical-functional differences of the L2 phones are expected to be fairly good, a new phonetic and phonological category is likely to be established eventually for the more deviant L2 phone. The phone perceived as less good exemplar of an L2 category will be perceived as a phonologically and phonetically equivalent variant, no new category will be established.

(3) if two L2 phonological categories are perceived as equivalents of the same L1 phonological category and are considered as equally good or poor exemplars of that category (*"one-category"*): Initially, both will be assimilated phonetically and phonologically to the same L1 category. A learning effect is expected to depend on whether both are perceived as good or poor exemplars of the L1 category; perceptual learning requires the learning of a new phonetic category for at least one phone and the establishment of a new phonological category.

(4) if no L1-L2 phonological assimilation takes place *("uncategorized")*: The perceptual learning of one (or two) new categories is expected to be relatively easy. On this point, there is an interesting difference between PAM and SLM's "new" sounds: In PAM, perceptual learning is determined not only by the similarity or dissimilarity of a given L2 phone to a L1 category but by the comparative relationships of sounds within the phonological system. Here, PAM makes an important distinction: If both contrasting phones are similar to different L1 phones and are thus relatively distant from each other, the L2 listener will easily recognize

relevant functional differences in L2 and two new categories are likely to be learned. If, however, the two uncategorizable L2 phones are perceived as similar to each other, the discrimination and the perception of lexical-functional properties will be difficult and a single new phonetic and phonological category will be learned, which may eventually be split into more than one category during further L2 learning. This tenet will also be central for the discussion of the present work.

Experimental studies using perceptual *assimilation tasks* (e.g. Hentschel 1989; Best et al. 2003; Strange, Levy & Law 2009) where subjects are asked to identify the L1 sound category, which is most similar to the L2 input, are implicitly motivated by these basic assumptions from SLM. However, a problematic aspect of the perceptual assimilation approach is that from a contrastive comparison of the segmental inventories of L1 and L2 it does not immediately become apparent when an L2 sound will be classified as different and when as similar. As is hypothesized in the present study, perceptual assimilation to L1 categories cannot account for all instances of misidentifications. Other mechanisms such as over-generalization, hypercorrection (see section 1.3.2) and the compensatory use of acoustic cues that are not used by native listeners of the target language, occur in L2 vowel perception and account for several instances of misperception and wrong categorization of L2 segments.

Little work has been done on the interaction of categories in a system as a whole. Frieda & Nozawa 2007 used a two-step approach with perceptual assimilation (experiment 1) and discrimination (experiment 2)). The current study will present an analysis of L2 learners' perception of the entire phonemic inventory of German, extending the central assumption of equivalence classification in SLM to *intra-lingual similarity* relationships between L2 sounds.

2.4.4 Exemplar-based theories of speech perception

A central issue in studying second language acquisition concerns the interaction between sound categories and the learner's lexicon and the question whether words are stored as episodic traces or as abstract representations (Hazan 2007). Exemplar-based models account for two dimensions of second language acquisition: the acquisition of the phonological system of the L2 and the processing of speech input by learners in L2. Exemplar theory as a model of perception and categorization was first introduced in psychology and subsequently extended and specified for speech sounds (e.g. Johnson 1997; Pierrehumbert 2001; Ettlinger & Johnson 2009; Bybee 2010).

Most phonologists distinguish independent abstract phonological representations for speech sounds from their phonetic surface form and assume a kind of mapping of discrete phonological representations onto more gradient phonetic forms (for discussion, see e.g. Lahiri 2007). This view is directly opposed to exemplar models assuming that phonetic details and variation are coded in detail in memory.

Tests of auditory word recognition and recognition memory have demonstrated that acoustic-phonetic details are part of listeners' long-term representation of speech (Goldinger 1998). Apparently, fine phonetic details such as voice, affect, or pitch range, closely associated with individual talkers' characteristics, seem to be stored as part of the listeners' memory of words (Pisoni 1973; Goldinger 1998, 2007; Winters et al. 2013). These results have led to the assumption that the mental representation(s) of a sound consist of the sum of all heard tokens containing linguistic and extra-linguistic properties such as acoustic, lexical, contextual, and social information.

These assumptions are not consistent with the so-called normalization theories which assume an abstract level of sound representations, but rather support *episodic or exemplar-based coding* models, a common approach to the study of categorization and memory in cognitive psychology (Johnson 2008: 383).

In a *feature-based approach* to speech perception, the extraction of relevant features from the acoustic signal is assumed, followed by a process of mapping or categorization of these features (Liberman & Mattingly 1989; Stevens 2002). E.g. for a vowel, the listener would extract the relevant cues (e.g. formants), normalize them to stored values and the fundamental frequency, and would then transform them into a set of more abstract features (Ettlinger & Johnson 2009).

In an *exemplar-based approach*, no extraction of features is assumed. Rather, each speech sound is compared to the collection of stored exemplars and is assigned to the category in which the most similar cases are found. Speech production is then the reproduction of an acoustic signal from heard tokens in similar contexts.

Earlier works claimed an either/or-approach concerning the abstract vs. episodic memory representation, while more recent works focus on models that combine episodic and abstract representations (Cutler & Weber 2007; Goldinger 2007; Ettlinger & Johnson 2009). While purely episodic models would require vast memory resources, abstractionist models cannot account for influence of surface details such as speaker idiosyncrasies on perception. Goldinger (2007) argues that an optimal theory would include stable abstract representations, combined with context-sensitive episodic traces. Ettlinger & Johnson (2009) present evidence that an exemplar-based model of speech perception makes more accurate predictions for L2 discrimination of German front high vowels by English, Turkish and French listeners, arguing that the existence of at least one front rounded vowel in the L1 facilitates discrimination in L2. A coexistence of both kinds of representations, abstract features and detailed information in stored exemplars of words, is theoretically possible.

Exemplar-based theories of human sound categorization define a given speech sound as a collection of all tokens, or exemplars, a listener has encountered so far. Speakers of a given language have detailed long-term memories of particular percepts, which are stored on a kind of *perceptual map*. This map refers to the articulatory/acoustic space and its dimensions of articulatory and acoustic contrasts (Pierrehumbert 2002). Additionally, a set of labels over this map is assumed. For vowels, the F1xF2 vowel plot is commonly used as a "map" to represent at least some parts of the high-dimensional information encoded in the articulatory or acoustic vowel map.

According to exemplar theory, we store long-term memories of particular percepts as locations on the cognitive map. These are the "exemplars". A large number of remembered percepts are categorized under a particular label from the label set. These percepts are grouped in one particular area of the perceptual map ("exemplar cloud", see Pierrehumbert 2002). In the center of this distribution area, the density of stored percepts is very high, at the margin less exemplars are stored (Pierrehumbert 2002: 113; see also Krumhansl 1978). Errors in establishing the label set and phonetic distributions lead to various kinds of over- or undergeneralization in the phonetic output. Exemplars are associated with each other at multiple levels of representation referring to linguistic and extra-linguistic properties (Ettlinger & Johnson 2009).

Exemplar clouds are associated with phonological units in words. The relative frequency of a given word and the lexical neighbourhood density, i.e. the number of words which can be created from a given word by minimal sound/letter substitutions, are assumed to have an influence on word recognition in that with low neighbourhood density, words are more easily recognized than words with a high density. Similarly, the relative *frequency* of a given segment is expected to have an influence on the relative recognition rate. Word recognition results form the matching of an incoming acoustic signal with the most similar stored representation of the signal. Similar stored representations will be activated more than less similar exemplars.

Allophonic relations are modelled as a high co-activation of all occurring surface forms. High rates of co-activation make sounds perceptually less distinct (Boomershine et al. 2008). If in L2, two categories exist in an area where L1 has only one category, then an incoming L2 signal may activate both categories. This co-activation may result in a relative loss of perceptual distinctness.

However, since every utterance is stored with its acoustic representation, the acoustic representation for a given signal may eventually be altered by additional L2 input. Obviously, rich input including phonetic variation and multiple tokens spoken by different speakers will facilitate perceptual learning in L2. Davidson (2007) argues that new categories of L2 are more likely to be learned by exposure to minimal pairs of lexical items containing difficult sounds. Bradlow & Bent (2003) demonstrated that talker-independent perceptual learning is facilitated by exposure to talker variability.

Language processing and language acquisition in a second language, especially with adults, is fundamentally different from processing and acquisition processes in the L1. While infants develop perceptual categories primarily due to exposure and statistical learning, learners of a second language dispose over lexical, phonetic, phonological, morphological, syntactic and semantic-pragmatic *knowledge* in their first language that has to be considered as abstract knowledge (Cutler & Weber 2007).

Metalinguistic knowledge *about* L2 phonology is developed from a variety of evidence such as directly from listening, from graphemic or orthographical representation, from direct interaction and feedback of native speakers or classmates, from behavioural evidence, direct instruction etc. If the range of variation of two sounds of the target language is in an area of a single category of the learner's L1, difficulties in the acquisition process are expected. Thus, speech processing for adults is claimed to be not primarily data-driven but knowledge-driven. Cutler & Weber (2007) distinguish knowledge that is directly derived from the learner's analysis of the input and meta-knowledge, i.e. the knowledge about the existence of certain sound contrasts in the target language. The interaction of *meta-knowledge*, (insufficient) directly acquired *knowledge* and the *input data* in a given speech situation may have drastic consequences on the recognition of spoken words resulting in confusion: While the given input does not provide sufficient information about a specific contrast, the stored knowledge includes this distinction. Abundant evidence for this "conflict" of different types of knowledge will be presented in the experimental part of the current study.

2.4.5 Comparison and conclusions

All three models, NLM, PAM, and SLM, agree that cross-language comparisons cannot be performed at a level of abstract phonological classes or features but must consider the actual phonetic realizations of non-native sounds and their perception by individual listeners.

While the Native Language Magnet (NLM) model's primary focus is the acquisition of a sound system in native language acquisition, Flege's Speech Learning Model (SLM) and Best's Perceptual Assimilation Model (PAM) mainly focus on a description of non-native sound perception. While models of non-native speech perception like NLM and PAM focus on *naïve listeners'* perception of L2 characteristics, SLM's focus is the *L2 learning process* of beginners *and* more experienced learners. Experimental evidence from numerous studies for allophonic, phonotactic or sub-phonemic (phonetic) influences from L1 on L2 perception is consistent with the claims of SLM and PAM (cf. Best & Strange 1992; Best & Tyler 2007). The relative difficulty in perception is considered to depend on the listeners' native language background and the specific relationship between structures of L1 and L2. The models posit some kind of *assimilation* of non-native sounds to existing native categories and are thereby referring to some notion of "*similarity*".

Neither PAM nor SLM have restricted their predictions on influences from L1 to phonemic contrasts in L2 (as originally suggested by the Contrastive Analysis Hypothesis). All three models have stressed the importance of non-contrastive *phonetic similarities* and sub-phonemic differences between L1 and L2. While SLM's focus is the phonetic level, within PAM the phonetic and phonological level in perception of phones is considered. Naïve listeners can recognize phonological distinctions and phonetic deviations in their L1, but often are not aware of which phonetic distinctions function phonologically in an unfamiliar target language (Best & Tyler 2007: 23).

A central tenet of PAM and SLM is that a sound's *relative perceptual difficulty* is largely predictable from a comparison of the L1 and L2 system. SLM refines its predictions by differntiating "new" L2 sounds from L2 sounds that are "similar" to L1 categories. A differentiation of predicted difficulties in terms of "category goodness" is also proposed in PAM-L2 (Best & Tyler 2007).

Exemplars as representatives of a given sound category are explicitly considered only in NLM. While NLM explicitly accounts for the relative perceptual difficulty of different exemplars within a category, other approaches refer to a more abstract model for sound discrimination in L2, though SLM considers the fact that variation due to phonetic context may have an influence on the perception of L2 sounds. PAM refers to a more feature-based account for sound discrimination: Two-category assimilation refers to native sound contrasts. The use of the same feature-based contrast in L1 and L2 results in easy discrimination. Similarly, SLM predicts that a phonetic difference, which functions contrastively in L2 but not in L1, will result in poor discrimination of non-native sounds.

Ettlinger & Johnson (2009) present experimental evidence supporting the assumptions of an *exemplar-based model* of vowel discrimination rather than a feature-based account. They argue that while there is converging evidence that exemplar-based representations are used for lower-level tasks like sound discrimination, featural contrast may better account for higher-order patterning of language (Ettlinger & Johnson 2009: 237f; Mielke 2008) and suggest an account that uses both approaches. In an exemplar-based approach to L2 speech perception, experience with the sound under consideration is the most influential factor in perception and discrimination, as numerous training studies convincingly show. However, it may be that discrimination is limited to the words and the context presented during the training phase.

In a bottom-up account of speech perception in L2, phonemes do not seem to be the appropriate unit for description, since they represent a very abstract level of characterization. An *exemplar-based approach* starts from the auditory coding of a speech signal and assumes category labels associated with a frequency distribution of remembered instances of that label. The strength of representation in a specific area of the map depends on the number and recency of stored exemplars (Pierrehumbert 2003: 132). The relative difficulty of a given input signal in perception depends on the strength of representation and on the patterns of activation of stored representations in L1 and L2. Co-activation of more than one category results in a decrease of perceptual distinctivity or higher similarity of the sounds involved.

To conclude, the notion of c*ross-language phonetic similarity* is commonly referred to in L2 speech perception models to predict initial and continuing difficulties in perceptual differentiation of L2 contrasts. However, cross-language phonetic similarity as predictor for the relative accuracy or difficulty in discrimination or identification of non-native segments has to be based on a measure independent of non-native segment identification and discrimination performance. A detailed discussion of measures to identify distances and similarities between vowel categories will be provided in section 12.2.4.

2.5 Conclusions

This chapter presented an overview of major issues and research areas in speech perception and discussed methodological approaches to the study of speech perception in first and second language acquisition. Fundamental aspects of human speech processing (section 2.1) and the development of speech perception in L1 and L2 (section 2.2) were presented and common experimental methods in L2 perception studies (section 2.3) and models of sound perception and category learning in L2 were reviewed and compared (section 2.4).

Despite the complexity and immense articulatory and acoustic variability of realizations of the same vowel category that varies along multiple parameters, listeners show an astounding capacity to identify speech sounds correctly. This capacity is due to the interaction of cognitive processes at different levels of speech processing such as bottom-up and top-down processing, speaker normalization or categorization but seems to be reduced in L2.

One of the most fundamental cognitive capacities to deal with variability in perceptual input is *categorization* (see also section 12.1). Two modes of perception – categorical perception and continuous perception (Wode 1990) – can be distinguished (see section 2.1.2 and 2.1.3). *Categorical perception* refers to perceptual sensitivity for between-category but not for within-category differences, whereas *continuous perception* refers to perceptual sensitivity for changes along a stimulus continuum. While in perceptual discrimination tasks requiring "same – different" judgements the *continuous mode* is activated, identification tasks – as used in the present study – rather involve the *categorical mode* of perception to enable listeners to "label" input stimuli as belonging to a specific category.

The phonetic vowel space can be understood as a high-dimensional *cognitive map* on which a metric of proximity and similarity in terms of articulatory and acoustic properties is defined (cf. Pierrehumbert 2003). Categorical perception of phonemes in a particular language arises from deactivating or rather reducing the sensitivity for irrelevant phonetic distinctions in the native language and increasing sensitivity for those phonetic details that are phonologically relevant. Learning the phonological and phonetic patterns of a specific language involves the learning of probability distributions over the parametric phonetic space. By exposure to the ambient language, statistical properties of speech are picked up. In other words, linguistic experience alters perception of speech, "warping" perception towards the language of the infant's surrounding (Kuhl 2000b). This *warping of the perceptual space* takes place in early infancy (see section 2.1.4). The acquisition of an L2 subsequently to L1 requires processes of re-allocation of perceptual sensitivity for those cues that are relevant in the target language (see section 12.5). The language-specific warping of the perceptual space in a listeners' L1 limits his/her capacity for adequate perception in L2.

Current models of perception and production in L2, such as Flege's Speech Learning Model (Flege 1995), Best's Perceptual Assimilation Model (Best 1995; Best/Tyler 2007), and Kuhl's Native Language Magnet Model (Kuhl 1995; Kuhl & Iverson 1995) suggest that perceptual accuracy limits L2 production and that sound perception in L2 is largely determined by similarity of L1 categories and categories of the target language. All current models refer to the construct of *similarity* that is conceived as relative or gradual to account for category-specific differences in difficulty or ease in L2 acquisition. However, – though essential for an understanding of the acquisition of sound patterns in L2 – relationships of *intra-lingual similarity* within the category system of the target language are not explicitly considered.

This thesis argues that *intra-lingual similarity* between vowel categories of the target language L2 has a crucial role in explaining category-specific differences in ease or difficulty in L2 vowel perception. The current study refers to the observation that discrimination and identification of L2 vowels is not equally good or poor for all non-native categories but that there are gradual differences in the learners' success and ease or difficulty in L2 acquisition. These inter-category differences may be motivated by language-specific, typological or universal forces as discussed in chapter 1.

The major challenge for an experimental investigation though refers to the question how to operationalize similarity between German vowel categories in phonetic as well as cognitive terms and how to identify those factors that can account for the observed relationships of similarity between L2 categories as established by L2 learners and for differences in ease or difficulty in L2 learning.

The next chapters will adress the following issues: Chapter 3 will discuss notions of similarity in second language acquisition research and phonetics (in articulatory, acoustic and perceptual terms). Chapter 4 will discuss the phonetic bases of vowel sounds and fundamental notions and theories on vowel inventory typology. Chapter 5 will present a description of the German vowel system that is based on articulatory and acoustic characteristics and offers first insights into phonetic similarities between German vowel categories in articulatory and acoustic terms.

3 Similarity in Cross-language Speech Perception

Similarity is one of the oldest and most basic concepts in describing and predicting processes of second language acquisition. The concept mostly refers to relations between structures of the target language and the learners' native language that indicate the relative likelihood of a given L1 structure to be transferred to L2. The operationalization of this construct has nevertheless so far not received sufficient attention in SLA research. Therefore, one of the major aims of this study is to integrate contributions from SLA research, phonological theory, empirical phonetics and cognitive psychology with experimental evidence for perceptual similarity of vowels in L2 acquisition. The data analysis will focus on intra-lingual similarities within the system of the target language due to cross-linguistic influences as well as universal biases and individual factors that contribute to the listeners' perception of L2 vowel sounds.

In this chapter, different notions and dimensions of similarity and the many ways they may influence and constrain the learnability of L2 sound structures will be discussed. Section 3.1 discusses different types of similarity relations in L2 acquisition and relates the notions of similarity, transferability and learnability with each other, section 3.2 discusses dimensions of phonetic similarity, section 3.3 discusses the impact of acoustic-phonetic structures on the language-specific interpretation of speech sounds and the listener's phonological representations. In section 3.4, the major underlying assumptions of the present study will be summarized.

3.1 Notions of similarity in second language acquisition research

This section will discuss different types of similarity relations and their role in L2 acquisition. The following considerations are largely inspired by Jarvis & Pavlenko's (2008: 174ff) and Ringbom's (2007) general considerations on similarity and cross-linguistic influence in language performance and cognition for multilingual individuals, but will be presented here with special focus on questions related to the acquisition of L2 sound structures.

The relative ease or difficulty of learning a second language may be intuitively estimated in terms of *similarity*, *congruence* or *difference* between structures of the target language and the learner's native language.

Cross-linguistic similarity of two languages has been discussed under many different terms such as language distance, congruence, typological proximity, psychotypology, or cross-linguistic similarity (e.g. Kellerman 1983; Ellis 1994; Eckman 2004). Cross-linguistic similarity or “inter-lingual similarity” is considered to have a number of positive effects on processes of comprehension and learning such as facilitating meaning-to-form mapping for familiar or easier recognizable structures, increasing the rate of association with already known forms and enhancing the rate of learning. Similarity may however also constrain the learnability of L2 structures (see section 3.1.4 and 2.4.2).

3.1.1 Specific and overall similarity

The relative distance or similarity of the learner’s native language and the target language largely depends on structural differences and similarities as they actually exist in the two languages involved and on the language-specific or individual perception of these properties.

Overall similarity between languages has to be differentiated from *specific similarity*, i.e. the existence of similarities and differences in specific components or parameters of a given language pair such as the size of phoneme inventories or the existence of certain phonemes or phonemic contrasts or the use and combination of specific phonetic features and acoustic cues a given contrast is based on in the respective language.

Bradlow, Clopper & Smiljanic (2007) undertook an attempt to compare 17 languages with respect to perceived *overall similarity* and visualized similarity by a two-dimensional Multidimensional Scaling (MDS) analysis. Short speech samples were presented to listeners who were asked to rate them. The samples were then compared with respect to some structural, mostly acoustic phonetic features that were correlated with the subjects’ ratings. The post-hoc identification of the physical and psychological parameters determining the MDS outcome and accounting for the listeners’ judgements is one of the crucial challenges of such experiments. A crucial point for overall comparability that is not satisfactorily considered in Bradlow et al.’s study is the choice of parameters that are used to compare languages with each other. In this kind of studies, subjects usually make *intuitive* judgements about the relative similarity of language samples that do not necessarily refer to linguistic criteria such as inventory size or particular phonetic features. A reduction to only two dimensions related to the existence of specific segments in the input sample and the geographical distribution of languages as offered by Bradlow and colleagues can certainly not account satisfactorily for the complex and multi-dimensional interaction of the many phonetic-phonological and general typological parameters that determine the perception by listeners of a given language. To conclude, an estimation of perceived overall similarity of languages is an interesting complement to a structural linguistic analysis, though some important methodological considerations may have been overlooked in Bradlow et al.'s study (for discussion of interpretability of dimensions in MDS, see section 7.3).

3.1.2 Objective and subjective similarity *or* phonetics and phonology

Another important differentiation of notions of similarity refers to what is similar or different in objective terms in contrast to what is subjectively perceived to be similar or different by the learners. While *objective similarity* can be defined as the actual degree of congruence between languages or specific linguistic structures, *subjective similarity* refers to the degree of congruence the L2 learner actually perceives or believes to exist. To put it differently, we may distinguish differences and similarities in the *language system* from differences and similarities as *perceived, learned and represented by the language user*. Referring to the acquisition of a second language vowel system, objective similarities or differences can be primarily described on a *phonetic* level in objective terms, such as the articulatory or acoustic properties of L2 speech sounds, but have to be distinguished from the *phonology*-driven subjective interpretation of differences and similarities of L1 and L2 vowel phonemes as perceived by the language learner.

In an experimental setting, similarity of physical characteristics of *stimuli* described in objective phonetic terms must be distinguished from their subjective phonology-driven interpretation as indicated by the participants' *responses*.

While objective similarities are constant over time, subjective similarities are never static but are subject to temporary and highly individual hypotheses, reactions and decisions of the L2 learner that may be affected by developmental changes induced by the learner's progress in acquiring the target language (see section 1.3.2). These subjective similarities may be established by the learner *consciously* or sub-consciously.

Moreover, there may be hypothetic similarities that are only *assumed* by the learner without having actually encountered evidence for it (Ringbom 2007; Jarvis & Pavlenko 2008: 179). The assumed similarity between e.g. L2 vowel categories indicate the insufficient perception of specific cues for L2 contrasts and the instability of mental representations of the categories involved.

Jarvis & Pavlenko (2008: 178) mention subjective similarities resulting from three general conditions that are also relevant for the perception of L2 sound structures: (1) the L2 user's failure to recognize some of the objective similarities that actually exist, (2) the L2 user's misperception of the nature of some of the similarities that exist across languages, and (3) the L2 user's assumption of the existence of some similarities that actually do not exist.

A crucial issue for the present study is the fact that subjective similarities appear to be largely *asymmetric*, applying in one direction more easily or more frequently than in the other (Eckman 2004: 530f; Ringbom 2007).

3.1.3 Inter-lingual and intra-lingual similarity

Theoretical linguistic studies are mostly concerned with similarities and differences *between* languages, especially when dealing with linguistic structures from a contrastive or typological or universal point of view. An important aspect observed in many studies on the acquisition of a second language phonology, is that the mere description of *inter-lingual* similarities or differences as provided by a contrastive analysis of L1 and L2 cannot account sufficiently for how L2 learners develop mental representation for sounds of the target language. On the other side, the consideration of *intra-lingual* similarity relations of L2 structures and the perceived inter-category relationships between L2 elements is commonly largely neglected in SLA studies. Intra-lingual similiarities are as crucial for an understanding of developmental processes of language acquisition and for the establishing of mental representations in the target language as is the description of inter-lingual similarities between L1 and L2.

The learners' perception of inter- and intra-lingual similarities may differ language-specifically in their form and range and may also change over time as learners achieve more proficiency in the target language. Intra-lingual relations are assumed to be rather established by more advanced learners with more language experience, whereas inter-lingual similarity shows to be stronger with beginners (Ringbom 2007: 89ff).

To sum up, while most studies focus on the inter-lingual relationship of items in L1 and L2, the intra-lingual similarities of items within the target language receive hardly any attention. Therefore, a special focus of the present study are relations of intra-lingual similarities between vowel categories within the target language as established by L2 listeners in the acquisition of the target language. Intra-lingual relations of similarity are of course largely influenced by a learner's L1 but are moreover strongly determined by individual learning processes, i.e. by an individual learner's hypotheses and strategies that result from language-specific, universal and individual factors (see section 1.3).

3.1.4 Similarity, transferability and learnability in L2 acquisition

As pointed out in the previous sections, a crucial point in studying processes of acquisition and mental representation in L2 is the differentiation of *differences* and *similarities* between languages as well as the distinction of *objective* and *subjective* similarities and differences between the languages involved. Contrastive analyses mainly focus on *differences* between languages, whereas L2 learners are more concerned with *similarities* they can establish between the target language and their native language or any other known language that may facilitate their learning task (Ringbom 2007: 1).

Effects of cross-linguistic similarity on transfer of L1 structures to L2 can be *performance-related* or *learning-related*. Linguistic structures may differ in their probability to be transferred and their probability to be learned. While *transferability* rather refers to performance-related aspects, *learnability* refers to aspects related to acquisitional processes and changes in L2 competence. For a thorough description of L2 learning, of L2 performance and of the mental organization of L2 sound structures, inter-lingual as well as intra-lingual similarity relations have to be considered together with objective similarities of L1 and L2 sound structures.

Inter-lingual identifications rather emerge from subjective *similarities* as found by the learners, whereas subjective *differences* typically lead learners to try to *avoid* transfer (Jarvis & Pavlenko 2008: 179). In other words,

> "*subjective similarity affects the degree to which the learner relies on the source language when learning and/or using the recipient language, and objective similarity can affect the likelihood that the transfer (i.e. whatever transfer may have occurred) is positive or negative*" (Jarvis & Pavlenko 2008: 178).

Learners may of course also *avoid* certain L2 structures because they perceive these structures as difficult. Therefore, the empirical interest in the impact of cross-linguistic similarities and differences has to be expanded from mere transfer phenomena to other aspects of *cross-linguistic influence* in L2 performance such as over-use, under-use or avoidance of specific linguistic structures due to structural properties and their subjective perception by the L2 learners (see section 1.1.1).

"Similarity" is one of their central concepts in models of cross-language speech perception (see section 2.4). In theoretical and empirical works on second language acquisition, several though mutually often contradictory predictions concerning transferability, relative difficulty and learnability of L2 structures similar to L1 structures have been formulated: On the one side, similar structures are expected to be easier learned or acquired in L2 and on the other side it may be exactly similarity between L1 and L2 structures that hinders the acquisition of L2 structures[7] (see section 1.1.2 for a discussion of similarity in L2 acquisition and section 2.4 for models of cross-linguistic speech perception).

---

[7] It is necessary to emphasize that such predictions do not hold in the same way for all linguistic components.

Though "perceived similarity" may be at first sight a rather fuzzy concept (cf. Ringbom 2007: 7), it may be elucidated by a combined approach, i.e. (1) a more objective analysis of specific linguistic structures together with (2) a thorough investigation of the many ways in which perceived similarity may manifest itself in the learners' performance either in natural language use or in experimental settings. In the present study, the perceived similarity of L2 German vowels is based on an analysis of "objective similarities" of phonetic and phonological characteristics of vowels in the German target system. However, as will become evident in chapter 4 and 5, vowels vary in a number of phonetic (articulatory and acoustic) and phonological parameters. The analyst's choice of specific parameters for a description and comparison of L1 and L2 will again be determined by his/her subjective perception and interpretation of "phonetic reality".

The following section will discuss major dimensions of phonetic similarity that have to be considered when describing and comparing vowel categories of the native and the target language system. A discussion of articulatory and acoustic parameters for vowel description and the analysis of their impact in L2 vowel perception of German vowels will follow in chapter 4 and 5.

3.2 Dimensions of phonetic similarity

Overall similarity of languages or the similarity of specific sound patterns in two or more languages can be described on different levels and dimensions such as phonetics, phonology, or language typology.

A comparison of sound structures in terms of phonetics can be carried out in three different dimensions: (1) articulatory (gestural) similarities at the *articulatory phonetic* level, (2) acoustic similarities at the *acoustic phonetic* level, and (3) perceptual similarities at the *auditory phonetic* level.

Figure 3.1 shows a schematic representation of the speech processing chain: the signal is produced by a speaker (articulation), is transmitted to a listener (acoustics) and perceived by a listener (perception). The acoustic signal that can be described in objective acoustic terms (phonetics) is interpreted in subjective terms by the listener (phonology). In other words, the perception of same acoustic signal and more specifically of the same acoustic cues in the signal may vary from listener to listener.

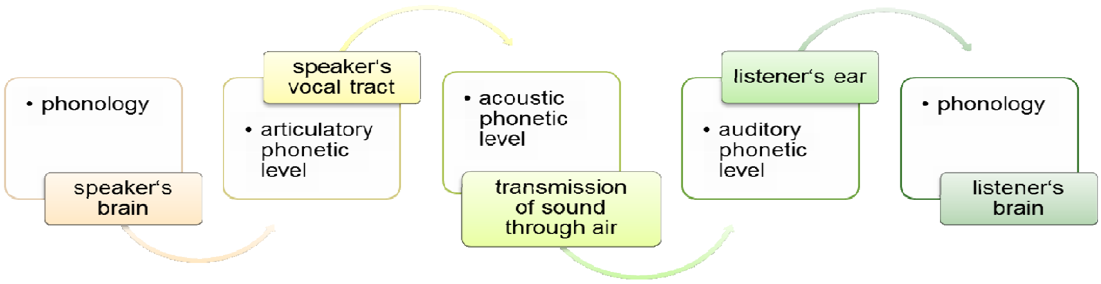

Figure 3.1: The interaction of phonetics and phonology in speech processing[8]

Acoustic phonetics never asks for the individual part that a listener actually takes in speech processing. Perception is a matter of phonology and may vary language-specifically as well as inter-individually. The listener's representation of the L1 phonological system and of the system of the L2 target language, his expectations and the subjective relationships of cross-linguistic similarities established by the language learner are the origin of most types of cross-linguistic influence in L2 acquisition. Moreover, we have to consider that the orthographic representation of sounds in the target language and the listener's native language may play a role in the acquisition and mental organization of L2 speech sounds. The following sections will briefly discuss different dimensions of objective and/or subjective similarity of vowels.

3.2.1 Articulatory similarity

Vowel sounds have often been described in terms of relative "articulatory targets" or acoustically in terms of relative "spectral targets". While an acoustic analysis can only indirectly reflect influences from articulators, e.g. lips, tongue, etc., the aim of articulatory descriptions is to describe the influence of specific articulators on acoustic and perceptual patterns (Wood 1979, 1986; Hoole 1999; Hoole & Mooshammer 2002).

Referring to cross-language articulatory similarity, several researchers (e.g. Best et al. 1988, 2001; Best & Strange 1992) describe phonetic similarity in terms of gestural constellations, i.e. the articulatory/gestural match between sounds, largely based on the framework of Browman & Goldstein's Articulatory Phonology (1992). The description of articulatory parameters for vowels may be more difficult than for consonants. In studies on L2 speech learning, acoustical analyses and results of direct assessment of perceived cross-language similarity are commonly provided, whereas measurements of articulatory parameters are

[8] Of course, the diagram represents a very simplistic model of speech perception disregarding the role of other components as grammar and lexicon or context and expectations in speech processing.

rarely presented. However, articulatory investigations are particularly relevant for the acquisition of L2 sound patterns, since the acquisition of perceptual and productive patterns of L2 occur in close interaction. To date, gestural comparisons of vowels in studies on cross-language speech perception are rather rare. Wood's (1979, 1986) work on the comparative description of vowels in different languages with special consideration of the constriction location is an interesting account that has not yet received sufficient attention (for Wood's articulatory-oriented approach, see section 4.7.6).

3.2.2 Acoustic similarity

The acoustic description of vowels usually includes spectral information (*formant frequencies*) and temporal structure (*duration*). Formant frequencies are generally used to situate vowel sounds in a two-dimensional (F1xF2) or three-dimensional (F1xF2xF3) space, based on physical units (frequency in Hertz) or on psycho-acoustical measures (Bark, Log, Mel). On the basis of these measures, relative distances between vowels can be specified within and across languages (Strange 2007: 39). Cross-language comparisons of vowel duration and spectral structure allow insights into the relative position of vowel sounds in the vowel space of the native and the target language. However, such descriptions represent a rather static approach. Dynamic aspects of the vocalic continuum and the crucial role of the transitional parts along with the influence of preceding and following segments are rarely considered. Thus, the description in a two- or three-dimensional space always represents a certain reduction of the full informational content of the speech signal.

Moreover, there is a considerable amount of variability among tokens of the same vowel category due to *contextual effects* (pre-vocalic and post-vocalic), speech rate, prosodic structure etc. In other words, cross-language spectral and temporal similarity is highly dependent on the phonetic and prosodic context a vowel is produced in.

Cross-language comparisons of vowel formant data usually involve acoustic descriptions of different speakers and are, at least in part, subject to the so-called *speaker normalization* problem which is due to the variability of formant frequency values caused by the individual shape of the speakers' vocal tract. Therefore, the transformation of the formant data into distances (e.g. F1-F0, F2-F1, F3-F2 in Bark) or ration measures is advisable to normalize vocal tract differences. However, as Strange (2007: 40) proposes, if the differences for the so-called "point vowels" /i/, /u/ and /a/ considered as defining the "extrema" of the vowel space are minimal, a direct comparison of the relative locations of intermediate vowels may be appropriate.

Strange (2007) describes two types of acoustic cross-language comparison for vowels that are common in most experimental studies: The first method is based on a comparison of the average formant frequencies of vowels produced in different phonetic and prosodic contexts in their *graphic representation*. Graphic representations such as an F1xF2 plot visualize information about language-specific patterns of spectral variation.

A *quantitative approach* to cross-language acoustic similarity may provide more reliable information about differences and central tendencies within and across languages. However, descriptions in terms of mere formant frequencies provide only limited information (see section 4.3.4). Therefore, Strange (2007: 45f) proposes the use of linear discriminant analysis to compare distributions of vowels in L1 and L2.

It is however necessary to emphasize that such comparisons are again only based on *acoustic* structures but cannot directly reflect *perceptual similarity* for an individual learner. A central issue in the current study is to what extent acoustic comparisons together with traditional phonological analyses of vowel sounds in L1 and L2 will predict perceptual assimilation patterns of L2 sounds.

3.2.3 The influence of orthography

A frequently underestimated factor is the influence of orthography in L2 category formation. It is well known that orthographical form creates the illusion of phonemic identity. For example, linguistically naïve speakers may estimate the quality of the two <e> in German "*Esel*" as equal due to their orthographical identity, although in Standard Austrian German one is pronounced as [e:] and one as [ɛ] and the second one my even be deleted due to syllabicity of the following lateral. Likewise, German natives may not be aware of phonetic differences between the two <b> in German "*Album*" and "*Kalb*", the last being affected by the process of final devoicing. The identical graphemic representation for different phonemes causes the illusion of phonemic and phonetic identity. Among literate adults such "illusions" about the identity of phones or phonemes and their orthographic representation frequently occur. Likewise, identical or similar graphemic representations of sounds in L1 and L2, e.g. of <u> - <ü>, <o> - <ö> or <ö> - <ä>, may give rise to perceptual "illusions" about L1 and L2 categories that will also have an influence on the acquisition of L2 (cf. Kerschhofer-Puhalo 2010a, 2010b, 2012, in press a).

3.2.4 Phonetic and phonological similarity

Strange's (2011) *Automatic Selective Perception Model* suggests two modes of online processing, a "phonological mode" and a "phonetic mode", and a different attentional focus either at a phonetic or a phonological level of analysis, explaining the complex interaction of linguistic experience and phonetic similarity relationships between L1 and L2 inventories. While the *phonological mode* of perception, which is used by adult listeners in the "bottom up"-process of recognizing L1 word forms, is fully automatic for L1 processing and usually ignores context-dependent phonetic variations, the *context-specific phonetic mode* involves the attunement to context-dependent allophonic details requiring access to the listeners' knowledge of phonetic and phonotactic patterns. The perception of word-forms is assumed to be accomplished by highly over-learned, automatic *selective perception routines* (SPRs), which are attuned to the most reliable information for detecting phonological segments and sequences in L1. Even under non-optimal listening conditions, these SPRs enable rapid and robust perception. The phonetic mode may be involved when listeners have to adjust to a very different "dialect" or variation of the language or in specific experimental tasks like "accent"-rating tasks or perceptual assimilation tasks, in which listeners are asked to make judgments about the "goodness" of L2 sounds with respect to L1 phonological categories. Strange's differentiation of phonological mode vs. phonetic mode parallels Wode's distinction of categorical mode vs. continuous mode of perception (see Wode 1990).

Strange (2011) assumes that at the beginning stages of L2 learning, a phonetic mode of perception is involved when processing phonetic sequences that include non-native segments or segments occurring in L1 and L2 with language-specific phonetic differences,. The phonetic mode requires more attentional focus, it is slower and can be less successful in non-optimal listening conditions. This is, however, partly in contradiction to other models (cf. Flege 1987, 1995; see section 2.4.2 and 2.4.5) assuming that the phonological mode dominates in initial stages of learning, where L2 sounds are perceived as equivalents to L1 categories, and that the successful acquisition of sub-phonemic details can only be performed by attention to phonetic differences.

It is moreover interesting to observe that in discrimination tasks, where the inter-stimulus interval (ISI) is relatively short, listeners can detect phonetic differences that they would not differentiate under "normal" conditions in L1 speech processing. Strange (2011) therefore distinguishes *automatic processing* and *attentional processing* of speech along a continuum of "effort" involved in the detection of relevant information that is requested in the specific task.

To conclude, perceptual similarity between vowel sounds depends not only on intrinsic characteristics of the signal but also on the language-specific warping of the perceptual space, the listener's general language knowledge and proficiency and on a specific conditions in a given speech situation or experimental setting.

Speech perception can be characterized as an active process of purposeful information-seeking. Natural language processing is a process that proceeds mainly in a *top-down*-direction, i.e. listeners are, driven by their expectations about the content of the linguistic message and the speaker's intentions, seeking sufficient information from the continuous speech signal to identify the string of words in an utterance and to decode the message encoded by the speaker. In perceptual experiments, however, listeners are mainly involved in *bottom-up*-procedures that differ from processes of online processing of continuous speech input. Therefore, experimental results and their relevance for "real-world" speech processing in L2 have to be (re)viewed very critically.

3.2.5 Perceptual similarity and difficulty in L2 acquisition

In models of L2 speech perception (see section 2.4), the *relative difficulty* of accurate perception and production of L2 sounds is predicted or explained as largely dependent on the *perceived similarity* of L1 and L2 sound categories. A contrastive analysis may help to predict or explain the relative difficulty in L2 perception but has to consider detailed phonetic data (vowel-specific description of formant frequencies, dynamic spectral variation and duration as well as variation due to speaker, context and prosodic structures) to account for articulatory, acoustic and auditive similarity. While nowadays most studies on L2 speech perception provide acoustic analyses of L1 and L2 elements, the availability of comparative articulatory data is very restricted. A thorough comparison of the phonetic structure of sounds in L1 and L2 has to include (1) articulatory (gestural) similarities, (2) acoustic similarities, and (3) perceptual similarities by direct experimental assessment (Strange 2007: 38ff).

A description of acoustic similarity of L1 and L2 sounds is an essential contribution to perceived similarity but will never be sufficient to predict patterns of cross-language perceptual assimilations or inter-category confusions. To assess cross-language *perceptual similarity* patterns, several techniques have been used in experiments with beginners and advanced learners of L2 (for an overview see also section 2.3 and Strange 2007). Conclusions on cross-language perceptual similarity can be derived from experimental investigations, e.g. by testing listeners' identification of L2 sounds with L1 or L2 categories or from assimilation patterns of L2 sounds to L1 categories. Goodness ratings as used in the Perceptual

Assimilation paradigm (see section 2.4.3) may reflect the listeners' ability to discriminate between non-native contrasts and their awareness of the "native-likeness" of L2 sounds. Obviously, the nature of the input stimuli (pre- and post-segmental contextual variation, citation forms vs. words in sentence context) exerts an influence on the output of such assimilation tasks. Therefore, conclusions about cross-language perceptual similarity will be more reliable, if segmental variation due to context, prosody, speech rate etc. is included in the analysis (Levy 2004; Strange 2007). It is important, however, to realize that experiments based on data from citation forms (word lists, non-words) or on data without contextual variation will never provide a complete picture of cross-language perceptual similarity (Strange 2007; Bohn & Steinlen 2003).

In perceptual *assimilation tasks*, subjects are asked to categorize L2 utterances in terms of L1 categories represented orthographically or by key words. Optionally, subjects are asked to rate their judgments (*goodness-rating*). A similar technique is used e.g. by Guion, Flege, Akahane-Yamada & Pruitt (2000) or Hentschel (1985).

The *transcriptional approach* is another way of assessing perceptual similarities: Subjects are exposed to non-native phones and are asked to use L1 orthographical labels to write down their perception of the L2 sound (the use of diacritics and verbal descriptors being allowed, e.g. Best et al. 2001 for Zulu consonants). In terms of the Perceptual Assimilation Model (Best 1995; Best & Tyler 2007, see section 2.4.3) such results could be interpreted as follows: The use of identical orthographic labels can be interpreted as an instance of Single-Category Formation and will thus reflect significant difficulties in perceptual discrimination. The use of diacritics and descriptors within a common label may reflect instances of Category-Goodness Assimilation that are expected to be most difficult in discrimination. In instances of a Two-Category Assimilation pattern, i.e. the use of two different orthographic labels from L1, accurate discrimination is expected. However, no quantification of the *degree of category goodness* is possible within this approach, though the degree of category goodness is of special relevance for vowels that are generally not perceived categorically; even the transcription of native vowels may be thus difficult for naïve listeners. Still, the approach is problematic for languages with a non-transparent or even inconsistent orthographic representation for vowels as e.g. English (Strange 2007: 48).

Flege and colleagues have used more quantitative techniques. Flege, Munro & Fox (1994) used a *paired-comparison technique*, i.e. a *dissimilarity rating* design for vowels where subjects were asked to rate the dissimilarity of vowel pairs with a 9-point Likert scale. Mean dissimilarity ratings of 405 judgments were then computed for each pair of Spanish and

English vowels. This technique does not require response labels, it does not involve orthographic representations and allows for further ordering of L2 contrasts in terms of degree of category goodness (cf. Strange 2007: 49). One disadvantage of this method is the large number of trials necessary to account for phonetic variation due to context and prosodic structure. For languages with a large vowel inventory, this approach is especially problematic. Another disadvantage is that subjects do not compare L2 phones against their own internal representations of phonetic categories. Rather, the experimental design requires a comparison of specific tokens of L1 and L2 phones. Thus, the results are not predictive enough for the listeners' perceptual assimilation patterns since the L1 productions may not correspond to their own internalized phonological category distributions.

To gain insights into perceived similarities within the German vowel system, the current study will use category confusion data from a perceptual identification task (see section 6.2).

3.3 Similarity, salience and difficulty in perception

3.3.1 Perceptual salience and phonemic contrast

From an auditory perspective, *perceptual similarity* could also be defined as lack of perceptual distinctiveness of L2 contrasts due to the learners' experience with L1 and L2. The detection of contrast between categories is crucial for the acquisition of segmental representations of sound categories in first and second language acquisition. The insufficient knowledge about distinctive characteristics of L2 sounds distinguishing vowels in a contrast pair explains why some vowels are perceived as particularly "similar" or even "equal". The insufficient discrimination of vowels in a contrast pair is responsible for instable perceptual categories which may result in *perceptual substitutions*, i.e. a sound event belonging to a specific sound category of the target language is misperceived or misinterpreted as belonging to another sound category from L1 or L2.

In perception, some distinctive properties of sounds are considered to be more salient than others. In a general sense, the notion of *salience* refers to a higher degree of distinctiveness of a given sound or feature or acoustic cue facilitating its detection in perception and the stability of contrasts a sound is involved in (Lindblom 1986; Flemming 2004). However, relative *perceptual salience* of speech sounds depends not only on signal-inherent properties of the stimuli but also refers to the listener's language experience. Referring to physical characteristics of vowels, perceptual salience is not only associated with loudness or intensity of the signal. Other characteristics that are assumed to enhance acoustic salience are degree of aperture (identifiable by F1 or F1-F0), a center of gravity effect of formant frequencies vowel-intrinsic duration or intrinsic pitch (Stevens 1989, 1998). These assumptions refer to the salience of specific cues within a given sound. However, as the discussion of phonemic vowel contrasts and their acoustic correlates as well as the analyses of experimental data will show, the identification of those acoustic cues that determine a given sound's salience is a challenging task.

Referring to the relative *salience of a contrast*, Burnham (1986) introduced the valuable notion that phonetic contrasts can vary on a continuum from "fragile" to "robust". Salience of contrast may account for differences in the relative difficulty that naïve listeners and L2 learners have in acquiring the ability to differentiate non-native contrasts. Describing developmental changes in speech perception, Burnham (1986) distinguishes *fragile* contrasts, which are lost in infancy due to *lack of exposure* to correspondent sounds, from *robust* contrasts, which are lost later around initial stages of (formal) language learning and training due to *lack of experience* with phonologically irrelevant contrasts.

Strange (2010) proposes an interesting distinction between two types of contrast salience, *auditory salience* and *perceptual salience*, that determine, at least in part, naïve adults' and late L2 learners' perceptual difficulties with non-native contrasts among consonants and vowels:

Auditory salience refers to the magnitude of the physiological response to a change from one lexical segment, tone or sequence of segments to another. This type of salience is described as being a function of language-universal characteristics of human listeners. While physiological responses are clearly observable with this type of salience (ERPs such as MMN), behavioral indices remain relatively unchanged by long-term linguistic experience. Differences in response rate or response accuracy vary between different phonetic contrasts as a function of physical differences between different phonetic contrasts rather than as a function of the listeners' language background (e.g. acoustic differences between /r/ and /l/ in American English, primarily F3 transition onset and direction, are auditorily relatively unsalient cues, which are difficult for Japanese as well as for American listeners when presented as non-speech).

Perceptual salience, on the other hand, is described to vary as a function of linguistic *experience* (Strange 2011). In L1, speech perception can be described as a purposeful information seeking activity based on highly over-learned *selective perception routines*, by which listeners detect the most reliable cues to identify phonetic segments and sequences (Strange 2010). These selective perception routines from L1 do not function adequately in L2 due to language-specific sub-phonemic differences and insufficient experience with L2 sounds. Listeners' linguistic experience determines the "strength" of behavioral and physiological responses to phonetic contrasts. The perceptual salience of a stimulus might be indexed by cortical ERP components like MMN or P3, and may be interpreted as being due to "neural commitment" (Kuhl et al. 2008). In experimental tasks, the relative perceptual salience of phonetic contrasts as a function of long-term linguistic experience is better reflected in behavioral perceptual tasks that are more demanding in terms of stimulus structure and memory constraints (Strange 2011).

For the present study we can hypothesize that relative perceptual *difficulty* of a given L2 contrast is expected to depend on the auditory salience of specific acoustic cues as well as on the perceptual salience determined by the listeners' language-experience, i.e. their L1 (for a detailed discussion of language-specific selective attention routines from L1 and their role in L2 learning, see also section 12.5). Some types of contrasts are expected to be easier to perceive and to acquire than others. These differences may be partly due to the higher salience of the sounds and acoustic cues involved. Perceptual salience is defined as the experience-driven sensitivity to acoustic cues indicating phonemic contrasts.

3.3.2 Similarity, phonemic contrasts and language-specific representations

A full account of the phonological acquisition of a vowel system in L2 must comprise a description of L2 perception. Each language uses only a reduced set of the full range of possible sounds that can be produced by the human vocal tract. The multi-dimensional vowel space can be described under an articulatory, an acoustic or a perceptual view.

The *multi-dimensional perceptual space* can be compared to a *cognitive map* that is a substrate of general auditory-perceptual capacities underlying speech perception on which a metric of proximity or similarity is defined (Pierrehumbert 2003: 128). Assuming a language-universal perceptual map may account for common tendencies and preferences in the evolution of language-specific sound systems (cf. Steriade 2001b). In other words, the "universal" perceptual vowel space can be assumed to be shaped by exposure to a particular language in early childhood resulting in a kind of "neural commitment" to the specific acoustic properties of this language (see section 2.2.1).

The acquisition of the phonemic contrasts of a language involves the establishment of *categories*, i.e. *labels* over the perceptual map, and the association of these labels with probability distributions over the space defined by the map (Pierrehumbert 2003: 128).

Learning the phonetic patterns of a language means learning the probability distributions over the parametric phonetic space. To master such probability distributions for sound categories, a considerably high amount of language experience must be available to the listener. The initial perceptual space is thus *warped* towards the specific conditions in a given language.

A crucial question refers moreover to whether the *peripheral auditory sensitivities* as such vary from language to language or whether a substrate of auditory-perceptual capacities underlying speech perception which is the same for all languages may be assumed. It is commonly assumed that the auditory sensitivity to language-specific patterns in speech, i.e. the universal perceptual space as such, is altered or "*warped*" in L1 acquisition (Kuhl 1991, 2000b; Iverson et al. 2003). However, language-specific perceptual sensitivities also arise from other dimensions of linguistic knowledge, e.g. from lexical knowledge. This commitment to the particular system of the native language, however, results in greater effort and less efficiency in foreign language processing (Zhang et al. 2005). By acquiring a writing system and orthographic representations of a particular language, the categorization of L1 sounds is further refined and settled towards the specific characteristics of a given language according to their linguistic and orthographic norms (Kerschhofer-Puhalo 2010b, 2012 and in press a).

Strange (2011) makes a gross distinction of three types of memory: (1) short-term memory *(traces)*, (2) long-term memory *(representations),* and (3) procedural memory *(perceptual routines)*. Current models of speech perception describe representations in terms of *phonological categorie*s, which are stored in long-term memory and which correspond to the learner's *knowledge* about the set of contrastive elements and their relationships in a given language. *Short-term memory traces* can be conceived as the product of habituation processes. *Procedural knowledge* reflects learned routines for detecting and attuning to task-relevant information specified by acoustic patterns.

Results and strategies in perceptual experiments are conditioned largely by the task itself and the input-quality. The inter-stimulus interval (ISI), for example, determines whether listeners can detect phonetic differences or not. When longer intervals are placed between stimuli, short-term memory traces are of little use and listeners can only rely on their abstract phonological representations.

3.3.3 Similarity, perceptual salience and asymmetry

A fundamental assumption of the present study is that errors in L2 vowel identification do not occur randomly but systematically and that vowel categories are supposed to be not equally “distant” or “different” from each other. Rather, some vowels are perceived as more “similar” to each other than others. Vowels that are perceived as “similar” are considered to share several common properties. Vowels that are perceived as “different” are supposed to differ in more such properties. On the basis of such common articulatory or acoustic properties it is possible to formulate hypotheses on the probability of perceptual confusions (for feature-based approaches to similarity and confusion probability, see section 12.2.2.2 and 12.2.4).

In language acquisition as well as in sound change, we find patterns of *directional asymmetry* in phonological behaviour. While a sound change or sound substitution in one direction is frequently occurring, the reverse may be rarely attested (Ohala 1993b; Chang, Plauché & Ohala 2001; Mielke 2002; Kaun 2004; Polka & Bohn 2003, 2010). Such uni-directional asymmetric confusions also occur in perception studies, in identification tasks as well as in discrimination tasks where subjects are asked to judge the members of a stimulus pair as “same” or “different”. The empirical results of the current study will provide ample evidence of asymmetry in perceptual confusion of L2 German vowel categories (see section 11.3).

Asymmetric pattterns have also been observed for category goodness ratings, i.e. judgments of “good” or “poor” exemplars of a category: The discrimination of differences among good exemplars of a category is found to be worse than between poor vs. good exemplars, corroborating the assumption that categories are organized around an ideal, prototypical representative of a phonetic category (Kuhl 1993; Kuhl & Iverson 1995; Best & Tyler 2007).

Polka & Bohn (2003, 2011) present evidence for patterns of asymmetry in vowel discrimination from several experiments with infants and adults: A frequently occurring pattern shows that changes from a “less peripheral” vowel towards a “more peripheral” one are easier to discriminate than vowel changes in the other direction. Directional asymmetries are also evident with adult listeners in L2 for vowel contrasts that are non-phonemic in their native language (Polka & Bohn 2011).

Polka & Bohn (2003, 2011) refer to “*peripherality*” as the relative location within the vowel space defined in terms of the two dimensions of vowel height (F1) and frontness/backness (F2). As the authors admit (Polka & Bohn 2011: 8), it is obvious that a description of vowel contrasts only in these two dimensions is insufficient and that the F1xF2 space cannot account for all vowel perception asymmetries (for discussion of the notion “vowel space” see section 4.7 and 12.1, for a critical discussion of “peripherality”, see also 12.3.3.1).

In a vowel discrimination task, Kerschhofer (1995) observed that Polish learners of German could discriminate the rounding contrast in short front mid vowels easier when the “unfamiliar” vowel /œ/ was presented first in the stimulus pair (only 9% wrong in /œ/ - /ɛ/)

than when it was presented in reverse order (39% wrong in /ɛ/ - /œ/) (for similar effects, see also Barry 1975).

These results show that there are *directional asymmetries* within a contrast pair: Perceptual confusion or substitution is more likely to ccour in one direction than in the other. Evidence for such asymmetrical patterns of phonological behaviour is found in first language acquisition, in synchronic substitutions, in diachronic sound change, in loanword phonology, and in foreign language acquisition, where two members of a pair of confused sounds can be considered as structurally *similar* but one member has an additional feature that the other member lacks. The listener is more likely to miss a distinctive part of the stimulus (e.g. rounding in front rounded vowels) than is to imagine its presence when it is actually not present. This effect is clearly observable in discrimination tasks, but we may expect similar effects in a vowel identification task as well.

The reasons for these asymmetric patterns may be signal-inherent as well as listener-related. In natural speech, temporal and spectral properties of acoustic events function as *cues* for sound differentiation. The intrinsic *salience* or *robustness* of these acoustic cues may vary along a continuum. Some acoustic cues are auditorily more salient and robust in perception than others. Irrespective of their intrinsic salience, all cues may be affected by some degree of degradation in transmission, but robust cues are less likely to be compromised to the point of losing their distinctiveness whereas non-robust cues may be degraded to the point where they lose their distinctiveness (Chang, Plauché & Ohala 2001: 80). Chang et al. (2001) argue that this is the reason for so-called *asymmetries* in perceptual confusions as observed in confusion matrices.

The reverse behaviour of erroneously “adding” a nonexistent cue into the speech signal is not expected to occur by Chang et al. (2001), but we know that it happens in language acquisition when learners are led by “wrong” hypotheses and strategies like hypercorrection or avoidance (see section 1.3.2). In L2 acquisition, we expect that – especially with intermediate learners who already have some language experience in L2 but have not yet established robust vowel categories – the imagined perception of a feature in spite of its actual absence may also occur. This type of imagined perception can be interpreted as a kind of hypercorrection, i.e. a strategy that is developed by L2 learners as a reaction to prior errors (see section 1.3.2 and 12.5).

The identification or discrimination of vowels in a foreign language could be described as a special case of “degradation in transmission” by which the discrimination of a given contrast becomes more difficult (García Lecumberri, Cooke & Cutler 2010). One major reason for errors in L2 perception is the imperfect knowledge of the target language together with weaker perception routines for sound categories and their distinctive acoustic cues. In minimal pairs, we can expect that a non-robust "extra" cue distinguishing the members of a minimal pair might be degraded to the point of being lost, resulting in a perceptual confusion of the two members in the pair. According to this view, the "extra" non-robust cue is an additional feature that gets lost in transmission. García Lecumberri et al. (2010) explain *asymmetry* in perception by perceptual "loss" of a specific non-robust cue that has a distinctive function in L2 but not in L1. However, as will be argued in section 12.4, asymmetries are not only due to weakness or loss of specific cues but rather by the language-specific setting of selective attention to acoustic cues.

From a theoretical perspective, two important questions arise with respect to the reasons for such patterns of perceptual asymmetry: (1) Is it the "loss" of a specific "additional" cue or feature that is responsible for asymmetry in confusion? (2) Or is the asymmetrical pattern due to markedness connected with the relative weakness of the cue and the (in)frequency of occurrence of a particular sound? Patterns of confusion asymmetry and unidirectional substitutions could be interpreted as a way of favouring the unmarked. Can we expect that effects of frequency and markedness are present even at the level of phoneme identification?

External evidence from perception in second language acquisition as well as from historical sound change and loanword phonology could provide answers to this question. Chang et al. (2001) favour the explanatory primacy of the relative salience in terms of acoustic/auditory cues over markedness in consonant confusion. However, it is argued here that in L2 acquisition it may rather be *perceptual salience* due to selective attention (see Strange 2011) than auditory salience of specific cues in the signal that determine language-specific or more general patterns of asymmetry in vowel perception.

Moreover, *frequency effects*, i.e. the relative frequency of occurrence of a given sound or sound contrast in the learners' input, have to be considered to shape the learners' knowledge about the vowel system of L2. Frequency of occurrence therefore has to be considered as a central explanatory factor in L2 vowel identification (see section 5.5 and 12.3.3.2). The relative frequency does have an effect on the perception in L2 insofar that frequency patterns influence the learner's knowledge about the L2 as well as his/her expectations. Vowels that are absent in the learners' L1 and that are less frequently occurring in the target language and/or the learners' input will be more difficult to discriminate and to identify correctly. The reason for these frequency patterns may lie in language-specific and/or in language-universal preferences due to relative auditory salience of intrinsic properties of the sounds involved.

To summarize, the present study is based on the assumption that both, auditory salience and perceptual salience of acoustic cues together with influence from L1-patterns, *frequency patterns* in the given input and *universal preferences* for certain sound structures can account for the susceptibility of L2 sounds for perceptual substitutions. While *perceptual salience* is mainly driven by listeners' language experience, *auditory salience* may be grounded in characteristics of the human vocal tract and the human peripheral auditory system.

3.4 Conclusions

To conclude, the following underlying assumptions and hypotheses on the perception of vowel contrasts in L2 and the L2 learners' performance in a vowel identification task can be formulated:

Vowel sounds differ with respect to several acoustic cues that serve to differentiate them in speech perception. Some vowels are phonetically more similar to each other than others. *Phonetic similarity* refers to shared articulatory and/or acoustic characteristics that may in perception diverge in attentional weight. In other words, some acoustic cues are more salient than others. Languages do not only differ with respect to number and quality of distinctive vowel phonemes but also in fine sub-phonemic phonetic details and the relative perceptual weight of specific acoustic cues.

The perceptual distinctiveness of a given vowel contrast does not only depend on specific properties in the signal but also on the constitution of the *vowel system* as a whole and the number and quality of "similar" vowels in the system under consideration. Vowel categories and vowel contrasts of L2 that do not exist in L1 are expected to cause difficulties in L2 acquisition. If both members of a vowel contrast are elements of the learner's L1, correct identification may be relatively easy. If one member of the contrast does not exist in L1, difficulties in perception will arise. However, inter-category similarities and differences in so called "fine phonetic details" between an L1 category and an equivalent L2 category may cause difficulties in correct identification. Due to perceived similarity, "new" L2 sounds may be perceptually assimilated to "similar" L1 categories (Flege 1987, 1995). These equivalence classifications due to similarity explain insufficient inter-category differentiation and perceptual confusion between categories of L2.

Perceptual difficulty is conceived here to be relative rather than absolute. There are no absolutely difficult sounds. Perceptual difficulty rather refers to items in *contrast*. The relative similarity of vowels in a contrast pair and the relative perceptual salience of specific cues in the signal will decide how difficult the perception of a given contrast is.

In L2 acquisition, some vowel categories are expected to be more stable in perception and representation than others, i.e. they are better differentiated from other categories in discrimination tasks and are more likely to be correctly identified in identification tasks. When perception is at risk, L2 learners will tend to confuse L2 categories and will fail to discriminate particular L2 vowel contrasts sufficiently. More "stable" categories, e.g. the point vowels /i u a/, will be acquired earlier than less stable categories. The relative stability of vowel contrasts however varies language-specifically.

Patterns of perceptual confusion can be explained by the relative proximity or similarity of the sound categories involved. Similar sounds share some common properties while less similar sounds differ in more such properties. Vowel contrasts may differ in perceptual distinctiveness or inter-category similarity. The discrimination of similar vowels will be more difficult due to the larger number of shared properties and/or due to the lower perceptual salience of those cues that differentiate the members of a contrast.

In vowel identification tasks, incorrect identifications due to perceptual confusion between categories may occur. With some vowel categories, perceptual confusion is more likely to occur than with others. *Category confusion* is expected to occur especially if (a) the members of a contrast pair share several common features, i.e. if they are "similar", (b) if the number of "similar" vowels within a given class is relatively high (or higher than in L1), and (c) if the members of a contrast are differentiated by less salient cues.

Salience refers to *auditory salience*, i.e. by general acoustic-phonetic factors determining its potential "loss" in transmission (Chang et al. 2001), as well as to *perceptual salience*, i.e. the language-specific use of specific cues in the learner's L1 and the listener's automatic selective attention routines (Strange 2011). Perceptual substitutions in L2 can partly be explained by "loss" during speech transmission of specific (additional) non-robust cues that have a distinctive function in L2 but not in L1. If a perceptual cue functions distinctively in L1, it is more likely to be used to differentiate contrasts in L2. The acoustic definition of perceptual salience for vowels remains however difficult. It is not only determined by specific characteristics of the signal but also by *listener-specific weighting* of cues in the signal. The listeners' prior linguistic experience in L1 and L2 conditions the detection of the relevant cues.

Together with acoustic properties of L2 sounds, the *orthographic representation* is expected to have an influence on the perceived quality and similarity of certain L2 categories and on the acquisition of L2 categories in language learning. Orthographic similarity such as <ü> and <u> is expected to contribute to category confusion.

Perceptual confusions do not occur at random but are considered to be systematic due to *perceptual substitution processes* (see section 1.2.4). Some contrasts are more likely to be affected by substitution processes than others. In other words, such contrasts are more difficult to differentiate. In case of insufficient differentiation, a given input may be incorrectly substituted by another category.

Moreover, *asymmetry* and *biases* regarding the direction of perceptual substitution processes is expected. *Biases* towards one specific category of a contrast pair are expected. Patterns of asymmetry and bias in perception frequently match cross-linguistic tendencies and preferences observed in sound change or first language acquisition. Some vowel categories are expected to be more preferred as perceptual targets, i.e. they are more likely to function as substitutes for similar vowel qualities than others. In an identification task, these categories will be preferred as response options, while others will be less frequently selected. In other words, members of a less preferred category are expected to "assimilate" perceptually to other categories.

The analysis of patterns of perceptual confusion between L2 vowel categories has to consider the existence of asymmetry in perception as well as the possible causes for these asymmetric relationships. It is hypothesized here that these specific perceptual patterns are physiologically and acoustically conditioned and are due to a universal preference for vowels with more robust acoustic cues. These universal preference patterns are embodied in the knowledge of the speakers as well as in the linguistic norms of a given language (cf. Dressler 1985: 292ff). Moreover, the relative frequency of occurrence of a given vowel contrast in L2 will influence the rate of acquisition of the vowel categories involved. The frequency of occurrence might be related to the typological preference or dispreference for a vowel sound in the languages of the world.

The discussion in the present and the following chapters will show that for a model of vowel perception in L2 a mere description of vowels in phonetic terms is insufficient. A phonetic analysis of vowel categories as will be offered in chapter 5 cannot provide a full account of the highly complex interaction of language-specific, individual or language-universal phonetic and phonological factors that have an impact on processes of vowel categorization, language acquisition and the mental representation of L2 vowel categories. The following chapter will discuss properties of vowels and vowel systems in terms of language-specific and typological cross-language comparisons. A discussion of the interaction of phonetic similarity and psychological similarity between vowel categories illustrated by evidence from experimental data of the present study will follow in chapter 12.

4 Vowel Sounds and Vowel Inventories

This chapter presents phonetic and phonological definitions of vowels, discusses the major phonetic and phonological dimensions in which vowel sounds differ articulatorily and acoustically and recapitulates the most prominent theories of vowel inventory typology.

4.1 The study of vowels in L2 acquisition

The acquisition of L2 vowel contrasts, their correct perceptual identification and the categorization of L2 vowel phonemes is a necessary condition for speech comprehension and comprehensibility in L2 that have received considerable scientific interest. The perception of L2-specific phonetic details and the acquisition of adequate long-term representations for L2-vowels are essential for the development of a native-like pronunciation in the target language. Difficulties in the perception of L2 vowel contrasts explain a significant part of the problems learners have in acquiring the sound categories of L2 (Flege 1995, 2002; Strange 2007). Studies on L2 speech intelligibility moreover demonstrate the high importance of native-like vowel pronunciation for L2 learners' speech comprehensibility and for the acceptability of L2 learners' speech production (Hirschfeld 1994; Bent, Bradlow & Smith 2007).

While the aim of most studies on L2 vowel perception is to demonstrate the language-specific perceptual mapping of acoustic-phonetic properties, studies on L2 vowel production generally focus on the description of spectral and/or durational characteristics of the vowels produced by learners of L2.

Several studies have focused on the assumption that some aspects of perceptual sensitivity in non-native speech discrimination and identification seem to be related to differences and *similarities* between *non-native* sounds and *native* speech patterns (e.g. Scholes 1967; Rochet 1995; Best et al. 2003; Escudero 2005), whereas others may reflect language-universal preferences in speech perception or production (Eckman 1977, 2008; White 1987; Major 1987a, 2008; Major & Kim 1999).

Additionally, several studies examine the influence of learner-related factors such as age, gender and others on success and learning rate in L2. The effect of L2 experience indicated e.g. by the length of residence in a country, where the target language is spoken, the age of learning or the age of arrival, and the effect of language experience on a native-like vowel production has received particular interest in empirical research (see chapter 8). However, studies in L2 vowel production and perception show that the learners' language background, the L1 phonemic inventory, its size, and phonetic features of L1 are substantially more

influential than individual factors as age of learning or length of residence (see e.g. Bohn 1998; Major 2008; Iverson & Evans 2007; McAllister, Flege & Piske 2002).
The languages of the world show an enormous variety of language-specific constellations of vowel sounds in their phonemic systems. However, vowel inventories apparently show recurrent patterns of *preferences* for some particular vowel qualities or "areas" in the so-called "vowel space". These common patterns, as will be discussed in the following sections, are largely due to acoustic and articulatory facts that exert an influence on the distribution of vowel sounds across languages as well as on common patterns in vowel perception with learners with different L1 language background.

> "*... because of the (genetically-transmitted) physical abilities and limitations of human speakers, some combinations and sequences of phonetic features are more difficult than others, and the substitutions that speakers make (in the mental processing of their speech) to ease these difficulties represent natural processes (...). The processes are there because the phonetic difficulties inherent in the speaker's physical makeup are there, and because the brain is capable of the kind of planning that can overcome such difficulties.*" (Donegan 1985: 26)

The basic articulatory and acoustic-phonetic characteristics of vowel sounds and theories on vowel system typology together with their relevance for phonemic distinctions in vowel systems will be discussed in this chapter. A description of the German vowel system which is based on these general considerations will be presented in chapter 5.

4.2 The definition of vowels – a functional phonetic approach

Most definitions of vowels comprise phonetic as well as functional aspects (cf. Pike 1943; Chomsky & Halle 1968; Ladefoged 1982; Ladefoged & Maddieson 1996: 281ff; Stevens 1998: 257ff) and refer to two major characteristics of vowel sounds: (1) Vowels are produced with no major constriction or occlusion in the vocal tract[9], and (2) they are syllabic, i.e. they function as syllable-centers. Vowels occupy a central position in human speech processing. Higher intrinsic intensity and duration of vowels allow vowel perception over longer distances than consonants and contribute to a better perception within words compared to consonant perception (Browman 1980), a fact that is consistent with the function of vowels as syllable centers. Vowels carry a significant amount of intrinsic phonetic information. The function of vowels as syllable nuclei, bearing lexical stress and rhythmic structures, is crucial

9 Beyond a certain degree of constriction in the vocal tract the sound produced is not a vowel anymore.

for word recognition (Cutler 2005). Prosodic contours bearing grammatical as well as emotional information are also conveyed via the continuous vowel signal.

From an articulatory point of view, vowels can be characterized as relative articulatory "targets" or gestural constellations (Strange 2007: 38). These typical articulatory constellations or "targets" may vary language-specifically. In terms of articulation, vowels differ in the positions of active articulators (jaw, tongue body, lips), the place of narrowing or "constriction" in the oral cavity, the degree of narrowing or constriction, and in additional gestures, e.g. gestures of the lips for rounded vowels or of the velum for nasal vs. oral contrasts.

Acoustically, vowels differ in several parameters such as formant frequency patterns, duration and fundamental frequency changes (Stevens 1998). Different articulatory constellations yield specific resonances in the vocal tract that are interpreted as realizations of a specific vowel category by listeners of a given language. However, the same acoustic output may be the result of different articulatory constellations. Therefore, a linear relationship between articulation and acoustics cannot be assumed.

The acoustic manifestation of vowels is commonly described in terms of formant frequencies and duration. One of the attributes that distinguish vowels from consonants is the spectrum of vowels that is characterised by several prominent peaks particularly in the mid-frequency range of about 800-3000 Hz[10] (Stevens 1989: 30). Changes in the frequencies of the first two or three vowel formants result in changes of perceived vowel quality.

Perceptually, and more specifically in terms of auditory categorization, the same phonetic input can be interpreted differently by listeners from diverse native languages due to language-specific differences in the categorization of vowel sounds. In other words, the acoustic cues involved in sound production and perception may contribute differently, i.e. language-specifically, to sound categorization. Variation in specific ranges of formant frequencies cause differences in perceived vowel quality but their distribution and phonological interpretation depends on the listeners' language background and experience.

[10] Stevens (1989: 30) admits that a spectral peak for F1 can also occur below this frequency range.

4.3 The phonetic and phonological description of vowels

Phonological descriptions of vowel sounds in the world's languages commonly refer to three basic dimensions or *major vowel features*:

(1) high – low or *Height*,

(2) back – front or *Backness*, and

(3) rounded – unrounded or *Rounding* referring to lip compression and protrusion.

While Rounding is commonly described as a binary feature, Height is frequently represented as a multi-valued feature. Frontness/Backness can be conceived as either a binary or a multivalued feature (cf. Ladefoged & Maddieson 1990: 94, 102).

In addition to these major features, Ladefoged & Maddieson (1990: 103f) describe *minor features* of vowel quality such as

(1) nasalisation, the most common minor vowel feature,

(2) features that involve special gestures of the tongue and associated structures (advanced tongue root (ATR)[11], pharyngealization, stridency, rhotacization, and fricatives),

(3) features that involve different phonation types (voiceless, breathy, laryngealized, or creaky vowels), and

(4) features that involve differences in the time domain, producing variations in length and diphthongization.

Features commonly refer to phonological distinctions but cannot directly be related to phonetic differences among vowel qualities due to the non-linear relation of acoustics and articulation. Differences between vowels of the same quality in different languages are interpreted as variations of the phonetic values of these major vowel features (Ladefoged & Maddieson 1990: 103; Disner 1983). In a cross-language comparison, the non-linearity between phonological values and acoustic and articulatory properties becomes particularly evident.

Schwartz et al. (1997a) present a typology of vowel systems based on the UPSID[12] material and describe vowel systems in a three-step analysis: They (1) distinguish *"primary"* from *"secondary"* constituents of a system, (2) describe each system in terms of its *"peripheral"* structure, and (3) describe its *"non-peripheral"* components. Within Schwartz et al.'s (1997a)

---

[11] Constriction degree or tenseness/laxness is not mentioned by Ladefoged & Maddieson (1990) but could also be interpreted as minor feature of vowel articulation in some languages.

[12] The UCLA Phonological Segment Inventory Database UPSID (Maddieson 1984) is based on the segment inventories of 317 languages from 20 families of languages, including at least one language from each group or sub-group of languages.

analysis, vowel systems are conceived as usually exploiting the *primary system* of sounds, which consists of primary vowels without secondary articulations (cf. minor vowel features in Ladefoged & Maddieson's terminology), such as lengthening, degree of constriction, breathy or creaky voice, nasality, ATR, etc. If the number of vowels in a system becomes too large, i.e. if it contains more than 9 sounds, a new dimension (*secondary system*) will be exploited (Schwartz et al. 1997a). Primary systems consist of 3 to 9 vowels, with a strong preference for 5 vowels, just as secondary systems that may consist of 1 to 7 secondary vowels (two third of the systems use 2 to 5 secondary vowels).

4.3.1 Frontness/Backness

Considering the interaction of the three basic parameters Height, Backness and Rounding that are commonly used for the phonological description of vowel quality, we find some interesting asymmetries: *Front-back symmetry* is the rule for primary systems but if *asymmetry* occurs in a system, more front than back vowels are expected. In around 30% of the cases, the number of front vowels is greater than the number of back vowels, especially in primary systems. While front vowels are usually unrounded (94%), back vowels are usually rounded (93.5%) (Maddieson 1984; Disner 1984).

With respect to the front/back-dimension, Maddieson's (1984: 124) analysis of the UPSID material shows that

- front vowels are a little more numerous than back vowels (40% vs. 37.8%),
- front vowels are usually unrounded (94%), back vowels are rounded (93.5%), and
- so-called "central" vowels are considerably less frequent (22.2%).

4.3.2 Height

"Vowel height" is generally referring to the relative position of the tongue body and is described as a multi-valued feature. The UPSID material distinguishes five degrees of vowel height (high, high-mid, mid, low-mid, low). Some interesting distributional patterns across languages are observed in the UPSID material that apparently reflect common preferences for certain vowel qualities and the co-varying of certain features across languages (Maddieson 1984: 124f):

- Vowels in the mid-range (high-mid, mid, low-mid, e.g. /e ɛ æ/) are a little more common than high vowels (40.5% vs. 39% of the sample).
- Low vowels are substantially less common (20.5%).

- Low vowels are usually central, i.e. neither front nor back (75.1%), and central vowels are usually low (69.4%).
- High front vowels are more frequent than high back vowels.
- Mid-range vowels are more commonly back than front if the lip position is unmarked (rounded if back, unrounded if front).

However, from an articulation-oriented view, height referring to the relative height of the tongue body is of restricted relevance particularly for back vowels:

> "*The randomness of [o ɔ a* ɑ*] tongue heights (Table I, Fig. 5) compared with the constancy of vocal tract configurations (Figs. 6 to 9) shows that tongue height is hardly a physiologically relevant parameter of vowel articulation. The physiological aspects are discussed in Paper II.*" Wood (1982a: 17)

This argument is crucial for the phonetic-phonological account presented in this study. An alternative and more critical approach to the conventional description of vowels in terms of height and frontness/backness will be presented in section 4.7.5 and 4.7.6.

4.3.3 Rounding

In terms of articulation, two kinds of rounding can be observed depending on the lip gesture: vertical lip *compression* and horizontal *protrusion* (Ladefoged & Maddieson 1990: 100ff). In most languages these parameters may be implemented jointly, often linked with the front-back dimension, resulting in a distinction between rounded (compressed or protruded) vowels vs. unrounded vowels. Unrounded vowels are considerably more frequent than rounded vowels (61.5% vs. 38.5% of all vowels described in UPSID).

As mentioned above, the great majority of languages show a predictable relationship between the phonetic dimensions of backness and rounding. Cross-linguistically, we can state a preference for back rounded and for front unrounded vowels. In 94% of the UPSID languages, front vowels are unrounded, whereas back vowels are usually rounded (93.5%) (Maddieson 1984: 124), i.e. front and low vowels are mostly pronounced with unrounded lips (e.g. [i e]), back non-low vowels ([u o]) are pronounced with lips in a rounded position (Maddieson 2008b: 50; 2013). However, even if the occurrence of lip rounding with back non-low vowels (e.g. [u o]) is the normal pattern, front rounded as well as back non-low unrounded vowels do occur in the vowel inventory of several languages. Front rounded vowels (e.g. [y ø œ]) are more frequent compared to back unrounded (e.g. [ɯ ɤ]). These qualities are often called "interior" vowels as opposed to "peripheral" vowels, i.e. front unrounded and back rounded qualities.

However, for notions such as "interior" or "central" it is quite difficult to identify distinct articulatory constellations. These notions are based on the view of a "continuous vowel space" (for discussion, see section 12.1). Mostly, the notion of a continuous vowel space refers to a two-dimensional F1xF2 chart or to the IPA quadrilateral representation of the vowel space that incorrectly describes the relation between articulation and acoustics specifically for back vowels. Nevertheless they are commonly used in the discussion of vowel systems to refer to the established phonological description of vowels in a system, but have to be viewed with great caution. It is necessary to distinguish between "central" as opposed to "peripheral" positions in an F1xF2-vowel chart and the actual articulatory constellations corresponding to a specific outcome of formant values F1, F2, F3 etc.

The frequency of occurrence in language systems of so called "*interior*" or "non-peripheral" vowels in the terminology of Schwartz et al. (1997a) decreases from high to low vowels, with the exception of ə-schwa, which is the preferred non-peripheral vowel (Schwartz et al. 1997a: 251). High *interior vowels* in primary systems (e.g. /ɨ/) are more likely to be "central" (55% of the high interior vowels). If they are not "central", they are more likely to be front rounded, e.g. /y/ (30%) than back unrounded /ɯ/ (15%). Table 4.1 shows that high non-peripheral vowels are considerably well balanced, with a vowel "in the middle" of the space, i.e. /ɨ/, occurring in about 50% of the languages analyzed, and the remaining 50% being about equally distributed between front rounded /y/ and back unrounded /ɯ/ (Schwartz et al. 1997a: 246).

[+front] [+round]	# languages	[central]	# languages	[+back] [-round]	# languages
/y/	25	/ɨ/	45	/ɯ/	23 (14 with /u/, 9 alone)
/ø/	22	/ə/	77	/ɤ/	14 (8 with /o/, 6 alone)
/œ/	3	/ɜ/	3	/ʌ/	4 (3 with /ɔ/, 1 alone)

Table 4.1: Number of languages with front rounded and back unrounded vowels (Schwartz et al. 1997a: 246)

Rounding and Height are also related in some aspects: Within the class of rounded vowels, higher vowels are described to be more rounded than lower vowels (see Linker 1982; Donegan 1978; though there are some exceptions, see Ladefoged & Maddieson 1990: 99f). The higher degree of jaw opening with lower vowels limits the amount of adjustment of the lips, since the lips are stretched vertically (Maddieson 2008b: 50; 2013).

Acoustically, lip spreading and lip protrusion lower all formants but have a stronger decreasing effect on F2. Lip rounding in high vowels has more dramatic acoustic consequences than in mid vowels due to the higher degree of lip activity (Stevens 1998: 290ff). With back rounded vowels, lip rounding has a decreasing effect on all formants and reduces the distance between the already close formants F1 and F2 even more while weakening the prominences of higher frequencies (Stevens 1998: 290). Lip rounding for non-low back vowels causes a slightly stronger decrease of F2 reducing the difference in frequency between F2 and F1 (Stevens 1998: 290). *u*-and *o*-like vowels therefore have approximately the same values for F1 and quite close F2 values, while showing clear differences in F3 (Stevens 1998: 281).

The effect of lip rounding in front vowels results in a closer proximity of F2 and F3 and a lowering of the F3-F2 center of gravity. With non-low front vowels (/i/ or /e/), the anterior end of the narrow front cavity is narrowed and lengthened for /y/ and /ø/ by lip rounding, which has a lowering effect on the frequencies associated with the front cavity.

These acoustic differences between rounded vowels are expected to cause differences in their phonological behaviour. Those vowel contrasts, for which lip rounding induces a relatively weak acoustic effect, are expected to be typologically dispreferred. Non-high front rounded vowels appear to be even less preferred.

The assumed decrease in perceptual salience from higher to lower front rounded vowels is consistent with the frequency and distribution of such vowels in the UPSID inventories, where the number of non-peripheral front rounded and back unrounded vowels decreases from higher to lower (Maddieson 1984, 2008b; Schwartz et al. 1997a: 246). In general, front rounded vowels are preferred over back unrounded. For high vowels, the series /i ɨ u/ is more preferred than the series /i y u/ (Schwartz et al. 1997b: 261), both systems occurring only if a certain number of vowels in the inventory is given.

In a set of 561 languages surveyed by Maddieson (cf. 2008a, b) for the World Atlas of Language Structures (WALS, Haspelmath et al. 2008) only 37 languages (6.6%) have one or more front rounded vowels. In this sample, 23 languages use high and mid front rounded qualities, 8 systems use only high and 6 only mid front rounded vowels. Front rounded vowels typically occur in languages with a larger than average vowel system. The mean number of vowel qualities in the set described in Maddieson (2008b) is 7.87, while the mean for languages without front rounded vowels is 5.85. Usually, if a high front rounded vowel occurs in a language, a front unrounded counterpart of the same height exists (e.g. Mandarin, Albanian).

occurrence of front rounded vowels	occurrence in %	number of languages
none	93.4%	524
high and mid	4.1%	23
only high	1.4%	8
only mid	1.1%	6
total		561

Table 4.2: Occurrence of front vowels with lip rounding in vowel inventories (World Atlas of Language Structures, Maddieson 2008b: 50; 2013)

31 of the 561 languages included in Maddieson's (2008a, b) WALS survey have high front rounded vowels and 29 include mid front rounded vowels. 23 of the 37 languages described have both high and mid front rounded vowels. Thus, Maddieson (2008a, b) observes a strong tendency towards a common occurrence of both high and mid qualities together in a system and concludes that there is no clear preference for front rounded vowels to be rather high than mid, a conclusion that differs from the analysis of the UPSID material presented in Maddieson (1984), which is based on a set of only 317 languages.

4.3.4 Dynamic spectral variation and context-dependent variation

Commonly, vowels do not occur in isolation but are embedded in the speech string. The preceding and following consonantal context as well as the speaking rate and the prosodic structure of an utterance cause systematic and significant acoustic variability. The nature and extent of contextual variability in phonetic realizations of a given vowel differs across languages (e.g. Bohn & Steinlen 2003; Strange et al. 2005). Acoustic and perceptual studies provide evidence for the existence of distinctive patterns of formant movements in consonant-vowel and vowel-consonant-*transitions* that contain important information for the distinction of co-articulated vowels (see Strange 1989; Andruski & Nearey 1992).

Apart from coarticulatory variation, *vowel-inherent dynamic variation* in the spectrum can be observed, which seems to provide important cues for vowel identification. Perception studies (e.g. Strange 1989; Strange, Jenkins & Johnson 1983) found that vowels are sub-optimally differentiated when acoustic information is entirely based on static spectral sections of the vowel target.

The perceptual importance of dynamic spectral information has been demonstrated by experimental studies with synthetic steady-state vowels and in "*silent-center*"-experiments, in which the centers of a syllable were silenced (e.g. Strange & Bohn 1998; for pre-linguistic infants see Bohn & Polka 2001). These experiments proved that correct vowel identification was possible on the base of the *transitions* to pre- and post-vocalic consonants. In general, vowel identification is not less accurate in silent-center syllables than in vowel-center syllables where the onset and offset of the syllable are removed, if durational information is available for the listeners. Further, transitions of vowels and their context also contain cues for the identification of consonants with respect to voicing and place and manner of articulation (Hillenbrand 2013).

Two types of dynamic spectral information appear to be relevant in vowel perception (cf. Strange & Bohn 1998; for an overview, see Hillenbrand 2013):

(1) Vowels can be differentiated by temporal characteristics of the *formant trajectories* of the CV- and VC-parts of the syllable. However, at least some of the parameters of formant trajectories in syllable onsets and offsets are considered to be intrinsic to the vowel itself and to be invariant over different consonantal contexts (cf. the *Dynamic Specification Theory* by Strange, Jenkins & Johnson 1983; Strange 1989; Strange & Bohn 1998: 49).

(2) According to Nearey's *Compound Target Theory* (Andruski & Nearey 1992; Nearey & Assmann 1986) vowels can be differentiated by *vowel-inherent spectral change* (VISC) within the vocalic nuclei of CVC-syllables with diphthongized vowels in many English dialects.

These studies only refer to English vowels. In Strange's (1989) more general interpretation, however, dynamic spectral information plays a relatively important role even for languages that have little or no diphthongization in vowel targets (Strange & Bohn 1998: 489). This view is corroborated by results in Strange & Bohn's (1998) study with North German vowels.

An underlying assumption of Strange's (1989) *Dynamic Specification Theory* is that – additionally to a postural "target" – vowels are understood as (dynamic) articulatory gestures (Browman & Goldstein 1992) with intrinsic timing characteristics and intrinsic styles of articulatory movement. These dynamic differences result in distinctive articulatory and acoustic patterns that are considered to differentiate vowels especially in larger vowel systems (Strange & Bohn 1998: 489).

To summarize, dynamic spectral variation provides important information for the differentiation of vowel sounds. *Time-varying spectral information* comprises the *transitions* into and out of a vocalic nucleus (CV- and -VC) as well as *vowel-inherent* information within the vocalic nucleus.

4.3.5 Phonetic duration and phonemic length

While some languages exhibit a *phonemic vowel length* contrast, other languages do not use phonological quantity oppositions. Contrastive vowel quantity refers to the use of duration to distinguish vowels with similar properties in a given vowel system. Numerous languages show a binary phonemic distinction between long and short vowels, occasionally a distinction of three contrasting vowel lengths is observed (e.g. Estonian). In UPSID[13], the probability of using length in a vowel system increases with the number of vowel quality contrasts. 53.8% of languages with ten or more vowel qualities use phonemic length.

In some respect, vowel height and length seem to be associated: In 18% of the UPSID languages with /e(:)/ and 19.6% with /o(:)/ the vowel only occurs as long compared to only 6.6% for /i(:)/, 4.9% for /u(:)/ and 2.8% for /a(:)/. Thus, mid vowels seem to be higher when lengthened and lower when shortened, or in other words, differences in vowel quantity are associated with differences in vowel quality (Maddieson 1984: 130). A universal relationship between "intrinsic vowel length" and "vowel height" has therefore been proposed, based on observations that lower vowels tend to be longer in duration than higher vowels (Lehiste 1970; Neweklowsky 1975; Maddieson 1984; Fischer-Jørgensen 1990; Laver 1994).

In all languages, *intrinsic vowel duration* has some function in conveying linguistic information to the listener (stress patterns, emotional cues etc.), even in languages without phonological quantity oppositions. Strange & Bohn (1998) demonstrated that perceptual errors increase for stimuli in which all syllable duration information was removed.

The duration of realizations of the same vowel phoneme may vary as a function of its position within larger units such as syllables and words and is further influenced by supra-segmental features. Moreover, there is between- and within-speaker variation in speaking rate with effects on the duration of individual vowel segments (Krull et al. 2003).

The perceived duration of vowels is influenced by spectral cues as well as by dynamic fundamental frequency and largely varies with the listener's language background and with patterns of co-occurrence of phonetic properties in the native language. The relative weight of these factors and the contribution of dynamic spectral variation together with durational and intonational properties in the speech signal vary language-specifically and, in L2 language acquisition, also inter-individually due to different extent of experience in L2 and the listeners' individual strategies (see section 1.3 and 8.7).

[13] The UPSID material includes vowel length contrasts only if they are linked to vowel quality differences.

Neweklowsky (1975) hypothesized that longer intrinsic length has to be considered as a compensatory mechanism for a lower intrinsic pitch. The question whether intrinsic vowel length is related to vowel height or more precisely to vowel quality in German, or in other words, whether differences in F1 indicating the degree of aperture are related to differences in duration, will be discussed in section 5.4.5.3.

4.3.6 Intrinsic pitch and dynamic fundamental frequency

Several studies have demonstrated a relationship between vowel height and fundamental frequency *F0* (Whalen & Levitt 1995; Connell 2002). Low vowels such as [a] tend to have a lower fundamental frequency than high vowels like [i] or [u]. This relationship of articulatory constellations and F0 is referred to as "intrinsic F0" (IF0) or "intrinsic pitch" and has been claimed to be universal. As articulatory correlates, differences in vowel height (high vs. low) or in degree of openness (close vs. open) have been proposed to account for differences in intrinsic F0. Considerable discussion, however, concerns the physical and cognitive motivation behind this effect. Explanations by acoustic or articulatory reasons (cf. Whalen & Levitt 1995; for a review, see e.g. Fischer-Jørgensen 1990 and Connell 2002) have been discussed as well as enhancement of differences between different vowel categories suggesting that intrinsic pitch would be a phonological phenomenon rather than a universal phonetic effect (Diehl & Kluender 1989; Diehl 1991; Kingston 1992; Kingston & Diehl 1994).

Whalen & Levitt (1995) present a comparative survey that investigated the effect of intrinsic pitch and vowel height for [a i u] in 31 languages from 11 language families considering the size of the language's vowel inventory and speakers' sex. Whalen & Levitt's choice of languages reflects a typologically representative range of languages with respect to the function of pitch (such as stress, contour and register tone or pitch accent). Similar patterns for intrinsic pitch are found in stress-timed, syllable-timed and tone languages. Whalen & Levitt (1995) therefore conclude that intrinsic pitch is a universal, phonetically based phenomenon that results directly from vowel articulation rather than being a strategy of deliberate enhancement of the signal.

Whalen & Levitt (1995) compare only results for high [i u] and low [a] that are considered to differ maximally in height. It is however necessary to ask whether intrinsic pitch can be considered as a gradient feature of vowel systems and whether differences in F0 vary consistently with intermediate degrees of opening (Turner & Verhoven 2011). Connell's

(2002) survey on African tone languages provides no strong evidence for or against the hypothesis of gradient variation of intrinsic pitch with vowel height.

Some studies (e.g. Pape & Mooshammer 2006; Van Hoof & Verhoven 2011) assume a language-specific effect of intrinsic pitch that may (at least partly) depend on the size of the vowel system and the need to enhance vowel contrasts, an idea that would support a mixed physiological-enhancement account on intrinsic pitch variation.

For speech perception, it is hypothesized that intrinsic pitch may contribute to vowel identification. Reinholt-Peterson (1986) demonstrated that in case of ambiguity F0 contributes to disambiguation. In Reinholt-Peterson's study with a synthesized vowel continuum between Danish [u:] and [o:], the listeners tended to identify an ambiguous vowel rather as [u:] when the synthesized signal showed a higher F0.

Fowler & Brown (1997) argue that listeners compensate for intrinsic F0 height differences, at least in Germanic languages like German or English: In order to reach the same pitch level for [i] and [a], [i] has to be produced with a higher F0 than [a].

Differences in intrinsic pitch have most commonly been associated with differences in vowel "*height*" and degree of *openness*. However, some terminological problems are observed in the literature with respect to "vowel height" and "openness". While "height" may refer to the effect of tongue height on F0 with a maximal contrast of high [i] vs. low [a], some authors also refer to differences in height among so called close tense vs. open lax vowel pairs as e.g. in German, considering them to contrast in their relative degree of height and openness. This may, however, be a misleading assumption, since the articulatory differences between e.g. [u] or [i] and [ʊ] or [ɪ] could be better described in terms of *constriction degree* than as differences in tongue height (for a detailed discussion, see section 5.4.3).

A differentiation in F0 with respect to jaw position (Fischer-Jørgensen 1990; Mooshammer et al. 2001) and constriction location (Moosmüller 2007: 118ff) may find more evidential support.

So far, talking of intrinsic F0 values gave the impression that F0 is a static vowel property. However, we have to be aware of the *dynamic variation* of F0 in the continuous speech signal and its relevance for speech perception. In speech production, variation of F0 is a result of coarticulatory variation, e.g. the F0 at the beginning of a vowel after a voiceless stop or fricative is higher than for voiced consonants (Löfqvist, Baer, McGarr & Story 1989; Ladefoged 2003: 86f). In speech perception, dynamic F0 variation may provide additional acoustic cues to the identity of speech sounds.

For didactic purposes in German as a foreign language, Cauneau (1992: 65f) recommends exercises that use variation of F0 as a kind of reinforcing factor to improve the perception and production of German tense vs. lax contrasts. She postulates that "tension" is higher in phrase-initial position or in phrase-final position with rising intonation. To facilitate the production and perception of tense vowels, Cauneau recommends the embedding of the difficult sound into a word in phrase-initial position or with rising intonation in phrase-final position, while by embedding into phrase-final position with falling intonation the perception and production of lax vowels should be improved.

However, even if the idea of this kind of perceptual "reinforcement" seems very appealing, a closer look at the literature on the effect of dynamic F0 variation on perceived vowel duration or quality gives a very inconclusive picture and reveals that Cauneau's assumption that rising intonation would enhance perceptual cues for qualitative differences or vowel duration cannot be confirmed by the available studies. For vowels, early studies using synthetic stimuli of a single vowel report that *dynamic F0* increases the perceived duration of vowels (Lehiste 1976; Pisoni 1976; Wang et al. 1976), Wang et al. (1976) used synthetic stimuli consisting of single vowels and non-speech sounds and found that a rising F0 contour increased the perceived length of vowels compared to vowels with a falling F0 for English listeners; these were perceived as longer than vowels with a level F0, whereas later studies could not find an effect of dynamic fundamental frequency or found an impact only under certain conditions, e.g. in certain contexts or in mono- vs. disyllabic words (van Dommelen 1993; Lehnert-LeHouillier 2010; Cumming 2010). Van Dommelen (1993) found that German listeners perceive falling F0 in the second half of isolated monosyllabic stimuli to be "longer", whereas falling F0 in disyllabic stimuli increased the number of "short" judgements.

To summarize, the results of investigations on dynamic F0 and perception of vowel duration depend very much on variables in the experimental design, e.g. the use of mono- vs. disyllabic words, of meaningful words vs. non-sense stimuli, level F0 vs. rising or falling F0 (Wang et al. 1976), or level F0 vs. rising-falling and falling-rising F0 (Lehiste 1976), steady decrease of F0 over the entire vowel or only in the second half (van Dommelen 1993), etc. (for a detailed overview for several languages, see Lehnert-LeHouillier 2010 and Cumming 2010).

In the present study, the dynamic variation of pitch was included in the experimental design and in the acoustic analysis of the input stimuli (see section 6.3). However, since only natural non-manipulated stimuli were used in the perception experiment of the current study, the effect of F0 variation could not be controlled in the listeners' input.

To summarize, at least three types of acoustic information are essential for the identification of vowels in continuous speech: (1) *spectral information* reflecting the articulatory activity to produce a vowel, (2) *dynamic spectral information* in CV- and VC-transitions reflecting coarticulatory effects of adjacent sounds, and (3) *durational information* reflecting information on the syllable level, possibly together with intonation cues, i.e. fundamental frequency F0.

4.4 Vowel inventories and phonetic universals

Languages across the world differ considerably with respect to number and types of vowels in their phonemic inventory. Detailed phonetic descriptions of vowel inventories and articulatory realizations in different languages distinguish an immense variety of vowel types.

Vowel inventories differ in the *number* of vowel phonemes, the *distributional patterns* within the vowel system and in *phonetic details*, i.e. their specific articulatory and acoustic quality.

A given language has only a very limited set of contrasting phonemes, the total segment inventory size clustering around a mean of 27 phonemes (Maddieson 1984). While some sounds occur very frequently in a large number of languages, others are rarely distributed in the languages of the world described so far (Greenberg 1963, 1966; Crothers 1978; Disner 1984; Maddieson 1984, 1997, 1999, 2008a).

The sound inventory of a given language is far from being a random sample of sounds. Rather, sound inventories can be considered as the result of a set of specific selective processes and restrictions, many of which can be explained in terms of universal constraints and preferences which are supposed to be grounded phonetically (Ohala 1983; Maddieson 1999; Flemming 2008).

When speaking of vowels in a vowel system, reference is often made to the notion of an abstract "vowel space". The articulatory-acoustic *vowel space* is an abstraction of the total set of possible articulatory constellations and the resulting resonances in the vocal tract, traditionally described in terms of formant frequency values (F1, F2, F3, ...).

The range of possible vowels in the vowel space is delimited by articulatory, acoustic and perceptual coordinates. Studies on language universals and typology attempt to define the limits within which variation across languages is confined, describing the relative frequency of various structures, based on the underlying assumption that more frequent patterns may in a sense be "better" than other less frequent ones. In this sense, *universals* can be understood as

> “*distributions of individual properties and of patterns of related properties* (…) *being studied over the universe of known languages*” (Maddieson 1999: 2522).

Maddieson (1997, 1999) distinguishes “*mechanical”* and *“ecological” constraints* to sound inventories. “Mechanical” constraints refer to mechanical limits on the range of articulatory and spectral diversity, i.e. to physiological limits in vowel articulation which restrict the number of possible vowel sounds, distinctive features and their combinations. Beyond the articulatory or spectral limits of this articulatory or acoustic vowel space it is not possible to produce sounds that could be designated or perceived as vowels. For example, beyond a certain degree of constriction in the vocal tract the sound produced will not be a vowel anymore. Other factors determining the specific arrangement of vowels in the vowel space can be assigned as “ecological”, i.e. as the result of the “ecological setting” in which a language developed or is presently used (cf. the construct of type adequacy and system adequacy, Dressler 1986: 292ff). Maddieson (1999: 2523) refers to “ecological” constraints to describe the direction of selection in certain directions within the range of what is mechanically possible, suggesting a scale from “more mechanical” to “more ecological” rather than a binary opposition to account for the physical and functional limits in the range of vowel sounds in inventories (Maddieson 1999: 2524).

Realizations of the same vowel type may diverge considerably in their concrete articulatory and acoustic manifestations. Even identical vowel phonemes may differ considerably in phonetic details in languages compared to each other (e.g. /u/ in German vs. /u/ in Polish or Arabic). Additionally to this kind of variation, there is a certain amount of vowel-intrinsic dynamic variation of phonetic parameters due to context, coarticulation and dynamic spectral differences (see above; for German, see section 5.4). Cross-language dissimilarity of vowel sounds can be defined by differences in terms of individual *gestural components* and their acoustic consequences associated with differences in *timing* relationships in these components (Strange 2007).

Therefore, the very common way of describing vowels in a two-dimensional F1xF2 chart with formant frequency values obtained from measurements at one steady point in time to represent all information necessary for vowel identification has proven to be insufficient to describe these characteristic features and the intrinsic variation of vowels in a given language. Numerous studies (e.g. Nearey & Assmann 1986; Miller 1989; Hillenbrand & Nearey 1999; Moosmüller 2007: 29ff.; Johnson 2008; Hillenbrand 2013) have demonstrated that representations of vowels as points in a two-dimensional vowel space based on formant measurements at a given point in time are too static and do not consider the role of higher formant frequencies, dynamic and temporal information. F1xF2 charts must be considered as oversimplification that do not reflect the full informational content of vowel spectra. Rather, phonemic contrasts between vowel phonemes in a given system result from a complex interplay of several phonetic dimensions (articulatory gestures and the resulting formant frequency values, dynamic spectral variation, duration, pitch, etc.). To describe the phonetic identity of a vowel, the dynamic change of articulatory and acoustic parameters has to be considered and many more dimensions than just a static value for F1 and F2 at a given point have to be included in the acoustic description of a vowel system.

The challenge for a phonological theory when describing vowel systems is to account for the full range of attested structures, representations and rules as found in the languages of the world and to characterize the many forms and patterns that are not attested. Assuming that the occurrence and phonological behaviour of sounds is not purely random, phonological theory should be capable to express the core similarities of different phonological patterns and feature interactions while allowing for the abundant wealth of variation that accompanies this core of similarities (Archangeli & Pulleyblank 1994: 3).

4.5 Phonology and vowel perception

To account for the fact that in perception the same phonetic input may be identified differently by listeners from different native languages, it is necessary to understand that phonetics and phonology act in close interaction and that they can be understood as two aspects of the same phenomenon. In cross-language vowel perception, phonetic facts are involved together with language-specific phonological interpretations, individual hypotheses and strategies and, presumably, with some constraints and preferences which may influence speech perception independently of a specific language.

The interplay between speech perception and phonology, i.e. between the perceptibility of certain sound structures and their distribution and behaviour in phonological systems across languages, has been a challenging topic for many researchers (e.g. Ohala 1993b; Boersma 1998; Hume & Johnson 2001a, 2001b; Flemming 2002; Mielke 2002).

According to functional phonological approaches, phonology is influenced by two main forces: a tendency towards *minimization of effort* in articulation and a tendency towards *maximization of perceptual clarity* to minimize ambiguity and confusion in perception (Dressler 1985: 287). However, since the speaker's primary intention is to be understood, the tendency towards clarity in perception is expected to be superior to ease in articulation (Dressler 1985; Moosmüller 2007: 4ff, 174ff).

The influence of speech perception on phonology becomes apparent in the phonological behaviour of sounds in alternations, in sound change and in language acquisition. It has been argued that the relative perceptual salience of speech sounds is relevant to explain their phonological behaviour (Hura et al. 1992; Kohler 1990, 1992; Ohala 1993b; Steriade 2001a, b; Mielke 2002). Salient sounds are less likely to be altered than perceptually weaker sounds.

The influence of phonology on perception becomes apparent in numerous studies on second language speech perception where listeners from different language backgrounds show systematic variance in the identification of invariant input stimuli. The perception of sound categories is not universal but is determined by the phonological system of the listeners' L1 (Lado 1957; Flege 1988b, 1992, 1995, 2002; Kuhl & Iverson 1995; Escudero 2005; Best et al. 2003). The fact that in speech perception listeners show a certain sensitivity to phonetic differences within a vowel category, i.e. the ability to discern "good" from "less good" exemplars of a category that differ phonetically (Kuhl 1998a, b), indicates that the perceptual system is sensitive to phonological and phonetic differences. Therefore we have to assume that the perceptual discrimination ability is based not only on phonological categories but also on phonetic differences and similarities in so-called sub-phonemic details between non-native stimuli and native speech patterns (Best et al. 2003; Kingston 2003).

In experimental studies on vowel perception, experimental context effects, e.g. the order of stimuli presentation, inter-stimuli interval (ISI), or the size of the stimuli set show a considerable influence on vowel categorization (Strange 2007: 39ff; Strange 2010).

A contrastive approach to L2 acquisition would base the prediction or explanation of problems in L2 perception on a comparative analysis of L1 and L2. However, as pointed out in chapter 1, a contrastive analysis cannot account for all difficulties in L2 perception. The reason why some sounds or even classes of sounds in L2 are acquired later or seem to be more difficult in perception than others is not to be explained by a merely contrastive approach. Therefore, other factors explaining patterns of *difficulty, preference* and *asymmetry* in perception, e.g. the higher perceptual *salience* of certain speech sounds or *similarities* within the L2 system, have to be considered.

Numerous studies have been undertaken to describe the relative difficulty of linguistic elements in L2 acquisition based on *markedness* theory and linguistic *universals* (e.g. Broselow 1983; Eckman 1977 [1987], 1991; Takahashi 1987; White 1987; Major 1987a, 2008; Major & Kim 1999). Phonetic universals can be described as patterns which are particularly common in languages described so far (see Greenberg 1978; Crothers 1978; Maddieson 1984, 1997, 1999). Common and regularly occurring phenomena are generally considered to be unmarked. Less common or dispreferred phenomena in the sound systems of the world's languages are considered to be more marked. Less marked phonetic or phonological structures of L2 will be easier to learn. Less marked structures of L1 may be harder to unlearn and will therefore be transferred more easily to the L2. It is not expected, however, that there is an absolute hierarchy of markedness in and between languages. Rather, Markedness Theory contributes to a better understanding of common tendencies in transfer, simplification, and over-generalization in foreign language acquisition. As Hayes (1999) points out, "*phonetics can serve here as a rich source of phonological explanation, since the typology matches the phonetic mechanisms so well*" (Hayes 1999: 9).

In the following sections, basic phonetic characteristics of vowel sounds together with common phonological patterns of vowel systems will be considered and phonetic and phonological motivation for recurrent patterns behind the complex spectrum of cross-linguistic variation will be presented.

4.6 System size and distributional patterns

The UPSID Database (Maddieson 1984) distinguishes 921 phonemes, 269 of them are vowels – yet a typical language utilizes only 3% of this potential set of sounds. For the present study, the UPSID material is an interesting reference to discuss the relative frequency of specific vowel contrasts under the view of a cross-linguistic comparison.

A majority of languages have at least three phonemic vowels, though some languages have been analyzed to have less than three vowel phonemes, e.g. Kabardian, Margi, or Ubykh as 2-vowel systems (Ladefoged & Maddieson 1990); the best known 2-vowel language is Kabardian (for a phonetic analysis, see Wood 1994 and Gordon & Applebaum 2006).

The total number of vowels in the UPSID inventories varies between 3 and 24 (Disner 1984: 126) with a mean of 8.7 (Maddieson 1984: 9). The most common number of vowels in a language is 5 (found in almost a third of the languages in UPSID). The mean number of distinctive vowel qualities in a language is also 5, but the number of vowel phonemes is not necessarily equal with the number of distinctive vowel qualities (Maddieson 1984: 127). Two thirds of the inventories have 5 to 7 vowel qualities, inventories with up to 10 different vowel qualities are still relatively common.

One might assume that smaller inventories contain more common sounds while larger inventories have a greater probability of including less frequent sounds. However, the relation between size and content of an inventory is a matter that concerns individual segments rather than a specific set of members of an inventory. In other words, a given sound type can be described as more likely to occur in a small or in a larger vowel set (Maddieson 1984: 10f).

Another interesting issue is the relation of the number of consonants and the number of vowels in a phonemic system. Languages with complex vowel inventories rarely show contrasts in secondary consonant articulations of "plain" vs. palatalized or labialized consonants. On the other hand, languages that use distinctive secondary articulation contrasts for consonants tend to avoid multiple vowel contrast, particularly avoiding phonemic contrasts for back rounded vs. front rounded vowels, e.g. vowels of the type /y/ or /ɯ/. Both types of contrasts are marked (Maddieson 1984: 38, 124f) and very rarely co-occur in language inventories (Kochetov 2008: 194f[14]).

4.7 Theories of vowel inventory typology

Numerous studies searched for systemic configurations of vowel systems, for the individual position of single vowels in the articulatory-acoustic vowel space, for the resulting vowel contrasts in a system and for reasons that explain the distribution of vowels in the articulatory-acoustic space. Although, cross-linguistically, vowel systems show a wide range of phonemic and phonetic variety, there seem to be certain *preferences* for some of the articulatory options available in the phonetic and acoustic vowel space. With respect to articulation, contrastivity and auditory salience, some vowels, it is argued, are "better" or more preferred than others.

Two questions are of particular interest for *theories on vowel inventory typology* (cf. Vaux & Samuels 2006):

(1) Why do vowels fall into particular areas of the produceable/perceivable vowel space?
(2) Why do some sounds seem to be more basic and prior in language acquisition and more preferred in synchronic and diachronic evolution of segment inventories?

4.7.1 Markedness Theory

Traditional *Markedness Theory* presents a hierarchy of linguistic structures based on descriptions of phonological behaviour in language acquisition, language use, sound change, and cross-linguistic distribution in the languages of the world (see section 1.1.3, 1.1.4 and

[14] Kochetov (2008) argues that no innate restrictions against having both types of contrasts in a language system need to be assumed, since language systems would shift to a more stable solution by "self-organization" avoiding either rounding contrasts in vowels or secondary articulation contrasts for consonants or none of these marked contrasts.

1.2). An important question is whether markedness is based on paradigmatic contrast or rather on intrinsic properties of vowel sounds themselves and whether difficulties arise from the categorization of sounds within a system or from their individual acoustic-auditive characteristics (Vaux & Samuels 2006: 13). Criticism to typological studies and Markedness theory mainly refers to the fact that it often limits itself to a mere description of what can be observed in sound inventories rather than explaining these facts. The same argument would hold for Dispersion Theories (see section 4.7.4).

A *functional approach* which explains preferences and constraints in vowel systems in terms of articulation, acoustics, perception and speech processing provides precious supplemental motivation for the existence of the observed patterns (cf. here the contributions of *Natural Phonology*, e.g. Dressler 1984, 1989, 1995; 2002; Donegan 2002b; Dressler & Dziubalska-Kołaczyk 1994).

4.7.2 The Natural Phonology of Vowels

According to Natural Phonology (NP), every regular sound *substitution* reflects the operation of one or more *phonological processes*. Natural Phonology (see section1.2.4, for a general description) predicts that "difficult" sounds are usually substituted by "less difficult" sounds, i.e. sounds that are easier to pronounce and better to perceive. The aim of phonological processes is to eliminate difficult features. Processes apply independently of a given language or a specific speaker but are considered to be language-universal. The phonology of a given language is based on the system of universal phonetic principles. Processes usually apply involuntarily and unconsciously. In production, a process becomes noticeable only in pronunciations of structures where it is not applied appropriately, e.g. when imitating foreign sounds or sound sequences.

Phonological processes work paradigmatically (for segments) and syntagmatically (for sound sequences). While paradigmatic processes usually are so-called foregrounding processes or "fortitions", syntagmatic processes work as backgrounding processes or "lenitions" (cf. Donegan & Stampe 1979: 149; Dressler 1985: 44). Both are articulatorily and perceptually motivated. Conflicts between paradigmatic and syntagmatic processes may arise by the counteracting interests towards ease of production and clarity of perception.

Donegan (1978: 59ff) in her work "On the Natural Phonology of Vowels" describes a set of *vocalic fortition processes* which govern the segmental phonology of vowels. Fortition processes are considered to typically optimize or maximize phonetic features of individual segments, they are typically context-free and frequently apply dissimilatively. Three pairs of fortition processes are described by Donegan (1978):

(1) The processes of *Tensing and Laxing* concern changes in the degree of colour of palatal and labial vowels. Tensing strengthens colour for a given vocalic height, whereas Laxing attenuates colour but imparts sonority for a given height. Both often interact with other phonological processes. Long and high vowels are especially susceptible to Tensing. The implicational conditions for Tensing are notated as [!long], [!higher] (read "the longer and/or the higher a vowel, the more it is susceptible to Tensing").

(2) *Lowering and Raising* affect changes of vowel height, a feature which is the most direct manifestation of sonority. Lowering increases sonority by decreasing the height of vowels by one degree. Its implicational conditions are [!achromatic], [!lax], [!mixed], [!long], whereas Raising applies under opposite conditions and has contradictory effects (Donegan 1978: 68ff).

(3) *Bleaching and Colouring* are processes that add or remove colour (palatality or/and labiality) without changing vowel height. Being functionally parallel to Laxing and Tensing they share many of their implicational conditions. Bleaching consists of two sub-processes, *Delabialization and Depalatalization*. There is no implicational relation between these two processes, though it seems that Depalatalization is the rarer one. Their implicational conditions are [!lower], [!lax], [mixed] (Donegan 1978: 83ff.).

It is important to stress that processes apply to (or are conditioned by) classes of segments, e.g. front rounded vowels or mid vowels, rather than individual segments. Segments affected by a certain process share particular phonetically-based features such as labiality, nasality, height, etc. In both, production and perception of speech, phonological features serve as mental classifiers.

The application of a natural process does not always result in exactly the same substitutions from speaker to speaker, language to language, style to style or time to time. The same process may apply with different degrees of generality to a class of sounds or to certain members of a class. For example, the process of delabialization may be limited to labio-palatals, but it may also apply to all labial vowels (like e.g. in child speech, Donegan 1978: 16). However, the general process [V] => [-labial] is usually applied only to labio-palatals [V, +palatal] => [-labial] or may be restricted only to non-high vowels [V, +palatal, -high] => [-labial].

Natural processes vary language-specifically but within strictly defined limits. These limits are provided by *implicational hierarchies* that correspond to the susceptibility of single segments to processes. Generally, there are strict hierarchic conditions limiting the application of natural processes. These conditions are unilaterally implicational and may refer not only to the presence or absence of a feature (like palatality) but also to the degree to which a feature is present. For example, the lower a vowel is, the more susceptible it is to delabialization: a higher vowel is unrounded only if any corresponding lower vowel is also unrounded, that is [y] => [i] implies [ø] => [e] implies [œ] => [æ] (the more susceptible must undergo the substitution if the less-susceptible does so). On the other side, unrounding of a lower vowel implies nothing about any higher vowel. These general implicational conditions might interact with implicational conditions on the environment of a certain sound, e.g. the segmental context.

The implicational conditions on process application as the processes themselves are universal but each process is subject to a number of possible conditions, to different degrees and in different combinations, i.e. interaction with other processes. Discovering these hierarchical conditions (and interactions) is crucial in the investigation of a language-specific system of processes as well as for cross-linguistic comparisons.

4.7.3 The Quantal Theory of Speech

The Q*uantal Theory of Speech* (Stevens 1972, 1989; Stevens & Keyser 2010) refers to the relation of articulatory parameters and their acoustic effects, linking discrete abstract phonological representations to the perception and production of speech signals (Stevens & Keyser 2010). Stevens (1972) observed a non-linear relation between variations in the articulatory dimension, e.g. variation in back cavity vs. front cavity length, and their consequences in the acoustic and perceptual dimensions. The sensitivity of the acoustic pattern to incremental articulatory steps varies, defining areas, where articulatory changes have minimal acoustic effects as opposed to areas, where minimal articulatory changes cause rapid changes in acoustic parameters ("*quantal effect*"). Regions of relative acoustic stability are separated thus from zones of more rapid pattern variation (Stevens 1972, 1989).

Moreover, the proximity of two formants, the so-called "*center of gravity*" of a particular vowel (cf. Chistovich & Lublinskaya 1979), is a crucial cue to vowel identification. The two most basic quantal effects reported by Stevens are the F1-F2 proximity for back vowels and the proximity of F2 and F3 for front-vowels. The back-vowel area is relatively stable such that language-dependent variations of the constriction location do not seem to change F1 and F2 much for /u/ and /o/. This is due to the strengthening effect of lip-rounding. Formant proximity in front vowels has to be viewed more cautiously, especially with respect to the group of front rounded vowels[15].

Zones of stability and instability provide conditions for ways of "optimizing" phonemic inventories. Stevens (1989) argued for the typological preference for *stable regions* in the articulatory space, so called "*hot spots*", where variability in production has minimal acoustic or perceptual impact. Referring to formant proximity, the point vowels /i u a/ could be viewed as maximally distinct and therefore typologically preferred. Their location in stable regions allows for less articulatory precision on part of the speaker without risking perceptual confusion, since these regions are auditorily clearly distinct from each other and are separated by regions of acoustic instability which involve a high range of acoustic change.

Criticism to the Quantal Theory (for critical reviews on the theory, see the contributions in Journal of Phonetics, 17, 1989) refers e.g. to the fact that the theory mainly refers to the three point-vowels /i u a/, while other very common vowels like /ɛ/ may be stable to perturbations in back-cavity length but are not bounded by regions of high instability. Thus, very frequent vowels like the front /ɛ/ are not predicted, while the occurrence of a less frequent vowel like /y/ is predicted by the theory (see e.g. Diehl (1989), for a critical comparison of the Quantal Theory (QT) and the *Theory of Adaptive Dispersion* (TAD) of Liljencrants & Lindblom 1972).

4.7.4 Vowel Dispersion Theories

Vowel Dispersion Theories (Liljencrants & Lindblom 1972; Lindblom 1986; Terbeek 1977; Flemming 1996, 2002; Schwartz et al. 1997b; de Boer 2000, 2001) predict an optimal arrangement for any given number of vowels in a system and an "even distribution" in the available phonetic vowel space to achieve a maximization of perceptual distinctiveness (for a review and comparison of the different approaches, see Vaux & Samuels 2006). The evolution of sound inventories is considered to be motivated by maximal perceptual distinctiveness at minimal articulatory cost. Maximally dispersed vowel systems show no unbalanced gaps in the primary/peripheral system gaining maximal contrast in the articulatory

[15] Fant (1989: 82f) mentions some doubt concerning the importance of F2-F3 proximity for the identification of front vowels. The preferred constriction place of /i/ in many languages seems to be 1-2 cm anterior to the proximity point of F2 and F3. At this point, lip rounding would effectively lower F3 much more than F2, creating an F3-F2 proximity for /y/ which is distinct from F4-F3 proximity for /i/. Rather, Fant suggests the relative high location of F2 supporting the upper group of formants as the distinctive feature for the entire range of front vowels. The stability of /i/ would be guaranteed anyway by the articulatory convenience and perceptual tolerance. The ease of adding lip-rounding for /y/, resulting in a proximity of F3-F2, would guarantee the distinction of /y/ and /i/ (Fant 1989: 83).

as well as in the perceptual space. In connection with the phonological framework of *Optimality Theory*, *Dispersion Theories* have gained more support (e.g. Flemming 2002, 2004).

Lindblom's *Dispersion Theory* (DT) and the later *Theory of Adaptive Dispersion* (TAD) (Liljencrants & Lindblom's 1972; Lindblom 1986) argue that in order to maintain sufficient contrast between elements in a system, the elements will be adaptively dispersed over the vowel space. Thus, optimization of a vowel system is considered to be achieved by maximization of perceptual distances among members of a system. The perceptual distance of two vowels is the Euclidian distance in terms of F2-F1 or in terms of F1/F2' (F2' refers to "F2 prime", that is a hypothesized perceptual integration of F2, F3 and F4). Larger inventories would expand the vowel space in increasing the distance between the point vowels. However, we may assume that there are other ways of optimally exploiting the vowel space than in increasing distances. The use of other than primary dimensions (e.g. length, nasalization, secondary articulations) can also contribute to maintain sufficient contrast between members of a system. Vowel systems predicted on the basis of maximal auditory dispersion have more high vowels than are attested in real systems. In a more recent simulation approach of preferred vowel systems, Diehl, Lindblom & Creeger (2003) find two ways to eliminate the problem: (1) auditory distances are calculated in background noise and (2) distances are calculated by using auditory representations that incorporate both spatial coding (excitation pattern) and temporal coding (dominant frequency) of spectral components.

Dispersion-Focalization Theory (DFT) (Schwartz et al. 1997b; Schwartz et al. 2005) calculates the optimality ("energy") of a vowel system as weighted by a combination of two separate auditory parameters: *dispersion*, i.e. maximization of auditory distance, and *focalization*, i.e. maximization of the importance of "focal" vowels such as e.g. /i/ and /y/. So-called "focal" vowels are considered to be preferred in vowel systems due to their inherent acoustic qualities, irrespective of their role in the system as a whole (see also Vaissière 2011). One or more formants that are close together are considered to enhance each other, guaranteeing the perceptual robustness of the vowel.

Dispersion theories argue that there are two types of "constraints" that determine phonological systems: While *global constraints* are based on the relations between elements in a system explaining preferences for those elements that are maximally distinct from each other, *local constraints* focus on the selection of regions in the articulatory-acoustic-perceptual vowel space, where elements exhibit preferred intrinsic properties (Schwartz et al.

2005). *Global dispersion*[16] refers to the distribution of vowels in a space while *local focalization* refers to intra-vowel spectral salience related to the proximity of formants (see also Vaissiére 2011).

Schwartz et al. (2005) argue that the so-called "peripheral" vowels (cf. Polka & Bohn 2003, 2011; Labov 1995: 458; Labov 2010: 114) are actually "focal" vowels which function as a reference or perceptual anchor for discrimination within a contrast due to their increased auditory salience. Articulatory parameters, however, are rather neglected in this approach in favour of acoustic characteristics of the speech signals.

4.7.5 The continuous vowel space vs. discrete locations of constriction

Two basic types of theories on vowel systems can be distinguished that differ fundamentally with respect to one basic assumption:

(1) Theories assuming a *continuous vowel space* describe vowels as points in the vowel space or rather in the vowel quadrilateral, in which articulatory parameters vary along a continuum (e.g. Bell 1867; Jones 1932; Liljencrants & Lindblom 1972; Schwartz et al. 1997a, b; IPA International Phonetic Association Handbook 1999). The position of a specific vowel in a given language is commonly described along the two dimensions front-back and high-low mainly referring to the position of the highest point of the tongue. Implicitly, this view assumes a direct correspondence between the position of the articulators (especially the tongue body) and the acoustic or auditory output observed in terms of acoustic resonances and measured in formant frequency values. Although this view seems at first sight very plausible, it has some shortcomings. It has already been stated that no linear relation between articulatory movements, their acoustic output and the phonemic status of a vowel (e.g. the position of the highest point of the tongue in [o] vs. [u]) which cannot be attributed only to speaker-individual differences, different speech styles and language varieties. A linear correspondence of the articulatory setting and the resulting formant frequencies cannot be upheld, especially if lip rounding is involved, by which all formant frequencies, especially F2, are lowered.

Wood (1979) leads these non-linearities between articulatory and acoustic data back to the choice of the wrong articulatory variables ("Height" and "Fronting") for description rather than to articulatory irregularities. According to Wood, the number of locations where the

[16] Global dispersion is abstractly equivalent to Lindblom et al.'s notion of "auditory dispersion" (Diehl, Lindblom & Greeger 2003). Auditory distance in DFT is calculated with respect to a formant-based space.

vocal tract is narrowed or constricted by the tongue, is relevant for the description of vowel articulation. While differences in formant frequencies are gradual, constriction locations are described as discrete. Constriction locations can be defined as areas of narrowing in the vocal tract.

(2) The view of *discrete regions of constriction* (Wood 1979; Stevens 1989) in vowel production can be traced back to ancient Indian grammarians (Wood 1979). This view refers to discrete regions in the articulatory vowel space but has received less attention with the advances in acoustical analysis and psycho-acoustical experimentation that have led to the common acceptance of a purely auditory or integrated acoustic-auditory description of vowels. Within a vowel theory that assumes no continuous vowel space but discrete or "quantal" regions (Stevens 1989, see 4.7.3), vowels are expected to occur in qualitatively distinct types. The location of narrowing or constriction in the vocal tract influences acoustics. Each vowel type could be defined by a specific constriction location with a stable articulatory-acoustic relation.

Wood (1979) describes four discrete *constriction locations* in vowel articulation: (1) a palatal constriction location along the *hard palate* for [i-ɛ]-like and [y-ø]-like vowels, (2) a constriction in the vicinity of the *soft palate* for [u-ʊ] and [ɨ]-like vowels, (3) a constriction in the *upper pharyngeal* region for [o-ɔ] and [ɤ]-like vowels, and (4) a constriction in the *lower pharyngeal* region for [ɑ-a-æ]-like vowels (Wood 1979: 30f). There is no direct evidence for a "central" constriction location.

F1 and F2 are only minimally sensitive to constriction shifts at these four locations. Based on these four constriction locations, the F1xF2-space could be divided into four relatively unambiguous areas defining a family of vowel qualities. Within each of these areas, different spectral configurations result from variation in the *degree of constriction*, corresponding to mandibular and lingual articulation, and the *degree of mouth-opening*, corresponding to labial and mandibular articulation (Wood 1979: 31). While the constriction locations are discrete, the degree of constriction, jaw opening and rounding may vary gradually with consequences on the formant frequencies.

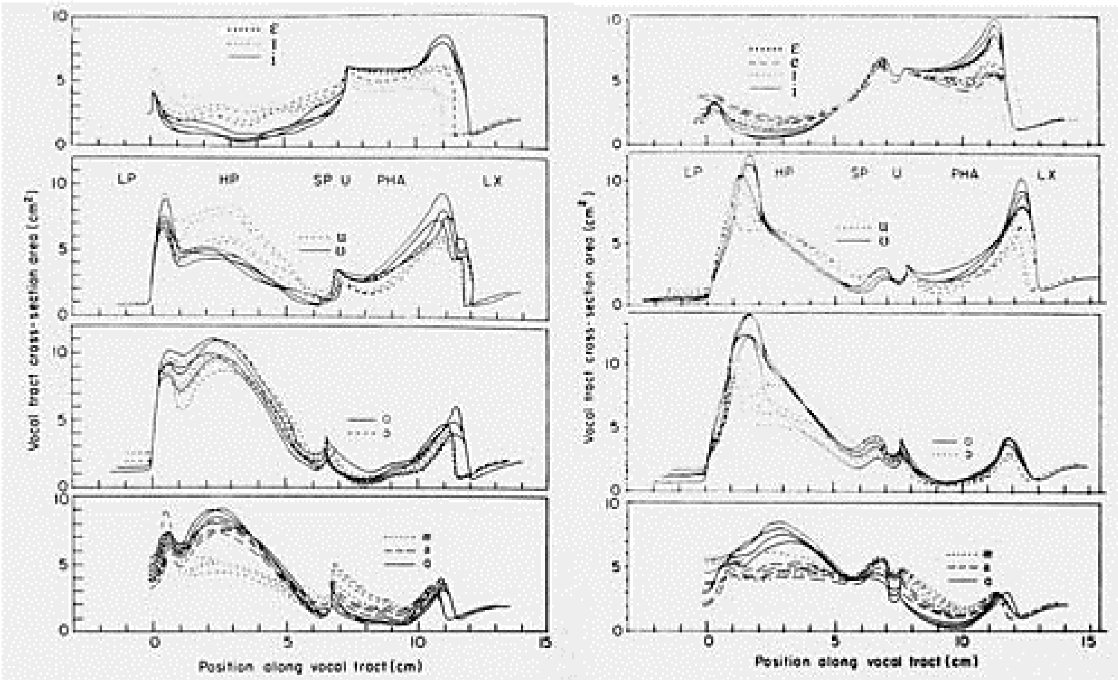

Figure 4.1: Area functions[17] of the vocal tract for vowels of Southern British English and Cairene Arabic classified according to four constriction locations (Wood 1979: 26[18])

For palatal vowels, there are language-specific differences in the exact location of constriction. Palatal vowels can be further differentiated in more or less fronted palatal vowels, e.g. the constriction in Arabic is more front than in English (27 vs. 35 mm behind the central incisors, cf. Figure 4.1and Figure 4.2). Instances of such a more fronted *pre-palatal* constriction[19] mostly occur in languages with a contrast of [i] and [y] or [ɨ], which are enhanced by a maximally high F3 in [i] (Wood 1979: 34; for acoustic evidence for a pre-palatal constriction in Standard Austrian German, see Moosmüller 2007).

At each constriction location, vowel timbres are varied by modifying the degree of constriction, tongue body posture, tongue blade elevation, lip activity and larynx depression.

17 The area functions are lined up from the central incisors (coordinate 0 cm). The letters identify parts of the vocal tract: LP lips; HP hard palate; SP soft palate; U uvula; PHA pharynx; LX larynx.

18 Figure by Sydney Wood http://swphonetics.files.wordpress.com/2014/03/1979figs1and2.jpg?w=640

19 "horizontal tongue position" (Mooshammer 1998) or "front raising" (Moosmüller 2007: 82).

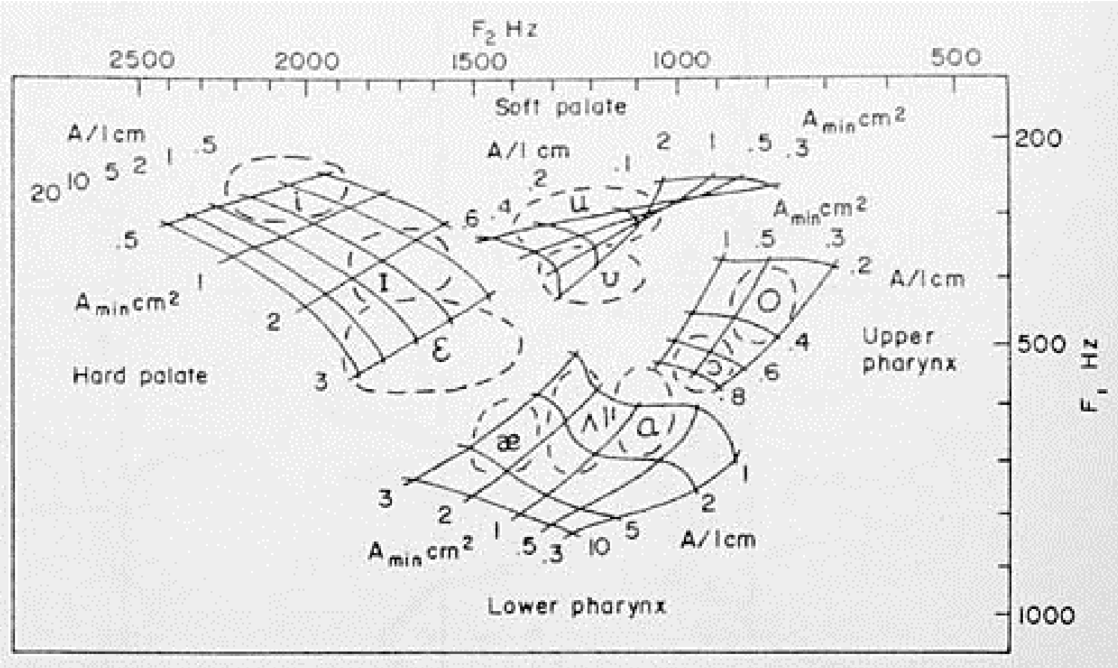

Figure 4.2: Frequencies of F1 and F2 generated by the three parameter model of the four preferred constriction locations[20]. Figure from Wood (1979: 30[21]) based on nomograms in Stevens & House (1955)

Figure 4.2 (reproduced from Wood 1979: 30) represents the F1 and F2 frequencies that are associated with each of the four constriction locations. Wood's (1979, 1982) comparison of data from different languages revealed that the resonance modes in these constriction areas are insensitive to some displacement of the constriction and that the preference for these locations seems to be universal.

> "*The four constriction locations conveniently divide the entire F1/F2 space into four relatively unamgibuous areas, each enclosing a definable family of vowel qualities. Within each area, the different spectra are obtained by varying the degree of constriction (corresponding to lingual and mandibular articulation). In natural speech the formant frequencies are also determined by tongue root movement in the lower pharynx, by tongue blade movement in the buccal cavity and by vertical larynx movement.*" (Wood 1979: 31)

These observations appear very compatible with the *Quantal Theory* of Speech (Stevens 1972, 1989), which makes reference to regions in the articulatory space where the acoustic parameters are relatively insensitive to changes in the articulatory parameters, i.e. where changes in articulation do not have much effect on the acoustic output, whereas in other areas slight differences in articulation may produce qualitatively different acoustic effects (Wood

[20] Distance from the source of constriction 12 cm for the hard palate, 8.5 cm for the soft palate, 6.5 cm for the upper pharynx, and 4.5 cm for the lower pharynx. The superimposed vowel areas are from a sample of Southern British English (Wood 1979:30).

[21] Figure by Sydney Wood http://swphonetics.files.wordpress.com/2014/02/20140206dgb01.jpg?w=640

1979: 41). According to Stevens's Quantal Theory, we seek for constriction locations in the vocal tract at places where vowel formant resonances are least sensitive to variability of constriction location.

The view of discrete regions in the vowel space provides an alternative perspective on many phonetic and phonological problems (e.g. vowel harmony) compared to the more established model of tongue articulation describing vowel systems in terms of height and fronting/backness. Its relevance for cross-language vowel perception and for the acquisition of vowel inventories in L2 acquisition will be discussed generally in the following section 4.7.6 and with respect to German in section 5.4.

Two alternative hypotheses are possible: (1) the location of the constriction is *language-universal*, i.e. category-boundaries are more or less the same in different languages, or (2) the relative location of discrete constriction regions in the vocal tract may *vary language-specifically*. While there are some regions that are not sensitive to slight articulatory changes, we may expect that in other regions the category boundaries between two vowel phonemes are affected by articulatory manipulations of the vocal tract. Spectral information may give us indirect evidence for the acoustic consequences of articulatory differences.

4.7.6 An articulatory-oriented approach

With the advances in acoustical analysis and psycho-acoustical experimentation techniques, the focus of scientific studies on vowel sounds and vowel systems shifted towards a more acoustic or acoustic-auditory view of vowels while neglecting articulatory aspects of vowel production. A strongly auditorily or acoustically-auditorily oriented view has also been propagated by the system of cardinal vowels for the phonetic description of vowel inventories (cf. Joos 1948; Jones 1932[22]; International Phonetic Association 1999). As a result, theories that are substantially based on the concept of a universal "vowel space" (e.g. *Vowel Dispersion Theories*) gained prominence in the typological description of vowel inventories.

Recently, however, the use of more advanced imaging methods for articulatory tracking, e.g. x-ray photography, Magnet Resonance Imaging (MRI) or Electro-Magnetic Midsagittal Articulography (EMMA, e.g. Hoole & Mooshammer 2002) contribute to a re-focussing on

[22] Jones (1932: 36) admits that the shape of diagrams representing the tongue positions for Cardinal Vowels is a compromise between scientific accuracy and more practical requirements. "*Scientific accuracy would require a diagram with curved sides somewhat as shown in Fig. 24. This shape, however, is inconvenient for use in practical teaching. Practical teaching can only be attained by means of a figure bounded by straight lines.*" (Jones 1932: 36). Therefore, an elliptic diagram with curved sides rather than straight lines (inspired by Jones 1932: 36, Fig. 24) will be used here for a more accurate representation with regard to tongue position of back vowels (see Figure 4.3).

articulation and allow for an integrated description of acoustic and articulatory data (e.g. Hoole & Mooshammer 2002; Pouplier & Stone 2005), providing deeper insights into the phonetic correlates of phonological distinctions as well as into articulatory-acoustic variability.

Feature-based theories are, however, challenged by the existence of phonetic variability in speech. For vowels, phonetic variability refers to differences in quality as well as quantity (duration). Alternatively to feature-based accounts and referring to the level of speech planning, one could assume a set of *articulatory gestures* instead of distinctive features. Gestures are assumed in several theoretical approaches: the *Motor Theory* (Liberman & Mattingly 1985), the *Dynamic Specification Theory* (Strange 1989b) and *Articulatory Phonology* (Browman & Goldstein 1989).

Within the framework of *Articulatory Phonology* (Browman & Goldstein 1989, 1992), phonemic contrasts are described in terms of gestural constellations or in terms of dynamic articulatory *gestures.* Within an articulatory approach, sounds or phonemes are modelled as bundles of features or gestures rather than in terms of single distinctive features (Browman & Goldstein 1989). Articulatory gestures are viewed as suitable to characterize phonological functions such as contrast.

Articulatory constellations may differ in terms of the *articulators* involved (e.g. closure of the lips vs. closure of the tongue tip vs. closure of the tongue body) or in terms of the presence or absence of a particular *gesture*, e.g. widening of the glottis to distinguish voiced vs. voiceless consonants or labial activity for [+/– labial] vowels or consonants. Articulatory constellations may also differ in terms of the exact *location* of the closure (e.g. alveolar vs. palato-alveolar). For example, the independent articulators, i.e. the upper lips, lower lips and jaw involved in lip closure for labial vowels or consonants, contribute to the lip closure gesture to different extents as a function of the context. In other words, instead of a set of distinctive features, a phonological inventory could also be described as a combinatorial system of gestures by calculating the number of gestures and the number of contrastive states they can go into.

Proctor (2007) proposes a typological analysis of sound inventories within Articulatory Phonology based on the concept of *entropy*. In terms of this approach, entropy in a phonological system can be increased either by increasing the number of contrastive elements, or by increasing the number of distinct states these units can enter into. The complexity of phonological inventories is increased according to a hierarchy of contrasts. More salient primitives will be used prior to less salient ones. In other words, contrasts between articulators or distinctive regions of the articulatory space in vowel production will be used prior to other types of distinctions.

This view is very compatible with the concepts of "hot spots" (Stevens 1989) and the asumption of distinctive constriction locations (cf. Wood 1979, 1982a, b). While Proctor's proposal mainly refers to consonant inventories and simple vowel systems (American Spanish), it will be applied here to more complex vowel inventories (see section 5.6, for a proposal for German vowels, see Table 5.9).

We may thus posit that the complexity of a vowel phoneme inventory can be described in terms of the number of contrastive elements and the gestures and distinct states involved, e.g. the number of distinctive regions of constriction, the degree of constriction or the temporal contrasts that cannot be ascribed to gestural coordination within a phoneme.

An integrated model of both, articulatory and acoustic characteristics of basis vowel qualities is proposed by Wood (1979, 1982a, b, 1986). A *primary dimension* of phonemic vowel contrasts refers to the constriction location and typical formant frequencies as described by Wood (1979, 1982a: 42f): Five-phoneme systems distinguish (1) [a]-like vowels with low-pharyngeal constrictions, (2) [e]-like vowels and (3) [i]-like vowels with spread lips and palatal constrictions, (4) rounded [u]-like vowels with velar constrictions, and (5) rounded [o]-like vowels with upper-pharyngeal constrictions (Wood 1979: 40). For palatal vowels, language-specific tendencies to either pre-palatal or mid-palatal tongue positions can be observed (Wood 1982a: 43). Languages contrasting [i] and [y] seem to prefer the pre-palatal position for both vowels (Wood 1979, 1982, 1986[23]). Only little variation is observed for the other three constrictions.

Wood (1986) argues that the component manoeuvres contributing to a given phoneme contrast fall into two functionally different groups: one component causing the major formant shifts that constitute the contrast, e.g. lip rounding for [y] vs. [i], and those manoeuvres that optimize the resonance conditions. However, both components are indispensible and not interchangeable. Therefore, the full complex of manoeuvres that contribute to the contrast and their spectral effects have to be regarded in physiological and acoustical studies (Wood 1986: 399f).

[23] For a detailed cross-language comparison of parameters contributing to the contrast of /i/ and /y/ such as the degree of lip rounding, larynx lowering, tongue location, degree of palatal narrowing, and tongue blade elevation, see Wood (1986).

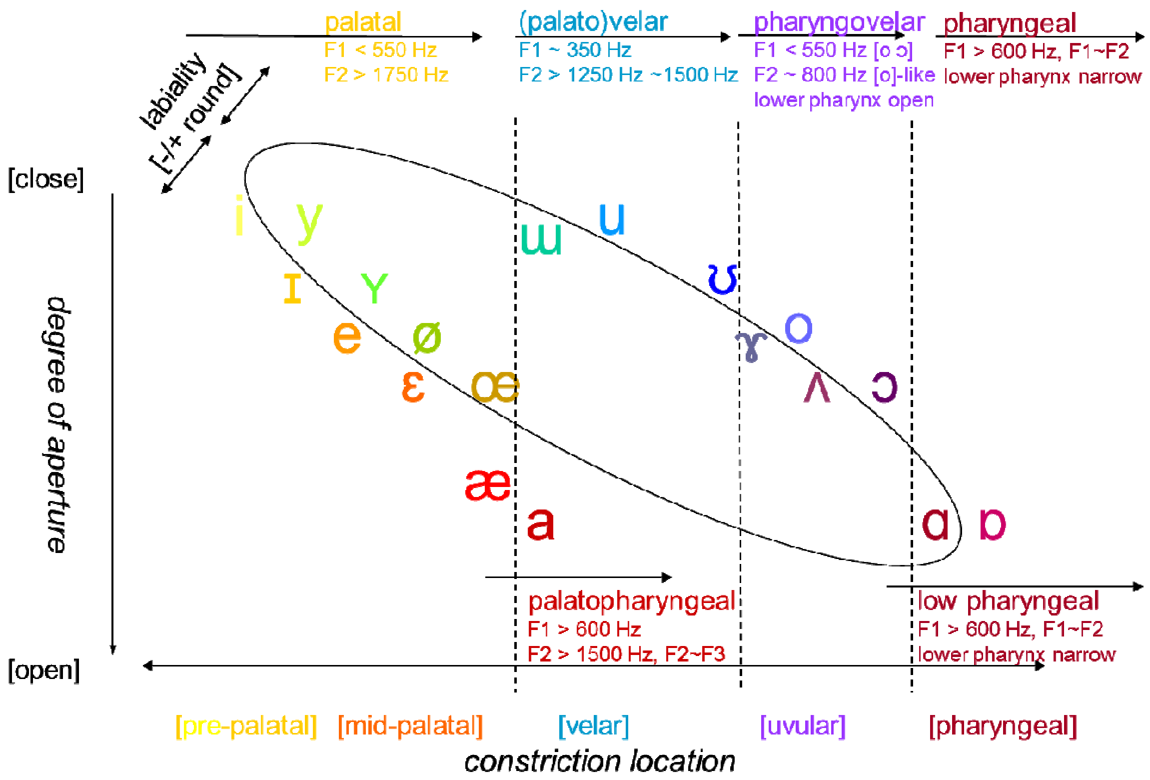

Figure 4.3: Articulatory-acoustic model of constriction locations and major vowel classes

The integrated articulatory-acoustic approach is visualized in Figure 4.3. The current study proposes an *elliptic representation* of the articulatory-acoustic vowel space instead of the common IPA vowel quadrilateral that provides an inadequate representation of articulatory constellations specifically with respect to back vowels and the concept of "tongue height" (see Wood 1975, 1979, 1982a, b). Figure 4.3 offers an account for the complex interplay of *articulatory constellations* and their *acoustic effects* (for formant frequencies, see Wood 1979) and therefore showed to be useful for the interpretation of cross-language vowel perception. The description of the German vowel system in the current study is based on this articulatory-oriented account (for discussion, see section 5.6, Figure 5.24 and Table 5.9).

Further contrasts, e.g. in systems with more than these five basic categories, are made by *secondary modifications* of the basic parameters, for example by the relative degree of mouth opening, the constriction grade, the degree of lip rounding or lip protrusion or by variation of the area of constriction. A further possibility is *temporal variation*.

For each of these *secondary parameters* a set of distinctive states is assumed. By a shift of these articulatory parameters from one distinctive state to another, the phoneme boundaries may be transgressed in perception. The number of modifying parameters and the set of distinctive states and the thresholds are assumed to vary language-specifically. Figure 5.24 will exemplify this for the German vowel system showing that primary phonemic distinctions can be described in terms of different constriction locations while the contrast between allophones is expressed by additional articulatory variables such as degree of constriction, degree of lip rounding, tongue blade depression, or larynx depression. For German, a set of four additional variables is proposed in section 5.6: mandibular aperture, lip rounding, constriction degree and phonemic length.

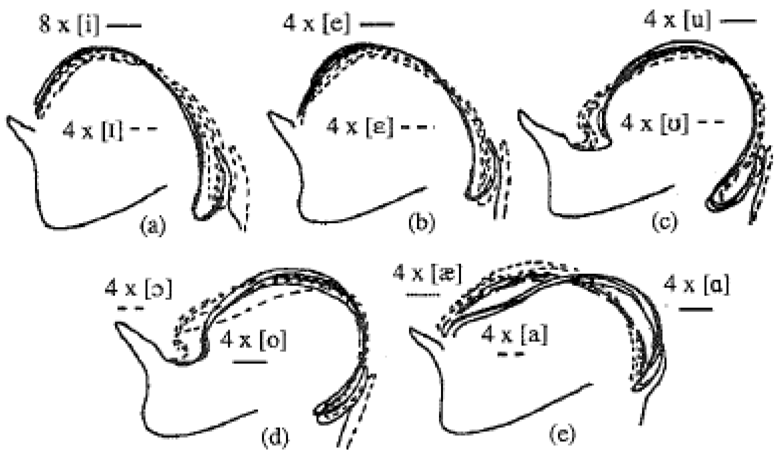

Figure 4.4: Tongue position relative to the mandible for stressed vowels by a speaker of Egyptian Arabic classed by constriction location. The profiles show the range of variation and the pre-palatal posture for palatal vowels (Wood 1979: 32, figure reproduced from Sydney Wood 1990: 197) [24].

For a language like Classical Arabic that has traditionally been characterized as a 3-vowel system of /i u a/, we may – according to this model - in a first step assume three constriction locations: (1) a palatal location for /i/, (2) a pharyngeal location for /a/, and (3) a velar location for /u/.

However, there is considerable allophonic variation observed for these three vowel qualities (for a detailed description, see section 10.4.1). Moreover, durational differences are phonemically functional in Arabic, resulting in a system of 3 short vowels /i u a/ and 3 long vowels /i: u: a:/. In some dialects, the two diphthongs /aj/ and /aw/ are realized as [e:] and [o:] (e.g. Egyptian Arabic (Cairene), Central Sudanese and some Levantine dialects). The system of Cairene can therefore be described as consisting of three short vowel qualities /i u a/ and 5 long ones /i: e: a: o: u:/. Wood (1979) describes the vowel phonemes and allophones for Egyptian Arabic in terms of the four constriction locations mentioned above. However,

[24] Wood (1979) distinguishes for Cairene Arabic at each of the five constriction locations a constricted and an unconstricted vowel quality, an assumption that is not in accordance with traditional descriptions of the Arabic vowel system (see section 10.4); for information on the sentences read by the Cairene Arabic subject, see Wood (1982: 41).

contrasting English and Arabic as in Figure 4.1, Wood observes a preference for the pre-palatal region in Arabic for [i], while English is described as preferring the mid-palatal region for palatal vowels. A pre-palatal constriction location has mostly been observed in languages with a contrast of [i] and [y] or [ɨ] that is enhanced by a maximally high F3 in [i]. While /y/ is not a phoneme in Arabic, the pre-palatal constriction location may nevertheless be useful to distinguish the phonemic contrast of palatal vowels in Egyptian Arabic together with the great range of allophonic variation for all vowel qualities.

To conclude, the articulatory approach to the constitution of vowel systems and phonemic contrasts presented here is considered as an alternative to the traditional description of phonemes in a vowel inventory in terms of height, fronting and rounding of the tongue.

> "*Articulatory features for use in phonology should reflect the preference for four constriction locations and the unique relationship between constricting tongue gestures, muscle situation and the degrees of sensitivity of vocal tract resonances to area perturbations at different parts of the vocal tract (…) … the features of tongue arch height and fronting of the established model are ambiguous in these respects and constitute a capricious medium for relating the different phases of speech production.*" (Wood (1979: 41).

4.8 Vowel perception in L2 acquisition – Levels of description

To describe or explain patterns of perceptual confusion of vowels in L2 acquisition we have to distinguish at least three descriptive levels:

(1) A basic component in describing, explaining or predicting patterns of perceptual substitutions in L2 is a *contrastive analysis* of the learner's native language and the target language. The use or non-use of specific phonemic oppositions and the articulatory gestures and acoustic cues that differentiate these oppositions vary language-specifically. Vowel phonemes which are not used in the learner's L1 will be difficult and may be confused with similar sounds that share several features with the target sound but lack others. These difficulties are *language-specific difficulties*, coming from the listeners' native-language.

(2) Difficulties in the perception of L2 contrasts can also be described as coming from the sounds and their intrinsic features and are here referred to as *vowel-specific difficulties*. These difficulties may either be (a) due to *vowel-inherent characteristics* attributed to a single vowel referring to intrinsic properties such as the relative perceptual salience or robustness of the relevant cues or (b) due to system-specific constellations and contrasts within the *vowel inventories* of the languages involved and the relation between vowel categories within a system.

(3) There is, however, one central component that has so far not been regarded sufficiently: *Listener-specific* factors other than the learners' knowledge of L1 comprise several person-related variables such as age, sex, experience with the target language (age of learning, type of input, length of exposure, experience with specific varieties of L2) and several others. These factors will be discussed in detail in chapter 8. Moreover, as discussed in section 2.3, the *experimental setting* has to be considered to understand the specific "behaviour" of learners in L2 performance. For the moment, it is important to note, that universal tendencies and language-specific strategies alone will never provide a full account for a learner's "behaviour" in L2 neither in an experimental setting nor in every-day communication and language learning. After all, it is the learner's individual constructive activity that determines his/her L2 performance and the acquisition progress (see chapter 8 and section 12.1 and 12.4).

To conclude, two major types of difficulties in L2 performance are distinguished:

(1) *difficulties coming from the sounds* and their inherent universal or language-specific characteristics and

(2) *difficulties coming from the listeners or language learners* and their auditory and perceptual capacities determined by

> (a) the learners' native language (L1),
>
> (b) the linguistic experience in L2 and the quality of the L2 input available during acquisition, and
>
> (c) the learner's individual hypotheses and strategies to compensate and overcome difficulties in the perception of L2 sounds.

In the present study, these two types of difficulties will be attributed as *signal-related biases* and *listener-related biases* (for a detailed discussion of different types of biases, see section 12.4).

4.9 Conclusions

To summarize, the sound inventory of a given language is far from being a random sample of sounds. Rather, sound inventories can be considered as the result of a specific set of selective processes, preferences and constraints, many of which can be explained in terms of universal constraints and preferences that are supposed to be grounded phonetically (Ohala 1983; Maddieson 1999; Flemming 2008) and become manifested in distributional patterns of vowel dispersion in phonetic systems as well as in patterns of frequency of occurrence (see section 4.4, 4.5, 4.6 and 4.7).

For a better understanding of native and non-native vowel identification and categorization, an integrated view of vowels and vowel systems in terms of phonetic parameters, physiological activity, resonator configurations and spectral output is necessary (cf. Wood 1979, 1982). A comprehensive model based on the shape of the entire vocal tract that relates physiology, i.e. articulation and acoustics, with phonology has been introduced in this chapter (see section 4.7.6) for a better understanding of the physical as well as phonological and psychological aspects of non-native vowel perception. Based on the *articulatory-acoustic model* (cf. Figure 4.3) that classifies vowel sounds in terms of constriction locations and of major vowel types, a description of the German vowel system integrating articulatory and acoustic aspects will be offered in the next chapter.

When speaking of vowel systems, reference is often made to the notion of an abstract "vowel space". The *articulatory-acoustic vowel space* is an abstraction of the total set of possible articulatory constellations and resulting resonances in the vocal tract, traditionally described in terms of formant frequency values (F1, F2, F3, ...). The range of possible vowels in the vowel space is delimited by articulatory, acoustic and perceptual coordinates. An acoustic vowel signal can be the result of more than only one articulatory constellation in the vocal tract. Therefore, a phonetic description of vowel sounds in a given language has to comprise articulatory gestures together with their acoustic and auditory effects. Traditional descriptions in terms of the distinctive features Frontness, Height and Rounding are – apart from difficulties to identify their distinct acoustic correlates – not sufficient for an understanding of the complex interplay of articulatory, acoustic and auditory aspects of phonemic elements in native and non-native vowel perception.

For the identification of vowels in continuous speech at least three types of acoustic information are essential: (1) *spectral information* reflecting the articulatory activity to produce a vowel, (2) *dynamic* spectral information in CV- and VC-*transitions* reflecting coarticulatory effects of adjacent sounds, and (3) *durational information* reflecting information on the syllable level, possibly together with intonation cues (F0).

A number of studies and theoretical accounts of vowel inventory typology (Liljencrants & Lindblom 1972; Lindblom 1986; Donegan 1978; Dressler 1985b; 1999; de Boer 2000, 2001; Flemming 2004; Schwartz et al. 2005) have argued that more *salient* and more *robust* categories are favoured in the languages of the world to distinguish vowel sounds resulting in the relatively high frequency of some vowels in the languages of the world.

All theoretical approaches discussed in section 4.7 argue that the constitution of vowel systems is determined by the relative "*salience*" and "*robustness*" of specific vowel contrasts and by the relative "*distance*" between elements in a system. In perception, more *salient contrasts* are expected to be easier to differentiate than less salient ones. Less salient cues might simply be not perceived by the listener resulting in a perceptual substitution with a similar sound lacking this property.

However, as discussed in section 3.2 and 3.3, it is necessary to differentiate *phoneticcue strength* of specific acoustic cues from *perceptual salience* due to the listener's selective attention processes (see section 12.5). Similarity between items as perceived by a given listener is strongly determined by perceptual salience. Perceptual salience however varies language-specifically. Contrast salience can be defined as sufficient distance or dissimilarity between two categories.

Referring to an *integrated articulatory-acoustic model* as introduced in this chapter, specific cues cannot be argued to be auditorily or perceptually salient as such. Rather, it should be argued that some articulatory constellations at so called "*hot spots*" in the human vocal tract (cf. Stevens' Quantal Theory (Stevens 1989), see section 4.7.3) produce acoustic signals of high perceptual stability and minimal ambiguity. Vowel qualities produced at these *hot spots* are fairly resistant to small articulatory deviations such as variation due to contextual effects that has hardly any discernable effect on the listener's interpretation of the output. In other less stable regions of the vocal tract, small articulatory deviations produce a large shift in acoustic output (*quantal effect*).

Languages seem to prefer stable "quantal" regions of the vowel space for the category types /a i u/ that are particularly common in the languages of the world. *Theories on Vowel Dispersion* (Liljencrants & Lindblom 1972; Lindblom 1986; Flemming 1996, 2002; Schwartz et al. 1997b; de Boer 2000, 2001) and *Focalization* (Schwartz et al. 1997b; Schwartz et al. 2005) argue for a (larger) set of so called "focal" vowels (see also Vaissière 2011) that are more robust and distinct in perception (see section 4.7.4).

Universal constraints and *preferences* are postulated to influence the constellation of sound inventories and to act in production and perception in first language acquisition, sound change, loanword phonology and the acquisition of a foreign language (see section 1.2). Assuming that the occurrence of specific vowel qualities in sound inventories and their phonological behaviour is not purely random, the challenge for a *phonological theory* is to account for the full range of *variation* of attested structures as well as for *constraints* on vowel systems in the languages of the world.

Different theoretical accounts such as stronger versions of *Markedness Theory* (e.g. Chomsky & Halle 1968; Trubetzkoy 1939) or *Dispersion Theories* (Liljencrants & Lindblom 1972;

Lindblom 1986; Flemming 1996, 2002; Schwartz et al. 1997b; de Boer 2000, 2001) are based on other than mere physical properties of sounds that are employed by speakers/listeners and languages to differentiate the phonemes in a system. A number of theoretical constructs such as markedness, preferences, universal constraints, maximal dispersion, symmetry, economy, etc. are used to explain the choice of particular elements in a given phoneme inventory and common or universal patterns of distribution and preferences of elements in language systems, whereas weaker definitions of markedness simply refer to cross-linguistic frequency of occurrence.

Perceived distance between two elements is understood here as inverse function of *similarity* (for a detailed discussion of notions of similarity and distance, see chapter 3 and 12). When two sounds are phonetically *similar*, they are expected to be subject to perceptual confusion. A lack of auditory salience of particular acoustic cues in a given signal or contrast may lead to a lack of perceptual contrastivity and a higher degree of *similarity* that can result in perceptual substitutions unless *selective attention weights* raise the perceptual salience of these cues (see section 12.5).

In L2 acquisition, the correct perception of new and similar sounds will be easier if the same or similar cues are used in the native and in the target language. However, as Flege's Speech Learning Model (Flege 1995) states, similarity may also hinder the correct perceptual distinction of L2 contrasts if the acoustic cues indicating phonemic distinctions are less salient (see section 2.4.2 and 3.3).

To conclude, the relative difficulty of sounds in L2 perception is thus determined by the *auditory salience* of physical properties of the signal as well as by *perceptual salience* that is driven by the listeners' phonology and language-experience, their *knowledge* about constraints and preferences and the specific setting of selective attention to specific cues that are of relevance in a given language system.

Substitutions are phonological processes by which one sound is substituted with another similar sound. In cases, where more than one category come into consideration as substitute, the "more salient", more frequently occurring and "less marked" sound category (see section 4.3 and 4.4) is expected to be selected as perceptual substitute rather than a less salient, less common and more marked sound. In other words, acoustic *cue weighting* goes along with processes of *category weighting* by probability rating for the occurrence of a specific category (cf. Pierrehumbert 2003). These processes of weighting and probability calculation are on the part of the listener. As a result of these processes, the listener will show preferences for specific categories and dispreferences for others that correlate with language-specific and/or universal patterns.

To summarize the considerations of the preceding chapters, in L2 acquisition *constraints and preferences* form part of the listeners' or learners' linguistic knowledge and are manifested at different levels of language and language systems:

(1) in general *physiological* conditions and limitations of human articulation and speech perception and their influence on the mental categorization of speech *signals*,

(2) in *language-specific* conditions and limitations on the phonological system under consideration, and

(3) in the L2 leaners' *individual interlanguage system* that is strongly determined by his/her linguistic knowledge in one, two or more languages.

5 The German Vowel System

This chapter presents a phonetic and phonological description of the German vowel system. The discussion will focus on aspects of system size and on the phonetic correlates of phonemic oppositions that are commonly described in terms of quantity and/or quality distinctions. The acoustic-phonetic data presented here for German vowel phonemes are based on an analysis of the input stimuli used in the experimental part of the present study.

The last section of this chapter presents an analysis of phonetic similarity of German vowel phonemes in terms of acoustic-phonetic properties by means of hierarchical clustering. Together with the contrastive analysis of the listeners' native languages (see chapter 10), phonetic similarity by hierarchical cluster analyses will serve as a base to predict perceived similarity of German vowel categories and perceptual substitutions occurring in L2 vowel perception.

5.1 Typological description

Typologically, German is one of the languages with a very rich vowel inventory. Three main features characterize the German vowel system: (1) the phonemic distinction of long tense and short lax vowels, e.g. *Wahn* 'delusion' *vs.Wann* 'when', (2) the existence of a contrast of front rounded vs. front non-roundeded vowels, e.g. *Tür* ' door' *vs. Tier* 'animal', and (3) the existence of diphthongs vs. monophthongs (Roelcke 1997: 23, 2003: 31). Another characteristic feature is the occurrence of vowel alternations (*Ablaut* and *Umlaut*) within word stems, historically derived from morphonological alternations. The German vowel repertoire in unstressed syllables is strongly reduced.

The German consonant inventory comprises a rich set of obstruents (plosives, fricatives and affricates). Complex consonant clusters (3-4 adjacent consonants in pre-vocalic or post-vocalic position within a syllable, e.g. */ʃɪmpfst /* – 'to insult' (2nd sg), */ʃtraikst/* – 'to strike' (2nd sg) are another characteristic feature of German.

German can be characterized as stress-timed language with free word stress. Word formation in German is described as synthetic-analytic.

5.2 System size

The vowel inventory of Standard German (SG) is most commonly described as consisting of a set of fourteen full vowels in stressed positions, including a series of typologically marked front rounded vowels /y: ʏ ø: œ/, which are also called "*Umlaut*"-vowels, and two vowels /ə, ɐ/ occurring only in unstressed position (e.g. Jørgensen 1969; UPSID Maddieson 1984: 265; Kohler 1999).

Most analyses of Standard German describe two a-qualities, a long and a short one.

The phonemic status of /ɛ:/ is controversially discussed. For some regional standards, /ɛ:/ can be described to be phonemic, while in several others the opposition between /e:/ and /ɛ:/ is neutralized (see Kleiner & Knöbl 2011: 6f).

The large set of long vowels historically developed from lengthening in open tone-syllables and spread through analogy (Roelcke 1997: 70ff; Ronneberger-Sibold 1999). In addition to quality, German uses quantity contrasts, i.e. length. For the seven contrasting pairs /i:/-/ɪ/, /e:/-/ɛ/, /ɑ:/-/a/, /u:/-/ʊ/, /o:/-/ɔ/, /y:/-/ʏ/, /ø:/-/œ/ patterns of systematic co-variance of quantity and tenseness are observed (the only exception is /ɛ:/): Short vowels are usually described to be *lax* as opposed to the long *tense* vowel qualities.

However, phoneticians and phonologists differ considerably in their views concerning the number of distinctive vowel phonemes in German as well as the nature of the phonetic and phonological features that differentiate these oppositions. Differences between the various descriptive accounts mostly refer to the phonemic status of the e-schwa /ə/ and /ɐ/, to the quality of a-qualities, to the status of /ɛ:/ (<ä>) and to the distinctive features and phonetic correlates that account for the pairs mentioned above.

The exact number of phonemes and the distinctive features are moreover considered to vary regionally, even in the standard varieties (Kleiner & Knöbl 2011).The number of vowel phonemes for Standard German (SG) has been set as 17 (Wiese 1996), 16 (Jørgensen 1969; UPSID Maddieson 1984: 265; Kohler 1999), 15 (Moulton 1962; Sendlmeier 1981; Iivonen 1987a, b) or 8 (Becker 1998). Becker (1998) claims that the German vowel system can be described as consisting of eight vowels that can occur as short or long, depending on their position in the syllable. For Standard Austrian German (SAG), 13 vowels are distinguished (Moosmüller 2007).

The major differences between *Standard German* (SG) and *Standard Austrian German* (SAG) concern

(1) the difference in quality between the long and the short /*a*/-vowel, which is neutralized in Standard Austrian German,

(2) the realization of /ɛ:/, which is merged with /e:/ in Standard Austrian German, e.g. <*ä*> in “Mädchen” (“girl”) is realized with [e:] in SAG and in some regions of Germany,

(3) the quality of the *e-schwa* in unstressed syllables, which has no phonemic status in Standard Austrian German, e.g. SAG [ge’ge:bɛn] vs. SG [gə’ge:bən], and

(4) the contrast of /i:/ - /ɪ/, /y:/ - /ʏ/, and /u:/ - /ʊ/ in quality and/or quantity, where tendencies towards neutralization are observed in SAG, but phonemically the contrast still exists (Moosmüller 2007; Brandstätter & Moosmüller, in press).

5.3 Distinctive features

As has been stated above, phonetic and phonological descriptions of the German vowel system differ considerably with respect to the phonological features and their phonetic correlates to distinguish German vowel phonemes.

Table 5.1summarizes the phonological features and labels that are most commonly used to refer to phonemic distinctions in German.

	front				back		
	unrounded		rounded				
	palatal				non-palatal[25]		
	non-labial		labio-palatal		labial		
	close[26] tense	open lax[27]	close tense	open lax	close tense	open lax	
high[28]	i:	i [ɪ]	y:	y [ʏ]	u:	u [ʊ]	chromatic[29]
mid[30]	e: ä: [ɛ:]	e [ɛ], (ä [ɛ])	ø:	ø [œ]	o:	o [ɔ]	
low					a: / ɑ:	a/ ɑ	achromatic
	long	short	long	short	long	short	

Table 5.1: Phonological features used for the description of German vowels in stressed syllables

[25] also called “velar“. The term “velar” is avoided here to refer to the class of back vowels because of articulatory differences concerning the constriction location for a-, o- and u-vowels.

[26] e.g. Kohler (1999). Another definition of [open] will be discussed in section 5.4.4 (see also Moosmüller 2007: 79ff).

[27] the [tense] vs. [lax] feature is commonly used only for high and mid vowels but not for the low vowel /ɑ/.

[28] referring to tongue height.

[29] referring to palatality and/or labiality, a feature that is also called “colour” or “chromaticity” (Donegan 1978).

[30] Several studies distinguish between higher mid (e.g. /e, o/ and lower mid /ɛ, ɔ/, e.g. UPSID Maddieson 1984).

The *tense vs. lax*-opposition and its phonetic correlates are one of the most controversially discussed issues in German phonology. Contrasts of the type /iː/ vs. /ɪ/ or /eː/ vs. /ɛ/ have traditionally been described as an opposition between *long tense vs. short lax* vowels (see Table 5.2). The duration contrast within these tense vs. lax vowel pairs is reportedly larger in German than in English (e.g. Iivonen, 1987a, b; Weiss 1976; Bohn & Flege 1990, 1992).

tense long			short lax		
Miete	*rent*	iː	Mitte	*middle*	ɪ
Meter	*meter*	eː	Mette	*Christmas Mass*	ɛ
Mode	*fashion*	oː	Motte	*moth*	ɔ
Mut	*audacity*	uː	Mutter	*mother*	ʊ
müde	*tired*	yː	Mütter	*mothers (pl)*	ʏ
möge	*may*	øː	möchte	*would like*	œ
Mate	*Paraguay tea*	aː	Matte	*mat*	a
Mär	*fairy tale*	ɛː	März	*March*	ɛ

Table 5.2: Minimal pairs in German stressed syllables

As has been pointed out already, long /ɛː/ <ä> falls out of this typical dichotomy of tense long vs. short lax vowels: It is either assumed that it has no short counterpart or that its counterpart is short /ɛ/.

Generally, the tense/lax-opposition seems to occur more commonly with non-low vowels and is rather rare with low vowels (Stevens 1998: 296). For German, the tense/lax opposition in terms of quality and quantity is rarely posited for a-vowels; therefore the same phonetic symbol for the long and the short a-vowel is commonly used in descriptions of the German system (e.g. Valaczkai 1998; Kohler 1999; see section 5.4.3 and 5.4.5.1). However, Stevens (1998: 296) describes that a tense/lax differentiation is also possible for low vowels: the lax counterpart of a low back [ɑ] is achieved by adjustment of the tongue body to a less extreme back position to yield [a]. [a] may however also be the lax counterpart of front [æ] (Stevens 1998: 296).

In monosyllabic words, only one consonant can follow a long tense vowel or a diphthong, whereas after lax vowels a more complex coda is allowed. In bisyllabic words, lax vowels are considered to be followed by an ambisyllabic consonant, resulting in a closed syllable (/muːt/ 'audacity' vs. /ˈmʊ.tɐ/ 'mother' vs. /ˈmʏ.tɐ/ 'mothers', cf. Ronneberger-Sibold 1999). Concerning the covariance of tenseness and phonological length, there is considerable disagreement about the question whether quality or quantity is the primary feature of the opposition. Two major concepts may be distinguished: (1) tenseness as a segmental feature vs. (2) tenseness refering to the domain of the syllable ("syllable-cut" theory). Segment-based theories differ with respect to the question which feature is primary, quality or quantity or both. Referring to the question of the underlying criteria of this opposition, the contrast has been described to be based on

1. quantity distinctions (Hertrich & Ackermann 1997 for SG), or
2. quality distinctions (referring to notions such as „tenseness" or articulatory differences and formant frequency values (for SG Iivonen 1987a, b; Wiese 1996; Kohler 1999; for SAG Moosmüller 2007), or
3. quantity and quality distinctions (Jessen et al. 1995 for SG), or
4. suprasegmental features, especially on syllabification and the so-called syllable cut[31] (for SG Vennemann 1991; Becker 1998), or
5. quantity relations of Middle and North Bavarian dialects (Ronneberger-Sibold 1999) referring to the abandoning of a three-way quantity distinction in Classical Middle High German, which has led to a quality opposition in Standard German (for SAG, see Moosmüller 2007).

To summarize, we observe two different traditions in the literature concerning the tense/lax contrast: (1) One that considers tenseness as a separate property and (2) one subsuming these contrasts with adjustments in height. Some authors even mix these traditions referring to /i/ vs. /ɪ/ and /u/ vs. /ʊ/ as a tenseness contrast while describing the contrast of /e/ vs. /ɛ/ and /o/ vs. /ɔ/ as a contrast in height, an analysis that would result in a distinction of four degrees of height rather than three (Wood 1992: 284).

It has to be added that – from a phonetic point of view – the tense-lax opposition clearly refers to more than only one articulatory feature and its acoustic effects (see section 5.4). Moreover, we have to consider that the phonetic features the opposition is based on may differ regionally, resulting in the differences described above for Standard German (SG) vs. Standard Austrian German (SAG).

For SAG, Moosmüller (2007) states that *duration* is not the primary relevant feature for the tense/lax opposition, since the durational dimension of vowel contrasts in production tasks is

---

[31] A long vowel has a long, continuous decrease of intensity. The transition to a following consonant or word boundary is called "smooth/soft cut". The connection between a long vowel and the following consonant is characterized as "loose contact". A short vowel is cut immediately after its intensity peak, a fact that is called "abrupt/sharp cut" or "close contact" (Trubetzkoy 1939; Ronneberger-Sibold 1999: 249; Mooshammer 2000).

not maintained in all task types and speaking styles (e.g. reading, word list, spontaneous speech), while quality differences are preserved. Moosmüller (2007: 78ff) concludes that *degree of constriction* is the best candidate to describe the so-called tense-lax opposition. This distinction would also account for those cases, where /ɑ/ and /a/ are not neutralized, since for /ɑ/ a higher degree of constriction in the pharyngeal region is observed (Valaczkai 1998; Pouplier et al. 2004; Moosmüller 2007: 81), while the feature [+/–front] is not appropriate in description.

To conclude, it is advisable to investigate the phonetic correlates of phonemic oppositions for a given regional variety. The phonetic characteristics of German phonological contrasts will be discussed in the following sections. The discussion will refer to the phonological description of phonemic oppositions as well as to their phonetic correlates and will focus on the description of Standard Austrian German, for which a detailed acoustic analysis has been presented by Moosmüller (2007). Results from the acoustic-phonetic analysis of the input stimuli of the present study will exemplify the acoustic issues discussed in this section.

The input stimuli are read by a male speaker of Southern German/Standard Austrian German. A detailed description of the structure of the input material is presented in section 6.3. The acoustic-phonetic analysis of the data (formant frequencies, duration, F0) is summarized in the present chapter and will serve as a base for predictions on similarity relationships and perceptual substitutions in L2 perception (see section 5.8). Figure 5.1 to 5.14 present acoustic measurements of two linear time-aligned tokens of logatoms of the type /pVt/ embedded in the phrase-final position of a constant carrier sentence (see 6.3)

5.4 Phonological contrasts and phonetic correlates

A specific vowel contrast can be produced by several different constellations in the vocal tract. For a complete understanding of the phonemic inventory of a given language, it is therefore essential to describe those factors that are constitutive for the phonemic elements of the system together with their acoustic-auditory effects rather than to confine oneself to a description in terms of phonological features. The description of the German vowel system presented here will be based on the articulatory-oriented approach to vowel classification that has been introduced in chapter 3. Table 4.3 presents a first overview over distinctive properties within an acoustic-articulatory-oriented approach as it has been proposed for Standard Austrian German based on an acoustic analysis by Moosmüller (2007).

	/i/	/ɪ/	/y/	/ʏ/	/e/	/ɛ/	/ø/	/œ/	/ɑ/[32]	/o/	/ɔ/	/u/	/ʊ/
constricted	+	–	+	–	+	–	+	–	+	+	–	+	–
open[33]	–	+	–	+	–	+	–	+	+	–	+	–	+
round	–	–	+	+	–	–	+	+	–	+	+	+	+
front	+	+	+	+	+	+	+	+	–	–	–	–	–
lower pharyngeal	–	–	–	–	–	–	–	–	+	–	–	–	–
velar	–	–	–	–	–	–	–	–	–	–	–	+	+
pre-palatal	+	+	+	+	–	–	–	–	–	–	–	–	–

Table 5.3: Feature matrix of the vowels assumed for Standard Austrian German based on an acoustic analysis (Moosmüller 2007: 121)

5.4.1 Constriction location

A primary differentiation of vowel qualities refers to the *constriction location*. By a constriction in the oral cavity, the resonances of the vocal tract are modified, resulting in variation of the acoustic parameters. Further modifications of the vocal tract resonances are induced by additional articulatory gestures like degree of oral aperture, lip rounding/protrusion etc.

Stevens (1972, 1989) identified zones in the vocal tract where spectral data are relatively insensitive to small articulatory variation. These observations are compatibel with the concept of discrete constriction locations defined by Wood (1979, 1991; see chapter 3).

Fant (2001: 45f) describes three major areas of tongue constriction for Swedish vowels: a front region with a constriction located less than 4 cm from the teeth, a mid region (4-7cm from the teeth) and a back region for all back vowels with a distance of more than 7cm[34].

Constrictions in the areas described by Wood, Fant and Stevens induce vowel spectra that are characterized by the proximity of two adjacent formants which Stevens (1989) interprets as perceptual enhancement ("center of gravity") of a specific vowel quality (see also Stevens & Keyser 1989; Stevens & Keyser 2010). A typical example for enhancement is lip rounding in non-low back vowels. But the proximity of adjacent formants is also characteristic for other vowel types. Non-low front vowels are characterized by the proximity of F2 and F3, or F3 and F4.

[32] Moosmüller (2007: 120) states that qualitative differences between /ɑ/-vowels occur only sporadically and that for SAG only one /ɑ/-vowel can be assumed.

[33] referring here to degree of oral aperture (Valaczkai 1998: 79; Wood: 1982: 145)

[34] Fant (2001: 45f) describes the vowel [u:] as located at the border between mid and back regions and [æ:] fitting best into the back category.

For German, as for other languages, we find at least four different constriction locations (Wood 1979, 1986, 1991): the area of the hard palate (for i- and e-like *palatal* vowels), the soft palate (for u-like *velar* vowels), the upper pharyngeal region (for o-like *uvular* vowels), and the lower pharyngeal area (for a-like *pharyngeal* vowels). There seems to be no evidence for so-called central vowels.

	Constriction location			
	[i]- and [e]-vowels palatal	[u]-vowels palatovelar	[o]-vowels pharyngovelar	[a]-vowels low pharyngeal
pharyngeal	-	-	+	+
palatal	+	+	-	-
velar	-	+	+	-

Table 5.4: Feature specifications for the four basic tongue body manoeuvres (Wood 1990: 199)

In some languages, we find a further distinction in the palatal region of the vocal tract between *pre-palatal* and *mid-palatal* vowels to differentiate the pre-palatal i- and y-vowels from the mid-palatal e- and ø-vowels (Wood 1986, 1991). While for larger systems like e.g. Swedish with three high vowels /y: ʉ i:/ (cf. Fant 2001), a distinction of mid-palatal and pre-palatal vowels is often described, we observe a pre-palatal location of constriction also in some languages with a smaller inventory like Russian or Egyptian Arabic. For Standard Austrian German, we may assume a pre-palatal and a mid-palatal region that serves to differentiate the i- and y-vowels from the e- and ø-vowels (Moosmüller 2007).

The following sections will describe the German vowel system referring to the basic classification of constriction location and to secondary distinctive properties as lip rounding, constriction degree etc. Examples of the acoustic-phonetic analysis of the input material that was used in the present study (see also chapter 5) are presented to illustrate the characteristic properties of vowel contrasts and to explain possible perceptual problems in L2 learning.

The figures in the following sections are based on the acoustic analysis of vowel items used in the perception test as described in chapter 6. For each vowel two items in the same context /pVt/, one with falling and one with rising F0, were analyzed. The formant frequency contour of the entire nucleus was measured in overlapping frames. The values presented here are mean frequencies calculated on a framewise comparison after time-alignment for each vowel type separately. The abscissa represents the time axis in terms of number of frames (NoFr) of the acoustic analysis. A direct comparison in terms of absolute duration of the two members presented in the diagram, however, is not possible, since linear time alignment only refers to vowels of the same category. For a comparison in terms of absolute duration see section 5.4.5.

5.4.1.1 Pharyngeal vowels

German pharyngeal *a*-vowels are located in the *lower pharynx*. Pharyngeal refers to the activity of the hyoglossal and/or pharyngeal constrictor that narrows the lower pharynx. A central characteristic of low vowels is the high F1. By constricting or narrowing the vocal tract in the lower pharyngeal region, F1 is raised above 600 Hz for low pharyngeal vowels like [ɑ a æ]-vowels (Wood 1990: 200; Stevens 1998: 274).

Low vowels with a narrowing or constriction in the lower pharyngeal region may fall into two groups: A low vowel with a constriction about 8 cm (of a total length of 16 cm) from the glottis resulting in a proximity of F1 and F2 is characterized as back /ɑ/, whereas a low vowel with a back cavity constriction of about 4 cm together with a smaller distance of F2 and F3 is associated with front /a/ or /æ/ (Stevens 1998: 275f). F2 is primarily associated with the relative length of the back cavity, exhibiting a minimum value at 8 cm (of 16 cm) back cavity length (Stevens 1998: 268ff, 274f). Thus, acoustically, two qualitatively different configurations of natural frequencies depending on cavity length are given: Either F2 is maximally low and close to F1, or F2 is maximally high and close to F3.

For German, traditionally, two a-qualities are described, a front /a/ and a back /ɑ/[35]. For /ɑ/, the degree of lip aperture and the degree of constriction is higher than for /a/ (Valaczkai 1998). Compared to /a/, a higher constriction in the pharynx is observed for /ɑ/ (Wood 1990: 196; Moosmüller 2007: 81; Valaczkai 1998; Pouplier et al. 2004).

Moosmüller's (2007: 112ff) acoustic analysis of Standard Austrian German shows no differences in constriction location between the short and the long a-phoneme and only occasionally differences in constriction degree and degree of lip aperture, even in formal speaking styles. Therefore, only one /ɑ/-quality is assumed for Standard Austrian German. For more formal styles, however, a distinction at least in terms of duration will be assumed here. The acoustic analysis of the input stimuli of the present study supports this assumption, see Figure 5.1 illustrating the formant movements over time for long /ɑː/ (209 ms[36]) and short

[35] The notions "front" and "back" are commonly used to refer to the "highest point of the tongue" but refer here primarily to the position of the *tongue body* for a-vowels. For front /a/, the tongue body is displaced forward, maintaining a narrowing in the lower pharynx, the tongue tip touching the lower incisors (Stevens 1998: 274).

[36] mean duration in the /pVt/-context.

/ɑ/ (116 ms). The relative proximity of F1 and F2 indicates that we deal with a "back" /ɑ/-quality for both the long and the short /ɑ/.

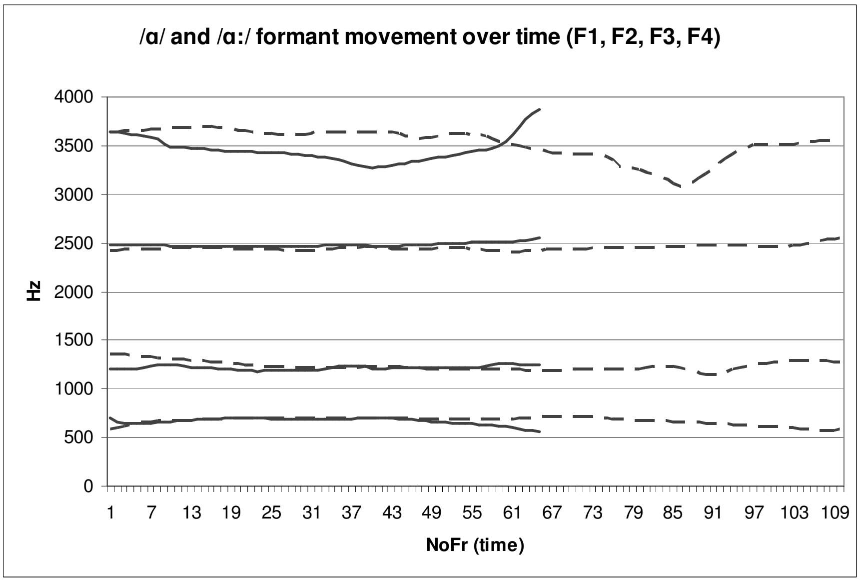

Figure 5.1: /ɑ/ (full line) and /ɑ:/ (dashed line) formant movement over time in the /pVt/-context. Mean formant values (F1, F2, F3, F4) and standard deviation (SD) in Hz, duration (in number of frames)[37]

5.4.1.2 Uvular vowels

The German uvular /o:/-vowels are located in the *upper pharyngeal region*. By a pharyngeovelar, i.e. uvular movement the upper pharynx there is a narrowing for [o, ɔ, ɤ]-like vowels (Wood 1990: 198). A constriction in this region causes a low F1, a low F2, and a high F3. For rounded [o]-vowels, F2 is about 800 Hz, while F1 is about 550 Hz for [o] and [ɔ] (Wood 1990: 201)[38] due to inactiveness of the hyoglossi and middle pharyngeal constrictors.

The spectral difference between /o:/ and /ɔ/ appears more conspicuous (see Figure 5.2), a fact that is mainly due to the higher degree of lip opening that is observed for /ɔ/, while for /o:/ lip protrusion is higher (Valaczkai 1998: 130, 132).

For /ɔ/, F1 and F2 are higher and F3 may be somewhat lower than for /o:/. However, the most stable difference between /ɔ/ and /o:/ seems to be F1, indicating a higher degree of opening

[37] /ɑ/: F1 668 Hz (34 SD), F2 1217 Hz (21 SD), F3 2483 Hz (18 SD), F4 3457 Hz (129 SD), 116 ms /ɑ:/: F1 673 Hz (39 SD), F2 1237 Hz (47 SD), F3 2451 Hz (24 SD), F4 3522 Hz (160 SD), 209 ms

[38] The values for F1 (550Hz) as described by Wood (1990: 201) are relatively high and not confirmed for German by the data in the present study where F1 is considerably lower (mean F1 335 Hz for /o:/ (10 Hz SD) and 452 Hz for /ɔ/ (21 Hz SD) in the /pVC/-context).

for /ɔ/. For SAG, Moosmüller (2007: 112) observes a tendency towards contrast neutralization of /o:/ and /ɔ/ in spontaneous speech.

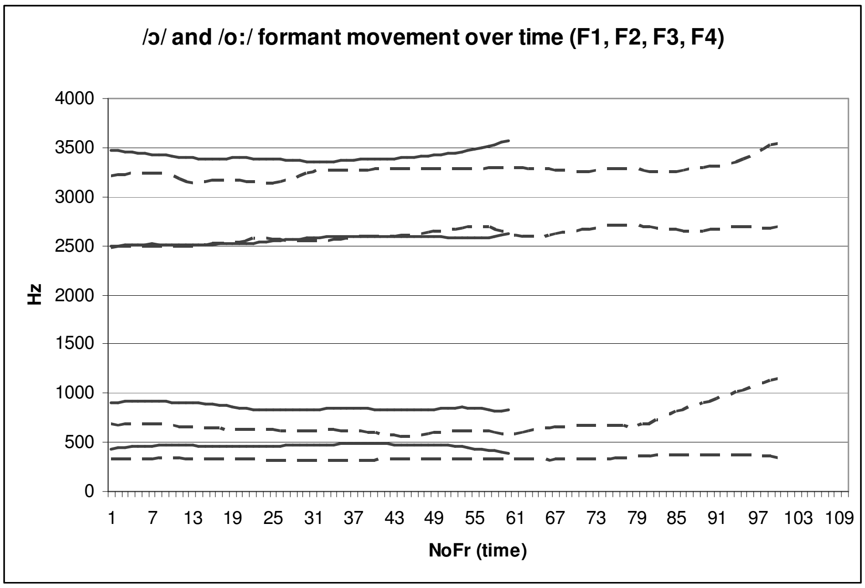

Figure 5.2: /ɔ/ (full line) and /o:/ (dashed line) formant movement over time in the /pVt/-context. Mean formant values (F1, F2, F3, F4) and standard deviation (SD) in Hz, duration (in number of frames)[39]

To summarize, we may expect that the German /ɔ/ vs. /o:/-contrast will be discriminated quite well by learners of L2 German. The differentiation of the comparatively open /ɔ/-quality is expected to be easier as compared to those qualities with a small degree of mouth aperture.
The contrast between /o:/- and /u:/-qualities (see below) may however cause problems in L2 perception due to the relative small spectral differences, especially if the uvular vowel in the L2 learners' native language is an /ɔ/ rather than an /o:/-quality and if durational differences do not function phonemically in the learners' L1.

5.4.1.3 Velar vowels

The pharyngovelar vowels /u:/ and /ʊ/ are located in the velar region of the vocal tract. Velar refers to the styloglossal activity drawing the tongue towards the nasopharynx. A constriction in the velar region, i.e. a backed tongue body, causes a lowering of F2 and the proximity of F1 and F2 together with a high F3. An acoustic consequence of a low F2 which is close to F1 is that the amplitudes of the higher frequency peaks in the spectrum are low relative to the

[39] /ɔ/: F1 461 Hz (20 SD), F2 857 Hz (33 SD), F3 2558 Hz (38 SD), F4 3413 Hz (49 SD), 114 ms /o:/: F1 337 Hz (19 SD), F2 695 Hz (138 SD), F3 2610 Hz (71 SD), F4 3263 Hz (73 SD), 211 ms

amplitudes of F1 and F2. Stevens (1998: 283) assumes that the higher frequencies may therefore play a less significant role in determining the quality of non-low back vowels.

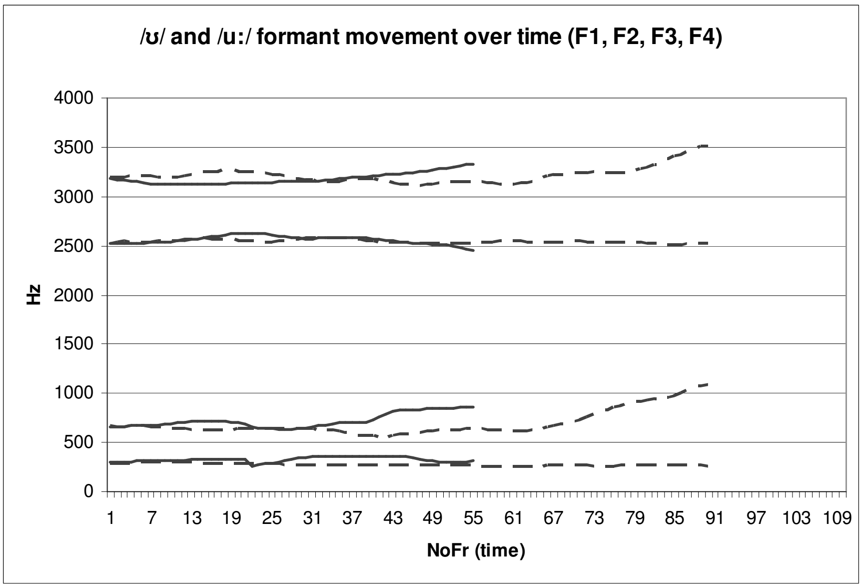

Figure 5.3: /ʊ/ (full line) and /u:/(dashed line) formant movement over time in the /pVt/-context. Mean formant values (F1, F2, F3, F4) and standard deviation (SD) in Hz, duration (in number of frames)[40]

The lower F1 for /u:/ (see Figure 5.4) may be explained by the higher degree of lip protrusion and the resulting lengthening of the vocal tract as compared to /o:/ (Valaczkai 1998: 130f).

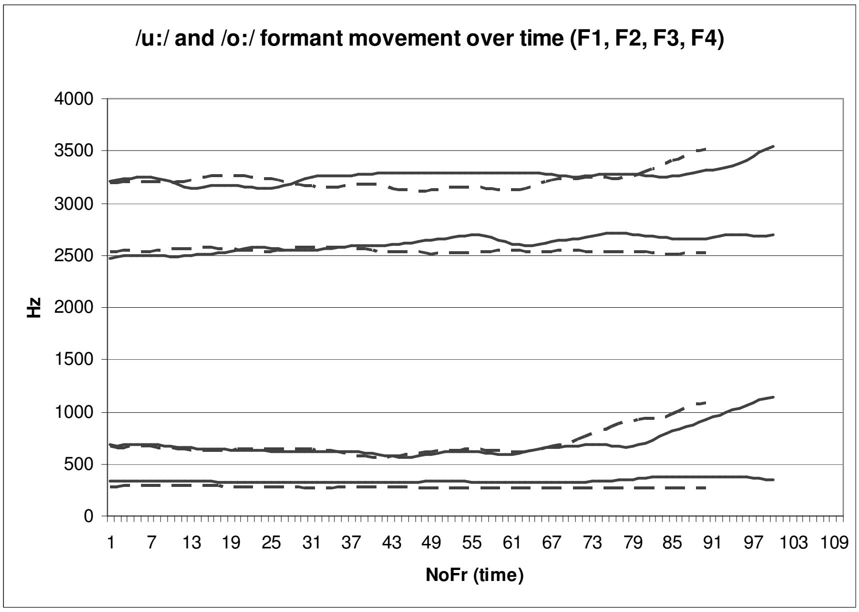

Figure 5.4: Comparison of /u:/ (dashed line) and /o:/ (full line) in the /pVt/-context

40 /ʊ/: F1 326 Hz (37 SD), F2 724 Hz (88 SD), F3 2558 Hz (52 SD), F4 3182 Hz (101 SD), 98 ms /u:/: F1 276 Hz (10 SD), F2 697 Hz (131 SD), F3 2544 Hz (18 SD), F4 3219 Hz (86 SD), 183 ms

Spectral differences between /u:/ vs. /o:/ and/u:/ vs. /ʊ/ are relatively small (see Figure 5.3 and Figure 5.4). Lip protrusion, generally leading to a decrease of formant values especially for F2, causes hardly any variation of the low F2 values with back vowels even with changes in the constriction location, i.e. there are only slight changes in F2 in an area of 4-8 cm from the glottis, although the constriction location is uvular for /o:/ and velar for /u:/. Therefore, we observe only small differences in F2 for uvular o-like and velar u-like vowels, at least in a Hertz-scaled description. Acoustically, these vowel qualities are distinguished by F1 and F3, which are both higher for /o:/ than for /u:/, while having very similar F2 values. F3 values change at a distance of about 6 cm from the glottis (cf. Fant 2004). This may explain the problems in perceptual discrimination of German o- and u-vowels that are often observed with L2 learners of German.

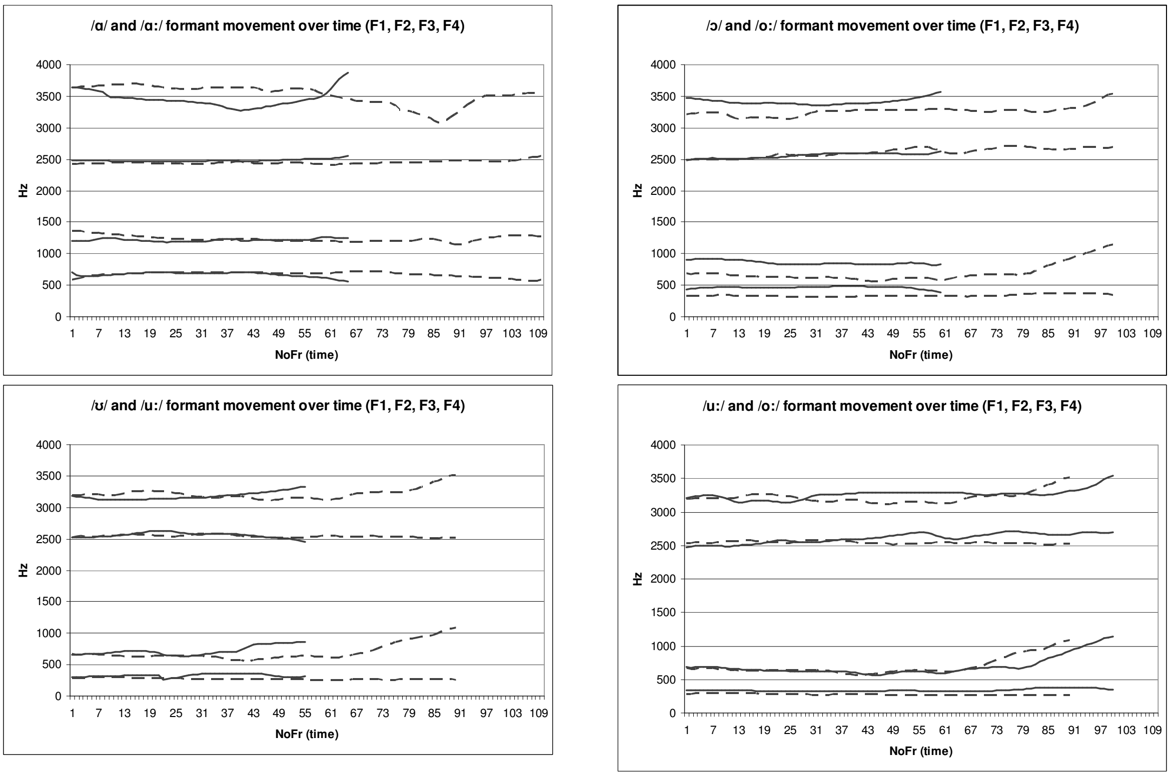

Figure 5.5: Comparison of back rounded vowels (F1, F2, F3, F4 in Hz, duration in number of frames)

While F2 is very similar for German /u:/ and /o:/ (see Figure 5.4), F1 and F3 are both higher for /o:/ than for /u:/ (see Figure 5.5). If however, listeners are not accustomed to the use of these distinctive acoustic cues, the differentiation of /u:/ and /o:/ will be at risk, since the spectral differences between these two qualities are not too conspicuous. Moreover, mentions, the amplitudes of the higher frequencies F3 and F4 for non-low back vowels are relatively low and are therefore less easily available for the listeners (Stevens 1998: 283), a fact that may also account for difficulties in the perceptual differentiation of these u- and o-qualities. Problems in the perceptual differentiation of velar u-like and uvular o-like qualities are expected especially if the /o:/-quality in the learners' L1 is an open [ɔ] rather than a more constricted [o].

5.4.1.4 Palatal vowels

German has eight front vowels that are located in the *front region* of the vocal tract. A differentiation of these eight vowels in terms of *pre-palatal* /i/- and /y/-vowels and *mid-palatal* /e/- and /ø/-vowels (cf. Wood 1979, 1982, 1986) is proposed here to be relevant at least for Standard Austrian German (see the acoustic analysis of Standard Austrian German by Moosmüller 2007).

Acoustically, pre-palatal *i*-vowels and mid-palatal *e*-vowels are clearly discerned by F1 indicating a difference in the degree of lip aperture. However, the most relevant parameter for the differentiation of the pre-palatal vs. the mid-palatal constriction location is F3.The *pre-palatal region* of the vocal tract is a highly sensitive area. Small articulatory displacements involve considerable changes of the acoustic signal. A constriction in the pre-palatal region causes a comparatively high F3. This high F3 may be lowered however by a shorter constriction length or changes of tongue height. A constriction in this region can cause a switch of cavity affiliation for F2 and F3 when the constriction degree is small and the constriction length is rather long (around 5 cm in more formal speaking styles), causing a substantial raise of F3 approximating F4 (Moosmüller 2007: 86).

Moosmüller's (2007) acoustic analysis shows that in Standard Austrian German the /i/-vowels are located exactly at or very near to a sensitive boundary where formant frequencies converge only by a slight displacement of the tongue position or by a reduction of constriction length which causes a considerable change in formant frequency values, especially for F3 (Moosmüller 2007: 47). It is important to note however that – in terms of acoustics –the pre-palatal constriction location can only be verified for /i:/; for unconstricted /ɪ/ as well as for rounded /y:/ and /ʏ/ (see 4.5.3), F3 is lower and affiliated to the back cavity due to a lower degree of constriction and/or due to lip rounding.

The more vowels a language has in the palatal region the more additional cues have to be used for phonemic differentiation. In German, further differentiations in the palatal region are established by differences in constriction degree and lip rounding (see section 4.5.2 and 4.5.3).

Perceptual problems with the differentiation of front vowels are expected in L2 acquisition of German and are assumed to be due to the high number of vowels in the front region of the vocal tract, especially if the listener's L1 does not have a comparable repertoire of front vowels and if the learners have not developed sufficient sensitivity to the acoustic cues indicating phonemic boundaries between pre-palatal and mid-palatal vowels in L2.

Figure 5.6 shows the frequency pattern over time for /i:/ and /ɪ/. The analysis of the speaker reading the input stimuli for the present study shows high values for F3 and the proximity of F3 and F4 for /i:/, clearly indicating the use of the pre-palatal constriction location for /i/-vowels.

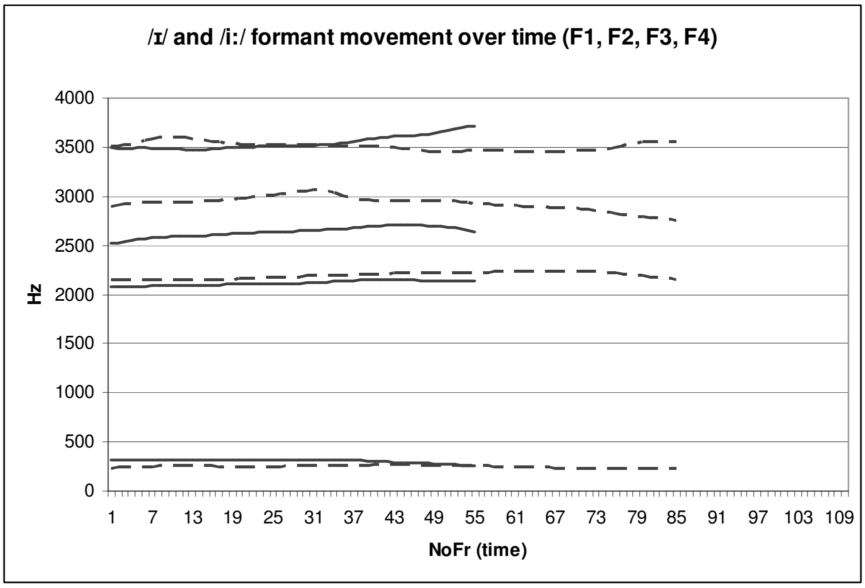

Figure 5.6: /ɪ/ (full line) and /i:/ (dashed line) formant movement over time in the /pVt/-context. Mean formant values (F1, F2, F3, F4) and standard deviation (SD) in Hz, duration (in number of frames)[41]

41 /ɪ/: F1 304 Hz (16 SD), F2 2117 Hz (23 SD), F3 2638 Hz (51 SD), F4 3546 Hz (70 SD), 100 ms /i:/: F1 248 Hz (12 SD), F2 2195 Hz (34 SD), F3 2929 Hz (72 SD), F4 3512 Hz (44 SD), 144 ms

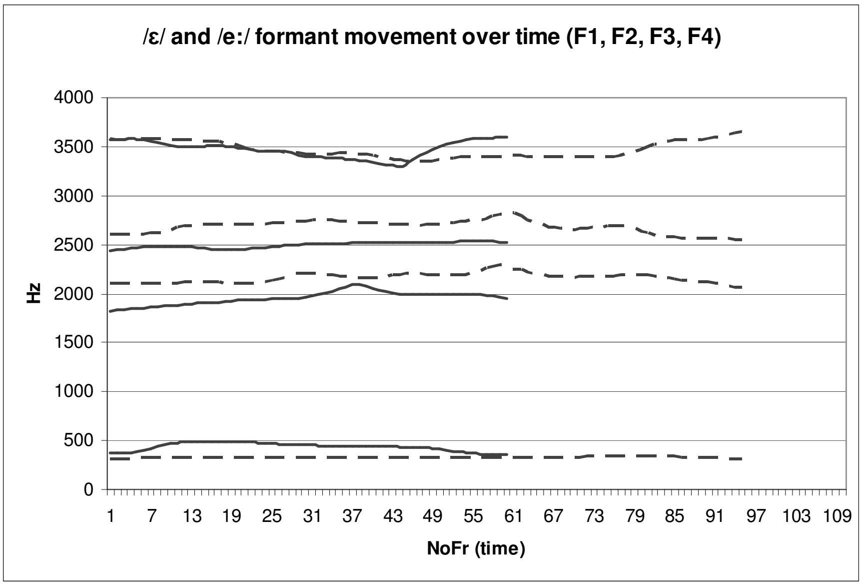

Figure 5.7: /ɛ/ (full line) and /e:/ (dashed line) formant movement over time in the /pVt/-context. Mean formant values (F1, F2, F3, F4) and standard deviation (SD) in Hz, duration (in number of frames)[42]

For /e:/, due to mid-palatal constriction and concomitant back cavity affiliation of F3, F3 is lower than for /i:/ (see Figure 5.9). A center of gravity is formed by F2 and F3. F2 for /i:/ is either lower or identical with /e:/, while F1 and F3 are more conspicuous cues for the differentiation of /i:/ and /e:/ (Moosmüller 2007: 100f).

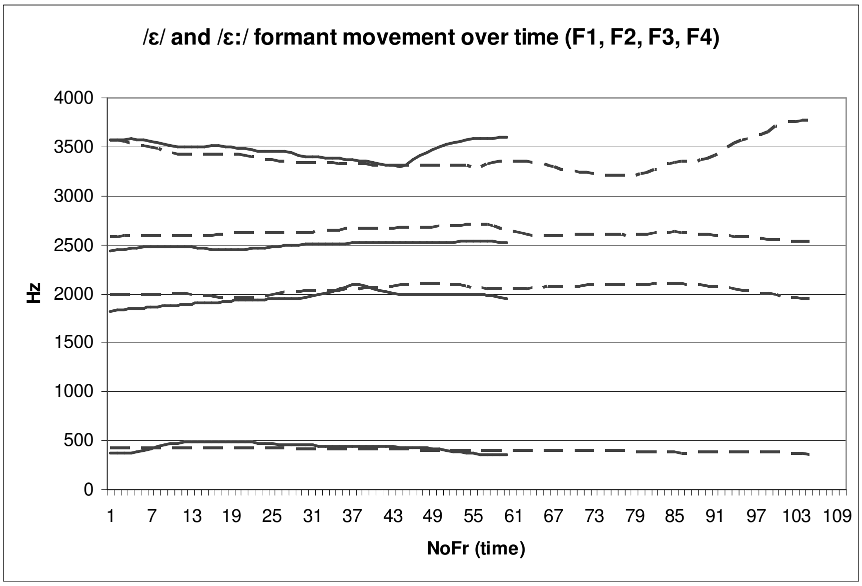

Figure 5.8: /ɛ/ (full line) and /ɛ:/ (dashed line) formant movement over time in the /pVt/-context. Mean formant values (F1, F2, F3, F4) and standard deviation (SD) in Hz, duration (in number of frames)[43]

[42] /ɛ/: F1 435 Hz (42 SD), F2 1953 Hz (65 SD), F3 2498 Hz (30 SD), F4 3471 Hz (87 SD), 112 ms /e:/: F1 331 Hz (6 SD), F2 2163 Hz (48 SD), F3 2685 Hz (66 SD), F4 3470 Hz (85 SD), 202 ms

[43] /ɛ/: F1 435 Hz (42 SD), F2 1953 Hz (65 SD), F3 2498 Hz (30 SD), F4 3471 Hz (87 SD), 112 ms /ɛ:/: F1 406 Hz (18 SD), F2 2043 Hz (47 SD), F3 2621 Hz (42 SD), F4 3387 Hz (133 SD), 191 ms

A comparison of the formant values of pre-palatal /i:/ and mid-palatal /e:/ (see Figure 5.9) shows that F2 and F4 are approximately equal with both qualities, while F1 is lower and F3 is higher for /i:/ than for /e:/. The high F3 value for /i:/ and the resulting proximity of F3 and F4 are characteristic for a language that uses a pre-palatal constriction location[44]. However, these acoustic cues may not function distinctively for listeners of a language that does not differentiate between mid-palatal and pre-palatal vowels. Moreover, German /e:/ is very constricted as compared to the /e/-phonemes in many other languages (e.g. Polish, SerBoCroatian, Albanian) resembling more an open [ɛ] (similar to German /ɛ/) than a more constricted [e]-quality. Difficulties in the differentiation of German /i:/ and /e:/ can therefore be expected in L2 perception.

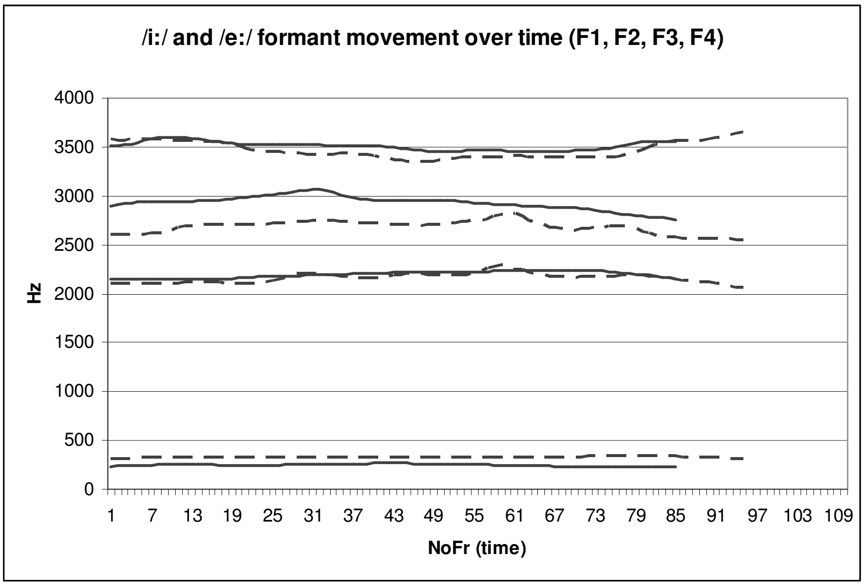

Figure 5.9: Comparison of /i:/ (full line) and /e:/ (dashed line) in the /pVt/-context

As has been mentioned above, German /ɛ:/ is an exception to the dichotomy of German vowels and does not occur in all regional varieties of German. However, it has been included in the present study. A comparison of formant values for /ɛ:/ of the input material (in the /pVt/-context) shows that it is more similar in quality to short /ɛ/ than to long /e:/ (see Figure 5.7 and Figure 5.8).

Previous perceptual experiments (e.g. Sendlmeier 1981; Hentschel 1986; Kerschhofer 1995) show that the distinction of /e:/ and /ɪ/ is difficult in perception for L2 learners of German.

---

[44] This effect is less conspicuous for /y:/ vs. /ø:/ due to lip-rounding (see Figure 5.11 and Figure 5.12).

For SAG, Moosmüller (2007: 102f) observes a clear acoustic differentiation of these qualities by at least one formant (F1 and/or F2) in spontaneous speech and even a differentiation by two or three formants in unstressed position. Interestingly, these differences are stronger in unstressed than in stressed positions (2007: 102f). However, in the present study, the acoustic values of the /e:/ and /ɪ/ input stimuli show considerable similarity in all formants as well as in duration even in the formal style of logatome reading (see Figure 5.10) even if in less formal styles the differentiation of /e:/ vs. /ɪ/ in terms of duration may be less conspicuous than in the material presented here. Therefore, perceptual difficulties in the differentiation of the contrast /e:/ vs. /ɪ/ are expected especially in spontaneous speech and with L2 learners who do not use vowel length for phonemic differentiation.

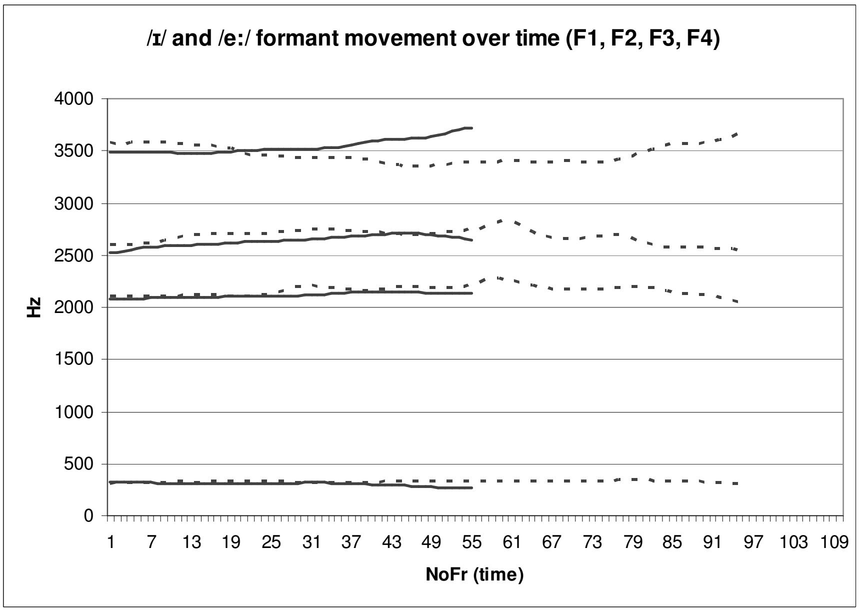

Figure 5.10: Comparison of /e:/ (dotted line) and /ɪ/ (full line) in the /pVt/-context

To summarize, five constriction locations are assumed for German, a pharyngeal location for ɑ-vowels, an uvular region for o-vowels /ɔ o:/, a velar region for u-vowels /ʊ u:/, and a palatal region that is divided in a pre-palatal region for i-vowels /ɪ i:/ and ü-vowels /ʏ y:/ and a mid-palatal region for e-vowels /ɛ e: ɛ:/ and ö-vowels /œ ø:/.

These primary distinctions manifest themselves in characteristic patterns in the formant frequencies.

For a language with more than 4 or 5 vowel qualities, additional parameters are used for further differentiation within the system. These are described in terms of additional articulatory-acoustic parameters in the following sections.

5.4.2 Rounding

Lip activity is an additional parameter for defining phonemic contrasts in German. Lip rounding has traditionally been described as a third dimension of the traditional vowel space being independent of high/low and front/back contrasts. Valaczkai (1998: 130ff) describes three parameters of lip activity: (1) lip opening, (2) lip spreading, and (3) lip protrusion (for the upper and the lower lips separately). While the velar and the uvular constriction location are most commonly accompanied by a labial constriction in many languages, this is less frequently the case for the palatal and the pharyngeal location. In German, however, palatal vowels may occur with or without additional lip constriction.

Lip rounding or lip protrusion causes a lengthening of the vocal tract up to 3 cm in total and has a lowering effect on all formants. Lip rounding or lip spreading by itself does not create a contrast but functions as a mean to increase or adjust the defining acoustic attribute of constriction location. As Stevens & Keyser (2010) emphasize, lip rounding or spreading is not the defining gesture for a distinctive contrast in a given language but rather an enhancing gesture. Lip rounding in non-low back vowels apparently serves to enhance the contrast to non-low front vowels which are often produced with lip spreading, thus strengthening the contrast to rounded back vowels.

For front rounded vowels, we generally expect lower formant frequencies as compared to their non-rounded cognates. The effect of lip rounding for non-low front vowels such as /i/ or /e/ is to narrow the anterior end of the front cavity and to lengthen this part of the vocal tract. The narrowing of the cross-sectional area and the increase in length has consequences on F2, F3, and F4. The acoustic result is a lowering of the frequencies of the front cavity resonances and a decrease of the bandwidth of F3 by narrowing lip aperture (Stevens 1998: 291f). Lower values for F2 and F3 are observed for mid-palatal ö-qualities as well as for pre-palatal ü-qualities (see Figure 5.11 and Figure 5.12). However, for a pre-palatal constriction at more than 12 cm from the glottis (of 16 cm), a decrease of F2 and an increase of F3 are observed (Stevens 1998: 281).

Stevens (1998: 294) describes a general tendency for a major broad spectral peak formed by F1 and F2 for back vowels and by F2 and F3 or F3 and F4 for front vowels that is more prominent for rounded than for unrounded qualities. Stevens considers this increased prominence of the spectral peak together with a lowered center of gravity of the peak as primary acoustic correlates of the rounding feature.

For rounded vowels, the larynx is generally lower than for vowels with spread lips. The larynx is moreover more depressed for tense/constricted vowels than for lax/unconstricted vowels (Wood 1979: 34). This larger laryngeal depression contributes to the lowering of F2 with rounded vowels, an effect that is stronger at the palatal constriction. However, as Hoole & Kroos (1998) show, there are considerable inter-category (front vs. back rounded vowels) and inter-speaker differences in lower larynx position while coarticulatory effects are relatively weak.

Generally, the degree of lip aperture in front rounded vowels is greater with unconstricted vowels than with constricted qualities. For /ʏ/-/y:/ and /ʊ/-/u:/, Valaczkai (1998: 131f) describes stronger lip protrusion for short unconstricted vowels, whereas with /œ/-/ø:/ and /ɔ/-/o:/ lip protrusion is stronger with the long more constricted vowels.

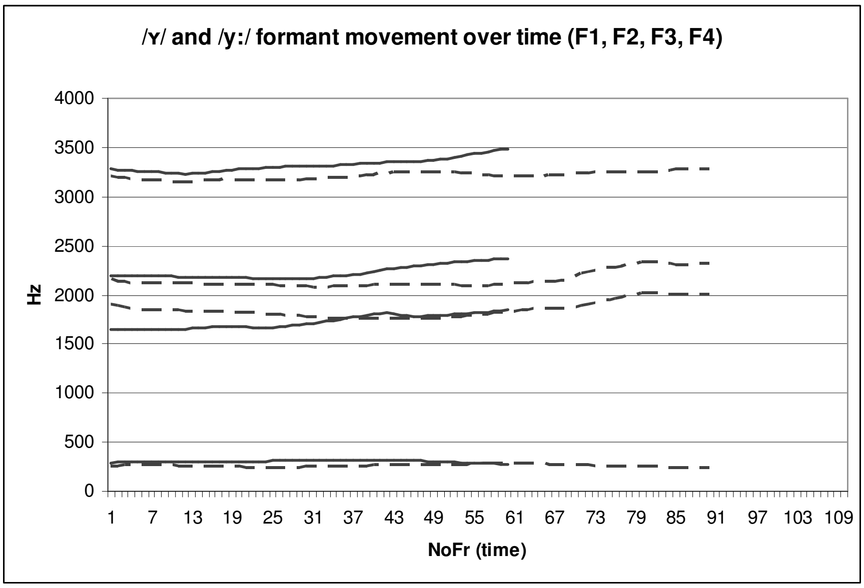

Figure 5.11: /ʏ/ (full line) and /y:/ (dashed line) formant movement over time in the /pVt/-context. Mean formant values (F1, F2, F3, F4) and standard deviation (SD) in Hz, duration (in number of frames)[45]

For /ʏ/, Moosmüller (2007: 93f) observes a higher F1, a lower F2 and a similar or higher F3 compared to the constricted counterpart /y:/. Spectral peaks are therefore better kept apart for /ʏ/ than for /y:/ (see Figure 5.11).

[45] /ʏ/ F1 304 Hz (12 SD), F2 1726 Hz (70 SD), F3 2229 Hz (67 SD), F4 3321 Hz (67 SD), 93 ms /y:/ F1 263 Hz (12 SD), F2 1851 Hz (82 SD), F3 2154 Hz (82 SD), F4 3216 Hz (39 SD), 146 ms

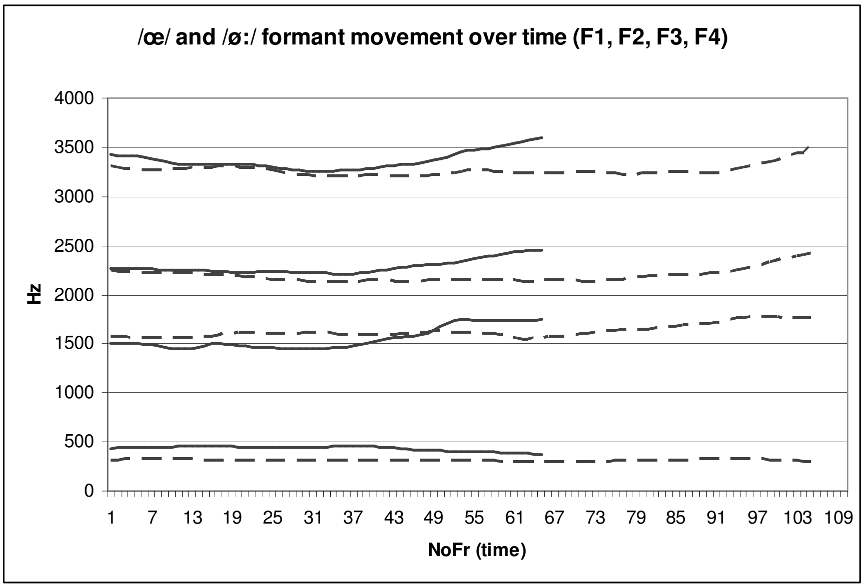

Figure 5.12: /œ/ (full line) and /ø:/ (dashed line) formant movement over time in the /pVt/-context. Mean formant values (F1, F2, F3, F4) and standard deviation (SD) in Hz, duration (in number of frames)[46]

For mid-palatal /ø:/, a moderate lip protrusion with compensatory larynx lowering is described causing a F2-F3 proximity. Unconstricted /œ/ is produced with a higher degree of lip opening than /ø:/ but with comparable lip protrusion (Valaczkai 1998). A higher F1, a lower F2 and a similar or equal F3 than for /ø:/ are therefore characteristic for /œ/ (see Figure 5.12).

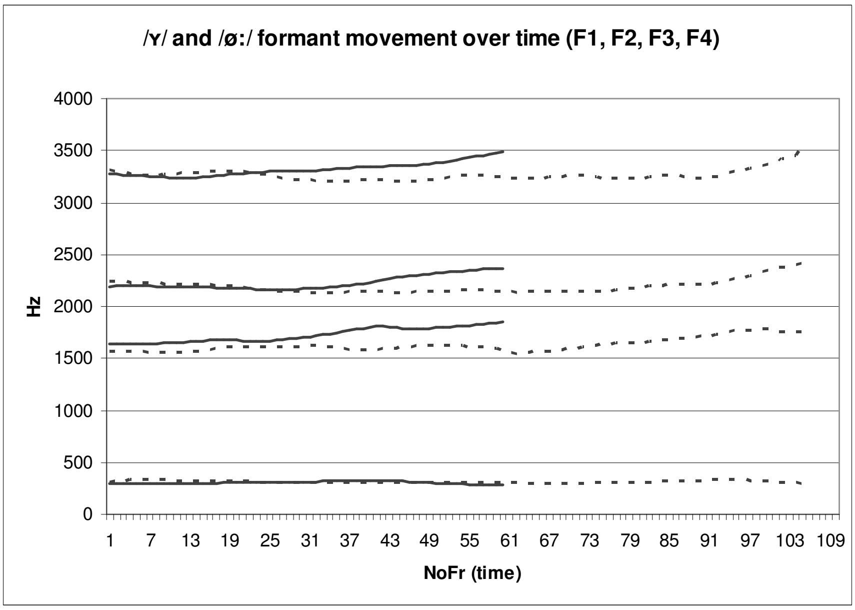

Figure 5.13: Comparison of /ʏ/ (full line) and /ø:/ (dotted line) in the /pVt/-context

[46] /œ/: F1 404 Hz (24 SD), F2 1626 Hz (111 SD), F3 2360 Hz (70 SD), F4 3512 Hz (93 SD), 118 ms /ø:/: F1 315 Hz (9 SD), F2 1629 Hz (64 SD), F3 2192 Hz (64 SD), F4 3265 Hz (53 SD), 194 ms

The comparison of /ʏ/ and /øː/ in Figure 5.13 shows a certain similarity of all formant values, especially of F1 (and F3), that might cause problems in discrimination if listeners do not use durational information as distinctive cue.

In general, we may say that the palatal region is quite crowded in German, since eight to nine phonemes are distinguished in this region. The acoustic cues that indicate phoneme boundaries in this region are language-specific and may therefore cause problems in L2 perception. Lip rounding of front vowels is an important device to guarantee additional contrasts. Other cues to these distinctions are given by degree of constriction, jaw opening and duration[47]. However, typologically, front rounded vowels are less common than front unrounded vowels. We therefore expect that the exact distinction of these vowel phonemes will cause difficulties especially to L2 learners of German who have a smaller inventory of palatal vowels and have no rounded front vowels in their native language. The differentiation of front rounded vs. non-rounded vowels and the perceptual interpretation of additional cues to German palatal oppositions may therefore be at risk.

5.4.3 Constriction degree

Degree of constriction ([+/– constricted]) can be viewed as equivalent to *tongue height* but has to be discerned from the degree of jaw opening (see section 5.4.4; Wood 1982a, b; Fischer-Jørgensen 1990; Valaczkai 1998; Moosmüller 2007). For front vowels, constriction degree is equivalent to tongue-palate distance. Variation in constriction degree renders monotonic changes in formant frequency values (Carré 2004). The relative *length of constriction* may cause further variation in the formant frequencies but has to be distinguished from the degree of constriction. For /i y e/, a strong constriction degree is observed (Ac < 1 cm^2) which is realized fairly consistently, while /ɛ ø œ/ show a large variation in constriction degree from 0.2 cm^2 to 4.0 cm^2 (Boë, Perrier & Bailly 1992). Concerning front vowels, the tongue-palate distance is bigger for front rounded vowels than for their unrounded counterparts, leading to four different tongue-palate distances in the two front regions (Wood 1986; Pouplier et al. 2004; see discussion in Moosmüller 2007: 83f).

However, the relative degree of constriction cannot account for all vocalic distinctions, especially in the front region. A mere distinction in terms of different degrees of tongue

[47] In previous studies (Halle & Stevens 1969; Wood 1975; Ladefoged & Maddieson 1996), a correlation of tongue height and *advanced tongue root (ATR)* was described for Germanic languages (cf. Wiese 1996: 20f for "Modern Standard German"). However, ATR as distinctive feature can be excluded for German as opposed to languages like Igbo or Akan, where ATR and tongue height vary independently of each other (Pouplier et al. 2004; Moosmüller 2007).

height, i.e. tongue palate distance, as distinctive feature is problematic since these distinctions cannot be sustained in weak prosodic positions and less formal speaking styles. Phonemic contrasts would therefore be at risk. As Moosmüller (2007: 83) argues, the /i/ vs. /e/ contrast is therefore not to be described in terms of a relative degree of constriction but by a further differentiation of the area of constriction as described above, i.e. by the differentiation of pre-palatal vs. mid-palatal vowels, together with a difference in degree of aperture.

5.4.4 Jaw and lip aperture

Jaw opening ([+/–open]) is another feature to differentiate vowels at a given constriction area. By mandibular movements, the magnitude of the mouth opening and the position of the tongue relative to the palate and the pharynx is varied (Wood 1990: 201). The relative aperture of the inner lips varies proportionally with the height of the mandible (Valaczkai 1998: 133).

Wood (1975) presented an early schematic analysis of the relative involvement of tongue and jaw in the height/tenseness opposition. While jaw opening (together with constriction location) refers to the distinction between /i/- and /e/-vowels, the "tense/lax" opposition is rather due to tongue position while the jaw is kept in a constant position, i.e. /i/ and /e/ have a higher tongue position than /ɪ/ and /ɛ/. [+open] refers to the lower mandible position (8-12 mm) of [e, ɛ, o, ɔ, a]-like vowels, whereas [-open] refers to the higher position (5-8 mm) of close vowels [i, ɪ, u, ʊ] (Wood 1990: 201). We can thus define the *degree of constriction* as a *lingual* feature (referring to "tongue height") and the *degree of jaw aperture* as *mandibular* feature.

For German, Valaczkai (1998: 133f) observes a higher value of vertical mandibular movement as well as of lip aperture for [ɛ:] and [e:] compared to [i:] (see also Wood 1982a: 46).

Theoretically, jaw opening in combination with constriction degree may differentiate four different German vowels: /i:/ [+constricted, -open jaw], /ɪ/ [-constricted, -open jaw], /e:/ [+constricted, +open jaw], /ɛ/ [-constricted, +open jaw]. However, these observations have to be further refined by the effect of lip rounding and the influence of consonantal coarticulatory effects as described for front vowels by Hoole & Mooshammer (2002: 7ff). Comparing jaw height and tongue height of German front vowels in different contexts in an articulatory EMMA study, Hoole & Mooshammer (2002) observed systematic differences in degree of jaw height: /i: y:/ < /e: ø:/ < /ɛ œ/. The values for /ɪ/ and /ʏ/ showed most variation with respect to consonantal context, being however rather close to /e:/ and /ø:/ values. Therefore, the simple scheme proposed by Wood has to be differentiated with respect to the influence of rounding and coarticulatory effects. Since the values for jaw height for /e:/ and /ø:/ are between those of /i:, y:/ and /ɛ, œ/ in Hoole & Mooshammer's data, we may conclude that the differentiation of e- and ö- vs. i- and ü-vowels by [+/–open jaw] is not so clear-cut as proposed by Wood.

Moreover, this pattern cannot be directly transferred to the back rounded vowels due to the effects of lip rounding. To differentiate back vowels, lip activity and the relative degree of lip aperture rather than jaw opening seem appropriate to account for acoustic differences. In constricted back vowels, the lips are more protruded than in their unconstricted counterparts. Therefore, we have to differentiate /u:/ and /o:/ with a high degree of lip rounding vs. /ʊ/ and /ɔ/ with less lip rounding and a higher degree of aperture (see Valaczkai 1998: 130). The lip configuration for /u:/ and /o:/ is in fact very similar, while it differs substantially between /o:/ and /ɔ/ and to a smaller degree between /u:/ and /ʊ/ (Valaczkai 1998).

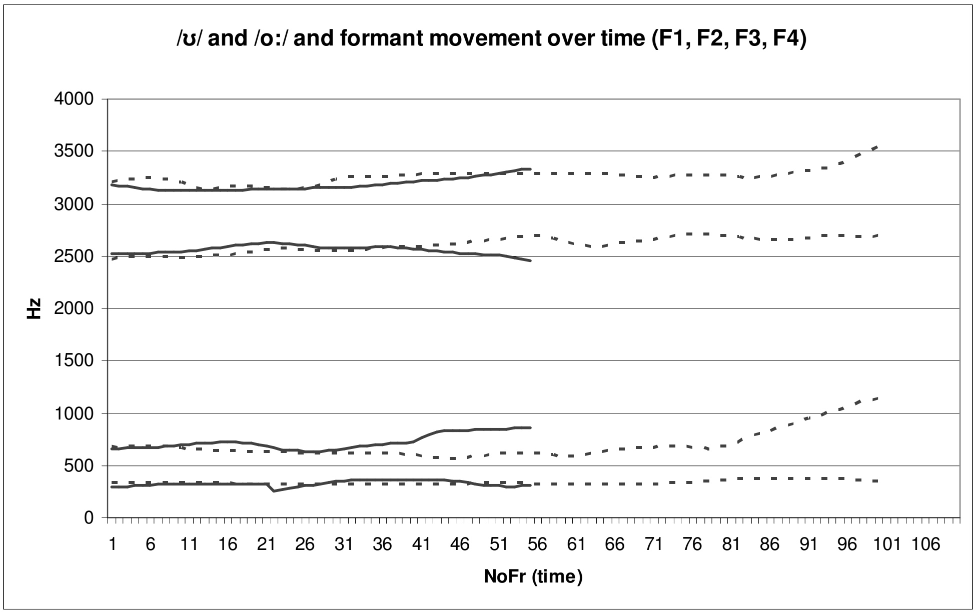

Figure 5.14: Comparison of /ʊ/ (full line) and /o:/ (dotted line) in the /pVt/-context

While the distinction of /u:/ as [+constricted, –open jaw, –open lips] and of /ɔ/ as [–constricted, +open jaw, +open lips] is evident, the description of /ʊ/ and /o:/ is less distinct with respect to jaw aperture and lip aperture. Even if /ʊ/ and /o:/ are differentiated by the articulatory features [+/–open] and [+/–constricted] in the feature matrix (see Table 5.9), their perceptual differentiation may be at risk due to the similar acoustic output of the articulatory constellations, if duration is not considered (see Figure 5.14).

	u:	ʊ	o:	ɔ
jaw open[48]	-	-	+	+
lips open	-	-	-	+
constricted	+	-	+	-
round	+	+	+	+

Table 5.5: Feature constellations for back rounded vowels

To summarize, jaw height is assumed to be a relevant feature to differentiate German vowels. However, describing the acoustic consequences of specific constellations of the vocal tract, jaw height has to be considered together with the effects of lip rounding and degree of lip aperture. While for front vowels the position of the jaw is higher for /i, ɪ/ than for /e, ɛ/, for back and front rounded vowels, the explanatory value of jaw opening is more restricted. For back rounded vowels, the degree of lip rounding and the relative degree of mouth aperture are more descriptive to account for the acoustic output. The status of constricted mid-palatal front vowels with respect to the feature [open] is difficult to define. Considering the available data, /e:/, /o:/ and /ø:/ will be preliminarily defined here as [-open], while /ɛ/, /ɔ/ and /œ/ are defined as [+open] due to a lower degree of jaw aperture together with less lip rounding compared to their constricted counterparts (see the feature matrix in Table 5.9).

5.4.5 Prosodic characteristics

5.4.5.1 Duration

Vowel duration is determined by the identity of a given vowel itself (intrinsic duration, see section 5.4.5.3) as well as by several contextual factors such as the identity of adjacent segments (voiceless vs. voiced consonants, stops vs. fricatives), the syllabic structure of a word (stressed vowels in bisyllabic words vs. monosyllabics), syllabic stress, pitch accent and pitch contour or the proximity to a syntactic boundary (Klatt 1976; van Santen 1992). As discussed in section 5.3, in most phonological accounts of German, vowel duration ([+/– long]) is considered as a phonetic cue contributing to the so-called “tense/lax“-opposition, though the relative relevance or weight of spectral and temporal cues is not satisfactorily defined. Most German vowel oppositions that are conventionally described as quantity oppositions are accompanied by significant spectral differences and the question whether durational or spectral differences provide the cues that German listeners use to distinguish

48 Valaczkai (1998: 134) observes a degree of jaw aperture of 107% for [u:] and 108% for [o:]. No values for [ɔ] and [ʊ] are given.

between members of a contrast pair is unanimously discussed (Becker 1998; Ramers 1988). Moreover, the distribution of these cues as well as their perceptual use is subject to regional variation. As Moosmüller (2007) argues, duration does not function as a distinctive feature in Standard Austrian German (SAG). In SAG, Moosmüller (2007) observes no intrinsic durational differences between high and low vowels and considers duration not to be the primary relevant feature for the tense/lax opposition, since the durational dimension of vowel contrasts is not maintained in all task types and speaking styles (sentence reading, word list, spontaneous speech), while quality differences are preserved. With decreasing formality of style and decreasing prosodic strength, length oppositions are described to be neutralized in SAG (Moosmüller 2007).

In the input stimuli of the present study, measurements of the mean duration of vowels in the test items show that length distinctions are pertained at least in the formal style of logatome reading (see Table 5.6 and Figure 5.15).

duration (sec)	*mean*	*SD*	*maximum*	*minimum*
ɑ	.104	.023	.147	.059
ɑ:	.197	.025	.243	.154
ɛ	.100	.018	.134	.073
e:	.175	.026	.210	.125
ɛ:	.192	.021	.228	.156
ɪ	.074	.022	.114	.040
i:	.150	.023	.191	.114
ɔ	.108	.016	.137	.078
o:	.186	.020	.217	.146
œ	.109	.018	.139	.083
ø:	.184	.019	.229	.159
ʊ	.087	.019	.124	.054
u:	.157	.027	.192	.114
ʏ	.083	.027	.132	.045
y:	.151	.024	.196	.115
total	*.137*	*.047*	*.243*	*.040*

Table 5.6: Mean duration values, standard deviation, minimum, and maximum (in sec) of the 270 test stimuli (each vowel occurring 18 times in changing context)

Figure 5.15 presents the mean duration values for all vowel categories in descending order based on measurements of all test items that were used in the present study. In the test items, vowels in logatomes were embedded in a constant carrier sentence. The vowels occur only in stressed syllable position and in the sentence focus (for a detailed description, see section 6.3).

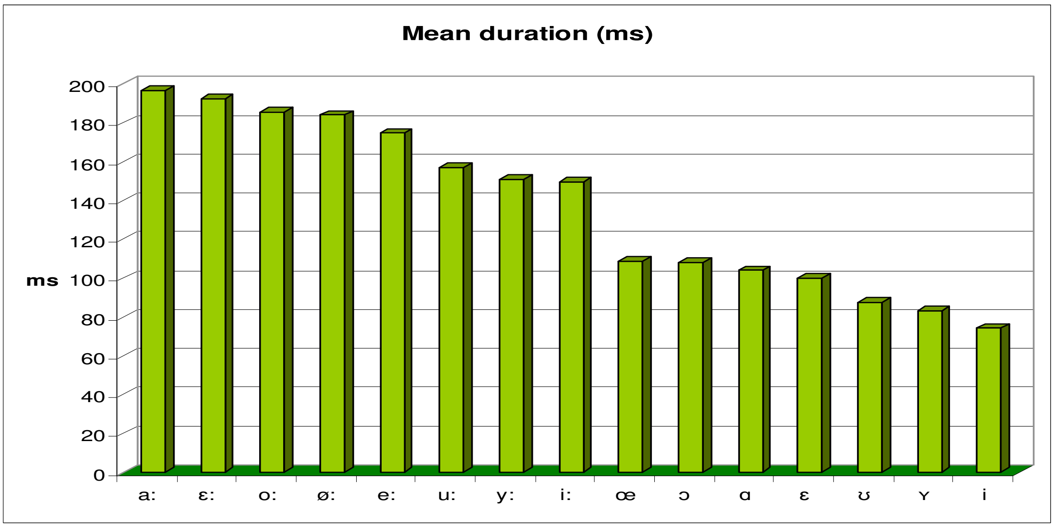

Figure 5.15: Mean duration values (in ms, in descending order) for all vowel items in all contexts (input stimuli of the present study)

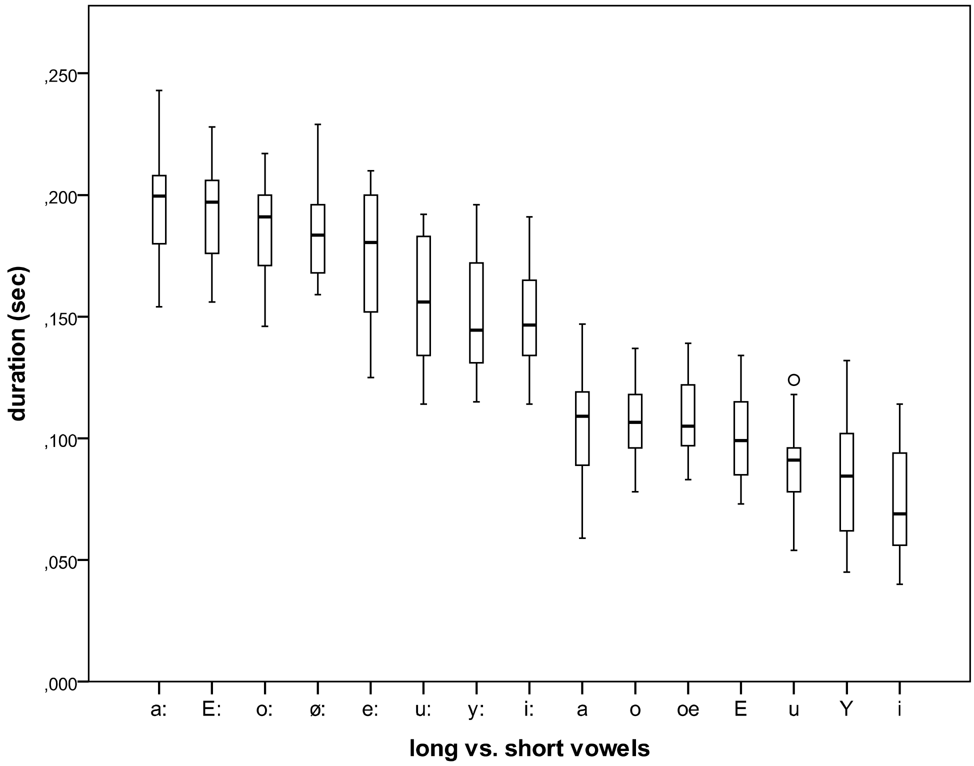

Figure 5.16: Duration (in sec) for German long and short vowels in the input stimuli of the present study (each vowel occurring in 18 logatomes embedded in sentences, vowels in changing pre-vocalic and post-vocalic context)

Figure 5.16 presents the boxplots of all vowels from the test stimuli. The items are read in the very formal speaking style of logatome reading and clearly differ in terms of duration/length. Moreover, we observe that in the data presented here the relative length of different vowel

qualities differs considerably within and among categories. Table 4.5 presents the mean duration values, minima, maxima, and standard deviation for each vowel category separately.
The data confirm the universal variation of different vowel categories with respect to intrinsic duration that has been discussed in section 5.4.3. Within the class of long vowels, duration varies together with degree of jaw aperture, i.e. open vowels are intrinsically longer than closed vowels. For the short unconstricted qualities, this pattern is less clearly given (see variation for the categories /ɑ ɔ œ/ in the boxplots in Figure 5.16 vs. mean values in Figure 5.15).

Many languages do not differentiate between long and short vowel phonemes at all, i.e. duration has no phonemic function in these languages. Therefore, even if there is a clear distinction of the duration values for long vs. short vowels of the same quality, we have to consider that in L2 vowel perception phonemic length distinctions might not be recognized as easily by L2 learners of German. Even if Bohn’s (1995) assumption that duration cues are easily accessible in vowel perception even for listeners’ whose L1 does not use length distinctions may be basically correct, we have to consider the substantial variation across speech styles and the fact that in less formal styles and in unstressed positions durational contrasts are frequently not preserved in German. We therefore have to expect that the perceptual distinction of German long vs. short vowels will pose problems in L2 perception that will, however, not be clearly identifiable as problems in detecting length differences, since German non-low vowels differ not only in quantity but also in quality, i.e. in degree of constriction.

5.4.5.2 Stress

A characteristic feature of German is the existence of stressed vs. unstressed position in a word and of primary vs. secondary word stress in words with more than two syllables. For German, commonly 14 to 15 vowel phonemes are described in stressed syllables, while in unstressed syllables only the so-called “reduced” qualities /ə/ and /ɐ/ are considered to occur (e.g. Jørgensen 1969; Kohler 1999). In Standard Austrian German, the /ə/-schwa (<e>) is generally realized as [ɛ], e.g. in [ˈmitɛ] “middle” in the final syllable, or as [e] in pre-tonic ‘be-‘ [be] or ‘ge-‘ [ge]). The /ɐ/-schwa corresponds to <–er> and is realized as [ɐ] in word-final syllables or in prefixes like ‘er-’ [ɐ], ‘ver-‘ [fɐ-] or ‘zer-‘ [tsɐ-].

The relative prosodic strength of vowels in stressed vs. unstressed syllables can result in differences in vowel quality and quantity. Phonetic correlates are differences in duration, intensity, fundamental frequency and changes of formant frequencies.

Vowels in secondary stress position are generally considered to differ from vowels in primary stress position as well as from vowels in unstressed position. This is especially the case in non-primary stress positions in German compounds, e.g. [ˈne:bɛnˌfa·rba·n] ‘Nebenfahrbahn’. While vowels in secondary stress position are considered to be reduced in quantity, their quality appears to be unchanged. A second group of vowels in non-primary stress position is given in words of foreign origin, e.g. /i/ in [siˈgna:l], /e/ in [ˌmediˈtsi:n] ‘Medizin’, /a/ in [fanˈto:m] ‘Phantom’, /o/ in [loˈka:l] ‘Lokal’, /u/ in [muˈsi:k] ‘Musik’, /y/ in [tyˈran] ‘Tyrann’, or /ø/ in [ˌøkonoˈmi:] ‘Ökonomie’.

For Standard Austrian German, Moosmüller (2007: 198, 204ff) describes quantitative as well as qualitative differences between vowels in stressed vs. unstressed syllables: In *quantity*, vowels in unstressed position are significantly shorter than in stressed positions. The *quality* of vowel phonemes is strongly influenced by its position in a word and by the prosodic strength of this position. In Moosmüller’s data, statistically significant qualitative differences between vowels in stressed vs. unstressed positions occur in spontaneous speech as well as in a sentence reading task. The strongest difference is observed for F2 for almost all vowel categories, the weakest difference is given for F3 (Moosmüller 2007: 205). This effect becomes most evident with o- and u-qualities. In cases where statistically significant differences of F2 occur between vowels in stressed vs. unstressed positions, F2 is lowered for front vowels and raised for back vowels. This effect is due to a reduction of constriction degree, constriction length and lip protrusion (for back rounded vowels). In general, the variability of F2 is considerably lower in unstressed than in stressed syllables (Moosmüller 2007: 208). For /ɑ/, however, F1 is the primary differentiating factor of stressed vs. unstressed vowels (indicating a higher degree of aperture in stressed than in unstressed positions). For other vowel categories, (if statistically significant differences can be observed), F1 is higher in unstressed positions (Moosmüller 2007: 206).

In positions bearing secondary stress[49], qualitative changes are observed as well that could result from differences in the degree of lip opening, degree of constriction, length of constriction and degree of lip protrusion. However, as Moosmüller (2007: 208-218) observed, the means by which vowels in secondary stress position are realized and distinguished from primary stress positions vary from speaker to speaker.

---

[49] Moosmüller analyzed vowels in primary stress position in content words and investigated the effects of secondary stress in compounds and morphologically complex words, but not in words of foreign origin.

It is important to stress here that even if there are qualitative and quantitative differences observed in stressed vs. unstressed position, these differences cannot be assumed to result in changes of the respective vowel category. The phonemic identity is preserved even in weak prosodic positions.

In the present investigation, only vowels in stressed position are considered. The test items in the listeners' input consist of one- and two-syllabic logatomes that are embedded in a constant carrier sentence and are always situated in a syntactically prominent position. The listeners' responses only refer to the stressed vowels in these logatomes.

5.4.5.3 Intrinsic pitch and intrinsic length

As has been discussed in section 4.3.5 and 4.3.6, different vowel qualities tend to show differences in fundamental frequency, a phenomenon that is referred to as *intrinsic pitch* (IF0). This phenomenon has been considered to be language-universal and related to vowel height (usually interpreted as tongue height). Another language-universal tendency is a correlation between the *intrinsic duration* of vowels and their height (interpreted as tongue height) that is independent of other factors such as coarticulation, i.e. "higher" vowels show a higher intrinsic duration than "low" vowels (Lehiste 1970; Neweklowsky 1975; Maddieson 1984; Fischer-Jørgensen 1990, Laver 1994). A strong negative correlation is moreover observed between tongue height and *F1*. These effects have been observed for German as well. However, a description in terms of degree of aperture rather than "tongue height" seems to be more appropriate. For German, Fischer-Jørgensen (1990) found a better agreement of F0 with jaw position than with tongue height. These results are confirmed by Mooshammer et al. (2001) in an EMMA study. However, the differentiation of long tense/constricted vs. short lax/unconstricted vowels in German makes direct conclusions on the relation between F0, F1 and duration more complicated. Fischer-Jørgensen's (1990: 108f) review of previous studies on German shows that in all studies reviewed F0 was higher for [ɪ, ʏ, ʊ] than for [i, y, u] (one exception) and that F1 was higher for [ɪ, ʏ, ʊ] than for [i, y, u] and very close to [e, ø, o], often even the same or slightly higher. Considering acoustic and articulatory data, Fischer-Jørgensen (1990) and Mooshammer et al. (2001) observed similar F0 values for tense-lax pairs, where the jaw rather than the tongue correlated with F0, because differences in jaw position for tense-lax pairs were smaller than differences in tongue position. However, Mooshammer et al. (2001) found several cases where the lax members of a pair had a slightly higher F0 despite a somewhat lower jaw position[50].

Stimuli items from the present study (only experiment 2) have been analysed with respect to F0, F1 and intrinsic duration[51]. The bisyllabic items in experiment 2 have the structure /bVtɛ/ in phrase-initial position ("bVtɛ bedeutet auch nichts") read by a male speaker. Each vowel occurred in 10 different pre-vocalic contexts (/b p d t g k v f z j/).

The question at which point fundamental frequency should be measured has been treated very differently in studies on this topic (for discussion, see Fischer-Jørgensen 1990: 121). In the present study, F0 has been measured at the beginning and the end of the vowel together with an average (in Hz) and the distance between beginning and end point (in Hz or in semitones). For the comparisons and correlation analyses presented here, mean F0 values of single input items were correlated with other phonetic parameters as duration and F1.

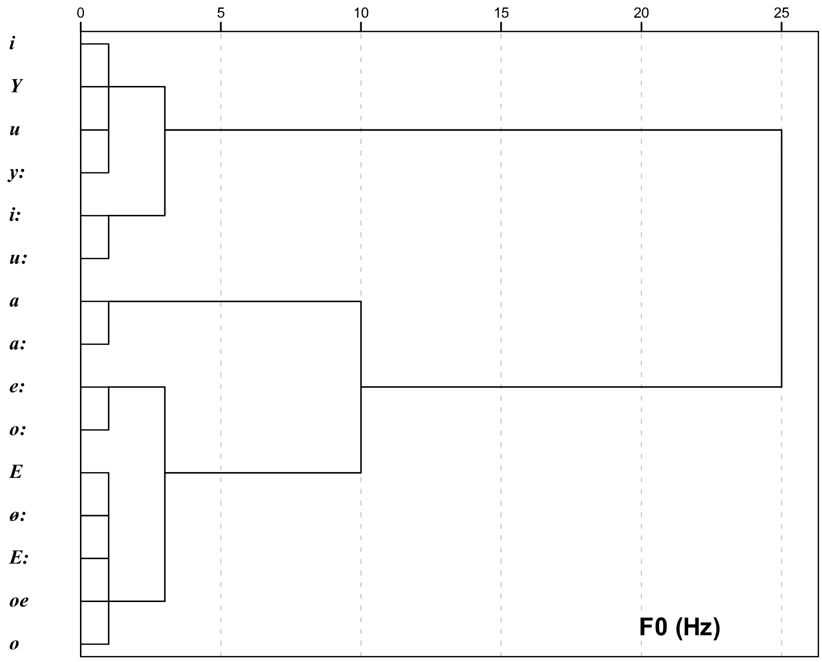

Figure 5.17: Hierarchical Clustering based on mean F0 values (in Hz) of items in experiment 2

The dendrogram in Figure 5.17 represents a hierarchical clustering (for details about this method, see section 5.7.2) of mean F0 values for vowel categories in experiment 2 and gives a first impression about the complex relation of F0 and vowel height. The dendrogram shows that i-, u- and y-qualities that are traditionally referred to as "high" vowels can be clearly

50 In the present study, a somewhat higher F0 is only observed for unconstricted /œ/ as compared to /ø:/ (see Figure 5.18).

51 In test 1, the logatome containing the vowel under investigation was situated in phrase-final position and varied according to sentence prosody (falling vs. rising tone).

differentiated from non-high vowels. The cluster of the [-open] vowels branches into two sub-clusters, one sub-cluster for long constricted /i:/ and /u:/ and the other one for constricted /y:/ and unconstricted /ɪ ʏ ʊ/. The other sub-cluster further branches into one sub-cluster for a-qualities and one for so-called "mid" e-, o- and ö-vowels.

short (N=70)	F0 (Hz)	duration (ms)	long (N=69)	F0 (Hz)	duration (ms)	ratio short/long	ratio long/short
ɑ	159	90	ɑ:	156	180	50	2
			ɛ:	173	180	50	---
ɛ	172	90	e:	179	160	56	1.8
ɪ	193	60	i:	202	140	43	2.3
œ	175	100	ø:	171	170	59	1.7
ʏ	193	60	y:	196	140	43	2.3
ɔ	168	100	o:	181	170	59	1.7
ʊ	198	80	u:	204	140	57	1.8
mean	180	83		183	160	52	1.9

Table 5.7: Mean duration (in ms) and mean F0 (in Hz) for vowels in open syllables in changing pre-vocalic context /CVtɛ/, 810 occurrences for each vowel category

A comparison of mean F0[52], mean F1 and duration of stimuli of the present study as presented Table 5.7 reveals a clear relationship between rising F0 and falling F1. In very general terms, an increase of F0 seems to go along with a decrease of F1. The Pearson correlation coefficient indicates a correlation of F0 and F1 that is significant at the .01 level (1-tailed) with a medium effect for short vowels (r = .345) and a considerably strong effect for long vowels[53] (r = .748). Moreover, if short and long vowels are considered separately, a significant correlation of *F1 and duration* is observed at the .01 level (1-tailed) with Pearson's correlation coefficient showing a medium to strong effect (r = .407) for short vowels and a strong effect (r = .507) for long vowels.

A correlation of *F0 and duration* can therefore be expected, if short unconstricted and long constricted qualities are analyzed separately. The mean duration of short unconstricted qualities is 84 ms, for long constricted vowels it is 160 ms (see Table 5.7).

A correlation of F0 and duration is significant at the .01 level (1-tailed) with a medium effect (r = .297) for short unconstricted vowels and a strong effect (r = .612) for long constricted vowels.

---

52 The relatively high F0 of the male speaker is probably due to the prominent prosodic position of the logatom in phrase-initial position and the formal situation in a speech lab.

53 69 long constricted and 70 short unconstricted vowel items were analyzed. For reasons of comparability with previous studies (for a comparison see Fischer-Jørgensen (1990: 110ff), /ɛ:/ was excluded from the correlation procedures but is included in the graphical representations.

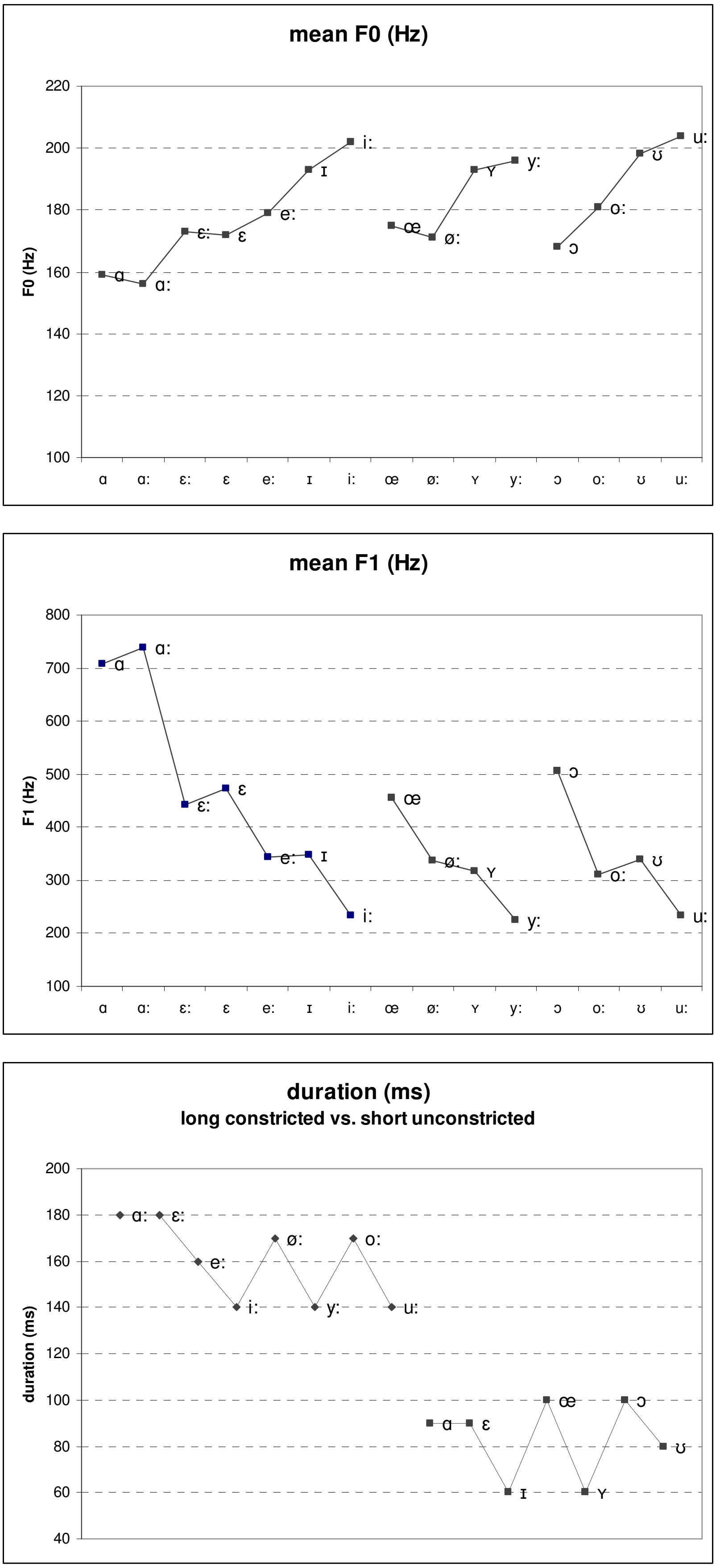

Figure 5.18: Comparison of German vowel categories in terms of mean F0, mean F1 and mean duration.

Studies on the tense/lax contrast in German describe a similar acoustic F0 for both members of these pairs although they are claimed to differ in terms of so-called articulatory "height". Fischer-Jørgensen (1990) discusses the fact that a direct correlation of intrinsic F0 (IF0) and vowel height interpreted as tongue height does not hold for German unless short lax vowels and long tense vowels are considered separately; a better agreement is found between IF0 and jaw position. Likely, Pape's (2009) acoustic and articulatory investigation of German tense vs. lax vowels showed no significant difference in intrinsic pitch comparing these vowel pairs. As Pape (2009) states, these different vowel classes show a similar acoustic F0, although their articulatory "height" is assumed to be different. While articulatory tongue height differs significantly for the tense/lax pairs in Pape's (2009) investigation, no significant F0 differences are found comparing the tense vs. lax vowel pairs.

For the vowel pairs that are traditionally described as "tense vs. lax" but are here denoted as [+/–constricted] and [+/–long], Fischer-Jørgensen's (1990) comparison of previous studies shows a higher F0 value for high unconstricted [ɪ ʏ ʊ] than for [i: y: u:]. Average F1 values for [ɪ ʏ ʊ] are usually higher than for [i: y: u:] and close to [e: ø: o:] and sometimes even higher, a fact that is easily explained by the higher degree of aperture of unconstricted qualities. While F1 values are definitely higher for [ɛ œ ɔ] than for [e: ø: o:], F0 averages are very similar, i.e. the same or slightly lower than F0 for [ɛ œ ɔ].

However, the present data shows that [ɪ ʏ ʊ] have a lower mean F0 but a higher mean F1 than their long constricted counterpart [i: y: u:]. For [ɛ œ ɔ] and [e: ø: o:] the situation is different: F0 values for [ɛ ɔ] are lower than for their constricted counterpart [e: o:], whereas F0 for [œ] is slightly higher than for [ø:].

Average F1 values from stimuli in the present study are considerably higher for [ɛ œ ɔ] than for [e: ø: o:] as well as higher for [ɪ ʏ ʊ] than for [i: y: u:] due to the higher degree of aperture for the unconstricted qualities compared to the constricted ones. A lower F1 for [o:] than for [ʊ] and a slightly lower F1 for [e:] than for [ɪ] is given, whereas F1 for [ø:] is slightly higher than for [œ].

A t-test for equality of means of the so-called tense/lax pairs showed significant differences in F0 for none of the pairs: /i:/-/ɪ/ (p = .185), /y:/-/ʏ/ (p = .624), /e:/-/ɛ/ (p = .148), /ø:/-/œ/ (p = .407), /u:/-/ʊ/ (p = .383), /o:/-/ɔ/ (p = .089), /ɑ:/-/ɑ/ (p = .646). Levene's test for equality of variances showed that equal variances can be assumed in all cases with the only exception of /u:/-/ʊ/ (p = .092 Levene's test).

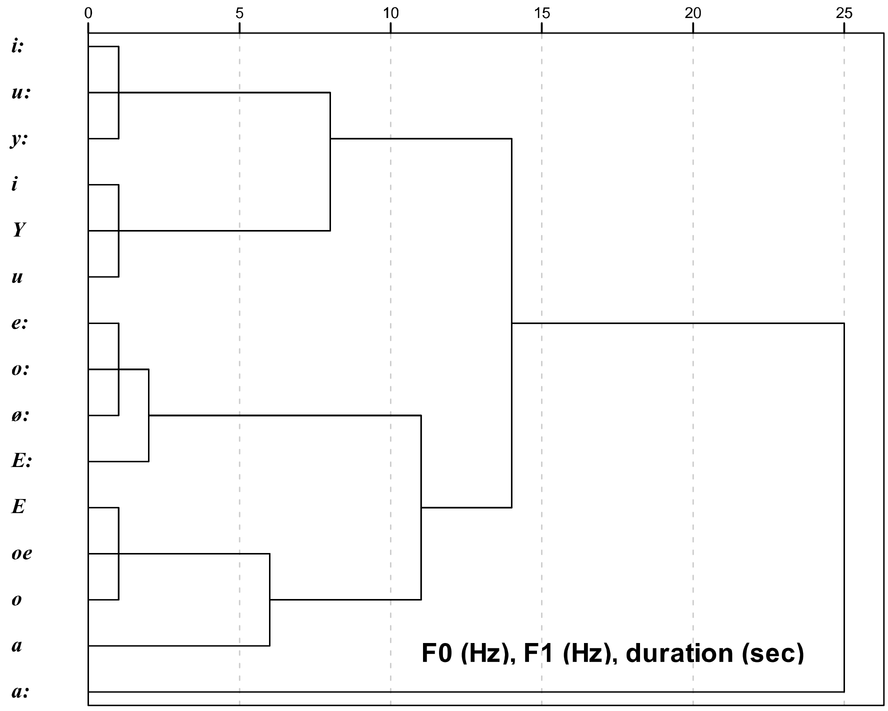

Figure 5.19: Hierarchical Clustering based on mean F0 values (in Hz), mean F1 (in Hz) and mean duration (in sec) of items in experiment 2

For Standard Austrian German, Moosmüller (2007: 118ff) observes that in stressed positions in read speech as well as in spontaneous speech, pre-palatal /i:/ vs. mid-palatal /e:/ as well as velar /u:/ vs. upper-pharyngeal /o:/ were differentiated in F0 by most speakers, while in unstressed position they were differentiated only by some speakers. Moosmüller concludes that F0 might act as an additional cue to differentiate vowels at different constriction locations, namely pre-palatal from mid-palatal vowels, upper-pharyngeal from velar vowels, and pre-palatal and velar vowels from lower-pharyngeal vowels in most formal speaking tasks. A look at the hcluster diagrams in Figure 5.17 and Figure 5.19 suggests that this hypothesis may also hold for the data of the present study.

To prove this hypothesis, an independent samples t-test was performed with the input data of the present study to test whether significant differences in F0 are given for pre-palatal vs-mid-palatal rounded and non-rounded qualities as well as for velar vs. uvular vowels. Levene's test for equality of variances showed that equal variances can be assumed in all cases. Only for /ɪ/ vs. /ɛ/, the *p*-values showed no statistically significant differences in F0 ($p = .926$). For all other pairs, significant differences in F0 were found: /i:/ - /e:/ ($p < .001$), /u:/ - /o:/ ($p < .001$), /y:/ - /ø:/ ($p < .001$), /ʊ/ - /ɔ/ ($p = .001$), /ʏ/ - /œ/ ($p = .004$).

The hierarchical clustering in Figure 5.19 visualizes the relative similarity of German vowel categories (here only in stressed syllables) with respect to all three parameters F0, F1 and duration. It shows that high front and non-front vowels, i.e. pre-palatal and velar qualities with a low degree of jaw aperture are clustered together, i.e. they can be considered as clearly distinct from all other qualities in terms of these three properties. Lower pharyngeal /ɑ:/ is the most distinctive category compared to all other categories.

The hcluster moreover shows that pre- and mid-palatal vowels do not occur in the same cluster and can therefore be considered as sufficiently distinct from each other.

For Danish, Robert-Peterson (1986) showed that listeners tended to identify an ambiguous vowel in a synthesized continuum between Danish /u:/ and /o:/ rather as /u:/ when the signal showed a higher F0. This observation that suggests that intrinsic F0 variation is relevant in vowel perception and may also be relevant for phonemic distinctions in L2 perception.

As Fischer-Jørgensen (1990) states, the relation between *F0* and vowel *height* or rather *jaw aperture*, which is most directly reflected in *F1*, together with *intensity* and *duration* has to be treated as one coherent problem. Within each property and between these properties there is a complex interplay of physiological, acoustic and psychoacoustic mechanisms and constraints (see also Neweklowsky 1975). However, the experimental isolation of influential parameters is a challenging task that cannot be performed in this study.

5.5 Frequency of occurrence

The learners' language experience is determined substantially by the available L2 input and the frequency of occurrence of a given sound contrast. The relative frequency of vowel qualities in the learners' input (language-specific or language-universal type frequency vs. language-specific token frequency) is expected to be of relevance to explain the relative difficulty of a given vowel quality. Learners will establish phonemic categories in L2 easier for more frequently occurring vowel qualities than for less frequently occurring ones.

Meier (1967) analyzed the token-frequency of German vowel *phonemes* in the so-called "100,000 sounds corpus", a text corpus from poetry and prose texts. In this corpus, 38,709 vowels (38.71%) and 61,291 consonants (61.29%) occurred[54].

The data (see Table 5.8 and Figure 5.20) show that long (and constricted) vowels are less frequent (8.33%) than short unconstricted vowels (13.57%). Meier's (1967) analysis describes long and short vowels and a subgroup of short but not "lax" vowels that usually occur in loanwords (*'Zitrone'*, *'Lokal'*), where they are realized as [+constricted] but not [+long]. Long vowel qualities (transcribed as <a. e. i. o. u. ö. ü.>) occur in non-stressed syllables and

[54] Vowels occurring in diphthongs are counted separately, i.e. as a short open vowel and a following short closed vowel, e.g. /a/ and /e/ in /ae/ are counted separately (Meier 1967: 250).

words of foreign origin are least frequent (0.54%). ə-schwa occurs with a high token frequency of 8% and is thus the most frequent sound in the corpus.

type	% tokens	example	% vowel tokens
a:	1.568	nach	4.05
a	2.856	das	7.38
a.	0.038	Alaun	0.10
e:	1.149	mehr	2.97
e.	0.066	jedoch	0.17
e:/e.	1.149	der	2.97
e=ä	2.464	des, nett, hätt	6.37
ä:	0.242	wäre	0.63
e-schwa	8.666	eine	22.39
i:	2.449	die	6.33
i	3.962	in	10.24
i.	0.25	Lineal	0.65
o:	0.971	so	2.51
o	1.288	von	3.33
o.	0.068	Lokal	0.18
ö:	0.226	möglich	0.58
ö	0.102	können	0.26
ö.	0	ökonomisch	0.00
u:	0.98	zu	2.53
u	2.109	und	5.45
u.	0.092	Ulan	0.24
u̯	0	Statue	0.00
ü:	0.441	für	1.14
ü	0.297	würde	0.77
ü.	0.006	Tyrann	0.02
ae	2.458	ein	6.35
ae	2.458	ein	6.35
ao	0.852	auf	2.20
ao	0.852	auf	2.20
oö	0.325	euch	0.84
oö	0.325	euch	0.84
Σ	38.709		100.00

Table 5.8: Token-frequency of German vowels in the "100,000 sounds corpus" (Meier 1967: 250, 253). 38,709 vowel tokens counted (38,71%)

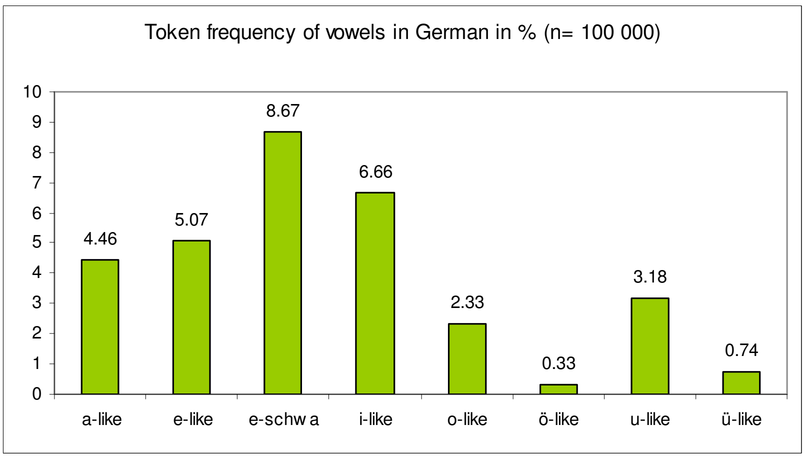

Figure 5.20: Token frequency in % of vowel sounds in German in Meier (1967: 250, 253)

An analysis of the frequency of vowel graphemes in the same corpus (N=124,503 graphemes) shows that <e> is by far the most frequent grapheme, followed by <a i o u>, while the so-called "Umlaut"-vowels are least frequently occurring (see Figure 5.21).

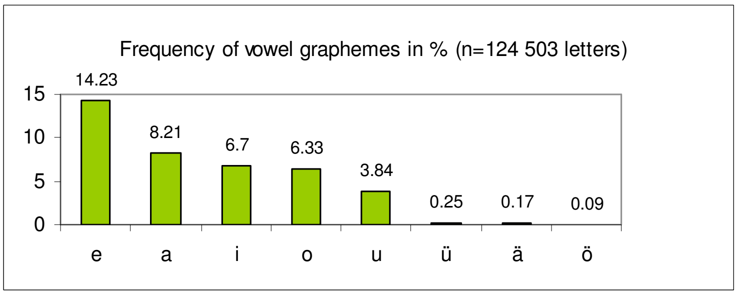

Figure 5.21: Frequency in % of vowel graphemes in Meier (1967: 334)

Figure 5.22 represents the token frequencies for German vowel phonemes (Schröder & Nadie 2014) from word-frequencies in the CELEX-database (Baayen, Piepenbrock & Gulikers 1995) derived by Aichert, Marquardt & Ziegler (2005[55]) indicating the number of vowels per 1,000,000 words in the CELEX corpus (word-frequency). All front rounded vowels as well /ɛː/ show very low frequency of occurrence scores in the CELEX-database.

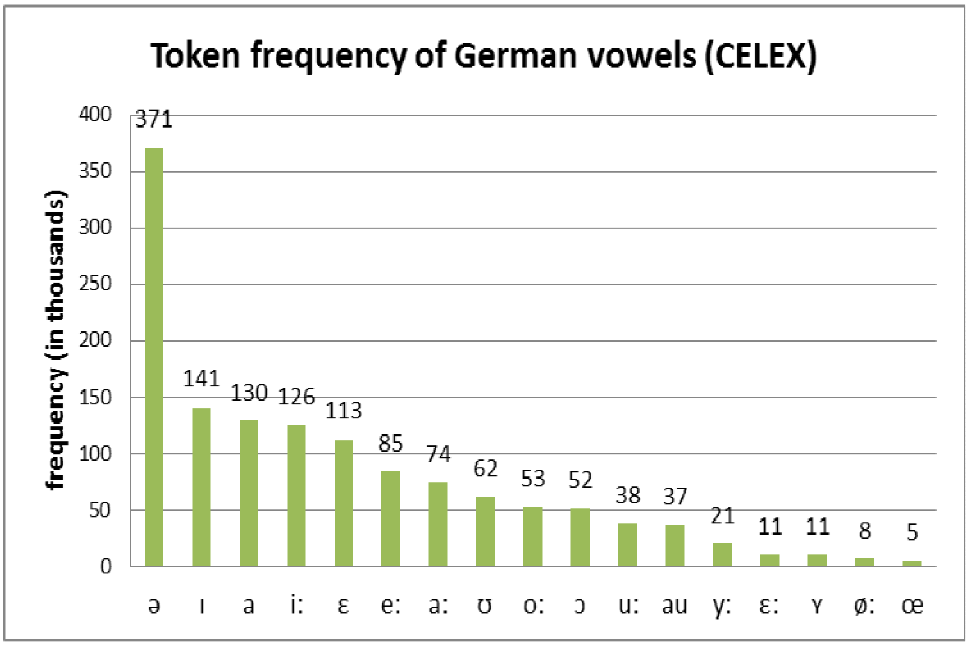

Figure 5.22: Vowel frequency (in thousands) per 1,000,000 words in the CELEX Database (Schröder & Nadie 2014, derived from word-frequency by Aichert, Marquardt & Ziegler 2005)

[55] Schröder & Stadie (2014), based on Aichert, Marquardt & Ziegler (2005) http://www.uni-potsdam.de/fileadmin/projects/treatmentlab/assets/Phonemfrequenzen_TreatmentLab.pdf

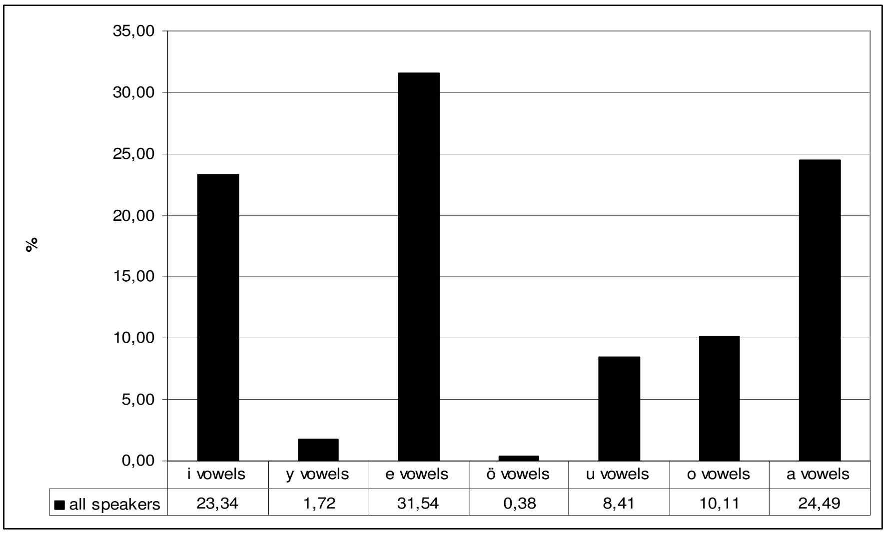

Figure 5.23: Frequency of occurrence in % of vowels in all prosodic positions in spontaneous speech (Standard Austrian German). Figure reproduced from Moosmüller (2007: 52).

Moosmüller (2007: 52) presents a frequency analysis of spontaneous speech for Standard Austrian German. In spontaneous speech, a clear predominance of front unrounded vowels is observed (31.5% e-vowels, 23.3% i-vowels), front rounded vowels are very rarely occurring, and back rounded vowels are comparatively less frequent than front vowels (10.1% o-vowels, 8.4% u-vowels). a-vowels have a frequency of 24.5% in the analysed material (see Figure 5.23).

To summarize, German vowel phonemes and corresponding graphemic representations are very unequally distributed in terms of token frequency as evident in Table 5.8 and Figure 5.20, Figure 5.21, Figure 5.23 and Figure 5.22. In written texts as well as in spontaneous speech, ü- and ö-sounds are very uncommon compared to the plain vowels. The e-qualities are the most common sounds, followed by a- and i-sounds. Back vowels are considerably less frequent. These differences in frequency of occurrence may have an effect on the acquisition rate of vowel categories in L2 acquisition. L2 vowel categories for sounds that are less frequently occurring in the input are expected to develop at a slower rate than more frequently occurring sounds (for discussion of the study's results, see section 12.3.3.2).

5.6 Phonemic vowel contrasts in German

To conclude, the phonemic differentiation of vowel contrasts in a given system refers to a set of primary and secondary gestural features that determine the acoustic resonances in the vocal tract and the interpretation by the listener.

Alternatively to the traditional representation in terms of the IPA vowel chart, an integrated articulatory-acoustic approach as presented in section 4.7.5 and 4.7.6 (cf. Wood, 1979, 1986, 1990; Moosmüller 2007) is proposed here for modelling the German vowel system as in Figure 5.24 (for visualization, see also Jones 1932: 36).

The present approach differs from more traditional accounts that are based on the three major features of Height, Frontness/Backness and Rounding. Alternatively to this view, the proposed description of the German system – as visualized in Figure 5.24 and summarized in the feature matrix in Table 5.9 – is based on a primary dimension of *constriction location*, i.e. the location of major narrowing in the vocal tract. For front vowels in the German system, a differentiation of pre-palatal and mid-palatal qualities is made.

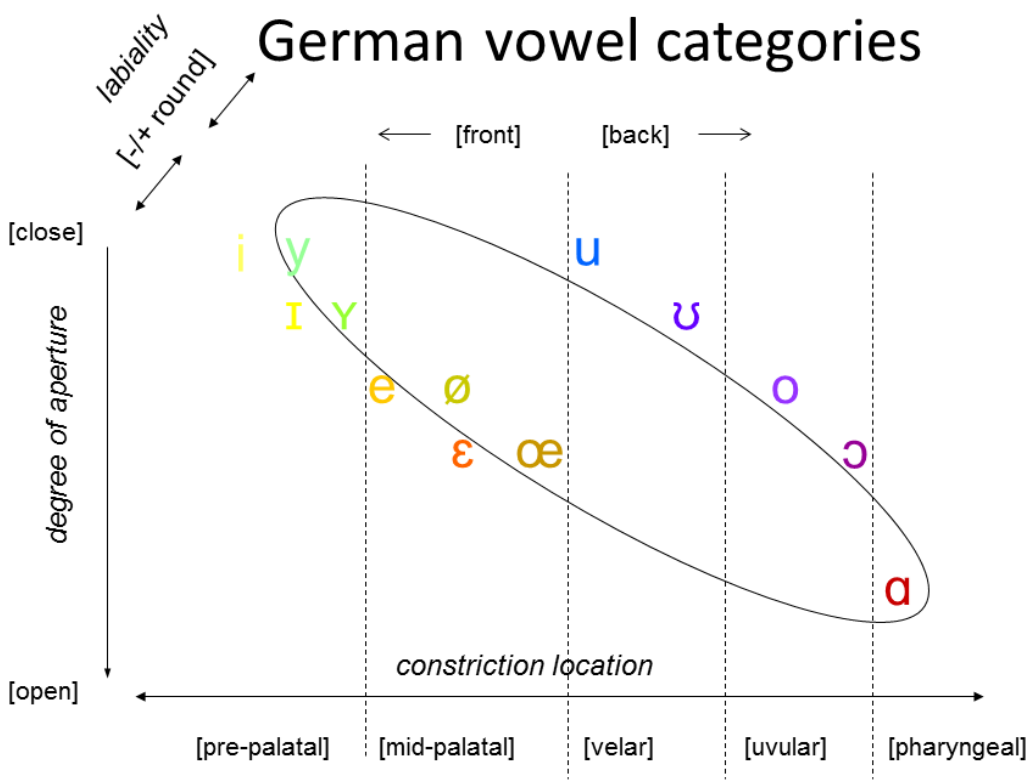

Figure 5.24: Schematic representation of the German vowel system

The traditional IPA quadrilateral cannot be considered as a precise model of the articulatory constellations in the vocal tract specifically with respect to the features Height and Backness. Especially with back vowels, Height[56] – whether conceived as binary or scalar – cannot provide precise constructs for articulatory differences between u-, o- and a-vowels (for discussion, see section 4.7.5 and 4.7.6, see Wood 1979, 1982a: 17). Rather than Height, it is the place of narrowing (constriction location) in the vocal tract that differentiates these qualities and determines acoustic resonances: *velar u-vowels* are located at a constriction location that is more front than *uvular o-vowels*; *u*- and *o*-vowels are more front than *pharyngeal a*-vowels.

Also, contrast pairs such as German /y:/ - /u:/ can be treated in analogy with /u:/ - /o:/ as differing in relative Frontness/Backness. The schematic diagram in Figure 5.24 reflects the articulatory proximity between [i]-like and [y]-like vowels and velar [u]-vowels more precisely than the IPA quadrilateral. The IPA chart rather suggests maximal distance of front [i]-like and [y]-like vowels and back [u]-vowels, even though in fact within the class of back vowels velar u-qualities are nearest to the front palatal constriction location. The schematic representation of the German vowel system proposed in Figure 5.24 ahows to be more appropriate than the traditional quadrilateral IPA chart to account for context-specific occurrences of [y]-like allophones of /u/-vowels as observed e.g. in several varieties of English, or for substitutions of the type /y:/ > /u:/ as frequently observed in L2 data.

The relevant feature specifications for the German vowel system can be summarized as in Table 5.9. Basic phonemic distinctions of German vowel phonemes refer to the *location of constriction* in the vocal tract. Further contrasts are expressed in terms of additional features:

- [+/– open] referring to *jaw aperture* (see 4.5.4),
- [+/– constricted] referring to *constriction degree* (see 4.5.3),
- [+/– round] referring to *lip rounding/protrusion* (see 4.5.2), and
- [+/– long] referring to *phonemic length* (see 4.5.6).

In this sense, contrasts like /e:/ - /ɪ/ or /ø:/ - /ʏ/ strongly refer to a difference in Frontness (pre-palatal vs. mid-palatal) together with constriction degree & phonemic length and less to differences in "tongue height" (see section 5.4.1 and 5.4.4).

These features are differential in nature: they refer to differences between categories (Serniclaes 2011). In L2 perception, the differences between categories and not the single

[56] for a critical discussion of "tongue height", see Ladefoged (1993: 221) and section 5.4.1 and 5.4.4.

categories themselves are considered to be difficult. It is the task of the L2 listeners to detect those contrasting phonetic properties and acoustic cues that indicate vowel contrasts in the target language.

constriction location									
pharyngeal	-		-	-	+				
uvular	-		-	+	-				
velar	-		+	-	-				
palatal	+		-	-	-				
pre-palatal	+	-	-	-	-				
mid-palatal	-	+	-	-	-				
						open[57]	constr	round	long
	i:					-	+	-	+
	ɪ					-	-	-	-
	y:					-	+	+	+
	ʏ					-	-	+	-
		e:				-	+	-	+
		ɛ				+	-	-	-
		ɛ:				+	-	-	+
		ø:				-	+	+	+
		œ				+	-	+	-
			u:			-	+	+	+
			ʊ			-	-	+	-
				o:		-	+	+	+
				ɔ		+	-	+	-
					ɑ:	+	+	-	+
					ɑ	+	+[58]	-	-

Table 5.9: Feature matrix for vowel contrasts in German

The present analysis of the German system as represented in Table 5.9 differs in a few aspects from the feature matrix that has been proposed by Moosmüller for Standard Austrian German (2007: 121, see Table 5.3 above). Differences concern the integration of the feature [+/–long] in the system and the treatment of mid-palatal and rounded vowels with respect to the [+/–open] feature (see discussion in section 4.5).

Even if duration did not prove to be discriminatory in less formal styles and in unstressed positions in Moosmüller's (2007) acoustic analysis of Standard Austrian German, the feature

57 [open] refers primarily to jaw opening for front vowels, where the position of the jaw which is higher for /i, ɪ/ than for /e, ɛ/. For back and front roundedvowels, the explanatory value of jaw opening is restricted, lip rounding and the relative degree of mouth aperture are more descriptive here.

58 The acoustic data suggest that /ɑ/ and /ɑ:/ are only differentiated in terms of [+/-long]. /ɑ:/-/ɑ/ and /ɛ:/-/ɛ/ are considered here to differ only in phonemic length.

[+/–long] is included in the present study in the feature matrix as well as in the set of response categories in the perception experiment. In the current study, the long vs. short differentiation between categories is supposed to be an important characteristic of the German vowel inventory that is explicitly realized in the input stimuli (see section 4.5.5). Phonemic length is particularly relevant to account for the contrasts /ɑ/ - /ɑ:/ and /ɛ/ - /ɛ:/. Phonemic length, moreover, can be considered as of strong psychological relevance for L2 listeners not only due to its physical existence in the speech data at least in more formal styles in stressed syllables but also due to its (though largely inconsistent) representation in German orthography.

The acoustic analysis of the input stimuli suggests that only one /ɑ/-quality is given that is further differentiated in terms of length. In other words, in Standard Austrian German, the feature [+/– front] does not seem to be appropriate to differentiate ɑ-vowels since this contrast is not realized by qualitative differences. Thus, /ɑ:/-/ɑ/ as well as /ɛ:/-/ɛ/ are described here as differing only in phonemic length[59].

Table 5.10 presents the mean formant frequency values (F1, F2, F3), the mean duration, and mean F0 of test items for all categories that were used in the present study[60]. Figure 5.25 provides a visual comparison of formant frequency values F1, F2 and F3 from Table 5.10.

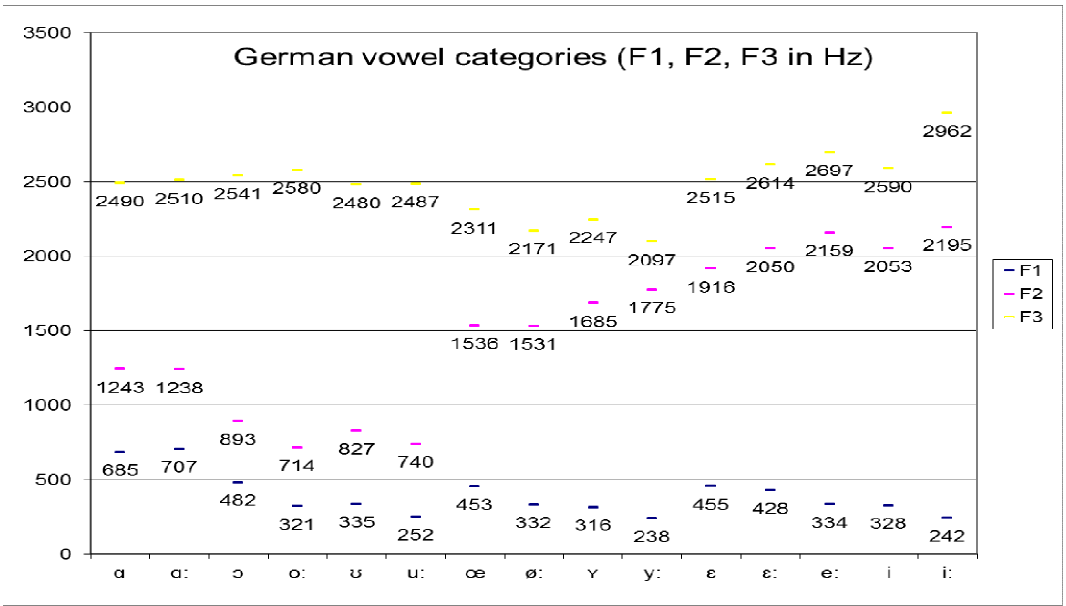

Figure 5.25: Comparison of mean formant freuquencies F1, F2, F3 (in Hz) for all input categories

[59] In Moosmüller's (2007) analysis, [long] is considered as a redundant feature because it co-varies with the feature [constricted] in contrasts of [+contricted]/+long] vs. [-constricted]/[-long]. For Standard Austrian German, Moosmüller does not assume the existence of a phonemic /ɛ:/.

[60] Each vowel category occurs 18 times in the test material and was produced in different pre-vocalic and post-vocalic context and in changing prosodic position (see section 6.3)

	F1 (Hz)	*SD F1*	F2 (Hz)	*SD F2*	F3 (Hz)	*SD F3*	*F0 (Hz)*	*SD F0*	duration (ms)	*SD duration*
ɑ	685	*39*	1243	*122*	2490	*84*	155	*14*	104	*23*
ɑ:	707	*41*	1238	*60*	2510	*87*	157	*11*	197	*25*
ɛ	455	*34*	1916	*58*	2515	*72*	165	*18*	100	*18*
e:	334	*15*	2159	*46*	2697	*87*	168	*18*	175	*26*
ɛ:	428	*24*	2050	*53*	2614	*57*	164	*16*	192	*21*
œ	453	*21*	1536	*56*	2311	*77*	171	*14*	109	*18*
ø:	332	*11*	1531	*85*	2171	*65*	163	*14*	184	*19*
ɪ	328	*29*	2053	*107*	2590	*106*	174	*47*	74	*22*
i:	242	*16*	2195	*48*	2962	*86*	182	*26*	150	*23*
ʏ	316	*17*	1685	*66*	2247	*67*	181	*19*	83	*27*
y:	238	*18*	1775	*98*	2097	*71*	182	*22*	151	*24*
ɔ	482	*38*	893	*107*	2541	*106*	167	*19*	108	*16*
o:	321	*81*	714	*82*	2580	*88*	169	*19*	186	*20*
ʊ	335	*20*	827	*169*	2480	*99*	185	*26*	87	*19*
u:	252	*23*	740	*96*	2487	*326*	186	*26*	157	*27*

Table 5.10: Mean formant frequency F1, F2, F3 (in Hz) and mean duration (in ms) for all input categories (test items in varying consonantal context realized by one male speaker)

5.7 Simulating patterns of perceptual similarity for German vowel categories

5.7.1 Phonetic similarity in terms of physical properties

Cross-language phonetic similarity can be described in three dimensions: (1) articulatory similarity, (2) acoustic similarity, and (3) perceptual similarity.

The similarity of German vowel phonemes in terms of articulatory and acoustic properties has been described in the previous sections and has been illustrated with examples from the input stimuli of the present study. The L2 listeners' responses collected in the empirical part of this study will describe the *perceptual similarity* relationships of German vowel categories as perceived by non-native listeners with different language-background. A description of *physical similarity* in terms of articulatory characteristics and acoustic properties can serve to predict possible difficulties in differentiating German vowel contrasts in L2 perception. The scope of these predictions will be extended here by means of a *hierarchical cluster analysis* based on the mean values of acoustic properties of the input stimuli. Different cluster solutions will be proposed grouping those categories together that are more similar with respect to specific acoustic properties. The clustering will serve to summarize and visualize the similarity relationships that can be established for German vowel phonemes. Based on this analysis, patterns of perceptual confusion will be predicted.

5.7.2 Hierarchical clustering of input stimuli

As has been discussed in detail in chapter 2 and 3, the representation of vowel categories as points in a two-dimensional F1xF2 space based on mean values for the first two vowel formants measured at a specific point in time provides an insufficient description of phonetic reality since it does not account for the immense phonetic inter- and intra-stimuli variation. Higher formants, duration and dynamic variation have to be included in the description of similarity relationships of vowels in a given language.

A hierarchical cluster analysis provides a possibility to establish similarity relations between vowel categories by grouping vowels into a certain number of discrete clusters that best agree, in a sense, with a matrix of symmetrized measures of proximity or distance. It represents a group of different items in a number of clusters and sub-clusters each of which contains those items that are found to be most similar to each other. Hierarchical clustering can thus offer a holistic representation of all vowel categories in a system, comparing them with respect to more than one property and reflecting either the *physical properties* of single items (if physical properties are given in the input) or the perceived *perceptual distances* between them (if distance measures are given in the input). In this section, similarity relationships in terms of acoustic properties will be described to account for the *physical distance* or similarity between phonemic categories. To describe *perceptual distances* between vowel categories, a Multidimensional Scaling procedure will be used based on the listeners' responses in the identification task (see chapter 6).

The clustering is based on the *mean values* of formant frequencies and duration for each of the 15 vowel categories. For the clustering procedure in SPSS Statistics, the pair-wise proximities between the clusters were computed with the squared Euclidean distance. The distance between the objects and each of the new clusters was calculated with the Average Linkage between Groups algorithm by which cluster proximity between subgroups is defined as the average pair-wise proximities of all pairs of points from different clusters. The variable values were standardized to z-scores (with a mean of *0* and a standard deviation of *1*).

The purpose of the analyses presented here is to model different representations of possible similarity relationships between German vowel categories as they could be established by listeners based on their specific interpretation of the acoustic cues available in the input. The resulting clustering can be used to simulate the listeners' possible weighting of acoustic cues by variation of the input data (F-means and/or F-distances (in Bark), with or without duration).

Different solutions were obtained that are determined by the choice of acoustic parameters in the input. The input items are described and compared in the cluster analysis in terms of the frequency measurements of the first three formants F1, F2, F3 (*mean Fs* in Hz/Bark), and/or in terms of formant frequency distances F3-F2, F2-F1 (*F-distances* in Hz/Bark) with or without durational information (in sec).

Four different solutions will be presented and discussed here, which differ in the input data used for analysis: Solution 1 (Figure 5.26) represents a hierarchical clustering (hcluster) based on the mean formant frequencies F1, F2, and F3 (mean Fs in Bark). Solution 2 (Figure 5.27) represents an hcluster based on the mean formant frequencies F1, F2, F3 (in Bark) and durational information (in sec). The hcluster in solution 3 (Figure 5.28) is based on the mean frequency distances F3-F2 and F2-F1 (F-distances in Bark) and durational information (in sec)[61]. Solution 4 (Figure 5.29) combines the mean formant frequencies F1, F2, F3 (mean Fs) with the mean formant frequency distances (F-distances).

The features tagged to the branches of the resulting dendrograms are a post-hoc interpretation of the hierarchical cluster and were not included in the input for cluster analysis.

No unique solution or absolute interpretation of acoustic similarity or distance can be presented here since the output of the obtained cluster solutions depends very much on the available input data.

As becomes evident from the dendrograms of the four solutions, an interpretation in terms of absolute formant frequencies differs considerably from an interpretation in terms of formant distances. Moreover, the use of durational differences in the input yields substantially different clusters in the output.

All item clusters are arranged hierarchically with individual items at the leaves and one single cluster at the root of the dendrogram. The distances between the single vowel categories are small in the leaves or far branches of the tree and large in the near branches. Minimally distant items are merged in a cluster in the leaves of the tree reflecting maximal acoustic *similarity*. The height at which two clusters are merged in the dendrogram reflects the relative *distance* of the two clusters. Earlier branching indicates larger distance and less similarity in

---

[61] A representation of vowels in an F1/F2 plots are most commonly used, though some researchers prefer other kinds of representation. Ladefoged & Maddieson (1990) suggest that the difference F2-F1 gives a better representation of frequencies. Some researchers, such as Traunmüller (1981) or Syrdal & Gopal (1986), suggest to use F1-F0 additionally to F2-F1 (F2-F1 serving as speaker-independent measure of vowel frontness). F1-F0 would then be used as a speaker-independent measure of vowel openness and F0 would serve to normalize the differences between male and female first formants). However, this would imply that a higher pitch of a speaker would cause a higher F1 for a given vowel quality than a lower pitch. A speaker-independent measure of vowel quality is still elusive (cf. Deterding 1997).

terms of physical properties of the speech signal. For items clustered together in the leaves of the dendrogram more perceptual confusion can be expected than for items belonging to a cluster that is differentiated earlier in the procedure.

All solutions show that the pharyngeal a-qualities are classified as maximally distant from all other categories and (if at all) are most closely associated with a cluster of [-front, +round] vowels. This is reflected by a separate branch indicating maximal distance of this sub-group in terms of quality from all other branches of the dendrogram and by the very late division of /ɑ/ and /ɑ:/ indicating minimal distance or maximal similarity within the class of pharyngeal vowels due to very small qualitative differences between long and short /ɑ/-vowels.

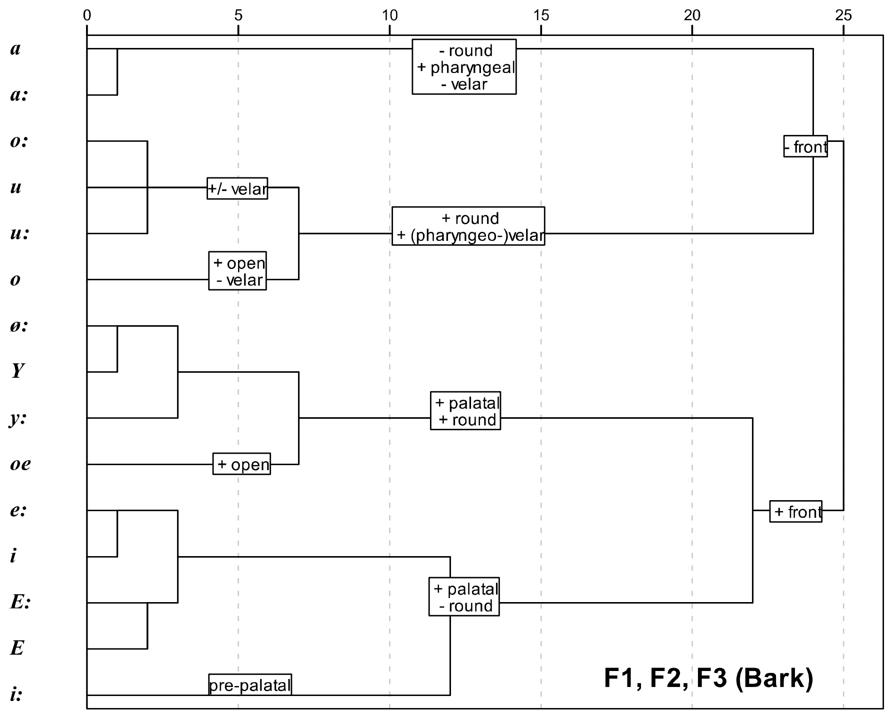

Figure 5.26: Solution 1 based on absolute formant frequencies (mean F1, F2, F3 in Bark): Dendrogram from the Hierarchical Cluster Analysis (average linkage between groups).

In solution 1 (mean Fs in Bark), front vowels are clearly divided from non-front vowels. Non-front vowels branch in non-rounded a-qualities vs. rounded o- and u-qualities. The clusters at the left margin of the dendrogram show that within a given vowel class, e.g. [+front, +round], [+front, -round], or [-front, +round] vowels, one quality is clearly distinctive from the others. For the palatal non-rounded vowels it is pre-palatal constricted /i:/ that is maximally distinct from all other members of this class. This is reflected by the relatively early branching of /i:/ compared to the other members of the palatal non-rounded sub-group. The spectral similarity of /e:/-/ɪ/ and /ɛ:/-/ɛ/ is clearly reflected in this model, since durational information is not included in this solution. The early branching of /i:/ is due to high F3 for /i:/ indicating a fronted pre-palatal constriction location. This clustering pattern can therefore be interpreted as maximal distinctness of /i:/ compared to all other palatal non-rounded vowels, whereas /e:/-/ɪ/ and /ɛ:/-/ɛ/ are least distinctive in terms of spectral information. Therefore, more perceptual confusion is expected to occur with /e:/ vs. /ɪ/ and /ɛ:/ vs. /ɛ/, than with /i:/ vs. any other palatal non-rounded qualities.

Within the class of rounded palatal vowels, it is /œ/ that is maximally distinct from all other members of the class due to a lower F1 indicating a higher degree of aperture. The spectral differences between /ø:/ and /ʏ/ are the weakest, while /y:/ is more distinct. However, /y:/ can be estimated as similar to /ʏ/ as well as /ø:/, i.e. insufficient differentiation can be expected with both vowel pairs, /y:/ vs. /ʏ/ and /y:/ vs. /ø:/.

A very similar pattern is observed within the class of back rounded vowels with unconstricted mid-palatal /ɔ/ vs. /o: ʊ u:/. /ɔ/ is maximally distinct from all other back rounded vowels due to its high F1 reflecting a high degree of aperture. The very flat cluster for the other three members of this class reveals that these qualities can be considered as very similar in terms of formant frequencies since no durational information is available in this solution.

/ɔ/ as well as /œ/ are the qualities with least lip rounding and the highest degree of aperture that clearly differ from all other members within the class of [+palatal, +round] or [-palatal, +round] vowels by their high F1 being maximally distinct from the values for /u:/ and /y:/ respectively. Therefore, a higher degree of differentiation and less perceptual confusion of /u:/ vs. /ɔ/ and of /y:/ vs. /œ/ is expected than between the other members within the class of back rounded or front rounded vowels respectively.

Solution 1 merges velar and uvular vowels into one cluster and mid- and pre-palatal into two clusters that are differentiated by acoustic patterns that seem to refer to lip rounding. Even if this solution is not conform with the distinction of velar vs. uvular and of mid- vs. pre-palatal vowels, it may serve as a good simulation of some perceptual difficulties that L2 learners may have in differentiating e.g. German /o:/-/ʊ/ or /u:/-/ʊ/ or even /o:/-/u:/, but less problems in differentiating /ɔ/ from /o:, ʊ, u:/. A similar pattern of under-differentiation could be expected for /e:/-/ɪ/, /ɛ:/-/ɛ/ and /e:/-/ɛ:/ and even /e:/-/i:/, but less often for /e:/-/ɛ/, as well as for /y:/-/ʏ/, /ø:/-/ʏ/ and /y:/-/ø:/.

Solution 2 (mean Fs + duration) shows a different pattern. Since durational information is available in this solution, the primary branching is strongly determined by differences in duration. /ɑ/-qualities have the highest intrinsic duration and are moreover clearly distinct in terms of other acoustic features and are therefore separated from all other categories very early in the dendrogram.

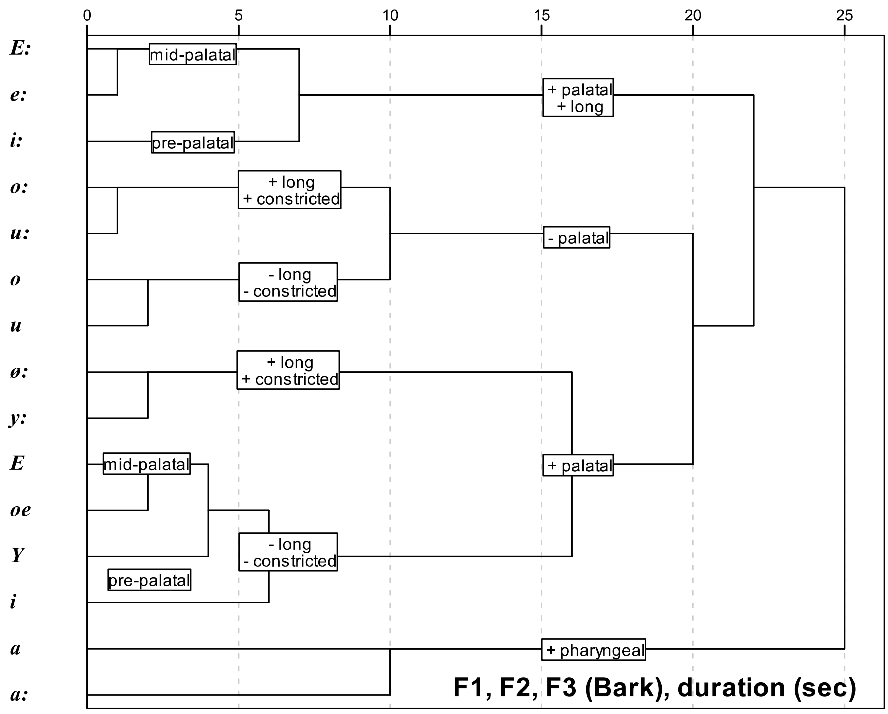

Figure 5.27: Solution 2 based in absolute formant frequencies (mean F1, F2, F3 in Bark) and duration (in sec): Dendrogram from the Hierarchical Cluster Analysis (average linkage between groups)

With the further branching in solution 2, a clear-cut interpretation in terms of phonological features is more difficult. For mere statistical criteria, solution 2 is not an optimal model, because the branches of the tree divide relatively early, i.e. they are closer to the right margin than in other solutions. Moreover, the model shows no clear distinctions with respect to specific features. However, solution 1 and 4 cannot simulate the under-differentiation of palatal rounded vs. non-rounded vowels that is frequently observed in L2 production and perception by L2 learners of German (for L1 Polish, see e.g. Hentschel 1986; Kerschhofer 1995; for American English listeners of French, see e.g. Levy & Strange 2008; Levy 2009). This pattern of "unrounding" is better reflected in solution 2 and 3, especially for short palatal qualities. In solution 2, within the cluster for short palatal vowels, /ɛ/ vs. /œ/ shows the highest degree of similarity, whereas in solution 3 (F-distances and duration) /ɛ/, /ɪ/ and /ʏ/ are clustered most closely together. Another similarity that becomes evident in solution 2 is the proximity of /e:/ and /ɛ:/.

Solution 2 and 3 show that in terms of relative formant distances front rounded vs. front unrounded vowels are acoustically quite similar and can therefore be considered to be perceptually similar as well. A high degree of under-differentiation of +/–rounded palatal vowels can therefore be expected.

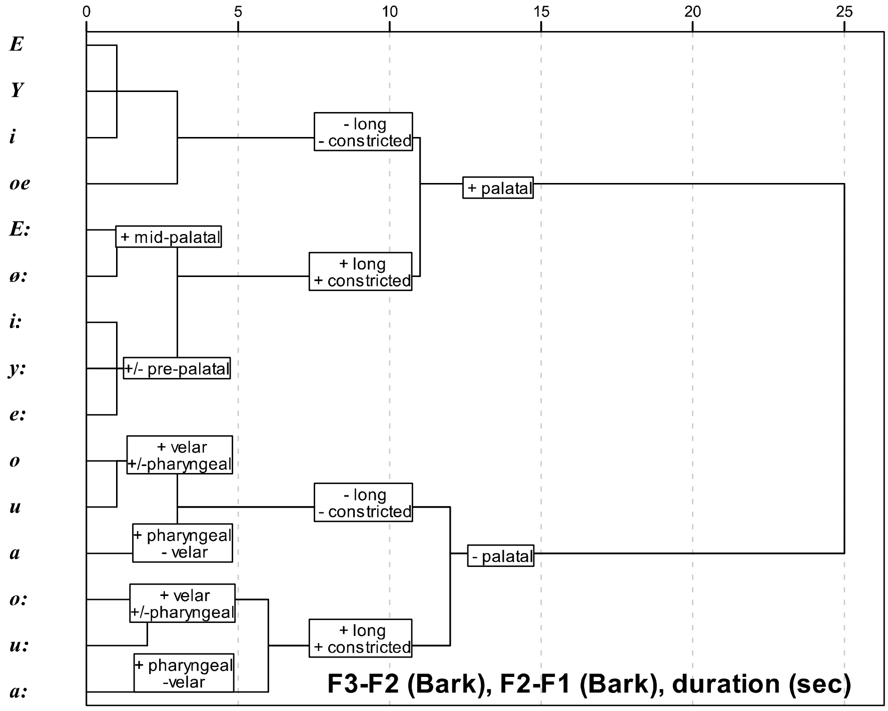

Figure 5.28: Solution 3 based in absolute formant frequencies (F-distances in Bark) and duration (in sec): Dendrogram from the Hierarchical Cluster Analysis (average linkage between groups)

In solution 3 (F-distances + duration), qualitative and quantitative information is integrated yielding a different output that can simulate perceptual problems related to quality differentiations within a given class of vowels of same quantity or constriction degree [+/– long] or [+/–constricted]. Here, all palatal vowels are subsumed in one cluster, reflecting the high F1-F2 distance together with the relative proximity of F2 and F3. For non-palatal vowels, F1-F2 distance is small while the distance of F3-F2 is higher. The cluster of palatal vowels is further divided into a group of short unconstricted palatal qualities (/ɛ ʏ ɪ œ/) and a group consisting of long constricted mid- and pre-palatal qualities (including unconstricted /ɛ:/). This [+long, +constricted] cluster is divided into two further groups: mid-palatal /ɛ: ø:/ vs. /i: y: e:/. This pattern suggests that the mid-palatal qualities /ɛ:/ and /ø:/ could be interpreted as more similar to each other than to the other members of the class of long front vowels. Presupposing that there is a primary distinction of long constricted vs. short unconstricted palatal vowels, solution 3 can therefore account for patterns of perceptual confusion among front rounded vs. non-rounded vowels (e.g. /i:/ vs. /y:/) as well as for a possible confusion of /i:/ and /e:/, whereas confusion of /e:/ vs. /ø:/ seems less likely. The non-palatal branch of the dendrogram in solution 3 is immediately divided into [-long, -constricted] vowels vs. [+long, +constricted] vowels. Within each of these sub-clusters, the o- and u-qualities are closer to each other than to the a-qualities, a pattern that is clearly understandable if we look at the acoustic data for these vowel categories.

Under-differentiation of palatal rounded and non-rounded vowels in perception as well as in production is a pattern that is relatively common with L2 learners of German who have no front rounded vowels in their L1. A frequently observed pattern of L2 German perception and production is the under-differentiation of /ɪ/-/ʏ/ or /ɛ/-/œ/ (e.g. for Polish natives, cf. Kerschhofer 1995).
In this sense, solution 2 and 3 may be non-optimal representations of the German vowel system, but they can nevertheless simulate some aspects of similarity relationships as established by non-native listeners of German. Both solutions could serve as model for a situation where a certain sensitivity to durational cues is given, whereas qualitative differentiations between vowels within a class appear to be less distinct.

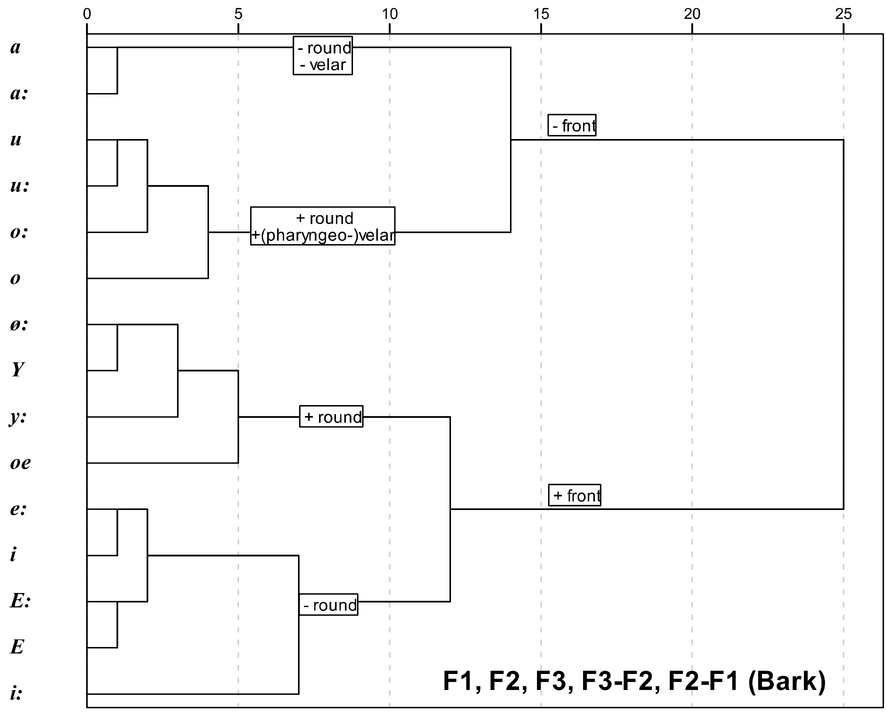

Figure 5.29: Solution 4 based on mean formant frequencies (F1, F2, F3 in Bark) and formant frequency distances (mean distance F2-F1, F3-F2 in Bark). Dendrogram from the Hierarchical Cluster Analysis (average linkage between groups)

Cluster patterns in the output of solution 4 (mean Fs + F-distances) are very similar to the output of solution 1 (mean Fs). However, the later branching of the sub-clusters in solution 4 indicates that from a statistic point of view, this solution is “better” than solution 1. Again, we observe four major clusters: front rounded vowels, front non-rounded vowels, a-qualities, and a cluster of velar and pharyngeovelar u- and o-qualities. Within each of these sub-clusters, one quality is less similar or more distinct from all other members of the same class. For front non-rounded vowels it is pre-palatal /i:/ that is maximally distant from other front non-rounded vowels. /ɛ/-/ɛ:/ and /ɪ/-/e:/ are less distinct (due to the non-availability of durational information), indicating that these qualities could cause perceptual difficulties if durational information is not available or not used for perceptual discrimination.

A similar pattern is observed with /œ/ vs. all other front rounded vowels (/ʏ/, /ø:/ and /y:/) and with /ɔ/ vs. all other non-front rounded vowels if durational information is not available. Within the class of non-front rounded vowels, /u:/ and /ʊ/ are most similar and very near to /o:/. Within the class of front rounded vowels, /ø:/ and /ʏ/ are most similar and near to /y:/, while /œ/ appears to be more distinct.

To summarize, the hierarchical cluster analysis technique is used here to *simulate* patterns of similarity relationships between vowel categories to predict perceptual difficulties in L2. The advantage of this approach is that the analysis is not based on abstract phonological features but on the physical properties of the input stimuli. The features tagged to the branches of the dendrograms are a post-hoc-interpretation of the cluster output. The solutions presented here show considerable differences that become evident by differences in the grouping of categories to clusters and leaves on the left margin of the dendrogram as well as by differences in the features that can be assumed to underly the branching patterns. The branching in the output and the resulting clusters in the leaves of the dendrogram are determined by the choice of the properties in the input (e.g. absolute formant frequencies and/or formant-distances, with or without durational information).

The conclusion that solution *i* is “better” than solution *j* would be inappropriate. None of these solutions can be considered to be the uniquely possible interpretation of acoustic similarity. Rather, each of these solutions represents some aspects of acoustic similarity that may account for similarity relationships as they are established by L2 learners of German on the base of their specific interpretation of the acoustic signal. For a discussion of the hclusters obtained and their relevance for the present study’s results, see section 12.5.

5.8 Hypotheses – Predictions vowel perception in L2 German

5.8.1 Constriction location

The discrimination of German vowel contrasts in cross-language speech perception is generally expected to be good in terms of different constriction locations, since a “*preference for the four constriction locations is apparently universal*” (Wood 1979: 31). Languages with more than four or five vowels make use of additional features for phonemic differentiations expressed by modifications of jaw opening, degree of constriction, lip activity or duration.

These differentiations may be perceptually less easily discernable than primary distinctions in terms of constriction location. In L2 speech perception, the perceptual distinction of primary differentiations in terms of constriction location may be “disturbed” by language-specific differences in the use of secondary differentiations such as constriction degree, jaw opening, lip activity or duration together with language-specific co-articulatory effects and their interpretation in a given language. Contrasts of “similar” phonemes may be easily differentiated in perception by native listeners of German by means of e.g. acoustic correlates to constriction degree or duration or simple by the native listeners’ use of linguistic knowledge in “top-down” speech perception. In L2 speech perception, however, these contrasts may not be differentiated adequately by non-native listeners causing inadequate phonemic categorizations due to language-specific articulatory differences, their acoustic effects and the resulting spectral patterns.

As the acoustic analysis and the hierarchical cluster solutions presented in the previous sections reveal, some contrasts, e.g. /u:/-/o:/, /o:/-/ʊ/, /i:/-/e:/, /e:/-/ɪ/, /y:/-/ø:/, or /ø:/-/ʏ/, show considerable spectral similarity and may therefore be difficult in cross-language perception even if they are not considered to be “similar” to each other under a phonological perspective. From a phonological point of view, these vowel pairs are not similar since they do not fall under the traditional dichotomy of “short lax” vs. “long tense” vowels. Articulatorily, distinctions like e.g. /e:/-/ɪ/, /ø:/-/ʏ/, and /o:/-/ʊ/ are differentiated in terms of constriction location (mid- vs. pre-palatal, velar vs. uvular) together with variation in the degree of constriction and phonemic length. Acoustically, however, these pairs show very similar spectral patterns, especially if durational aspects are not considered. These patterns result from the articulatory constellations for given vowel phonemes. Native listeners are expected to have less problems in differentiating such contrasts due to sufficient experience and the interaction of top-down and bottom-up processes in speech processing. Less experienced non-native listeners will have considerably more problems in disambiguating the input signal especially with contrast pairs that are acoustically similar. For example, for L2 learners of some languages, e.g. Arabic or Farsi, difficulties with the perceptual differentiation of vowels at specific locations are frequently reported by teachers of German. These difficulties concern the differentiation of German pre-palatal i-and y-vowels vs. mid-palatal e- and ö-vowels as well as the contrast of velar u- vs. uvular o-vowels.

A look at the description of the German vowel system shows that there are several facts that may explain such under-differentiation of certain German vowel pairs. The class of back rounded vowels in German consists of /u:/ and /ʊ/ in the velar region and of /o:/ and /ɔ/

located in the upper-pharyngeal region of the vocal tract. As Table 5.11 shows, German u- and o-vowels share a feature [+velar] as common feature but differ with respect to the features [+palatal] for u-qualities and [+pharyngeal] for o-qualities (cf. Wood 1990: 199). Acoustically, u- and o-qualities show considerable similarity (see section 4.5.1.3 and 4.5.4). Variation of F2 is relatively small within a range of about 4-6 cm from the glottis (for a vocal tract of 14 cm, cf. Stevens 1998: 281) and F1 is only slightly higher for /o:/ than for /u:/. Moreover, the relative low amplitude of higher frequencies of non-low back vowels reduces the differentiating role of F3 and F4 in disambiguating the acoustic signal. As the hcluster solutions suggest, the differentiation of non-low back vowels and here especially of /u:/-/ʊ/ and /u:/-/o:/ is expected to be difficult.

	constriction location	
	u-vowels palato-velar	o-vowels pharyngeo-velar
pharyngeal	-	+
palatal	+	-
velar	+	+

Table 5.11: Feature specifications for non-low back vowels

In the palatal region, German distinguishes 8 to 9 different vowel phonemes. At least for Standard Austrian German, a primary distinction of these qualities in terms of constriction location is assumed, i.e. mid-palatal and pre-palatal vowels are distinguished (Moosmüller 2007). Vowel contrasts in this region are differentiated by location of constriction in the mid- vs. pre-palatal region (see Table 5.9) and by means of several additional features (constriction degree, rounding, length).

The pre-palatal constriction location is acoustically relatively instable and moreover less frequently exploited cross-linguistically. In other words, in languages that do not use the pre-palatal region for constriction, the mid-palatal region is used for i-vowels or may be more precisely, articulatory variation of the place of constriction in the palatal region is not used for phonemic contrast of i- vs. e-vowels. As the hcluster solutions 1 and 4 show, pre-palatal /i:/ is clearly differentiated by its spectral characteristics from other vowels in the palatal region by a high F3, even if durational information is not available. With the other pre-palatal vowels, however, F3 is lowered, with /ɪ/ due to its unconstricted nature and with /ʏ/ and /y:/ mainly due to the effect of lip rounding lowering all formants. Under-differentiation of pre-palatal and mid-palatal vowels in L2 perception can therefore expected.

	constriction location			
	i-vowels	ü-vowels	e-vowels	ö-vowels
mid-palatal	-	-	+	+
pre-palatal	+	+	-	-
round	-	+	-	+

Table 5.12: Feature specifications for palatal vowels

Therefore, confusion of the type /u:/-/o:/, /o:/-/ʊ/, /i:/-/e:/, /e:/-/ɪ/, /y:/-/ø:/, or /ø:/-/ʏ/ can be considered as instances of under-differentiation with respect to constriction location, i.e. of an under-differentiation of pre-palatal vs. mid-palatal or velar vs. uvular, but they show, however, a certain influence of other properties such as constriction degree and duration.

Even if they are phonemically distinct, they show considerable similarity in terms of acoustic patterns. For German natives, these contrasts are generally not difficult, because native listeners disambiguate the acoustic signal by their linguistic knowledge (i.e. phonology, morphology, lexicon) of the given language and by the context in the respective speech situation. In L2 perception, however, by co-action of language-specific differences in articulation and similar spectral patterns (which may be combined with the irrelevance of durational cues in L1) together with insufficient linguistic knowledge of the target language, this perceived similarity may result in the learners' failure to distinguish L2 phonemes adequately. In an experimental design with logatoms in the input, this effect is expected to increase.

The experimental data will show if the reported discrimination problems with respect to constriction location especially in the class of palatal vowels and with velar u- vs. uvular o-qualities can be empirically verified with non-native learners of German from different native languages.

5.8.2 Constriction degree and phonemic length

The phonemic contrast between the so-called "short lax" vs. "long tense" vowels, which are here described in terms of degree of constriction and phonemic length, is sufficiently expressed for native listeners of German by acoustic differences in quality (constriction degree) and quantity (phonemic length), but this is not necessarily the case for L2 learners. The irrelevance of durational differences in the learners' L1 together with differences in the distribution of secondary cues or even the absence of functionally relevant distinctions such as [+/–constricted] in the listeners' L1 together with potential coarticulatory effects in L1 and L2 may result in the insufficient differentiation of these contrasts. Moreover, the fact that durational differences in German can be neutralized in less formal speech styles and in unstressed position may even decrease the perceptual discriminability of these vowel contrasts. The isolation of specific acoustic cues that are responsible for the perceptual under-differentiation in L2 can, however, not be performed in the present study that is based on a perception experiment with natural input stimuli but has to be investigated in more details in a study with synthetically manipulated speech sounds.

5.8.3 Rounding

It has already been mentioned that the differentiation of velar vs. uvular vowels is described as difficult for learners of some languages. The exact reasons for perceptual under-differentiation in this region of the vocal tract have to receive special attention. In terms of acoustics, velar and uvular qualities are mainly distinguished by higher formant frequencies (with a relatively low amplitude), whereas differences in F1 and F2 are comparatively small. One reason could be that for non-native listeners the acoustic consequences of lip rounding apparently have a kind of further "blurring" effect on category boundaries in this region of the vocal tract.

A similar effect is moreover expected for the class of front rounded vowels that occur rather seldom cross-linguistically and are considered to be universally marked.

The hcluster solutions suggest however that within the class of front rounded vowels some contrasts will be more distinct in terms of spectral similarity, especially when durational information is not considered. While /øː/ and /ʏ/ are most similar in terms of acoustic properties, the contrast of /yː/ vs. /œ/ is expected to be relatively clear. The hcluster solutions suggest that a differentiation of /yː/-/ʏ/ or /ʏ/-/øː/ and even of /yː/-/øː/ will be less clear for learners of L2 German. /œ/ appears to be most distinct in terms of spectral similarity from all other members of this class.

A similar pattern is predicted for back rounded vowels: While for /oː/ and /ʊ/ the highest spectral similarity is observed, /uː/ and /ɔ/ appear to be more distinct.

The fact that /œ/ and /ɔ/ are more distinct in terms of spectral information is mainly due to their higher degree of jaw aperture and a lower degree of lip activity, i.e. they are less rounded than other members of their class. For /œ/, this low degree of lip protrusion may even result in a perceptual unrounding in L2, as has been shown e.g. for Polish learners of German. Kerschhofer (1995) showed that perceptual substitutions by unrounding, i.e. delabialization, occur most frequently with unconstricted open /œ/. This frequently reported delabialization of front rounded vowels is not directly predicted by the hierarchical clustering, but becomes most evident in hcluster solution 2, where /œ/ clusters together with /ɛ/. Occasionally, perceptual confusion of front rounded vowels may also interfere with difficulties in differentiating back rounded vowels due to the "blurring" effect of lip rounding, even if the hcluster solutions do not show such a pattern in the clustering output. Confusion of German front vs. back rounded vowels may therefore be assumed to be not only due to acoustic similarity but may be co-conditioned by other factors such as graphemic representation.

Cross-linguistically, front rounded vowels are considered to be more "marked" than non-rounded palatals; nevertheless four different rounded palatal qualities are included in the German vowel inventory. Therefore, considerable difficulties are expected specifically with this vowel class in L2 perception of German. In interaction with the insufficient discrimination of pre- vs. mid-palatal rounded vowels, as discussed above, perceptual difficulties are expected particularly with the class of front rounded vowels.

5.8.4 Frequency of occurrence

Perceptual category formation, i.e. the establishment of long-term representations for vowel categories in L2, is an important pre-condition for correct vowel identification. The frequency with which a particular vowel occurs in a language or the learner's input is assumed to have an influence on the learnability of a given vowel category. For less frequently occurring vowels, the process of category formation will be more difficult.

The relative frequency of certain vowel qualities in the target language (*type frequency*) and the frequency of occurrence in the learners' L2 input (*token frequency*) are considered to influence the acquisition of German vowel categories.

From the considerations on vowel frequency in German in section 5.5, the following expectations are derived:

German /ɛ:/ is expected to be relatively difficult for learners of German due to its rare occurrence or even non-existence in many regional varieties of German and even in most regional standards.

Front rounded vowels are universally marked and are moreover – compared to other vowel categories – relatively infrequent in German texts (see Table 5.8 (Meier 1967), Figure 5.20, Figure 5.21, Figure 5.22 and Figure 5.23).

It can therefore be expected that less frequently occurring vowels (e.g. front rounded vowels) in the available L2 input will be more difficult in perception than qualities that occur very frequently in the learners' input. The data will show whether differences within the class of front rounded vowels are given with respect to their percentage of correct identifications and whether these differences can be explained by their acoustic characteristics or cross-linguistic patterns of frequency and distribution.

5.8.5 System size and category distribution

System size is another factor that is expected to be of relevance in the acquisition of an L2 vowel system. While German has a very large vowel system differentiating five different constriction locations (see Table 5.9), the native languages of most of the L2 learners who participated in the present study, have considerably less vowel categories.

Learners from a native language with a large vowel inventory are expected to differ in their acquisition process from learners with a smaller inventory of vowel contrasts. In L2 acquisition, the crucial question is whether a large L1 vowel repertoire facilitates the acquisition of another large vowel inventory or whether it inhibits the correct acquisition due to *perceptual assimilation* and *equivalence classification* to one of the numerous existing L1 categories. Iverson's & Evans' (2007) hypotheses that a "dense L1 vowel space" like German would interfere with learning another complex vowel system like English due to assimilation to L1 categories and that learners with a smaller L2 vowel inventory would learn fast due to "*more unused areas of the vowel space*" (2007: 1625) were falsified in their empirical study. Rather, it seems that a larger vowel inventory makes individuals more sensitive to gradient categorical differences between vowels. Possibly, this greater sensitivity to category differences also facilitates the discrimination of new vowel categories. However, another possible explanation would be that *equivalence classifications* (Flege 1988) between sounds of two larger systems become less apparent and that the probability of "hitting" a quite adequate form in L2 is bigger with a large L1 system as starting point. Therefore, the contrastive analyses provided for each of the learner's languages in chapter 10 will consider not only the system size of the participants' native languages but also those "areas", where less or even no vowel categories or phonemic oppositions exist in the learners' native system.

5.9 Conclusions – Perceptual similarity of German vowels

Languages differ with respect to the size and quality of their phoneme inventories, the set of articulatory constellations and acoustic properties of distinctive speech sounds, and their linguistic interpretation by the speakers/listeners of a language.

The choice of *articulatory gestures* to realize vowel contrasts in a system can vary language-specifically. Articulatory constellations in the vocal tract produce *acoustic patterns* that function distinctively for phonemic differentiations in a given language. A second major component of inter-lingual variation is the language-specific *perceptual* interpretation and *categorization* of the resonances in the vocal tract that is determined by the linguistic experience of the listeners.

Perceptual under-differentiation of L2 contrasts is expected if a non-robust or less salient cue of functional relevance in L2 (cf. section 4.8 and 4.9) to phonemic distinctions in a contrast pair is of relatively less salience or even non-existent in the listeners' L1. Such perceptually non-salient cues might be degraded to the point of being "lost" in the process of speech transmission (either by quality of the input signal or by linguistic knowledge of the L2 learner, cf. Chang, Plauché & Ohala 2001), resulting in the perceptual confusion of the two members of a contrast pair. In case of acoustic ambiguity, listeners will decide for one member of the contrast pair. In terms of traditional *feature-based accounts*, the member without the respective distinctive cue will be the preferred response category in a perception test, i.e. it will replace the more complex or more marked member of the pair in perceptual substitutions rather than the other way round. Thus, patterns of a*symmetry*, i.e. biases or preferences for the less marked member of the contrast pair, are expected to occur in L2 perception. However, as will be discussed in detail in chapter 12, this feature-based view is just one way to describe phonetic differences and perceptual distances. Alternative approaches referring to *spatial representations* of the perceptual vowel space and *perceived* inter-category distances and similarities will be discussed in sections 7.3, 9.5, chapter 11 and chapter 12.

In L2 acquisition, a correlation between the learner's ability to produce a given L2 sound contrast with his/her ability to discriminate the sounds involved is assumed. To summarize the considerations from chapter 2, 3 and 4, we have to state that the listener's behavioural reaction to specific vowel contrasts in an identification task is assumed not only to vary language-specifically but that it is also influenced by intrinsic properties of the sounds in the speech signal, by the listeners' language experience (in L1 and L2) and by their individual hypotheses about the target language system and their perceptual strategies.

Difficulties in L2 perception are not considered to result from mere *transfer* of L1 sound structures to L2 but are supposed to be based on *similarity* relationships established by the individual learners which are determined by linguistic knowledge in L1 and L2, by the input signal itself and by *universal preferences* for certain sound patterns.

The perceived similarity of L2 vowel categories is influenced by (1) the characteristics of the phonemic categories of L2 and their physical manifestations, (2) the listeners' L1 and their knowledge and individual hypotheses in L2, and (3) by language-independent preferences for universally or typologically less marked vowel qualities.

Based on the phonetic description of German vowels and the hierarchical clustering presented in this chapter, it is hypothesized that specific phonemic contrasts of German will cause more difficulties to L2 learners of German, especially if these contrasts are not phonemic in the learners' L1 or if the acoustic cues for phonemic distinctions are interpreted differently in the learners' L1. Difficulties are expected with respect to

(1) contrasts in phonemic length and constriction degree,
(2) the differentiation between palatal rounded and non-rounded vowels,
(3) the differentiation within the class of palatal vowels wit respect to the pre-palatal vs. mid-palatal constriction location,
(4) and presumably also the differentiation of velar from uvular vowels.

These basic predictions are not yet based on a contrastive description of a specific native language but are expected to hold independently of the learners' L1 for the L2 perception of German vowels. They are grounded in the description of the phonetic properties of vowel sounds. A contrastive analysis for each L1 sub-sample will serve to refine these predictions on a language-specific base (see chapter 10).

6 The Empirical Study

This chapter presents the study's experimental design and argues for the specific choice of method and experimental setting to investigate relationships of perceived similarity, cross-linguistic influence and universal preferences between L2 vowel categories. It presents the structure of the stimuli and response categories and the sample of participants. Methods of data interpretation will be introduced in chapter 7.

6.1 Preliminary considerations

The major aim of this study is to gain insights into the learners' long-term representation of L2 German vowel categories as determined by their L1 vowel inventory, their knowledge of the L2 system and their construction of contrasts and similarities between L2 categories that are influenced by language-specific and language-universal factors (see section 1.1.4 and 1.2) as well as by learner-related factors (see chapter 8) such as quantity and quality of L2 input and the individual construction of a learner's interlanguage.

The study is based on the assumption that L2 vowel categories are specified in long-term memory representations. These mental representations of L2 categories form part of a learner's interlanguage developing in the course of the L2 acquisition process. Interlanguages (see section 1.1.1 and 1.3) are conceived as emerging linguistic systems developed by language learners approximating eventually the target language system influenced by the system of their native language L1 (and other previously or subsequently acquired languages).

The basic assumptions of the current work as discussed in the previous sections regard the following issues:

(1) The correct categorization of vowels in L2 is considered to be generally more difficult for L2 learners than for German native speakers.

(2) The perception of German vowels in L2 will vary systematically according to the listeners' L1.

(3) The task of learning a rich and complex phoneme inventory as the vowel system of German may be fundamentally different for learners whose L1 vowel inventory is smaller (as e.g. Polish or SerBoCroatian) than for learners with a larger and complex L1 vowel inventory (as e.g. English or Hungarian).

(4) Some German vowels are considered to be more difficult than others in L2 perception and will therefore be more susceptible to perceptual substitutions in an identification task

irrespective of the listeners' native language. These difficulties cannot be explained by a mere contrastive description of differences between the learners' L1 and L2 German.

(5) The status of a given L2 category as more or less *marked* is expected to partly explain its relative difficulty in L2 perception (see section 1.1.3, 1.2 and 1.3). The correct identification or categorization of German vowels that are generally referred to as unmarked, more preferred and more common in the languages of the world, may be significantly better than the identification of less preferred, more marked sounds.

(6) *Inter-lingual similarity between L1 and L2* categories has to be distinguished from *intra-lingual similarities* between members of the set of phonemic categories of the target language L2. Intra-lingual similarities between L2 categories determine their probability to undergo perceptual substitutions.

(7) The *correct identification* of non-native vowel phonemes is more difficult than their perceptual discrimination from similar L1 qualities and from other L2 categories.

(8) For *auditory discrimination*, listeners may use sub-phonemic phonetic differences that are perceived sub-consciously to differentiate stimuli in a discrimination task. However, when the attentive focus lies on the phonological function of a speech segment these sub-phonemic details are not consciously available to the listeners. The relevance of particular acoustic cues as phonemically functional or sub-phonemic and the range of phonetically different realizations of items (tokens) for the same category (types) vary language-specifically.

(9) The *wrong identification* of a given vowel stimulus as belonging to another category than the intended is regarded as evidence for *confusion between L2 categories* due to insufficient perceptual discrimination and instable mental representations for L2 contrasts. Wrong identifications of input stimuli are considered as caused by *perceptual substitution* processes.

(10) In a vowel identification task, more experienced learners are expected to perform better than less experienced learners, i.e. a learning effect is expected in L2 perception. In initial stages of L2 learning, L2 vowels may be "assimilated" to similar L1 categories by learners with less experience in L2 (see section 2.4). This effect is evident in so-called perceptual assimilation tasks, where only L1 categories are offered as response options (see section 2.3.5.3). However, by increased experience in L2, perceptual assimilation to L1 categories is expected to decrease with more experience in the target language. Therefore, inter-category *confusion* of L2 categories *within the L2 system* is observed even with more advanced learners of German. Therefore, in addition to inter-lingual similarities

between elements of L1 and L2, relationships of *intra-lingual category similarity* as perceived and established by the L2 learners are a crucial issue for an understanding of phonological acquisition processes in L2.

(11) The occurrence of *perceptual confusion* between L2 categories provides valuable evidence for *perceived similarities* between vowel categories *within* the L2 system. Evidence for inter-category confusion is found in oral and written production but will be systematically collected here in a perceptual identification task. Data collected in an identification task reveal category confusion and relations of similarity between categories of the target language German.

The study's main interest are processes of L2 vowel categorization and the acquisition of L2-specific vowel contrasts. The main focus lies on inter-category relationships of *perceived similarities* between L2 categories, i.e. *intra-lingual similiarities*, rather than similarities between L1 categories and vowel sounds of the target language German.

To gain evidence for similarity of linguistic items, different experimental methods have been used such as rating of objects along a similarity/dissimilarity scale, same/different judgments in discrimination tasks, pair ranking, the anchor point method, and last but not least confusion data from identification tasks (cf. Backhaus 2011b: 223ff; for discussion of experimental designs for measuring similarity, see also section 3.2.3).

In the current study, neither a perceptual assimilation task (see section 2.3.5.3), where listeners identify L2 input stimuli in terms of L1 categories and rate the "goodness of fit", nor a discrimination task (see 2.3.5.1), where listeners have to judge members of a stimulus pair as "same" or "different", seem appropriate for the research question of interest. The data and analyses presented here do not investigate the learners' sensitivity to specific sub-phonemic cues in the input signal, but rather on the L2 learners' conceptions of L2 categories and the German vowel system as a whole. Influences of sub-phonemic phonetic details could be better investigated by carrying out experiments with synthetically manipulated stimuli where variation of particular sub-phonemic details can be controlled for.

Therefore, an identification task with natural vowel stimuli, where listeners have to identify German vowel stimuli as belonging to German response categories, will provide perceptual confusion data that give insights into *intra*-lingual relationships of contrast and similarity between L2 German vowel categories as perceived by learners from different language backgrounds. The following section will describe the method and test procedure and the participants of the study.

6.2 Method and data collection

6.2.1 Identification task

The experimental data of the present study were collected in a category identification task. A closed-set identification task (cf. e.g. Iverson & Evans 2007, for a similar test design) was chosen as most appropriate to investigate the acquisition of the complex vowel inventory of German in L2 acquisition.

While most previous empirical studies on L2 vowel perception have focused only on a sub-set of L2 vowel contrasts, the major goal of the present study is to describe inter-category relations of similarity and contrast within the complete system of German vowel categories and the relative perceptual difficulty and category confusion for single categories. Therefore, the complete set of 15 German vowel categories, including the controversially discussed /ɛ:/, served as input *stimuli* in the study. /ɛ:/ was included in the stimuli set as well as in the response set, because it is realized at least in some regional varieties of German, because there is an orthographic representation <ä> and because it is commonly taught in L2 German to be a vowel phoneme of German.

The set of *response* options consisted of vowel categories from L2. All 15 German vowel categories and 3 diphthongs were offered as response options to the subjects.

Multiple natural tokens of German vowels were presented to the subjects in nonsense-words. These nonsense-words were embedded in a constant carrier-sentence. Each of the 15 vowel phonemes occurred 18 times in changing pre- and post-vocalic context in the test material. In total, 270 German vowel tokens were presented to the subjects.

The participants were asked to identify the vowel contained in each of the nonsense-words and to mark it on a response form containing a table with all German vowel phonemes and three diphthongs (cf. Table 6.1).

The stimuli consisted of naturally spoken speech material. 15 German vowel phonemes were presented to the subjects in a large variety of different contexts (pre-vocalic, post-vocalic and prosodic variation). Thus, the input material contained a considerable range of natural phonetic variation as given in everyday speech. The use of natural speech input in this study was motivated by several characteristics of the language learners who were tested: The subjects were volunteers. Most participants were tested in their spare time after their language courses or in private places. Very few of them had experience or affinity to linguistic experiments. A test situation in a laboratory with synthetic stimuli would have discouraged many of them to undergo the test procedure.

This type of identification task has been selected for the intended investigation because of several advantages:

(1) The results obtained in the test procedure can be clearly scored as “wrong” or “correct”, i.e. *quantitative results* are obtained.

(2) Quantitative results enable a *qualitative analysis*. The relative *difficulty* of vowel stimuli can be described quantitatively and qualitatively. Results from identification tasks allow general conclusions about the susceptibility of certain sounds for wrong identification. L2-sounds which are frequently identified incorrectly by the subjects are considered to be more difficult than others.

(3) Moreover, the output of identification tasks provides direct evidence for *perceptual substitutions*. The subjects have to decide which particular response option they think is correct. Thus, in case of wrong identification, the data directly reveal perceptual substitutions.

(4) The results do not only reveal which vowels sounds are most prone for perceptual substitutions but show which response categories are preferred substitutes for difficult vowel items. Those vowel categories which are perceived by the learner as most “similar” or even “equal” to the input vowel are expected to be chosen as preferred response category. Thus, the method provides evidence for the *perceived similarity* between certain categories without asking explicitly for similarity judgements.

(5) An identification task with response categories from L2 German instead of L1 categories can be applied with German learners from different languages. Behavioural data from speakers with different language background for the same set of stimuli allow a *cross-linguistic comparison* of perceptual patterns on a language-specific, a segmental, a context-sensitive and a person-related level.

(6) The correct identification of vowels may vary as a function of phonetic context. Repetitions of a vowel in a different pre- and post-vocalic context allow the control of *contextual effects*. The repetition of several instances of the same sound category guarantees *stability* and *reliability* in the test results (see below). The presentation of multiple tokens of each vowel category in different phonetic and prosodic context provides more reliable evidence for the identification of L2 vowels and guarantees *robustness* and within- and across-learner *consistency*.

(7) The repetition of the test procedure with a *control group* of German native speakers guarantees the *reliability* of the test design.

(8) Moreover, results from identification tasks can be *related to segmental production* data to explain problems in oral and written production caused by misperception of L2 vowels.

The test results are subject to a detailed quantitative and qualitative data analysis as described in chapter 7. Instances of wrong or correct perception and categorization are tabulated in a *confusion matrix*, revealing areas of *difficulty* and confusion among categories as well as *preferences* for specific categories as response options. Confusion matrices are converted into similarity matrices and *distance matrices*. The distance matrix will be used for *Multidimensional Scaling*, revealing the perceived *similarity* or dissimilarity of vowel categories by L2 learners with different L1s.

6.2.2 Response categories

An important question of the test design is the choice of appropriate response labels presented to the subjects. The response labels should be unambiguous for L2 learners even if they are not fully familiar with phoneme categories and spelling conventions of the target language or with phonetic symbols (Flege 2003a).

It is necessary to be aware of the fact that the representation of vowels in German orthography is quite inconsistent and that several many-to-one and one-to-many relations exist between phonemes and graphemes in German. This fact renders the acquisition of German phoneme categories and orthographic conventions as well as the choice of response labels more difficult.

In the present study, the input stimuli as well as the output response categories consisted of 15 German vowel phonemes (plus three diphthongs in the output). The response alternatives consist of the total set of 15 German full vowels /ɑ, ɑ:, ɛ, e:, ɪ, i:, ɔ, o:, ʊ, u:, œ, ø:, ʏ, y:, ɛ:/ and of three diphthongs /ai, oi, au/.

Although [ɛ:] is not actively used by German native speakers in many regional varieties, it was included in the input material and presented as a response alternative because even listeners who do not themselves use this vowel may nevertheless have a perceptual category for it (cf. Sendlmeier 1981; Strange & Bohn 1998: 503).

The decision to add diphthongs as response alternatives was motivated by results from previous studies and experience with learners of some of the languages involved, e.g. Polish, because diphthongs showed to be possible candidates for perceptual substitutions for some vowels with learners of some languages (cf. Hentschel 1986; Kerschhofer 1995).

The set of response categories and their arrangement on the response sheet remained constant throughout the test procedure. The response options were presented in their common orthographic representation, long vowels were additionally indicated as "lang" and set in a broader field (see Table 6.1).

a	**a**$_{\textbf{lang}}$	**e**	**e**$_{\textbf{lang}}$	**i**	**i**$_{\textbf{lang}}$	**o**	**o**$_{\textbf{lang}}$	**u**	**u**$_{\textbf{lang}}$	**ö**	**ö**$_{\textbf{lang}}$	**ü**	**ü**$_{\textbf{lang}}$	**ä**$_{\textbf{lang}}$	**ei**	**eu**	**au**

Table 6.1: Response labels in the identification task

Unlike many previous studies which considered only a subset of a vowel system and presented only a small number of L2 categories to the participants, the aim of the present study is a survey of German vowel perception by learners from different native languages. The study intends to investigate the acquisition of the German vowel inventory as a whole and therefore includes the complete set of German vowels in the input stimuli as well as in the set of response options. Only categories from L2 German were presented in the input as well as in the output. Native categories were not involved in the test procedure. This choice was motivated by two main arguments:

While many studies using L2 identification tasks (e.g. Hentschel 1986; Levy/Strange 2008; Levy 2009) present L1 categories as response items to the participants and ask which L1 category is the most similar to the L2 input (*perceptual assimilation task*), the response options in this study consisted only of L2 German phoneme categories. Perceptual assimilation tasks may be appropriate to reflect the situation of naïve listeners or learners in initial stages of learning and may allow conclusions on production data. However, advanced learners generally know about the phonemic inventory of the target language and are not expected to look for direct correspondences with L1 phonemes. Therefore, no L1 items were presented to the participants in this study and the subjects were informed that the input items only contain vowels of the target language German. A further advantage of this decision is that the same test material could be used with learners from any language, allowing direct comparability between subjects with different L1s.

For each vowel category, a contrastive analysis of a given L1 and L2 German would predict a specific set of possible candidates for perceptual substitutions, as the number and type of categories that could be possible candidates for substitution may vary by the listeners' L1. Depending on the actual L2 input category and the native language considered, the response options would have to change with every test item, according to the native language and the substitutions expected. This would cause significant complications in the test procedure.

Therefore, the set of response options was the same for all L1 sub-samples. The complete set of 15 German vowel categories and three diphthongs was presented to the subjects.

6.2.3 Test procedure

Task and method were chosen to provide experimental evidence for false identification and confusion among vowel categories of L2 German. The test neither aimed to gain evidence for perceptual assimilation of L2 sounds with L1 categories (as e.g. Hentschel 1986; Strange 2007; Frieda & Nozawa 2007: 84ff.) nor to test the learners' sensitivity to variation in specific phonetic details. The main intention of the test was to collect evidence for how L2 learners deal with the numerous vowel categories of German in perception.

The participants were informed about the general aim of the perception test when asked to participate. They were told that the aim of the experiment was to describe how learners perceive the numerous and partly difficult vowel sounds of German and to collect data on the relative difficulty of German vowel for learners from different L1 backgrounds.

The subjects were tested individually or in small groups, either in homogenous or in language-mixed groups. The test was carried out in a quiet room, in most cases the tests were performed after class in one of the classrooms where the participants usually had their German classes.

Participants were instructed to listen to an input sentence and to identify the vowel in the nonsense-word by choosing one of the German vowel labels in the answering form. The subjects responded by checking one of the response options on the answer form. No repeated listening option was offered.

Before beginning the experiment, the subjects were familiarized with the task by listening to the response categories listed on the response leaflet and to three familiarization tokens. Each vowel category was presented in isolation with reference to the presentation on the answer form before the test started. Participants were then listening to three familiarization tokens in a sentence structure similar to the test sentences. Responses from the familiarization were not used for data analysis. The experiment was carried out in one standard rate for all participants, i.e. the listeners had no influence on the speed of stimulus presentation.

6.3 Stimuli

Each of the 15 German vowels phonemes (/a, a:, ɛ, e:, ɪ, i:, ɔ, o:, ʊ, u:, œ, ø:, ʏ, y:, ɛ:/) were embedded in logatomes which were embedded in a constant carrier-sentence. Each vowel occurred 18 times in different consonantal context.

A considerable amount of contextual variation was included in the data. Each German vowel was presented in 18 repetitions to the participants, i.e. every vowel phoneme occurred 18 times in varying pre- and post-vocalic context and in changing prosodic structures. In total, the test material consisted of 270 test items. The input contained no diphthongs. However, diphthongs were presented as response options.

The stimuli were read by a male native speaker of Standard Austrian German producing each of the test items containing one of the 15 German vowel phonemes (/a, a:, ɛ, e:, ɪ, i:, ɔ, o:, ʊ, u:, œ, ø:, ʏ, y:, ɛ:/), each in 18 different consonantal contexts. The material was digitally recorded in a sound studio at a sampling rate of 44.1 kHz, 24 bit. The data served for acoustical analysis and for presentation to the listeners in the experiment.

Test items (carrier sentence with embedded word) were presented to the listeners with a constant inter-stimuli-interval of 2.5 sec[62]. This time span was used by the subjects to mark the answer on the response sheet.

6.3.1 Item structure

The test material was divided in two parts, experiment 1 and experiment 2. The carrier-sentence and the non-word differed in structure in experiment 1 and experiment 2: The structure of the logatomes in experiment 1 was /pVC/ and /CVtə/ in experiment 2. The carrier sentence consisted only of words with unstressed qualities and diphthongs, thus no other stressed monophthong occurred in the sentence.

Experiment 1 consisted of 120 test items with varying *post-vocalic* context /pVC/ (see Table 6.2).

Experiment 1	labial	alveolar
plosive	pVp	pVt
fricative	pVf	pVs

Table 6.2: Item structure in experiment 1 (post-vocalic and intonational variation)

The post-vocalic consonants in experiment 1 were labial or alveolar plosives or fricatives. Velar consonants were not included because of the allophonic distribution of fricatives /x/ and /ç/. The carrier-sentence in experiment 1 was produced either with falling or with rising intonation: "*Er meint heute pa:t.*" (falling) and *"Er meint heute pa:t?"* (rising). Thus, each

[62] The length of inter-stimuli interval (ISI) determines perceptual processing: While an ISI of 500 ms listeners induces phonetic processing, an ISI of 1500 ms allows phonological processing (Werker & Logan 1985).

vowel type was presented twice in a specific post-vocalic context, once with rising and once with falling intonation (4 contexts x 2 intonation contours x 18 categories = 120).

Experiment 2 consisted of 150 test items in varying *pre-vocalic* contexts (see Table 6.3). In experiment 2, the stimulus word was the first word in the carrier-sentence which was produced with falling intonation: “*da:te bedeutet auch nichts.*“ The pre-vocalic consonant in the stimulus word varied in voice, place and manner of articulation.

Experiment 2	bilabial	labiodental	alveolar	palatal	velar
plosive lenis	b		d		g
plosive fortis	p		t		k
fricative voiced		v	z	j	
fricative unvoiced		f			

Table 6.3: Item structure in experiment 2 (pre-vocalic variation of context)

6.3.2 Speaker

The speaker was a 29-year-old male who was chosen as a non-dialectal speaker of Standard Austrian German or more generally of Standard Southern German. He had lived in Austria (Vienna), Germany (München), and in a few non-German speaking countries during his life. His father is Austrian, his mother Swiss. Typical characteristics of Standard Austrian German are observed in his speech such as a slight tendency towards monophthongization of diphthongs in the carrier-sentence in “auch” and “meint”, final /ə/ is realized as [ɛ] as is usual in Vienna. /ə/ in the pretonic syllable “be-“ is realized as [e]. However, the /z/ in word-initial pre-vocalic position is realized consistently as [z] and voiceless /s/ is never used word-initially due to the instructions he received and the formal speaking style (reading logatomes in a lab situation).

6.3.3 Acoustic analysis

The speech material was digitalized at 44.1 kHz, 32 Bit. Temporal and spectral measurements of the acoustic input were performed using the workstation STx developed for acoustic analysis at the Acoustic Research Institute of the Austrian Academy of Science (http://www.kfs.oeaw.ac.at).

The vowels were segmented manually. The first positive zero crossing or the end of aperiodicity was determined as starting point of the vowel phase. The last full period similar to the preceding ones or the start of aperiodicity indicated the end of the vowel. Formant frequency values for F1, F2 and F3 were extracted using *Linear Predictive Coding* algorithm

(LPC, Markel & Gray 1976). A 46 ms long gliding Hanning window was applied at a sampling frequency of 44.1 kHz with an overlap of 95%, using 56+4 coefficients providing sufficient measurement points even for fast formant transitions and short signal segments.
Depending on the respective duration of the vowel segments, this method renders about 20-150 measurements per vowel. If necessary, formant frequency traces were corrected or aligned to the correct order. In case of interruptions or missing parts, the formant values were edited with STx. The data obtained were exported for further quantitative and qualitative analysis.
For all vowel items, the *formant frequency* contour of the entire nucleus was analysed. This is considered to describe the spectro-temporal changes throughout the nucleus best and is especially useful for short vowels, where no steady state portion may be found. Applying this method, transitions are included in the mean values. This method accounts for the full information content of the vowel continuum and is more reliable than measurements of a (quasi-)steady state portion of the vowel continuum, especially with short vowels.
Additionally, *duration* measurements and an analysis of F0 were performed. Vocalic duration was defined as the interval between the first positive zero crossing or the end of aperiodicity and the last full period similar to the preceding ones or the start of aperiodicity indicating the end of the vowel.
Fundamental frequency measurements were performed with the autocorrelation method SIFT (*Simplified Inverse Filter Tracking*, cf. Markel & Gray 1976: 276) which was synchronized with the formant frequency measurements. For a few very short vowels another method of autocorrelation (FOAC - F0 detection using autocorrelation, cf. Boersma 1993) seemed to be more appropriate.

6.4 Sample structure

6.4.1 Participants

The sample described here consists of 173 adult learners of German who participated in the test as unpaid volunteers. The sample consists of 37% male and 63% female participants. All of them were reporting normal hearing, speech, language and reading abilities in their native language as well as in the target language German. The subjects' mean age was 24.8 years (8.2 standard deviation), the youngest participant was 15, the oldest 60. Most subjects were undergraduate students or academics.

The sample provides 46,710 possible responses (270 stimuli x 173 participants). For each vowel category 3114 responses were collected. From the total of 46,710 responses, those instances, where subjects did not give a response for a particular stimulus presented or gave more than one response, were excluded from the final analysis of the data. The analysis presented here is based on 45,794 responses by 173 subjects.

6.4.2 Language background (L1)

The choice of the listeners' native languages was mainly motivated by pedagogical considerations. Results of the study should provide insights into the acquisition process and learning problems of learners of German as a foreign language. One practical aim of this study is to deduce pedagogical implications for teachers of German in Austria for the most common immigrant languages by describing L2 learners' difficulties in vowel perception and their possible consequences on acquisition processes in oral and written language production in L2 German (cf. Kerschhofer-Puhalo 2009, 2010b, 2012, in press a).
Therefore, the sample should represent the most common languages of immigrants and their descendants in Austria. The population in Austrian primary schools was considered to reflect the current situation and future development in Austria. Therefore, the ten most common languages at Viennese primary schools were chosen for this investigation from the ranking in the "Wiener Schulpflichtmatrik 2006" (Stadtschulrat Wien 2007).

L1	subjects	beginners	advanced learners	very experienced learners	Rank at Viennese primary schools in 2006
Albanian	12	1	9	2	5
Arabic	10	7	2	1	3
English	13	6	7	0	7
Farsi	4	3	1	0	9
Hungarian	26	11	11	4	8
Mandarin	8	2	6	0	10
Polish	31	22	9	0	4
Romanian	12	2	9	1	6
SerBoCroatian	33	17	14	2	1
Turkish	24	12	10	2	2
all L2 listeners	173	83	78	12	

Table 6.4: Native languages, position in the ranking list and number of participants (beginners, advanced and very experienced L2 learners) per language

Ten L1-defined sub-samples were formed according to the participants' first languages Albanian, Arabic, English, Farsi, Hungarian, Mandarin, Polish, Romanian, SerBoCroatian, and Turkish. Additionally, a control group of 18 L1 German adults was tested. As participants were volunteers, the number of participants varied considerably across language subgroups (see Table 6.4).

Participants were recruited and tested in Vienna, Budapest, Rijeka and Wrocław. The Polish sample was tested at the University of Wrocław and consisted of 9 students of German in their second year of German studies (advanced learners) and of 22 students from other subjects (beginners). The Hungarian sample was partly recruited at the German department of ELTE University in Budapest (11 advanced learners studying German in the second year). 11 beginners were tested in a private language school in Budapest and 4 subjects were tested in Vienna (two Hungarian natives were born in Transylvania/Romania). 20 native speakers of Croatian were recruited at the German department of the Pedagogical Faculty Rijeka (14 advanced learners) and the Juridical Faculty (6 beginners) in Rijeka (Croatia). The second part of the SerBoCroatian sample was tested in Vienna. All other subjects – beginners and advanced learners, all not below level A2 of the Common European Reference Frame for Languages (Council of Europe 2001)) – were recruited in Vienna at the Wiener Internationale Hochschulkurse and the Vorstudienlehrgang der Wiener Universitäten or via personal contacts.

The total sample consisted of 173 learners of German at different levels of language proficiency (beginners, advanced and very experienced learners). The ten L1-defined sub-samples were heterogeneous not only with respect to their first language but differed also in other variables such as level of proficiency, age, language experience, knowledge of other foreign languages, quality and quantity of L2 input, length of residence in a German speaking country and others. Chapter 8 will discuss the influence of these learner-related variables on individual differences in the overall success in the identification task.

6.4.3 Questionnaire

It is hypothesized that the way L2 sounds are perceived is influenced by a bundle of factors:

(1) factors related to the *experimental setting* (see section 2.3 and 6.5),
(2) *language-related* factors conditioned by the relation of L1 and L2 sound categories (e.g. the relation between L1 and L2, perceived overal language distance (cf. Bradlow et al. 2007), similarity of particular L1 and L2 categories etc.),
(3) factors related to *vowel-specific* characteristics (see chapter 4, 5 and 11), and
(4) factors related to *individual variables* (age, exposure to L2, current use of L1 and L2, general experience in language learning etc., see chapter 8).

A questionnaire was designed in order to elicit information on subject-related variables that account for the substantial amount of intra-group variability among participants within the L1-defined sub-samples.

Before performing the identification task, the participants were asked to complete the questionnaire. Based on the questions and answers in the questionnaire, SPSS variables were formed and analyzed (see chapter 8). The questionnaire allows to compose homogeneous subject groups and to filter subgroups with respect to individual factors such as

a) their first language (language, country of birth, regional variety),
b) their language experience in L2 (so called "age of learning" (AOL), i.e. the age at which L2 German learning began),
c) level of proficiency,
d) their general experience in language learning (bilingualism, number of foreign languages acquired),
e) length and kind of language instruction in German (school, language course, university, private contacts, …),
f) length of residence (LOR) in a German speaking country and regional varieties of German the learners had been exposed to,
g) formal teaching in language, phonetics and pronunciation training,
h) age, and
i) sex.

A detailed analysis of the influence of person-related variables on the learners' performance in the experiments will be presented in chapter 8.

6.4.4 Native German control group

The experiment was carried out with a control group of 18 native speakers of German from different regions of Austria and Germany (13 persons were born in Vienna, two in Vorarlberg/Austria, one in Upper Austria, one in Thüringen/Germany, one in Oberfranken/Germany). The native listeners' overall rate of identification accuracy varied between 0 and 9.6% wrong responses (mean 3.2%). The id_scores varied category-specifically between 93% and 99%. Higher id_wrongs were observed only for /ɛ:/ (79% wrong responses for /ɛ:/). Interestingly, many native subjects of Southern German varieties tended to substitute /ɛ:/ by /e:/, even some persons having linguistic background knowledge, whereas others – linguistically educated persons as well as persons with no linguistic knowledge but from regions where <ä> is realized /ɛ:/ (e.g. Vorarlberg (most Western part of

Austria, Alemanic dialect) and some parts of Germany, e.g. Thüringen) were sensitive for the contrast /ɛ:/-/e:/. All native German participants correctly recognized that there were no instances of diphthongs in the test input. A detailed description of the results from the L1 control group is presented in section 10.2.

The results from the control group show that all input stimuli are highly identifiable by native speakers of different regional varieties of German. The identification error rate lies clearly below the 15% mark that was established in other studies with native speakers for listening experiments (e.g. Peterson & Barney 1952; Hillenbrand et al. 1995). The test items can therefore be considered to be good exemplars for German native vowel categories.

6.5 Task evaluation: objectivity, reliability, and validity

Objectivity, reliability and validity are the three most important quality criteria for empirical studies. Experimental results should be independent of the investigator (*objectivity*) and should be reproducible in repetition (*reliability*), e.g. in a test-retest condition. Reliability is a precondition for *validity*. *Validity* refers to the best available approximation of the falsity or truth of propositions (Brewer 2002: 3). *Construct validity* refers to the extent to which the construct, i.e. the object we intend to measure, can be operationalized and measured.

Internal validity of conclusions from empirical studies is given if the dependent variable can be directly explained by the independent variable (e.g. by the vowel presented or by the learner's L1) and if alternative explanations for the results found can be excluded. A high degree of internal validity can be achieved by a high degree of control over experimental conditions, e.g. with tests in a laboratory, the use of controlled stimuli (e.g. synthetic vowel stimuli with systematic changes of a specific phonetic feature), or a homogeneous sample (e.g. level of language proficiency or length of residence etc.).

External validity refers to the question of whether an effect (and its underlying processes) as observed in one research setting could be obtained in other settings, with different participants or different research procedures (Brewer 2002: 10). In other words, it refers to the question whether the results obtained in one setting with a specific sample can be generalized to other samples, e.g. to the target population (from which the sample has been drawn) as well as to other populations or conditions. Inferences about cause-effect relationships are considered to possess external validity if they can be generalized from a given setting to other populations and conditions. The generalizability of the experimental results is limited by all situational specifics (e.g. time, location, noise, ...) and, last but not least, by the sample itself, i.e. by the *participants* of the study. Different aspects of external validity and the question about where, when, and to whom the results of a research are to be generalized can be examined; robustness, ecological validity and relevance are the most important aspects here (Brewer 2000: 10).

Relevance refers to the question whether the findings are related to phenomena that actually occur in real life and whether research results can be transmitted to application (Brewer 2000: 12).

Robustness refers to the replicability of results, i.e. to whether particular results are reproducible in different settings, with different subjects or in different contexts. Robust results would be replicable despite of significant variation in the test conditions (Brewer 2002: 10). Brewer (2002: 11) distinguishes generalizability across multiple populations and settings from generalizability to a particular population. Robust phenomena may hold up for the population at large (e.g. a language-universal preference for specific vowel qualities), but may not be obtained for a specific sub-population (e.g. individual learners or learners with a particular native language may show different preferences). The robustness of an effect can be tested empirically by systematic variation of potentially moderating factors.

A high degree of external *validity* can be achieved by a high number of participants and maximal heterogeneity of the sample (e.g. age, sex, L1, length of L2 learning, etc) as well as by the variation of research settings. The higher the number of participants and the degree of heterogeneity, the higher the degree of external validity, i.e. of generalizability of the results obtained.

In a certain sense, internal and external validity seem to be in contradiction to each other. An increase in the degree of internal validity (e.g. by the use of controlled minimally varying stimuli or a heterogeneous sample) limits the generalizability and external validity of the results.

Ecological validity refers to methods, materials and research settings approximating situations in everyday life and conditions that are typical for the population at large (Brewer 2000: 12). Thus, results obtained from perception experiments in a laboratory setting with university students studying German might not be reproducible in a real-life speech situation with L2 learners who had acquired the target language in a naturalistic setting with limited or even no language instruction. Especially for the study of L2 acquisition with regard to possible pedagogical implications of the results, ecological validity is an important criteria. To increase the ecological validity of an empirical study, a "*representative design*" (Brunswik 1956) with probabilistic samplings of subjects and situations should be chosen. Brewer (2002: 12) distinguishes *mundane realism* (referring to the extent to which a research setting resembles events in everyday life) and *psychological realism*, which refers to the psychological processes occurring in an experiment and their resemblance to psychological processes in everyday life. However, while internal and external validity are necessary criteria for the overall validity of a study, ecological validity is not (Shadish et al. 2002).

Another problem in experimental research is *reactivity*: Subjects usually are aware of the fact that they are observed or tested in a research study. Reactivity effects like "demand characteristics", "experimenter expectancies", or "evaluation apprehension" derive from the fact that participants are seeking cues in the research setting to understand what they are expected to do (Brewer 2002: 8). The independent variable may contain cues which influence the participants' guesses about what would be the proper response in the research setting (e.g. by variation in the experimental instruction). By this, the *construct validity* of the independent variable could be compromised.

The way a study is designed and conducted is largely connected with the conclusions that can be drawn from the results. Thus, "*we cannot speak of the validity or invalidity of research per se*" (Brewer 2000: 3), but have to evaluate the conclusions of research with respect to the *purposes the study* has been undertaken.

The purpose of the present study is (1) to test assumptions and predictions of contrastive linguistics, second language research and markedness theory on a sample of L2 learners of German from different languages, and (2) to derive pedagogical implications from the results obtained for an improvement of teaching practice in oral and written L2 speech production.

Therefore, in the present study, external validity (at the cost of internal validity), and here especially ecological validity and relevance for the teaching practice have been of primary interest. While a test design with synthetic vowel stimuli with controlled variation of specific acoustic features and the use of non-orthographic response labels (e.g. pictures representing key-words) could guarantee maximal internal validity, a big disadvantage of this method would be that the test design and the stimuli are very far from perceptual processes in real life. A more "realistic" test design was chosen here in which participants listened in a "natural setting" (the test was mostly performed in rooms where the participants usually received their German classes) to natural speech stimuli and were asked to relate them to semi-orthographic representations they are familiar with.

The repetition of the test procedure with a *control group* of 18 German native speakers guarantees the *validity* and *reliability* of the test design.

A natural variance of distribution of vowel realizations is given in multiple repetitions of the vowel stimuli in varying contexts (15 vowels *x* 18 contexts). The repetition of multiple tokens of each vowel category in varying phonetic and prosodic context provides more reliable evidence for the identification of L2 vowels and guarantees within- and across-learner *consistency* and *stability* in the results. Since the correct identification of vowels may vary as a function of phonetic context, repetitions of the same vowel in varying pre- and post-vocalic context allow for the control of *contextual effects*.

A critical issue is *construct validity*, i.e. the extent to which the construct we intend to measure can be operationalized and measured. The underlying assumption of identification and categorization tasks is that listeners have mental categories for speech sounds and that their behaviour in experimental tasks allows direct conclusions about the mental representation of speech sounds (Labov 1995: 349ff). However, it is necessary to be aware of the fact that such tasks demand a conscious reaction of the participants and that the listeners' *response behaviour* in the task – just as in any other experimental design that tests perception by active responses of the participants – is not only conditioned by the input stimuli but also by previous experiences and current assumptions, expectations and hypotheses that may determine the listeners' current conceptualization of perceptual categories in L1 and L2. This effect may limit *internal validity* to a certain degree, since the listeners' behaviour in the identification experiment cannot be explained by one single independent variable. However, for the reasons explained above, a set of natural stimuli instead of a highly controlled set of synthetic stimuli were used in the study.

To achieve a high degree of *external validity*, a high number of non-native listeners (173 persons) from 10 different native languages and different proficiency in L2 German (beginners, advanced, and very experienced learners) were tested.

Conceptual difficulties with the notion "native speaker" and its relevance for "second language research" should be briefly mentioned here. Though, the notion of the "native speaker" and his/her identity is intuitively easy to grasp from a sociolinguistic view, the psycholinguistic dimension of this notion is more difficult to define. In linguistic literature, two competing notions are common: (1) the "*ideal speaker-hearer, in a completely homogeneous speech community, who knows its language perfectly ...*" (Chomsky 1965: 3) and the "*real native speaker, i.e. an observable language user*" (Sharwood Smith 2011). However, both notions show to be inherently vague, when it comes to decide which criteria to select in order to define a specific group of "native speakers" of a particular language and their "speech community". This issue becomes particularly relevant when investigating so-called pluricentric languages like German, English or Arabic, for which more than one standard variety exists. For the sample of the present study, this issue will be discussed for those languages where it appears relevant (for further discussion of the notion "native speaker", see also section 10.1 and 12.1.3.1).

6.6 Conclusions

This chapter described the design and experimental setting of the study. Underlying assumptions concerning methodological decisions and their empirical relevance were discussed in section 6.1. An identification task with natural non-manipulated input stimuli was chosen to gain evidence for L2 learners' perceptual difficulties with the German vowel system. A detailed description of the test procedure, the input and response material and the participants was provided in 6.2, 6.3 and 6.4. Advantages of the method and basic issues of task evaluation (objectivity, reliability and validity) were discussed in 6.5. The next chapter will describe the procedures of data analysis providing a discussion of quantitative and qualitative ways of data analysis and interpretation.

7 Data Analysis

This chapter provides a description of different methods of data analysis and interpretation of the experimental results obtained in the perceptual identification experiment. Detailed analyses of the data will follow in chapter 8, 9, 10 and 11, where the experimental results are presented focussing on different aspects: Chapter 8 will discuss *learner-specific factors* and their influence on the listeners' performance in the perception experiment. Chapter 9 will provide a cross-language overview which is followed in chapter 10 by a detailed *language-specific* presentation and analysis of the data for each of the L2 listeners' native languages. A *vowel-specific analysis* and a discussion of general patterns of difficulty, preferences, similarity and asymmetry for the major vowel classes is presented in chapter 11.

The present chapter will discuss the methods and *procedures of data analysis* and the goals, scope and limitations of data analysis and data interpretation.

7.1 Identification, categorization and descriptive statistics

The listeners' correct and wrong responses for each of the input stimuli were collected and tabulated in a response matrix. For all vowel phonemes tested, *identification scores* (*id_scores*) were calculated by tallying the number of correct answers and wrong identifications, i.e. the number of instances where a given input stimulus was identified as belonging to a specific response category. The number of instances of wrong or correct identification was divided by the number of valid responses and multiplied by 100. These values were summarized in a *confusion matrix*. Identification scores in the confusion matrix indicate the percentage of wrong categorizations (*id_wrong*) or correct identifications (*id_correct*) for each vowel category.

Identification scores were moreover calculated for each of the response labels (*id_V scores*), obtaining an identification score (id_V_i , $id\_V_{j, \ldots,}$) indicating the number/percentage of instances a given input quality was identified as belonging to a specific category, *e.g. id_a, id_y, id_œ* etc. *Id_V* scores will be discussed in detail for each of the listeners' native languages in chapter 10 and in a cross-language comparison in chapter 11.

To summarize, *id_wrong* scores represent the percentage of wrong identifications, id_correct indicated the percentage of instances a given input category is identified incorrectly and reflect the relative *difficulty* of a specific category; *id_V* scores indicate how often a specific category is selected as response option and reflect the listeners' *preferences* for a specific response category.

%	ɑ	ɑ:	ɛ:	ɛ	e:	ɪ	i:	ɔ	o:	ʊ	u:	œ	ø:	ʏ	y:	‹ei›	‹eu›	‹au›	wrong	corr
a	**76**	22	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	24	**76**
a:	15	**84**	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	16	**84**
ε:	0	1	**33**	9	49	0	1	0	0	0	0	0	0	0	0	5	0	0	67	**33**
ε	1	0	7	**74**	15	1	0	0	0	0	0	0	0	0	0	1	0	0	26	**74**
e:	0	0	15	7	**55**	4	12	0	0	0	0	0	1	0	1	5	0	0	45	**55**
ɪ	0	0	1	10	4	**68**	11	0	0	0	0	0	0	2	1	1	0	0	32	**68**
i:	0	0	0	1	3	14	**80**	0	0	0	0	0	0	0	1	1	0	0	20	**80**
ɔ	0	0	0	0	0	0	0	**74**	20	1	0	2	1	0	0	0	1	0	26	**74**
o:	0	0	0	0	0	0	0	8	**61**	4	12	2	6	1	2	0	1	1	39	**61**
ʊ	0	0	0	0	0	0	0	9	3	**66**	13	2	1	5	1	0	0	0	34	**66**
u:	0	0	0	0	0	0	0	0	2	15	**71**	1	1	4	6	0	0	0	29	**71**
œ	0	0	2	5	3	0	0	2	1	2	1	**59**	14	6	2	0	1	0	41	**59**
ø:	0	0	0	0	0	0	0	0	1	2	5	7	**45**	9	28	0	1	0	55	**45**
ʏ	0	0	0	0	0	1	0	0	0	8	3	9	4	**58**	15	0	1	0	42	**58**
y:	0	0	0	0	0	0	2	0	0	2	3	1	3	16	**71**	0	1	0	29	**71**
Σ	*6*	*7*	*4*	*7*	*9*	*6*	*7*	*6*	*6*	*7*	*7*	*6*	*5*	*7*	*9*	*.9*	*.5*	*.1*	*35*	***64***

Table 7.1: Confusion matrix in % for non-native listeners (N=173, 46,694 responses): Percentage of wrong and correct (in bold) identifications

The matrix in Table 7.1 presents the *id_wrong*, *id_correct* and *id_V* scores in % for all response categories for all 173 non-native participants. Input stimuli are presented in rows, selected response categories are presented in columns in semi-orthographic representation. The percentage of correct identifications is indicated in bold figures. Values below 0.5% are not indicated in the table.

7.2 Analyzing perceptual substitutions

7.2.1 Identification and confusion matrix

For data analysis, each of the listeners' wrong and correct responses is tabulated in an identification matrix, which is then transformed into a confusion matrix (see Table 7.1 in % and Table 7.2 in tokens). A *confusion matrix* tabulates presented stimuli (input) and selected categories (output responses) together in a matrix and indicates how often each of the input categories was identified correctly or incorrectly as one of the given response options. In a confusion each row in the matrix corresponds to one of the tested vowel categories (*presented input*), each column corresponds to one of the response labels available to the listeners (*selected output*). The confusion matrix can be represented in percent (as in Table 7.1) or in terms of individual tokens (as in Table 7.2).

Each vowel was presented 18 times to the participants. Table 7.2 shows the results in tokens for the control group of native German listeners and is to be read as follows: The participants in the German control group identified 4672 tokens of the 4860 input stimuli correctly. In total, for each vowel type 324 (18 x 18) answers could be obtained from native German listeners. In total, 188 tokens were identified incorrectly by German native listeners. In 316 cases, /ɑ/ was identified correctly as /ɑ/. In 8 cases /ɑ/-stimuli were wrongly categorized as /ɑ:/.

token	ɑ	ɑ:	ɛ:	ɛ	e:	ɪ	i:	ɔ	o:	ʊ	u:	œ	ø:	ʏ	y:	‹ei›	‹eu›	‹au›	wrong	correct
a		8																	8	**316**
a:	5				2														7	**317**
ɛ:				1	64										1	1			67	**257**
ɛ			8		8							2							18	**306**
e:			22	1			1												24	**300**
ɪ				2			3							2					7	**317**
i:						2		1											3	**321**
ɔ									7										7	**317**
o:								3			4								7	**317**
ʊ								1			5								6	**318**
u:										3									3	**321**
œ													6						6	**318**
ø:									1			3			2				6	**318**
ʏ						1				1		2	1		8				13	**311**
y:											1		2	3					6	**318**
Σ	*5*	*8*	*30*	*3*	*74*	*3*	*4*	*5*	*8*	*4*	*10*	*7*	*9*	*5*	*11*	*1*	*0*	*0*	*188*	***4672***

Table 7.2: Confusion Matrix in tokens for the native German control group (4860 responses)

In the *token*-oriented representation in Table 7.2, only the wrong responses are tabulated in the inner fields of the table, whereas in Table 7.1 the *percentage* of correct answers is included in the table's diagonal as well as in the right-most column (printed in bold).

Confusion matrices provide a possibility to describe the data quantitatively as well as qualitatively. The matrices can be produced for individual learners, for groups of learners (e.g. with the same native language), or for larger groups of subjects (e.g. for beginners vs. advanced learners), for certain vowels or for vowels in a specific context. In the present study, matrices will be produced and interpreted mainly for all listeners of a given language sub-group, based on the assumption of common language-specific patterns of perceptual confusion.

7.2.2 Difficulties

Instances of wrong categorizations are assumed to reveal incomplete and instable perceptual categories for the respective input category and insufficient differentiation from those categories by which the input was incorrectly substituted by the listeners in the identification task. *Id_wrong* scores reflect the susceptibility of a vowel category for perceptual substitutions and thus indicate its relative *difficulty*. The higher the *id_wrong* score, the more difficulties the learners are supposed to have in identifying the respective vowel category correctly. The lower the *id_wrong* score, the more stable a category is, i.e. the less perceptual confusion and perceptual substitution processes are observed. Thus, identification scores reflect the relative *difficulty* of a specific sound category.

The relative difficulty of a specific vowel category is assumed to depend (1) on the vowel category and its physical properties and (2) on the listeners' native language and the mental representation of corresponding categories in the L1, the L2 and the learners' interlanguage system.

A comparison of difficulties between vowel categories or sample sub-groups will be graphically represented with bar charts and classical box-and-whisker plots.

Bar charts are used to represent the mean values for *id_wrong/id_correct* scores to indicate the listeners' *difficulties* for each of the input categories (e.g. Figure 7.1 for the responses of all 173 non-native listeners). In bar charts, *id_scores* are usually presented in descending order, i.e. the left-most category is the most difficult category.

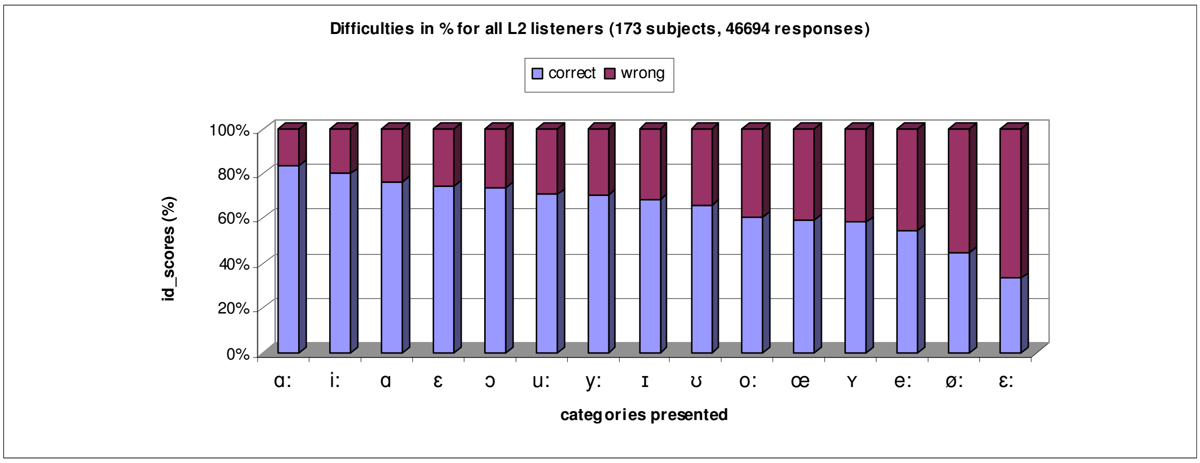

Figure 7.1: Barchart representing the identification scores (*id_wrong*) for each vowel category for all 173 non-native listeners

Boxplot diagrams are used to represent the range of variation of listeners' id_wrong scores. The box-and-whisker diagrams present considerably more information than a simple comparison of statistical means, visualising the envelope of the 50% central region, the median, i.e. the 50% percentile, the upper and the lower quartile and the smallest and largest value observed. The bottom and the top of each box represent the 25- and 75-percentile, the band near the middle box represents the median, and the ends of the whiskers represent the lowest and highest value within the 1.5 interquartile range of the lower/upper quartile. Observations outside this range are outliers plotted as dots or extrema plotted as stars.

The boxplot diagram in Figure 7.2 compares the *id_wrong* scores (in %) for all non-native listeners for each of the input categories separately and for all categories together (mean id_wrong in %) providing a comprehensive overview over the overall difficulty of a given vowel category (e.g. the id_wrong scores for /i:/ are considerably lower than for /ɛ:/ or /ø:/).

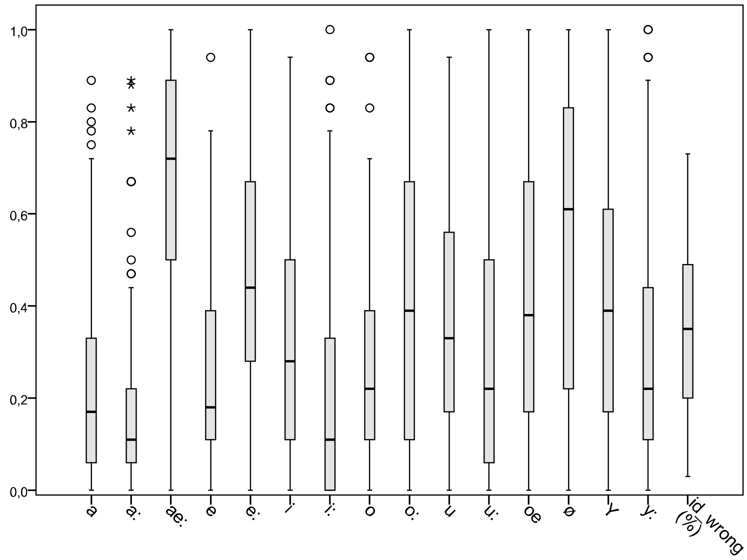

Figure 7.2: Boxplots representing the identification scores (*id_wrong)* of all 173 non-native listeners indicating the relative difficulty of a vowel category

To summarize, the listeners' *difficulties* with the correct identification of German vowels are evident in three types of data presentation: the confusion matrix for a qualitative and quantitative presentation and the barchart and boxplot diagrams for a quantitative representation. The percentage of wrong and correct identifications in a confusion matrix gives insights into two aspects of L2 vowel perception: (1) which *input categories* are identified correctly or incorrectly and therefore considered to be more easy or difficult to identify (*difficulties*), and (2) which *response categories* are more or less preferred in the *output*, i.e. the listeners' answers (*preferences*). Language-specific patterns of *category confusion* are evident in the listeners' wrong identifications reflecting the relative *difficulty* of a given vowel category as well as in the listeners' *preferences* for specific response categories.

7.2.3 Preferences

In a confusion matrix, the horizontal lines of the table give information about how often a given response label is selected for a specific input category and provides thus information about patterns of perceptual *preferences* for each of the presented response categories separately. In Table 7.1, the *id_Vs* for all 173 non-native participants from the 10 language subgroups are tabulated together. The confusion matrix shows wrong and correct responses in the same row. Correct answers are marked in bold and presented not only in the *id_correct* column but also in the diagonal of the table. The last line in Table 7.1 represents the *preferences* of the listeners for a specific response category, i.e. the *percentage* of cases where this category was selected as response option, irrespectively of whether the answer was wrong or correct. In Table 7.2, where correct answers are only presented in the left-most column, the last row tallies only the number of instances a specific category has been falsely selected by native listeners of German. The *id_V score*, i.e. the number of instances a specific category was selected as response option for a given input, tells us something about the participants' *preferences* for specific categories. The last line in the confusion matrix gives the total number of instances in which this specific category was selected as response option. We see here that the diphthongs <ei, eu, au> are the least preferred response options (since diphthongs did not occur in the input). <ä:> (/ɛ:/) is also a clearly dispreferred response category. Among the front rounded vowels the *ü*-qualities are more preferred than the *ö*-qualities. These overall preferences are visualized in *bar charts* for each of the learners' native languages (see Figure 7.3).

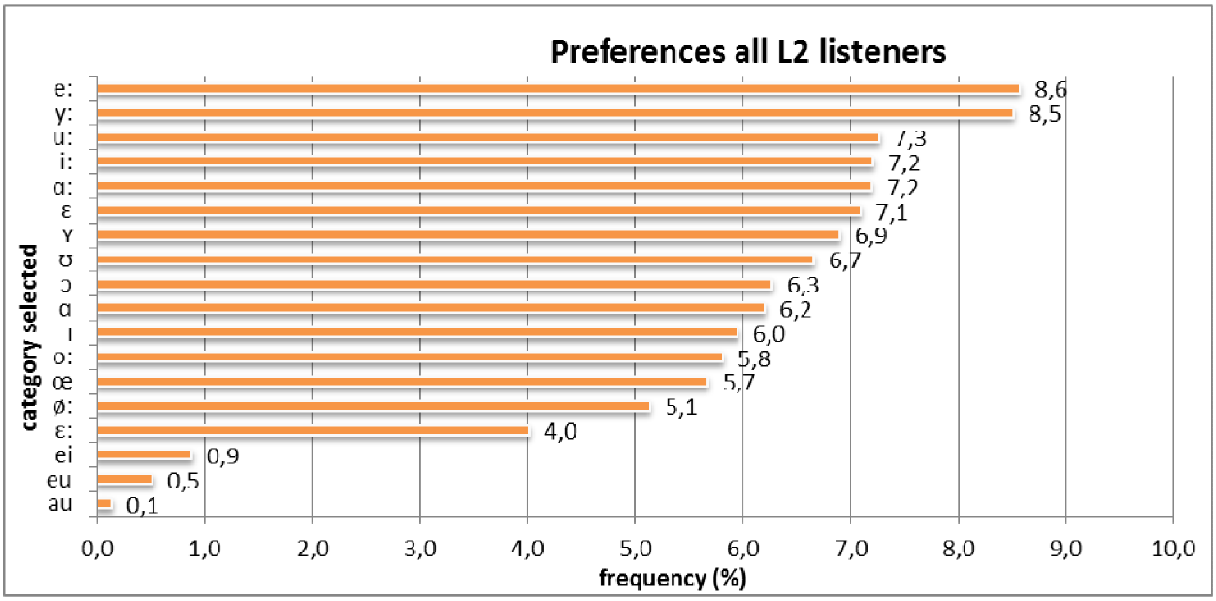

Figure 7.3: Barchart representing the preferences per vowel category for all 173 non-native listeners

If a specific response category is selected more often than others, it is considered to be more preferred and to function as a kind of "perceptual magnet" category that seems to attract more

responses from other categories (as will be discussed for each language individually in chapter 10). Patterns of *preferences* are expected to depend on the vowel quality itself as well as on the listeners' native language but they are also conditioned by the participants' individual or temporary hypotheses and strategies.
Preferences are moreover graphically represented by differences in type size in the two-dimensional MDS solution (see section 7.3).

7.2.4 Patterns of perceptual confusion

The confusion matrix summarizes the occurrence of perceptual confusion between vowel categories and patterns of vowel substitutions. For reasons of clarity and conspicuity, the percentage of correct identifications (id_correct) will be represented in bold numbers, the percentage of wrong identifications below the 5.5% mark will be printed in grey, id_scores below 0.5% are not indicated.
A crucial assumption of the current study refers to the expectatio that vowels that substitute for each other are in one or more aspects perceptually *similar*. The identification of phonetic correlates of these similarities is a matter of secondary interpretation.
In terms of a feature-based approach (see Tversky 1977), intercategory c*onfusion* can be viewed as a selective substitution of features of a given contrast. Mermelstein (1976: 2) considered similarity among consonants to be a monotonic function of the number of shared features. However, referring to perceptual similarity and distance between vowel categories in L2, it is not only the question of how many features distinguish a given contrast pair but rather a matter of the interaction of phonological features, their physical (articulatory and acoustic) correlates of these features and other properties of the categories involved such as graphemic similarity, frequency of occurrence, perceptual salience and the relative perceptual weighting that determine the listeners' interpretation of the speech signal.
It is crucial to recognize that confusion matrices are based on listeners' *responses* but do not provide direct information about the physical properties of the input stimuli. Any structure or pattern in the confusion matrix reflects interrelations between the stimuli as they are represented *psychologically* by the listeners – not as they are measured physically (Shepard 1972).
In this sense, perceptual confusion data provide valuable insights into the learners' *perceptual vowel space* and their metalinguistic theories and knowledge about the vowel system of the target language. In the present study, confusion matrices are used to derive measures of perceptual inter-category similarity between L2 vowel categories. Measures of similarity

(sim_scores) and distance are derived from the confusion data in order to map the listeners' underlying "*perceptual vowel space*" by evidence from the confusions observed.

To summarize, patterns of confusion and the occurrence of perceptual substitutions are considered as indicator of *psychological similarity* between categories as perceived by the listeners. A look at the confusion matrix reveals that some vowels are very frequently confused with each other. Vowel categories, whose members are more often substituted for each other, are considered to be perceived as more "similar" by the listeners and are therefore closer to each other in the listeners' perceptual space. Methods to parametrize the concept of perceptual similarity and to derive measures of psychological similarity from perceptual confusion data will be discussed in the following sections in this chapter and in section 12.2.4.

7.2.5 Similarity and distance

From the listeners' confusion, measures of perceived similarity and distance are derived. In a confusion matrix, each entry p_{ij} represents the relative frequency with which the i^{th} stimulus leads to a response belonging to the j^{th} stimulus. In this raw form, the entries p_{ij} are not entirely suitable as measures of proximity. Shepard (1972) in his work on the psychological representation of speech sounds suggests a method to compute more specific estimates of perceptual distance from confusion matrices. In a first step, the *similarity* of two vowels is calculated with Shepard's formula (1972: 73) from the response proportions p of each vowel pair under consideration. *Distance* measures are then derived in a second step from the similarity values.

A derived estimate of the *psychological proximity* or *similarity* s_{ij} is suggested by Shepard (1972) in terms of the relevant numbers p_{ij}, p_{ji}, p_{ii}, p_{jj} from the confusion matrix, calculating the total number of confusions between i and j and dividing it by the total number of correct responses to these same two stimuli:

$$s_{ij} = \frac{p_{ij} + p_{ji}}{p_{ii} + p_{jj}}$$

With Shepard's formula, similarity scores (*sim_scores*) are calculated and tabulated in a similarity matrix. Table 7.3 presents the similarity matrix for the German control group. *Sim_scores* are values between 1 and 0. The diagonal shows similarity values of 1. Zero values indicate that no confusion of the two categories was observed in the data. The lower the *sim_score*, the less similarity of the two respective vowel qualities is given.

V	ɑ	ɑ:	ɛ:	ɛ	e:	ɪ	i:	ɔ	o:	ʊ	u:	œ	ø:	ʏ	y:
ɑ	1.000														
ɑ:	0.021	1.000													
ɛ:	0.000	0.000	1.000												
ɛ	0.000	0.000	0.016	1.000											
e:	0.000	0.003	0.154	0.015	1.000										
ɪ	0.000	0.000	0.000	0.003	0.000	1.000									
i:	0.000	0.000	0.000	0.000	0.002	0.008	1.000								
ɔ	0.000	0.000	0.000	0.000	0.000	0.000	0.002	1.000							
o:	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.016	1.000						
ʊ	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.002	0.000	1.000					
u:	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.006	0.013	1.000				
œ	0.000	0.000	0.000	0.003	0.000	0.000	0.000	0.000	0.000	0.000	0.000	1.000			
ø:	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.002	0.000	0.000	0.014	1.000		
ʏ	0.000	0.000	0.000	0.000	0.000	0.005	0.000	0.000	0.000	0.002	0.000	0.003	0.002	1.000	
y:	0.000	0.000	0.002	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.002	0.000	0.006	0.017	1.000

Table 7.3: Similarity matrix for the German native listeners

This procedure is based on Shepard's law (Shepard 1972) stating that the relationship between perceptual distance and similarity is exponential, i.e. that similarity is a decreasing function of interclass distance.

V	ɑ	ɑ:	ɛ:	ɛ	e:	ɪ	i:	ɔ	o:	ʊ	u:	œ	ø:	ʏ	y:
ɑ	0.00														
ɑ:	1.47	0.00													
ɛ:	4.86	4.42	0.00												
ɛ	5.30	5.65	1.91	0.00											
e:	6.22	5.58	0.32	1.80	0.00										
ɪ	7.70	8.44	4.13	2.53	2.85	0.00									
i:	7.38	7.82	4.89	4.93	2.18	1.76	0.00								
ɔ	6.49	7.79	8.09	7.32	6.89	6.99	6.38	0.00							
o:	8.35	7.69	6.56	7.63	8.17	6.88	6.56	1.57	0.00						
ʊ	6.30	8.43	7.32	7.67	8.22	6.24	7.02	2.70	2.82	0.00					
u:	7.32	8.46	8.06	7.01	8.25	7.66	7.34	5.62	2.22	1.61	0.00				
œ	8.32	6.77	3.51	3.26	3.61	5.38	5.64	3.42	3.49	3.39	4.50	0.00			
ø:	11.51	11.51	4.73	6.00	4.72	5.58	5.94	4.83	2.67	3.65	2.84	1.57	0.00		
ʏ	8.33	7.68	7.24	7.61	4.89	3.67	4.95	5.31	4.19	2.32	2.93	2.08	2.05	0.00	
y:	11.51	7.76	5.66	6.60	4.67	4.95	4.21	7.30	4.17	3.80	2.70	3.61	1.30	1.41	0.00

Table 7.4: Distance matrix for all language sub-groups derived from similarity scores by Shepard's formula based on the responses of all 173 non-native listeners

To obtain *perceptual distance* values *d* from these similarity scores, the negative of the natural logarithm of similarity was calculated (Johnson 2003: 68).

$$d_{ij} = -\ln(s_{ij})$$

The resulting values for distances are tabulated in a *distance matrix* (see Table 7.4). Each cell of the distance matrix codes the relationship between one distinct pair of items in terms of a distance measure. Distance scores tend to infinity.

Table 7.4 presents the *distance matrix* for the response data of all L2 subjects. Here, the diagonal shows zero values. Low distance values indicate that the two items are very close to each other, e.g. /ɑ/ and /ɑ:/ with a distance of 1.47, or /y:/ and /ø:/ with a distance of 1.3. Higher distance scores indicate a larger distance of the vowel types involved, see e.g. the distance score of 11.51 between /ɑ/ and /y:/. The high scores indicate that these qualities are perceived as maximally distinct and are rarely – if at all – confused in perception.

In a further step of analysis, a *Multidimensional Scaling* procedure is performed based on the distance matrices for each of the sub-groups under investigation. This procedure will be discussed in details in section 7.3.

7.2.6 Symmetry and asymmetry

By converting confusion matrices into distance matrices, two types of difficulties have to be considered (Terbeek 1977: 12f): (1) zero-values due to non-occurring confusions between two given categories and (2) asymmetric patterns of confusion and substitution.

In many instances, too many zero-values may exist in large parts of the confusion matrix due to non-occurring confusions between two given categories (see e.g. the confusion matrix for German natives in Table 7.2 and the derived similarity matrix in Table 7.3), especially for very advanced learners or for languages with a vowel system like e.g. Hungarian that is very similar in structure to the German system. Zero values may cause problems with the calculation of the negative logarithm *-ln*.

A second critical point refers to the fact that confusion patterns usually appear to be *asymmetric*, i.e. the number of times a vowel *i* is identified as *j* is not equivalent with the number of times *j* is identified as *i* (Tversky 1977; Nosofsky 1986, 1987). As Shepard (1972: 73) argues, by means of his formula, the desired condition of symmetry $s_{ij} = s_{ji} = 1$ is achieved for all *i* and *j*. However, by this procedure, information about the asymmetric relationship between input stimuli and their substitutes gets lost. This point must not be neglected, since the asymmetric relationship between similar vowels is one of the central questions in the present study. It is hypothesized here that these asymmetric relations largely depend on two types of factors: (1) the language-specific bias for specific categories, and (2) the language-universal bias determining the relative preference for specific categories.

Therefore, data analysis will refer to the relative distance between vowel categories under the *preliminary assumption* that similarity and distance are *symmetrical*. These symmetrical distances will serve to map the perceptual vowel space for L2 learners by Multidimensional Scaling (see the following section). Asymmetric confusion patterns are subject to analysis in a further step. In a second step, a qualitative analysis and discussion of perceptual substitutions in the confusion matrix will discuss *asymmetric relations* between vowels; Relations of asymmetry are discussed in detail for each of the listeners' native languages in chapter 10, their possible reasons and origin will be discussed in chapter 11 and 12. A detailed theoretical discussion of the asymmetric nature of similarity relations between categories in a system and theoretical models to account for asymmetries and bias in the acquisition of a second language will be provided in chapter 12.

7.3 Multidimensional Scaling (MDS)

7.3.1 Mapping the perceptual space

The human perceptual space is commonly conceived to be multi-dimensional, i.e. *objects* are perceived and differentiated in terms of different *properties* that vary across more than one dimension. *Multidimensional Scaling (MDS)* is a method to map the human perceptual space. MDS is based on data of *perceived similarity* between objects – here vowel categories as represented in the similarity matrix – and yields a spatial configuration of the perceptual space of one or more dimensions.

In geometric models of similarity (see section 12.2.2 and 12.2.3), each vowel category is situated in a geometric space of two or more dimensions. In an MDS representation of the perceptual space under consideration, distances between points correspond to the perceived similarity between objects or categories. Referring to the perceptual vowel space, similar vowels are closer to each other than dissimilar vowels.

MDS is a useful device to visualize relationships between vowels that are not readily evident in a confusion matrix. The method has been used in a number of studies to model the perceptual similarity of phonemes (e.g. Shepard 1972; Terbeek 1977; Kewley-Port & Atal 1989; Iverson & Kuhl 1995; Fox, Flege & Munro 1995; Francis & Nusbaum 2002).

An advantage of this method is that the relevant phonetic or phonological variables of the input stimuli determining their position in the perceptual space can be unknown and that there is no external influence exerted on the results (e.g. by selecting specific properties and excluding others). A possible disadvantage of the method is related to a certain loss of information contained in the original data compiled in the confusion matrix, the similarity matrix or the distance matrix (e.g. loss of information about the asymmetric bias in substitution, e.g. /e:/ > /i:/ vs. /i:/ > /e:/) and to difficulties in the interpretation of the results since no direct correspondence between dimensions of the perceptual space and the physical properties of the objects can be expected.

Several studies using MDS attempted to establish correspondences between dimensions in the MDS and articulatory or acoustic parameters of speech sounds, mostly by searching for correlations between phonetic parameters and the items' position in the MDS space (e.g. Fox, Flege & Munro 1995; Kewley-Port & Atal 1989; for a critical discussion, see Shepard 1972, 1980; Terbeek 1977). In many of these studies, dimensions in MDS are posited to correspond closely to phonological features or to acoustic measurements, such as formant frequencies of vowels, reflecting their distribution in the vowel space. As Kewley-Port & Atal (1989) posit, even sub-phonemic differences among vowels could be effectively mapped with MDS, though this position has to be viewed very critically.

Fox, Flege & Munro (1995) used an MDS analysis to describe the similarity or dissimilarity of three Spanish and seven English vowels asking for the underlying dimensions by which their English and Spanish subjects rated the similarity or dissimilarity of vowel contrasts. Fox et al. (1995) posited that the number of dimensions to describe the listeners' performance best can vary language-specifically. They argued for a two-dimensional, non-metric solution for Spanish learners of English and a three-dimensional, non-metric solution for the English natives. By addition of a further dimension, no sufficient increase of cumulative variance was observed. Interpreting the perceptual dimensions in the MDS, Fox et al. (1995) compared the coordinates of each vowel in the spatial solution with their acoustic properties in a correlation analysis. The most salient perceptual dimension for both languages seemed to be vowel height. For English listeners they posited that this dimension was most significantly correlated with duration, indicating a language-dependent sensitivity to this phonetic feature, whereas duration showed no significant correlation with any of the Spanish dimensions. For the Spanish listeners, only two dimensions were assumed to be salient enough in the correlation analysis for being easily interpreted. For English listeners, Fox et al. (1995) found that dimension 2 seemed to correlate most with a "front/back"-distinction and dimension 3 correlated with the Euclidean distance measure between categories – a fact they related to a mid/non-mid dimension referring to "central" vowels.

It is important, however, in the interpretation of a MDS to avoid this "correlation-causation-trap". As Shepard (1972) explicitly points out, the MDS analysis itself refers only to *psychological similarity* and differs from more psychophysical approaches, in that it is based exclusively on the empirical data of the listeners' judgments. Similarity of two given items is is an exponentially decreasing function of interclass distance and is derived from the reported confusion matrix. Perceptual distance corresponds to the distance between classes *i* and *j* in the MDS spatial representation. MDS analysis does not start with a preconception of variables such as acoustic or articulatory parameters that are assumed to be relevant. It even does not specify which variables will be relevant for establishing a relation between the data and the variables. Rather, the frequency of confusion of phonemes with each other functions as primary measure for *psychological similarity* that has to be distinguished from similarity in terms of the stimuli's *physical properties* (Shepard 1972) as other factors such as e.g. graphemic representations, the categories' frequency of occurrence, the existence of equivalents in L1, the listeners' proficiency in L2 also have an influence on perceived similarity and inter-category confusion. MDS is regarded here as a very useful technique for visualizing psychological similarity. In a second step it can be interpreted in terms of acoustic or articulatory aspects.

Additional evidence for the validity of spatial representations of similarity or distance in MDS solutions may be sought by the relations with external variables such *as phonological features* or *physical properties* of sounds. In the current study, an analysis of physical similarities is provided by *cluster analyses* of different combinations of *acoustic properties* of input stimuli iin the perception experiment (see section 5.7). The outputs of the hierarchical cluster analyses in section 5.7 will represent additional evidence for the relative similarity of German vowels in terms of their acoustic properties. It can be expected that physically similar input stimuli will be more easily confused by the listeners than more distant vowel tokens.

7.3.2 Distance measures and scale level

To map objects in the perceptual space, data has to be collected that reflect the perceived similarity of objects of interest (here German vowel categories). There are several different experimental methods to measure similarity of pairs of objects, such as pair ranking, the anchor point method, or the rating of objects along a similarity/dissimilarity scale (Backhaus 2011b: 223ff).

In the present study, MDS was performed on the basis of the distance measures derived from the confusion matrix for each of the listeners' native languages. The analysis is based on the distance matrix calculated with Shepard's formula for each of the listeners' native languages separately. The *Euclidean distance* is used as distance measure. The distance matrices derived from the confusion matrices are analyzed separately for each of the listeners' native languages, based on the assumption of an inherent perceptual space common to all subjects of a given language sub-group.

Using the ALSCAL procedure implemented in SPSS 19, an *ordinal non-metric MDS* solution (see Norušis 2007: 350ff) is preferred here for several reasons: In a confusion matrix, the number of instances where input stimuli were identified as belonging to a specific category is counted. The input categories as well as the response labels used in the experiment are *nominal* categories, for which central tendencies can only be described by the mode, i.e. the label that is most frequently observed in the data set. By calculating similarity scores from these categorical data, we obtain higher scaled non-metric *ordinal* data. Ordinal values reflect the relative ranking along a non-metric scale of similarity but are not indicating directly the magnitude of difference. MDS does not require metric data to derive a configuration from the data, the ranking of distances is sufficient. However, the results of MDS, i.e. the derived distances between objects in the perceptual space, are always *metric* (Backhaus et al. 2011b: 221).

7.3.3 Number and interpretability of perceptual dimensions

A description of the perceptual space in terms of a multidimensional representation is not only determined by the scale level but also by the number of relevant dimensions and their orientation in the space. Therefore, a central question for data interpretation concerns the number and quality of *dimensions* of the perceptual vowel space. As was outlined above, the number of relevant dimensions determining the perceptual vowel space is a priori unknown and has to be set by the researcher. In L2 vowel perception, the learner-specific weighting of these dimensions can vary language-specifically, individually or even temporarily and is moreover be influenced by context-specific phonetic factors. Therefore, in the present study, no a priori assumptions are made concerning the number of dimensions that would best describe the listeners' perceptual behaviour. Rather, one-, two- and three-dimensional solutions will be discussed for their descriptive value for each of the languages under consideration.

As discussed above, Multidimensional Scaling transforms lower scaled non-metric data into higher-scaled metric data. Raising the scale level requires a procedure of data compression. To determine the appropriate number of dimensions and to assure a stable solution, a *data compression coefficient Q* is calculated for the transformation of ordinal-scaled data into metric data (Backhaus et al. 2011b: 227ff) by the formula

$$Q = \frac{K * (K-1)/2}{K * R}$$

where K is indicating the number of objects and R the number of dimensions. In other words, Q is the relation between the number of (dis-)similarities, i.e. the input, and the number of coordinates, i.e. the output data.

Q has to be higher than 1, i.e. the number of input data has to be higher than the number of output data, and should be lower than 2 to guarantee a stable solution. Compression is increased when the number of objects is higher and is decreased by a smaller number of objects. As a rule of thumb, Backhaus et al. (2011b: 227ff) recommend a Q-value ≥ 2 to yield a stable solution. For a two-dimensional solution, a minimum of 8 objects is recommended, in a three-dimensional solution a minimum of 12 objects would be required (Backhaus et al. 2011b). In the present study, we compare 15 objects, i.e. 15 German vowel categories, and obtain a Q-value of 3.5 for a two-dimensional solution (R = 2) and 2.33 for a three-dimensional solution (R = 3).

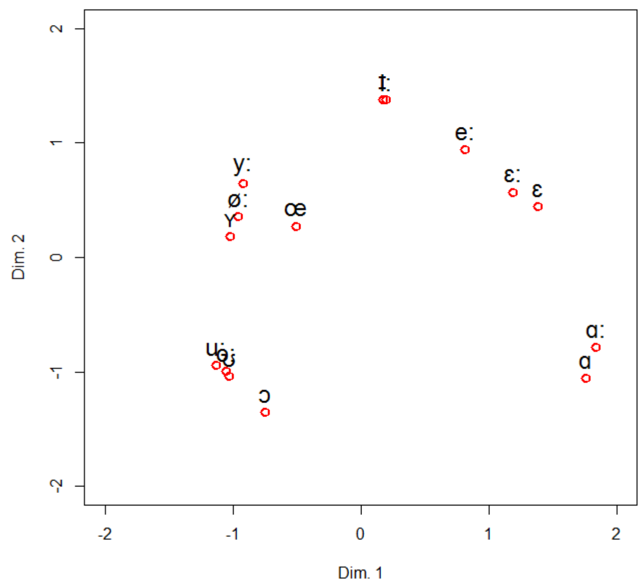

Figure 7.4: Two-dimensional non-rotated MDS solution representing the perceptual vowel space for all 173 non-native listeners computed from the distance matrix in Table 7.4 (RSQ .905)

Figure 7.4 presents a two-dimensional and Figure 7.5 a three-dimensional non-metric MDS solution for all non-native listeners participating in the study.

Though a three-dimensional solution may differentiate more aspects of relevant phonetic and phonological components that determine the perceived similarity between categories in L2, a two-dimensional solution may be preferred for reasons of better visual representation.

For reasons mentioned above, a non-metric solution derived from ordinal similarity scores was preferred in the present analysis. A comparison of goodness of fit for metric and non-metric MDS solutions for all language sub-groups is provided in Figure 9.16.

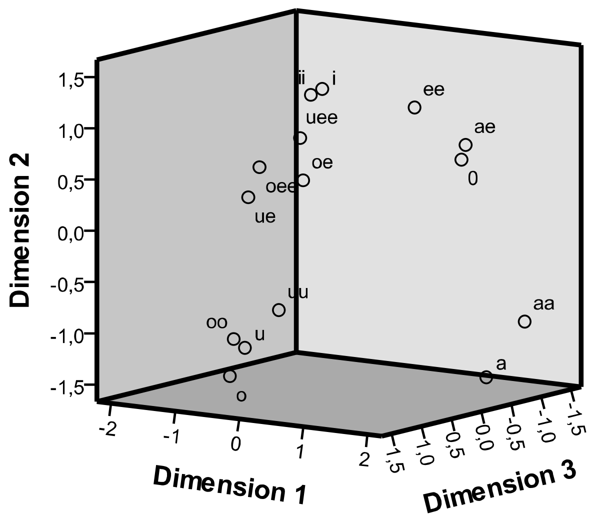

Figure 7.5: Three-dimensional non-metric MDS solution representing the perceptual vowel space for all 173 non-native listeners computed from the distance matrix in Table 7.4 (RSQ .926) [63]

A look at the MDS representations in Figure 7.4 and Figure 7.5 shows that the grouping of items does seems to reflect natural classes of vowel sounds, e.g. the class of front rounded vowels, front non-rounded vowels or back vowels. However, it is important to stress that the description of the items' positions and of dimensions in the MDS in terms of either phonetic parameters or phonological features is a matter of post-hoc interpretation.

7.3.4 Goodness of fit

The MDS procedure implemented in SPSS starts with a first configuration that is improved in several subsequent iterations of the procedure until a minimal *stress value* is reached indicating the *goodness of fit* of a given configuration. *Stress* values refer to the difference between distances of objects and disparities ($d_{kl} - \hat{d}_{kl}$). The higher the stress value, the lower the "goodness of fit" of the model. Ideally, the ranking of distances is equal with the ranking

[63] In the three-dimensional MDS representations produced in SPSS no IPA fonts can be produced. Long vowels are represented by double letters, e.g. /u:/ - <uu>, /e:/ - <ee>. Front vowels /y: ʏ ø: œ/ are represented as <uee ue oee oe>, /ɛ:/ as <ae>; in SPSS 19 /ɛ/ is replaced by <0> ([sic!] presumably due to a software error).

of dissimilarities and should be close to zero (Backhaus et al. 2011b: 222). Increasing the number of dimensions decreases the stress values.

Stress values in the ALSCAL procedure in SPSS are defined by *Kruskal's stress formula 1*. Table 7.5 offers an interpretation of Kruskal stress 1 values (Backhaus et al. 2011b: 225).

goodness of fit	*Kruskal's stress 1*
poor	.2
sufficient	.1
good	.05
very good	.025
perfect	.0

Table 7.5: Goodness of fit for Kruskal's stress formula 1 (Kruskal & Carmone 1973, cited in Backhaus et al. 2011b: 225).

Another measure for goodness of fit obtained in SPSS is *RSQ* given for an *n*-dimensional solution: *R-squared (RSQ)* is the squared correlation of distances and disparities, giving the proportion of variance of the scaled data (disparities) in the partition (row, matrix, or entire data) which is accounted for by their corresponding distances (Backhaus et al. 2011b). Values above 0.9 indicate a very high goodness of fit.

A comparison of the language sub-groups shows that stress values and RSQ (*R*-squared or R^2) and Kruskal stress values may vary considerably across language sub-groups for each dimension added (see Table 7.6 and Table 7.7).

Kruskal's Stress 1 (non-metric MDS)	*Dim 1*	*Dim 2*	*Dim 3*	*Dim 4*
Albanian	.445	.230	.147	.082
Arabic	.386	.163	.118	.073
English	.340	.140	.100	.062
Farsi	.436	.194	.134	.115
Hungarian	.575	.242	.165	.108
Mandarin	.399	.162	.113	.091
Polish	.318	.185	.104	.085
Romanian	.463	.228	.152	.110
SerBoCroatian	.480	.209	.121	.073
Turkish	.512	.250	.162	.105
L1 German	.560	.311	.212	.141
all L2 listeners	.307	.136	.102	.058

Table 7.6: Comparison of stress values (Kruskal's stress 1 formula) for ordinal-scaled MDS solutions by language sub-groups, the German control group and full L2 sample

Table 7.6 summarizes the Kruskal stress 1 values for the non-metric solution for the ordinal-scaled MDS representations based on distance values for each of the language sub-groups, for the German control group and for the "all L2" sample showing that for most language sub-samples "sufficient" stress values are obtained in three- or more-dimensional solutions.

For all languages, the RSQ values were higher in a two-dimensional non-metric solution than in a two-dimensional metric solution (for a comparison, see Figure 9.16). A visual comparison of RSQ values for one-, two-, three- and four-dimensional solutions is provided in Figure 9.16 in section 9.5.2. However, not for all languages values above 0.9 are not obtained, even in a four-dimensional solution. Table 7.7 compares the *RSQ values* for non-metric MDS solutions for all language groups tested. The *RSQ values* obtained (see Table 7.7) vary considerably between the language sub-samples. While for some languages, a very good fit is already obtained in a two-dimensional non-metric solution (e.g. *RSQ* for English 0.918, Arabic 0.882 or Mandarin 0.865), for other languages a percentage above 0.9 is only obtained in a four-dimensional solution. In the case of Farsi, Hungarian, Romanian, Turkish and German, the 0.9-mark is not reached even in a four-dimensional solution.

RSQ (non-metric MDS)	*Dim 1*	*Dim 2*	*Dim 3*	*Dim 4*
Albanian	.505	.733	.814	.928
Arabic	.655	.882	.914	.952
English	.608	.918	.940	.966
Farsi	.538	.835	.884	.893
Hungarian	.148	.671	.799	.842
Mandarin	.635	.865	.911	.931
Polish	.689	.844	.922	.935
Romanian	.476	.751	.830	.876
SerBoCroatian	.302	.782	.878	.940
Turkish	.405	.693	.801	.896
L1 German	.170	.461	.624	.773
all non-native listeners	.730	.905	.926	.967

Table 7.7: R-squared (RSQ) values for the ordinal-scaled MDS analysis based on distance values for each of the language sub-groups, for the German control group and for all L2 subjects together.

For L1 Hungarian and for the native German control group the values are particularly low in the first dimension (0.148 Hungarian, 0.170 German) and keep low after the addition of a second or third dimension. This effect is strongly conditioned by the very low error rate and the resulting high number of *zero*-values in the similarity matrix causing equal values for a number of vowels in the distance matrix for these two languages.

It is however important to note that the fact that a good fit is obtained in terms of two (or three) dimensions does not mean that these dimensions can be interpreted as directly reflecting variations in two specific distinctive phonetic or phonological feature dimensions only (as e.g. suggested by Fox et al. 1995). Distinctions in terms of other features may be (at least partially) preserved in the data but are not directly evident in a two-dimensional representation. Therefore, a two- or three-dimensional description of the MDS solutions in terms of only two distinctive features has to be regarded as oversimplification of the complex relationships of similarity within the perceptual space (Shepard 1972: 77). As was outlined above, the inter-category similarity as perceived by the listeners is determined by the listeners' perceptual weighting of several properties of the categories under consideration, such as articulatory and acoustic properties but also orthographic characteristics, frequency effects or lexical considerations (for a detailed theoretical discussion, see section 12.4). Therefore, a two- or three-dimensional solution will never be able to represent the "true" psychological dimensionality of the objects under consideration in its full range. Of course, a solution on a space of higher dimensionality might be sought by which the fraction of variance accounted for will increase. However, simply increasing the number of perceptual dimensions is not an appropriate solution to the problem of factor multiplicity to explain perceptual similarity. As Table 7.7 and Figure 9.16 show, even by a four-dimensional solution, it will not be possible to account for the full variance in the data.

To summarize, MDS may furnish a useful reduction of the original nominal confusion data, transforming non-metric data into higher scale metric data and making it available in an easily accessible, explicit forma. It provides a visualization of distances between items in a more-dimensional space in terms of the position of these objects in the human psychological perceptual space. The relative distances between vowel categories in the MDS solution are considered to reflect the similarity or dissimilarity of objects and the susceptibility to perceptual confusion. Categories that cluster more tightly in the MDS representations are more similar and therefore more difficult to discriminate. A spatial representation of vowel points in a space of two or more dimensions can be interpreted as corresponding to the psychological space, providing a map of the "perceptual vowel space". The perceived distance or similarity between categories may vary language-specifically. It is necessary to stress, however, that perceptual maps cannot be directly related to acoustic-phonetic data, but are based on the *perceived* distance or similarity of the input signal as derived from the confusion matrix, i.e. they are derived from the listeners' *behaviour* in the experimental situation.

In a post-hoc interpretation, phonological features (which are nominal in nature) or physical variables of vowels, e.g. formant frequencies (which are inherently continuous), can facilitate the interpretation of the data and may help to define the continuous psychological space at least partially, though a direct relation to projections of the recovered points on any reference axis or dimension in a MDS solution must not be taken as a matter of course (Shepard 1972).

7.3.5 Rotation

A fact that must not be overlooked is that the position of the reference axes in the MDS solution and the orientation of the points in the spatial representation are totally arbitrary. For the purpose of interpretation of MDS, it is quite permissible to seek for new axes corresponding to external variables on which the projections of the points are correlated with the variable (Shepard 1972: 80). The dimensions can be reflected, translated, permuted or rotated, and even be rescaled by the same scaling factor (Norušis 2007: 348).

For vowels, the frequency values of the first two or three formants have often been proposed as relevant variables. Fox et al. (1995) calculated Pearson product-moment correlations between the vowel coordinates in the perceptual-spatial representation and several other acoustic parameters such as duration, *VOT, F0, F1–F0, F2–F1, F3–F2*, and several dynamic properties as *F1-, F2-,* and *F3*-slopes from onset to midpoint, midpoint to offset and onset to offset, and distance from a "neutral" vowel (F1=500, F2=1500) for each of these parameters.

However, even if a clear relationship between physically measurable properties of the stimuli and the "psychological structure" would become evident in the correlation procedure, it is still important to be aware that this is just an interpretation of the psychological relationships of similarity or distance between the vowel points. Concepts that rely even less on acoustic phonetic facts, using notions like "peripheral" vs. "central" vowels cannot be corroborated by the MDS output.

Johnson (2012: 147) computed MDS solutions of the data from Terbeek (1977) for four languages (English, German, Thai, Turkish) and rotated the vowel spaces so that the line connecting /i/ and /u/ was horizontal, revealing similarities between the perceptual vowel spaces and the acoustic vowel space as represented in an F1xF2 chart.

In the present study, the MDS coordinates obtained in the analysis discussed above were rotated so that the line connecting the points for /i:/ and /a:/ is in a position of -45° in the space.

If necessary, the obtained solution was moreover flipped along the diagonal axis. However, a rotation of the solution does not change the relationship between the vowel categories (Norušis 2007: 348). The same Euclidean distances between objects are given before and after rotation and/or reflection, only the coordinates in the space are changing. Figure 7.4

represents the non-rotated MDS solution for the L1 German control group, Figure 7.6 shows the rotated form. The coordinates of the vowel points are different, the distances between objects, however, remain the same.

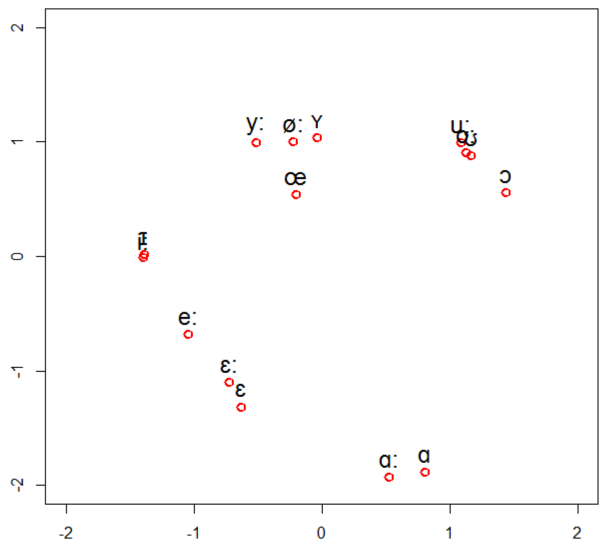

Figure 7.6: Two-dimensional MDS solution for all L2 listeners after rotation and reflection representing the perceptual vowel space for all 173 non-native listeners computed from the distance matrix in Table 7.4

A smaller distance between vowels in the space points to a higher degree of perceived similarity between these vowels. The visual representation of the perceptual vowel space in Figure 7.6 shows that the distance between back rounded vowels is very small, revealing a high degree of perceptual similarity between these vowels with the only exception of /ɔ/, which appears to be more distinct from /u:/, /ʊ/ and /o:/. Another cluster of similar vowels is formed by front rounded vowels. [a]-vowels are clearly distinct from all other categories. A detailed discussion of results for the “all L2” sample is provided in chapter 9.

All MDS representations in chapter 9 and 10 are rotated and if necessary flipped. In Figure 7.7 additional information concerning the relative *preference* for specific German vowel categories is incorporated in the graphic representation of the vowel space: The *type size* of the category label indicates the relative preference for a given category: smaller symbols show that the respective category was selected less often as response option than categories with bigger symbols, e.g. /e:/ is clearly preferred to /ɛ:/, i.e. /e:/ is more often selected as response option than /ɛ:/.

7.3.6 Interpretation of spatial representations

The challenge in the linguistic interpretation of spatial representations of similarity in the perceptual vowel space is to find those acoustic and physical parameters that are psychologically relevant in sound recognition or confusion. It is important to emphasize that the listeners' response behaviour may not only be determined by mere acoustic features but also by other vowel properties (e.g. articulatory properties such as rounding or orthography).

It is necessary to note that the distances and similarities between different speech sounds are not equally distributed and that the sound space is also not uniformly populated (see chapter 4). The similarity between two palatal vowels will be substantially higher than the similarity between a palatal and a pharyngeal vowel. The listener's individual perceptual space is determined by *differential weighting* of acoustic cues to vowel identity. These cues and weights vary according to the listeners' native language but are also influenced by the learners' proficiency in the L2, the L2 input so far available to the learners and their (sub-conscious) processing of acoustic cues in phoneme decision (see section 2.1, 2.2 and 12.5). The role of physical properties of vowel sounds can be simulated to a certain degree by variation of acoustic input information (e.g. formant frequencies or distances, durational information etc.) for Hierarchical Cluster analysis (as in section 5.7.2), though modelling the specific weighting of such cues is a more complicated matter (for discussion, see section 12.3.3 and 12.5).

Using confusion data, the MDS approach assumes the distance between distinct tokens of the same phonemic category to be zero, even if we know that there may be considerable sub-phonemic differences between single tokens of the same vowel category (Mermelstein 1976: 94).

To account for inter-category and intra-category variation of id_scores, the MDS results have therefore to be considered together with other types of evidence, such as the direct interpretation of patterns in the confusion matrix and a comparison of sim_scores for contrast pairs. The present study will focus on three major aspects: (1) *difficulties*, (2) *preferences*, and (3) inter-category *similarities* as reflected by the choice of response categories in the vowel identification task and evident in similarity scores and MDS representations.

There are different approaches to the interpretation of spatial representations as provided by MDS solutions. In a second step, for a post-hoc interpretation the solutions are reviewed

(1) identifying vowel *features* in L2 that correspond best to the dimensions obtained in terms of (a) traditional *phonological* features like [front/back], [+/–round] or [high/low], or (b)

articulatory features referring to constriction location or constriction degree, degree of aperture or lip activity (as discussed in 4.7.4, 4.7.5 and 5.4),

(2) identifying *acoustic correlates* that correspond best to the dimensions and distances in the MDS representations obtained before or after rotation,

(3) interpreting the position of the vowels in the dimensions with reference to "similar" *L1 vowels in the listeners' native language*, assuming that similar L2 vowels are clustering around the nearest L1 sound.

It is important to stress that for a qualitative interpretation of a given MDS solution it is necessary to consider the relative *distances* between items in the spatial representation rather than the position of one specific item in the space. In other words, MDS refers to perceptual *distance* and similarity relations *between* items but is not meant to provide information about the absolute position of an item in the phonetic or psychological vowel space.

Similarity is, after all, a concept of the listener. Not only acoustic properties of the input signal and their psychophysical correlates but also the listeners' concept of L2 categories, individual considerations, expectations and (temporary) hypotheses about typical characteristics of vowels and their occurrence in the input stimuli influence the participants' reaction to the stimuli.

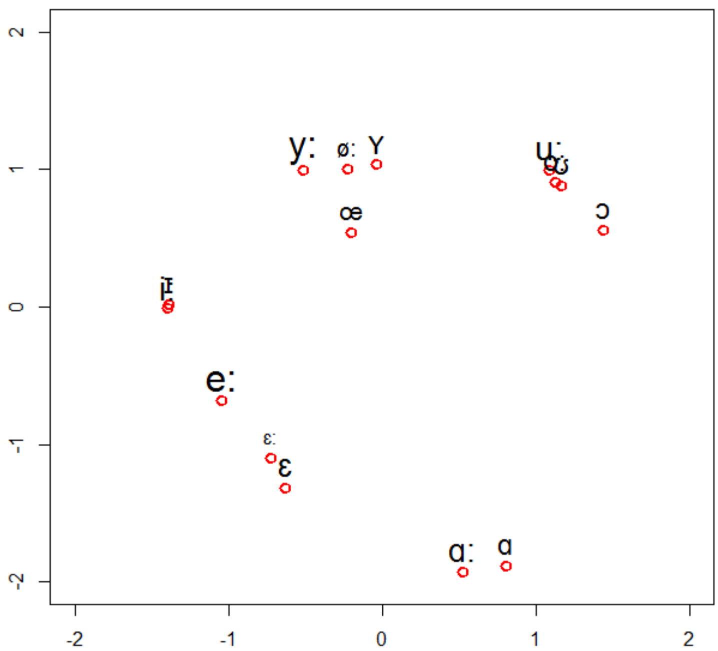

Figure 7.7: Perceptual vowel space for non-native listeners representing listeners' preferences for categories by type size (the smaller the less frequently selected)

7.4 Conclusions

To summarize, a detailed analysis of vowel perception in L2 acquisition has to include an analysis of the input signal (*stimuli*), the listeners' decisions for a specific output category (*responses*) and a bundle of factors that determine the listeners' decision such as the listeners' L1, the experimental design (including characteristics the input stimuli), properties of the input signal, number and types of response categories, person-related factors such as language experience in L2 (cf. chapter 8), and the possible influence of more general factors that are commonly summarized under the terminus markedness and universal preferences (see section 1.2).

Table 7.8 summarizes all procedures for the analysis of the German vowel system and the input stimuli (INPUT analysis) and the methods to analyse the listeners' responses (OUTPUT analysis) as well as different aspects of data interpretation (GENERAL analysis).

A detailed analysis of the German vowel *input data* in terms of phonological, articulatory and acoustic characteristics of German vowel categories was provided in chapter 5. A Hierarchical Cluster analysis on the basis of acoustic measurements was performed in order to group the full set of German vowel phonemes into sub-groups that are considered to be phonetically similar with respect to specific articulatory and acoustic properties (see section 5.7). Different hierarchical cluster solutions are provided that serve as language-independent simulations of possible areas of difficulty and category confusion (see section 5.8).

The vowel system of the listeners' L1 is described contrastively in chapter 10 (see the language-specific descriptions in sections 10.n.2 and 10.n.3 respectively) to predict possible language-specific perceptual problems with German vowels.

The *output data* was analysed as follows: The learners' *difficulties* are described by the calculation of identification scores (*id_scores*) providing information about instances of wrong and correct identifications per vowel. Patterns of confusion in a given language sub-group are considered to reflect individual and language-specific *difficulties* in the L2 perception. The full range of wrong and correct identifications is subsumed in a *confusion matrix*. Confusion matrices reveal areas of instability and incomplete category formation. Wrong responses indicate difficulties in category formation and confusion among categories. Confusion matrices also indicate the frequency by which a certain category is selected as response option indicating *preferences* for specific response categories and patterns of perceptual substitutions (*id_V scores*).

Using Shepard's (1972) method for calculating the relative *similarity* of two given categories *i* and *j* from identification scores in a confusion matrix, *similarity scores* (*sim_scores*) are obtained. These are transformed into values of perceptual *distance* by taking the negative of the natural logarithm of similarity and are tabulated in a distance matrix. A *Multidimensional Scaling* analysis is calculated on the basis of these distance values. The method of Multidimensional Scaling (MDS) is used to *map the L2 listeners' perceptual vowel space* based on the observed confusion patterns. These perceptual vowel maps will visualize psychological relationships of similarity and distance between vowels. In a further step, possible correlations between dimensions of the perceptual map, phonological and phonetic properties and patterns of difficulty and category preferences may be considered.

INPUT ANALYSIS Phonetic and phonological similarity of German vowel categories	**OUTPUT ANALYSIS** Language-specific patterns in vowel categorization and perceived similarity of German vowels in L2
Phonological description of German vowels • articulatory constellations • phonological features and their phonetic correlates (constriction location, +/–round, +/–constricted, +/–open, +/–long) • type and token frequency	**Confusion Matrix and Descriptive Statistics** • difficulties (wrong identifications of presented stimuli) • preferences for specific response categories • perceptual vowel substitutions and confusion patterns • patterns of asymmetry and response bias
Acoustic analysis of input stimuli • formant frequencies • suprasegmentals (duration, F0)	**Cross-language vowel perception – Data analysis and interpretation** • difficulties and category confusion • similarity and distance • asymmetry and preferences • perceptual salience and phonetic ambiguity
Hierarchical Cluster Analysis • simulating phonetic ambiguity and perceptual similarity/distance of input stimuli derived from acoustic properties • post-hoc interpretation in terms of distinctive features	**Similarity Matrix & Distance Matrix** • perceived similarity and distance between L2 categories derived from listeners' responses • interpretation in terms of Contrastive Analysis of L1 + L2, distinctive features and acoustic properties
Contrastive Analysis L1 – L2 German • vowel inventory • system size, phonemic contrasts and distinctive features • articulatory and acoustic phonetic characteristics	**Multidimensional Scaling** • similarity in terms of perceived distance • language-specific perceptual vowel maps • phonetic similarity vs. psychological similarity
Predicting perceptual similarity and confusion, difficulty and preferences from • Hierarchical Clustering: predicting relationships of inter-category similarity in terms of acoustic properties • Contrastive Analysis of L1 and L2 • type-frequency in German • cross-linguistic distribution and universal markedness of vowel qualities and vowel contrasts	**Analysis of patterns of perceptual similarity and confusion, difficulty and preferences** • language-specific confusion patterns • mapping the listeners' multidimensional perceptual space in a geometric MDS solution • accounting for asymmetries by selective attention and biases • stimulus bias vs. response biases • signal-related vs. listener-related biases

GENERAL ANALYSIS

- **Learner-related factors** (chapter 8)
- **Cross-linguistic comparison and interpretation of general patterns** of difficulty, category confusion, category preferences and asymmetry in perceptual substitutions
- **Language-specific analysis and interpretation** (chapter 10)
- **Vowel-specific analysis and interpretation** (chapter 11)
- **Modelling language-specific vowel perception** as effect of phonetic and psychological similarity (chapter 12)
- **Modelling the acquisition of L2 vowel contrasts** by selective attention and attentional learning, i.e. re-allocation of selective attention and attentional weight (chapter 12)

Table 7.8: Schematic representation of data analysis and interpretation of results

8 Learner-Related Factors and Speaker-Specific Analysis

Phonetic learning in second language acquisition and the performance of L2 learners in experimental studies is determined by the complex interplay of different types of variables such as (1) *language-specific variables* determined by the sound patterns of the learners' native language and the target language, e.g. the number of vowel categories in the listeners' L1 vs. L2, the existence of specific contrasts in L1 vs. L2, and the occurrence of specific category types in L1 vs. L2, (2) *personal variables* other than the listeners' L1 such as language experience and level of proficiency in L2, and (3) *methodological variables* and their influence on the participants' performance in an experiment (experimental setting, stimuli, task type).

Methodological aspects have been discussed in chapter 6 and 7, language-specific results will be presented in chapter 10 and a cross-language comparison and discussion of general patterns in the data is provided in chapter 9 and 11. This chapter will focus on person-related variables that have an influence on L2 learners' success in the perception experiment.

Difficulties in L2 perception cannot be understood by referring only to linguistic factors without considering *who* the learners are and what kind of language experience and linguistic environment they have been exposed to in the course of L2 acquisition (Odlin 1989: 128). Based on the analysis of person-related information collected in a questionnaire, this chapter will discuss the influence of learner-related variables on individual differences in the listeners' relative success in the study's perception task.

8.1 Collection and analysis of personal variables

The subjects participating in the present study were asked in a questionnaire about their age, sex, their L1, language experience and instruction in L2 and other foreign languages, language instruction, the learning setting and other learner-related variables.

Based on the participants' answers in the questionnaire, personal variables were formed for a statistical comparison that allow to compose homogeneous subject groups and to filter sub-groups with respect to individual factors such as (a) the learners' *first language*, (b) the *level of proficiency*, (c) *language experience* in L2, i.e. length of residence (LOR) in a German speaking country, the age at which L2 German learning began (AOL), and the length of learning (LOL)), (d) length and kind of *language instruction* in German (school, language course, university, private contacts, …), (e) *formal instruction* in language, phonetics and pronunciation training, (f) general *experience in language learning* (bilingualism, number of

foreign languages acquired), (f) personal contact with German native speakers, (h) *age*, and (i) *sex*. These factors were related to the listeners' performance in the perception experiment.
Identification scores for each person participating in the study were obtained in the perception test. The subjects' overall identification score for all German vowel qualities, i.e. the percentage of wrong identifications (id_wrong), served as *dependent variable* to describe the learners' success in the L2 perception experiment. *Person-related information* such as age, length of instruction in German level of proficiency, duration of residence in a German speaking country served as *independent variables*.
A detailed statistical data analysis was carried out to identify those learner-related variables that may explain the participants' success in the experiment. Their influence was tested by statistical procedures, e.g. correlation coefficients, mean comparisons and analysis of variance. The results of this analysis will be discussed in the following sections[64].

8.2 Internal vs. external factors

Most studies in second language acquisition consider the influence of *external* factors related to the circumstances of language acquisition and language use such as age of arrival, length of residence, teacher, class, or amount of use of L1 and L2 on the learners' performance in foreign language acquisition (Flege, Frieda & Nozawa 1997; Piske, McKay & Flege 2001; Flege & Liu 2001). Several other variables such as age, sex and its biological and social correlates, talent and affective components such as language attitude and motivation, learning style and personality (e.g. extro-/introvertedness, empathy, fear) are discussed in the literature as learner-*internal* factors (Gardner 1991; Larsen-Freeman & Long 1991; Ellis 1994; Bohn 1998; Grotjahn 1998; Lightbown & Spada 1999). "Talent" is a frequently used notion to account for individual variances in performance in L2-related experimental tasks and in overall success in language learning. Studies with a cognitive, psychological or neurological approach examine other internal factors such as motivation, affective components and even personality factors such as empathy and extroversion (see Guiora et al. 1972; Larsen-Freeman & Long 1991; Elliot 1995a; Bohn 1998; Moyer 1999; Dörnyei 2005, 2008; Jilka 2009).
Learner-internal factors receive particular scientific interest to account for individual differences in language acquisition and to explain differences in quality of performance and rate of acquisition of L2 structures within the same group of L2 learners. However, a clear definition of learner-internal variables is very difficult as influences of cognitive, social and

[64] The Farsi and Mandarin samples were excluded from some of the statistical procedures in this chapter due to small sample size.

psychological features strongly interact and superficial proficiency and person-inherent factors like talent or motivation are very hard to differentiate in the learners' performance. Additionally, the complex interplay of several other variables that influence the learners' performance in an experimental setting (e.g. task type, motivation, attention) makes it difficult to isolate the effects of single variables experimentally (Jilka et al. 2007a; Jilka 2009). Therefore, the following sections will focus on external rather than internal learner differences to describe differences in performance in the L2 perception test.

8.3 Between-group differences

Differences between sub-groups of L2 listeners are expected to be primarily related to the learners' first language and to the extent and quality of experience with the target language. *Language-specific patterns* in L2 speech perception are expected to be due to the influence of L1 sound patterns and can be explained by a comparative description of the learners' first language and the target language. Learners with the same first language are expected to exhibit similar patterns of sound perception in L2. However, the description of patterns in L2 perception may also reveal some perceptual patterns that are shared by learners from different L1s and that may be described as *language-universal* aspects of L2 perception. These cross-linguistic tendencies are discussed in chapter 4, 9 and 11.

To account for individual differences in test performance, *learner-related factors* such as the learners' language experience, age, training, language aptitude or "talent" but also the learners' individual strategies and hypotheses about the L2 sound system have to be considered. Many studies have investigated the influence of learner-related factors on the degree of "foreign accent" in L2 speech production (e.g. Piske, MacKay & Flege 2001; Moyer 2004; Ioup 2008). Three major types of learner-related variables have been studied in such works: (1) characteristics of the learners' *first language*, (2) *experience* in the target language, and (3) the *age* at the onset of learning. But even in studies where the factors of L1, age, and language experience were controlled (e.g. Bohn & Flege 1990) considerable individual differences between subjects could be observed, indicating that more than these learner-related variables must be taken into consideration to account for individual differences in performance in L2 speech (Bohn 1995, 1998; Markham 1997; Bongaerts et al. 1997; Moyer 1999, 2004).

For L2 *perception*, the same factors as for speech production have to be taken into consideration. However, the relative strength of influence of learner-related variables on performance in L2 speech perception may differ. We can expect that characteristics of L1 and

L2 as well as length and extent of L2 experience will have an influence on the perception of L2 sounds, while other factors such as age, general experience in language learning, language aptitude, bilingualism or motivation may be less influential. Moreover, individual and temporary strategies and hypotheses are observed in perception that indicate individual variation in vowel categorization but also to instances of hypercorrect responses in the test setting (see section 1.3.2; for evidence in the current data, see section 10.9.4 and 12.5).

8.4 The influence of L1

Numerous studies have reported that L2 learners differ in their perception of L2 sound structures according to the sound patterns of their native language. We can therefore assume that the learners' L1 will have a major influence on their performance in the test.

Table 8.1 shows the identification scores (*id_wrong*) of all L1 sub-groups (number of subjects N, mean, standard deviation (SD), minimum and maximum and percentiles).

	N	*mean*	*median*	*SD*	*min*	*max*	*percentile 25*	*percentile 50*	*percentile 75*
Albanian	12	.35	.34	.14	.14	.64	.24	.34	.45
Arabic	10	.38	.46	.20	.09	.63	.13	.46	.52
English	14	.47	.46	.12	.26	.70	.37	.46	.59
Farsi	4	.53	.54	.18	.30	.73	.34	.54	.70
Hungarian	26	.12	.09	.09	.03	.40	.06	.09	.17
Mandarin	6	.52	.57	.16	.21	.68	.41	.57	.63
Polish	31	.52	.53	.12	.15	.69	.49	.53	.61
Romanian	12	.31	.33	.11	.06	.47	.23	.33	.37
SerBoCroatian	33	.31	.29	.13	.09	.56	.21	.29	.41
Turkish	24	.32	.32	.12	.11	.51	.22	.32	.40

Table 8.1: Percentage wrong identifications per language (number of subjects, mean, median, standard deviation, minimum, maximum and 25-, 50- and 75-percentile)

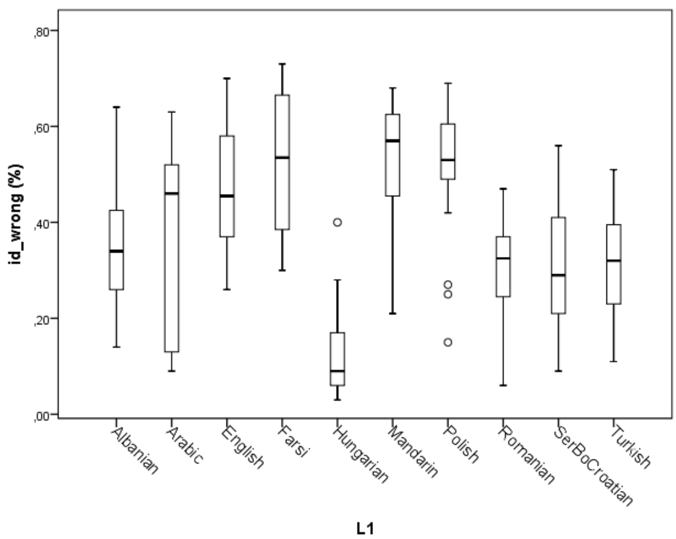

Figure 8.1: Boxplot diagram: Percentage of wrong identifications for each language sub-group

The boxplot diagram in Figure 8.1 represents the range of id_wrong scores in percent for each of the language sub-groups.

To compare the variability of id_wrong means for all language sub-groups, an analysis of variance (ANOVA) was performed. The overall id_wrong scores (percentage of wrong identifications) of 162 subjects from eight different L1s were included in the ANOVA procedure[65]. The ANOVA results reveal a significant effect of L1 on overall performance in the perception test: $F(7, 154) = 23.41$, $p < .001$. In a post-hoc test, *pair-wise comparisons* of all eight language-groups were carried out with $p < .05$ as criterion for significance (Bonferroni correction) to identify those groups that vary significantly from each other. The post-hoc procedure compares all combinations of language sub-groups with respect to significant differences in the listeners' test performance (see Table 8.2). For some language pairs, a significant effect of L1 on identification scores was found ($p < .05$[66]), while other languages do not differ significantly with respect to the subjects' performance in the test: e.g. SerBoCroatian (SBC) subjects differ significantly from Polish and English; SerBoCroatian subjects (with a mean error rate of 31%) perform significantly better than the Polish (mean error rate 52%) or English subjects (47%) but show no significant differences in error rate compared to Turkish, Arabic, Albanian, Romanian and Hungarian natives.

	Alb	Arab	Engl	Hung	Pol	Rom	SBC	Turk
Alb								
Arab								
Engl								
Hung	<.001	<.001	<.001					
Pol	.002			<.001				
Rom			.047	.001	<.001			
SBC			.003	<.001	<.001			
Turk			.011	<.001	<.001			

Table 8.2: *p*-values for language pairs that differ significantly ($p < .05$) from each other in percentage of wrong identifications. For all other language pairs no significant effect was found

A comparison of mean id_wrong scores is presented in Table 8.3. The results presented in Table 8.3 are read e.g. as "Albanian L2 listeners show less wrong identifications than Arabic, English or Polish listeners but more than Hungarian, Romanian, SerBoCroatian und Turkish participants". Turkish subjects show significantly better results than Polish and English subjects but perform significantly worse than Hungarian subjects. Hungarians perform

[65] Farsi and Mandarin subjects were excluded due to the small size of these samples.

[66] *p*-values higher than .01 were found only for Romanian-English (.047) and English-Turkish (.011). All other pairs showed *p*-values below .01.

significantly better than all other subjects, a fact that may be explained by the complex vowel system of Hungarian which is very similar in size and structure to German but also by the high number of advanced learners in the Hungarian sample.

	Alb	Arab	Engl	Hung	Pol	Rom	SBC	Turk
Alb		+	+	-	+	-	-	-
Arab	-		+	-	+	-	-	-
Engl	-	-		-	+	-	-	-
Hung	+	+	+		+	+	+	
Pol	-	-	-	-		-	-	+
Rom	+	+	+	-	-		-	+
SBC	+	+	+	-	+	+		+
Turk	+	+	+	-	+	-	-	

Table 8.3: Comparison of language groups with respect to differences in L2 perception performance, i.e. percentage of wrong identifications, p <.05 (/+/ lower id_wrong, /-/ higher id_wrong)

It is commonly assumed that the rate of id_wrong scores varies basically as a function of language experience and proficiency in L2 but is also conditioned by the listeners' native vowel system, more specifically by L1 system size of the learners' native language (e.g. Iverson & Evans 2007) and the relevance of specific dimensions for phonological contrasts. However, the data show that no direct correlation between the number of vowel phonemes in the listeners' L1 and their success in L2 perception can be postulated, as e.g. the example of the Arabic sub-sample shows: Arabic is reported to have a very small vowel system[67], but the rate of wrong identifications does not differ significantly compared to other languages apart from Hungarian. This example shows that a mere comparison of id_wrong scores between language sub-samples is too simplistic when the listeners' L1 system and their level of proficiency and experience in L2 are not considered.

It is important to note that the differences between language sub-groups described here refer to a mere *quantitative* comparison of the participants' error scores in the test procedure. An interpretation of these results does not allow for undifferentiated general statements like "vowel identification in German is easier for SerBoCroatian than for English native speakers". The results merely indicate that the English natives participating in this study were less successful in vowel identification than the SerBoCroatian participants. This can be explained partly by the specific constellation of the language samples, since the samples vary not only in number of subjects but also in the relative number of beginners and advanced learners ("level of proficiency", see below) and with respect to the subjects' length of

[67] for a different account of the vowel system of e.g. Egyptian (Cairene) Arabic, see section 8.4 and 4.7.6.

language contact and duration of formal instruction. Since the variable “level of proficiency” or “language experience” interferes with most other person-related variables and was only controlled in the Hungarian and the Polish sample (a group of beginners vs. a group of advanced learners who participated in a teacher education program for German were tested), a comparison of between-group differences with respect to the subjects’ L1 and their success in the test does not allow for general conclusions for all native speakers of a given language.

The qualitative analysis presented in the following chapters will provide more insights into the relative degree of difficulty of specific vowel qualities, though the same restrictions with respect to sample size are valid there. However, even if the learners’ L1 has a significant effect on L2 vowel identification, several other learner-dependent variables have to be considered to account for the enormous individual variability of L2 learners’ performance in the perception test.

8.5 Sex

105 of the 173 participants were female and 61 male, 7 persons did not indicate their sex. The statistical information about the id_wrong scores of male and female subjects is represented in Table 8.4 and is visualized in the box plot diagram in Figure 9.2.

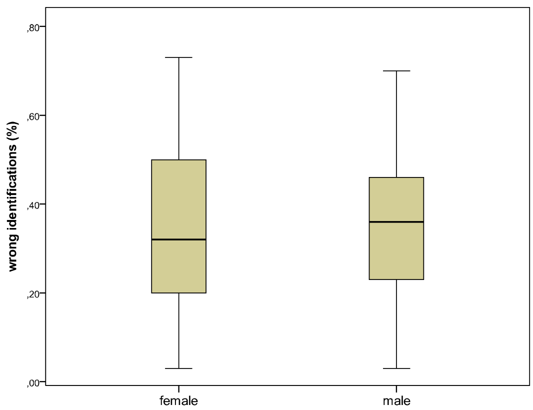

Figure 8.2: Wrong identifications (%) by 105 female and 61 male L2 listeners (see Table 8.4)

		male	female
N		61	105
mean		.35	.34
median		.36	.32
standard deviation SD		.17	.18
minimum		.03	.03
maximum		.70	.73
percentile	25	.21	.20
	50	.36	.32
	75	.46	.50

Table 8.4: Comparison of wrong identifications (%) by female and male L2 listeners (N=166)

A *t*-test showed no significant effect of sex on the overall percentage of wrong responses (p = .95).

8.6 Level of proficiency

Participants who took part in a German course at the time of the test had to indicate their current course level. 149 subjects were attending formal instruction at the time the experiment took place. Moreover, for all subjects, the level of proficiency was rated by the experimentator[68]. Thus, two variables concerning the level of proficiency were obtained concerning (a) the learners' *course level* (N = 149), and (b) the *estimated level of proficiency* (N = 173).

The correlation between the learners' *course level* (variable a) and their success in the test was tested for the subgroup of those participants who indicated a course level (N = 149). The correlation coefficient between the two variables (Spearman's *rho* = -.229, p = .005) is negative, in other words, with higher course levels the performance in the test improves, i.e. the percentage of wrong identifications decreases.

Based on the estimated *level of proficiency* (variable b), the sample was divided into a group of 83 beginners and a group of 90 advanced learners. The group of advanced learners was further divided into two subgroups of 78 advanced and 12 very advanced subjects. Very advanced persons were identified after having completed the test and can be described as persons with very good test results, a very high oral proficiency and fluency in German and rich language experience (long length of residence and long length of instruction).

Clear differences are observed between these three groups: A mean of 42% wrong identifications is observed with beginners, 32% with advanced learners, and 11% with very advanced learners (see Figure 8.3 and Table 8.5).

% wrong	N	mean	SD	min	max
beginners	83	42	18	0.06	73
advanced	78	32	16	0.03	63
very advanced	12	11	0.06	0.03	23
all L2 listeners	173	35	18	0.03	73

Table 8.5: Wrong identifications (%) for all vowel categories for beginners, advanced and very advanced learners (number of subjects, mean, standard deviation, minimum and maximum)

[68] Participants of a course up to A2 were rated as beginners, all others as advanced learners.

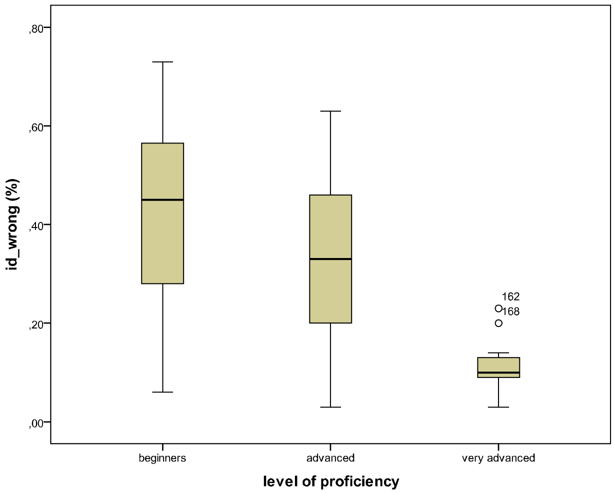

Figure 8.3: Id_wrong scores for beginners, advanced and very advanced learners

An ANOVA showed that the three groups differed significantly with respect to the subjects' level of proficiency in L2 ($p < .001$) and their success in the perception test. Not surprisingly, the data confirms the assumption that L2 vowel identification improves with general proficiency in L2.

8.6.1 Beginners vs. advanced learners

Although advanced learners generally show better identification scores than beginners, there are however some beginners who show a very accurate perception of German vowel categories with a minimum of only 6% wrong identifications (see Table 8.5).

It is expected that this effect may be somewhat different with the universally marked class of front labial vowels. To compare the mean id_wrong scores for labial vowels of all three groups of proficiency (beginners, advanced and very advanced learners), an ANOVA was performed. The results reveal a significant effect of level of proficiency on overall success for front labial vowels ($F(2, 170) = 18.93$, $p < .001$).

% wrong front labial Vs	N	mean	SD	min	max
beginners	83	51	25	0.04	96
advanced	78	38	21	0.00	79
very advanced	12	11	0.07	0.03	25
all L2 listeners	173	42	25	0.00	96

Table 8.6: Wrong identifications (%) for front labial vowels (/y ʏ ø œ/) for beginners, advanced and very advanced learners (number of subjects, mean, standard deviation, minimum and maximum)

Interestingly, a comparison of the id_wrong scores of (/y: ʏ ø: œ/) for the three levels of proficiency shows that the mean percentage of wrong identifications as well as the standard deviation is generally higher for beginners and advanced learners but not for very advanced learners (see Table 8.6). In other words, difficulties with front rounded vowels are observed even with very advanced learners and seem to persist in more advanced stages of L2 German acquisition.

8.6.2 Very advanced learners

The group of very advanced learners consisted of 12 persons showing very high oral proficiency, very good test results and a high amount of L2 German language experience. For this group, a mean percentage of 11% wrong identifications is observed (min 3%, max 23%). Very advanced learners differ significantly in their results from beginners and advanced learners. To account for these differences in performance, three major factors seem to be of relevance: A comparison of the variables age of learning (AOL), length of residence (LOR) and length of learning/instruction (LOL) of the 12 very advanced learners as compared to the rest of the sample showed significant differences with respect to all three variables.

The mean *length of residence* (LOR) was 16 years for the very advanced learners (vs. an average of 1.5 years of residence for the rest of the sample, N = 158). Thus, with respect to length of residence, very advanced learners differ significantly from the rest of the sample ($p < .001$).

The *age of learning* (AOL) is another significant difference between the two groups ($p = .004$). Very advanced learners started to learn German at an average age of 13 years, while for the rest of the sample the mean age of onset was 19 years.

The mean *length of formal instruction* for the very advanced group was 7 years, which differs significantly from the average length of instruction in the non-advanced group (only 3 years, $p = .030$).

To summarize, the results for the group of very advanced learners show that early beginning of learning, length of residence and length of formal instruction appear to contribute substantially to successful perception in L2, an effect associated with the learners' high amount of overall language experience with the target language.

8.7 Language experience

In the past decades, *language experience* has been accepted as one of the most influential factors on success in L2 learning. A quantification of language experience is however difficult and no standardized definitions exist for "experienced" vs. "inexperienced" L2 learners. Several studies hypothesize that the *length of residence* (LOR) in an L2 speaking surrounding will predict the extent of language experience. Therefore, many empirical studies collect data about the biographic and social background of the participants to draw conclusions on a person's language experience.

In the questionnaire of the present study, information about quality and quantity of L2 German language experience in terms of several variables was collected: The participants were asked to indicate (1) their *age of beginning* of learning German (AOL), (2) the length of *formal instruction* in German and other foreign languages, (3) the *length of residence* in a German speaking country (LOR), (4) *personal contacts* to German speaking persons, and (5) whether they had received *pronunciation training* during their German classes. The effect of these variables will be discussed in this section.

8.7.1 Length of residence

Length of residence (LOR) has been claimed in several studies to reflect the amount of L2 input. However, one has to be aware that the length of residence in a German speaking country neither gives us direct information about the quantity or quality of German input and language experiences nor about the learners' level of proficiency in the target language and does moreover not directly reflect the frequency of use of L2 and L1.

Significant success in L2 perceptual learning has been reported for late (adult) learners after 6-12 months of target language immersion (Best & Tyler 2007). Some studies report perceptual benefits of language experience for L2 listeners already after 6 months in an L2-speaking country (e.g. Aoyama et al. 2004), while other studies have defined the critical span of length for experienced learners with 2, 3, 5 or even more years of residence in an L2 environment. Significant effects of length of residence on foreign accents between groups of learners who had lived in the United States for averages of 1.1 and 5.5 years could not be observed (Flege 1988a; Flege, Yeni-Komshian & Liu 1999; Piske, MacKay & Flege 2001). As other studies had also demonstrated, experience of more than 5 years seems to involve no significant additional effect of learning at the phonetic, phonological, morphological or syntactic level (Patkowski 1980, 1990; Johnson & Newport 1989; Bohn & Flege 1990). However, the significance of LOR as a factor for improvement in L2 performance in production or perception could not be proved straightforward (Moyer 1999). It is only given if learners receive a substantial amount of *native speaker input* (Flege & Liu 2001). Therefore, LOR itself is considered to be only one of several factors that can provide an index on the amount or nature of L2 experience together with the use of L2 and L1 in everyday life and formal instruction in the target language.

In the present study, the mean length of residence in a German speaking country was 2.5 years (SD 5.97, min .0, max 36 years). 17% of the subjects had never stayed in a German speaking country. The median was 0.25 years, i.e. around 2.5 months of residence in a German speaking country. This is largely explained by the fact that many learners who were tested in Austria were contacted during their German classes that they attended in the first few months of residence in Vienna.

The length of residence in a German speaking country for 170 subjects (missing values for 3 persons) was correlated with the percentage of wrong identifications in the perception test. A Pearson correlation between the subjects' performance in the test (id_wrong scores) and LOR showed a very weak effect (r = .124, p = .108).

A regrouping of subjects into two groups was carried out: One sub-sample consisted of 117 subjects who spent less than 6 months in a German speaking country and was compared to a group of 53 persons with more than 6 months LOR. A t-test which was performed to compare the mean id_wrong scores of the two sub-groups (36% for the short-term residence group vs. 32% in the long-term group) showed no significant differences in their success in the test (p = .165).

In a re-analysis of the data (N = 170), the influence of *long-term residence* (> 5 years) was analysed (cf. Bohn & Flege (1990) who compared two groups of L2 learners: one with 6 months of residence and a group of subjects with more than 5 years length of residence).

29 participants had never lived in a German speaking country and were excluded from this analysis. The remaining 141 persons were divided in a long-term residence group of 24 persons who had lived more than five years in a German speaking country (5 to 36 years) and a short-term residence group of 117 persons who spent less than 5 years in a German speaking country. The mean id_wrong scores for these two groups were compared in a t-test. Assuming that variances are equal (Levene-test, p = .234), no statistically significant differences between groups were found at an α-level of .05 (p = .099)[69], i.e. long-term residence (more than five years) does not have a statistically significant effect on the learners' performance in the perception test, a result that is also supported by the qualitative analysis of the data. However, even if long-term residence would show a statistically significant effect, we could not assume that length of residence is the only causal factor behind the differences between groups. Other factors such as amount and quality of L2 input, general language awareness and experience in foreign language acquisition, motivation and personal involvement amongst

---

[69] Due to the large differences in sample size for the two sub-groups these results have of course to be treated with care.

others have to be taken into account to describe the learners' experience and proficiency in L2.

8.7.2 Formal instruction

Length of formal instruction is another relevant factor for language experience in L2. Participants had to indicate the length of formal instruction in German in the questionnaire. The mean lenght of formal instruction was 3.5 years (SD 4.1). 11 persons had received two months of formal instruction, 59 persons (34%) received less than 6 months instruction and only 11 persons had more than 10 years of language instruction (max 21 years). Thus, the length of formal instruction is a variable that is not normally distributed. The data were therefore recategorized into six groups (see Table 8.7). A correlation analysis was run to determine the relationship of length of formal instruction and id_wrong scores.
A correlation of length of formal instruction in German and performance in the test shows a weak effect of formal instruction (Spearman's *rho* = -.383, $p < .001$), i.e. the more instruction the learners had received, the lower the id_wrong scores in the test.

length of formal instruction	frequency	%	cumul %
< 6 months	59	34.1	34.1
0.5 -1 year	21	12.1	46.2
1-3 years	25	14.5	60.7
3-5 years	21	12.1	72.8
7-10 years	36	20.8	93.6
more than 10 years	11	6.4	100.0
all L2 listeners	173	100.0	100.0

Table 8.7: Duration of formal instruction in L2 German (N = 173)

In the Hungarian and the Polish sample and also in parts of the SerBoCroatian sample, the group of advanced learners consisted of students who were enrolled in a German teacher education program in the second year of their studies (N = 26). A comparison of the mean id_wrong scores of the group of advanced students participating in a teacher education program for German (22% id_wrong) and non-participants (40% id_wrong) shows a clear difference between these two groups. While for the group of students participating in a German teacher education program a significant medium effect of length of instruction can be observed (Spearman's *rho* = -.452, $p < .05$), for the group of non-participants no such effect can be found (Spearman's *rho* = .164, $p = .070$). The effect of formal instruction for the teacher students can most probably be explained by their generally higher level of proficiency and by the total length of formal instruction for students enrolled in a German teacher education program.

8.7.3 Informal language experience

A further item in the questionnaire referred to the participants' *personal contacts* with German native speakers. 71 persons declared to have no personal contacts (their mean id_wrong score was 39%), while 92 participants had personal contacts with German native speakers (34% id_wrong). Information on the length and extent of personal contacts was not collected in the questionnaire. A t-test showed no significant influence of personal contacts on the performance in the test (p = .084).

8.7.4 Age, onset, and duration of L2 learning

The relationship between the *age of learning* a second language and the performance in L2 perception and production is an issue of particular interest in many studies on second language acquisition. The learner's age at which they were first exposed to L2 is often referred to as "age of learning" (AOL) or "age of arrival" (AOA) for immigrant learners. Many studies investigated the relation of AOL or AOA and accent in L2 speech by relating L2 production data to the variables AOL or AOA (e.g. Flege, Munro & McKay 1995b; Flege, Yeni-Komshian & Liu 1999; Piske, Flege, MacKay & Meador 2002; Flege & MacKay 2004). This research line is strongly connected with the assumption that "late learners" acquiring L2 after early childhood typically show differences in L2 production and perception compared to "early" learners. It is commonly accepted that "early learners", who were exposed to L2 already in (early) childhood, are more successful than so-called "late learners", who started L2 learning in late adolescence or adulthood, an effect that is observed best in L2 pronunciation and perception. The view that "earlier is better", i.e. that early bilinguals generally succeed better in producing and perceiving an L2 than late bilinguals (see Long 1990 and Flege 1999, for reviews), is strongly supported by several studies (e.g. Munro, Flege & MacKay 1996; Baker, Trofimovich, Mack & Flege 2002), though some studies show that even adults can under certain conditions develop high proficiency in L2 and can achieve even native-like performance in L2 production (Bongaerts et al. 1997; Flege & MacKay 2004; Birdsong 2005). Similar patterns of results have also been found for perception of L2 vowels (e.g. Gottfried 1984; Ingram & Park 1997; Baker, Trofimovich, Mack & Flege 2002). Baker et al. (2002) examined the discrimination of English vowels by Korean learners and found that early learners showed better results than late learners but did not differ significantly from English native speakers.

Such age effects have been replicated in several studies and have been related to the idea of a critical period for L2 learning. In the last decades, Lenneberg's (1966, 1967) *Critical Period Hypothesis* (CPH) that claimed in its original version that a full command of a second language is only possible between the age of 2 to 12, has consequently received a variety of different interpretations. Implicitly or explicitly, many researchers and theorists accept some version of CPH to account for age-related differences in speech perception and production. There are, however, studies investigating the influence of age on language learning that falsified this hypothesis (see Flege & MacKay 2011, for a review) and showed that native-like patterns can be achieved even by adult language-learners.

However, it is worth noting that production and perception are usually not differentiated in discussions about the critical period hypothesis. Developmental studies with young infants show that the development of the two systems is tightly coupled and that a polymodal mapping of auditory and motor information (Kuhl 2000: 11854) is to be assumed rather than a model in which perceptual representations are specified in motor terms as previously assumed (cf. e.g. Liberman & Mattingly 1985).

However, when investigating the effect of AOA/AOL, the interaction with other factors such as length of residence (LOR) and use of L1 has to be considered (Piske, MacKay & Flege 2001). Flege, MacKay & Meador (1999) compared early and late Italian learners of English who differed in extent of L1 use (high vs. low use of L1) and observed that "early low" and "early high" groups showed better performance in discrimination tasks than late learners but did not differ significantly from English natives or one another. Flege & MacKay (2004) found that "early low" learners who seldom used their L1, did not differ from English natives in the perception of English vowel contrasts, but that the "early high" learners who continued to use their L1 differed significantly from native English listeners in perceiving English vowels. These results show a clear effect of AOA and L1 use, indicating that beginning to learn L2 in early childhood does not guarantee a native-like perception of L2 vowels. On the other hand, there are also late learners, who were found to perceive L2 English vowels accurately.

In the present study, the subjects were asked for their *age* at the time they had started to learn German (AOL), the *length of formal instruction* (LOL) they had received in German and about their *length of residence* (LOR) in a German speaking country (see discussion below). For those learners who were tested in Austria, AOL and AOA may differ if the learners received language instruction before arriving to Austria. Therefore, the age of first contact with German (here AOL) must be distinguished from the age of arrival (AOA) and the "length of learning" (LOL), i.e. the total duration of formal instruction a person had received in German.

172 participants indicated at which age they started to learn German (AOL) (see Figure 8.4). The mean age of beginning was 18.7 years (SD 7.5, median 19, modal 14). The youngest participant was first exposed to German at the age of two, the oldest began to learn German at the age of 46.

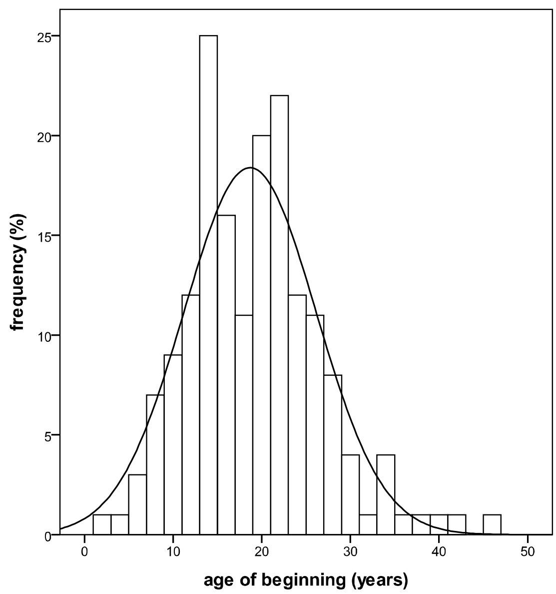

Figure 8.4: Age of beginning (mean 18.7 years, SD 7.5 years, median 19 years, modal 14 years).

A correlation of the variables AOL and id_wrong scores in the test revealed a weak effect of age with a Pearson's correlation coefficient of $r = .273$, $p < .001$.

Hence, it cannot be straightforwardly relied on the assumption that the older a person is when starting to learn German, the harder it is for him or her to categorize German vowels correctly. This proofs that the very general assumption that "earlier is better", i.e. the lower AOL or AOA the better, is too simplistic. Rather than reducing the relative success to the age factor, the quantity and quality of language *input* and language *use* have to be considered as crucial variables in the development of perceptual skills in L2 together with other language-related skills.

8.8 Experience in foreign language learning und metalinguistic knowledge

8.8.1 German as first foreign language

With respect to foreign language acquisition, 141 participants indicated that German is not their first foreign language acquired ("L3+"), while for 28 subjects German was their first foreign language, i.e. "L2" in a proper sense (N = 169). For only 2 subjects German was the only foreign language acquired so far. Most subjects had learned 2-3 foreign languages.

The mean error rate for subjects with German as their first foreign language ("L2") was 22% as opposed to 38% errors with "L3+" subjects who had acquired German after another foreign language. A t-test for equality of means shows a significant difference between L2 und L3+. (p <.001). That is, persons who had learned German as a first foreign language (L2) showed significantly less wrong categorizations (21% id_wrong) than persons who had acquired German after another foreign language (38% id_wrong for L3+).

It is however necessary to consider that this relation is closely related to the age of beginning (AOL) and the length of instruction (LOL) in German: Participants who had learned German as a first foreign language started at a mean age of 9 years, while the acquisition of German for the L3+ group began at a mean age of 21 years. Subjects with L2 German had received formal instruction in German for an average of 10 years while the mean length of instruction for L3+-subjects was only 2 years. These differences explain the results of the t-test.

To summarize, learners who learned German as L2 (before learning another language) perform significantly better than those who learned German after another foreign language. However, this effect is strongly connected with the total length of instruction in German and the age at the onset of learning.

8.8.2 Bilingualism and multilingualism

It is a commonly accepted view (or is it a myth?) that persons with more experience in language learning and especially bilinguals who are used to listen to more than one language from early childhood on have some advantage in learning another language compared to monolingual persons, especially with respect to phonetic skills. This effect could be explained by the growth of language awareness or, more specifically, of *phonological awareness* (McBride-Chang 1995; Gillon 2004) induced by the acquisition of a second language sound system. Gut (2010) posits a larger repertoire of phonetic-phonological knowledge, phonological awareness and learning awareness and an increased cognitive flexibility that may facilitate the acquisition of phonological patterns in later acquired languages (L3 or Ln). However, the influence of a previously acquired foreign language (L2) on later acquired languages (L3 or Ln) is a research field that until recently has hardly received scientific interest (for a review, see Cabrelli Amaro 2012 and the special issue of the International Journal of Multilingualism 7/1, 2010).

The question that arises is whether persons with more experience in language learning would perform better in perceptual tasks related to phonetic skills, or more precisely, whether they generally have a better aptitude to discern sound contrasts irrespectively of the language concerned and whether the learning of foreign languages enhances phonetic skills in general?

To investigate the question whether language learning experiences could also improve perceptual abilities and the listeners' performance in L2 in general and their success in the identification task, a Pearson correlation test for the *number of foreign languages* acquired and the percentage of wrong corrections was performed. It showed that the number of foreign languages a person has learned has a rather weak effect on overall success in the identification task (r = -.164, p = .032). That is, the assumption that the more languages a person has learned, the lower the percentage of wrong answers will be in the identification task is too simplistic and cannot be confirmed in this form.

8.8.3 Phonetic instruction and pronunciation training

It was further considered whether *pronunciation training* and *phonetic knowledge* could have a positive effect on the success rate in the test that might be due to an increase of phonological awareness through instructions in phonetics (Piske 2008). 44 subjects indicated to have theoretical knowledge in phonetics, whereas 127 persons received no phonetic instructions (N=171).

A t-test shows that theoretical knowledge in phonetics has a significant effect on performance in the perception test decreasing the failure rate for 8%. The mean failure rate of persons with no phonetic instruction was 37% vs. 29% for persons with phonetic instructions.

Most of the subjects with *phonetic knowledge* are university students of German. Especially the sample of Hungarian, Polish and SerBoCroatian advanced learners consists of students in their second year of study participating in a German teacher education program. All of them had received an introduction to phonetics at the time the test was performed. Therefore, the sample of persons with phonetic knowledge was divided into two groups dividing persons studying German at university level (N = 26, 25 of them had phonetics) vs. persons not studying German (N = 121, 14 of them had phonetics). The two sub-groups of persons with phonetic knowledge differed with respect to their involvement in university studies of German. A t-test comparing the means of persons not studying German in 2 subgroups – with and without phonetics – showed no significant effect of phonetic instruction (p = .4), indicating that in fact it is rather the combination of high proficiency in German, formal instruction and phonetic knowledge that accounts for the better results of the group with phonetic knowledge.

84 participants declared that they had received some kind of *pronunciation training* during their language instruction (the intensity of phonetic training could not be quantified). 85 subjects indicated that they had never received any pronunciation training. A t-test comparison of the two groups showed no significant effect of pronunciation training (p = .168) on overall test results.

8.9 Conclusions – The interaction of learner-related variables

The learners' success in L2 vowel identification is assumed to depend substantially on their experience with the target language. Language experience has to be described in a quantitative and a qualitative dimension. In this chapter, person-related variables were analysed that are considered to be determinant factors for experience and proficiency in L2 in many studies on foreign language acquisition. However, the statistical analysis of learner-related variables showed that only some variables appear to have a significant effect on correct vowel identification in the perception test.

An ANOVA yielded a significant main effect of the listeners' *native language* (F (7, 154) = 23.41, $p < .001$) on L2 learners' performance in the perception test. A comparison of the L1 sub-groups showed significant id_wrong score differences between some but not all languages (see Table 8.2 and Table 8.3), though differences in sample size and sample structure may account for some of these quantitative differences between language groups.

A significant effect on success in the perception test was found for the subjects' self-estimated *level of proficiency*. The three sub-groups of beginners, advanced and very advanced learners differed significantly from each other with respect to identification scores in the test. A comparison of the id_scores for these three groups showed that beginners had a higher failure rate in the test than advanced learners and that very advanced learners were significantly more successful than learners at all other levels. The group of very advanced learners differed in three variables from the rest of the sample: length of residence, age of learning and length of formal instruction, all together contributing to overall success in the perception task.

The effect of *age of learning* (AOL), i.e. the age at which learners had first contact with the target language German, is discussed as a crucial variable in SLA literature. An early extensive exposure to L2 in an environment providing rich L2 input (as is typically given for children who immigrate in early childhood) as well as the extent of use of the L1 seem to contribute to learners' success in learning L2 sound patterns (Flege & Liu 2001). However, the very general assumption "the earlier the better", i.e. the lower the age of learning (AOL) or AOA (age of arrival) the better the achievement in L2 will be, could not be fully proved by the data. In the the present study, only a weak correlation of AOL and performance in the perception task was found ($r = .273$, $p < .001$), given that the present sample consisted mainly of participants who started to learn German after early childhood (mean age of learning onset 18.7 years). Rather than age it is the extent of *language experience* and the quality of input that seems to be the crucial factor determining the learners' success in L2 production and perception. *Length of residence* is discussed in the literature as one important factor contributing to language experience. In the present study, length of residence in a German speaking country, duration of formal instruction, informal experience and pronunciation teaching were considered to determine the learners' experience in L2 German. However, no significant effect of length of residence on the listeners' performance in the perception test could be found.

Duration of formal instruction had a weak effect on the subjects' performance in the test (Spearman's $rho = -.383$, $p < .001$), a stronger effect was observed for the group of university students of German (Spearman's $rho = -.452$, $p < .05$).

Neither *informal experience*, i.e. personal contacts to German native speakers, nor *pronunciation training* nor general language learning experience or "*multilingualism*", i.e. experience in more than one foreign language, nor early bilingualism showed significant effects on the listeners' performance.

The sub-group of *very advanced learners* showed a very low failure rate in the perception test and differs significantly with respect to length of learning, age of beginning and length of residence from the rest of the sample. The high performance rate for this group of learners indicates that these three factors contribute substantially to correctness in perceptual categorization, indicating that a combination of formal and informal experience with the target language contributes to high performance scores in the perception task.

To conclude, the learners' native language, the learners' level of proficiency and extent of overall experience in L2 and duration of formal instruction in L2 are factors influencing the learners' success in the perception experiment. However, the interpretation of statistical results from experimental tasks in L2 requires awareness of the fact that personal skills and experiences, motivations and individual strategies cannot be determined in a quantitative sense, but nevertheless have to be considered to fully account for an individual's success in the acquisition of a foreign language sound system.

9 Cross-language Analysis

This chapter provides a summary of the study's results for non-native listeners from different language background and a general discussion of major tendencies in L2 perceptual substitutions observed across the language sub-groups. The data analysis is based on 45,794 valid responses by 173 L2 listeners from a total number of 46,694 possible response items. 900 cases had to be excluded because the participants either gave no answer or more than one answer. The full sample of 173 non-native participants will be referred to by the label "all L2".

The data show a considerable range of variation between categories, across languages and within the language sub-sets that will be discussed in this and the following chapters.

This chapter will discuss (1) L2 listeners' *difficulties* and the ratio of wrong and correct identifications with respect to (a) language-specific variation, (b) vowel-specific variation, and (c) context-dependent differences, (2) the listeners' *preferences* for specific response categories in a cross-language comparison and a vowel-specific comparison, followed by (3) a summary of *confusion patterns* and similarity scores for the full L2-sample, and (4) a comparison of *perceptual vowel maps* for each of the L2 listeners' native languages in form of two-dimensional *MDS representations*. This chapter provides an overview over major trends in L2 vowel identification. A detailed language-specific analysis will be presented in chapter 10. A comparison with the scores of the native German control group (for details, see section 10.2) will be offered if relevant.

9.1 Difficulties

Difficulties in perceiving a given category correctly are reflected by the number of wrong stimuli identifications for a given vowel category. For non-native listeners (N=173), the mean id_wrong score is 35.2% with a standard deviation of 18.1%, ranging from a minimum of 3% to a maximum of 73% wrong identifications. From a total of 45,794 valid L2 responses 35% (16,069 responses) were wrong and 65% (29,725 responses) were identified correctly. In other words, in more than a third of the cases, non-native listeners could not identify the input stimuli correctly.

The percentage of correct answers varies considerably across as well as within the language sub-groups. Table 8.1 summarized the percentage of wrong identifications for each of the language sub-groups. Figure 9.1 summarizes the rate of correct identifications for the "all

L2"-sample. A comparison of the range of variation between and within language sub-groups is presented in Figure 9.2.

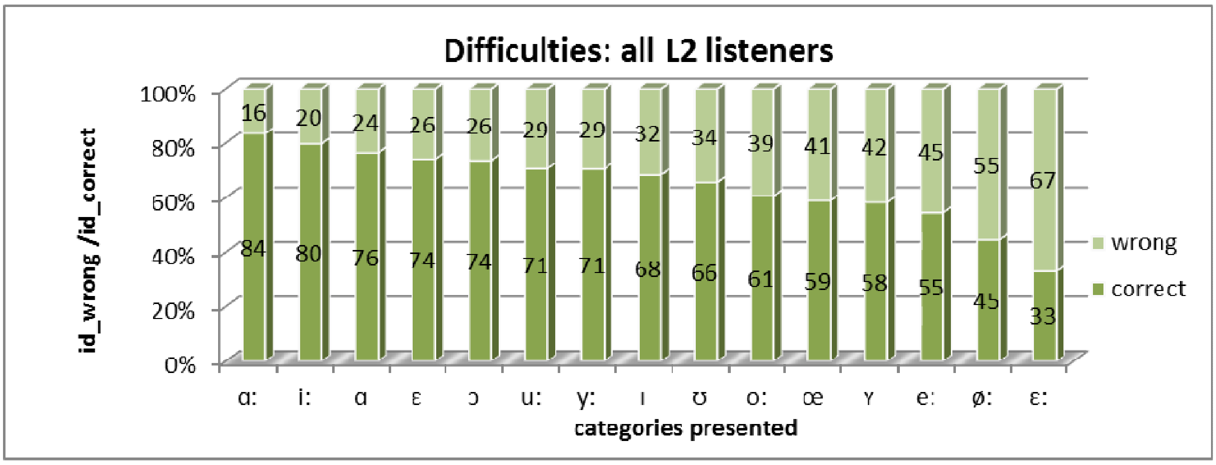

Figure 9.1: Percentage correct identifications in descending order for all L2 listeners (173 non-native listeners, 10 L1 sub-groups, 45,794 responses)

The results of the L1 German control group show that only 3.9% of the native listeners' responses were wrong (see section 10.2). The average percentage of 3.9% wrong responses in the L1 German group is mainly due to errors for /ɛ:/-stimuli (20.7% wrong, 79.3% correct), whereas a maximum of 99.1% correct answers is observed for /i:/ and /u:/. All other categories in the L1 German sample show id_correct scores above 92%, a rate that is comparable with listening experiments with native listeners in other studies, e.g. the classical study of Peterson & Barney (1952) or Hillenbrand et al. (1995).

9.1.1 Language-specific variation

The id_wrong scores for non-native listeners vary across languages as well as within a given language. In the L2 sample, the lowest percentage of wrong categorizations is obtained for Hungarian listeners (12% id_wrong), whereas Farsi (52.6%) and Polish (52.1%) listeners show very high error rates followed by Mandarin (48.8%) and English (48.2%) listeners.

The box-and-whisker plots in Figure 9.2 compare the inter- and intra-language variation for all language sub-groups tested. The boxes represent the interquartile range (25% - 75%) of the listeners' results, i.e. the range of *id_wrong* scores that is obtained by 50% of the listeners of a given language sub-group in the sample.

Least difficulties are observed for Hungarian listeners, as indicated by the very low median, whereas Mandarin, Polish, Farsi and Arabic show high median values. The interquartile range is highest for Arabic, indicating a very high range of inter-personal variability, and lowest in the Hungarian sample.

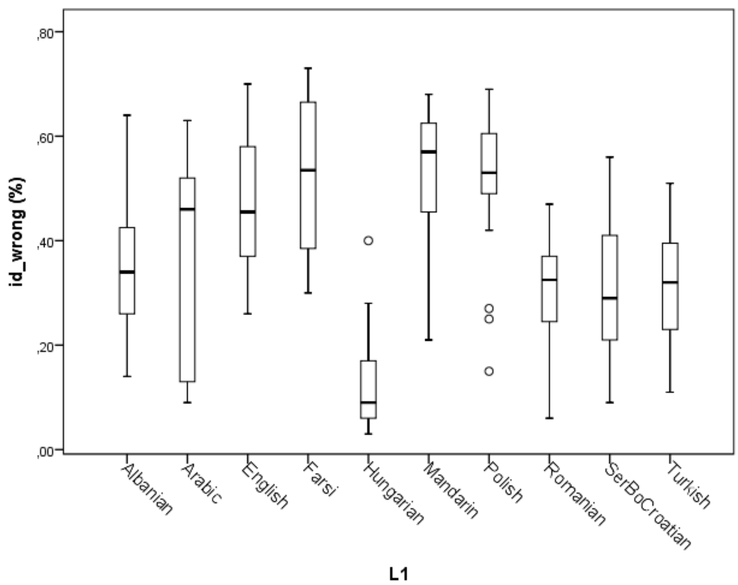

Figure 9.2: Boxplots representing the percentage of wrong identifications per language sub-groups

It is important to underline that the percentage of wrong and correct identifications is strongly influenced by learner-related characteristics (see chapter 8) and the specific composition of the language sub-samples that vary with respect to the listeners' language experience and level of proficiency in L2 German; the sub-samples were not homogenous with respect to the number of participants nor the ratio of beginners and advanced learners: e.g. the high median and high interquartile range in the Polish sample is due to a higher number of beginners than advanced learners in the sample. Thus the results do not allow general statements in the sense of "the acquisition of the German vowel system is more difficult for native listeners of language X than language Y". The effect of the listeners' native language on the relative degree of difficulty is hardly quantifiable and can – if at all – only be described by a qualitative comparison of the listeners' L1 and the L2 German vowel system, which is presented for each of the language sub-samples in chapter 10.

9.1.2 Category-specific variation

The number of wrong identifications varies considerably across vowel categories. A category-specific comparison is provided for the full sample of 173 non-native listeners in Figure 9.1, which summarizes the id_wrong and id_correct scores for each of the categories tested. The results indicate that some German vowel categories are particularly difficult to identify correctly, whereas others are comparatively robust in L2 perception. Moreover, a considerable range of intra-category variation is observed (see Figure 9.3).

The *id_wrong* rates are highest for /ɛ:/ (66.9%), /ø:/ (55.4%) and /e:/ (45.4%). High id_wrong scores above 40% are moreover observed for front rounded categories, with the only exception of /y:/ that shows an error rate of only 29.4%.

German a-qualities are particularly stable in perception as are /i:/ (79.8% correct), /ɛ/ (74.2%), and /ɔ/ (73.7%). These categories can therefore be considered as relatively easy in L2 perception, whereas front rounded vowels (except /y:/) and the contrast /e:/ - /ɛ:/[70] are comparatively more difficult to identify correctly.

The box-and-whisker diagram in Figure 9.3 shows the range of intra-category variation for each of the vowel categories for all 173 non-native listeners. The lowest range of variation within the interquartile range is observed for /i:/, whereas the median and the interquartile range for /ɛ:/ (<ä>) are generally higher than for all other categories. A high median together with a generally high interquartile range of wrong identifications are moreover given for /ø:/, but also for /e:/ and /o:/, whereas for /ɑ/, /ɑ:/, /ɛ/, /ɔ/ or /y:/ a high median and less intra-category variation indicated by the low interquartile range are observed.

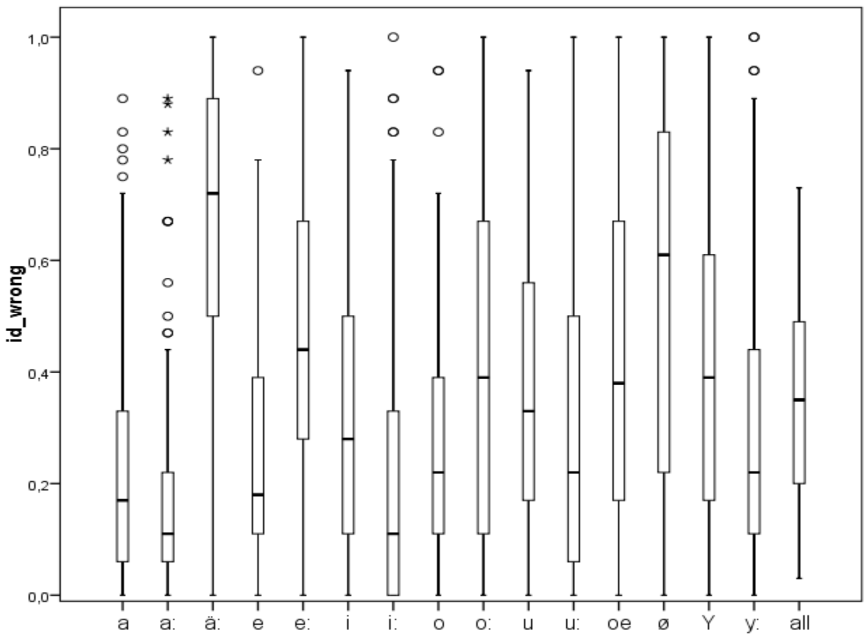

Figure 9.3: Range of intra-category variation of id_wrong scores for 173 non-native listeners

70 The qualitative interpretation of the confusion matrix (see section 9.3) will show that difficulties with /e:/ are not only due to the contrast with /ɛ:/ but that there are other interfering categories, e.g. /i:/, that increase the error rate for /e:/.

To compare the vowel-specific range of variation across language sub-groups, the id_wrong scores for each of the languages under investigation are summarized in Table 9.1, representing the percentage of wrong identifications (*id_wrong*) per input category for each language sub-group. The column "all L2" indicates the percentage of wrong identifications per category for all 173 non-native listeners. "German" refers to the L1 German control group. The last row in Table 9.1 indicates the mean percentage of wrong responses for each of the language sub-groups.

% wrong	Alb	Arab	Engl	Farsi	Hung	Mand	Pol	Rom	SBC	Turk	all L2	German
ɑ	38	15	27	25	15	41	38	28	12	20	24	3
ɑ:	15	25	36	47	6	25	18	10	13	14	16	2
ɛ:	64	54	54	80	46	70	80	69	69	82	67	21
ɛ	41	21	28	26	10	51	36	25	21	23	26	6
e:	32	56	57	69	21	50	76	28	42	37	45	7
ɪ	36	31	40	57	8	45	37	54	21	41	32	2
i:	18	27	36	44	1	28	35	16	15	14	20	1
ɔ	35	14	33	32	9	54	37	34	15	33	26	2
o:	26	51	47	67	11	41	65	26	40	33	39	2
ʊ	37	41	60	53	11	61	45	37	24	30	34	2
u:	22	51	65	58	3	40	39	20	28	17	29	1
œ	39	49	51	51	7	59	79	21	34	34	41	2
ø:	42	55	72	83	13	57	84	45	68	44	55	2
ʏ	53	36	63	51	15	64	61	32	34	39	42	4
y:	25	38	53	43	6	46	51	21	27	11	29	2
mean	*35*	*38*	*48*	*53*	*12*	*49*	*52*	*31*	*31*	*31*	*35*	*4*

Table 9.1: Wrong identifications (%) for each input category and each language sub-group

The results for the L1 German control group show that best results are obtained for /i:/ and /u:/, both with 99.1% correct answers by L1 German listeners, whereas only 79.3% correct responses are obtained for /ɛ:/. For all other input categories, the error rate of L1 German listeners lies below 7.4% (for a detailed vowel-specific analysis, see chapter 11). The average percentage of 3.9% wrong responses in total for the L1 German sample is mainly due to errors with /ɛ:/-stimuli: 20.7% of all [ɛ:] occurrences were incorrectly identified, and 79.3% were identified correctly.

The spider chart in Figure 9.4 visualizes the range of language-specific variation for German vowel categories for each language sub-group (see Table 9.1) to provide more insights into inter-category differences and commonalities.

The diagram shows that the error rate for /ɛ:/ but also for /e:/ is considerably higher than for other categories in all language sub-groups. Moreover, the error rate for /ø:/ but also /o:/ is higher than for other categories in most language sub-groups, an effect that is due to the interaction of several factors. The error rate is particularly low in all sub-samples for /i:/ and German a-vowels.

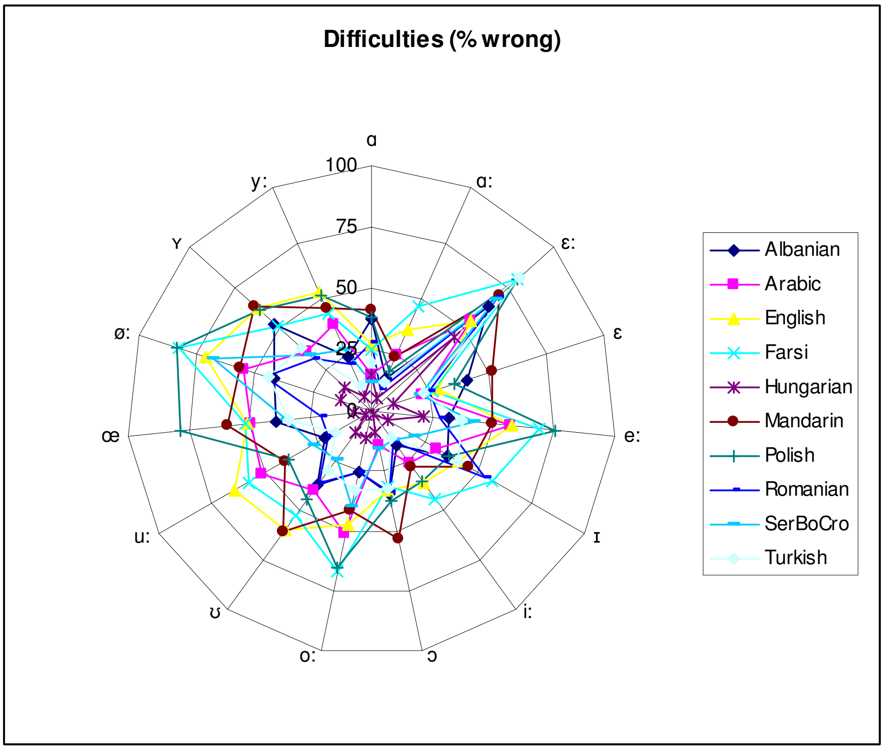

Figure 9.4: Percentage wrong identifications in cross-language comparison

Figure 9.5 compares the range of variation for each of the tested vowel categories and the id_wrong scores within a given language sub-sample. The box-and-whisker diagrams show interesting differences in the range of inter-speaker variation and inter- and intra-language variation. The effect of the vowel category on the error rate is clearly evident for /ɛ:/ or /ø:/, whereas for other categories such effects are less clearly recognizable.

To summarize, the percentage of wrong categorizations is considered to be an indirect indicator for the relative difficulty of a given vowel category. However, the degree of difficulty of German vowels for learners of different languages cannot be compared directly by the id_wrong scores obtained. For a correct interpretation of the values obtained it is important to remember that the number of subjects and their level of proficiency and experience in L2 German vary for each of the language sub-samples. Therefore, a direct inter-language comparison of the id_wrong scores of the language sub-groups is only of limited significance. Within a given language sub-sample, however, the proportions of difficulties observed for German vowel categories can be interpreted as a direct indicator for the relative difficulty of vowel categories for L2 learners with the respective native language. For a detailed discussion of the relative difficulty of vowel categories, see the vowel-specific analysis in chapter 11.

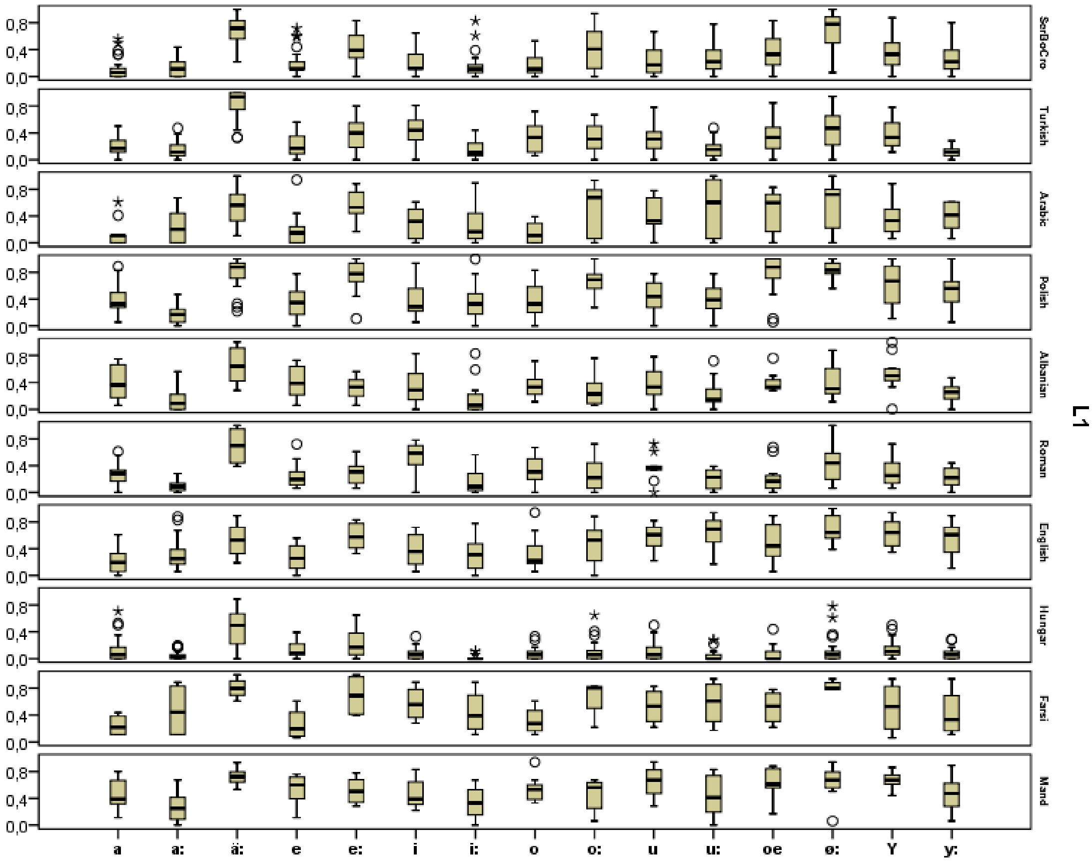

Figure 9.5: Comparison of vowel-specific variation in id_wrong scores across language sub-groups

9.1.3 Context-dependent variation

Acoustic cues for vowel identification are not only provided by spectral information of formant frequencies, their distance to F0 and intrinsic duration but also by information beyond the actual "segment" that is commonly referred to as context-sensitive variation due to coarticulation effects between neighbouring parts in the speech string. Vowel formants are sensitive to their consonantal context (Stevens & House 1963) and adjacent consonants do have an effect on the dynamics of formant frequencies in the CV- and VC-transitions of the signal. Context-sensitive variation caused by adjacent consonants results in more overlap of vowel formant frequencies between vowel classes. However, this type of variation can serve as disambiguating information in vowel identification as demonstrated by Strange, Jenkins & Johnson (1983) in experiments with various modifications of /*bVb*/-syllables, where parts of the signal were deleted or temporary information manipulated. Their results show that listeners use spectral information contained in initial and final portions of the signal to reconstruct the vowel's identity even when the vocalic part of the signal is actually deleted ("silent-center condition"), whereas the identification of isolated vowels showed to be equal or even worse than for vowels in context. These results indicate that context-sensitive variation does not really "disturb", but even can enhance the listeners' capacity to identify a given vowel.

In the present study, test items varied with respect to vowels and their consonantal context. The input stimuli consisted of /pVC/ non-words in experiment 1 and /CVtə/ in experiment 2, where the consonants varied systematically with respect to place and manner of articulation (see section 6.3). Figure 9.6 and Figure 9.7 summarize the *context-specific* id_wrong scores. While variation of post-vocalic context in experiment 1 does not cause significant differences in mean overall id_wrong scores, more variation is observed due to pre-vocalic context in experiment 2.

A detailed discussion of the influence of pre- and post-vocalic context would require a language-specific discussion of the data for each vowel category and each consonant category which is beyond the limits of this study. An analysis of the influence of pre- and post-vocalic consonants on the percentage wrong identifications is therefore offered here only for the full L2 sample but not for each of the language sub-groups.

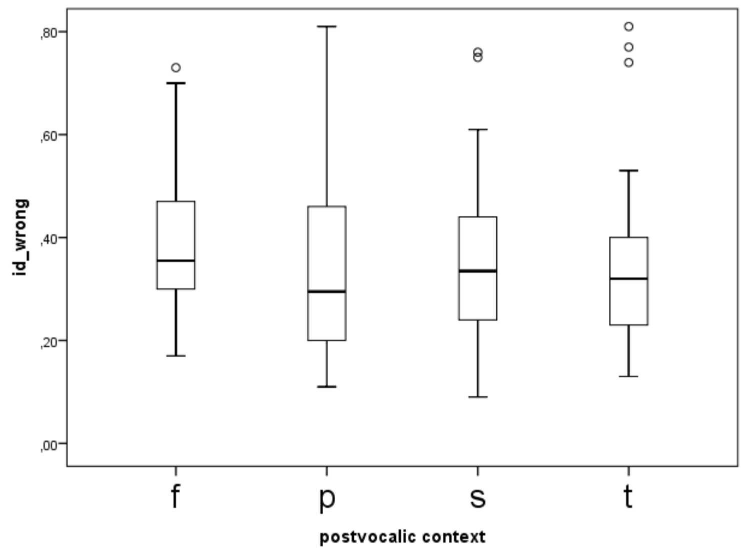

Figure 9.6: Percentage wrong identifications for post-vocalic consonant environments /pVC/ in experiment 1 (each of the 15 vowels occurring twice in 4 different post-vocalic contexts, 120 stimuli, 173 L2 subjects)

In experiment 1 (see Figure 9.6), where the pre-vocalic context is always /p/ and the post-vocalic context can be /p f t s/, the highest average id_wrong score is observed for post-vocalic /_f/ (mean .4, SD .14, min .17, max .73). However, no substantial differences to other groups were found: /_s/ (mean .35, SD .17, min .09, max .76), /_p/ (mean .35, SD .19, min .11, max .81), /_t/ (mean .35, SD .17, min .13, max .81).

Similarly, variation of pre-consonantal context does not cause considerable variation in id_wrong scores with the only exception of /j/ (see Table 9.2). The highest id_wrong score is observed in test items with pre-vocalic /j/ (mean .40, SD .14, min .14, max .70). In other words, in a cross-language comparison, the correct identification of German vowels seems to be most difficult in syllables with pre-vocalic /j/. Least variation is observed for pre-vocalic /t/, though this context does not seem to be significantly "easier" for L2 listeners.

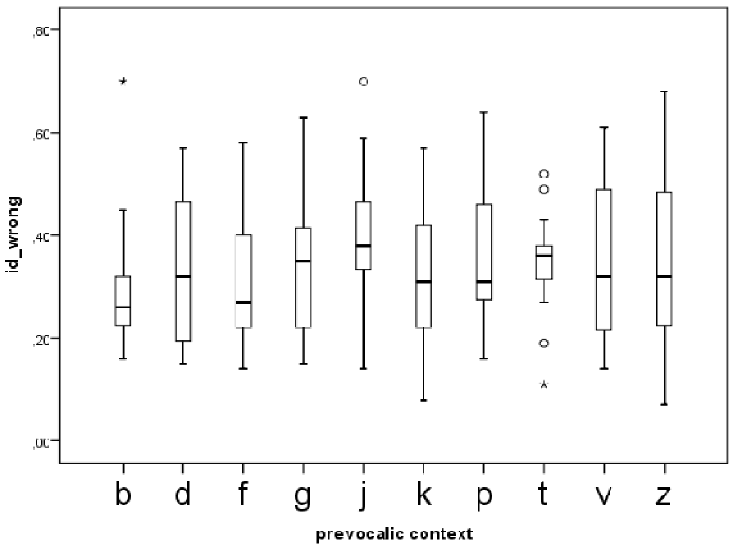

Figure 9.7: Percentage-wrong identifications for pre-vocalic consonant environments /CVtə/ in experiment 2 (15 input stimuli, 10 consonants, each vowel occurring once, 150 stimuli, 173 subjects)

pre-vocalic C	mean	min	max	SD
b	0.30	0.16	0.70	0.14
d	0.33	0.15	0.57	0.15
f	0.31	0.14	0.58	0.14
g	0.34	0.15	0.63	0.14
j	0.40	0.14	0.70	0.14
k	0.33	0.08	0.57	0.15
p	0.36	0.16	0.64	0.14
t	0.34	0.11	0.52	0.10
v	0.35	0.14	0.61	0.16
z	0.37	0.07	0.68	0.18

Table 9.2: Comparison of id_wrong scores by pre-vocalic context, mean, minimum, maximum and standard deviation SD (each vowel occurring once in each context, 150 stimuli, 173 subjects)

Some vowel categories are more sensitive to consonantal context variation than others and that some consonantal contexts, e.g. pre-vocalic /j/ causes higher id_wrong scores.

Figure 9.8 summarizes the impact of post-vocalic consonant variation of the results in experiment 1. No strong differences due to post-vocalic context are found in the overall rate of correct vowel identifications.

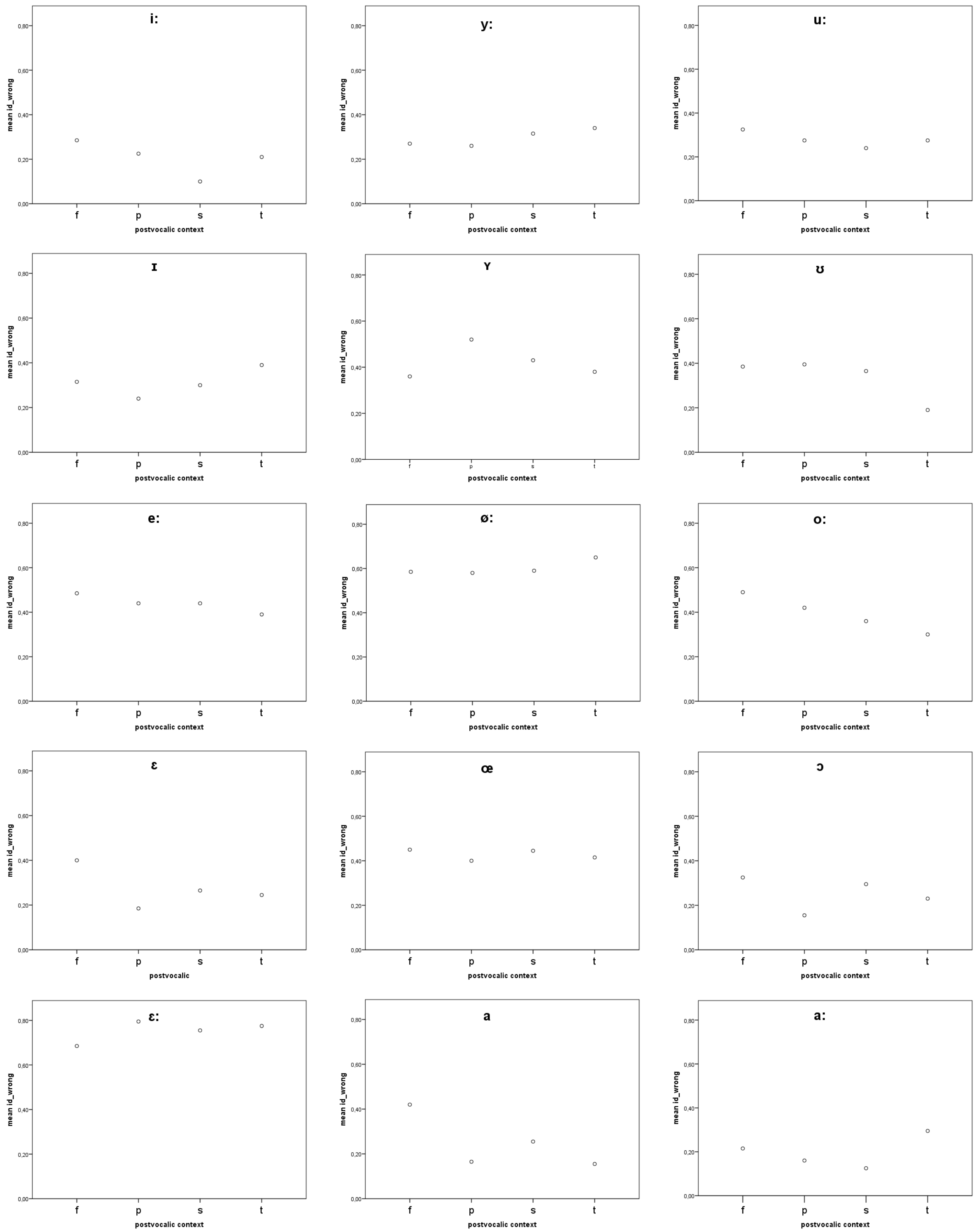

Figure 9.8: Comparison of id_wrong scores by post-vocalic consonant variation (each vowel occurring twice in a given context, 173 L2 listeners)

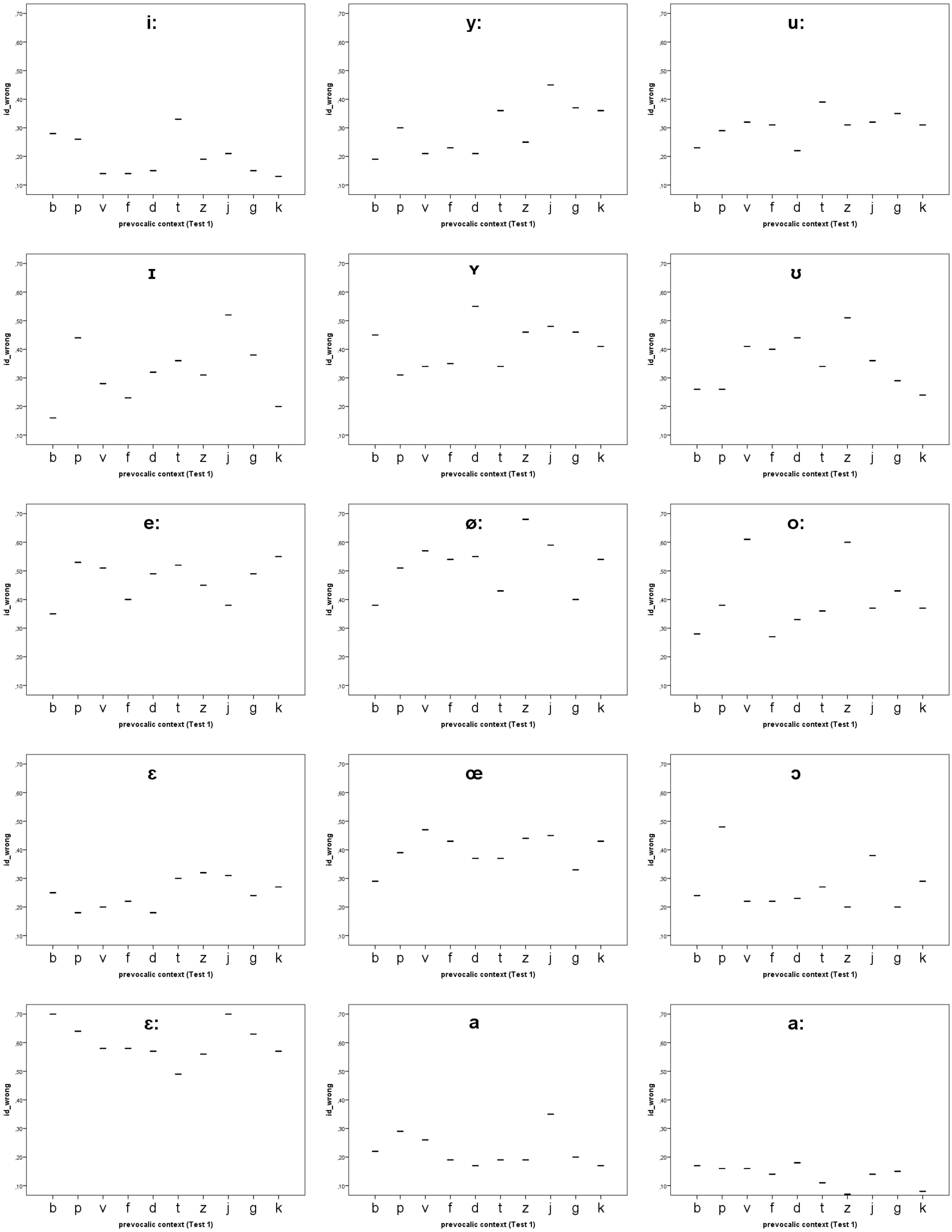

Figure 9.9: Comparison of id_wrong scores by pre-vocalic consonant variation (each vowel occurring once in a given context, 173 L2 listeners)

Figure 9.9 summarizes the impact of pre-vocalic consonant variation on the percentage of wrong identifications in experiment 2. It visualizes differences in id_wrong scores by vowel categories as well as by pre-vocalic contexts. Though for some vowel categories relatively little variation is observed, e.g. /a:/ or /ɛ/ show no significant variation with respect to context, more variation is given for higher, especially for some of the non-low rounded vowels such as /o:/, /ø:/, /ʏ/, /ʊ/ or /ʏ/, where pre-vocalic /j/ but also alveolar contexts (especially word-initial /z/) seem to have some effect on the identification of the following vowel. Since a given CV-structure was presented only once for each category and the analysis in Figure 9.9 was performed without dividing the sample into language sub-groups, the current data is not sufficient to make conclusions on the effect of pre-vocalic context variation.

The impact of variation of consonant context on vowel identification must be assumed to vary language-specifically. Detailed studies on context variation and vowel perception are, however, quite rare for most of the listeners' native languages analysed in the present study and studies on the effects of context variation in L2 learning are even rarer. Most studies are available for learners and native listeners of English (e.g. Strange et al. 2005; Strange et al. 2007, see section 10.5).

Therefore, the analysis in this chapter provides general insights into the influence of consonantal context on vowel identification on which further hypotheses on the impact of context on L2 vowel identification can be developed that would not only consider the influence of context on the error rate but also on the direction of substitutions. A detailed analysis has to consider not only place and manner of articulation but also the perception of specific acoustic cues for length, height or labiality in a given context and its impact on error rate by learners of a given native language. Future research on the influence of pre- and post-vocalic context could moreover extend the set of contexts by post-vocalic velars and pre- and post-vocalic nasals.

9.2 Preferences

The L2 listeners' response patterns vary not only in percentage of correct identifications but also with respect to the frequency a given response category was selected in the identification task. Some categories are selected more frequently than others and are therefore considered to be more "preferred" by the listeners.

Preference patterns vary language-specifically and are considered to be motivated by several factors such as language-specific differences in vowel inventories, frequency of occurrence, vowel-inherent category-specific characteristics and the listeners' experience and level of proficiency and (temporary) interlanguage hypotheses about the German vowel system (for a detailed discussion see chapter 11 and 12).

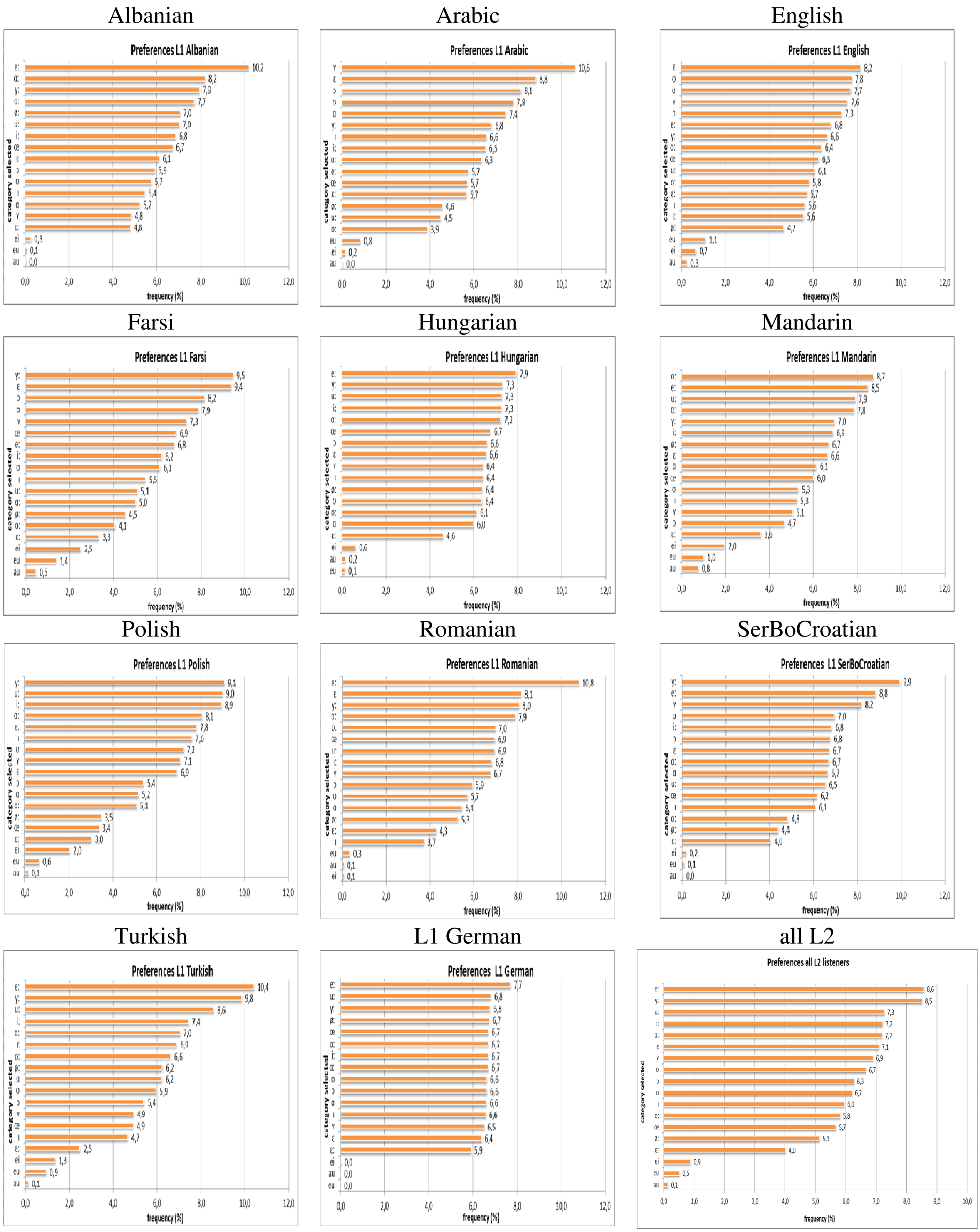

Figure 9.10: Comparison of language-specific variation in preference ranks[71]

[71] Preference scores are discussed in details for each of the language sub-samples in chapter 10 (see section 10.n.4.2 respectively)

Figure 9.11 represents the preference patterns for all 173 L2 participants and provides interesting insights into *asymmetries* in the L2 listeners' response behaviour and their preferences for specific response options.

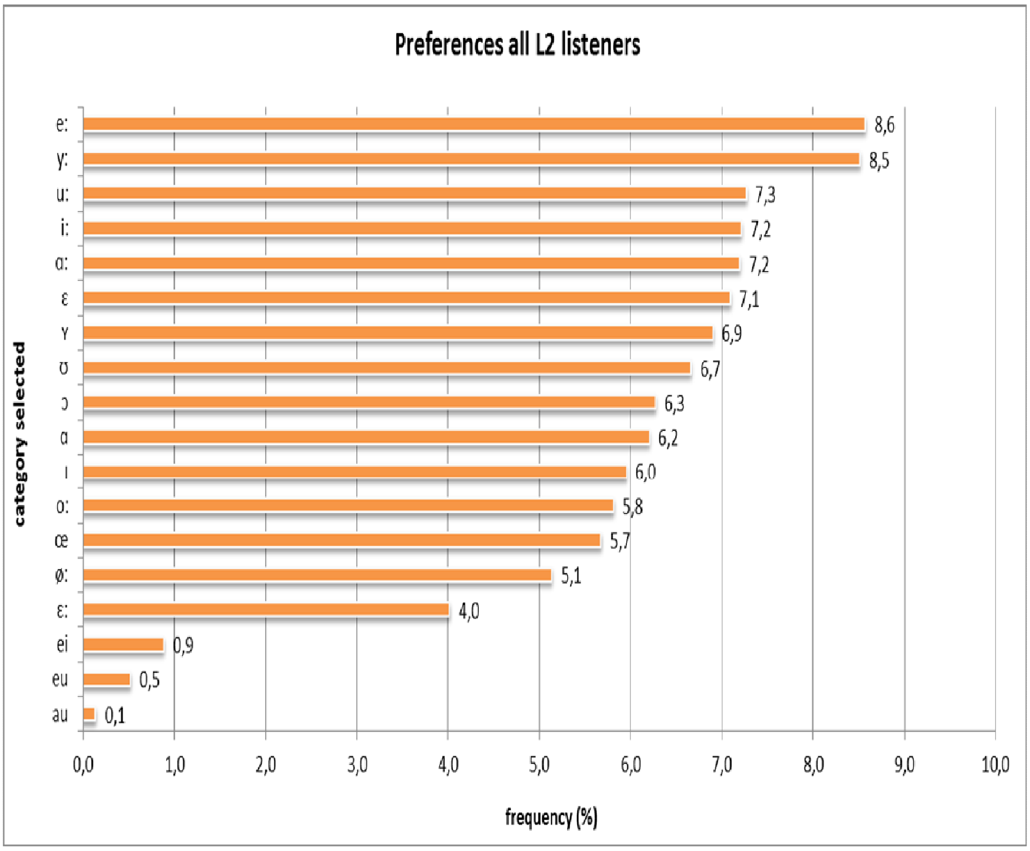

Figure 9.11: Preference ranking: Percentage category selected by all 173 L2 listeners (46.694 responses)

The comparison of language-specific variation in preference ranks in Figure 9.10 shows that in some L1 sub-samples pref_scores spread very evenly across vowel categories (e.g. Hungarian, English, German), whereas other samples show considerable inter-category variation in pref_scores, see the data for Arabic, Farsi or Polish but also for Romanian, SerBoCroatian or Turkish, where specific categories receive considerably higher pref_scores. This spread of inter-category varation in pref_scores reflects two aspects:

(1) Low inter-category variation in pref_scores, i.e. evenly dispersed pref_scores, indicate low overall confusion rate, which is of course expected in the L1 German sample but is also observed for other languages with a higher number of vowel qualities in their system such as Hungarian and Englisch.

(2) The results for other language sub-samples show that /y:/ and/or /e:/ receive very high pref_scores for reasons that will be discussed in detail in section 11.1.3 and 11.1.4 (for a language-specific comparison, see also Table 9.4).

A qualitative comparison of the preference ranking for each of the L1 sub-samples is presented in Table 9.3 and Table 9.4, visualized in the spider diagram in Figure 9.12.
/ɛ:/ and diphthongs (which do not occur in the input) are clearly non-preferred response options, followed by ö-qualities /œ/ and /ø:/, and by /o:/ and /ɪ/. /e:/ is the category that is most often selected, an effect that is due to the non-preference for /ɛ:/ as response category. Interestingly, the subjects seem to prefer phonemically long over short response categories. Especially /u:/, /i:/ and /a:/ are very attractive targets, while their unconstricted short counterparts are less often selected.

Albanian		*Arabic*		*English*		*Farsi*		*Hungarian*		*Mandarin*		*Polish*		*Romanian*		*SerBoCro*		*Turkish*		*all L2*	
e:	10.2	ʏ	10.6	ɛ	8.2	y:	9.5	e:	7.9	o:	8.7	y:	9.1	e:	10.8	y:	9.9	e:	10.4	e:	8.6
ɑ:	8.2	ɛ	8.8	ʊ	7.8	ɛ	9.4	y:	7.3	e:	8.5	u:	9.0	ɛ	8.1	e:	8.8	y:	9.8	y:	8.5
y:	7.9	ɔ	8.1	ɑ	7.7	ɔ	8.2	u:	7.3	u:	7.9	i:	8.9	y:	8.0	ʏ	8.2	u:	8.6	u:	7.3
o:	7.7	ʊ	7.8	ʏ	7.6	ɑ	7.9	i:	7.3	ɑ:	7.8	ɑ:	8.1	ɑ:	7.9	ʊ	7.0	i:	7.4	i:	7.2
ø:	7.0	ɑ	7.4	ɔ	7.3	ʏ	7.3	ɑ:	7.2	y:	7.0	e:	7.8	o:	7.0	i:	6.8	ɑ:	7.0	ɑ:	7.2
u:	7.0	y:	6.8	e:	6.8	œ	6.9	œ	6.7	i:	6.9	ɪ	7.6	œ	6.9	ɔ	6.8	ɛ	6.9	ɛ	7.1
i:	6.8	ɪ	6.6	y:	6.6	e:	6.8	ɔ	6.6	ø:	6.7	ʊ	7.2	u:	6.9	ɛ	6.7	o:	6.6	ʏ	6.9
œ	6.7	i:	6.5	ɑ:	6.4	i:	6.2	ɛ	6.6	ɛ	6.6	ʏ	7.1	i:	6.8	ɑ:	6.7	ø:	6.2	ʊ	6.7
ɛ	6.1	ɑ:	6.3	œ	6.3	ʊ	6.1	ʏ	6.4	ɑ	6.1	ɛ	6.9	ʏ	6.7	ɑ	6.7	ɑ	6.2	ɔ	6.3
ɔ	5.9	e:	5.7	u:	6.1	ɪ	5.5	ɪ	6.4	œ	6.0	ɔ	5.4	ɔ	5.9	u:	6.5	ʊ	5.9	ɑ	6.2
ʊ	5.7	œ	5.7	o:	5.8	u:	5.1	ø:	6.4	ʊ	5.3	ɑ	5.2	ʊ	5.7	œ	6.2	ɔ	5.4	ɪ	6.0
ɪ	5.4	ɛ:	5.7	ɛ:	5.7	ɑ:	5.0	ʊ	6.4	ɪ	5.3	o:	5.1	ɑ	5.4	ɪ	6.1	ʏ	4.9	o:	5.8
ɑ	5.2	ø:	4.6	ɪ	5.6	ø:	4.5	o:	6.1	ʏ	5.1	ø:	3.5	ø:	5.3	o:	4.8	œ	4.9	œ	5.7
ʏ	4.8	u:	4.5	i:	5.6	o:	4.1	ɑ	6.0	ɔ	4.7	œ	3.4	ɛ:	4.3	ø:	4.4	ɪ	4.7	ø:	5.1
ɛ:	4.8	o:	3.9	ø:	4.7	ɛ:	3.3	ɛ:	4.6	ɛ:	3.6	ɛ:	3.0	ɪ	3.7	ɛ:	4.0	ɛ:	2.5	ɛ:	4.0
ei	0.3	eu	0.8	eu	1.1	ei	2.5	ei	0.6	ei	2.0	ei	2.0	eu	0.3	ei	0.2	ei	1.3	ei	0.9
eu	0.1	ei	0.2	ei	0.7	eu	1.4	au	0.2	eu	1.0	eu	0.6	au	0.1	eu	0.1	eu	0.9	eu	0.5
au	0.0	au	0.0	au	0.3	au	0.5	eu	0.1	au	0.8	au	0.1	ei	0.1	au	0.0	au	0.1	au	0.1

Table 9.3: Comparative preference ranking by listeners' L1 (% selected)

Another interesting fact is that /y:/ is a highly preferred response category. In some language sub-groups (Polish, Farsi and SBC), it is even the most preferred category (see Table 9.3). The qualitative interpretation of the data shows that the preference for /y:/ is strongly connected with low preference scores for ö-qualities (/ø:/ and /œ/). A detailed description of language-specific preference patterns will show that there are asymmetric relationships between vowel categories and that some categories seem to be specifically attractive and exert a kind of "magnet effect" for other categories (for details, see chapter 11).
Table 9.4 summarizes and compares preference ranks for each of the L1 sub-samples. A comparison of the differing preference ranks for each of the categories tested in Table 9.4

shows that /y:/ and /e:/ are generally ranked very high in all language sub-groups, indicating that they are among the most preferred response categories, whereas /ɛ:/ receives very low ranks. /y:/ and /e:/ function as perceptual substitutes for other categories, such as front rounded vowels for /y:/ and /ɛ:/-stimuli for /e:/. For all other categories, considerable differences between languages are observed that have to be interpreted language-specifically (for a detailed discussion, see chapter 10 and 11).

	Alb	*Pol*	*Arab*	*Farsi*	*Mand*	*Engl*	*Hung*	*Rom*	*Turk*	*SBC*	*Germ*	*all L2*
ɑ	13	11	5	4	9	3	14	12	9	9	11	10
ɑ:	2	4	9	12	4	8	5	4	5	8	8	5
ɛ	9	9	2	2	8	1	8	2	6	7	14	6
e:	1	5	10	7	2	6	1	1	1	2	1	1
ɛ:	15	15	12	15	15	12	15	14	15	15	15	15
ɪ	12	6	7	10	12	13	10	15	14	12	12	11
i:	7	3	8	8	6	14	4	8	4	5	7	4
ɔ	10	10	3	3	14	5	7	10	11	6	10	9
o:	4	12	15	14	1	11	13	5	7	13	6	12
ʊ	11	7	4	9	11	2	12	11	10	4	9	8
u:	6	2	14	11	3	10	3	7	3	10	2	3
œ	8	14	11	6	10	9	6	6	13	11	5	13
ø:	5	13	13	13	7	15	11	13	8	14	4	14
ʏ	14	8	1	5	13	4	9	9	12	3	13	7
y:	3	1	6	1	5	7	2	3	2	1	3	2

Table 9.4: Cross-language comparison by preference ranks (1=most frequently selected, 15=least frequently selected)

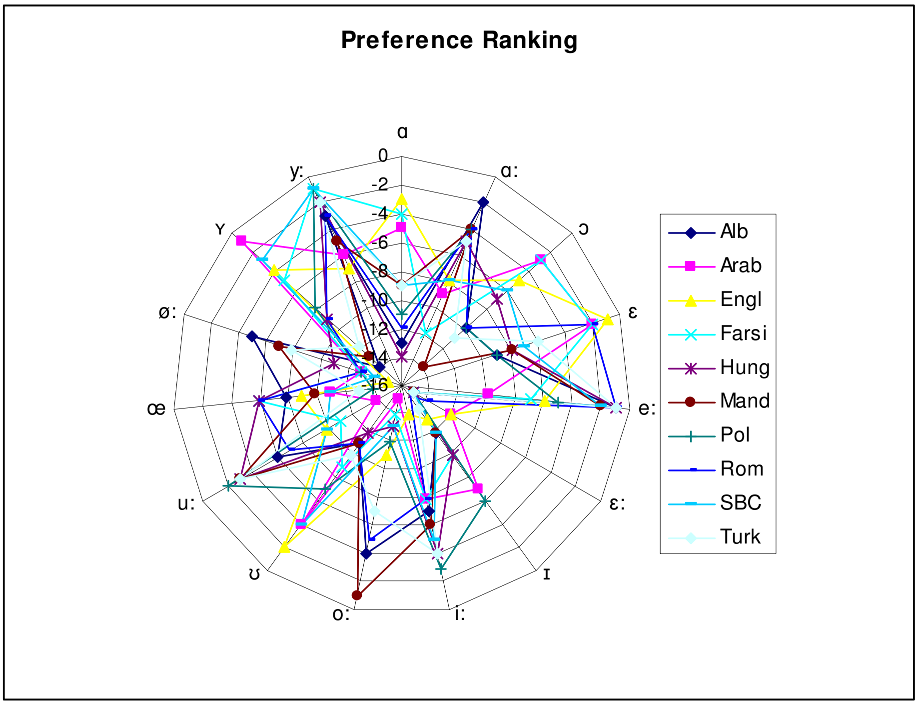

Figure 9.12: Cross-language comparison of preference ranks (more peripheral = more preferred)

The spider diagram in Figure 9.12 shows that /ɛ:/ is the least preferred category in all language sub-groups. Rather low pref_scores are moreover obtained in almost all languages (except Albanian) for /ø:/, whereas a generally high preference in most language sub-groups is observed for /y:/, /ɑ:/, /i:/, /u:/, and /e:/. Divergent effects between language sub-groups are especially observed for /o:/, which has high pref_scores in Mandarin, Albanian, Romanian and Turkish but very low pref_scores in all other languages.

9.3 Patterns of confusion

Wrong and correct identifications of input stimuli were compiled in a confusion matrix. Table 9.5 presents the results for all 173 non-native subjects in percent. Input stimuli are presented in rows, selected response categories in columns. Correct identifications are printed in bold, values <0.5% are not indicated, values <5.5% are printed in grey.

%	**a**	**a:**	**ɛ:**	**ɛ**	**e:**	**ɪ**	**i:**	**ɔ**	**o:**	**ʊ**	**u:**	**œ**	**ø:**	**ʏ**	**y:**	**‹ei›**	**‹eu›**	**‹ei›**	wrong	**corr**
a	**77**	22																	24	**77**
a:	15	**84**																	16	**84**
ɛ:		1	**33**	9	49		1									5			67	**33**
ɛ	1		7	**74**	15	1										1			26	**74**
e:			15	7	**55**	4	12						1		1	5			45	**55**
ɪ			1	10	4	**68**	11							2	1	1			32	**68**
i:				1	3	14	**80**								1	1			20	**80**
ɔ								**74**	20	1		2	1				1		26	**74**
o:								8	**61**	4	12	2	6	1	2		1	1	39	**61**
ʊ								9	3	**66**	13	2	1	5	1				34	**66**
u:									2	15	**71**	1	1	4	6				29	**71**
œ			2	5	3			2	1	2	1	**59**	14	6	2		1		41	**59**
ø:									1	2	5	7	**45**	9	28		1		55	**45**
ʏ						1				8	3	9	4	**58**	15		1		42	**58**
y:							2			2	3	1	3	16	**71**		1		29	**71**
Σ	*6*	*7*	*4*	*7*	*9*	*6*	*7*	*6*	*6*	*7*	*7*	*6*	*5*	*7*	*9*	*.9*	*.5*	*.1*	*35*	***65***

Table 9.5: Confusion matrix for all 173 non-native listeners (id_scores < 0.5% not indicated, <5.5% in grey)

The data for non-native listeners show three major areas of category confusion, where differences in vowel quality are perceived incorrectly: (1) confusion between *e- and i*-vowels, (2) confusion among *o- and u*-vowels, and (3) confusion among *ü- and ö*-vowels. These patterns of confusion will be described in more details in the vowel-specific analysis in chapter 11.

9.4 Similarity and distance

From the number of wrong and correct identifications in the confusion matrix, similarity scores are derived by Shepard's formula (for a detailed discussion, see section 7.2.5). The similarity matrix in Table 9.6 shows the similarity scores (sim_scores) for the full sample of non-native listeners (173 subjects).

The highest sim_scores are obtained for the contrast /ɛ:/ - /e:/ (0.725) followed by /ø:/ - /y:/ (0.273) and several other contrast pairs that differ only in phonemic length and constriction degree: /ʏ/ - /y:/ (0.243), /ɑ/ - /ɑ:/ (0.230), /ɔ/ - /o:/ (0.208), /œ/ - /ø:/ (0.208), /ʊ/ - /u:/ (0.201), /ɪ/ - /i:/ (0.172), /ɛ/ - /e:/ (0.166), and /ɛ/ - /ɛ:/ (0.147).

The more perceptual confusions occurred in the data the higher their degree of similarity (sim_scores). The sim_scores are subject to a transformation into distance scores (d_{ij} = -ln(s_{ij}) as described in section 7.2.5, on which the Multidimensional Scaling procedure is based.

V	a	a:	ɛ:	ɛ	e:	ɪ	i:	ɔ	o:	ʊ	u:	œ	ø:	ʏ	y:
a	1.000														
ɑ:	0.230	1.000													
ɛ:	0.008	0.012	1.000												
ɛ	0.005	0.004	0.147	1.000											
e:	0.002	0.004	0.725	0.166	1.000										
ɪ	0.000	0.000	0.016	0.079	0.058	1.000									
i:	0.001	0.000	0.008	0.007	0.113	0.172	1.000								
ɔ	0.002	0.000	0.000	0.001	0.001	0.001	0.002	1.000							
o:	0.000	0.000	0.001	0.000	0.000	0.001	0.001	0.208	1.000						
ʊ	0.002	0.000	0.001	0.000	0.000	0.002	0.001	0.067	0.060	1.000					
u:	0.001	0.000	0.000	0.001	0.000	0.000	0.001	0.004	0.109	0.201	1.000				
œ	0.000	0.001	0.030	0.039	0.027	0.005	0.004	0.033	0.031	0.034	0.011	1.000			
ø:	0.000	0.000	0.009	0.002	0.009	0.004	0.003	0.008	0.069	0.026	0.059	0.208	1.000		
ʏ	0.000	0.000	0.001	0.000	0.008	0.026	0.007	0.005	0.015	0.099	0.054	0.124	0.129	1.000	
y:	0.000	0.000	0.003	0.001	0.009	0.007	0.015	0.001	0.016	0.022	0.067	0.027	0.273	0.243	1.000

Table 9.6: Similarity matrix for all 173 non-native listeners

9.5 Mapping the perceptual vowel space in L2 German

9.5.1 Multidimensional Scaling

The MDS solutions presented in Figure 9.13, Figure 9.14 and Figure 9.15 are calculated from the similarity and distance scores for all 173 non-native listeners. The purpose of this procedure is to derive a low-dimensional spatial representation of the multidimensional vowel space as defined by several language-specific weighting of phonetic and phonological

parameters. The derived spatial MDS representation of the *perceptual* vowel space is based on the L2 listeners' confusion rates, from which similarity scores and distance values are deduced (for a detailed discussion, see section 7.2.5). Distances in the MDS solutions reflect the relative similarity of items in the perceptual space. A detailed discussion of the function of similarity and distance in the perceptual space will follow in chapter 12.
The representations of the perceptual vowel space in Figure 9.13, Figure 9.14 and Figure 9.15 differ in number of dimensions and in goodness of fit, indicated here by their RSQ values (see Table 7.7). The higher the RSQ-value, the better the solution (for details, see section 7.3.4).

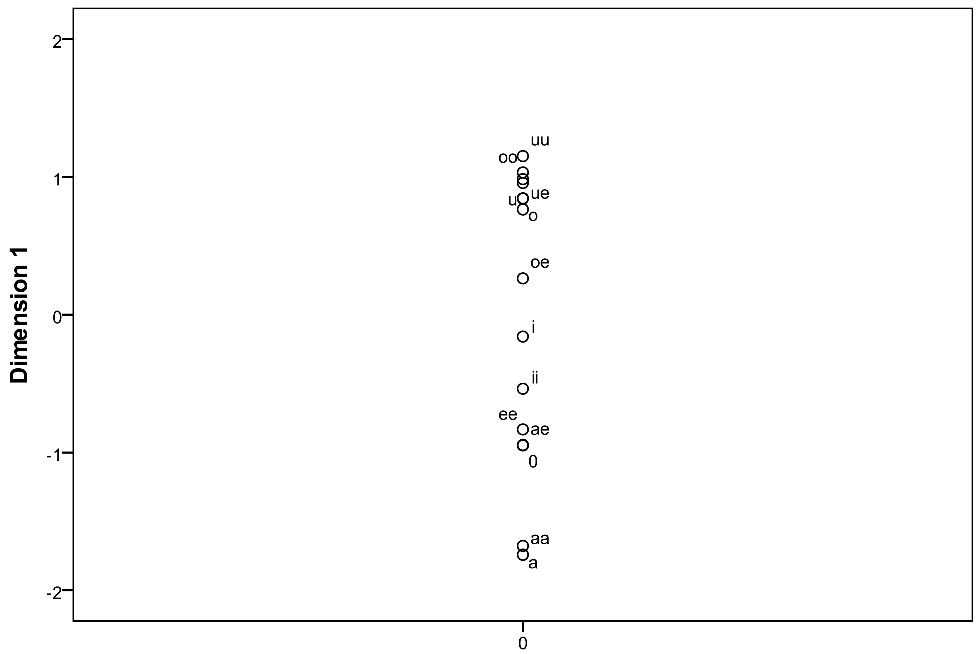

Figure 9.13: One-dimensional MDS solution of the perceptual vowel space for all 173 non-native listeners (RSQ .730)

The one-dimensional MDS representation in Figure 9.13 (RSQ .730) suggests a maximal distance between *a*- and *u*-vowels and high similarity of rounded (front and back) vowels at one side of the continuum. While front non-rounded vowels spread considerably along the one-dimensional axis, front and back rounded vowels (except /ɔ/) cluster strongly together, indicating a higher degree of perceived similarity between rounded categories. Of course this one-dimensional representation does not give sufficient insights into the listeners' perceptual representation and does not differentiate by listeners' native languages but provides a first impression on the relative similarity of categories in the perceptual space.

In the two-dimensional solution in Figure 9.14 (RSQ .905), four major clusters of German perceptual vowel categories can be identified: (1) *pharyngeal a*-vowels, (2) *back rounded u-* and *o*-qualities, where /ɔ/ is differentiated from other back rounded qualities, (3) a slightly more differentiated cluster of *front rounded* vowels, and (4) a cluster for *palatal* vowels that is divided in three sub-clusters: (a) *i*-qualities, (b) *ɛ*-qualities, and (c) /e:/ in an intermediate position between *i*- and *ɛ*-qualities.

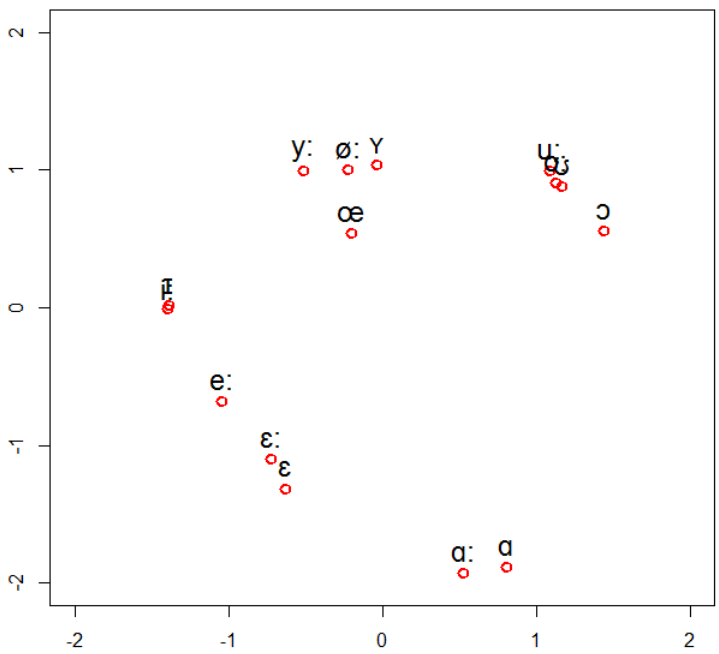

Figure 9.14: Two-dimensional MDS solution of the perceptual vowel space for all 173 non-native listeners (RSQ .905)

The three-dimensional representation for all L2 listeners in Figure 9.15 (RSQ .926) gives a similar impression as the two-dimensional solution but differs slightly with respect to the position of front rounded vowels that are located more closely to front vowels here than in the two-dimensional solution. Moreover, it shows a stronger differentiation within this class as well as within the class of back rounded vowels. The distance between /ɔ/ and all other back rounded categories is also evident in the two-dimensional representation, but the three-dimensional solution shows more differentiation within the cluster of back rounded vowels for /u:/, /o:/ and /ʊ/.

However, for a differentiated interpretation of spatial representations of the perceptual vowel space it is necessary to consider *not the position* of a category as such but rather the relative *distances* between different categories. For a detailed discussion of the interpretation of MDS solutions see section 7.3 and especially section 7.3.4, 7.3.5 and 7.3.6. Theoretical aspects of the relation of similarity and distance in human perception will be discussed in chapter 12.

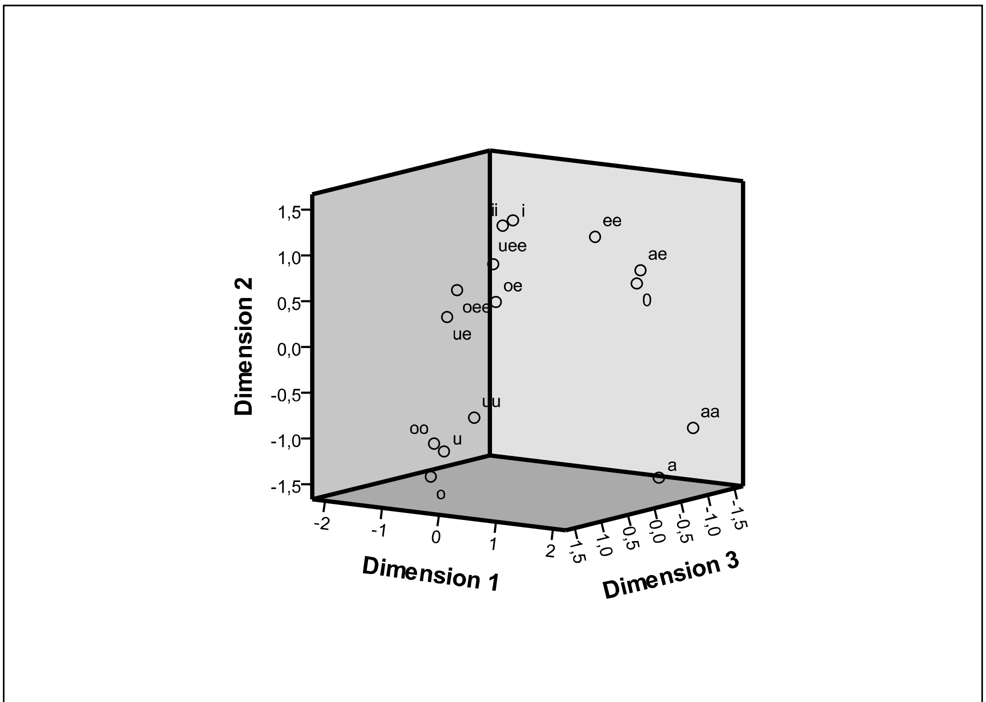

Figure 9.15: Three-dimensional MDS solution of the perceptual vowel space for all 173 non-native listeners (RSQ .926)

9.5.2 Cross-language comparison of perceptual vowel maps

As discussed in section 7.3.3, 7.3.4 and 7.3.6, it is important to note that spatial representations provided by MDS are on the one side a good method to visualize relationships of similarity between categories for listeners' of a given sub-sample, but, on the other side, are also a form of data reduction. MDS is a way to transform the original nominal confusion data into a higher and more explicit scale level converting them into metric (and also symmetric) measures of similarity or distance but reducing information about asymmetries in confusion patterns as evident in the confusion matrix. On the other side, MDS is a valuable method to operationalize, compare and visualize relationships of similarity between German vowel categories.

The advantages of this method to describe and compare relationships of similarity and perceptual distances between categories become evident in a cross-language comparison of MDS solutions for the ten language sub-groups. The cross-language comparison of two-dimensional MDS representations in Figure 9.18 shows that the major regions of similarity or confusion surface quite differently in the language-specific MDS charts and are strongly related to the listeners' L1 vowel system. The cross-language comparison presented here gives first insights into major patterns of confusion and typical constellations in the perceptual vowel space of non-native listeners of German. The language-specific MDS solutions and typical confusion patterns of listeners from a given language background will be discussed in details in the following chapter.

Two- and three-dimensional MDS solutions differ in goodness of fit, which is indicated by the measures of Kruskal's stress 1 and accounted variance R-squared (RSQ) (as discussed in 7.3.4). In general, non-metric solutions and solutions of higher dimensionality receive better RSQ scores (see the comparison in Figure 9.16). MDS representations for L1 sub-samples differ in goodness of fit. A language-specific comparison of RSQ values for non-metric (ordinal) MDS solutions is presented in Table 9.7.

The MDS solutions for the "all L2"-sample received in section 9.5.1 receive very high RSQ scores indicating a high goodness of fit due to considerable spreading of responses and the high number of participants, whereas the scores for the L2 sub-samples differ considerably in goodness of fit (see Table 9.7). In general, non-metric solutions receive better RSQ scores as the comparison of metric and non-metric solutions and degrees of dimensionality in Figure 9.16 shows.

RSQ non-metric MDS	*1Dim*	*2 Dim*	*3 Dim*	*id_correct (%)*
Albanian	.505	.733	.814	65
Arabic	.655	.882	.914	63
English	.608	.918	.940	52
Farsi	.538	.835	.884	48
Hungarian	.148	.671	.799	88
Mandarin	.635	.865	.911	51
Polish	.689	.844	.922	48
Romanian	.476	.751	.830	69
SerBoCroatian	.302	.782	.878	69
Turkish	.405	.693	.801	69
L1 German	.170	.461	.624	96
all non-native listeners	.730	.905	.926	65

Table 9.7: Percentage correct identifications and R-squared (RSQ) values for the ordinal-scaled (non-metric) MDS analysis based on distance values for each of the language sub-groups, for the German control group and for all non-native listeners

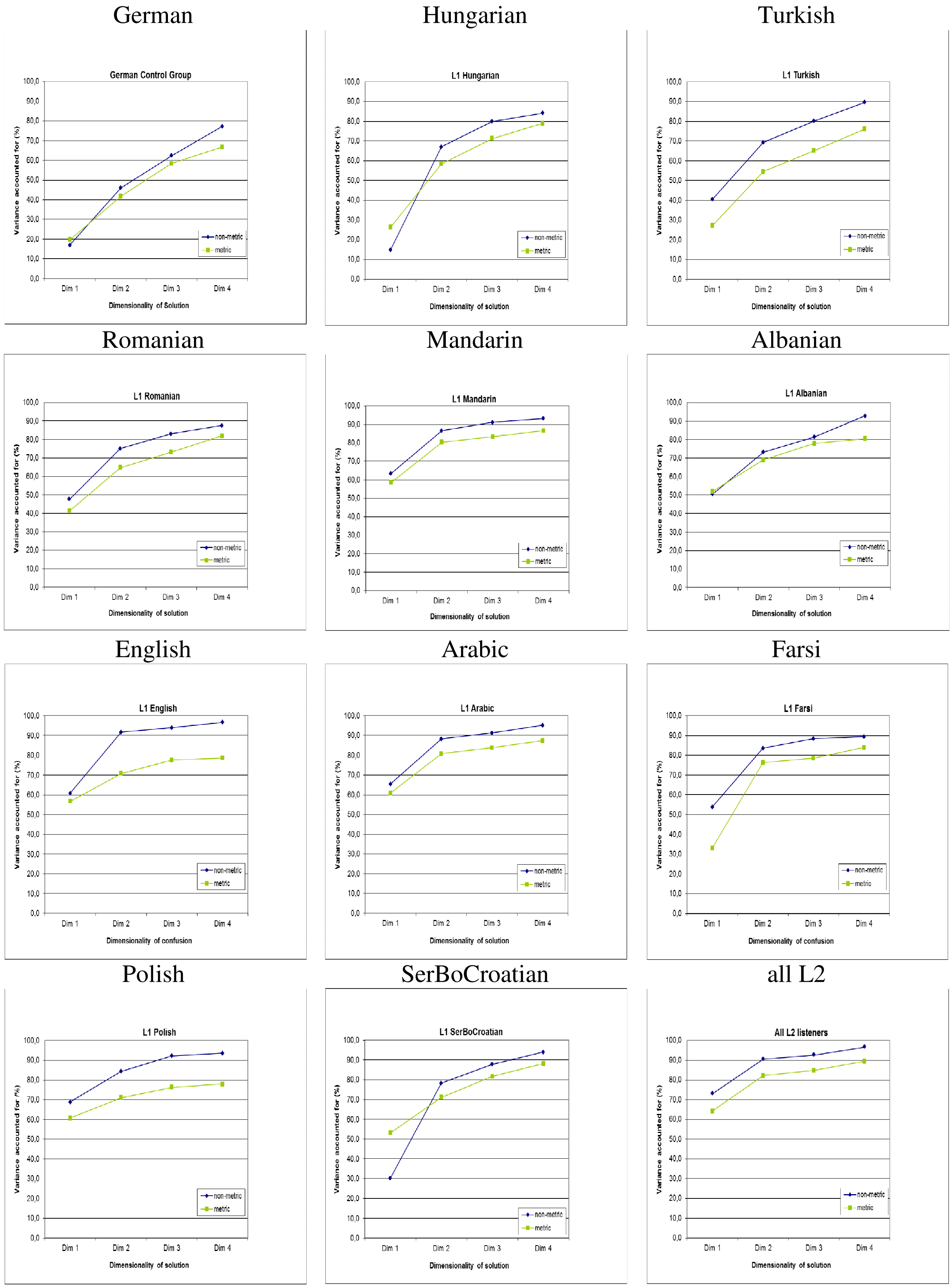

Figure 9.16: Comparison of RSQ values for metric (green) and non-metric (blue) one- two-, three- and four-dimensional MDS solutions

Figure 9.17 presents a *schematic* representation and comparison of the listeners' native language vowel systems. Vowels of allophonic or disputed status are printed in grey (for a detailed discussion of vowel categories in a given language, see chapter 10). The comparison shows that there are different types of systems: There are larger systems such as German, Hungarian and English with more than 12 vowel phonemes and smaller systems with only 5, 6 or 7 vowel types, such as e.g. SerBoCroatian, Polish or Farsi. There are systems with front rounded /y/- and /ø/-vowels as e.g. German, Hungarian or Turkish and systems with only one high front rounded vowel /y/ as e.g. Albanian or Mandarin. High non-peripheral /ɨ/ occurs in Polish and Romanian, whereas English, Arabic or Farsi have no high front rounded vowel. Moreover, we find languages with phonemic length distinctions whereas in others no such oppositions occur. In the current sample, German is the language with the largest system, where *ü*- and *ö*-types and phonemic length distinctions co-occur.

Figure 9.18 provides a cross-language comparison of all two-dimensional MDS solutions for each of the L2 sub-samples, the native German control group and for all non-native listeners ("all L2"). The comparison provides interesting insights into the language-specific mental representation of the German vowel system by L2 learners, though the results for Farsi and Mandarin have to be viewed with caution due to the small sample size.

The comparison of two-dimensional MDS representations in Figure 9.18 shows that the relative perceptual distance or similarity between vowel categories and the clustering of vowels belonging to a given class, e.g. back vowels, front non-rounded vowels and front rounded vowels, differ significantly between the language sub-groups. The most conspicuous difference between the sub-groups concerns the status of *front rounded vowels*, which are either perceptually "assimilated" to the class of back rounded vowels as in the English or Arabic sub-sample or perceived as a class of its own as in the Hungarian or native German sample, but also in the data for Turkish or Romanian listeners. The intermediate position of front rounded vowels in the Polish or the SerBoCroatian sample is due to confusions of front rounded vowels with front non-rounded as well as back rounded categories.

However, not all front rounded vowels are equally predisposed to substitutions by front non-rounded vowels. /œ/ is most frequently substituted by front non-rounded qualities, mostly by /ɛ/, whereas for /ø:/ substitutions with front non-rounded vowels hardly ever occur.

Moreover, the perceptual distances within the class of front non-rounded vowels are not equal. In general, /y:/ and /œ/ are differentiated best, whereas /ø:/ and /ʏ/ are more susceptible to confusions with other members of this class.

German

i:	y:		u:
ɪ	ʏ		ʊ
e:	ø:	ə	o:
ɛ(:)	œ		ɔ
		ɐ	ɑ(:)

Hungarian

i:	y:		u:
i	y		u
e:	ø:		o:
ɛ	ø		o
		a:	ɒ

Turkish

i	y	ɯ	u
e	ø		o
			ɑ

Romanian

i	ɨ	u
e	ə	o
	a	

Mandarin

i	y		u
	ə	ɤ	
	a		

Albanian

i	y		u
e	ə		o
	a		

English

i:			u:
ɪ			ʊ
e^{i}	ə		o^{u}
ɛ	ɜ		ʌ/ɔ
æ		a	ɑ/ɒ

Arabic

i:	u:
i	u
(e:)	(o:)
æ/a	ɑ:

Farsi

i:	u:
e	o
æ	ɑ/ɒ

Polish

i	ɨ	u
ɛ		ɔ
		ɑ

SerBoCroatian

i(:)		u(:)
e(:)		o(:)
	a(:)	

Figure 9.17: Schematic comparison of vowel systems (vowels of allophonic or disputed status in grey)

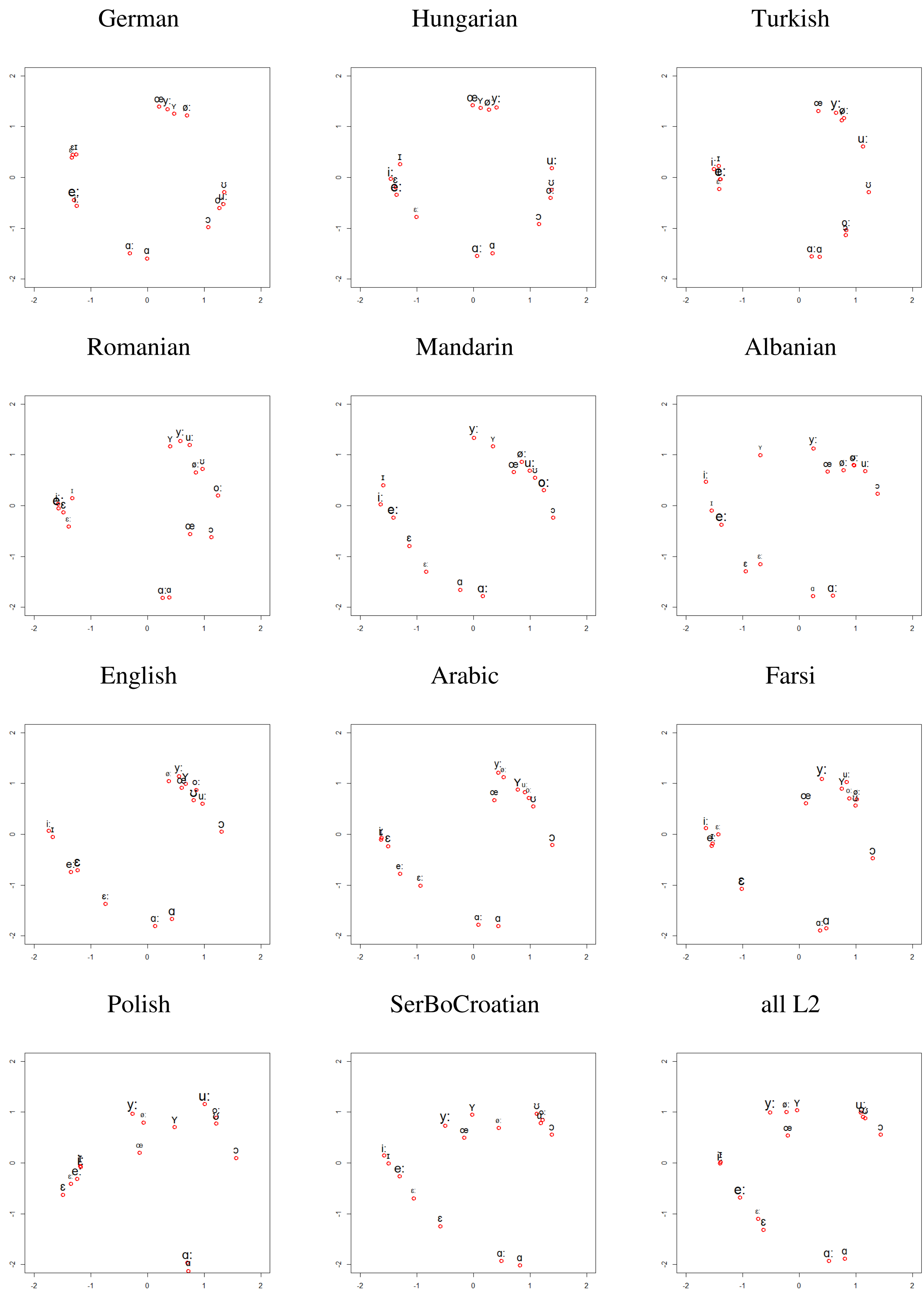

Figure 9.18: Comparison of language-specific two-dimensional MDS solutions

The two-dimensional MDS solutions for German, Hungarian and Turkish in row 1 of Figure 9.18 show the perceptual vowel spaces for listeners from languages, where front rounded /y/ and /ø/-vowels have phonemic status. The German and the Hungarian perceptual vowel space pattern quite similarly, which is largely due to a generally low number of wrong identifications and limited spreading of responses in both samples (see the confusion matrices of German in Table 10.3 and Hungarian in Table 10.42). In Hungarian and German, long and short vowel phonemes are differentiated, in German this distinction co-occurs with degree of constriction [+/– constricted]. Since front rounded vowels are occasionally confused with *u*-vowels by Turkish listeners but not by Hungarians, the *u*- and *o*-cluster is more distinct in the Hungarian sample, while /u:/ is closer to front rounded vowels for Turkish; this becomes more evident in the three-dimensional solution for Turkish.

The two-dimensional perceptual vowel maps for Turkish and Romanian listeners surface quite similarly. In both samples, front rounded vowels are better discriminated than in other samples. The capacity of Romanian listeners to differentiate front rounded vowels correctly can be accounted for by the fact that *y*- and *ø*-qualities occur in Romanian in several modern borrowings and are therefore more familiar to Romanian listeners.

The two-dimensional MDS solutions for Turkish, Romanian, Farsi, Polish and SerBoCroatian show a strong clustering of *i*- and *e*-qualities, indicating perceptual confusions within the class of front non-rounded vowels. These results are strongly related to a high spread of responses for German /e:/-stimuli within the class of *e*- and *i*-qualities and frequent substitutions by long /ɛ:/ and /i:/. By substitutions of this type, the listeners reveal a capacity to identify phonemic length but show difficulties to distinguish between pre-palatal and mid-palatal vowels. Difficulties in differentiating /i:/ and /e:/ are particularly common with Turkish, SerBoCroatian, Farsi and Polish listeners but not with Romanians. Moreover, strong asymmetries of mid-palatal /e:/ > pre-palatal /i:/ that occur substantially more often than /i:/ > /e:/ are observed in Farsi, Hungarian, Mandarin, Polish, SerBoCroatian and Turkish.

Another type of structuring the perceptual vowel space is observed for English, Arabic and Farsi in the third row of Figure 9.18. Here, a clustering of front rounded ü- and ö-vowels with back rounded u- and o-categories is observed indicating a significant portion of perceptual confusion of back and front rounded vowels. English, Arabic and Farsi listeners have no front non-peripheral qualities in their L1 system and therefore show considerable difficulties in establishing new perceptual categories for front rounded vowels. These difficulties are strongly connected with difficulties in German pronunciation and also surface in written production as evident in examples from Arabic learners presented in section 10.4.3.

Difficulties to distinguish front rounded and back rounded vowels are moreover reinforced by languages like Arabic that allow considerable allophonic spreading mainly for short unconstricted vowels. Similarly, several varieties of English show considerable allophonic variation for high u-vowels ranging from [u]-like to [y]-like qualities (cf. Cox 1999; Cox & Palethorpe 2001; Hillenbrand et al. 2001; Hawkins & Midgley 2005; Harrington et al. 2008; Labov, Ash & Boberg 2006; Strange et al. 2005; Strange et al. 2007; for discussion, see section 10.5.4.3).

To conclude, a comparison of the vowel systems in Figure 9.17 and the two-dimensional MDS solutions in Figure 9.18 shows that typical problems in differentiating front rounded vs. front non-rounded vowels and front rounded vowels vs. back rounded vowels seem not grounded in the overall number of vowel types in a given L1 system but rather in the type of phonemic oppositions and the range of allophonic variation in L1. The existence of phonemic distinctions in a given dimension, e.g. [+/– round] for front vowels, seems to influence the listeners' capacity to establish differentiated categories in the L2 German system. Hungarian and Turkish listeners, but also Mandarin, Albanian and even Romanian listeners, who are familiar with rounded vs. non-rounded front vowels distinguish front rounded vowels from back rounded qualities significantly better than English, Arabic or Farsi listeners. Interestingly, for the latter group, the phonemic differentiation of front vs. back rounded vowels, i.e. back rounded u- and o-qualities vs. front rounded ü- and ö-qualities is signficantly more difficult.

As the following chapters will show, this type of confusion between front *ü*- and *ö*-vowels and back rounded *u*- and *o*-vowels cannot directly be accounted for by acoustic characteristics of the stimuli alone, but is strongly connected with other factors determining the listeners' mental representations of L2 categories. Phonological oppositions, phonetic variation but also orthographic factors have to be considered in the analysis. In other words, no one-to-one relation between the acoustic signal and the listeners' performance in an L2 identification task can be assumed.

To conclude, the listeners' mental representations of L2 vowel categories are not only determined by acoustic properties of the input stimuli but also by specific characteristics of the response options. The stimulus and response relationship and the interaction of similarity and bias of stimuli and responses have to be considered for a complete analysis of the listeners' performance in an L2 perception task. A more detailed analysis of the direction of substitutions will be offered in the vowel-specific analysis in chapter 11 and the language-specific discussion in chapter 10.

9.6 Conclusions

To summarize, the results show a considerable range of inter-category range of variation in difficulties and type and direction of perceptual substitutions across languages and within each of the L1 sub-groups. The data show strong language-specific (see chapter 10) and category-specific (see chapter 11) differences in the percentage of wrong and correct identifications. *Difficulties* in identifying German vowels corrrectly and *preferences* for specific response categories vary considerably between and within the language sub-samples. The listeners' responses show *asymmetries* in the direction of perceptual substitution processes, e.g. /e:/ > /i:/ is more frequently observed than /i:/ > /e:/.

Asymmetries between German vowel categories may concern (1) the percentage of wrong identifications (*difficulties*) and the occurrence of confusion between categories, (2) variation in *preferences*, i.e. the listeners' bias for specific response categories (see preference rankings), and (3) asymmetries in the *direction of perceptual substitutions* as tabulated in the confusion matrices (for a detailed discussion, see chapter 11).

The statistical method of Multidimensional Scaling serves as valuable device to visualize relationships of perceptual similarity between German vowel categories and comparing the perceptual vowel spaces of L2 listeners from different native languages. Asymmetries in confusion patterns and perceptual substitutions are however not directly reflected in the perceptual vowel maps, as they are based on distance scores and Shepard's formula for similarity which yields symmetric values for inter-category similarities and distances. Therefore, MDS cannot replace a detailed analysis of patterns of perceptual confusion as provided in the confusion matrix for each of the listeners' native languages.

A language-specific analysis of the study's results for each of the language sub-groups will be presented in the following chapter. A vowel-specific discussion that can reveal possible reasons for perceptual confusions beyond the mere influence of the listeners' L1 will be offered in chapter 11. The relation of input signal and response options and the interaction of category-specific phonetic properties, phonological features, orthographic representations will be discussed in the final chapter of this dissertation. In chapter 12, the many aspects of the phenomenon of language-specific structuring of the human perceptual vowel space will be discussed theoretically and will be supported by the experimental data from the current study.

10 Language-specific Analysis

10.1 Analysis of L2 listeners' responses

This chapter presents a language-specific analysis of L2 confusion data collected in an identification task. The data analysis will follow the procedure described in chapter 7. For each of the participants' native languages, the analysis in the following sections will offer

(1) a description of *general characteristics* of the language under discussion,

(2) a *phonological and phonetic description* of the vowel system,

(3) a brief *contrastive outline* describing possible areas of difficulty with L2 German vowel perception, and

(4) a detailed *analysis of the results* of the perception experiment.

The discussion and interpretation of the results for each language sub-group will have the following structure:

(1) *Difficulties* that learners of a given native language have in correctly identifying the input stimuli will be represented in bar charts. Identification scores represent the number of correct responses (*id_correct*) or the number of wrong responses (*id_wrong*) for a given input category. The bar charts compare the relative degree of difficulty of the input categories for learners of German from a given language sub-group. Boxplot diagrams will show the range of interpersonal variation within a language sub-group for each of the categories tested.

(2) *Preference patterns* will be represented by bar charts for each of the language sub-groups. The preference scores (*pref_scores*) describe the frequency a given category is selected irrespective of whether this choice is wrong or correct. Preferences are also represented in the last row of the confusion matrix and are visualized in the tow-dimensional MDS solutions by differences in type size.

(3) *Patterns of confusion* will be summarized in a *confusion matrix* for each of the language sub-groups. The confusion matrix represents the id_wrong and id_correct scores for each of the input and response categories and will be interpreted in terms of number and quality of perceptual substitutions, i.e. in terms of wrong and correct categorizations for input stimuli belonging to a given vowel type and the response categories that were selected for these stimuli. Possible patterns of *asymmetry* or uni-directional substitution patterns between two categories will be discussed if observed.

(4) *Similarity and distance*: Based on Shepard's formula (see section 7.2.5), *similarity scores* (sim_scores) are calculated for each vowel contrast, which are summarized in a *similarity*

matrix. Higher sim_scores indicate higher perceptual similarity between categories and a higher percentage of inter-category confusion, while lower sim_score indicate better perceptual differentiation[72]. Sim_scores are discussed not only for individual categories but also for classes of sounds (e.g. back vowels or front rounded vowels). From the similarity scores calculated by Shepard's formula (see section 7.2.5) *distance scores* are derived by the formula $d_{ij} = -\ln(s_{ij})$. These distance scores will not be represented in a separate matrix as they are directly derived from the similarity matrix.

(5) *The perceptual vowel map*: One-, two- and three dimensional MDS solutions (see section 7.3) are derived from similarity and distance scores to represent the learners' perceptual vowel map for a given language sub-group. The MDS representations differ in *goodness of fit* (see section 7.3.4) which is indicated here by RSQ. The higher the *RSQ values*, the better the MDS solution (for discussion, see section 7.3.3). Differences in type size in the MDS solution indicate patterns of perceptual *preferences*, i.e. more preferred response categories are larger in type size than less preferred ones.

L1 % id_wrong	N	min	max	mean	SD	# / V
Albanian	12	14	64	34,8	13,7	204
Arabic	10	9	63	37,6	19,7	176
English	14	26	70	46,9	12,4	261
Farsi	4	30	73	52,5	18,4	72
Hungarian	26	3	40	11,8	9,0	409
Mandarin	7	21	68	51,7	16,3	105
Polish	31	15	69	52,1	12,2	551
Romanian	12	6	47	31,1	11,4	215
SerBoCroatian	33	9	56	30,9	13,5	589
Turkish	24	11	51	31,6	11,6	416

Table 10.1: Percentage wrong responses for all L1 sub-groups (number of subjects, minimum, maximum, mean, standard deviation SD, mean number of valid responses for each vowel category (#/V))

Table 10.1 presents the number of participants for each of the non-native subgroups. Larger samples are available for Polish, SerBoCroatian, Hungarian or Turkish; smaller samples are available for Albanian, English, Romanian or Arabic. For Farsi and Mandarin, for which the number of participants is below 10, the sample allows only limited generalizations of the

[72] The sim_scores relate stimuli and responses. No values for diphthongs are indicated in the similarity matrix as they do not occur as stimuli in the input.

results. For these languages, the results can offer only a first insight into regions of difficulty with German vowels for learners' of the respective native language.

An important question related to sampling in L2 studies is the control of individual differences and inter-speaker variation that may be due to a number of internal as well as external factors influencing language experience and language competence in L1 and L2 (see chapter 8). The language sub-samples in the current study are not balanced with respect to the number of beginners and advanced participants and the subjects' level of proficiency in L2 German, which is an influential factor but could not be systematically controlled in the process of subject acquisition. A statistically solid comparison of beginners and advanced learners can therefore not be provided.

One of the more frequently considered factors in sampling refers to the origin of the participants in a study and possible influences of regional variation within their native language. Intra-language variation between regional varieties and its possible effects on the acquisition of phonetics and phonology in L2 must of course be taken into consideration not only for poly-centric languages such as Arabic and English, but also for smaller languages such as Albanian with its two main dialect groups, or for languages whose status as national or poly-centric languages have been under discussion because of recent political developments, e.g. for the language that will be referred to here as SerBoCroatian.

We commonly rely on the assumption that listeners from the "same language" background will behave similarly in a perceptual categorization task, but this assumption as well as the notion of the "native speaker" itself has to be viewed with caution, not only because of regional and sociolinguistic variation within a speech community. The term "native speaker" and its implicit assumptions have been challenged in recent years (cf. e.g. Escudero & Sharwood-Smith 2001; Cook 2003; Davies 2003; Paikeday 2003; Sharwood-Smith 2011, see section 12.1.3.1) by the view of languages as dynamic systems (Cook 2003; de Bot & Larsen-Freeman 2011) and by evidence from studies showing that perception and production in L1 may change as an effect of the acquisition of L2, L3 or Ln (cf. Cabrelli Amaro 2012, Wrembel 2014, and the special issue of the International Journal of Multilingualism 7/1, 2010).

An inter-individual differentiation of the experiment's results will not be provided due to the high number of participants in the sample. However, to account for individual factors that may influence the L2 learners' success, a discussion of *learner-specific factors* influencing the listeners' overall performance in the experiment was presented in chapter 8.

A homogeneous sampling with respect to all these influencing factors is outside the scope of the present study (for methodological difficulties of homogeneous sampling, see also section 8.7). Rather, variation within a given L1 sample due to regional varieties, different stages of L2 acquisition or different contexts of L2 learning will be tolerated here to offer a more wide-spread description of *possible* difficulties and category confusion for speakers of one "language". The present study is to be understood as a survey study that provides data for ten typologically different languages and lends itself to develop research questions for further studies on L2 vowel identification.

The central aim of the current study is to present a *cross-language comparison* of L2 vowel identification for the full vowel system of German for ten different languages with the aim to identify language-specific as well as language-general tendencies for non-native listeners in L2 vowel perception: A further aim is to provide new methodological approaches to the notion of similarity and to model the language-specific perceptual vowel space in L2 German. The experimental design of the present study is not intended to test the role of specific sub-phonemic or context-specific phonetic details but to provide insights into the mental representation and organization of the German vowel system as a set of categories to be acquired by learners from different native languages.

The data analysis for the languages under investigation will show that there are many open questions related to the phonological status or to phonetic details for specific vowel sounds in a given language and/or its varieties as discussed in the relevant literature (e.g. the number of vowel phonemes in Mandarin). In the description of some languages treated in this study, several questions arise for which currently no reliable phonetic data are available. For each of the languages under investigation the results obtained in the present study reveal several research questions for further research and language-specific studies of specific aspects of vowel perception and cross-linguistic influence in the acquisition of German as a foreign language.

In the following sections, each language will be discussed separately. In chapter 11, a *vowel-specific* cross-language comparison of the results will be offered for each vowel class. A *cross-language comparison* of results for all L1 sub-groups describing common tendencies and language-specific tendencies was presented in chapter 9. Preceding the discussion of the results for non-native listeners, the results for the native German control group will be presented in the next section.

10.2 German – The L1 control group

10.2.1 Results and interpretation

18 German native speakers of different regional origin participated in the study. 4860 valid responses were obtained. 324 valid responses were collected for each vowel category (no missing answers).

10.2.1.1 Difficulties

96.1% of the German native speakers' responses were categorized correctly and 3.9% were incorrectly classified. Figure 10.1 compares the number of wrong and correct responses for each of the input categories. The mean percentage of wrong identifications ranges from very high scores for /i:/ and /u:/ (99.1%) to the lowest id_correct score of 79.3% for /ɛ:/ (see Figure 10.1). Interestingly, the most difficult categories were e-vowel categories; all e-qualities received id_correct scores below 95%. This effect is mainly due to a higher number of inter-category confusion between /e:/ and /ɛ:/. Occasionally, short /ɛ/-stimuli were categorized as /e:/ or /ɛ:/ (8 cases each).

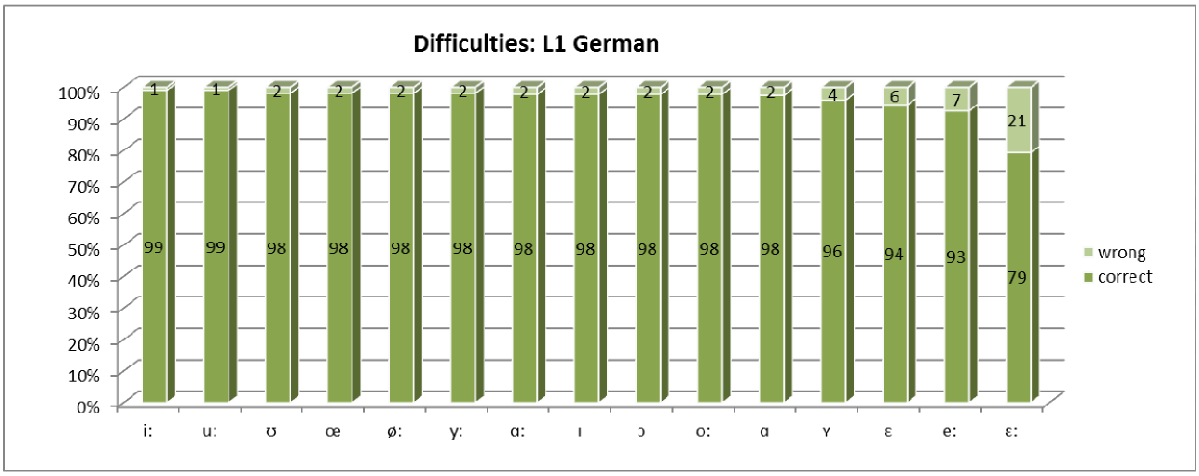

Figure 10.1: Percentage correct and wrong identification for the L1 German control group (N=18, 4860 responses)

Within the class of front rounded vowels, /œ ø y:/ received id_correct scores of 98.1%, whereas for /ʏ/-stimuli an id_correct of only 96% was obtained.

The id_scores obtained here are comparable with the percentage of correct answers in other studies with native speakers. Peterson & Barney (1952) had a minimum of 87.0% correct for /ɑ/-stimuli and a maximum of 99.9% for /i/ in their listening experiment. Hillenbrand, Getty, Clark & Wheeler (1995) obtained a minimum of 82.0% for /ɔ/ and a maximum of 99.6% for /i/ and excluded stimuli that had an identification error rate of 15% or higher. In the present study, higher id_wrong scores are only obtained for /ɛ:/. However, none of these stimuli were excluded from the input material.

10.2.1.2 Preferences

Preference scores for the L1 German control group are presented in Figure 10.2. The pref_scores for monophthong categories vary within a range of 1.8%, with a minimum of 5.9% for /ɛ:/ and a maximum of 7.7% for /e:/. Compared to the groups of non-native participants this range of variation is very low and responses are almost equally dispersed over all response categories with the only exception of /e:/ and /ɛ:/. Long categories seem to be slightly more preferred by German native listeners than their short counterparts. A diphthong was selected only in one single instance. For the class of front rounded vowels no substantial differences are found with respect to preferences for a specific category, though /y:/ (6.8%) and /ø:/ (6.7%) are slightly more frequently selected than their short equivalents. /ʏ/ is not only the front rounded category with the highest id_wrong score but also the least preferred front rounded category (6.5%). For back rounded vowels, /u:/ is slightly more frequently selected than other members of this class.

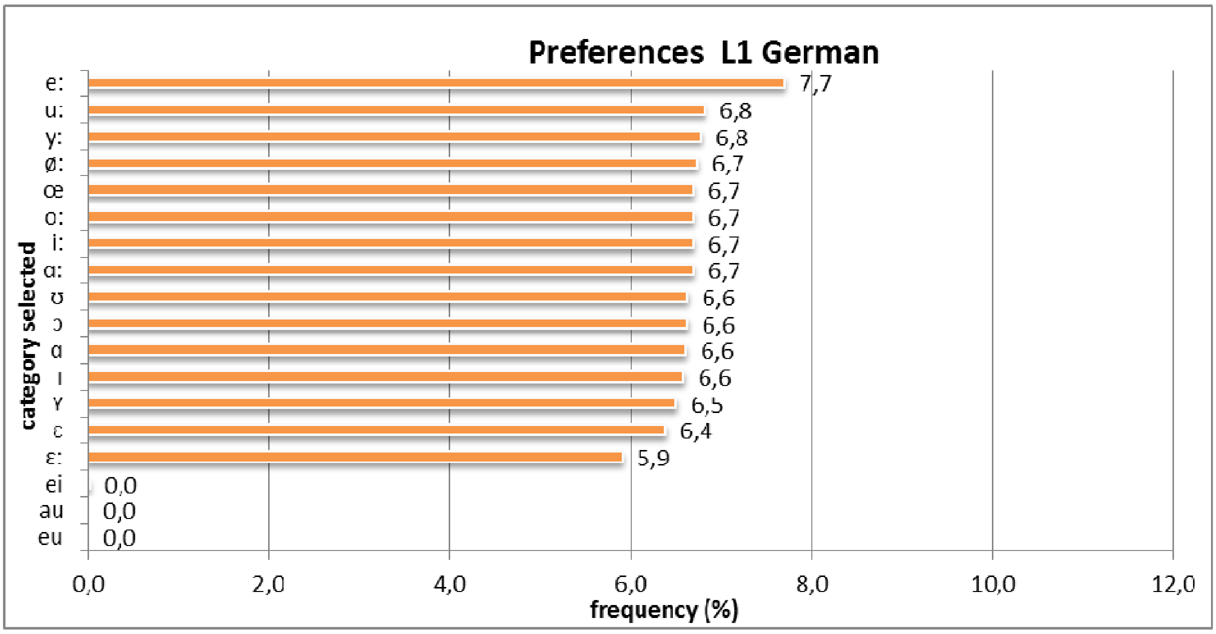

Figure 10.2: Preference scores for the L1 German control group

10.2.1.3 Patterns of confusion

A quantitative and qualitative representation of inter-category confusion is presented in Table 10.2 and Table 10.3. Table 10.2 represents the number of instances a given input-category was classified as one of the presented output-categories in tokens, whereas Table 10.3

represents the percentage of wrong and correct categorizations as one of the presented response categories in percent (values below 0.5% not indicated).

For a-vowels, occasional instances of wrong classification of vowel length are observed.

Most difficulties are observed for the long e-qualities. A diphthong was selected only in one single case (<ei> for /ɛ:/). The largest part of wrong categorizations observed in the data is due to an inappropriate recognition of distinctions between long constricted vs. short unconstricted vowels at the same constriction location, e.g. /ʏ / - /y:/, /ɔ/ - /o:/, /ʊ/ - /u:/ or /œ/ - /ø:/.

The highest percentage of wrong identifications is observed for e-qualities. Responses for /ɛ:/ and /e:/ clearly show an asymmetric response pattern. While /ɛ:/ was substituted by /e:/ in 20% of the cases, inverse substitutions of /e:/ > /ɛ:/ occurred only in 7% of the cases. The strong tendency to substitute /e:/ for /ɛ:/ reflects the reduction of the /e:/ - /ɛ:/ contrast that is observed in many regional standard varieties of German. While many speakers can differentiate the contrast in more formal styles of production (e.g. in a word reading task), the opposition would be weakened in more spontaneous speech styles and could be described as a "phantom opposition" (Maas 1999: 174f).

#	a	a:	ɛ:	ɛ	e:	ɪ	i:	ɔ	o:	ʊ	u:	œ	ø:	ʏ	y:	‹ei›	‹eu›	‹ei›	wrong	**correct**
a		8																	8	**316**
a:	5				2														7	**317**
ɛ:				1	64										1	1			67	**257**
ɛ			8		8							2							18	**306**
e:			22	1			1												24	**300**
ɪ				2			3							2					7	**317**
i:						2		1											3	**321**
ɔ									7										7	**317**
o:								3			4								7	**317**
ʊ								1			5								6	**318**
u:										3									3	**321**
œ													6						6	**318**
ø:									1			3			2				6	**318**
ʏ						1				1		2	1		8				13	**311**
y:											1		2	3					6	**318**

Table 10.2: Confusion matrix in tokens. Wrong and correct identifications for German native listeners (N=18, 4860 responses)

The present perceptual study shows that in non-word items the /e:/ - /ɛ:/ contrast is reduced, i.e. the contrast is not perceived by many L1 German listeners. However, for /ɛ:/-stimuli variation with respect to the participants regional origin is observed, especially in the response behaviour of subjects from Vorarlberg and Thüringen. These subjects as well as linguistically trained people (two subjects) were able to identify /ɛ:/ correctly. The other subjects in the control group showed uncertainties in the correct identification of /ɛ:/ that are also reflected by wrong or presumably also hypercorrect identifications of /ɛ:/ with /e:/ (6.8%).

%	ɑ	ɑ:	ɛ:	ɛ	e:	ɪ	i:	ɔ	o:	ʊ	u:	œ	ø:	ʏ	y:	wrong	**correct**
a	**98**	3														2,5	**97,5**
a:	2	**98**			1											2,2	**97,8**
ɛ:			**79**		20											20,7	**79,3**
ɛ			3	**94**	3							1				5,6	**94,4**
e:			7		**93**											7,4	**92,6**
ɪ				1		**98**	1							1		2,2	**97,8**
i:						1	**99**	0								0,9	**99,1**
ɔ								**98**	2							2,2	**97,8**
o:								1	**98**		1					2,2	**97,8**
ʊ										**98**	2					1,9	**98,1**
u:										1	**99**					0,9	**99,1**
œ												**98**	2			1,9	**98,1**
ø:												1	**98**		1	1,9	**98,1**
ʏ												1		**96**	3	4,0	**96,0**
y:													1	1	**98**	1,9	**98,1**
Σ	6,6	6,7	5,9	6,4	7,7	6,6	6,7	6,6	6,7	6,6	6,8	6,7	6,7	6,5	6,8	*3,9*	***96,1***

Table 10.3: Confusion matrix in %. Wrong and correct identifications for German native listeners (N=18, 4860 responses). (values < 0.5 omitted)

Qualitative difficulties are slightly more frequent with /o:/-stimuli and with ü- and ö-vowels than with other categories. The largest range of different categories is selected for /ʏ/-stimuli: /y:/ (8 cases), /œ/ (2 cases), and /ø: ʊ ɪ/ (1 case each).

“Perceptual” unrounding, i.e. the substitution of a [+round] input stimulus with a [-round] response category is only once observed with /ʏ/ (/ʏ/ > /ɪ/ (1 case).

The identification of /i:/ with /ɔ/ and of /ɛ:/ with /y:/ could be interpreted as the participant’s error in the choice of the correct response option on the work sheet rather than as instance of perceptual substitution.

10.2.1.4 Similarity and distance

The similarity matrix in Table 10.4 represents the similarity scores calculated by Shepard’s formula (see section) for all vowel pairs. Sim_scores represent the relative similarity of the two vowels involved in a given contrast.

The highest sim_score is observed for /e:/ - /ɛ:/ (0.639). Higher scores are also obtained for the short-long pairs /œ/ - /ø:/ (0.348), /ɑ/ - /ɑ:/ (0.347), /ʏ/ - /y:/ (0.325), /ɛ/ - /e:/ (0.306), /ɔ/ - /o:/ (0.285), /ʊ/ - /u:/ (0.280), /ɛ/ - /ɛ:/ (0.260), and /ɪ/ - /i:/ (0.243). Instances of confusion between front rounded vowels are reflected by sim_scores between 0.2 and 0.1 for /ʏ/ - /œ/

(0.186), /ʏ/ - /ø:/ (0.126), and /y:/ - /ø:/ (0.161). For all other contrast pairs, lower sim_scores are obtained that are due to only a few instances of wrong categorization.

V	ɑ	ɑ:	ɛ:	ɛ	e:	ɪ	i:	ɔ	o:	ʊ	u:	œ	ø:	ʏ	y:
a	1.000														
ɑ:	0.347	1.000													
ɛ:	0.010	0.000	1.000												
ɛ	0.008	0.000	0.260	1.000											
e:	0.000	0.003	0.639	0.306	1.000										
ɪ	0.000	0.000	0.015	0.081	0.048	1.000									
i:	0.000	0.000	0.000	0.000	0.006	0.243	1.000								
ɔ	0.000	0.000	0.000	0.000	0.000	0.000	0.000	1.000							
o:	0.000	0.000	0.000	0.000	0.000	0.000	0.003	0.285	1.000						
ʊ	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.080	0.050	1.000					
u:	0.000	0.000	0.000	0.000	0.000	0.000	0.003	0.000	0.032	0.280	1.000				
œ	0.000	0.000	0.040	0.000	0.008	0.000	0.000	0.042	0.014	0.008	0.011	1.000			
ø:	0.000	0.000	0.010	0.000	0.000	0.000	0.000	0.008	0.070	0.008	0.004	0.348	1.000		
ʏ	0.000	0.000	0.000	0.000	0.004	0.004	0.004	0.000	0.012	0.013	0.000	0.186	0.126	1.000	
y:	0.000	0.000	0.000	0.000	0.000	0.004	0.003	0.003	0.007	0.004	0.013	0.018	0.161	0.325	1.000

Table 10.4: Similarity matrix for native German listeners

10.2.1.5 The perceptual vowel map for native German listeners (L1 control group)

From the similarity scores calculated by Shepard's formula, distance scores and non-metric MDS solutions were derived. For the non-metric MDS solutions, RSQ values of 0.170 in dimension 1, 0.461 for the two-dimensional solution, and 0.624 for the three-dimensional solution are obtained.

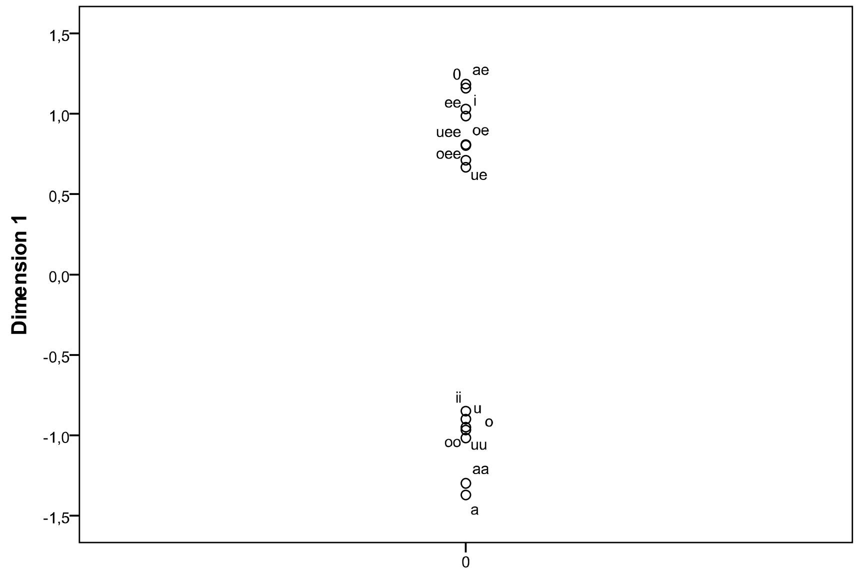

Figure 10.3: One-dimensional MDS representation of the perceptual map of German vowels for native German listeners (control group) (N=18, 4860 responses, RSQ .170)

The one-dimensional MDS representation (Figure 10.3) has a very low RSQ of 0.170 and must therefore be considered as of limited significance.

The two-dimensional MDS solution in Figure 10.4 represents the perceptual space for the German vowel system as perceived by the native German participants. The MDS solution is based on distance scores calculated from the similarity scores presented above. Differences in type size represent the relative preference ranking for each of the categories, though in the L1 German sample no strong asymmetries in preference scores are observed and the response categories are selected with almost equal frequency.

The diagram shows four to five major clusters: (1) German a-vowels that are well distinguished from all other categories, (2) back rounded vowels that are perceptually near to each other due to single cases of wrong identifications in the identification test but are sufficiently differentiated from all other vowel qualities, (3) front rounded vowels that are in a few cases confused with each other but are sufficiently distinguished from other qualities, and (4) front non-rounded vowels that divide into two sub-clusters: (a) long constricted /e:/ and /i:/ and (b) unconstricted /ɛ/ and /ɪ/ together with long unconstricted /ɛ:/.

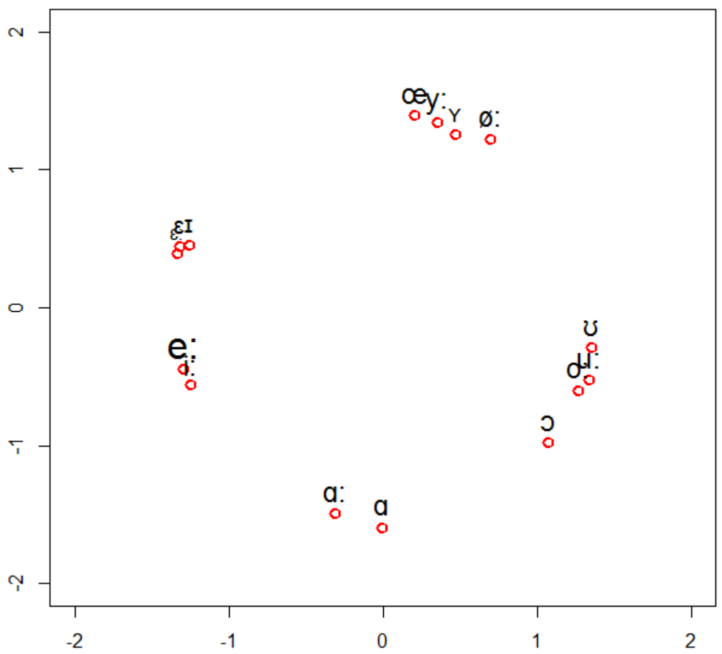

Figure 10.4: Two-dimensional MDS representation of the perceptual vowel map for native German listeners (control group) (N=18, 4860 responses, RSQ .461)

The two-dimensional MDS solution for native German listeners shows a very low RSQ of 0.461 indicating a rather low goodness of fit. This effect is strongly conditioned by the very low error rate in the sample and by the resulting high number of zero-values in the similarity matrix causing equal values in the distance matrix for these two languages. As discussed in section 7.3 a low error rate decreases the goodness of fit for the MDS solution obtained. However, the solution in its present form can be regarded as good basis of comparison of system as perceived by German native speakers compared to the L2 learners' sub-groups, especially for a comparison with those languages that have significantly less vowel phonemes in their L1 system.

The three-dimensional solution gives a slightly more differentiated representation for the class of front rounded vowels. Moreover, e-qualities are more differentiated in this solution.

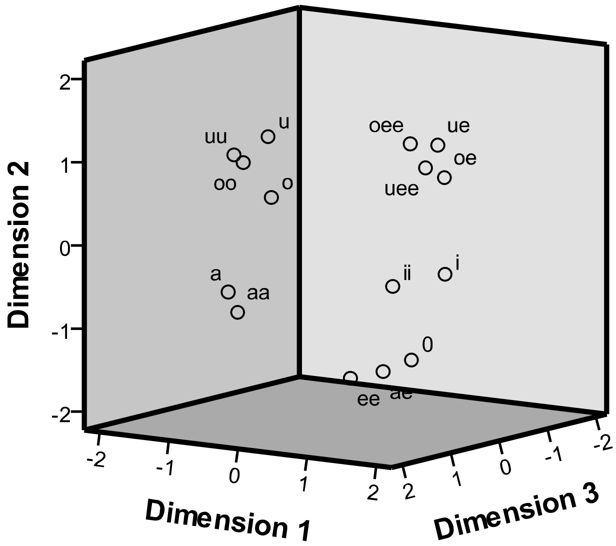

Figure 10.5: Three-dimensional MDS representation of the perceptual vowel map for native German listeners (control group) (N=18, 4860 responses, RSQ .624)

10.2.2 Summary

To summarize, the native German control group shows an overall percentage of correct responses of 96%. No strong asymmetries are observed in the listeners' preferences for specific response categories with the only exception of /e:/ and /ɛ:/. 20% of the responses for /ɛ:/ stimuli were attracted by /e:/ (vs. 7% /ɛ:/ > /e:/). Due to these difficulties, e-vowels receive the lowest percentage correct responses, followed by short unconstricted /ʏ/. All other categories receive high id_correct scores above 97%. Apart from difficulties with e-categories, wrong identifications were mostly due to the incorrect categorizations in terms of constriction degree and phonemic length, whereas wrong categorizations in terms of constriction location occur only very marginally (see Table 10.3).

10.3 Albanian

10.3.1 General characteristics

Albanian is spoken by around six million people in Albania, in Kosovo, some parts of Montenegro, Serbia, Italy and Macedonia and by about one million emigrants in Turkey, North America, Europe, and Australia. Albanian forms a branch of its own within the family of Indo-European languages.

Two main dialects are distinguished: Gheg in the North and Tosk in the South. Gheg is spoken in Kosovo and the outer North-East of Albania, in Middle-Albania, in some parts of Macedonia, and around Zadar. Tosk comprises the varieties in the North and East of Albania, the Labian and the Camen varieties. Almost all diaspora varieties belong genetically to the Tosk group (Fiedler 2003: 750).

Modern Standard Albanian as a unified standard of both dialects is relatively young and was proclaimed in 1972 in Tirana. It is mainly based on the Tosk variety, but still contains some Gheg features. Before 1972, both Tosk and Gheg had been used in their literary form (Fiedler 2003). Differences between Gheg and Tosk refer to phonetics, phonology, morphology, syntax, and the lexicon. Significant differences are observed in the vowel inventory.

10.3.2 Phonological and phonetic description

Modern Standard Albanian uses seven graphemes <a, e, ë, i, o, u, y> to distinguish seven vowel phonemes (see Table 10.5). The vowel systems differentiates the five basic vowels /i e u o a/, a front rounded vowel /y/ and a so-called schwa vowel that is orthographically represented as <ë>. The consonant inventory of Modern Standard Albanian consists of 29 consonants (see Table 10.6).

	front			back
high	i	y		u
mid	e		ə	o
Low			a[73]	

Table 10.5: Albanian vowels in terms of traditional phonological features

[73] For discussion of /a/ as a front or back vowel see Moosmüller & Granser (2003).

	bilabial	labio-dental	dental	alveolar	post-alveolar	palatal	velar	glottal
plosives	p b			t d		c ɟ	k g	
nasals	m		n			ɲ	ŋ	
fricative		f v	θ ð	s z	ʃ ʒ			h
affricate				ʦ ʣ	ʧ ʤ			
lateral				l ɫ				
trill				r				
flap				ɾ				
approximant						j		

Table 10.6: Albanian consonant inventory

Phonological differences between Gheg and Tosk (see Lloshi 1999: 285) mainly refer to the status and quality of the vowel <ë>, which is very frequent in Tosk where it occurs also in stressed position, and its Gheg nasal correspondent <â>, (e.g. *hân – hënë* 'moon'), the existence of phonemically relevant vowel quantity in Gheg, the existence of nasal qualities (only in Gheg), the diphthong *ue* or long *u* in Gheg and its Tosk equivalent *ua* (e.g. *due, du – dua* 'I want', *grue, gru – grua* 'woman'), and the very rare occurrence of /ø:/ and /æ:/ as phonemes in North(West)-Gheg. South Tosk and diaspora varieties do not use /y/ (Fiedler 2003: 751).

In Tosk, vowel nasalization only occurs before nasal consonants, whereas in Gheg the occurrence of nasal vowels is not context-dependent.

A considerable range of phonetic variation can be observed among speakers from different regions especially for the phonemes /u/, /o/, /a/ and <ë>. While variability of /o/ and /u/ is solely phonetically motivated by context, the variation of <ë> and /a/ is motivated by other factors (Moosmüller & Granser 2003).

<ë> in unstressed syllables can occur in both varieties, its phonetic realization shows no dialectal differences in unstressed positions (Granser & Moosmüller 2001). While in some regional varieties <ë> is preserved in final unstressed positions (Eastern North-Tosk), it is deleted in final unstressed position in others (Western North-Tosk). In Gheg and in South-Tosk, <ë> is deleted in final positions with compensatory lengthening of the former penultima (Fiedler 2003: 752).

Especially, for the schwa vowel <ë>[74], a high amount of variation among regional varieties is observed. The status of <ë> as a phoneme is canonically restricted to the South-Tosk variety

[74] The Albanian schwa will be represented here as /ë/which corresponds to the IPA notation [ə].

where it occurs in stressed syllables (Granser & Moosmüller 2001) and corresponds to a Gheg nasal [ã] or nasal [e] (Fiedler 2003). Tosk speakers tend to realize <ë> as a front vowel, while South-Gheg speakers rather use back and central qualities and North-Gheg speakers clearly prefer a so-called "central" articulation for the schwa. But note that there is a high within-speaker and within-item variability observed among speakers from all varieties (Moosmüller & Granser 2003). For Gheg varieties, <ë> is commonly described as a back rounded quality /ü/, though phonetically there is considerable variation among speakers of Gheg varieties (Granser & Moosmüller 2001: 317).

In an acoustic analysis of realizations of <ë>, Moosmüller & Granser (2003) observe realizations ranging from front [ɛ] to back non-rounded non-nasal [ʌ]. Moosmüller & Granser (2006) describe a range of variability from back [ɤ, ʌ] to front [e, æ] and even front rounded [œ]. The realization as [ʌ] is clearly different from Gheg dialectal varieties where schwa in stressed positions is realized as a back nasalized rounded vowel (Beci 1995, cited in Moosmüller & Granser 2003). To summarize, the high variability of schwa realizations is clearly not phonetically motivated by context factors but rather due to regional or even social variation.

Similarly to schwa, the vowel /a/ exhibits a high degree of phonetic variability (Moosmüller & Granser 2003). A tendency towards back articulation as [ɑ] and sometimes even rounding can be observed. While Tosk speakers show a tendency towards front articulation, with Gheg speakers more back articulation is observed with a considerable degree of within-item variability in both varieties. Gheg dialects distinguish a front [a] and a back vowel [ɑ] (Beci 1995, cited in Moosmüller & Granser 2003). In Gheg, frontness usually correlates with short vowel duration while backness is correlated with longer duration. Long [ɑ:] mostly results from schwa deletion with compensatory lengthening of the former penultima as in /plak/ [plak] "old man" vs. /plakë/ [plɑ:k] "old woman". This is typical for Albanian speakers from Kosovo who preserve this feature even in the Standard variety (Fiedler 2003). Beci (1995, cited in Moosmüller & Granser 2006) reports that Gheg dialects distinguish four different low vowels: a short front /a/ and a long back /ɑ:/ differing only in the front/back dimension and each occurring as oral or nasal vowel.

Moosmüller & Granser (2003) observe considerable variation of F2 values for /a/ reflecting articulatory differences ranging from back over central to front articulation of /a/.

The other vowel qualities /e, i, y/ show less phonetic variability and therefore do not convey any regional or social information (Moosmüller & Granser 2003).

The existence of phonological length distinction is considered to be one of the most conspicuous differences between the Gheg and the Tosk vowel system which is reflected in both varieties in documents written before Modern Standard Albanian was established in 1972 (Fiedler 2003: 752). Phonetically, Albanian vowels clearly differ in duration (e.g. vowels in open syllables are longer than in closed syllables) but it is not clear whether these differences in duration are phonologically relevant length distinctions. Stressed syllables usually are tenser and stronger in their articulation. In Standard-Albanian and in dialects which do not use phonological length, the longer duration is an important perceptual cue for stressed syllables (Fiedler 2003: 760). South-Tosk varieties and the diaspora varieties Arbëresh and Arvanitë show phonologically relevant differences in quantity. Generally, length distinctions only occur in stressed syllables. Fiedler (2003) mentions some Gheg varieties that exhibit even three degrees of length with distinctions between [+/–long] and [+/–short]. Fiedler (2003) mentions several positional effects on vowel quantity: Vowels are longer in open syllables than in closed syllables, longer before /r/, /ɫ/ and /j/ than before plosives, longer before voiced consonants than before voiceless, longer before single consonants than before consonant clusters, and longer if they are close to the stressed syllables compared to more distant syllables within a word. The more syllables a word contains, the shorter the single syllables will be.

Albanian has falling and rising diphthongs like /ua, ue, ye, ie (je)/ and even triphthongs such as /uaj yej iej (jej)/. In Standard Albanian, rising diphthongs only occur with /j/ and "peripheral" vowels.

Word accent in Albanian is dynamic and falls mainly on the penultima or the final syllable. Final open and closed syllables containing /a i u/ usually are stressed, final syllables containing schwa are never stressed. Final syllables with /o/ or /e/ tend to be unstressed if open and stressed if closed (Trommer & Grimm 2004).

10.3.3 Contrastive analysis

The phonological and phonetic description above reveals several similarities as well as differences between the German and the Albanian vowel system: Both systems have a complete set of vowels at the five major constriction locations. Similar to German, Albanian possesses a high front rounded vowel /y/.

"New" for Albanian learners of German are the phonemes /œ/, /øː/ and /ɛː/ (if we assume the existence of /ɛː/ as a phoneme in German) and the graphemes <ö>, <ü> and <ä>. Though /y/

exists as a phoneme in Albanian, its orthographic representation and the association with the German grapheme <ü> may cause difficulties.

The pronunciation of /œ/ and /ø:/ is generally described as difficult for Albanian learners. In pronunciation, [e]-qualities are reported as substitutes for both front rounded qualities (Schader 2008, 2009, 2010; Granser 2010). /ø/ and /œ/ are reported to be pronounced as e-qualities by Albanian learners of German (Schader 2008, 2010). Therefore, difficulties in the perceptual differentiation of /œ/, /ø:/ and other "similar" German vowel sounds are expected. Possible perceptual substitutes would be either front non-rounded vowels of the same height or the "similar" phoneme /y/ which exists in Albanian and German.

The correct distinction of short lax /ɛ/ and /ɔ/from their constricted counterparts /e/ and /o/, respectively, is described as difficult in pronunciation (Schader 2008, 2010; Granser 2010). Therefore, we generally may expect difficulties with the correct identification of different e- and o-qualities in German.

Since duration has no phonological relevance in Albanian, difficulties are expected also with respect to vowel length distinctions in German.

With persons deeply ingrained in Gheg dialects, dialectal influence may result in the insufficient distinction of German /ɔ/ vs. /ɑ/ or /a/ (Schader 2008, 2010).

10.3.4 Results and discussion

Twelve Albanian subjects were tested, one beginner, nine advanced and two very advanced learners. Nine participants are speakers of Gheg and three of Tosk. Due to the small number of Tosk speakers the data set is not large enough for a reliable comparison of Tosk and Gheg.

The data set discussed here consists of 3240 valid responses (in 174 cases answers were missing). For each vowel category, 201 to 206 valid responses were delivered.

10.3.4.1 Difficulties

The diagrams in Figure 10.6 and Figure 10.7 show the relative difficulty of German vowel categories for L1 Albanian learners.

The highest id_correct scores are observed for German /a:/ (85%), /i:/ (82%) and /u:/ (78%). Most difficulties are observed with /ɛ:/ (36.% correct). Other rather difficult categories are /ʏ/ (47%), /ø:/ (58%), /ɛ/ (59%), and /œ/ (61%). /y:/ (75% correct) is a comparatively stable category and shows clearly less wrong answers than all other palatal rounded vowels, which can be explained by the existence of /y:/ in L1 Albanian but also by the cross-language preference for this particular category within the class of palatal rounded vowels.

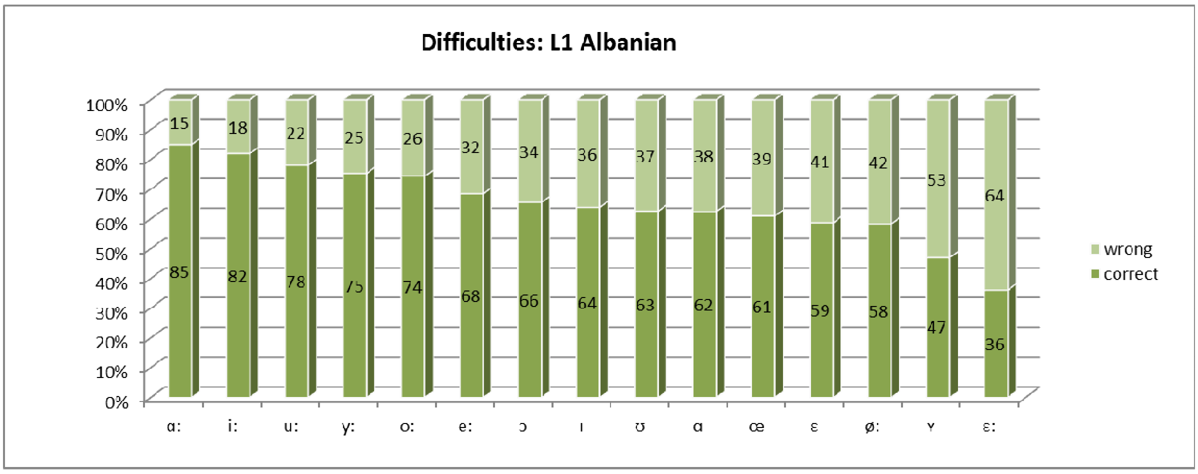

Figure 10.6: Correct and wrong identifications in % for L1 Albanian listeners

The listeners' mean id_wrong score was 35% (14% standard deviation), the median was 34%. The listener with the highest percentage of correct answers had an id_wrong of 14%, the highest id_wrong score was 64% wrong responses.

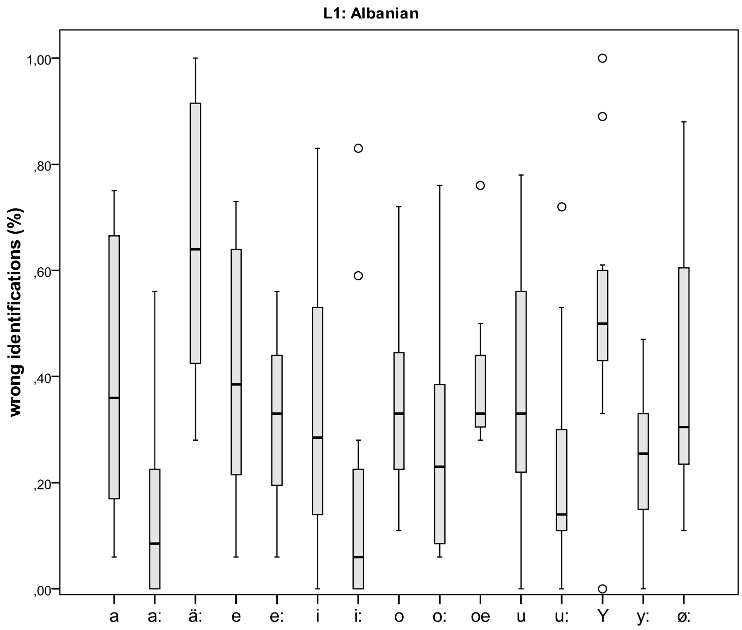

Figure 10.7: Boxplot diagram with id_wrong scores for L1 Albanian listeners

The boxplot diagram in Figure 10.7 shows the range of interpersonal variation for each of the categories tested. Low id_wrong scores together with least variation are given for /a:/, /i:/ and /y:/. These categories can therefore be considered to be most stable. Most inter-personal variation is observed with /ʏ/ which has a high id_wrong score (53% wrong) with two extreme values above 80% and a minimum value of 0% wrong identifications[75].

10.3.4.2 Preferences

Figure 10.8 shows the preference patterns for Albanian learners. The preference scores for monophthong categories vary within a range of 5.4% (min /ɛ:/ 4.8%, max /e:/ 10.2%). Clear differences between categories are observed. /e:/, /a:/ and /y:/ are the most preferred response options, followed by /o:/, /ø:/ and /u:/, while /ʏ/ and /ɛ:/ are the least preferred response categories.

The most preferred response category is /e:/ and the least preferred /ɛ:/. The main reason for the preference for /e:/ is a strong tendency to identify /ɛ:/ as /e:/ in 52% of its occurrences (see Table 10.7).

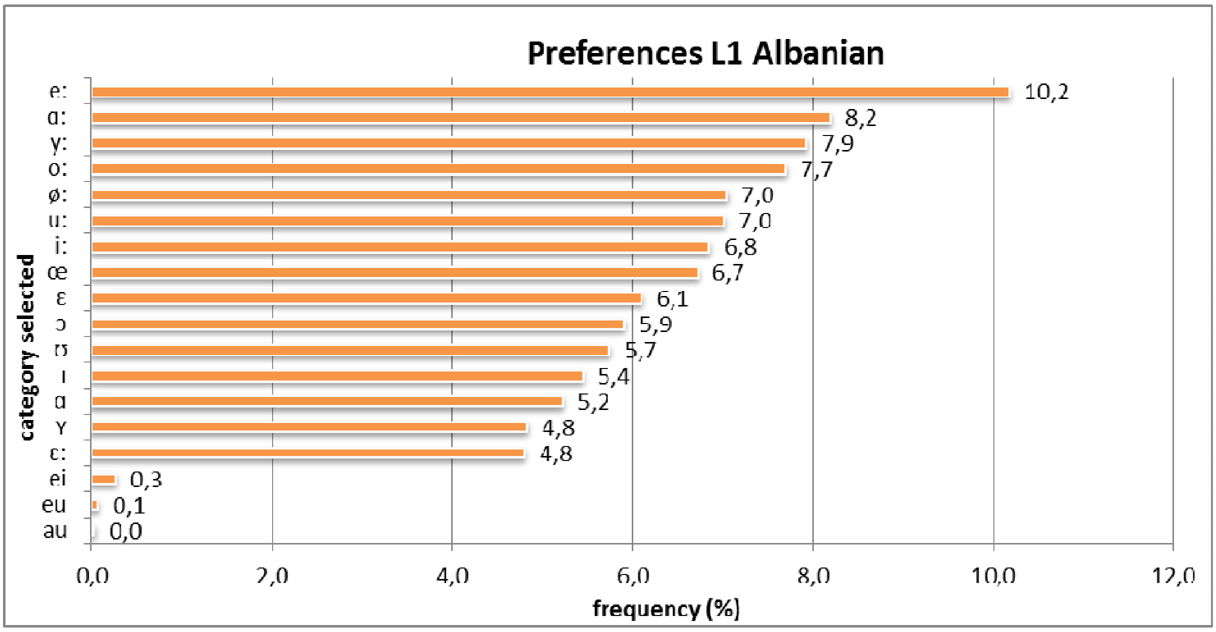

Figure 10.8: Preferred response categories for L1 Albanian listeners

Another interesting pattern is a general preference for long rather than short response categories. All long vowel categories are selected more frequently (10.2 to 6.8%) than the short categories (6.7 to 4.8%), indicating general uncertainties with the distinction of phonemic length and its acoustic correlates.

[75] One participant (Alb3) identified all /y:/ occurrences correctly but distributed all answers for /ʏ/ across several categories other than /ʏ/.

Diphthongs are selected as response category in only eight cases: <au> only once for /o:/, <eu> in two cases (for /ɛ/ and /o:/), and <ei> in six cases for different categories: /a:/ (1), /ɛ/ (2), /e:/, /ɪ/ and /i:/ (each only once) and /ʏ/ (2 cases).

Within the class of front rounded vowels, /ʏ/ is the least preferred response option (4.8%) followed by /œ/ (6.7%) and /ø:/ (7.1%). The most preferred category is /y:/ (7.9%) which has a cognate in Albanian and seems to act as a kind of attractor for other front rounded vowels, especially for /œ/ and /ø:/ (see the qualitative discussion below).

To summarize the quantitative analysis of wrong identifications, a clear asymmetric relation between the given response labels is observed. Long categories are generally more preferred than short ones. /ɛ:/ and /ʏ/ are the least preferred categories. A qualitative description of perceptual substitutions will be given in the next section.

10.3.4.3 Patterns of confusion

Patterns of category confusion for Albanian learners of German are presented in Table 10.7.

ɑ-qualities are very stable and clearly distinct from other vowel qualities in perception. Phonemic length distinctions, however, cause some difficulties: Long /ɑ:/ is not identified correctly in 15% of the cases. 37% of the /ɑ/-stimuli are hypercorrectly identified as long /ɑ:/.

/ɛ:/ and its orthographic representation <ä:> cause a considerable number of wrong identifications: it is substituted with /ɛ/ in 11%, with /e:/ in 52%, and with <ei> in 1% of the cases. Problems with /ɛ:/ interfere with the correct identification of short /ɛ/. However, the detection of length is performed correctly in the larger part of the cases, i.e. confusion of long /ɛ:/ with long /e:/ and vice versa is observed more frequently (52% /ɛ:/ > /e:/ vs. 15% /e:/ > /ɛ:/) than the substitution /ɛ:/ > /ɛ/ (11%).

The contrast of /e:/ vs. /i:/ is clearly distinguished by Albanian learners of German, i.e. No substitutions of /i:/ > /e:/ are observed, and /e:/ > /i:/ occurs in only 1% of the cases.

/ɪ/ (64% correct) is not clearly differentiated from other palatal categories and is identified either as /i:/ (19%), as /ɛ/ (9%), or as /e:/ (5%).

%	ɑ	ɑ:	ɛ:	ɛ	e:	ɪ	i:	ɔ	o:	ʊ	u:	œ	ø:	ʏ	y:	‹ei›	‹eu›	‹au›	wrong	**corr**
a	**62**	37	1																38	**62**
a:	15	**85**																	15	**85**
ɛ:			**36**	11	52											1			64	**36**
ɛ	1		14	**59**	25														41	**59**
e:			15	14	**68**	1	1												32	**68**
ɪ			1	9	5	**64**	19												36	**64**
i:						17	**82**												18	**82**
ɔ								**66**	31			3							34	**66**
o:								9	**74**	2	4	1	6		1				26	**74**
ʊ								10	4	**63**	20	1	1	1					37	**63**
u:									1	20	**78**				1				22	**78**
œ			4					2	1		1	**61**	27	1	1				39	**61**
ø:			1					1	3			14	**58**	5	17				42	**58**
ʏ									1			19	8	**47**	22	1			53	**47**
y:												1	4	18	**75**				25	**75**
Σ	*5*	*8*	*5*	*6*	*10*	*5*	*7*	*6*	*8*	*6*	*7*	*7*	*7*	*5*	*8*	*.3*	*.1*	*.0*	*34.7*	***65.3***

Table 10.7: Confusion matrix for Albanian listeners (N=12, 3066 responses). Correct (in bold) and wrong identifications in % (presented stimuli in rows, selected categories in columns in semi-orthographic representation, values < 0.5 not indicated)

While the qualitative differentiation in terms of constriction location of velar u-vowels vs. uvular o-vowels does not cause many difficulties to Albanian listeners of German, the differentiation of phonemic length and degree of constriction of o- and u-qualities is more difficult. Most dispersion is observed for /ʊ/, which is correctly identified in 63% of the cases and incorrectly categorized as /u:/ (20%) or /ɔ/ (10%) but only in 4% as /o:/.

/u:/ is identified correctly in 78% of its occurrences. /o:/ is no attractive candidate for /u:/ (only 1%).

/o:/ (74% correct) is incorrectly identified as /ɔ/ (9%), /ʊ/ (2%), or /u:/ (4%). Interestingly, another attractive substitute for /o:/ is front /ø:/ (6%). A parallel pattern is observed with /ø:/ that is categorized as /o:/ in 3% of the cases indicating a perceived similarity of /o:/ and /ø:/ for Albanian listeners. However, this pattern may also be due to the orthographic similarity of o- and ö-vowels and a tendency towards hypercorrect answers.

Considerable confusion is observed among *front rounded vowels* with respect to quality and quantity:

/y:/ is identified correctly in 75%, it is incorrectly substituted with /ʏ/ (18%), /ø:/ (4%), and /œ/ (1%), and attracts responses from /ʏ/ and /ø:/ (/ʏ/ > /y:/ 22%, /ø:/ > /y:/ 17%).

Among front rounded qualities, /ʏ/ (53% wrong) is the most difficult category that is not clearly distinguished from other members of this class. The listeners' responses for /ʏ/ are distributed over all other front rounded categories (/y:/ 22%, /œ/ 19%, /ø:/ 8%), a pattern that is observed with almost all listeners.

Similarly, answers for /ø:/ (42% wrong) are distributed across all front palatal categories (/y:/ 17%, /œ/ 14%, and less often /ʏ/ 5%) but also to /o:/ (3%) and /ɔ/ (1%). /ø:/ is attracted to /y:/ in 35 cases (17%), whereas /y:/ is substituted with /ø:/ only in 9 cases (4%). This perceptual asymmetry can be interpreted as an "equivalence classification" due to transfer of Albanian /y/, i.e. the "new" quality /ø:/ seems to be identified as similar or as a poor exemplar of native Albanian /y/.

Responses for /œ/ show a different pattern: /œ/ is correctly identified in 61% of the cases and substituted with /ø:/ in 27% and /ɛ:/ <ä> in 4%. A substitution of /œ/ with /y:/ and /ʏ/ but also with /u:/ and /o:/ is observed in only 1% of the cases respectively. A categorization of /œ/ as /ɔ/ occurs in only 2% of the cases. A tendency towards a categorization with non-front qualities is observed only with five advanced learners of the twelve participants and might be related to the general perceptual confusion among front rounded vowels.

In production in L2 German, the front rounded vowels /ø:/ or /œ/ are reported to be pronounced as e-qualities by native speakers of Albanian (Schader 2008, 2009,). However, these substitutions do not seem to be primarily motivated from perception since in the present data substitutions of [œ] > /ɛ/ are not very frequent and [œ] or [ø:] > /e:/ does not occur. Rather, we find an interference of the orthographic representation, where [œ] is identified as /ɛ:/ <ä> in eight cases (4%), a substitution pattern that could of course be due to an underlying perceptual substitution of /œ/ with /ɛ/, but is also likely to be explained by the graphemic influence of <ä:>.

The two-dimensional perceptual map in the MDS solution for Albanian (see Figure 10.10) also does not suggest a perceptual proximity of German /ø:/ and /œ/ to /e:/ or /ɛ/. To conclude, perceptual "unrounding" of German front rounded vowels is only observed for short unconstricted /œ/, where it seems partly influenced by German orthographic patterns. To summarize, a look at the confusion matrix shows category confusion for three major classes of vowels: (1) front unrounded vowels, (2) front rounded vowels, and (3) back rounded vowels. The analysis of the similarity scores for these three classes reveals some interesting aspects of category confusion that will be discussed in the next section.

10.3.4.4 Similarity and distance

Based on Shepard's formula, similarity scores are calculated for each vowel contrast (see section 7.2.5). Table 10.8 presents the similarity scores for L1 Albanian participants.

	ɑ	ɑ:	ɛ:	ɛ	e:	ɪ	i:	ɔ	o:	ʊ	u:	œ	ø:	ʏ	y:
a	1.000														
ɑ:	0.347	1.000													
ɛ:	0.010	0.000	1.000												
ɛ	0.008	0.000	0.260	1.000											
e:	0.000	0.003	0.639	0.306	1.000										
ɪ	0.000	0.000	0.015	0.081	0.048	1.000									
i:	0.000	0.000	0.000	0.000	0.006	0.243	1.000								
ɔ	0.000	0.000	0.000	0.000	0.000	0.000	0.000	1.000							
o:	0.000	0.000	0.000	0.000	0.000	0.000	0.003	0.285	1.000						
ʊ	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.080	0.050	1.000					
u:	0.000	0.000	0.000	0.000	0.000	0.000	0.003	0.000	0.032	0.280	1.000				
œ	0.000	0.000	0.040	0.000	0.008	0.000	0.000	0.042	0.014	0.008	0.011	1.000			
ø:	0.000	0.000	0.010	0.000	0.000	0.000	0.000	0.008	0.070	0.008	0.004	0.348	1.000		
ʏ	0.000	0.000	0.000	0.000	0.004	0.004	0.004	0.000	0.012	0.013	0.000	0.186	0.126	1.000	
y:	0.000	0.000	0.000	0.000	0.000	0.004	0.003	0.003	0.007	0.004	0.013	0.018	0.161	0.325	1.000

Table 10.8: Similarity matrix for L1 Albanian listeners

In the following analysis, the similarity scores will be discussed by place of constriction.

For *front unrounded vowels*, the highest similarity scores are observed with contrast pairs involving different e-qualities. The contrast /e:/ - /ɛ:/ is particularly difficult because of the special status of German /ɛ:/ <ä> (sim_score 0.639). Sim_scores for contrast pairs involving /ɪ/ are also quite high. The most difficult differentiation with vowel pairs involving /ɪ/ is the contrast of /i:/ - /ɪ/ (0.243), whereas /i:/ is differentiated well from all other front vowels, even from /e:/.

front unrounded Vs	*sim_scores*
e: - ɛ:	.639
e: - ɛ	.306
ɛ - ɛ:	.260
i: - ɪ	.243
ɪ - ɛ	.081
ɪ - e:	.048
ɪ - ɛ:	.015
i: - e:	.006
i: - ɛ	.000
i: - ɛ:	.000

Table 10.9: Similarity scores of German front unrounded vowels for L1 Albanian listeners

For *front rounded vowels*, the highest sim_scores are given for the vowel pairs /œ/ - /ø:/ (0.348) and /ʏ/ - /y:/ (0.325). In the /œ/ - /ø:/ and the /ʏ/ - /y:/ contrast pairs, "only" the degree of constriction together with durational differences indicating phonemic length is not sufficiently differentiated by L1 Albanian listeners. For /œ/ - /ʏ/ (0.186) and /ø:/ - /y:/ (0.161), differences in constriction location (mid- vs. pre-palatal) are not appropriately distinguished. The substitution of German /ø:/ with /y:/ is moreover motivated by an equivalence classification with Albanian /y/.

front rounded Vs	*sim_scores*
œ - ø:	.348
ʏ - y:	.325
œ - ʏ	.186
ø: - y:	.161
ʏ - ø:	.126
œ - y:	.018

Table 10.10: Similarity scores of German front rounded vowels for L1 Albanian listeners

The contrasts /ʏ/ - /ø:/ and /œ/ - /y:/ have lower similarity scores and are therefore better discriminated than the contrasts above. /œ/ - /y:/ which are most distinctive in terms of number of features have the lowest sim_score (0.018) in this class. However, there are cases where perceptual under-differentiation of /ʏ/ - /ø:/ and /œ/ - /y:/ occurs, i.e. neither differences in constriction location nor in constriction degree and length are appropriately distinguished by the listeners. We can conclude for front rounded vowels that the difference between the pre-palatal and the mid-palatal constriction location is better discriminated by Albanian listeners than differences in phonemic length or degree of constriction.

Sim_scores for velar and uvular vowels (Table 10.11) are clearly below the values for front vowels indicating that these qualities are better differentiated in perception than front vowels. The sim_scores reveal that the highest similarity is found with vowel pairs that are located at the same constriction region but differ with respect to constriction degree and phonemic length, i.e. with /o:/ - /ɔ/ (0.285) and /u:/ - /ʊ/ (0.280). All other contrast pairs show a considerably lower degree of similarity below the 0.01 mark, even if some instances of confusion are observed in the data.

velar and uvular Vs	*sim_scores*
o: - ɔ	.285
u: - ʊ	.280
ʊ - ɔ	.080
ø: - o:	.070
ʊ - o:	.050
œ - ɔ	.042
u: - o:	.032
o: - œ	.014
ʊ - ʏ	.013
u: - y:	.013
o: - ʏ	.012
u: - œ	.011

Table 10.11: Similarity scores for German velar and uvular vowels for L1 Albanian listeners

10.3.4.5 The perceptual vowel map for L1 Albanian listeners

From the similarity scores calculated by Shepard's formula, *distance scores* and one-, two- and three-dimensional MDS solutions were derived. For the non-metric MDS solutions, RSQ values of 0.505 in dimension 1, 0.733 in dimension 2, and 0.814 are obtained.

The one-dimensional solution (0.505 RSQ) in Figure 10.9 shows a strict separation of rounded vs. non-rounded vowels.

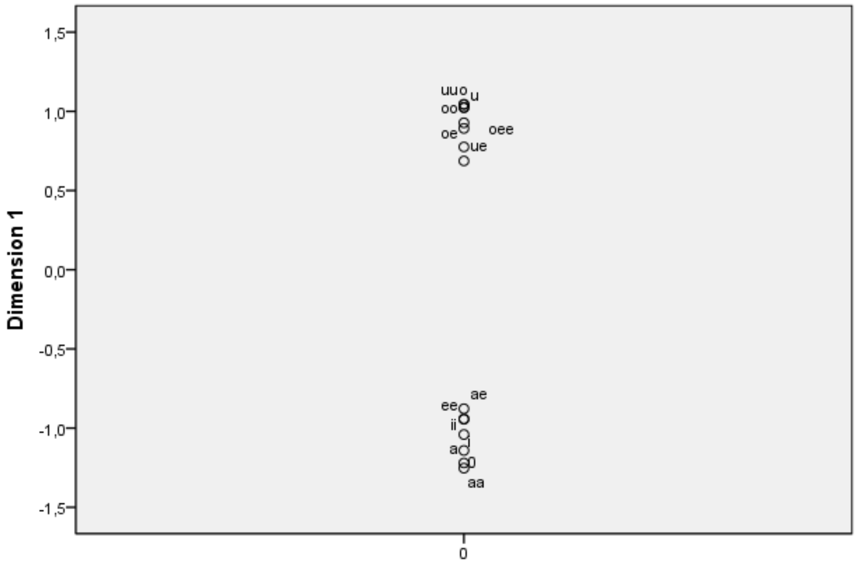

Figure 10.9: One-dimensional MDS representation of the perceptual vowel space for L1 Albanian learners of German (N= 12, 3066 responses, RSQ .505)

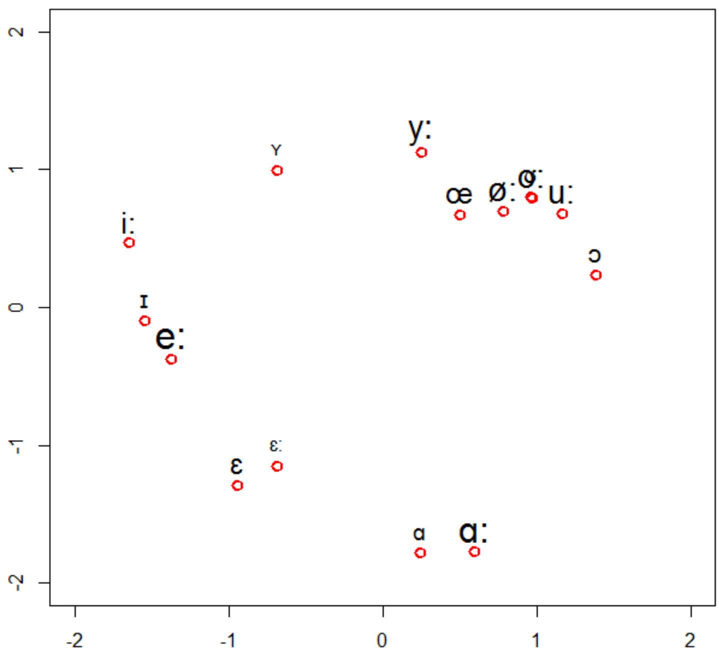

Figure 10.10: Two-dimensional MDS representation of the perceptual map of German vowels for L1 Albanian listeners (N= 12, 3066 responses, RSQ .733)

The two-dimensional MDS solution in Figure 10.10 represents the perceptual vowel map for L1 Albanian learners of German. Differences in type size represent patterns of perceptual preference. The map demonstrates a high degree of perceptual similarity of velar and uvular vowels and of front rounded vowel categories. These patterns of similarity also become evident in the confusion matrix in Table 10.7.

The two-dimensional MDS vowel map shows a cluster of rounded vowels in the upper right field of the MDS space that comprises front as well as back vowels. a-vowels are maximally contrastive from all other categories, the phonemic length distinction is however less clearly perceived by Albanian listeners. The two-dimensional MDS vowel map further shows that /iː/ is clearly differentiated from /ɛ/ and /ɛː/ but that /eː/ and /ɪ/ are perceived as more similar to each other and to /iː/, which results mainly from substitutions of /ɪ/ > /eː/ (5%).

Within this cluster of rounded vowels, /oː/ and /ʊ/ show the smallest perceptual distance, although the confusion matrix does not show a high percentage of substitutions for these two categories. /øː/ and /uː/ are perceived as close to /oː/ and /ʊ/. Pre-palatal /yː/ and /ʏ/ are differentiated well from /øː/ and /œ/. Within the class of front rounded vowels, /øː/ and /œ/ are closest to each other. /ɔ/ is distinguished well from all other rounded vowels. /ʏ/ appears to be perceptually more similar to front non-rounded vowels than all other labio-palatals. This effect in the vowel map is due to single substitutions of /ʏ/ with /eː/, /ɪ/ and /iː/, which are below the 0.5 level and therefore do not appear in the confusion matrix in Table 10.7.It may be hypothesized that these wrong categorizations are not only due to the phonetic similarity of these categories but are largely induced by orthographic factors, especially by the German graphemes <ü> and <ö> that do not exist in Albanian. In Albanian, /y/ is represented orthographically by <y>, while in German it is represented graphemically by <ü>. For German /øː/ and /œ/ there are no phonemic or graphemic equivalents in Albanian.

In the three-dimensional solution (RSQ .814), the cluster of rounded vowels is better differentiated than in the two-dimensional solution: ü-vowels are clearly differentiated from other rounded vowels, as are /uː/ and /ɔ/, a fact that can be explained by the existence of equivalent vowels in Albanian, whereas /ʊ/, /oː/, /øː/ and /œ/ are grouped more closely together.

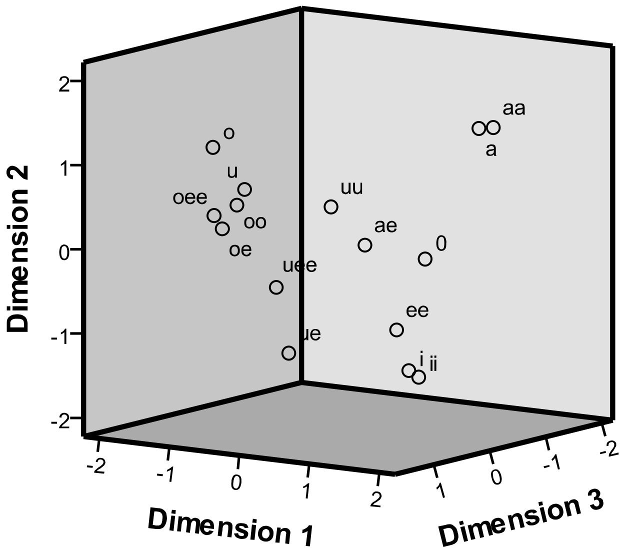

Figure 10.11: Three-dimensional MDS representation of the perceptual vowel space for L1 Albanian learners of German (N= 12, 3066 responses, RSQ .814)

10.3.5 Summary

To summarize, for Albanian learners of German, /ø:/, /ʏ/ and /ɛ:/ are most difficult in perception. These categories are most frequently substituted, least preferred as response categories and show most dispersion in perceptual substitutions.

Another interesting pattern is a higher preference for long constricted vowels. Long categories are more preferred as response categories and show more perceptual stability. /e:/, /a:/ and /y:/ are the most preferred response options, followed by /o:/, /ø:/ and /u:/, while /ʏ/ and /ɛ:/ are least preferred.

German rounded vowels are clearly distinguished from non-rounded vowels by Albanian listeners. The differentiation of front rounded vs. front unrounded vowels generally causes no difficulties. However, front rounded vowels are perceived to be more similar to back rounded vowels than to front vowels. In other words, no perceptual substitutions of front rounded by front unrounded vowels due to insufficient identification of the constriction location can be assumed.

Within the class of German front rounded vowels, /œ/ and /y:/ are differentiated best in perception. The perceptual differentiation of /ø:/ and /ʏ/, however, seems to be more difficult (a fact that becomes more evident in the confusion matrix than in the MDS vowel map). Some confusion is moreover observed with /œ/ <ö> and /ɛ:/ <ä:>, which is assumed to be graphically motivated rather than caused by perceptual difficulties. The reported articulatory substitution of /œ/ with unrounded /ɛ/ (Schader 2010) does not seem to be conditioned by “perceptual unrounding” of /œ/, since substitutions of /œ/ with front unrounded vowels are not observed in the data, with the only exception of /œ/ > /ɛ/ (4%), which is probably due to orthographic reasons.

Problems in differentiating vowel pairs with respect to phonemic length and constriction degree become evident in the data. For the short unconstricted qualities /ɪ/, /ʊ/, /ʏ/ and the contrast pairs /ʊ/ vs. /o:/ and /ʏ/ vs. /ø:/ we find difficulties with respect to the constriction location and phonemic length and constriction degree. For /ɪ/, /ʊ/, /ʏ/ we observe general problems in their contrastive differentiation especially with /e:/, /o:/ and /ø:/ respectively. Here, perceptual difficulties with phonemic length go together with difficulties in quality distinctions.

Patterns of *asymmetry* in perceptual substitutions are observed especially for short unconstricted qualities: (1) /ʏ/ > /œ/ (19%) vs. /œ/ > /ʏ/ (only 1%), (2) /ʊ/ > /ɔ/ (10%) vs. /ɔ/ > /ʊ/ (0%), and (3) /ɪ/ > /ɛ/ (9%) vs. /ɛ/ > /ɪ/ (< .5%). Interestingly, in all these cases the “lower” unconstricted category functions as perceptual target. In other words, unconstricted vowels are perceptually associated to qualities with a higher degree of aperture and are perceived to be “more back” than the intended input category, i.e. pre-palatal > palatal (/ʏ/ > /œ/ and /ɪ/ > /ɛ/) and velar > uvular (/ʊ/ > /ɔ/).

For long constricted /ø:/, however, an asymmetric substitution pattern in the opposite direction is observed: 17% /ø:/ > /y:/ vs. 4% /y:/ > /ø:/. Here, /y:/ which has an L1 equivalent (Albanian /y/) attracts responses for /ø:/ stimuli but not for /œ/. In other words, the pre-palatal category /y/ functions as perceptual magnet for German /ø:/.

To summarize, the perceptual vowel space of Albanian learners of German shows a clear differentiation of rounded vs. non-rounded vowels. A high degree of perceptual similarity can be observed for all rounded vowels. Front non-rounded vowels are better differentiated in L2 perception as is reflected in bigger distances in the two- and three-dimensional MDS solution. Category confusion due to insufficient identification of constriction location is mainly observed for unconstricted high vowels (pre-palatal > palatal and velar > uvular).

Most difficulties are observed with the differentiation within the class of front rounded vowels and with the differentiation of front rounded vs. back rounded vowel categories. Long constricted vowels are clearly preferred as response categories and function as perceptual targets for short unconstricted qualities, whereas short unconstricted qualities, especially high and more front qualities, are less preferred and more prone to perceptual substitutions.

10.4 Arabic

10.4.1 General characteristics (Classical Arabic, Modern Standard Arabic and regional varieties)

Arabic is spoken by more than 200 million people as a first language mostly living in 20 countries of the Middle East and North Africa and by speakers in Iran, Turkey, Chad, some regions in the South-West of the Sahara, Central Asia, francophone West Africa and by speakers in Arab communities in Europe and America. It is thus the sixth most widely spoken language in the world. It is moreover used as a liturgical language by about one billion Muslims around the world. Arabic belongs to the group of Semitic languages. Traditionally, Arabic has been claimed to belong to the South-Semitic or South-West-Semitic branch, while more recent studies propose that Arabic belongs to the North-West-Semitic branch within Central Semitic (Faber 1997: 8f, cited in Watson 2007: 6).

Arabic is one of the prime examples for *diglossia* (Ferguson 1959, 1991). Classical Arabic (CA) refers to the medieval variant of the language as codified in the Qur'an. The use of Classical Arabic is nowadays confined to some rare contexts and situations like reciting the Qur'an or classical poetry. Modern Standard Arabic (MSA), a modernized version of Classical Arabic. It characterizes the literary language which is used in all Arab countries today in science, teaching, literature, theatre, press, radio, television and cinema. However, it is not the language of the people in everyday life. All Arabs, even the most educated, are dialect speakers, and the use of MSA is confined to certain situations.

Literary Arabic developed with the establishment of grammatical norms from the eighth century on. The primary sources for standardization were the Qur'an and pre-islamic poetry (Freeman 1996: 1). With the dissemination of the Qur'an, the language gained enormously in importance and spread not only in Arabia but all over the lands of Islam. Classical Arabic (CA) has remained unchanged over the centuries, but the vernacular Arabic dialects have developed markedly during this period and show a considerable range of linguistic variation.

While morphology and syntax have basically remained unchanged for many centuries, lexis and stylistics of Modern Standard Arabic are clearly different from Classical Arabic as well as from the huge variety of vernacular forms spoken in all parts of the Arab speaking world. Due to its written form and to its use in the media, MSA is mostly intelligible for literate persons across the Arab world from Morocco to Iraq.

Contemporary spoken dialects, which remain mostly unwritten, are used in daily life and represent the regional and social characteristics of speakers. They differ significantly from each other in phonology, morphology, syntax and lexicon, rendering mutual intelligibility of dialects from the most Eastern and most Western regions of the Arab-speaking world almost impossible.

Four major groups of regional dialects may be distinguished: (a) dialects from the Maghreb region (Morocco, Algeria, Tunesia, and western Libya, (b) Egyptian dialects including Egypt, eastern Libya and the Sudan, (c) the Levante dialects (Syria, Lebanon, Jordan, and Palestine), and (d) the Arabic from the Arabian Peninsula and the Persian Gulf (Saudi Arabia, Iraq, Yemen, Oman, Qatar, Bahrain, the Arabic Emirates and Kuwait) (Freeman 1996: 2).

An additional differentiation that is based on another dimension refers to the distinction of urban, rural and Bedouin varieties, which has consisted throughout history (Behnstedt 2005: 42ff) and is still of relevance. Even though belonging to different dialect groups, the speech of a Cairene and a Damascene citizen may be closer to each other than that of an Egyptian Bedouin and a Cairene speaker (Freeman 1996: 2).

10.4.2 Phonological and phonetic description

In this section, a description of the phonological system of Modern Standard Arabic will be presented followed by a summary of results and comparisons of studies on regional contemporary dialects and stylistic varieties such as the Qur'an recitation style.

Modern Standard Arabic (MSA) has an inventory of 36 phonemes with 28 (29) consonants (including the ا *alif* according to traditional grammatical descriptions), six vowels and two diphthongs. While the number of consonants is clearly above the mean of 22.8 in UPSID (Maddieson 1984), the number of vowel phonemes is quite below the UPSID average of 8.7. Syllable structures in Classical Arabic are limited to CV, CVC and CVCC, where the vowel may be long or short. However, Arabic varieties display a rich inventory of syllable types. North African dialects differ from most other varieties in allowing complex syllable structures where consonants may appear as syllable nuclei (e.g. Shaw, Gafos, Hoole & Zeroual 2009).

Consonants may occur in single or geminate, i.e. lengthened forms. The pronunciation of several consonant phonemes differs in dialectal variants, e.g. [dʒ] ~ [ʒ] ~ [g]; /q/ > [ʔ], [g]; /d/ > [z]; /θ/ > [s], [t]; /ð/ > [z], [d], [x ɣ] ~ [χ ʁ] (Newman & Verhoeven 2002: 78).

The nature of the emphatic consonants and their articulatory realization is a matter of intense debate in the Arabic linguistic literature. Generally, "emphatic" consonants have a plain counterpart. The so-called "emphatic" sounds, a series of pharyngealized consonants /tˤ dˤ ðˤ sˤ lˤ/, are realized with Retracted Tongue Root, involving simultaneous pharyngealization and a certain degree of velarization (Thelwall & Akram Sa'adeddin 1990: 52f), exerting a strong coarticulatory effect on sounds in preceding and following syllables. Back consonants /x ɣ q ħ ʕ/ have a similar effect on neighbouring vowels. Two approximants are found in Arabic: the voiced palatal approximant [j] and the voiced labial-velar [w][76].

The number of emphatic consonants varies from dialect to dialect (Watson 2007: 10f). The targets, triggers, direction, and domain of pharyngealization differ across varieties. Vowels in the vicinity of emphatic consonants are lower, retracted or more "centralized" than in a non-emphatic context (Al-Ani 1970: 44ff; Kaye 2011: 564).

	bi-labial	labio-dental	dental	alveo-lar	postal-alveolar	palatal	velar	uvu-lar	pharyngal	glottal
Plosives	b		t d tˤ dˤ				k (g)	q		ʔ
Nasals	m		n							
Fricative		f	θ ð ð̠ˤ (~zˤ)	s z sˤ	ʃ	ʒ~dʒ~g [77]	x ɣ ~ χ ʁ[78]		ħ ʕ	h
Lateral				l (lˤ)						
Trill			r							
Semi-vowels						j	w			

Table 10.12: Consonantal phoneme inventory for Standard Arabic

The vowel inventory of Classical Arabic is traditionally described as consisting of a set of three distinct vowel qualities (only 5.4% of languages in the UPSID material have only three vowels), which is in clear contrast with the very rich vowel system of German. The Arabic system includes the three most common vowels /i/, /u/ and /a/. Duration differences are phonemically functional. The system therefore consists of three short vowels /i/, /u/, /a/ and three long vowels /i:/, /u:/, /a:/. Vowels generally do not occur word-initially but are preceded by the glottal stop أ *alif*. Vowel length therefore does not occur in word-initial position, it is more common in medial than in final word position (Al-Ani 1970: 75).

---

76 Watson (2007) and Thelwall & Akram Sa'adeddin (1990) denote the semi-vowel /w/ as velar. Rastegar-El Zarka (1997: 29) describes /w/ as a labial glide with velarization for Egyptian Arabic.

77 The realization of ج differs regionally (cf. Newman 2002: 69), e.g. [dʒ] on the Arabian Peninsula and the Persian Gulf, [g] in Egypt, [ʒ] in large parts of North Africa and the Levant.

78 The classification of خ and غ concerning place of articulation differs in phonological descriptions and is also subject to regional variation, e.g. uvular [χ ʁ] in Watson (2007:13) or velar [x ɣ] in Thelwall & Akram Sa'adeddin (1990) or velar /x/ and uvular /ɣ/ in Al-Ani (1970).

Some descriptions consider the diphthongs /ay/ and /aw/ as vowel phonemes in MSA. All contemporary *dialects* of Arabic have at least three long vowels /a:/, /i:/, /u:/, but differ with respect to the phonetic realization of diphthongs. While the diphthongs coalesced historically in Cairene (Egypt), Central Sudanese and many dialects in the Levant and are realized as /e:/ and /o:/ (or /i:/ and /u:/ in Moroccan dialects (see Kaye 2011: 565), other dialects such as San'ani or dialects from the Arabic Peninsula have preserved the diphthongs in their original form in all contexts (Watson 2007: 22f).

As this coalescence or monophthongization in Cairene and several other dialects is considered to be a historical process which is not productive any more, a *five-member long vowel system* can be posited for these dialects: /i: e: a: o: u:/. In this group of dialects, the system consists thus of a larger number of long vowels than short vowels.

Despite the restricted number of vowel phonemes, considerable allophonic variation can be observed resulting in a number of different phonetic qualities. The phonetic quality of long as well as short vowels displays considerable context-sensitive allophonic variation. Moreover, the vowels and their allophones differ regionally in the various dialects of contemporary Arabic. Especially in the neighbourhood of emphatic/pharyngealized consonants, vowels display phonetic differences compared to their plain counterparts. These differences manifest themselves among others in an increase in F1 and a lowering in F2 compared to the vowels in the context of plain consonants (Al-Ani 1970: 49; Newman & Verhoeven 2002: 80).

Arabic short vowels are commonly described as more "central" than their long counterparts, though the term "less constricted" may be preferred. Strong differences in constriction location varying between the front and the back region of the oral cave are observed for short /a/ in its many allophonic variants [a ɑ æ ɐ ɛ e ə]. For short /u/ the constriction location appears to be different as well (see also Figure 4.4 for constriction locations in Egyptian Arabic as documented by Wood 1979, 1990). In very broad acoustic terms, this corresponds to a higher F1 and a lower F2 for front vowels and a higher F1 and F2 for non-low back vowels. This tendency becomes apparent especially with high vowels and in connected speech (Newman & Verhoeven 2002). The reported duration is relatively high in most studies, probably due to elicitation in isolation or in single syllables, e.g. 100-150 ms for short vowels and 225-350 ms for long vowels (Al-Ani 1970: 75).

Acoustic characteristics of long and short vowels in *Qur'an recitation style* are presented in Figure 10.12 (Newman & Verhoeven 2002: 81). In recitation style, a clear distinction in terms of length is realized, while the acoustic difference between short and long vowels is not statistically relevant. However, it is important to note, that in local varieties and especially in connected speech, the formant values differ considerably from these ideal forms. In a comparison of vowels realized by Egyptians, Sudanese and Saudi speakers, Alghamdi (1998) shows that there are dialect-specific quality differences in the distribution of short and long vowels in the vowel space. In all varieties studied by Alghamdi (1998)[79], short vowels differed acoustically significantly from long vowels and are described as occupying the more "central" regions of the vowel space compared to long vowels. The major differences between the dialects were found in the frequency of F1. The duration of vowels was very similar in all these dialects.

For long vowels, the opposition between /i:/ and /u:/ exists in all dialects. In Modern Standard Arabic, /i:/ and /u:/ are phonetically quite similar to their short counterparts. /a:/ differs clearly from short /a/ in phonetic quality in that short /a/ is articulated more front than /a:/ (for a comparison of acoustical and physiological results from X-ray, see Al-Ani 1970: 25ff).

For long /a:/ in isolation, the tongue position seems lower and somewhat more retracted than with /a/. With high vowels, especially with /u/ and /u:/, the articulatory differences are smaller.

Even if /a/ is commonly regarded as the most stable and conservative of the three short vowels, it is nevertheless changed or even deleted in many contemporary dialects: /a/ can be realized as [a ɑ æ ɐ ɛ e ə] depending on consonantal context and position in a word. /a/ is realized as [æ] in the context of plain (non-pharyngealized) coronal consonants /θ/, /ð/, /n/, /t/, /d/, /s/, /z/, /l/, /ʃ/ and /dʒ~g~ʒ/, and in the context of labials /m/, /b/, /f/, glottals and pharyngeals (/h, ʔ/ and /ħ, ʕ/, and /k/, /j/, /w/. /a/ in unstressed syllables may be deleted, e.g. */ya+maħammad/* => /yamħammad/ 'oh Mohammad!' (Kaye 2011: 565). In Classical Arabic, several lexical doublets and even triplets occur for words with short /a/, e.g. */lasˤsˤ/ ~ /lisˤsˤ/ ~ /lusˤsˤ/* 'thief' (Kaye 2011: 565).

In general, short vowels seem to be more susceptible to change than long vowels in contemporary Arabic dialects. In some dialects, short /i/ and /u/ have collapsed to schwa, e.g. in North Mesopotamia, Bedouin dialects of Maghreb, and some Mauretanian dialects. In these dialects, the short-vowel system consists of only two vowels [a] and semi-closed [ə] (Watson 2007: 21). In several other systems, the opposition between /i/ and /u/ is reduced in certain contexts, e.g. in a few Syrian dialects and, to a lesser extent, in some Sudanese dialects and Cairene (Watson 2007: 21f). For instance, in Damascus Arabic, Classical /i/ and /u/ both merge into [ə].

---

79 Alghamdi (1998) compared speakers from Egypt, Saudi Arabia and Sudan.

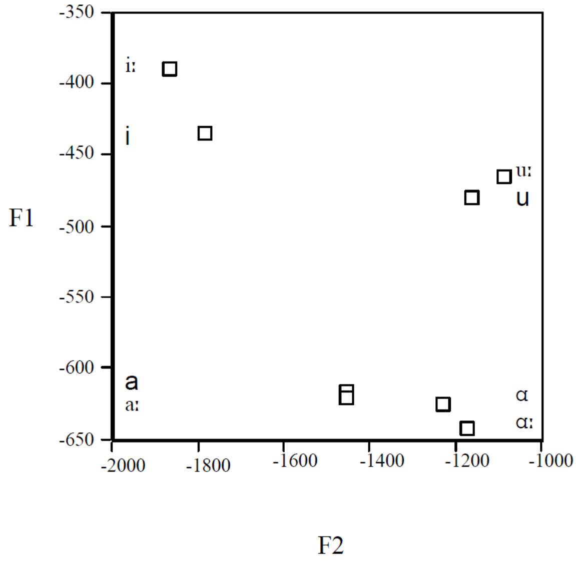

Figure 10.12: Average formant values (F1xF2 in Hz) for 400 short and long vowels in Qur'an recitation style (figure reproduced from Newman & Verhoeven 2002: 81)

Kaye (2011) mentions two phonological processes for qualitative changes of long vowels: (1) "/a:/-raising" by which /a:/ as in */ʕiba:d/* 'slaves' is changed to */ʕibe:d/* or */ʕibi:d/*, usually due to the umlauting influence of /i/, e.g. in Syro-Lebanese dialects */ba:b/* > [be:b] or [bi:b] 'door', or phonetic qualities in between or adjacent to those phones (Kaye 2011: 566), and (2) a process of delabialization explaining a change from /u:/ > /i:/ via an intermediate state of /y/ as in */ru:m/* > */ri:m/* 'Rome' (cf. also the rhyming option of /u:/ to /i:/ in Qur'an Arabic).

Newman & Verhoeven (2002) present a comparison of acoustic-phonetic results from various variants. Referring to Modern Standard Arabic, the values presented in Figure 10.13 are realized by speakers from different dialectal regions: Al-Ani (1970) presented a description of data by himself and other Iraqi speakers (plus two Jordanian speakers), Belkaid (1984) presented Tunesian data. Ghazeli included informants from six countries (Algeria, Tunisia, Libya, Egypt, Jordan, and Iraq). Abou Haydar (1994) compared data from informants from eight countries (Jordan, Lebanon, Qatar, Sudan, Syria, Tunisia, United Arab Emirates) and Alghamdi (1998) compared speakers from Egypt, Saudi Arabia and Sudan. It is important to not that data are elicited in different types of task but mostly vowels are realized in isolation or in single syllables.

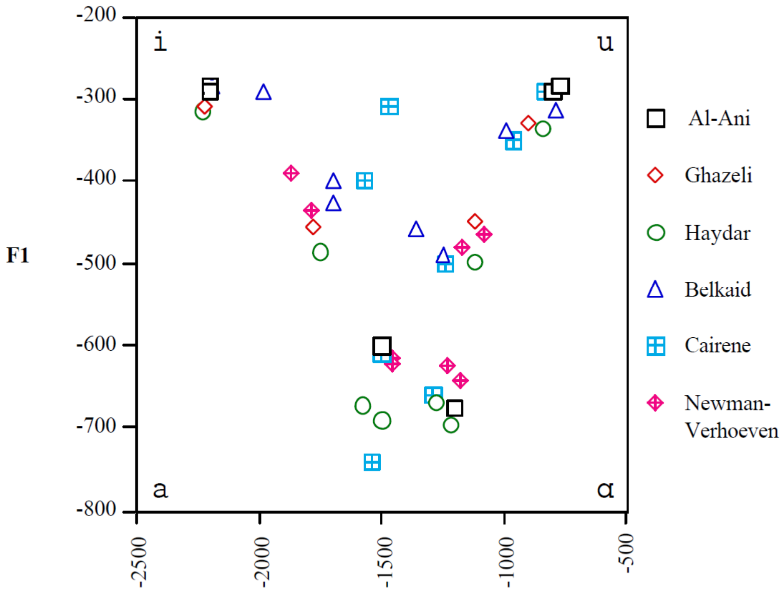

Figure 10.13: Scatter plot of formant values (F1xF2 in Hz) for six Standard Arabic vowels in six different studies (Al-Ani 1970, Ghazeli 1979, Abou Haydar 1991, Belkaid 1984, Newman & Verhoeven 2002 (Qur'an vs. Cairene)) from a meta-analysis by Newman & Verhoeven (2002) (figure reproduced from Newman & Verhoeven 2002: 83)

Newman & Verhoeven (2002) present data from connected read speech by a speaker of the Egyptian variety of Arabic (Cairene) (Figure 10.14) demonstrating clear differences from the Qur'an recitation style as in Figure 10.12.

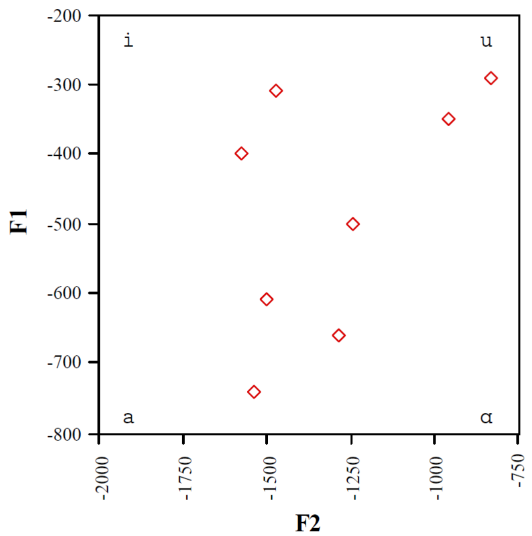

Figure 10.14: Cairene vowels in connected speech by F1xF2 in Hz (figure reproduced from Newman & Verhoeven 2002)

For Iraqi Arabic, Al-Ani (1970: 23f; see also Kaye 2011: 565) describes allophonic variation for the three basic vowel qualities as follows: The most common allophone of /i/ and /iː/ is a high front short or long respectively unrounded vowel. A centralized and slightly lower [ɨ] occurs in the neighbourhood of pharyngealized consonants except /lˤ/. [ɪ] is realized next to /ʕ/ and /ɣ/, and [i] elsewhere.

The most common allophone of /u/ and /u:/ is a – short or long respectively – high back rounded vowel. It is realized as, short or long respectively, [ʊ] in the neighbourhood of pharyngealized consonants except /lˤ/, and as [u] elsewhere.

Al-Ani (1970: 24) describes /a/ as retracted to [ɑ] in the vicinity of pharyngealized consonants or /r/, /q/. In final positions and not adjacent to pharyngealized consonants or /r/, /q/, /ʕ/, or /ɣ/, he reports pronunciation as [ə] (typical for Iraq and Saudi Arabia) and as [ʌ] next to /ʕ/ and /ɣ/. Realization as [a] is reported elsewhere. For long /a:/, he reports the same distribution except the realization as [ə].

Though it is generally claimed that high front and back vowels /i u/ and /i: u:/ differ mainly in quantity, whereas /a a:/ are described to vary in quantity and quality (Al-Ani 1970: 75), this may hold for vowels in isolation of single syllables as in the studies mentioned above, but not for connected speech, as in the Cairene data (Figure 10.14) collected by Newman & Verhoeven (2002), where both /i/ and /i:/ are described as considerably lower and more "central" than in other studies. Especially the spread of F2 values described for /i/ and /i:/ in studies with speakers from different countries is remarkably high (see Figure 10.14). Note however, that the realization of the diphthongs */ay/ and */aw/ as /e:/ and /o:/ in Cairene are not considered in their study.

Abou Haydar (1994) compared data from informants from eight different countries (Jordan, Lebanon, Qatar, Sudan, Syria, Tunisia, United Arab Emirates). The data show a considerable spread of measured values[80] across regional varieties of Arabic as represented in Table 10.13.

	Regional variation Abou Haydar (1994)		Cairene (Newman & Verhoeven 2002)	
	F1	F2	F1	F2
i:	280 - 335	1990 - 2530	290	1940
i	415 - 565	1400 - 2135	375	1575
u:	260 - 380	620 - 990	290	830
u	430 - 580	1005 - 1240	360	912
a:	610 - 770	1280 - 1780	610	1500
a	620 - 780	1390 - 1680	683	1435

Table 10.13: Formant frequency values (F1, F2 in Hz) in eight regional varieties of Arabic (Jordan, Lebanon, Qatar, Sudan, Syria, Tunisia, United Arab Emirates) measured by Abou Haydar (1994), cited in Newman & Verhoeven (2002: 87) and measurements for Cairene Arabic by Newman & Verhoeven (2002).

[80] No information is available in Newman & Verhoeven (2002) that indicate whether the values are normalized.

For palatal vowels, Wood (1979) posits a pre-palatal constriction in articulation for Egyptian Arabic speakers which is located at about 27 mm behind the central incisors. The position of the tongue relative to the mandible stressed vowels [i ɪ e ɛ] of the Egyptian Arabic subject described in Wood (1979: 32) is more fronted than for Southern British English (about 35-40 mm distance from the central incisors for American and British subjects, Wood 1979: 34). However, the comparison in Table 10.13 and Figure 10.14 shows that there are considerable differences between regional varieties with respect to F2 values for /i/ (or [ɪ]) and /i:/ suggesting that the constriction location could also vary regionally between the pre-palatal and mid-palatal area of the vocal tract. However, it has to be kept in mind that the phonetic quality of Arabic vowels is strongly influenced by the phonetic context which might also have an influence on the constriction location.

To summarize, Arabic shows rich vowel allophony in terms of a high degree of regional variation in colloquial dialects, stylistic variation and considerable context-sensitive variation especially but not only in the context of pharyngealized consonants.

10.4.3 Contrastive analysis

The limited set of three short and three to five long vowel phonemes in Arabic together with a high range of allophonic regional, stylistic and contextual variation can account for a number of difficulties generally observed with L1 Arab learners of German. Due to this range of allophonic variation and category overlap, the perceptual vowel space of Arab native speakers can be considered to be divided in less but wider vowel areas causing several difficulties with languages that have a larger vowel system and a more limited range of variation. Disambiguation of variation is performed by local contextual information (e.g. of consonantal context such as plain vs. emphatic consonants) and by top-down information processes. The considerable range of allophonic variation of several Arabic vowel phonemes, especially of /a/ and /u/, is of special relevance in L2 acquisition of German in perception and production, causing confusion between German vowel categories and their phonemic, morphological and lexical function (for a more detailed discussion, see Kerschhofer-Puhalo 2010b, 2012, in press a and b)).

An interesting source for external evidence of difficulties with the perception and production of German vowel phonemes are written texts where the interaction of incorrect perception, difficulties in pronunciation, morphological structures (derived forms, e.g. plurals) and lexical representations become particularly well evident. The examples (1) to (4) are single words or texts written by Arab learners of German (in italics) from different Arab countries (these learners did not participate in the study).

The examples (1) to (4) show considerable difficulties with respect to phonemic differences that are based on the *constriction location*. This concerns especially the distinction of pre-palatal vs. palatal and velar vs. uvular vowels.

(1) L1 Lybian Arabic

/kɔpf/	→ *Kepf*	‘head’
/hals/	→ *Hels*	‘neck’
/hand/	→ *Hend*	‘hand’
/fɪŋgɐ/	→ *Fankr*	‘finger’
/bain/	→ *Been*	‘leg’
/fu:s/	→ *Faß ~ Foes*	‘foot’
/tse:hɛ/	→ *Zaha*	‘toe’
/brʊst/	→ *Brast ~ Brüst*	‘breast’
/bɑɔχ/	→ *Baich*	‘belly’
/rʏkɛn/	→ *Rakan ~ Rüken*	‘back’

(2) L1 Egyptian Arabic

Ich hape zeiwe kendir. Ich bein verherTiTte. Ich bein 30 yare alte>

(3) L1 Iraqi Arabic[81]

In der Früh *kaffee trinken*
Am Morgen *ich gege in den Schule kors*
Am Vormittag *ich arpaeten in zu Hause .pots*
Zu Mittag *ich kochen mittag essen*
Am Nachmittag *ich gehe kauffen in Billa*
Am Abend *ich brengen meine Tochter in cindrkarten*
In der Nacht *gehe ins Bett schlafein.*

(4) L1 Egyptian Arabic

Ich gehe spitseren	Ich gehe spazieren
meine Fase ist kareink	Mein Fuß ist krank
Desempir	Dezember
toua	teuer
Sie kafet	Sie kauft
gorose	groß
Dutsch	Deutsch
arbiuten	arbeiten
Kindr	Kinder
Arbsch	Arabisch
Ich slefin	Ich schlafe
Sie lubet Möuick	Sie liebt Musik
1 Keind	1 Kind

[81] Words in italics are written by the learner, the beginning of the sentence was printed on the work sheet.

However, as in (1) and (4) there are instances of category confusion that are not that readily evident by a mere analysis in terms of constriction location. Typically, learners who show this kind of difficulties in written performance acquired German not (or not only) in a formal setting. Their written production and their individual way of pronouncing German words is directly reflecting areas of difficulties in L2 pronunciation. Reasons for these difficulties are manifold:

One of the most prominent coarticulatory effects in Arabic is the fronting of /a/ to [æ] in the context of plain (non-pharyngealized) coronal consonants /θ/, /ð/, /n/, /t/, /d/, /s/, /z/, /l/, /ʃ/ and /dʒ~g~ʒ/, labials /m/, /b/, /f/, glottals and pharyngeals (/h, ʔ/ and /ħ, ʕ/), and /k/, /j/, /w/. The high range of allophonic variation for /a/ and its allophones has severe effects on the perception of German a- and e-vowels, especially for short unconstricted qualities and in non-stressed syllables, as in example (1) *<Hels> - Hals* ‘neck’, *<Hend> - Hand* ‘hand’, <Fankr> - Finger ‘finger’ and in example (4): *<kareink> - krank* ‘ill’, *<slefin> - schlafen* ‘sleep’.

Another effect frequently observed with Arab learners of German is the confusion of o- and u-qualities, e.g. *<pots> - putzen* ‘clean’ or *<kors> - Kurs* ‘course’ that can be explained by two components: (1) the high allophonic range of /u/ in Arabic together with the possible interaction of [o:] from monophthongized /aw/ in many regional varieties and (2) the acoustic similarity of the four German u- and o-vowels as discussed in section 5.4.1 and 5.4.2.

Front rounded vowels do not occur in the Arabic system and are therefore expected to cause difficulties in perception. In L2 pronunciation, ü- and ö-vowels are frequently commonly substituted by back rounded vowels. Therefore, substitutions with back vowels may also be expected in L2 German perception.

Phonemic length distinctions do occur in Arabic and are – similarly to German – associated with qualitative differences that can be described in terms of differences in degree of constriction [+/–constricted] or especially for back vowels also by differences in constriction location (e.g. velar vs. uvular for u-vowels and front vs. back a-vowels). It is expected that the recognition of phonemic length as such will not be a difficult feature in the acquisition of L2 German, but that the qualitative distinction of long constricted vs. short unconstricted vowels may cause difficulties for Arab learners of German.

The graphemic representation of German vowels and the very limited number of orthographic equivalents in Arabic is another substantial factor for difficulties. The restricted number of vowel graphemes and their use in the Arabic writing system is expected to cause additional problems with the correct choice of graphemes for German vowels. In the Arabic writing system, the long vowels /i:/, /u:/ and /a:/ are represented by the letters أ و ي. Short vowels can be represented by vowel diacritics that are added to the consonant letters. Diacritics are used especially in the Qur'an and other religious texts or poetry but are usually not used in texts of every-day life. Another interfering factor is the strongly limited occurrence of consonant clusters in Arabic especially in onset position that are frequently avoided by vowel insertion. Example (4) shows instances of epenthesis to avoid C-clusters (e.g. <*kafet*>, <*gorose*>) as well as instances where (short) vowels are not realized, indicating the learner's use of a hypercorrect strategy. The deletion of vowels in written German (e.g. *Arabisch* as <*Arbsch*>) can be explained by the non-obligatory use of diacritics for short vowels in the Arabic script. Occurring problems with the orthographic representation of vowel contrasts in German vowels are therefore not only explained by difficulties with German orthographic conventions but also by the insufficient familiarity or limited fluency with the Latin based alphabet in general. For learners with proficiency in writing English even more difficulties may arise due to experiences with English orthographic conventions that largely differ from German representations.

To summarize, the major problem for a contrastive analysis of German and Arabic to predict or explain problems in German pronunciation and perception is the high regional and context-sensitive variability of vowels across Arabic dialects. The data presented in the current study is based on the test results of learners of different regional origin. Therefore, the data can function only as a possibility for a first insight into difficulties of Arab learners with German vowel categories, describing the major patterns of perceptual substitutions. It is of course not posited here that the patterns observed will be equal for speakers of all modern dialects of Arabic. A complete analysis would ideally link a thorough context-sensitive phonetic, i.e. acoustic and articulatory analysis of the learners' native vowel phonemes and their allophones in one given dialect with the perceptual data observed in L2 but this is beyond the scope of the present study.

To conclude, difficulties of Arab learners in perceiving and producing German vowel categories correctly are conditioned by the three major factors: (1) the high number of vowels in the German vowel system, (2) the high intrinsic variability of different allophones of the same vowel phoneme in Arabic and the "acoustic overlap" of several German vowel phonemes with Arabic allophones (see Figure 10.13 and Table 10.13), and (3) insufficient familiarity and difficulties with the Latin alphabet and the graphemic representation of German vowels to explain why the acquisition of the German vowel system causes so many difficulties for Arab learners, even for experienced learners.

In studies of regional varieties providing acoustic data as presented above, the influence of consonantal context that has considerable influence on the articulatory and acoustic characteristics of individual vowel realizations has so far not been sufficiently regarded. Further research on the acquisition of German vowel perception and production would require a detailed *context-sensitive analysis* of vowel realization of vowels for speakers of specific *regional varieties*; however, this type of acoustic data is currently not available. The following sections will therefore present an analysis of experimental data for the full sample of Arab native speaker irrespective of their regional origin. As the current study is intended as a large-scale survey study, the current data can provide an overview of the wide range of inter-category confusion of German vowels that is of particular interest for teachers of German as well as for further studies on vowel identification and the discrimination of particular vowel contrasts considering not only regional variation but also effects of speech style and task-type (read vs. spontaneous speech) and vowel-specific effects of consonantal context (place of articulation and plain vs. emphatic consonants).

10.4.4 Results and discussion

Ten native Arabic subjects were tested, seven beginners, two advanced and one very advanced learner. The sample consists of six speakers of Egyptian Arabic (three beginners, three advanced), two Iraqi, one participant from Tunisia (beginner) and one from Sudan (beginner).

In general, the test setting seemed particularly demanding for some Arab participants probably due to problems with the unfamiliar writing system and insufficient experience with spoken German for learners who studied German abroad. Data of participants who showed an extensive number of wrong identifications are therefore not included in the present analysis.

The data set discussed here consists of 2634 valid responses (2700 possible responses, 66 missing answers). For each vowel category, 168 to 180 valid responses were delivered.

10.4.4.1 Difficulties

The diagram in Figure 10.16 shows the relative difficulty of German vowel categories for L1 Arabic learners: /ɔ/, /ɑ/ and /ɛ/ are the categories that are most correctly identified by Arabic listeners, though /ɔ/ (86% correct) and /ɛ/ (79%) do not occur in the Arabic vowel system. The most difficult categories for Arab learners are /e:/, /ø:/, /ɛ:/, /o:/ and /œ/. The highest percentage of incorrect answers is observed for /e:/ (44%), whereas in all other language subgroups tested in this study /ɛ:/ shows higher id_wrong scores than /e:/. This high percentage of wrong identifications is due to insufficient differentiation of /e:/ vs. /ɛ:/ *and* /e:/ vs. /i:/ (cf. the confusion matrix in Table 10.14).

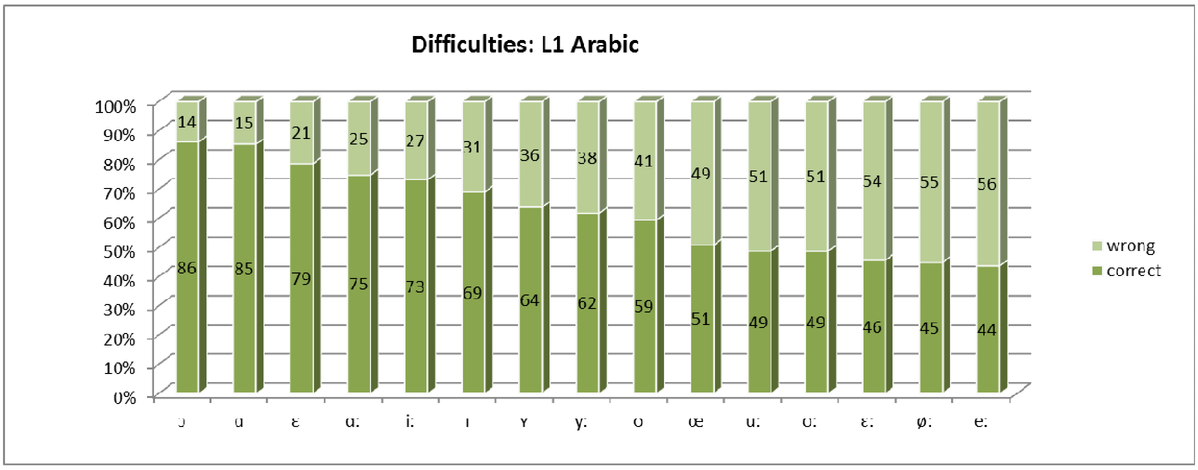

Figure 10.16: Percentage of wrong and correct identifications by Arabic listeners

The listeners' mean id_wrong score was 38% (20% standard deviation), the median was 46%. The listener with the highest percentage of correct answers had an id_wrong of 9%, the highest id_wrong score was 63% wrong responses.

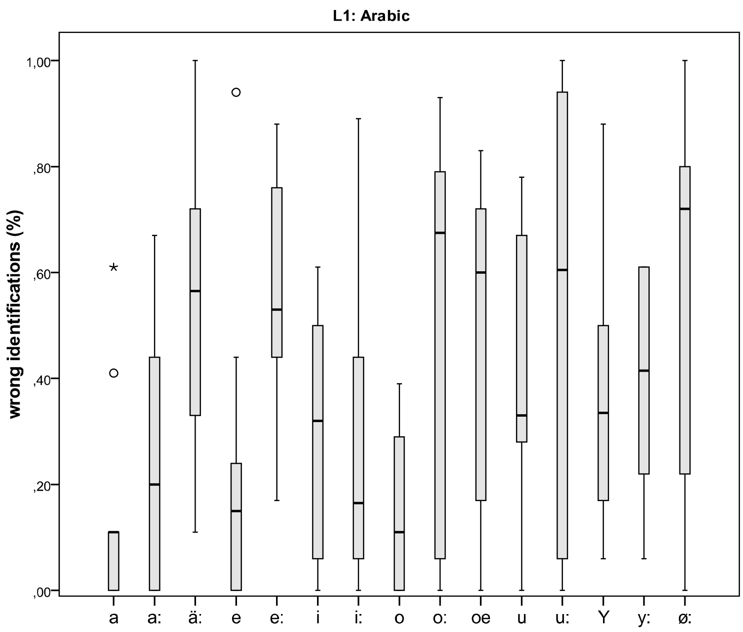

Figure 10.17: Boxplot diagram with id_wrong scores for L1 Arabic listeners

The boxplot diagram in Figure 10.17 shows a remarkably high percentage of correct answers especially for /ɑ/ for which moreover a very low inter-speaker variability of id_scores is observed. It can be assumed that the /ɑ/-stimuli presented in the test are recognized as very good exemplars of /ɑ/. A low degree of variation together with a high average percentage of correct answers is moreover observed with /ɔ/ and /ɛ/. For all other categories we find considerable quantitative and qualitative inter-individual variation. The widest range of inter-individual variation is observed for /u:/, /o:/ and /ø:/.

10.4.4.2 Preferences

The preference ranking for L1 Arabic listeners (see Figure 10.18) shows interesting preference patterns. The preference scores for monophthong categories vary within a range of 6.7% (min /o:/ 3.9%, max /ʏ/ 10.6%). Surprisingly, the most frequently selected category is /ʏ/ (10.6%), followed by /ɛ/ (8.8%) and /ɔ/ (8.1%). Phonemically long categories are considerably less frequently chosen as response option than short categories. Interestingly, this pattern holds consistently for all long vs. short categories with the only exception of /i:/ (6.5%) vs. /ɪ/ (6.6%). The least frequently chosen categories are /o:/ (3.9%), /u:/ (4.5%) and /ø:/ (4.6%), which also show the widest range of inter-individual variation.

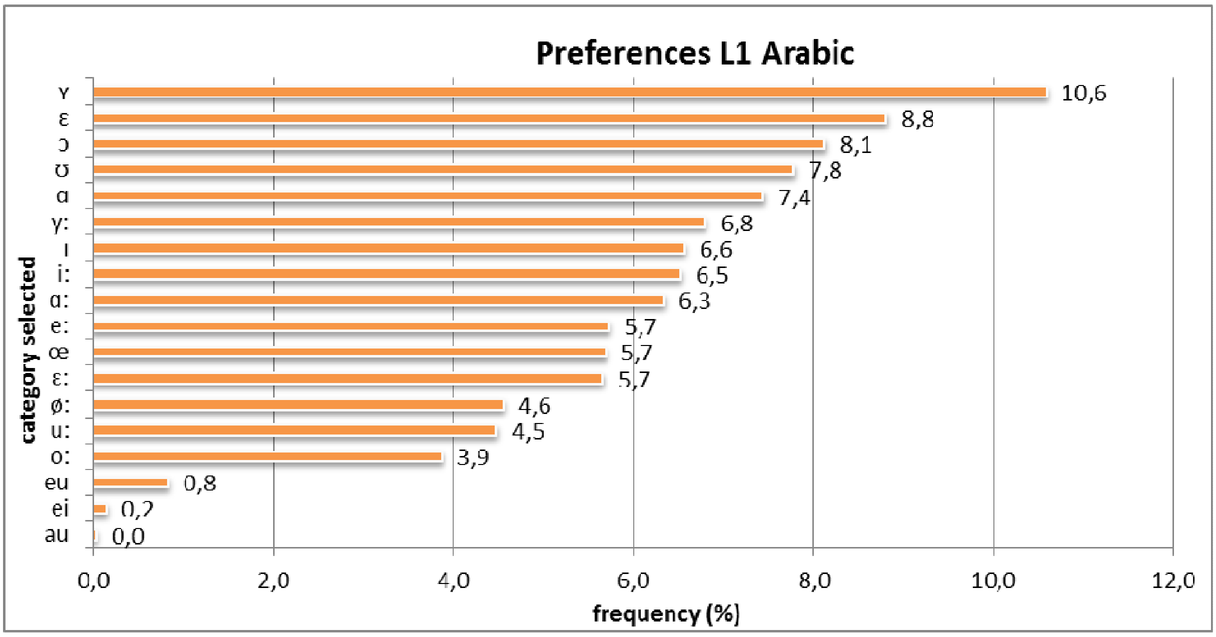

Figure 10.18: Preferences for L1 Arabic listeners

Diphthongs are only selected by beginners but not by advanced learners. Within the class of diphthongs, <eu> is the category that is most often selected by the participants (0.8%), while <ei> (0.2%) and <au> (one case) are selected only occasionally. This pattern can be explained by the fact, that Arab learners of German seem to have a clear representation for the diphthongs /ai/ and /au/ that occur in Standard Arabic and in many dialects and can therefore be transferred to German, whereas the perceptual representation of /ɔi/ seems to be more vague for Arab learners. It is worth to note moreover that advanced learners do not select diphthongs as response options.

10.4.4.3 Patterns of confusion

A first look at the confusion matrix for L1 Arabic listeners in Table 10.14 reveals considerable confusion among German vowel categories. Two major areas of confusion have to be considered in more details: (1) confusion among rounded vowels, and (2) confusion among front non-rounded vowels.

%	ɑ	ɑ:	ɛ:	ɛ	e:	ɪ	i:	ɔ	o:	ʊ	u:	œ	ø:	ʏ	y:	‹ei›	‹eu›	‹au›	wrong	corr
a	**85**	12	2							1									15	**85**
a:	23	**75**	2																25	**75**
ɛ:	1	5	**46**	17	27	1	3									1			54	**46**
ɛ		1	7	**79**	5	5	3												21	**79**
e:	1	3	24	15	**44**	4	10												56	**44**
ɪ			2	17	3	**69**	7							1			1		31	**69**
i:					6	19	**73**							1	1				27	**73**
ɔ								**86**	5	3	1	4					1		14	**86**
o:								16	**49**	3	6	5	10	5	4		2		51	**49**
ʊ							1	12	3	**59**	1	7	1	16	1		1		41	**59**
u:								1	3	18	**49**	1	5	12	10		1		51	**49**
œ			2	5	1		1	6	1	8	1	**51**	6	15	2	1	2	1	49	**51**
ø:			1							2	6	9	**45**	19	17		2		55	**45**
ʏ						1		2	1	16	1	8	2	**64**	5		1		36	**64**
y:										3	4		1	26	**62**	1	4		38	**62**
	7	*6*	*6*	*9*	*6*	*7*	*7*	*8*	*4*	*8*	*4*	*6*	*5*	*11*	*7*	*.2*	*.8*	*.0*	*37.5*	***62.5***

Table 10.14: Confusion matrix for Arabic listeners (N=10, 2634 responses). Correct (in bold) and wrong identifications in % for (presented stimuli in rows, selected categories in columns in semi-orthographic representation, values < .5 not indicated)

/ɑ/ is identified correctly in 85% of the cases. A substitution of /ɑ/ with /u/ is performed only by one participant of Egyptian origin who identifies /ɑ/ correctly in 16 cases but twice as /u/. This substitution could be due to a personal hypothesis of the learner enhanced by his knowledge of the English writing system. However, confusion of short /ɑ/ with /ʊ/ and even with /ɪ/ is quite common with Arab learners of German as e.g. evident in examples presented in section 10.4.3 by a Lybian learner[82] who has no knowledge of English showing category confusion of short /a/ with <u> and even with <ü> that both seem to be perceived as similar to German /ɑ/.

[82] This Lybian learner did not participate in the study and has no knowledge of English.

Long /ɑ:/ is perceived correctly in 75% of its occurrences and is substituted with /ɑ/ in 25% of the cases. Substitutions of /ɑ/ and /ɑ:/ with /ɛ:/ are performed only by one advanced learner (ArF13, Cairene, 25 years residence in Vienna).

Within the category of e-vowels, /ɛ/ is the most stable category (79% correct), while /ɛ:/ (46% correct) and /e:/ (44% correct) appear equally difficult. For /e:/, no substantial improvement is observed with advanced learners (41% correct for beginners, 47% correct for advanced learners). Difficulties with the correct identification of /e:/ result in substitutions with the whole range of German e- and i-qualities and even with a-qualities observed with beginners as well as with advanced learners. Similar difficulties are observed for /ɛ/ and /ɛ:/. Again, the full range from a-, e- and i-qualities are used as substitutes for these two categories. Interestingly, with advanced learners, /ɛ/ is not an attractive substitute for long /ɛ:/ (0%), whereas in 29% of the beginners' responses /ɛ:/ is substituted with /ɛ/. Here, as with a-qualities we observe the advanced learners' capability to discern phonemic short and long vowels, though this effect cannot be generalized due to the small number of only four advanced learners in the sample. A substitution of /ɛ:/ with a-qualities is observed occasionally (/ɑ/ 2 cases, /ɑ:/ 8 cases).

To summarize, the detection of contrasts between German a-, e- and i-categories appears to be difficult for Arabic learners. While short unconstricted /ɛ/ seems to be relatively easy, the differentiation of /e:/ and /ɛ:/ causes considerable difficulties. Substitutions of e-vowels with a-qualities are unidirectional. No substitutions of a- with e-qualities are observed in the data with the only exception of one advanced participant, who tends to substitute /ɛ/ and /ɛ:/ with /i:/ in 6 and 5 instances respectively. These substitutions may be interpreted as individual hyper-corrective strategy.

/i:/ is identified correctly in 73% of the cases and is substituted with /ɪ/ (19%) or /e:/ (6%).

The id_correct score for short unconstricted /ɪ/ is not much lower (69% correct) than for /i:/. The incorrect responses for /ɪ/ are dispersed over several categories: /ɛ/ (17%), /i:/ (7%), /e:/ (only 3%) and /ɛ/ (2%). However, with advanced learners, /ɪ/ is mostly confused with /i:/, all other categories do not seem to be attractive response options for more experienced listeners.

/ɔ/ is identified correctly in 73% of the cases and, interestingly, is confused with /o:/ only by advanced learners. Beginners rather perceive similarities with /ʊ/ (5%), /u:/ (1%), and /œ/ (7%).

The id_correct score for /o:/ differs remarkably between beginners (23% correct) and advanced learners (85%). Advanced learners show only problems with the contrast of /ɔ/ vs.

/o:/, whereas the beginners' responses are dispersed over all front and back rounded categories. Even <eu> is chosen in 3%.
A similar effect is observed for /ʊ/ (59% correct): Here, beginners and advanced learners mostly seem to perceive that the input signal is phonemically short, its quality however is not identified correctly: /ʏ/ (16%), /ɔ/ (12%) and /œ/ (7%) are the preferred response options while their long counterparts do not seem to be attractive substitutes.
/u:/ is identified correctly in 49% of the occurrences and is substituted with /ʊ/ (18%), /ʏ/ (12%) or /y:/ (10%) and less frequently with /o:/ (3%) and /ø:/ (5%). With advanced learners, no confusion between /o:/ and /u:/ is observed in either direction.
For unconstricted /œ/ (51% correct), a wide range of different qualities is chosen by Arabic participants from back round, front rounded and even front non-rounded categories. A clear preference, however, is observed for the short categories: /ʏ/ (15%), /ʊ/ (8%), /ɔ/ (6%), and /ɛ/ (5%, less often for advanced learners). In a few cases, diphthongs are chosen as response option.
/ø:/ (45% correct) is substituted with /ʏ/ (19%) and /y:/ (17%) and less often with /œ/ (9%) and /u:/ (6%). Occasional substitutions with /ʊ/ and <eu> are observed with beginners.
For /ʏ/ (64% correct), the most prominent substitute seems to be /ʊ/ (16%). However, substitution patterns vary between beginners and advanced learners: Beginners (56% correct answers for /ʏ/) substitute /ʏ/ most frequently with /ʊ/ (25%) and considerably less often with /u:/, /œ/, /ø:/ or /y:/. With advanced learners, the category boundary between /ʏ/ (75% correct) and /ʊ/ seems to be quite well established but the differentiation of /ʏ/ vs. /œ/ (11%) and /y:/ (7%) is not yet accomplished correctly.
Within the class of front rounded vowels, /y:/ (62% correct) is the one that is least confused with other qualities. Its most frequent substitute is /ʏ/ (26%). /u:/ is less often selected.
Several interesting asymmetries are observed that concern the differentiation of long constricted /e:/ - /i:/, /o:/ - /u:/, /ø:/ - /y:/ and short unconstricted /ɛ/ - /ɪ/, /ɔ/ - /ʊ/, and /œ/ - /ʏ/: For long constricted vowels, the front member of the opposition generally seems to be preferred, /e:/ > /i:/ (10%) vs. /i:/ - /e:/ (6%), /o:/ > /u:/ (6%) vs. /u:/ > /o:/ (3%), /ø:/ > /y:/ (17%) vs. /y:/ > /ø:/ (1%). For short unconstricted vowels, the opposite direction seems more common for /ɛ/ - /ɪ/ (/ɪ/ > /ɛ/ (17%) vs. /ɛ/ > /ɪ/ (5%)) and /ɔ/ - /ʊ/ (/ʊ/ > /ɔ/ (12%) vs. /ɔ/ > /ʊ/ (3%)) but not for /œ/ - /ʏ/ (/ʏ/ > /œ/ (8%) vs. /œ/ - /ʏ/ (15%)).
Asymmetric patterns of confusion are also observed for the contrast pairs /u:/ - /y:/ and /o:/ - /ø:/: hypercorrect /u:/ > /y:/ (10%), /u:/ > /ʏ/ (12%) vs. /y:/ > /u:/ (4%) and /y:/ > /ʊ/ (3%).

This asymmetric effect is stronger for o- and ö-vowels: /o:/ > /ø:/ (10%), /o:/ - /œ/ (5%) vs. /ø:/ > /o:/ (0%) /œ/ > /o:/ (0%), because ö-stimuli are mainly attracted by ü-responses.

To summarize, the differentiation of vowels within the class of rounded vowels, both front and back, causes considerable difficulties to Arabic learners of German. Similarly, difficulties are observed within the class of front vowels. Apart from /œ/, German front rounded vowels are mainly substituted with back qualities, suggesting that the primary differentiation of German vowels for Arab learners relies on the distinction [+/–round] as is also reflected in the one-dimensional MDS solution (see Figure 10.19 below). Within the class of front rounded vowels, short unconstricted /œ/ is the quality with the highest degree of aperture among front rounded vowels due to the lowest degree of lip rounding. A perceptual unrounding of /œ/ can thus be explained by the intrinsic characteristics of the vowel itself.

None of the advanced learners chooses a diphthong for any of the input stimuli. With beginners however, diphthongs, especially <eu>, seem to be a response alternative for front as well as for back rounded vowels, an effect that may be due with the familiarity of orthographic conventions as in French words.

10.4.4.4 Similarity and distance

The confusion matrix in Table 10.14 shows a clear differentiation of rounded vs. non-rounded vowels, while the perceptual differentiation of front vs. back rounded vowels appears to be more difficult for Arab learners.

V	ɑ	ɑ:	ɛ:	ɛ	e:	ɪ	i:	ɔ	o:	ʊ	u:	œ	ø:	ʏ	y:
a	1.000														
ɑ:	0.216	1.000													
ɛ:	0.021	0.057	1.000												
ɛ	0.000	0.007	0.193	1.000											
e:	0.004	0.024	0.567	0.167	1.000										
ɪ	0.000	0.000	0.030	0.143	0.066	1.000									
i:	0.000	0.000	0.024	0.023	0.136	0.188	1.000								
ɔ	0.000	0.000	0.000	0.000	0.000	0.000	0.000	1.000							
o:	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.153	1.000						
ʊ	0.008	0.000	0.000	0.000	0.000	0.000	0.004	0.104	0.053	1.000					
u:	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.013	0.091	0.172	1.000				
œ	0.000	0.000	0.018	0.035	0.006	0.000	0.005	0.074	0.059	0.137	0.017	1.000			
ø:	0.000	0.000	0.006	0.000	0.000	0.000	0.000	0.000	0.108	0.032	0.116	0.148	1.000		
ʏ	0.000	0.000	0.000	0.000	0.000	0.009	0.004	0.011	0.053	0.264	0.118	0.202	0.188	1.000	
y:	0.000	0.000	0.000	0.000	0.000	0.000	0.004	0.000	0.032	0.032	0.130	0.020	0.165	0.246	1.000

Table 10.15: Similarity matrix for L1 Arabic listeners

A comparison of the similarity scores in Table 10.15 shows that the most difficult contrast for Arabic listeners is German /e:/ - /ɛ:/ (0.567) which is not surprising since this pair causes problems to all language sub-groups and even to native speakers of German.

The contrast of /ʊ/ - /ʏ/ has the second highest sim_score (0.264) and may be explained by the short duration of the input signal together with the "blurring" effect of lip-rounding that is enhanced by the general high degree of perceived similarity.

Apart from these two contrast pairs, the highest degree of similarity is observed with pairs differing in *phonemic length* (/ɑ/ -/ɑ:/) *and constriction degree* (sim_scores between 0.246 and 0.148), as with /ʏ/ - /y:/, /ɛ/ - /ɛ:/, /i:/ - /ɪ/, /u:/ - /ʊ/, and /e:/ - /ɛ/. Moreover, the contrast pairs /œ/ - /ʏ/ (0.202), /ʏ/ - /ø:/ (0.188), and /ø:/ - /y:/ (0.165) have high sim_scores reflecting difficulties with the distinction of front rounded vowels that differ in constriction location [mid-palatal] vs. [pre-palatal] together with differences in length and constriction degree for /ʏ/ - /ø:/.

Difficulties with respect to the *constriction location* (sim_scores below 0.143) are observed for several vowel pairs. The contrast of [pre-palatal] vs. [mid-palatal], [velar] vs. [palatal] and [uvular] vs. [palatal] is not sufficiently distinguished causing perceptual substitutions of front rounded vowels with velar and uvular rounded qualities. The contrast [uvular] vs. [velar] has sim_scores below 0.104.

There is hardly any confusion of rounded with non-roundeded vowels, with only two exceptions where similarity scores are relatively low: /ɑ/ vs. /ʊ/ (sim_score 0.008) and /œ/ vs. /ɛ/ (0.035) or /ɛ:/ (0.018).

non-rounded Vs	*sim_scores*
e: - ɛ:	.567
ɑ - ɑ:	.216
ɛ - ɛ:	.193
i: - ɪ	.188
e: - ɛ	.167
ɪ - ɛ	.143
i: - e:	.136
i – e:	.066
ɛ: - ɑ:	.057
ɪ - ɛ:	.030
i: - ɛ:	.024
e: - ɑ:	.024
i: - ɛ	.023
ɛ: - ɑ	.021
ɑ - ʊ	.008
ɛ - ɑ:	.007
e: - ɑ	.004

Table 10.16: Relative difficulty of non-rounded vowels in terms of similarity scores for L1 Arabic listeners

The sim_scores in Table 10.15 will therefore be discussed here for rounded vs. non-rounded vowels separately. Table 10.16 and Table 10.17 show the similarity scores for rounded and non-rounded vowels separately indicating the relative difficulty of these contrast for L1 Arabic listeners.

A common phenomenon with Arab learners of German is the confusion of o- and u-qualities. In the data of the present study we observe, moreover, category confusion between front rounded and back rounded qualities. With advanced learners, confusion within the class of rounded vowels is less strong and observed mainly with short vowels. Table 10.17 summarizes the sim_scores for vowel contrasts involving rounded vowels.

rounded Vs	*sim_scores*
ʊ - ʏ	.264
ʏ - y:	.246
œ - ʏ	.202
ʏ - ø:	.188
u: - ʊ	.172
ø: - y:	.165
o: - ɔ	.153
œ - ø:	.148
œ - ʊ	.137
u: - y:	.130
u: - ʏ	.118
u: - ø:	.116
ø: - o:	.108
ʊ - ɔ	.104
u: - o:	.091
œ - ɔ	.074
o: - œ	.059
o: - ʏ	.053
ʊ - o:	.053
œ - ɛ	.035
o: - y:	.032
ʊ - ø:	.032
ʊ - y:	.032
œ - y:	.020
œ - ɛ:	.018
u: - œ	.017
u: - ɔ	.013
ʏ - ɔ	.011
ɑ - ʊ	.008

Table 10.17: Relative difficulty of rounded vowels in terms of similarity scores for L1 Arabic listeners

The highest sim_scores within the class of rounded vowels are obtained for vowel contrasts involving short unconstricted pre-palatal /ʏ/. The /ʊ/ - /ʏ/ contrast has the highest sim_score (0.264), presumably due to the short intrinsic length of the stimuli for both qualities. /ʏ/ is moreover perceived as similar to other front palatal rounded vowels.

Interestingly, /ʏ/ is the most preferred response category for L1 Arabic learners (10.6%) and the category with the lowest id_wrong score among front rounded vowels, attracting a fairly high percentage of responses for /œ/ and /ø:/-stimuli. The high sim_scores discussed in the preceding section result from the fact that /ʏ/ can function as perceptual substitute for all rounded vowels except /ɔ/.

The MDS solutions presented in the next section visualize the low perceptual distinctivity of categories within the class of German rounded vowels.

10.4.4.5 The perceptual vowel map for L1 Arabic listeners

From the similarity scores calculated by Shepard's formula, distance scores and MDS solutions were derived. For the non-metric MDS solutions, RSQ values of 0.655 in dimension 1, 0.882 in dimension 2, and 0.914 are obtained.

The one-dimensional MDS solution in Figure 10.19 shows a clear differentiation of rounded vs. non-rounded vowels. /œ/ is the only quality that is differentiated in this dimension from the cluster of all other rounded vowels.

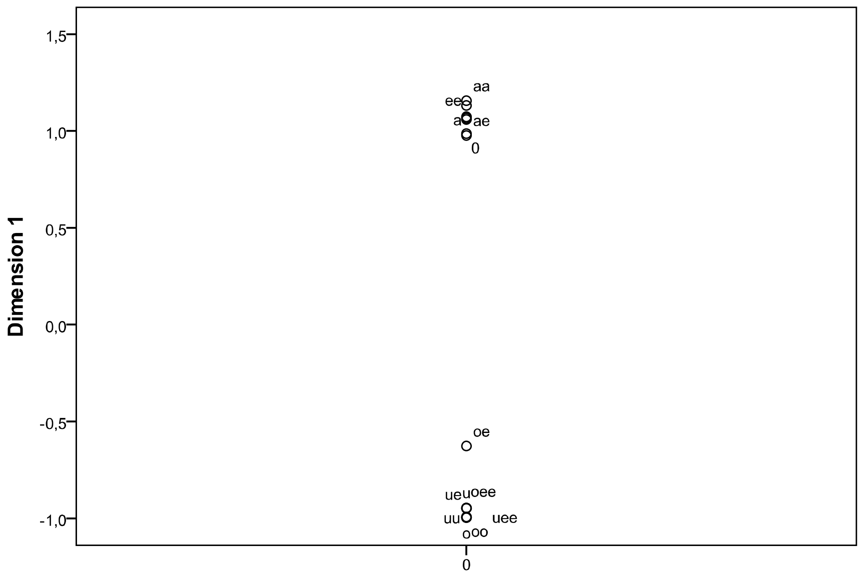

Figure 10.19: One-dimensional MDS representation of the perceptual map of German vowels for L1 Arabic listeners (N=10, 2634 responses, RSQ .655)

In the two-dimensional MDS solution in Figure 10.20, a spatial representation of perceived similarity relations is combined with a representation of preference patterns indicated by type size. The solution offered shows that rounding is the major distinctive feature dividing the perceptual vowel space for German into rounded vs. non-rounded vowels. It moreover shows that Arabic listeners differentiate clearly between pharyngeal vowels and other classes but that mid-palatal and pre-palatal categories are less distinctive from each other. Within the class of rounded vowels we observe problems in the differentiation of front vs. back rounded vowels.

/ɔ/ received the lowest id_wrong scores within the class of rounded vowels and can therefore considered to be most distinctive from all other rounded qualities. It is moreover one of the most preferred response categories. The perceptual distance of /œ/ and /ɔ/ from other rounded vowels becomes evident in the MDS solution. For /œ/ this distance can be explained by instances of perceptual substitution with front vowels and for /ɔ/ it is due to the high percentage of correct identifications.

The two-dimensional perceptual map reveals moreover a high degree of perceived similarity for /y:/ and /ø:/, for /u:/ and /o:/ and for /i:/ and /ɪ/. /i:/ and /e:/, however, are perceptually more distant in the vowel map, in spite of several examples of perceptual confusion for these two qualities. The larger distance in the map is to be explained by the large range of different perceptual substitutes (/ɑ ɑ: ɛ ɛ: ɪ i:/) for /e:/, whereas /i:/ is substituted only by /ɪ/ (19%) and /e:/ (6%).

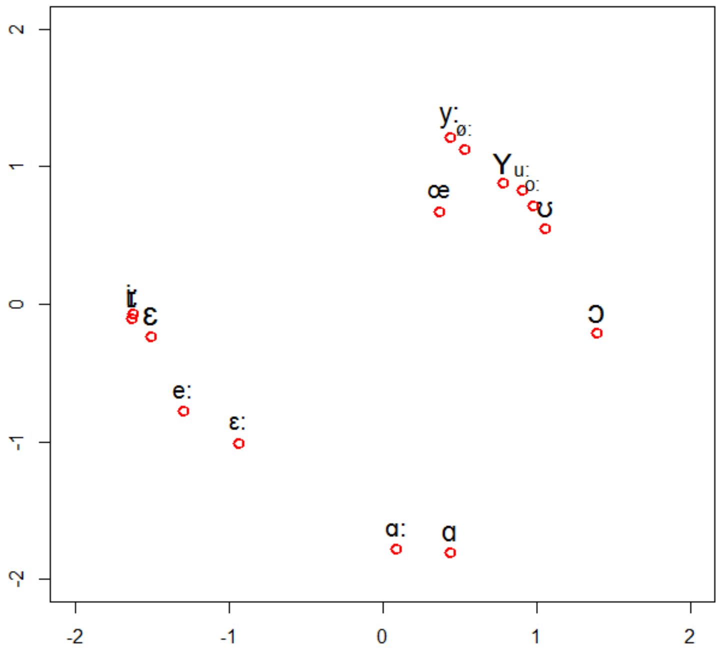

Figure 10.20: Two-dimensional MDS representation of the perceptual map of German vowels for L1 Arabic listeners (N=10, 2634 responses, RSQ .882)

The three-dimensional MDS solution in Figure 10.21 provides a more differentiated representation for the class of rounded vowels as perceived by Arabic native listeners: /y:/, /u:/ and /ø:/ are clustered very closely together, indicating a certain degree of perceptual sensitivity of Arabic listeners to length distinctions as evident in the confusion patterns though the similarity relations between /y:/ - /ø:/ and /y:/ - /u:/ are asymmetric in nature. As in the two-dimensional solution, /ɔ/ and /œ/ are better differentiated from other rounded qualities and i-vowels almost overlap in both representations.

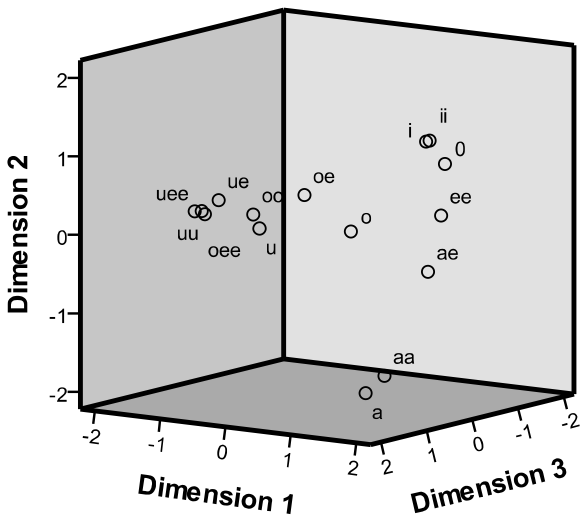

Figure 10.21: Three-dimensional MDS representation of the perceptual map of German vowels for L1 Arabic listeners (N=10, 2634 responses, RSQ .914)

10.4.5 Summary

To summarize, native listeners of Arabic show least difficulties with German /ɔ/, /ɑ/ and /ɛ/, though /ɔ/ (86% correct) and /ɛ/ (79% correct) do not occur in the Arabic vowel system. This fact can be explained by the “new vs. similar” dichotomy (Flege 1988c).

Most difficulties are observed for the so-called mid categories /e:/, /ø:/, /ɛ:/ and /o:/.

The inter-category differentiation of the short unconstricted rounded qualities /ɔ/, /ʊ/, /œ/ and /ʏ/ is especially difficult.

Generally, short unconstricted vowels are more often selected than long qualities. Interestingly, /ʏ/ is the most preferred category followed by short /ɛ ɔ ʊ ɑ/. The least preferred categories are long constricted rounded vowels: /o: u: ø:/.

The differentiation of vowels within the class of rounded vowels, both front and back, causes considerable difficulties to Arabic learners of German. Similarly, difficulties are observed within the class of front vowels. Apart from /œ/, German front rounded vowels are mainly substituted with back qualities,

For /ɑ/ and /ɑ:/, hardly any difficulties with respect to quality are given and only occasional interference with <ä> is observed.

While /ɛ/ is a very stable category, more problems are observed for /ɛ:/ and /e:/.

In general, Arabic learners of German seem to be sensitive to phonemic length distinctions, but the qualitative differentiation of palatal vs. pre-palatal qualities (/i: ɪ/ and /e: ɛ ɛ:/) causes several problems in identification.

Despite acoustic overlap, /ɪ/ vs. /e:/ can be distinguished quite well from each other, indicating that subjects are sensitive to the duration cue, though more difficulties are observed with beginners.

/ɔ/ is the least difficult category for Arabic learners, it is moreover one of the three most preferred response options (after /ʏ/ and /ɛ/). Severe problems are observed within the class of front rounded vowels as well as with the differentiation of front rounded vs. non-rounded vowels. Within the class of front rounded vowels, /y:/ could be expected to be the most stable and least marked category showing the smallest range of competing substitutes at different constriction locations. However, /y:/ is not the least difficult category within the class of front rounded vowels. Rather, /ʏ/ shows the lowest number of wrong identifications and the highest preference score. /ø:/ is particularly difficult for Arabic listeners. A substitution of /ø:/ with /y:/ (17%) or /ʏ/ (19%) is a common strategy, whereas /y:/ is hardly ever substituted by /ø:/ (only 1%).

The contrast of long constricted uvular /o:/ and velar /u:/ is relatively well distinguished and only a small asymmetry effect towards /u:/ is observed.

To summarize, the greatest portion of observed wrong vowel identifications is due to the insufficient differentiation of differences in constriction location with respect for contrasts of mid- vs. pre-palatal, palatal vs. velar/uvular, and uvular vs. velar vowel qualities. Moreover, within the class of rounded vowels, considerable difficulties the differentiation of front vs. backare observed. Similarly, difficulties are observed within the class of front vowels. Apart from /œ/, German front rounded vowels are mainly substituted with back qualities. The perceived similarity is highest with contrast pairs that differ in phonemic length and constriction degree and lower with differences concerning constriction location. Asymmetric patterns are observed for these contrast pairs. For the class of long constricted vowels /e:/ - /i:/, /o:/ - /u:/, /ø:/ - /y:/ and for short unconstricted /œ/ - /ʏ/ the higher more front members of a contrast pair attract responses from the lower quality. With short unconstricted /ɛ/ - /ɪ/ and /ɔ/ - /ʊ/, the opposite pattern is observed and the lower more open quality is preferred.

10.5 English

10.5.1 General characteristics

English belongs to the western branch of the Germanic languages within the Indo-European language group and is related to German, Dutch, Yiddish, Afrikaans and Frisian, though it has lost many of the characteristic features of this language family.

Over the past centuries, English has expanded to countries and regions all over the world and has established itself as the leading language in the world, serving as lingua franca in many aspects of daily life in business, academia, journalism, sports etc. English is spoken or learnt as a native language not only on the British Isles and North America, but also in Australia, New Zealand and South Africa. It has the status of an official language in India, Singapore, the Philippines, Western Samoa, Tanzania and Cameroon and the status of the only official language in around thirty countries such as Ghana, Liberia, Nigeria, Uganda, Zimbabwe, Jamaica and the Bahamas (Siemund 2003; Finegan 2011). Moreover, in several countries, many of which are former or current colonies or dependencies of the United Kingdom, various pidgins, creoles and creole-based varieties are spoken.

The widespread use and distribution of English has led to the emergence of several distinct regional varieties. Therefore, the term English cannot be used to describe one homogeneous linguistic system, but is rather used as a cover term for two co-existing standard varieties and at least a dozen well known and recognized regional varieties. The focus of the present description will be on American, English and Australian English, since natives from these three varieties participated in the perception experiment.

Two standard varieties are distinguished: Standard British English (SBE) and (Standard) American English (AE). Standard American English comprises educated varieties spoken in the United States and Canada. The educated varieties in England, Wales, Australia, New Zealand and South Africa have traditionally been subsumed under the British standard, with Ireland and Scotland having more complex affiliations (Finegan 2011: 78). Between the two standard varieties relatively small differences are observed (mainly in pronunciation, lexis and spelling), but substantial differences are rather found between standard and non-standard systems and between different regional non-standard varieties, especially in pronunciation (Siemund 2003). In North America, no uniform accent exists, even in standard varieties there is considerable latitude in pronunciation though a kind of "network standard" that is essentially based on a form of inland northern dialect (Finegan 2011) can be identified in broadcasting.

For the British Isles, the so called RP (Received Pronunciation) has long been considered as standard pronunciation associated with the broadcast company BBC and prestigious 'public' schools, though it is actually spoken only by three to five percent of England's population (Trudgill & Hannah 2002: 9). So-called Estuary English (McArthur 2002) is another increasingly favoured accent varying along a continuum between RP and the regional dialect of London (Finegan 2003: 79).

A major difference between Standard American and Standard British English is the number of phonemes in the system. While the systems of consonants in the varieties of English vary only according to a few well-defined parameters, the extent of variation found in the vowel systems of English regional varieties is considerably larger and more difficult to describe concisely (see section 10.5.2). A sharp north-south cline in the differences of vowel quality and vowel quantity is found in both Standard American and Standard British English (Siemund 2003: 7).

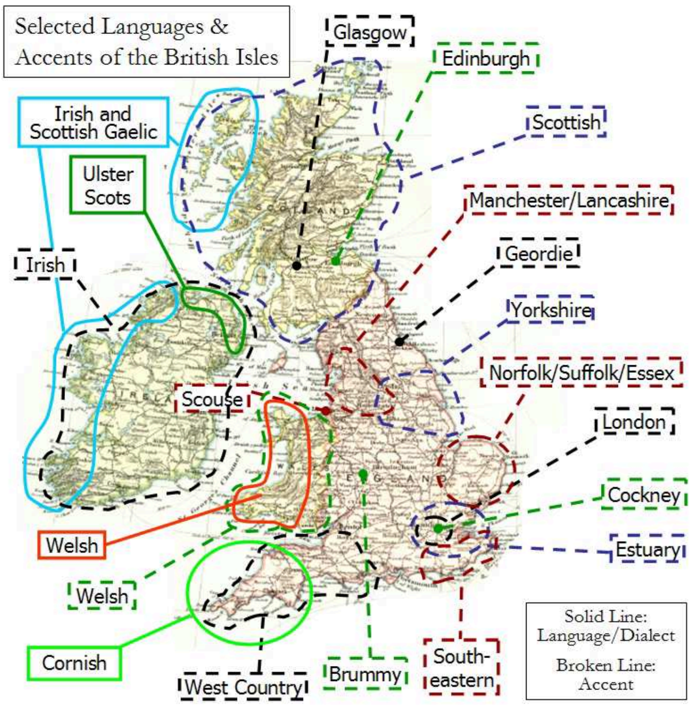

Figure 10.22: Accents and Dialects of English on the British Isles[83]

[83]http://upload.wikimedia.org/wikipedia/en/archive/3/34/20060828182648!Selected_languages_and_accents_of_the_british_isles2_rjl.jpg (accessed 2014-11-09)

In Great Britain, two important dialect boundaries are discerned: one between England and Scotland and one between the north of England and the central/southern part of England. Moreover, several regional accents are found.

In the Atlas of North American English (ANAE), Labov, Ash & Boberg (2006) distinguish seven major dialect regions: New England, New York and the Mid Atlantic States, the North, the Midland, the South, the West, and Canada. Labov et al.'s description of regional dialects is mainly based on ongoing sound changes active in the 1990s (see Figure 10.24). In total, Labov et al. (2006) distinguish 19 dialects and 27 mapping features for North America.

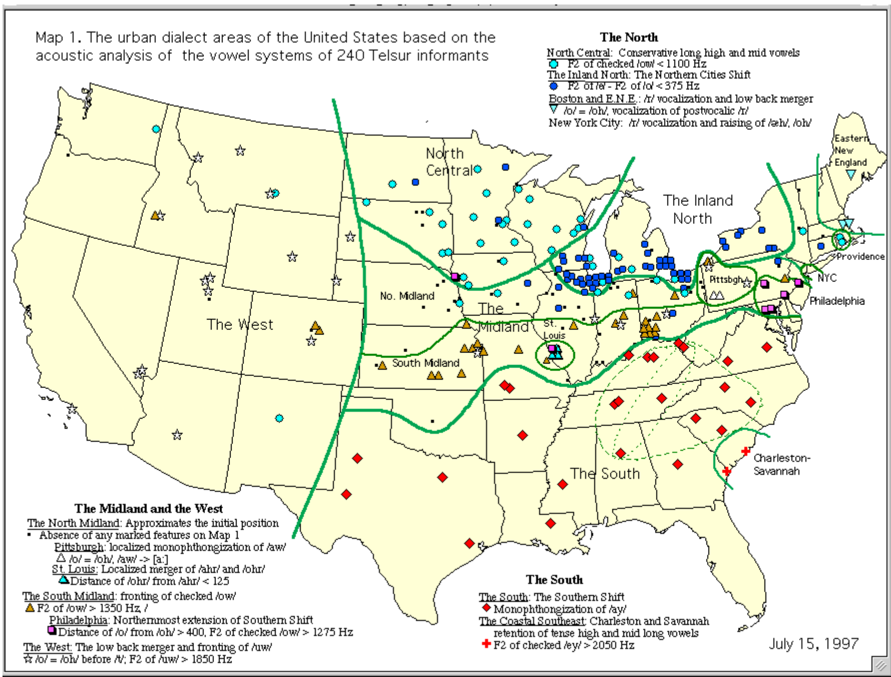

Figure 10.23: Urban dialect areas of the United States based on an acoustic analysis in the Telsur study[84] (figure reproduced from Labov, Ash & Boberg 1997[85])

Three types of active sound changes are described in ANAE for dialects of North America: (1) mergers and splits, (2) the fronting of back vowels, and (3) chain shifts by which vowels "rotate" within and across sub-systems. The vowel systems in different regions of North America are described as rotating in opposite directions by ongoing sound changes and chain shifts such as the Northern Cities Shift in the inland North, The Southern Shift in the South, the Low Back Merger in the West, or the Canadian Shift (see Figure 10.23 illustrating two

---

[84] The Telsur Project is a telephone survey of ongoing sound changes in North American English (cf. http://www.ling.upenn.edu/phono_atlas/home.html) (accessed 2014-11-09)

[85] http://www.ling.upenn.edu/phono_atlas/NationalMap/NatMap1.html (accessed 2014-11-09)

prominent types of chain shifts of North America). While some sound changes are limited to one specific dialect, others spread across several dialects.

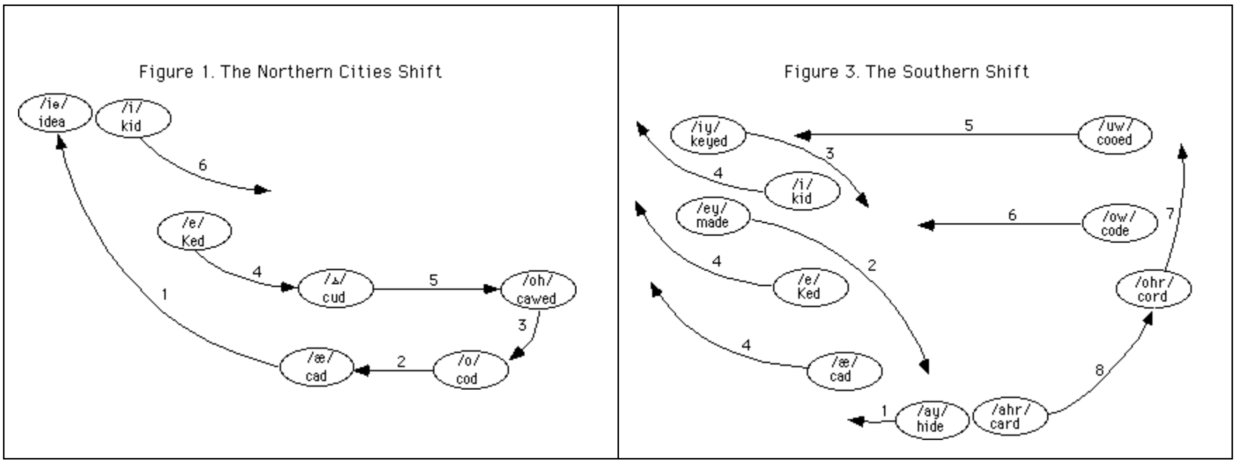

Figure 10.24: Chain shifts in North American English (figures reproduced from Labov, Ash & Boberg 1997[86])

10.5.2 Phonological and phonetic description

While English consonants (see Table 10.18)[87] have remained relatively stable throughout history and in contemporary regional varieties, the vowel systems exhibit striking instability, which is manifested historically and in the present-day form of the two standards and regional varieties.

	Bilabial	Labiodental	Dental	Alveolar	Postal-alveolar	Palatal	Velar	Glottal
Plosives	p b			t d			k g	
Affricates					tʃ dʒ			
Nasal	m			n			ŋ	
Fricative		f v	θ ð	s z	ʃ ʒ			h
Approximant				ɹ		j	w	
Lateral Approximant				l				

Table 10.18: English consonants (Ladefoged 1999; Hillenbrand 2003)

86 http://babel.ling.upenn.edu/phono_atlas/ICSLP4/Figure_1.GIF, http://babel.ling.upenn.edu/phono_atlas/ICSLP4/Figure_3.GIF (accessed 2014-11-09)

87 The pronunciation of consonants tends to vary socially more than regionally (Finegan 2011), with the exception of post-vocalic /r/ and intervocalic /t/.

The vowel systems in present-day English and its regional varieties may vary with respect to four basic aspects (Moulton 1974: 72f): (1) the *number of phonemes* in the system of a given regional variety, (2) the occurrence of phonemes in specific *words and contexts*, (3) the exact *phonetic realization* of a specific phoneme that may differ in regional variants, and (4) *variation* or even diffusion and indeterminacies in parts of a given vowel system.
Table 10.19 summarizes the vowel phonemes of British and American English and shows allophonic variants as realized in different regional varieties.

phoneme	allophonic variants[88]	example
/i/	[iː i i$^{\text{i}}$ i̞ i$^{\text{ə}}$]	*heed, beat*
/ɪ/	[ɪ ɪ.]	*hid, bit*
/e/	[eɪ e$^{\text{i}}$ e ẹ. e$^{\text{ə}}$]	*hayed, bait*
/ɛ/	[ɛ ẹ.]	*head, bet*
/æ/	[æ æ.]	*had, bat*
/u/	[uː u u$^{\text{u}}$ u̞. u$^{\text{ə}}$]	*who'd, boot*
/ʊ/	[ʊ ʊ.]	*hood, put*
/o/	[oʊ əʊ o$^{\text{u}}$ o o̞. o$^{\text{ə}}$]	*hoed, board, boat*
/ɔ/	[ɔ ɔ. ɔː]	*hot, bought*
/ɒ/	[ɒ]	*hod, bother*
/ʌ/	[ʌ ʌ.]	*hut, bud, but*
/ɑ/	[ɑ ɑ.]	*hard, father*
/ɜ/	[ɜ ɜ. ɝ]	*herd, bird*
/ai/	[æɪ aɪ ɔɪ a$^{\text{e̞}}$ a$^{\text{ə}}$ əɪ][89]	*hide, height, bite*
/ɔi/	[ɔɪ oɪ ɔ$^{\text{e̞}}$ ɔ$^{\text{ə}}$]	*boy*
/au/	[au̞ æu̞ əu̞]	*how, bout*
/ju/[90]	[ju u]	hued, cue

Table 10.19: English vowel phonemes and their allophones

The vowels of English are traditionally divided into two sets of so-called "tense" /i e ɑ o u/ vs. "lax" /ɪ ɛ æ ʊ ʌ/ vowels. However, phonetic differences between the two series /i e ɑ o u/ and /ɪ ɛ æ ʊ ʌ/ cannot simply be described as a difference in muscular tension. In terms of duration, /i u/ as in *beat, boot* tend to be longer than the corresponding short vowels /ɪ ʊ/ in similar contexts, e.g. *bit, put*. In the vowel pairs /i ɪ/, or /u ʊ/, the "lax" vowel is shorter, lower and less constricted than its "tense" counterpart. However, in phonetic terms, "tense" and "lax" vowels differ not only in length or rather duration but also in quality, i.e. in articulatory differences in position of tongue, constriction degree, lip position and acoustically in terms of spectral properties and dynamic spectral changes. Vowel-intrinsic spectral change (VISC)

---

88 for discussion, see Ladefoged (1993: 76, 82) and Moulton (1974: 73ff, 91)

89 For discussion of allophones see Moulton (1974: 77).

90 This diphthong can also be treated as a sequence of consonant+vowel

throughout the nucleus proves to be perceptually relevant for several English vowels (Nearey & Assmann 1986), especially in slowly articulated citation utterances.

The phonetic realization of vowels plays a crucial role in differentiating regional varieties of English around the world.

For American English, there is no one accepted standard vowel system that is valid for the entire area of North America. Educated speakers in different parts of North America use quite different vowel systems differing with respect to the four aspects mentioned (Moulton 1974: 72f).

For Standard British English, six short vowels /ɪ e æ ʌ ɒ ʊ/ (*bit, bet, bat, but, pot, put*) and five long vowels /iː ɜː ɑː ɔː uː/ (*beat, bird, pass, board, boot*) are described (Roach 2000). The diphthongs in SBE can be divided in three types gliding into /ə/, /ɪ/ or /ʊ/, where the first part is longer and stronger than the second part (Roach 2000: 21): (1) up-gliding /eɪ aɪ ɔɪ/, (2) up-gliding /əʊ aʊ/, and (3) "ingliding" /ɪə eə ʊə/ for vowels before rhotic /r/, though /ʊə/ is increasingly rare, since many speakers realize [ɔː] instead, e.g. *moored* [mʊəd] ~ [mɔːd] (Roach 2000: 21f). Moreover, a set of triphthongs is posited for the non-r-realizing dialects composed of any of the five closing diphthongs ending in /-ɪ/ or /-ʊ/ with /-ə/ added to it, as in *player, fire, lower, power, loyal* (Roach 2000: 24).

Australian English is very similar at the phonological level to South East Urban British English. The vowels system consists of 11 monophthongs (ignoring [ə]) and five closing diphthongs. Australian English is a non-rhotic variety: no syllable-final /r/ is realized and /Vr/ sequences are replaced by centring diphthongs. Traditionally, Australian English has been described as a stylistic continuum ranging from "Cultivated" and "General" to "Broad".

The distribution of "tense" and "lax" vowels is strongly related to stress patterns and syllable structure (for discussion, see Ladefoged 1993: 86). All vowels and diphthongs can occur in closed stressed syllables, but /ɪ ɛ æ ʊ ʌ/ do not occur in stressed open syllables.

The duration of vowels differs according to their position in stressed or non-stressed syllables and according to their post-vocalic context: Roughly, stressed vowels appear short before voiceless consonants, half-long before voiced consonants, and long in word-final position (Moulton 1974: 75), e.g. *leaf* [lif] – *leave* [li.v] – *lee* [liː].

Differences between SBE and Standard American English are especially evident in the sub-system of low and back vowels (see Figure 10.25). Speakers of most forms of British English have a tense vowel /ɑ/ that is realized in both open and closed syllables (*e.g. calm, car, card*) and a lax /ɒ/ that occurs only in closed syllables (e.g. *cod, common, con*). In most varieties of American English, there is no such "tense-lax" contrast before /ɹ/, the sound produced is somewhere between the two qualities (see Figure 10.25). Before /ŋ/ and /ʃ/, only the lax vowels occur in most varieties (for discussion, see Ladefoged 1993: 87f).

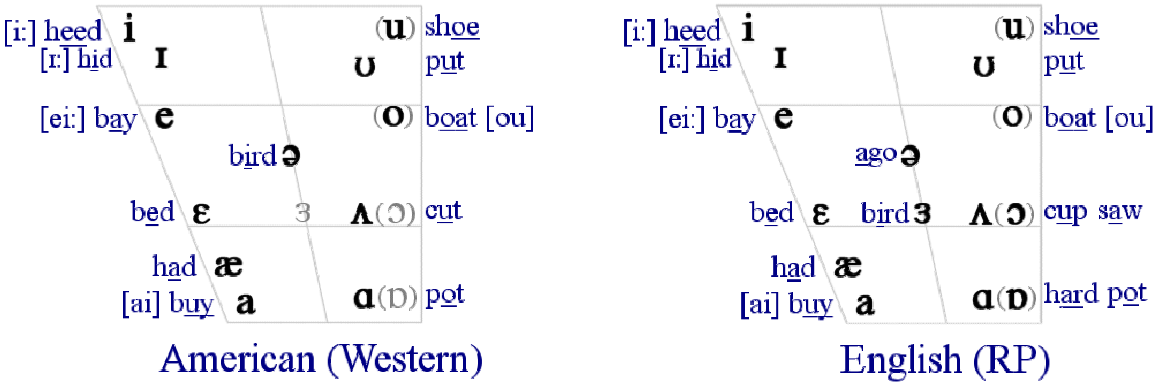

Figure 10.25: Comparison of American and English vowel inventories (figures reproduced from Smith 2004/2013[91])

British English and American English have /a/ and /æ/, which may be differently distributed. In most forms of American English, front /æ/ as in *hat, cam* and back /ʌ/ as in *hut, come* have no corresponding "tense" vowel, but both are shorter than /ɑ/ as in *spa*. Words with post-vocalic fricatives /f θ s/ or nasals tend to be realized with [æ] by Americans and with [ɑ:] in Southern British English or [ɑ] in Northern England, e.g. *bath*.

Most speakers of British English have an additional lax vowel /ɒ/ as in *cod, common, con* as counterpart to /ɑ/ as in *calm, car, card*. Standard British English /ɒ/ corresponds to either /a(:)/ or /ɔ(:)/ in Standard American English, e.g. *clock* SBE /klɒk/ vs. SAE /kla:k/. The contrast between /ɒ/ and /ɔ/ as in *cot, sock, dollar* vs. *caught, stalk, caller* is not distinguished by speakers in wide regions of North America (more than half of the geographical territory, though still only by a minority of the population, see Labov et al. 2006).

/ɜ/ (as in *bird*) is described as central, moderately rounded, and constricted with lateral tongue constriction in most American dialects (Moulton 1974: 92). In its realization as [ɝ] the additional feature of rhotacization is involved (for discussion, see Ladefoged 1993: 84).

Rhotacized vowels as in *beard, heard, hard* are absent in non-rhotic dialects, as is typical of the North American South and the New England region in North America as well as for British RP. In these varieties, the preceding vowel is usually lengthened and often glides into a schwa sound, resulting in the series of SBE diphthongs /ɪə/, /eə/ and /ʊə/ that are not used in Standard American English.

91 http://members.peak.org/~jeremy/dictionaryclassic/chapters/pronunciation.php (accessed 2014-11-09)

In the context of intervocalic *r*-sounds, American English has fewer vowel distinctions before /r/, causing *merry – marry – Mary* to rhyme, as do *mirror – nearer* and *furry – hurry*. Moreover, in many American English varieties, rhotacization occurs for vowels followed by [ɹ], as in '*beard, bared, bard, board, poor, tire, hour*'.

	SHORT		LONG					
			upgliding[92]				ingliding[93]	
			front upgliding		back upgliding			
	V		Vy		Vw		Vh	
nucleus	front	back	front	back	front	back	front	back
high	i	u	iy		iw	uw		
mid	e	ʌ	ey	oy		ow		oh
low	æ	o[94]		ay		aw	oh	ah
high	*bit*	*put*	*beat*		*suit*	*boot*		
mid	*bet*	*but*	*bait*	*boy*		*boat*		*bought*
low	*bat*	*pot*		*bite*		*bout*	*halve*	*father*

Table 10.20: Organization of North American English vowels in initial position in the ANAE notation system (Labov et al. 2006; Labov 2010: 11)

The symbols /i e u o/ and /ai au/ as traditionally used in many works on English phonology are somewhat misleading, suggesting that they refer to a set of monophthongs and a set of diphthongs. However, a considerable amount of diphthongization of phonemic monophthongs /i e u o/ is observed, especially in American English (see Hillenbrand et al. 1995; Hillenbrand 2003; for spectral movement in Canadian English, see Nearey & Assmann 1986). In very general terms, diphthongization increases along two dimensions: (1) a decrease in degree of height or jaw aperture, as in *boot – boat – bout* [bu$^{(u)}$t] – [bo^{u}t] – [baut], and (2) along the dimension 'long – half-long – short', as in *leet* [lit] – *lead* [li.^{i}d] – *lee* [li:^{i}] (Moulton 1974:

92 Middle English high and mid long vowels /i: e: u: o:/ underwent diphthongization in many English dialects. The notation integrates diphthongized vowels into the subsets of "true" diphthongs /ay oy aw/ that show a parallel participation in chain shifts (Labov 2010: 12).

93 The subset of long ingliding vowels notated here with /h/ is realized phonetically as length for low vowels and as the inglide [ə] for mid and high vowels.

94 /o/ is pronounced as [ɑ] in most North American dialects. The originat back rounded [ɒ] is retained in Eastern New England, Canada and Western Pennsylvania (after the merger with /oh/. A process of unrounding is not assumed in those dialects as it did in Western New England. Unrounding of /o/ plays a major role in the reconstruction of the Northern Cities Shift (Labov 2010: 12).

76). The characteristic feature of diphthongization of what is phonemically considered to be a monophthong is also transferred into German in L2 acquisition by English native speakers.

A parallel development of diphthongized vowels and the "true" diphthongs /ai ɔi au/ in sound change and in chain shifts is postulated by Labov (2010: 12) and Labov et al. (2006). It is most evident in the Southern Shift (see Figure 10.24), which is common in the southern US, and in the South of England, Australia, New Zealand and South Africa. However, there are several dialects that did not develop such diphthongization for monophthongal long /i: e: u: o:/ as in Scotland, Carribean English, traditional upper-class Charleston English and forms of English with a German or Scandinavian substrate as in Eastern Pennsylvania, Wisconsin or Minnesota (Labov 2010: 12).

Labov (2010: 11ff) provides a representation of the English vowel system[95] that is posited as the initial stage for North American English, from which all current sound changes are assumed.

Major patterns of vowel shifting in American English are identified in the Atlas of North American English (Labov et al. 2006). The vowel system of the northern dialect of American English is characterized by the Northern Cities Chain Shift (see Figure 10.24), currently occurring in large urban centres on both sides of the U.S./Canada border, merging among others low back vowels /ɔ/ and /ɑ/, e.g. *walk/wok* or *dawn/don*.

The system of the Southern dialect of American English is characterized by the Southern Vowel Shift (see Figure 10.24), fronting the back vowels /u/ and /o/, the fronting and raising of /ɪ/ and /ɛ/ and lowering and backing of /i/ and /e/ (Labov et al. 2006).

The Midland, Western and New England dialects are characterized by a merger of the low-back vowels /ɑ/ and /ɔ/, creating homophones such as *caught* – *cot* or *dawn* – *Don*.

The Midland dialect shows no other distinct features than the merger of /ɑ/ and /ɔ/, a feature that is not found in the Mid-Atlantic dialect, where these two vowels are more distinct due to raising of /ɔ/.

Several studies dealt with the acoustic description of English vowels in regional varieties: Deterding (1997) presents an acoustic analysis for monophthong vowels in Standard British English pronunciation for five male and five female SBE speakers from a digital speech database (from a set of monologues, newsreading and commentary spoken by BBC broadcasters). Deterding's average values for F1-F3 (see Table 10.21 and Figure 10.26) are based on about ten tokens for each of the eleven monophthong vowels of SBE realized by

95 The transcriptions used in many dialect studies are mostly not in accordance with the principles of the IPA but – as is argued – are a good representation for the dynamics of English sound change.

each speaker. /e/ and /o/ are not included in the data because of the considerable amount of diphthongization for these two qualities.

	F1	F2	F3
i:	280	2249	2765
ɪ	367	1757	2556
e	494	1650	2547
æ	690	1550	2463
ɜ:	478	1346	2488
u:	316	1191	2408
ʊ	379	1173	2445
ɔ:	415	828	2619
ɒ	558	1047	2481
ʌ	644	1259	2551
ɑ:	646	1155	2490

Table 10.21: Formant frequencies F1, F2, F3 for five male speakers of Southern British English (Deterding 1997)

A pioneering study on the acoustic description of Australian English vowels was done by Bernard (1970), presenting a speaker-averaged representation of formant frequencies F1-F3 based on the vocalic nuclei in /hVd/ syllables (see also acoustic data presented by Bernard & Mannell 1986) for the 11 monophthongs as produced by more than 200 adult male speakers (see also Mitchell & Delbridge 1965, and more recently, Cox 1999; Cox & Palethorpe 2001).

An overview of Canadian English phonetics was presented in the Atlas of North American English by Labov et al. (2006). A more detailed analysis and discussion of regional variation in Canada is presented by Hagiwara (2006) and Boberg (2008), suggesting a more refined division of Standard Canadian English at the phonetic level into six regions.

One of the most frequently cited papers for vowels of English is the seminal study of Peterson & Barney (1952) reporting the results of an acoustic analysis of 10 American English vowels in a /hVd/ context produced by 33 men, 28 women and 15 children and providing acoustic measurements from narrow-band spectra consisting of formant frequencies F1-F3, formant amplitudes, and fundamental frequency at one steady-state point in time. The /hVd/ signals were also presented to listeners in an identification task, revealing a strong relationship between formant frequency patterns and the identification of the intended vowel. However, considerable inter-speaker variability of formant frequencies and a substantial overlap among adjacent vowels was found in the data.

While Peterson & Barney's measurements have played a central role in the characterization of American English vowels as well as in theories of vowel recognition, there are however some methodological limitations to the database: Measurements were taken at one single time slice, duration measurements were not made, and information about dynamic properties such as duration and spectral change is not available. Moreover, the data were collected more than 50 years ago, the subjects were not screened for dialect, the sampling with respect to dialectal variation is neither controlled nor balanced but approximates rather the speech of the American eastern seaboard than General American English (for reviews, see Clopper, Pisoni & de Jong 2005; Hillenbrand, Getty, Clark & Wheeler 1995).

Though Peterson & Barney's data show a significant overlap of acoustic values in an F1xF2 chart (see Figure 10.26), especially for /ɑ/ - /ɔ/ and /ʊ/ - /ɜ/, the listening tests show that a significant "crowding" of vowels in some areas of the static F1xF2 representation of the vowel space is not accompanied by an increase in perceptual confusion of vowels in these regions. The fact that listeners can identify vowels correctly despite this overlap in the F x F2 dimensions indicates that static measures of F1 and F2 are only poor predictors for vowel identification since they cannot render the dynamic spectral properties of vowels that are relevant perceptual cues for vowel identification.

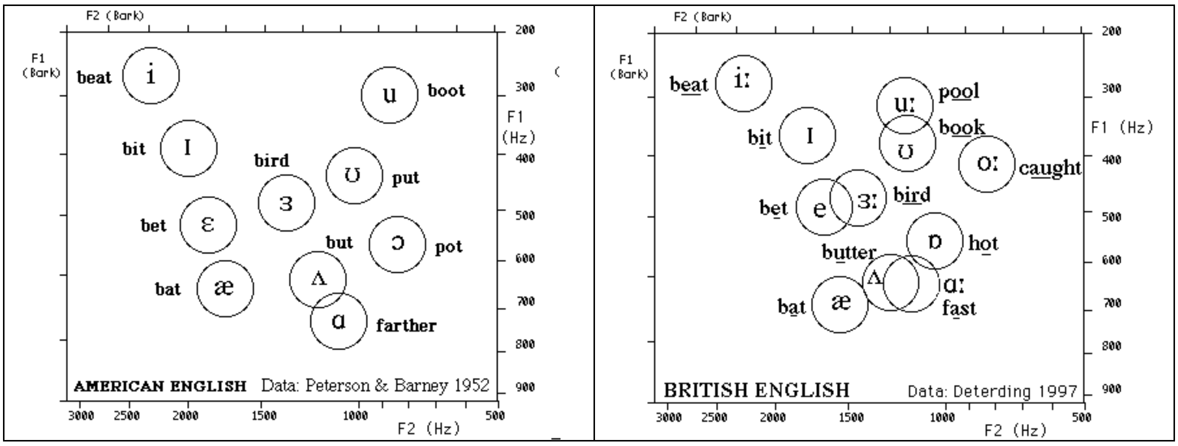

Figure 10.26: Comparison of American and British English (/e/ and /o/ omitted). Psychoacoustic version of the F1xF2 formant chart based on the Bark scale (figures provided by courtesy of Antti Iivonen[96]).

In order to control for some of the dialect-related problems, Hillenbrand, Getty, Clark & Wheeler (1995) replicated the Peterson & Barney study for vowel production and listening tests with 45 men, 48 women, and 46 ten- to 12-year-old children (27 boys, 19 girls) from the northern Midwest. The majority of their speakers were from the Michigan lower peninsula,

---

96 http://www.helsinki.fi/puhetieteet/projektit/vokaalikartat/kuvat/Ameng.gif (accessed 2013-03-03) http://www.helsinki.fi/puhetieteet/projektit/vokaalikartat/kuvat/british_vowels.gif (accessed 2013-03-03) Data from Deterding (1997) were obtained from a digital speech database from fluent speech (approximately 10 occurrences per vowel type). The American English data in the Peterson & Barney (1952) study are mean formant values of vowels in /hVd/ words realized by 33 American male speakers. The key words as presented here were not included in the original data.

the remainder from other areas in the upper Midwest (Illinois, Wisconsin, Minnesota, northern Ohio, northern Indiana). Formant contours for F1-F4, fundamental frequency and duration were measured. For reasons of comparability, the steady-state measurements from the Peterson & Barney's study were repeated, but additionally patterns of formant frequency change were investigated by measurements of formant patterns at 20% and 80% of vowel duration to show dynamic spectral changes in the vowel continuum and their role in vowel recognition.

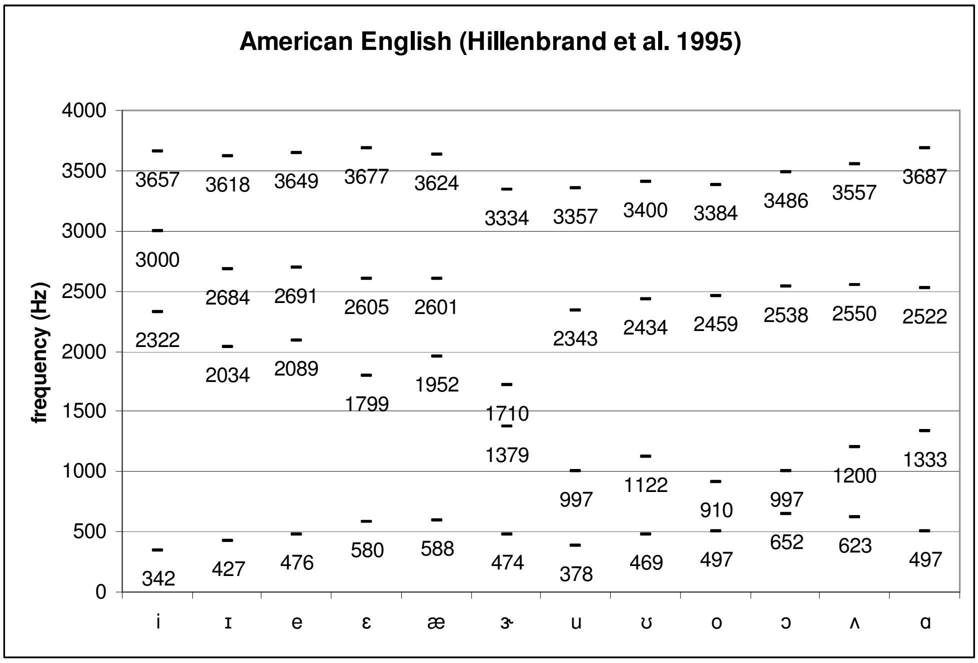

Figure 10.27: Average formant frequencies F1-F3 (in Hz) of American English vowels in the /hVd/ context produced by 45 men in the study of Hillenbrand et al. (1995). Formant values sampled at steady-state times for a sub-sample of tokens that were well identified in the listening study (tokens producing an error rate > 15% were excluded).

Hillenbrand et al.'s (1995) results (see Figure 10.27) show gross similarities with Peterson & Barney's (1952) data as well as some differences in the formant patterns, especially for /æ/ and /ɛ/. Moreover, they show a significant portion of dynamic spectral change for all vowel phonemes.

Table 10.22 summarizes the most important quantity differences for AE vowel pairs from Hillenbrand et al. (1995) in the /hVd/ context. Relative duration differences in AE help to differentiate spectrally similar front vowels but not back vowels (see the results of discriminant analyses by Strange et al. 2009). In vowel recognition, duration can play a central role for some vowel contrasts, e.g. /ɑ/ - /ɔ/ - /ʌ/ and /æ/ - /ɛ/, but appears of less importance for others, e,.g. /i/ - /ɪ/, /u/ - /ʊ/ and /ɪ/ - /e/ - /ɛ/ (Hillenbrand, Clark & Houde 2000; Hillenbrand 2003).

V contrast	long : short ratio
/i/ > /ɪ/	1.26
/e/ > /ɛ/	1.36
/æ/ > /ɛ/	1.39
/ɔ/ > /ɑ/	1.05
/o/ > /ʊ/	1.23
/ɑ/ > /ʌ/	1.35
/u/ > /ʊ/	1.23

Table 10.22: Pairs of spectrally similar American English vowels that differ in average duration. Ratio of the longer to the shorter vowel in /hVd/ syllables (Hillenbrand et al. 1995; Hillenbrand 2003: 125)

Hagiwara (1997) provides formant frequency data for speakers from southern California and found data that diverges from the findings of Hillenbrand et al. (1995) particularly with respect to the high back vowels /u/ and /ʊ/ that were fronted as is common for southern Californian speakers (see also Thomas 2001). Hillenbrand, Clark & Nearey (2001) investigated the effects of consonant environment on vowel formant patterns for AE vowels and found similar results for fronted high back vowels (see below). Ladefoged's (1999) phonetic description of American English is also based on a southern Californian dialect.

Clopper, Pisoni & de Jong (2005) compared the acoustic characteristics of the vowel systems of six regional varieties of North American English from New England, Mid-Atlantic, North, Midland, South and West. The data show consistent inter-speaker variation due to region of origin and gender. An effect of the speakers' region of origin was particularly evident with respect to the production of low vowels and high back vowels. For Northern speakers, shifted low vowels were observed that are consistent with the Northern Cities Chain Shift. Speakers from New England, the Midland and Western speakers produced the low back merger. Considerable gender-specific differences were found in the extent of the Northern Cities Chain Shift[97] and the magnitude of the back-vowel fronting.

Jacewicz, Fox & Salmons (2011) present a study on three regional varieties of American English (western North Carolina, central Ohio, southeastern Wisconsin, representing three major dialect zones), confirming a common set of systematic changes in all three dialect

[97] Later stages of the Northern Cities Chain Shift, particularly backing of /ɛ/ and /ʌ/ were found among females but not males. Significant effects of back vowel fronting of /o/ and /u/ was found for males, but only for /o/ with females. Across all dialect regions, /u/ fronting seems more advanced for females, while /o/ fronting remains predominantly a Southern phenomenon. For males, however, /u/ fronting is mostly a Southern and Western feature that has not spread to northern and eastern dialects. A merger or partial merger of /ɑ/ and /ɔ/ is found for New England, Mid-Atlantic, Midland, and Western dialects. Note however, that these findings refer only to F1/F2 comparisons and do not include dynamic spectral patterns.

areas: the gradual lowering of /ɪ ɛ æ/ across three age groups (from 20 to 65 years) and increased monophthongization, i.e. reduction of dynamic spectral changes in the "gliding" of these vowels, which is evident in formal and informal speech styles and is more evident with each younger generation of speakers, regardless of dialect-specific vowel dispersion patterns.

Several acoustic studies reveal considerable spectral overlap in the F1xF2 plain for adjacent English vowels especially for lower (back) vowels (e.g. Peterson & Barney 1952; Bernard & Mannell 1986). However, studies on English vowel recognition show a strong relationship between formant frequency movements and vowel identity, since spectrally adjacent AE vowels vary systematically in relative duration, and dynamic spectral patterns constitute clear differences between adjacent vowel qualities (Nearey & Assmann 1986; Flege, Munro & Fox 1994; Hillenbrand et al. 1995; Hillenbrand & Nearey 1999; Hillenbrand 2003). Spectral change in vowel recognition is considered to be a secondary but quite important cue to vowel identification (for review, see Nearey & Assmann 1986; Nearey 1989; Strange 1989b; Morrison & Assmann 2013). In other words, additionally to formant frequency patterns, the interaction of dynamic spectral information and duration is important for vowel recognition, specifically for the qualities involved in the so-called "tense-lax" oppositions in the English vowel system that are adjacent to each other in an F1/F2 representation. Even if intrinsic differences between "tense" and "lax" vowels can be considered as (phonologically redundant) secondary phonetic cues to vowel identity, English listeners seem to rely on both types of perceptual cues (Bohn & Flege 1990).

The use of these cues may however vary between different English varieties. Escudero & Boersma (2004) studied the /i/ - /ɪ/ contrast (*sheep – ship*) in Scottish English and Southern British English and found that there is a large difference in quality, especially in F1 values (treated as acoustic correlate of differences in vowel height), for Scottish and Southern English [ɪ]. While in most varieties of Scottish English /i/ and /ɪ/ are realized almost equally long, spectral differences in terms of F1 between the two vowel qualities are substantially larger than in Southern British English, where a large durational difference is found. In a vowel identification task with stimuli varying in duration and F1, native Scottish English listeners focus on quality differences almost exclusively, while Southern British English listeners identify the vowels by focusing on both quality and length, even if length differences are significantly higher in their variety (Escudero 2002; Escudero & Boersma 2004; see also Bohn & Flege 1990).

Coarticulatory patterns are another important perceptual cue in English vowel recognition. Particularly for coronal consonants strong coarticulatory effects on back vowels are observed.

A study by Hillenbrand, Clark & Nearey (2001) investigated the effects of consonant environment on vowel formant patterns for AE vowels. Their results show consistent acoustic effects of consonantal environment on vowel formant frequencies: (1) a general tendency towards "centralization" for vowels not produced in isolation, (2) large upward shifts in F2 for /u/ (500-600 Hz) and /ʊ/ (200-300 Hz) following coronals, (3) an upward shift in F2 for /ɑ/ and /ʌ/ (about 100 Hz) following coronals, (4) an upward shift in F2 (about 100 Hz) for back vowels following velars, (5) a downward shift in F2 (85-100 Hz) for front vowels following labials, and (6) a tendency toward lower F1 values for vowels in voiced environment. Despite this coarticulatory variation, the Hillenbrand et al.'s listening experiment revealed relative perceptual stability. A plausible account for these observations would be that (native) listeners internalize knowledge about the effects of context on vowel formants and invoke this in perception (Hillenbrand et al. 2001).

The fronting of high back vowels, as observed by Hillenbrand et al. (2001) has been described for several other varieties of English: for British English (Hawkins & Midgley 2005; Harrington et al. 2008), Australian English (Cox 1999; Cox & Palethorpe 2001), and for many varieties of North American English (Labov, Ash & Boberg 2006; Strange et al. 2005; Strange et al. 2007). Strange et al. (2005) find that AE mid to high back vowels are extremely fronted in alveolar contexts with little change in mid-low and low long vowels. Strange and colleagues posit the existence of fronted allophones for back rounded vowels in coronal contexts in rapid and continuous speech styles that are close to front rounded vowels, and suggest that in L2 acquisition of German or French front rounded vowels are considered to be acoustically similar to AE allophonic "fronted" back vowels and are therefore identified as back rather than front vowels by English listeners (e.g. Strange et al. 2005; Strange 2007).

10.5.3 Contrastive analysis

In his contrastive analysis of the English and the German vowel systems, Moulton (1974: 91ff) identifies the following major areas of difficulties in German pronunciation for (American) English learners:

Front rounded /y: ʏ ø: œ/, which represent a combination of the features [front] and [round] that never co-occur in English, constitute a major problem in L2 acquisition of German. /ø:/ and /œ/ are moreover vaguely similar to AE /ɜ/ for which allophonic lip rounding is observed, causing substitutions of /ø:/ and /œ/ by /ɜ/ in L2 German pronunciation (Moulton 1974: 92).

German /i: e: u: o:/ cause considerable problems in pronunciation to English learners of German. Before voiceless consonants, English natives would use the English allophones /i u/

and (perhaps) /e o/, but they would be too open and too short compared to the German equivalents (Moulton 1974: 92). For /e: o:/ the English diphthongized allophones are more likely to be used, especially before voiced consonants and in word-final position. /ɪ ɛ ʊ/ would represent only minor difficulties. Comparisons of mid vowels /e/ and /o/ of (North) German and AE show that the German vowels are higher and closer to NG high vowels than their AE equivalents (Strange et al. 2004).
The opposition of the German long and short a-vowels are a problem to English learners of German, though their actual realization depends very much on the learners' regional variety.
German /ɔ/ is described as difficult for almost all English learners. There are a few English varieties that distinguish "tense" and "lax" vowels in the mid back vowels (/hol/ *hole* vs. /hɔl/ *whole*); for speakers of these varieties the realization of the German opposition would be easy, whereas a majority of Americans have no /ɔ/ (lax mid back [o̜] in Moulton's transcription (1974: 92)) but only /ɑ/ or /ɒ/ (low back [ɔ] as in *cost, fall, dawn* in Moulton's transcription) that is markedly lower and longer in German words like *Kost, voll, Bonn*.
American English speakers of some varieties have no /ɑ/ - /ɔ/ opposition at all and use a single phoneme /ɒ/ instead, giving a series of homophone words (e.g. *caught, caller, paw* vs. *cot, collar, pa*). These speakers my have difficulties with the long-short opposition of a-vowels but also with the distinction of a-vowels from /ɔ/, resulting in a homophoneous pronunciation of German *Kamm, kam, komm* - 'comb, came, come!' (Moulton 1974: 93).
Several experimental studies dealt with the perception of German vowels by English native speakers (e.g. Bohn & Flege 1990, 1992; Bohn 1994; Strange, Bohn, Nishi & Trent 2005; Strange, Levy & Law II 2009) comparing the listeners' responses with acoustic properties of vowel qualities under investigation. Many of these studies are settled within the Perceptual Assimilation paradigm, where listeners are asked to identify a German input vowel with one of the offered categories of his native language and to judge the goodness of fit. An example for results of this type of study is presented in Table 10.23 (cf. Strange et al. 2004) where the two most frequently selected AE vowels are presented for each of the North German input vowels together with the percentage of cases where the respective categorization was performed in citation form and in sentence contexts.
Perceptual studies focusing on the perception of German (or French) front rounded vowels presented in one or more phonetic contexts have shown persistent perceptual difficulties for AE listeners for some front rounded vs. back rounded vowels and for some contrasts among front vowels (e.g. Flege 1987; Polka 1995; Best et al. 2003; Strange, Bohn, Nishi & Trent 2004, 2005; Levy 2009; Strange, Levy & Law II 2009).

Acoustically, German front rounded vowels were found to be "intermediate" between front and back AE vowels (see Strange et al. 2004; Strange 2007). In perception however, AE listeners categorize German (and French) front rounded vowels routinely as perceptually more similar to back native categories, even if the stimuli were considered to be poor exemplars of back AE vowels. In discrimination tasks, AE listeners have more difficulty to discriminate front rounded vowels from back than from front vowels (Strange et al. 2004; Strange et al. 2005; Strange et al. 2007).

In a perceptual assimilation task with North German vowel stimuli, Strange et al. (2004) found that the four front rounded vowels were not consistently assimilated to any one category, though all four German vowels /y: ø: ʏ œ/ were considered more similar to AE back vowels /u: ʊ ʌ/ (though judged as very poor tokens of these categories) than to front ones (see Table 10.23). For /y:/, the second most frequently selected response category was AE /i:/ in citation-style syllables but /ʊ/ in sentence contexts. German /œ/ was assimilated most often to AE /ʌ/ and /ɛ/. German /i: ɪ ɑ:/, which are acoustically very similar to their AE counterparts in terms of both spectral and durational cues were very consistently assimilated to their AE equivalents (Table 10.23). German /u:/ was a bit less consistently identified with AE /u:/. German a-vowels were assimilated as moderately to very good exemplars of three spectrally overlapping AE categories /ɑ: ɔ: ʌ/. While German /o:/ was assimilated to its AE counterpart /o(ʊ)/, /e:/ was not identified as most similar to its AE counterpart but rather assimilated as a fairly good exemplar to /i:/ (and occasionally to /ɪ/). /ɛ/ was very consistently accepted as an equivalent to AE /ɛ/, while /ɔ/ was considered to be a moderately good exemplar to AE /ɑ: ~ ɔ:/ (both categories collapsed[98]).

In the citation form syllables, German long categories were assimilated to long AE categories in 88% of the cases (range 46% - 99% for 7 vowels), whereas short vowels were assimilated to short AE vowels only 62% of the time (range 9% - 99% for 7 vowels), indicating that /ø: ʊ ɔ ɑ/ were assimilated more often to the spectrally closest AE vowel. Strange et al. (2004) conclude that when cross-language spectral similarity is in conflict with temporal similarity, spectral similarity seems to be the stronger dimension for AE listeners.

A detailed context-sensitive acoustic analyses and comparisons of AE and (North) German (e.g. Bohn & Flege 1990; Strange 2007; Strange et al. 2005; Strange et al. 2009) show that AE front vowels vary only little across prosodic or phonetic contexts, whereas AE back

98 The response categories /ɑ:/ and /ɔ:/ (a contrast that is partially or completely neutralized in many regional varieties) were collapsed in Strange et al.'s study, since many participants found it difficult to differentiate these vowels and to use the appropriate key words.

vowels vary extensively in the front-back dimension as a function of consonantal context. Acoustically, (North) German front rounded vowels are sufficiently deviant from any of the AE front vowel tokens, but show a high degree of similarity to some allophones of AE back vowels (Strange et al. 2005; Strange et al. 2007).

	most frequent category ($)		**2nd most frequent category ($)**		**most frequent category**		**2nd most frequent category**	
NG vowel	AE vowel (syllables)	% ($)	AE (syllables)	% ($)	AE vowel (sentence)	%	AE (sentence)	%
y:	u:	69	i:	24	u:	87	ʊ	7
ø:	ʊ	37	u:	30	u:	43	ʊ	28
ʏ	ʊ	56	ʌ	20	ʊ	56	ʌ	24
œ	ʌ	62	ɛ	30	ʌ	80	ɛ	10
e:	i:	66	eɪ	23	eɪ	53	i:	30
o:	oʊ	89	ɔ:	5	oʊ	81	u:	6
ɛ	ɛ	97	ɪ	1	ɛ	93	i:	1
ɔ	ɑ:-ɔ:	90	ʌ	8	ɑ:-ɔ:	64	ʌ	23
ʊ	oʊ	42	ɑ:	22	oʊ	67	ʊ	10
i:	i:	97	ɪ	1	i:	91	u:	2
u:	u:	86	oʊ	6	u:	72	ʊ	11
ɪ	ɪ	97	i:	1	ɪ	87	ɛ	3
ɑ:	ɑ:-ɔ:	98	ʌ	1	ɑ: ɔ:	93	oʊ	1
a	ɑ:-ɔ:	79	ʌ	12	ɑ: ɔ:	74	ʌ	17

Table 10.23: Perceptual Assimilation of North German vowels in citation-form syllables and in sentence context (Strange et al. 2004, Tab. III and Tab. VII ibid.)

A context-sensitive analysis of perceptual assimilation patterns by AE listeners by Strange et al. (2005) and Strange et al. (2009) shows how experience with phonetic variation of phonological categories in AE listeners' native language determines the learners' perception and production in German or French.

Confusions between high and mid German vowels can also be expected to be due to the amount to which duration is used to differentiate spectrally-adjacent vowel categories, especially for /i:/ - /e:/, /y:/ - /ø:/ and /u:/ - /o:/, which do not differ in duration.

Finally, difficulties in the perceptual identification of German vowels are expected to be not only caused by the differences between the German and the English vowel systems but also to a certain extent by large differences in orthographic representations between the two languages and by inconsistencies within the English as well as the German system for several vowels, e.g. for English /i:/ represented as English <ee> or <ea> vs. German <i> or <ie>.

Therefore, it is expected that orthographic differences and inconsistencies may cause transfer and "spelling perceptions"[99].

The results of the current study will present more data on patterns of category confusion, but will focus on confusion patterns *within* the German system instead of using the perceptual assimilation paradigm that has been used in other studies, since the perceptual assimilation paradigm is not considered to be appropriate for the description of vowel perception by more advanced learners who already have developed perceptual categories for L2. Previous studies only dealt with North German vowel realizations, while the present study uses stimuli of a Southern German variety.

10.5.4 Results and discussion

15 native English subjects were tested, seven beginners and eight advanced learners. All subjects were recruited while attending a language course in Vienna.

Six participants were US citizens from Arizona, California, Florida, Iowa, Massachusetts, and Ohio. Four persons were from Britain (Windsor, Birmingham, Bedford, London) and two from Australia (Sydney and North-West Australia). Five persons indicated Vienna as their permanent place of residence at the time the experiment took place. Six persons (two from Britain, four from the US) indicated to have a bilingual family background but were raised and literalized in English. Four persons indicated to be born in Hongkong, Singapur, Vietnam or Malaysia, but declared to be native speakers of English and were therefore included in the sample.

Due to the small size of the regional sub-samples, a solid comparison of perceptual differences between speakers of American English, British English or Australian English cannot be provided. Therefore, the data available will be discussed for all English natives together to provide insights into major areas of difficulty and category confusion and perceptual similarity for L1 English learners of German and to enable a comparison with language-specific patterns from other language sub-groups studied here. Unlike many previous studies that focussed only on a few vowel contrasts, e.g. front vowels, the present study investigates the full German vowel system, using an experimental design that shows the similarity relationships between vowels of the target language vowel system as a whole.

---

[99] in analogy to spelling pronunciations.

10.5.4.1 Difficulties

The percentage of wrong and correct identifications by L1 English natives for each vowel category is represented in Figure 10.28.

In total, 48.2 % of the valid reponses were correct, and 51.8% were wrong. The highest percentage of wrong responses is observed for /ø:/ (only 28% correct), followed by long constricted /u:/ (35%), short unconstricted /ʏ/ (37% correct) and /ʊ/ (40% correct). Least difficulties are observed for /ɑ/ (73% correct) and /ɛ/ (72% correct), followed by /ɔ/ (67%), /i:/ (64%), /ɑ:/ (64%), and /ɪ/ (60% correct). With the only exception of /i:/ - /ɪ/ and /y:/ - /ʏ/, short unconstricted qualities show better id_scores than their long constricted counterparts.

The listeners' mean id_wrong score was 47% (12% standard deviation), the median 46%. The listener with the highest percentage of correct answers had an id_wrong of 26%, the highest id_wrong score was 70% wrong responses.

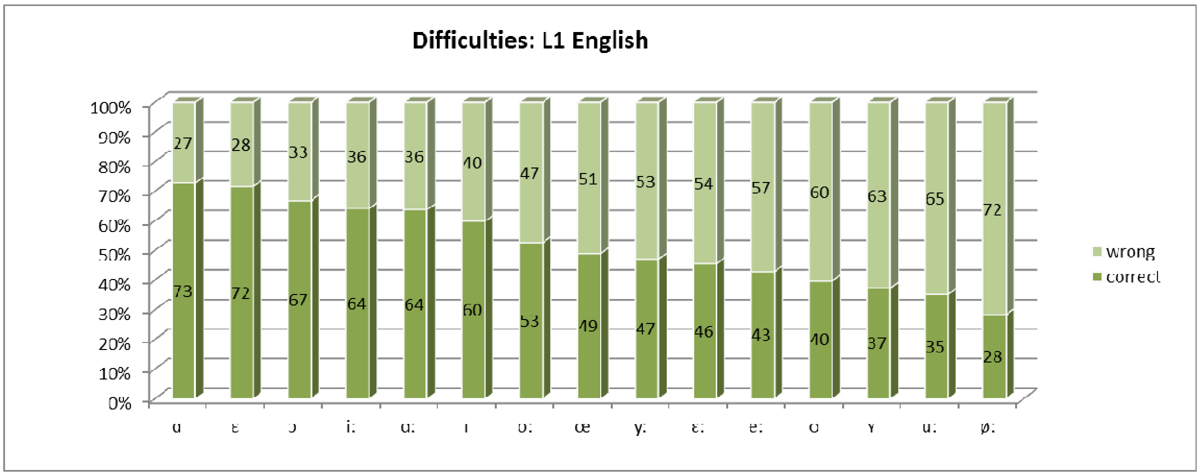

Figure 10.28: Difficulties for L1 English listeners (N= 14, 4049 responses)

The boxplot diagram in Figure 10.29 compares the id_wrong scores of all L1 English subjects and reveals interesting differences between categories. While /ɑ ɑ: ɛ ɪ i: ɔ o: œ/ are quite well differentiated from other categories by many subjects, there are other categories that cause severe difficulties in correct identification to most participants: For /ø:/, /ʏ/, /e:/ and /ɛ:/, most difficulties are observed as indicated by an id_correct score below 50%. Moreover, /u:/ and /ʊ/ cause considerable difficulties to many listeners. Front rounded vowels generally cause more problems than other categories. Most difficulties are observed for /u:/, /ø:/ and /ʏ/, whereas /y:/ and /œ/ seem to be better discriminated at least by some subjects.

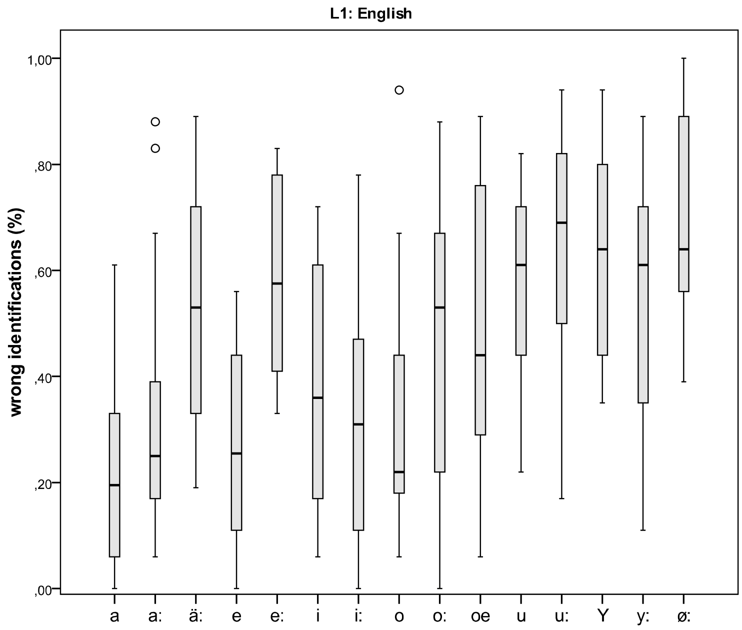

Figure 10.29: Wrong identifications by native English listeners

10.5.4.2 Preferences

The preference scores of L1 English listeners for monophthong categories (see Figure 10.30) vary within a moderate range of only 3.5%. The most preferred category is /ɛ/ (8.2%), the least preferred monophthong category is /ø:/ (4.7%). The short unconstricted categories /ɛ ʊ ɑ ʏ ɔ œ/ are more frequently selected than their long constricted counterparts. Only for /ɪ/ - /i:/ similar pref_scores are observed for both members of the contrast.

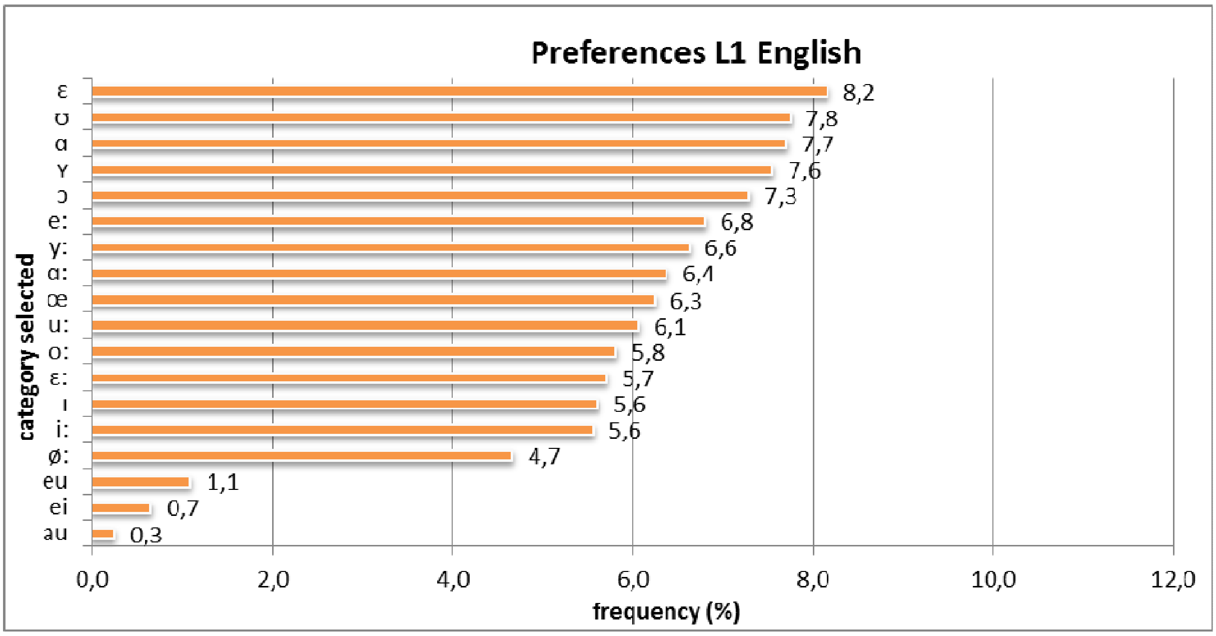

Figure 10.30: Preferences for L1 English listeners

Compared to most other language sub-groups, L1 English listeners select diphthongs more frequently. The most preferred diphthongal category is <eu> (1.1%), followed by <ei> (0.7%) and <au> (0.3%) (for a detailed discussion, see below). Interestingly, diphthongs are selected for long as well as short input stimuli, most frequently for /i:/, /ø:/, /œ/ and /ʏ/.

10.5.4.3 Patterns of confusion

Patterns of category confusion for English learners of German are represented in Table 10.24. A first look at the confusion matrix shows two major areas of difficulty in the perceptual vowel space: (1) confusion between *front non-rounded* e- and i-vowels, and (2) a high amount of confusion between *all rounded vowel categories*. Moreover, some problems with the differentiation of a- and e-vowels and a fairly high amount of diphthongization are observed for some types of vowels.

%	ɑ	ɑ:	ɛ:	ɛ	e:	ɪ	i:	ɔ	o:	ʊ	u:	œ	ø:	ʏ	y:	‹ei›	‹eu›	‹au›	wrong	correct
a	**73**	22	2	1						2	1								27	**73**
a:	34	**64**	1															1	36	**64**
ɛ:	2	6	**46**	12	30	1	1									2			54	**46**
ɛ	3	2	6	**72**	13	2	1									1			28	**72**
e:	1	2	30	15	**43**	3	5									1			57	**43**
ɪ			1	17	6	**60**	12			1						2			40	**60**
i:				4	10	17	**64**									4			36	**64**
ɔ	1							**67**	18	3	2	4	1	2			1		33	**67**
o:								20	**53**	3	4	6	7	4	1		1	2	47	**53**
ʊ								19	7	**40**	12	4	4	10	4				60	**40**
u:								2	6	22	**35**	1	3	15	14		2		65	**35**
œ				1				1	2	10	4	**49**	15	9	3		4		51	**49**
ø:				1					1	9	7	19	**28**	12	19		4		72	**28**
ʏ						1		1		22	13	8	5	**37**	10		3		63	**37**
y:									1	4	13	4	6	24	**47**		2		53	**47**
Σ	*8*	*6*	*6*	*8*	*7*	*6*	*6*	*7*	*6*	*8*	*6*	*6*	*5*	*8*	*7*	*.7*	*1*	*.3*	*48.2*	***51.8***

Table 10.24: Confusion matrix for L1 English listeners (N=15, 3920 responses). Wrong and correct (in bold) identifications in % (presented stimuli in rows, selected categories in columns in semi-orthographic representation, values < 0.5% not indicated, values < 5.5% in grey)

For a-vowels, most difficulties concern the differentiation of long vs. short vowels. /ɑ/ is identified correctly in 73% of its occurrences and is substituted by /ɑ:/ in 22%. Long /ɑ:/ is identified correctly as /ɑ:/ in 64% and in 34% incorrectly as /ɑ/. Some native English listeners seem to have problems to differentiate a-vowels from e-qualities (mainly ɛ-qualities) and u-vowels, which might be due to orthographic interference. Confusion of short /ɑ/ with u-vowels is most probably due to English orthographical conventions, where short /ʌ/ is represented as <u> (e.g. *but*). Instances of a-vowels that are categorized as e-qualities (as well as categorizations of e-qualities as a-vowels) can be explained by the interference of English /æ/ and its orthographic representation <a>.

Considerable inter-category confusion is observed for e- and i-vowels which could also at least partially be influenced by orthography: Within the class of front non-rounded vowels, the highest id_correct score is obtained for short /ɛ/ (72%), followed by /i:/ (64%) and /ɪ/ (60%). Long /e:/ (43% correct) and /ɛ:/ (46% correct) cause considerably more problems. For all e-qualities, identifications with a-vowels are observed, but they occur more frequently for unconstricted /ɛ:/ and /ɛ/ than for /e:/. Moreover, this id_pattern seems to be asymmetric in direction. Identifications of i-vowels with /e:/ or /ɛ/ but not /ɛ:/ (<ä>) are quite frequently observed, while substitutions in the opposite direction (e- qualities classified as i-vowels) are much less common. For all e- and i-vowels, diphthongs are selected in some cases, especially for /i:/ (11 cases, 4%) and /ɛ:/ (6 cases) but less often for /ɛ e: ɪ/ (2, 3, and 4 cases, respectively). Contrary to many other languages, /i:/ does not seem to act as a strong attractor for /e:/-stimuli: /e:/ serves as substitute for /i:/-stimuli in 10%, whereas /i:/ is a substitute for /e:/ in only 5% of the cases. Probably, this is again due to interference of English orthography, where [i:] can be represented as <ee> (e.g. English *see* /si:/).

Long constricted /i:/ (64% id_correct) is identified as /ɪ/ (17%), /e:/ (10%), and less frequently as /ɛ/ (4%) or <ei> (4%). The relative high percentage of /i:/ > <ei> is an interesting pattern that could possibly be explained by the interference of orthographic differences between English and German.

Short unconstricted /ɪ/ (60% id_correct) is substituted more frequently by short unconstricted /ɛ/ (17%) than by /i:/ (12%). /ɪ/ is moreover substituted by /e:/ (6%) and in 2 cases as <ä:>. In 2% of the cases, /ɪ/-stimuli were categorized as <ei>.

Within the class of front non-rounded vowels, /e:/ has the lowest id_correct score (only 43% correct), being most frequently substituted by /ɛ:/ (30%) and /ɛ/ (15%). Other candidates for perceptual substitution are /i:/ (5%), /ɪ/ (3%), /ɑ:/ (2%), <ei> (1%), and /ɑ/ (1%).

Long unconstricted /ɛ:/ (only 46% correct) is most frequently identified as /e:/ (30%), followed by /ɛ/ (12%), /ɑ:/ (6%), /ɑ/ (2%), <ei> (2%, 6 cases) and occasionally by i-qualities (1% each). Interestingly, the confusion of /e:/ and /ɛ:/ is symmetrical, i.e. one category does not attract substantially more responses than the other.

Short unconstricted /ɛ/ has a considerably higher id_correct score (72%) than the other e-qualities. It is mostly confused with long constricted /e:/ (13%) or <ä:> (6%), occasionally also with a-vowels (/ɑ/ 3% and /ɑ:/ 2%, probably due to interference of English /æ/), or i-vowels and <ei> (only two cases).

As in many previous studies, considerable inter-category confusion is observed for all rounded vowels. While the contrast of rounded vs. non-rounded vowels is clearly differentiated by native English listeners, within the class of rounded vowels front rounded and back rounded vowels many perceptual difficulties are observed. Least difficulties are observed for o-qualities (/ɔ/ 67% and /o:/ 53% id_correct), most problems occur with /ø:/ (27% correct), /u:/ (35% correct), and /ʏ/ (37% correct). For all rounded categories except /ʊ/ a few substitutions with <eu> are observed. <au> is selected in 6 cases (2%) for /o:/ and in one case each for /ɔ/ and /ʊ/.

o-qualities show the best results within the class of rounded vowels.

/ɔ/ is identified correctly in 67% of its occurrences and most frequently substituted by /o:/ (18%); all other rounded categories are only occasionally selected: /œ/ (4%), /ʊ/ (3%), /ʏ/ (2%), /u:/ (2%), /ø:/ (1%); an identification with /ɑ/ occurs only in three cases, probably due to the merger of /ɑ/ and /ɔ/ in some regional varieties; /e:/ and /y:/ are selected in only one case each.

Long constricted /o:/ is identified correctly in 53% of its occurrences and is incorrectly categorized with /ɔ/ (20%), but also with /ø:/ (7%), /œ/ (6%), /ʏ/ (4%), /u:/ (4%), /ʊ/ (4%), and occasionally with <au> (6 cases) and <eu> (2 cases).

Short unconstricted /ʊ/ (40% correct) is substituted with /ɔ/ (19%), /u:/ (12%), /ʏ/ (10%), or /o:/ (7%), but also with /œ/, /ø:/ and /y:/ (each 4%).

For /u:/ even more difficulties than for /ʊ/ are observed (only 35% correct), but the responses are a little less widespread. The most preferred substitutes for /u:/ are /ʊ/ (22%), /ʏ/ (15%), and /y:/ (14%). Other rounded categories are much less preferred: /o:/ (6%), /ø:/ (3%), /ɔ/ (2%), <eu> (2%), and /œ/ (1%).

/ø:/ is the category that causes most problems to native English listeners (only 28% correct), though the responses are not so equally dispersed over rounded categories as for most of the other rounded vowels. The most preferred response options for /ø:/-stimuli are /y:/ and /œ/ (19% each), followed by /ʏ/ (12%), /ʊ/ (9%), and /u:/ (7%). <eu> is selected in 10 cases (4%), /ɛ/ and /o:/ are also occasionally selected.

Most wide-spread are the responses for /œ/, though /œ/ has the highest id_correct score (49% correct) of all front rounded vowels: Incorrect responses for /œ/ are caused mainly by substitutions with /ø:/ (15%), /ʊ/ (10%), and /ʏ/ (9%), but also with /u:/ (4%), <eu> (4%), /y:/ (3%), /o:/ (2%), /ɔ/ (1%), and /ɛ/ (3 cases, 1%).

For /y:/, an id_correct score of 47% is obtained. Its most frequently selected substitutes are /ʏ/ (24%) and /u:/ (13%), followed by /ø:/ (6%), /œ/ (4%), /ʊ/ (4%), <eu> (2%), and /o:/ (1%, 3 cases).

Short unconstricted /ʏ/ (37% id_correct) is very frequently substituted with /ʊ/ (22%), but also with /u:/ (13%), /y:/ (10%), and /œ/ (8%). Less frequently selected categories for /ʏ/ are /ø:/ (5%), <eu> (3%), and /ɪ/ and /ɔ/ (1% each).

To summarize, the most significant area of difficulties for native English listeners is the class of rounded vowels and their differentiation with respect to the correct identification of the constriction location in the front-back dimension and their exact quality and quantity. Responses for rounded vowel stimuli are slightly more attracted by the short unconstricted qualities /ʊ ʏ ɔ/, as is reflected by their higher pref_scores (/ʊ/ 7.8%, /ʏ/ 7.6%, /ɔ/ 7.2%). The high amount of inter-category confusion between rounded vowels attenuates the patterns of perceptual asymmetry that is found in many other languages. Only for the /y:/ - /ø:/ contrast, a high attraction of /y:/ is observed (/ø:/ > /y:/ 19% vs. /y:/ > /ø:/ only 6%). For other categories, no strong asymmetries can be observed.

Several acoustic studies on English vowels have reported a high amount of coarticulatory variation for high back vowels with extreme fronting of u-vowels especially in positions after coronals (see section 10.6.3). Strange and colleagues assume that native AE listeners may consider ü-stimuli to be allophonic variants of u-vowels, which would explain the high amount of inter-category confusion in the data of the present study, where German u- and ü-qualities are found to have a high attraction effect for each other, though an asymmetric pattern does not become straightly apparent. To prove whether a perceptual fronting can be observed in the context of coronals /d t z/, a comparison of the responses for /u:/ and /ʊ/-stimuli in varying pre-vocalic context is represented in Table 10.25 and Table 10.26 (responses different from ü-vowels printed in grey).

The responses show that ü-vowels (/y:/ or /ʏ/) are more preferred for /u:/-stimuli than for /ʊ/-stimuli. In total, /u:/-stimuli were substituted with front ü-qualities in 50 cases, whereas for /ʊ/-stimuli only 28 cases of substitutions with front ü-qualities are observed. In coronal contexts, a higher number of perceptual substitutions of /u:/ with ü-categories is only observed for /t_/ and /z_/ but not for /d_/. For /ʊ/-stimuli, a slightly higher number of substitutions with ü-qualities is observed in /t_/ and /z_/ contexts is observed, though they are generally less numerous for /ʊ/ (only 28 cases). The consonantal context is probably not the only factor influencing the choice of response categories. Rather, we find inter-subject differences: some subjects seem to prefer ü-categories (e.g. E13[100] and E15[101]), while others never select ü-categories for u-stimuli (e.g. E12[102]). These differences in selection behaviour do not seem to depend on the learners' country of origin nor on their proficiency in L2 German, though the influence of consonantal context has to be tested on a larger sample with more controlled stimuli.

/u:/	*C_*	E1	E2	E3	E4	E5	E6	E7	E8	E9	E10	E11	E12	E13	E14	E15	*ü-resp*	*% ü*
pu:te	*p*	y:	u:	u:	ʊ	ʊ	ʊ	ʏ	ø:	u:	y:	u:	ʊ	ʏ	ʊ	ʏ	*4*	8,0
bu:te	*b*	eʊ	u:	u:	ʊ	ʊ	œ	u:	-	u:	ø:	u:	ʊ	y:	ʏ	ʏ	*3*	6,0
fu:te	*f*	ʊ	u:	ʊ	u:	u:	-	u:	o:	u:	u:	ʏ	ʊ	y:	ʏ	ʏ	*4*	8,0
wu:te	*w*	o	u:	y:	ʏ	u:	ʏ	ʊ	y:	ʏ	ʊ	ʊ	ʊ	ʊ	-	ʏ	*6*	6,0
tu:te	***t***	ʏ	ʊ	ʊ	ʊ	ʊ	ʏ	ʏ	y:	u:	y:	ʏ	u:	y:	ʏ	ʏ	***9***	18,0
du:te	***d***	y:	u:	ʊ	u:	u:	u:	u:	u:	y:	u:	ö:	ʊ	u:	ʊ	eʊ	***2***	4,0
su:te	***z***	y:	o:	y:	ʊ	ʏ	-	y:	y:	y:	y:	u:	u:	y:	u:	ʏ	***9***	18,0
ju:te	*j*	u:	y:	u:	eʊ	u:	-	ʏ	u:	u:	ʊ	ʊ	ʊ	ʏ	ʏ	ʊ	*4*	8,0
ku:te	*k*	ʏ	u:	u:	ʏ	ʊ	y:	u:	o:	u:	y:	u:	ʊ	u:	ʏ	ø:	*5*	10,0
gu:te	*g*	ʏ	u:	u:	ʊ	u:	œ	ʏ	y:	ʏ	u:	ʊ	ʊ	ʊ	ʊ	o:	*4*	8,0

Table 10.25: Responses for /u:/-stimuli in varying pre-vocalic context

/ʊ/	*C_*	E1	E2	E3	E4	E5	E6	E7	E8	E9	E10	E11	E12	E13	E14	E15	*ü-resp*	*% ü*
pʊtte	*p*	ʊ	ʊ	ʊ	ʊ	ʊ	ʊ	ʊ	u:	ʊ	u:	u:	ʊ	ɔ	ʊ	ʏ	*1*	3,6
bʊtte	*b*	u:	ʊ	u:	ʊ	ʊ	ʊ	ʊ	-	ʊ	u:	ʊ	ʊ	ʊ	ʊ	u:	*0*	0,0
fʊtte	*f*	u:	ʊ	ø:	ʊ	ɔ	ʊ	ʊ	o:	ʊ	ʊ	ʏ	ɔ	ʏ	ʊ	ʊ	*2*	7,1
wʊtte	*w*	ʏ	ʏ	ʊ	ʊ	ɔ	œ	œ	ɔ	ʏ	ʊ	ʊ	o:	ʏ	ɔ	u:	*4*	14,3
tʊtte	***t***	u:	u:	y:	ʊ	ʊ	ʏ	ʏ	u:	ʊ	y:	ɔ	u:	ʏ	ʏ	ʊ	***6***	21,4
dʊtte	***d***	ʊ	ʊ	o:	y:	ʊ	ɔ	ø:	ʊ	ʊ	ʊ	ʊ	ɔ	ɔ	ɔ	œ	***1***	3,6
sʊtte	***z***	u:	ʏ	ø:	ʏ	ɔ	ʏ	ø:	y:	ɔ	ʊ	ɔ	ʊ	ɔ	ɔ	ʏ	***5***	17,9
jʊtte	*j*	y:	y:	œ	ʊ	ʊ	-	ʊ	ʊ	ʏ	ø:	ʊ	ʊ	ʊ	ʊ	ʊ	*3*	10,7
kʊtte	*k*	ʏ	ʊ	-	u:	ʊ	ʊ	ʊ	ø:	ʊ	ʏ	ɔ	ʊ	ʏ	ʏ	u:	*4*	14,3
gʊtte	*g*	y:	u:	u:	ʊ	ʊ	-	ʊ	y:	u:	u:	ʊ	ʊ	œ	ʊ	ʊ	*2*	7,1

Table 10.26: Responses for /ʊ/-stimuli in varying pre-vocalic context

10.5.4.4 Similarity and distance

The similarity scores in Table 10.27 reflect the perceptual distance for German vowels as perceived by L1 English learners. The highest sim_score is observed for /e:/ - /ɛ:/ (0.673). All other contrast pairs with higher sim_scores above 0.2 are either pairs differing in phonemic length & constriction degree such as /ʊ/ - /u:/ (0.451), /œ/ - /ø:/ (0.437) /ʏ/ - /y:/ (0.396), /ɔ/ - /o:/ (0.314), /ɛ/ - /e:/ (0.241), /ɪ/ - /i:/ (0.238), or only differing in phonemic length like /ɑ/ - /ɑ:/ (0.408). Higher sim_scores are moreover observed for contrasts involving long /u:/ (/u:/ - /y:/ (0.331), /u:/ - /ʏ/ (0.376), and /u:/ - /ø:/ (0.162)), and for other vowel pairs that are

100 British English background.

101 American English background.

102 American English background.

phonetically near in terms of quality or quantity: /y:/ - /ø:/ (0.326), /ʏ/ - /ø:/ (0.265), /ʊ/ - /ɔ/ (0.210), /œ/ - /ʏ/ (0.193), /ʊ/ - /ø:/ (0.183), /œ/ - /ʊ/ (0.162), /ɪ/ - /ɛ/ (0.145), /i:/ - /e:/ (0.133), /u:/ - /o:/ (0.110), /ʊ/ - /o:/ (0.103).

V	ɑ	ɑ:	ɛ:	ɛ	e:	ɪ	i:	ɔ	o:	ʊ	u:	œ	ø:	ʏ	y:
a	1.000														
ɑ:	0.408	1.000													
ɛ:	0.032	0.063	1.000												
ɛ	0.029	0.014	0.150	1.000											
e:	0.007	0.022	0.673	0.241	1.000										
ɪ	0.003	0.000	0.018	0.145	0.089	1.000									
i:	0.000	0.000	0.010	0.034	0.133	0.238	1.000								
ɔ	0.008	0.000	0.000	0.000	0.003	0.000	0.000	1.000							
o:	0.000	0.000	0.000	0.000	0.000	0.003	0.000	0.314	1.000						
ʊ	0.017	0.000	0.004	0.003	0.000	0.008	0.004	0.210	0.103	1.000					
u:	0.010	0.000	0.000	0.004	0.005	0.000	0.000	0.030	0.110	0.451	1.000				
œ	0.000	0.000	0.004	0.010	0.004	0.000	0.003	0.046	0.080	0.162	0.065	1.000			
ø:	0.000	0.000	0.005	0.011	0.005	0.004	0.004	0.008	0.090	0.183	0.162	0.437	1.000		
ʏ	0.000	0.000	0.000	0.000	0.000	0.008	0.004	0.029	0.047	0.414	0.376	0.193	0.265	1.000	
y:	0.000	0.003	0.000	0.003	0.000	0.000	0.000	0.003	0.023	0.092	0.331	0.068	0.326	0.396	1.000

Table 10.27: Similarity matrix for L1 English listeners

Rounding is a highly distinctive feature for native English listeners. Contrasts of a rounded vs. non-rounded vowels are rarely confused and receive very low sim_scores all below the 0.02 mark (see below), while high sim_scores are obtained for contrasts of front rounded vs. back rounded vowels (see Table 10.28). The highest sim_score for a round/non-rounded contrast is obtained for /ɑ/ - /ʊ/ (0.017) to which English orthographical conventions may contribute.

round V contrasts	*front (f), back (b), front-back (fb)*	*sim_scores*
ʊ - u:	b	0.451
œ - ø:	f	0.437
ʊ - ʏ	fb	0.414
ʏ - y:	f	0.396
u: - ʏ	fb	0.376
u: - y:	fb	0.331
ø: - y:	f	0.326
ɔ - o:	b	0.314
ʏ - ø:	f	0.265
ʊ - ɔ	b	0.210
œ - ʏ	f	0.193
ʊ - ø:	fb	0.183
œ - ʊ	fb	0.162
u: - ø:	fb	0.162
u: - o:	b	0.110
ʊ - o:	b	0.103

Table 10.28: Sim_scores > 0.1 for [+round] categories

For contrasts of non-rounded vowels, the highest sim_scores are observed for /e:/ - /ɛ:/ (0.673) and /ɑ/ - /ɑ:/ (0.408). Moreover, considerable confusion is observed for contrasts of i-and e-vowels that is reflected by high sim_scores (see Table 10.29). Contrasts of a- and e-qualities, for which confusion might be induced by orthographic interference, show values below 0.1.

non-rounded V contrasts	*front-front (f) front-back (fb)*	*sim_scores*
eː - ɛː	f	0.673
ɑ - ɑː	b	0.408
eː - ɛ	f	0.241
iː - ɪ	f	0.238
ɛ - ɛː	f	0.150
ɪ - ɛ	f	0.145
iː - eː	f	0.133
ɪ - eː	f	0.089
ɛː - ɑː	fb	0.063
iː - ɛ	f	0.034
ɛː - ɑ	fb	0.032
ɛ - ɑ	fb	0.029
eː - ɑː	fb	0.022
ɪ - ɛː	f	0.018
ɛ - ɑː	fb	0.014
iː - ɛː	f	0.010
eː - ɑ	fb	0.007
ɪ - ɑ	fb	0.003

Table 10.29: Sim_scores for [– round] categories

Within the class of non-rounded vowels (see Table 10.29), the highest sim_scores are observed for /eː/ - /ɛː/ (0.673), followed by /ɑ/ - /ɑː/ (0.408). Sim_scores above 0.1 are observed for contrasts between front non-rounded vowels: /iː/ - /ɪ/ (0.238), /ɛ/ - /ɛː/ (0.150), /ɪ/ - /ɛ/ (0.145), /iː/ - /eː/ (0.133). Contrasts involving ɑ-vowels have sim_scores below 0.07, mainly due to ɑ-vowel > e-vowel substitutions, which do not occur in the opposite direction.

back V contrasts	*rounded (r) - non-rounded (nr)*	*sim_scores*
ʊ - uː	r	0.451
ɑ - ɑː	nr	0.408
ɔ - oː	r	0.314
ʊ - ɔ	r	0.210
uː - oː	r	0.110
ʊ - oː	r	0.103
ɔ - uː	r	0.030
ɑ - ʊ	rnr	0.017
ɔ - ɑ	rnr	0.008

Table 10.30: Sim_scores for [+back] categories

Table 10.30 summarizes the sim_scores for back vowel contrasts. The highest scores are observed for contrast pairs whose members differ only in constriction degree and phonemic length (/ʊ/ - /uː/ (0.451), /ɑ/ - /ɑː/ (0.408), /ɔ/ - /oː/ (0.314)). For qualities that differ in constriction location (velar vs. uvular), sim_scores between 0.3 and 0.1 are obtained: /ʊ/ - /ɔ/ (0.210), /uː/ - /oː/ (0.110), /ʊ/ - /oː/ (0.103). All other back vowel contrasts are significantly better kept apart (values clearly below 0.1). The high degree of perceived similarity between front and back vowels that do not differ in terms of rounding will be discussed in the next section.

10.5.4.5 The perceptual vowel map for L1 English listeners

From the similarity scores calculated by Shepard's formula, distance scores and MDS solutions were derived. For the non-metric MDS solutions, RSQ values of 0.608 for the one-dimensional solution, 0.918 for a two-dimensional solution, and 0.940 for a three-dimensional solution are obtained.

Figure 10.31 represents the one-dimensional solution non-metric MDS solution for English learners of German with an RSQ value of 0.608. The representation clearly shows that German rounded and non-rounded categories are clearly kept apart by English natives.

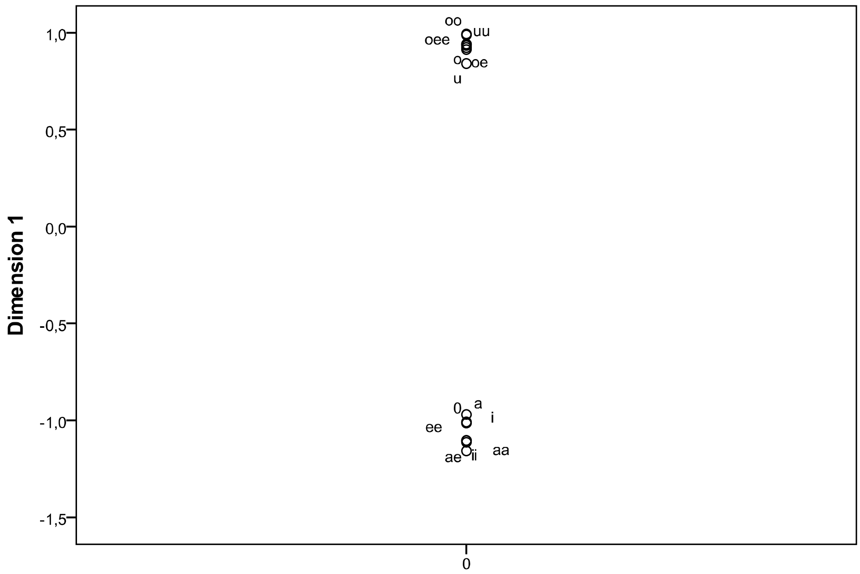

Figure 10.31: One-dimensional MDS representation of the perceptual map of German vowels for L1 English listeners (N=14, 4920 responses, RSQ .608).

The two-dimensional MDS representation of the perceptual vowel map for native English listeners in Figure 10.32 shows the major areas of difficulty for English learners of German as discussed in the previous sections. In the two-dimensional MDS solution, a spatial representation of perceived similarity relations is combined with a representation of preference patterns indicated by type size. The tight cluster of rounded vowels in the upper right part of the diagram shows the high degree of perceived similarity for all rounded vowels and the higher preference for short categories in this region of the perceptual vowel space. The diagram reflects the high degree of difficulty in differentiating front rounded from back rounded vowels. The relative distance of /ɔ/ from all other rounded vowels is mainly due to a lower percentage of substitutions with other rounded categories and to occasional substitutions of /ɔ/ with /ɑ/ (3 cases) that might be caused by orthographic interference.

Moreover, the two-dimensional solution shows that German contrasts of long constricted vs. short unconstricted vowels like /i:/ - /ɪ/ or /e:/ - /ɛ/ are kept apart quite well, while for /ɛ:/ a higher perceptual distance from other e-categories is observed.

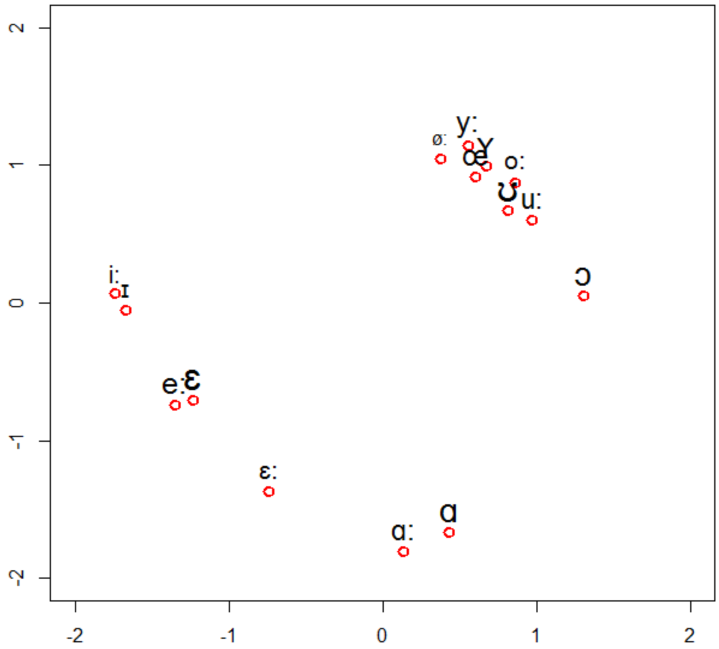

Figure 10.32: Two-dimensional MDS representation of the perceptual map of German vowels for L1 English listeners (N=14, 4920 responses, RSQ .918)

The three-dimensional MDS solution gives a slightly more differentiated impression of similarity relations for the class of rounded vowels. Front rounded vowels cluster slightly closer together compared to back rounded vowels. Again, short unconstricted /ɔ/ is differentiated best from all other rounded categories.

To summarize, contrasts of front non-rounded vowels vs. back vowels cause generally less difficulties to native English listeners, whereas the differentiation of rounded vowels with respect to the front/back dimension as well as to the exact identification of the constriction location is difficult.

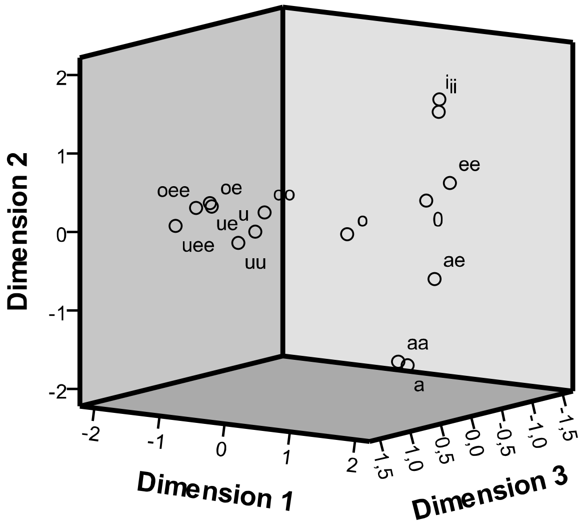

Figure 10.33: Three-dimensional MDS representation of the perceptual map of German vowels for L1 English listeners (N=14, 4920 responses, RSQ .940)

10.5.5 Summary

The most conspicuous effect found in the data of English listeners is the high amount of perceptual confusion of German rounded vowels with respect to their constriction location in the front or the back part of the oral cavity (see Figure 10.32). As the contrastive analysis in section 10.5.3 shows, front rounded vowels do not occur in the English phonemic system and the combination of the features [front] and [round] has no distinctive function in the English system. Acoustically, German ü- and ö-qualities show vague similarities with English /ɜ/ for which considerable allophonic variation with respect to lip rounding is observed. A large part of wrong categorizations observed are due to difficulties with the correct identification of ü- and ö-vowels. The two-dimensional and the three-dimensional MDS solution show a high degree of overlap for front rounded and back rounded vowels.

/ø:/ is found to be the most difficult category in the present data (72% wrong). Major difficulties are moreover found for /u:/ (65% wrong), /ʏ/ (63% wrong), /ʊ/ (60% wrong), and /e:/ (57% wrong), /ɛ:/ (54% wrong), /y:/ (53% wrong), and /œ/ (51% wrong). Generally, short unconstricted categories are more preferred response options than their long counterparts. <eu> (1.1%) and <ei> (0.7%) are relatively frequent response categories, not only for long input stimuli. Only attenuated asymmetry effects are observed for the English language sub-group, with the only exception of /y:/ - /ø:/ (/ø:/ > /y:/ (19%) vs. /y:/ > /ø:/ (6%)) and partly for /o:/ - /ø:/ (/o:/ > /ø:/ (7%) vs. /ø:/ - /o:/ (1%)).

Many previous comparative studies on German-English vowel perception (e.g. Flege 1987; Polka 1995; Best et al. 2003; Strange et al. 2004; Strange, Bohn, Nishi & Trent 2005) have found that native English listeners have considerable problems with front rounded and also with some back rounded vowel qualities. Moreover, several studies (e.g. Hillenbrand, Clark & Nearey 2001; Strange et al. 2005; Strange 2007; Harrington et al. 2008) have shown that English high back vowels can vary considerably along the front-back dimension and that they can be fronted in many regional varieties of English, especially in coronal contexts and in rapid and continuous speech with an average increase of F2 values by 2-3 barks and that allophones of AE back rounded vowels in coronal contexts are phonetically very similar to NG front rounded vowels, while mid-low and low back vowels are not affected (Strange et al. 2007; Levy & Strange 2008; Levy 2009; Strange 2010) and AE front vowels show only little variation across phonetic or prosodic contexts.

Perceptual assimilation experiments (e.g. Strange et al. 2004; Strange et al. 2005) using forced choice identification tasks with English response categories and judgments for goodness of fit show that the four (North) German (NG) front rounded vowels, which show vague acoustic similarities with English /ɜ/ and considerable allophonic variation with respect to lip rounding, are assimilated overwhelmingly to back categories by AE listeners (see Table 10.23). Only tokens of NG [œ] are heard as intermediate between AE front [ɛ] and back [ʊ, ʌ]. Despite the context-sensitive allophonic variation observed with English u-vowels, Strange et al. (2005) found no striking context-dependent variation in perceptual assimilation of NG vowels to AE vowels. The researchers concluded that AE listeners would perceptually assimilate front rounded vowels on the basis of their experience with acoustic variability in their L1 and apply their experience with allophonic fronting of back vowels in certain contexts.

In the present data, a higher percentage of perceptual fronting of /y:/ is only observed in the context of pre-vocalic /t/ and /z/ but not /d/. For /ʊ/, a higher percentage of substitutions with front ü-vowels is observed in the context of pre-vocalic /w t z j/ and even /k/. Generally, both categories /u:/ and /y:/ have a strong mutual attraction effect which seems to work in both directions for no strong asymmetries become evident.

In pronunciation, mid /e/ and /o/ tend to be diphthongized in English, especially in AE varieties. However, in the present study, “perceptual diphthongization” is relatively seldom found for /e:/ and /o:/-stimuli (only 1% both). The highest percentage of diphthongal categorizations is observed for German /i:/ (4%), /œ/ (4%), /ø:/ (4%), and /ʏ/ (3%).

Difficulties in the perceptual identification of German vowels are expected to be not only caused by phonological and phonetic differences between the German and the English vowel systems but also to differences in the *orthographic systems* of the two languages. Though many experimental studies on German-English vowel perception have tried to avoid this factor, it cannot be denied that orthography does influence the mental representation of vowel categories in a given language. In the present study, an interference of English orthography is observed for some categories, especially for /ɑ/ ~ <u, u:>, /e/-qualities ~ <a>, i-qualities ~ /e:/ (10%) and /ɛ/ (17%), and for /u:/ ~ /o:/ (6%).

Despite the fact that there are more empirical studies on English-German interferences in vowel perception than for any other language studied in the current work, there are several open research questions. Further studies will have to focus on differences between regional varieties of English and their impact on perceptual uncertainties in German but also on a more detailed investigation of the effect of context on vowel perception, but these questions are far beyond the scope of the present study. Another interesting issue would be whether English listeners show differences in perceptual categorization with stimuli from North vs. South German varieties.

To summarize, as the results of empirical studies on the perception of German vowels by English natives show, neither a contrastive analysis in terms of phonological categories nor the comparison of acoustic data for the two languages are sufficient to account for all difficulties observed with English native listeners. The established inter-category relationships of perceptual distance and similarity can only be described if the complex interaction of phonological representations, acoustic and articulatory patterns, and the influence of the complex orthographic systems in both languages are included in the analysis.

10.6 Farsi

10.6.1 General characteristics

Farsi/Modern Persian is an Iranian language belonging to the Indo-Iranian branch of the Indo-European language family. It is spoken by 80 to 100 million people in Iran, Afghanistan, Tajikistan and Uzbekistan and by minorities in countries of South-West and Central Asia as well as by emigrants in several Western countries.

The Modern Persian language is divided into three dialect groups, Farsi (Western Persian, spoken in Iran and by some minorities in Iraq and the Persian golf), Dari (Eastern Persian, spoken in Afghanistan) and Tajik. The following description will refer to contemporary Farsi as spoken in Iran.

10.6.2 Phonological and phonetic description

The vowel system of Modern Farsi consists of three front and three back vowels (Toosarvandani 2004; Mumm 2009) and is generally considered to be a quality-based system in which quantity is not distinctive.

	front	back
high	i	u
mid	e	o
low	a	ɑ

Table 10.31: The vowel system of Modern Persian (Farsi) in terms of traditional phonological features

In the Persian writing system, which is based on the Arabic system, only consonants are written. The historically "long" vowels /a:/, /u:/ and /i:/ are represented by the letters آ ی وُ . Short vowels are marked as diacritics added to consonant characters and are not used in texts of every-day life.

Historically, early New Persian had a system of five long vowels /i: e: a: o: u:/ and three short vowels /i u a/. Long /e:/ and /o:/ merged with /i:/ and /u:/ [103].

[103] Eastern varieties apparently preserve the distinction of /o:/ and /u:/ and of /e:/ and /i:/, whereas in Standard Tajik the length distinction has disappeared merging /i:/ with /i/ and /u:/ with /u/ (Perry 2005). In Dari varieties, the system is the least changed: As in Iran, short high vowels are lowered to mid vowels, which consequently are opposed to the retained long mid vowels /e:/ and /o:/ (Windfuhr 2011: 457f).

The historical opposition of vowel length in Classical Persian has largely disappeared in all but open non-final syllables, while qualitative differences have arisen between the once identical short and long vowels resulting in a system of six vowel qualities /i u e o ɑ a/.

The phonetic quality of Modern Farsi vowels was described recently by Salehi et al. (2008) in an experimental study with 60 Iranian subjects (18-24 years old, 30 male, 30 female). The mean formant frequencies F1, F2, F3 for the male speakers are represented in Figure 10.34. Salehi et al. (2008) interpret the acoustic data to distinguish anterior [i e æ] from posterior [u o a] (see Figure 10.34). The relative proximity of F1 and F2 for /a:/ points to a back quality [ɑ], while the relative proximity of F2 and F3 reveals a front quality [æ] for /a/. in pronunciation, /a:/ is frequently pronounced as [ɒ] which is one of the characteristic features in the pronunciation of Farsi learners of German (see).

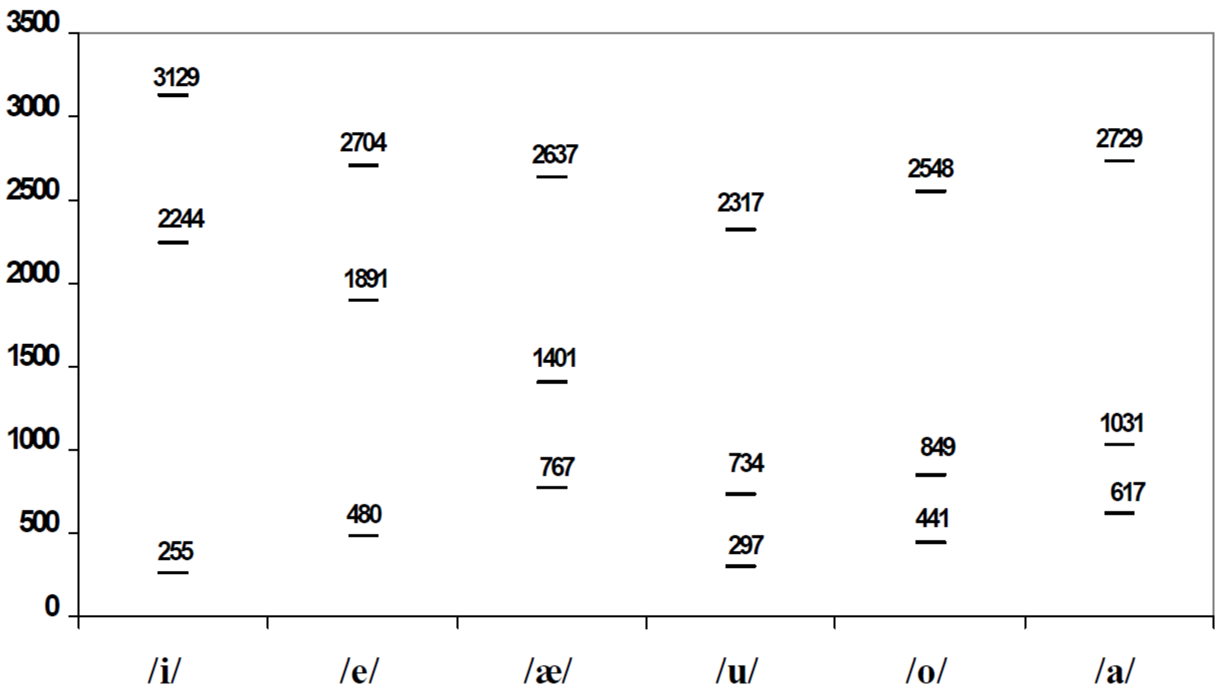

Figure 10.34: Formant frequencies of Persian male speakers produced on data from Salehi et al. (2008: 39)

Historical length differences in Modern Farsi are reported to be preserved only in open, non-final syllables (Toosarvandani 2004). In contemporary Farsi, there is some systematic variation in vowel duration preserved under certain conditions: The historically long vowels /i u ɑ/ – often referred to as "stable" vowels (e.g. Toosarvandani 2004) – remain the same in all environments, whereas the so-called "unstable" vowels /a e o/ may differ in their duration depending on syllable-position and may undergo assimilation and syncopic processes in open syllables (Mumm 2009: 13). The three unstable vowels /a e o/ are described as long in closed or word-final syllables and as short in open, non-final, unstressed syllables, e.g. *se.dá* 'sound' – *se:f.tár* 'harder' (Toosarvandani 2004: 242). The stable vowels /i u ɑ/ always maintain their length. Some studies suggest that the duration of short "unstable" vowels in closed unstressed syllables is intermediate between closed stressed syllables and open unstressed syllables (Toosarvandani 2004) and argue that phonemic length distinctions are neutralized in most positions, so that realizations of so-called long vowels [i u ɑ] are not longer than the so-called short [æ e o] (Rohany Rahbar 2009: 3).

[+front]		[-front]	
/i:/ [i]	ی	/u:/ [u]	او
/ɪ/ [e]	إ	/ʊ/ [o]	أ
/a/ [æ]	ا	/a:/ [ɑ]/[ɒ]	آ

Table 10.32: Persian vowel phonemes, their phonetic quality and orthographic representation[104]

Although the Perso-Arabic writing system as well as many transliteration systems, in their attempt to preserve historical distinctions, suggest the existence of a long-short distinction in contemporary speech, most phonological descriptions agree that phonological length is not the (primary) relevant feature in Modern Farsi. However, there is considerable disagreement on the issue whether quality and/or quantity are active features in contemporary Farsi.

Three types of approaches to the question of quantity and/or quality contrasts are found in the literature (for a critical review, see Rohany Rahbar 2009): a historically oriented quantity-based where quantity is active and a three-vowel system plus length contrast is assumed (/i/ [e], /i:/ [i], /u/ [o], /u:/ [u], /a/ [a], /a:/ [ɑ]), where short vowels differ in phonetic quality (Hayes 1979; Windfuhr 1979), a quality-based approach in which quantity is not phonologically active (/i e u o a ɑ/), and an approach, where quality and quantity both are active (/i: e u: o a a:/) (e.g. Toosarvandani 2004; Windfuhr 2011). Rohany Rahbar (2009) reconsiders the assumptions about the relevance of quality and/or quantity for Modern Farsi and argues with examples where the so-called short vowels can even be longer in duration than the so-called long ones. She suggests that quality is the relevant feature in the vowel system, referring to the decisive role of vowel height harmony as an active phonological

[104] /a, e, o/ are represented by diacritics. In word-initial position, these vowels are represented by *alif* ا. In word-final position, /a/ and /e/ are represented by ه, /o/ is represented by و.

process in Modern Farsi by which a mid vowel is raised to a high vowel, and proposes that the active qualitative feature is not height but tenseness/laxness resulting in a categorization of tense /i u ɑ/ and lax /e o a/. In a more recent approach, Windfuhr (2011: 449f) argues for "stability" as the contemporary feature, indicated partly by the lengthening of /e o a/ (< short /i u a/) when stressed and their shortening and partially conditioned assimilation when unstressed.

Modern Farsi has two diphthongs /aw/ and /ay/. While /aw/ tends increasingly to be monophthongized to [o:] in Iranian Farsi, even in higher registers (Windfuhr 2011: 450), /ay/ usually preserves its diphthongal character.

The final syllable is usually stressed in Modern Farsi. Suffixes are also stressed while enclitica are not stressed. However, there are several exceptions to this rule, e.g. in prefixed verbs where the prefix carries stress.

Syllable structures in Farsi allow no complex consonant clusters and vowels in the onset: V, CV, CVC, CVCC, VC, and VCC are allowed syllable structures. Initial and medial consonant clusters are forbidden in Persian and are split by the insertion of an [e] in words of foreign origin (Rohany Rahbar 2009: 10f).

To summarize, the question whether quality or/and quantity are distinctive in Persian phonology is a controversially discussed issue. The quantity distinction of long vs. short vowels is a residue of historical patterns enhanced by the use of the Perso-Arabic script suggesting an inventory of three short and three long vowels. For contemporary Farsi, it is assumed here that the contrast of the historically long and short vowels of Persian is to be considered not as a difference in length but in quality. The differences in phonetic quality (Figure 10.34) suggest that the quality approach is more appropriate for a synchronic description of the Farsi system and that quantity does not seem to be the active feature for vowel contrast. The system can be described as consisting of 6 monophthongs [i e æ ɑ/ɒ o u]. More detailed acoustic-phonetic research on vowel quality and durational effects considering the influence of open vs. closed syllable structure and word stress has to be performed for Farsi vowels.

10.6.3 Contrastive analysis

Persian has a relatively small vowel inventory of only six different qualities. Therefore, considerable problems in vowel perception are expected in L2 German due to the restricted number of vowel phonemes in Farsi. Due to the conventions of the Persian writing system and the non-obligatory graphemic representation of short vowels, the L2 learners' general awareness of the contrastive function of vowel qualities may be reduced, especially for German mid- and prepalatal e- and i-qualities and for back rounded u- and o-qualities. These difficulties may be enhanced by the specific set of distinctive features in German, especially the use of phonemic length distinctions associated with contrasts in constriction degree. Moreover, problems with the orthographic representation of vowel contrasts in German are expected that are not only explained by difficulties with German orthographic conventions but could also be due to the learners' limited fluency in the Latin based alphabet in general.

The phonetic quality of Persian a-qualities clearly differs from German a-qualities. Salehi et al. (2008) argue that /a:/ is a back quality because of the relatively high proximity of F1 and F2. In pronunciation, /a:/ is described as [ɒ] or [ɑ] rather than [a]. With less advanced learners of German, a transfer of this native quality for German a-vowels is observed in pronunciation. The historically short /a/-vowel in Farsi is mostly pronounced as fronted [æ] which clearly differs from Farsi /e/.

Front rounded vowels do not exist at all in Modern Farsi. Difficulties with these qualities are therefore expected in L2 German. Experiences from teaching learners with L1 German moreover show that the differentiation of o- and u-qualities and of e- and i-qualities is difficult for Farsi native speakers in perception.

10.6.4 Results and discussion

Four Farsi subjects were tested, three beginners and one advanced learner. All subjects were born in Iran. The advanced learner has been living in Vienna for 32 years showing nevertheless a relatively low id_correct score of 60%. The beginners' length of residence varies between one and six months, their id_correct score ranges from 30% to 73% of the responses given.

Because of the very small sample and the open questions in the phonological and phonetic analysis of the Farsi vowel system discussed above, the data presented here can only provide a first insight into perceptual processes of Farsi learners of German but are not to be generalized.

The data set discussed here consists of 1079 valid responses (x missing answers). For each vowel category, 71 to 72 valid responses were delivered.

10.6.4.1 Difficulties

The highest percentage of correct answers is observed for /a/ (75%), /ɛ/ (74%) and /ɔ/ (68%). These qualities are considered to be most stable in L2 perception. The lowest id_correct scores are obtained for /ø:/ (17% correct), /ɛ:/ (20%), /e:/ (31%) and /o:/ (33%).

Within the class of front rounded vowels, for /y:/ least difficulties are observed, while for /ø:/ the highest percentage of wrong identifications is given.

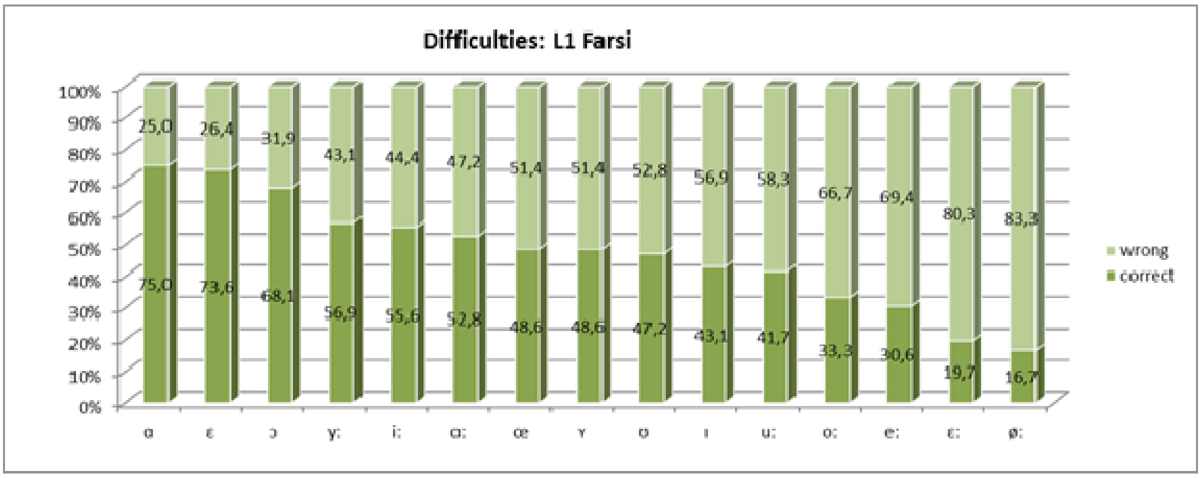

Figure 10.35: Percentage of correct and wrong identifications by Farsi listeners (N=4, 1079 stimuli)

While /ø:/ and /ɛ:/ are difficult to all speakers, with most other categories the inter-speaker variability is considerably high (see the boxplots in Figure 10.36). Least variability is observed with /ɑ/.

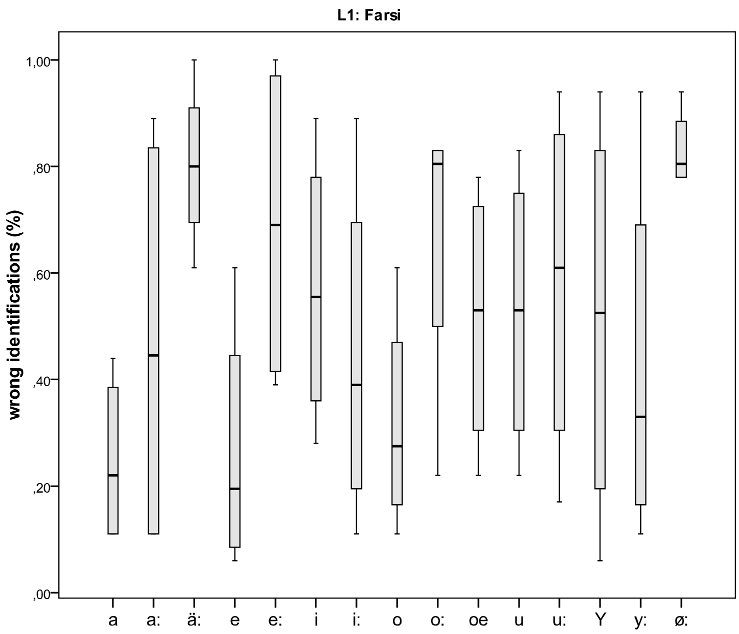

Figure 10.36: Id_wrong scores for Farsi listeners

The listeners' mean id_wrong score was 53% (18% standard deviation), the median was 54%. The listener with the highest percentage of correct answers had an id_wrong of 30%, the highest id_wrong score was 73% wrong responses.

10.6.4.2 Preferences

The preference scores for monophthong categories vary considerably within a range of 6.1% (min /ɛ:/ 3.3%, max /y:/ 9.5%). The most preferred response categories are /y:/ and /ɛ/ followed by /ɔ/ and /ɑ/. These vowels also show a higher percentage of correct answers than all other categories. In general, the short unconstricted categories are preferred over their long counterparts, with the only exception of /i:/ which is more often selected than /ɪ/. The least preferred response options are /ø:/, /o:/ and /ɛ:/, which also show very low id_correct scores, and diphthongs.

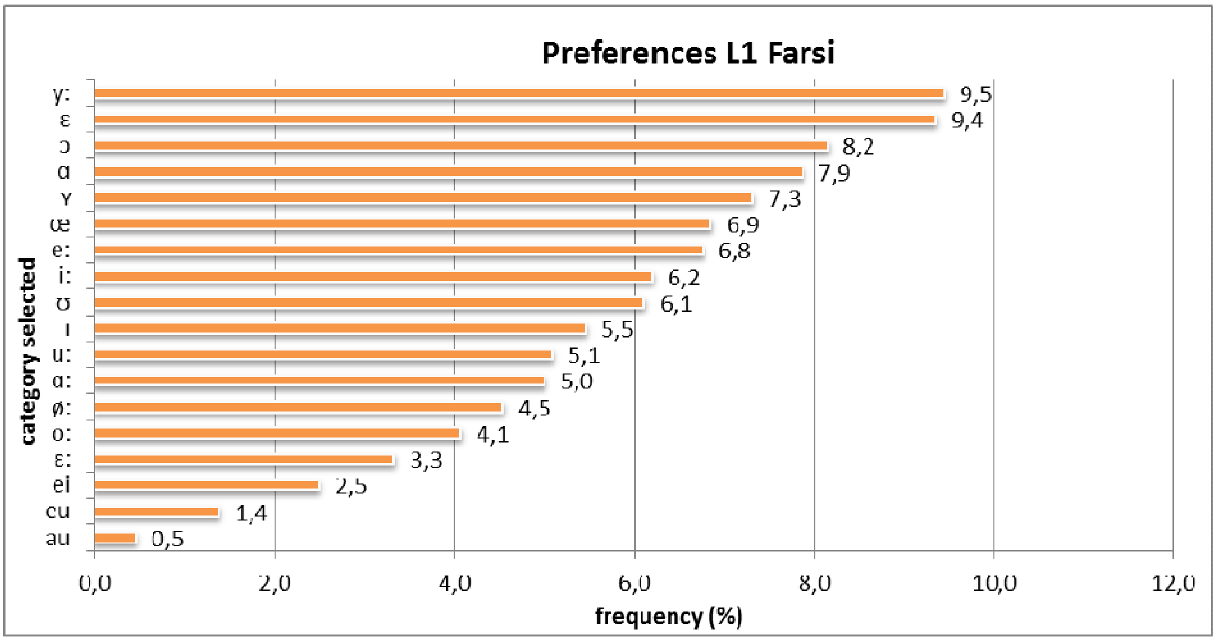

Figure 10.37: Preferred response categories for L1 Farsi listeners

Compared to other language sub-groups, the Farsi learners in this sample show a strong tendency to diphthongize German monophthongs perceptually. <ei> is selected as response option in 2.5% of the cases, <eu> in 1.4% and <au> in 0.5% (only for a-qualities) of the cases. In Iranian Farsi a strong tendency to monophthongize /ow/ > /o:/ is observed, while the diphthong /ei/ is preserved in Farsi. Difficulties with the correct recognition of German diphthongs and the hypercorrect answers observed in the present data may be explained by the absence of equivalent diphthongs in the learners' native language and personal strategies of the participants (particularly Per4).

10.6.4.3 Patterns of confusion

A first look at the confusion matrix (see Figure 10.38) reveals two main areas of difficulty: (1) difficulties with respect to quality distinctions within the class of front non-rounded vowels, especially with e-qualities, and (2) a very high percentage of confusion within the class of front as well as back rounded vowels. The distinction of front vs. back rounded vowels appears to be specifically difficult for Farsi listeners.

%	ɑ	ɑ:	ɛ:	ɛ	e:	ɪ	i:	ɔ	o:	ʊ	u:	œ	ø:	ʏ	y:	‹ei›	‹eu›	‹au›	wrong	**corr**
a	**75**	21						1										3	25	**75**
a:	42	**53**		1														4	47	**53**
ɛ:			**20**	18	49		4								1	4	3		80	**20**
ɛ	1	1	4	**74**	14	3										1	1		26	**74**
e:			13	21	**31**	6	17					1				13			69	**31**
ɪ			4	22	3	**43**	17					1				10			57	**43**
i:					4	29	**56**								3	8			44	**56**
ɔ								**68**	13			10	6			1	3		32	**68**
o:			1					24	**33**	3	10	7	14	3	4		1		67	**33**
ʊ				1				14	6	**47**	10	1	1	10	10				53	**47**
u:									1	28	**42**		8	6	15				58	**42**
œ			8	3	1			14	7	1	1	**49**	11		1		3		51	**49**
ø:									1	3	6	21	**17**	10	38		6		83	**17**
ʏ						1		1		8	4	10	10	**49**	13		4		51	**49**
y:										1	4	3	1	33	**57**				43	**57**
	8	*5*	*3*	*9*	*7*	*5*	*6*	*8*	*4*	*6*	*5*	*7*	*5*	*7*	*9*	*2.5*	*1.4*	*.5*	*52.5*	***47.5***

Figure 10.38: Confusion matrix for L1 Farsi listeners (N=4, 1079 responses). Correct (in bold) and wrong identifications in % (presented stimuli in rows, selected categories in columns in semi-orthographic representation, values <0.5 not indicated)

For German a-vowels, most difficulties are due to wrong categorizations with respect to phonemic length: For /ɑ/, 75% of the stimuli are identified correctly and mainly substitutions with /ɑ:/ (21%) are observed.

/ɑ:/ (53% correct) is categorized as short /ɑ/ in 42% of its occurrences. The advanced learner (Per1) and one of the beginners (Per4) occasionally use <au> as substitute for /ɑ/ (3%) and /ɑ:/ (4%).

There are considerable problems with the correct identification of German e-qualities: For /ɛ:/, only 20% correct identifications and a high percentage of substitutions with /e:/ (49%) as well as /ɛ/ (18%), but also with /i:/ (4%), <ei> (4%) and <eu> (3%) are observed.

/e:/ is identified correctly in only 31% of the cases and is substituted with /ɛ/ (21%), /i:/ (17%), /ɛ:/ (13%), /ɪ/ (6%), and <ei> (13%).

Diphthongs for e-qualities are especially selected by one participant (Per4) and could of course be due to a personal strategy of this learner rather than to a more general strategy of Farsi listeners.

/ɛ/ has a very high percentage of correct identifications (74%) and shows less variation in substitution patterns. It is substituted with /e:/ (14%), /ɛ:/ (4%), and /ɪ/ (3%).

The very low id_correct scores for /e:/ and /ɛ:/ (together with the very high sim_score for this contrast pair) indicate that these qualities are one of the most difficult contrasts for L1 Farsi learners. While unconstricted /ɛ:/ is mainly substituted by /e:/ (49%), long constricted /e:/ is rather substituted by /i:/ (17%), whereas perceptual substitutions of /e:/ > /i:/ are significantly less common (only 4%). Thus, in both cases we observe a kind of perceptual attraction effect of higher and more constricted categories /e:/ and /i:/ respectively, that is not observed for the short unconstricted cognates /ɪ/ and /ɛ/.

The responses for unconstricted short /ɪ/ (43% correct) spread over the class of e- and i-vowels: Substitutes are /ɛ/ (22%), /i:/ (17%), /ɛ:/ (4%), /e:/ (3%), and <ei> (10%, only by Per4).

Less variation is observed with /i:/ (56% correct), which is identified as /ɪ/ (29%), <ei> (8%, only Per4), /e:/ (4%) and /y:/ (3%). Apparently e-qualities are not a very attractive response alternative for /i:/. The substitution of /i:/ with /y:/ is applied in only two instances by two different subjects and can be interpreted as hypercorrect reaction or inversion of a process of perceptual unrounding which is observed more frequently with /œ/.

Unconstricted short /ɔ/ is considerably more often identified correctly (68%) than its constricted counterpart /o:/. For /ɔ/, we find /o:/ (13%), /œ/ (10%), /ø:/ (6%) and <eu> (3%, two subjects) selected as response options.

Long constricted /o:/ is one of the categories that shows most variation in perceptual substitution patterns. It is correctly identified in only 33% of the cases and is substituted with /ɔ/ (24%), /ø:/ (14%), /u:/ (10%), /y:/ (4%), /œ/ (7%), /ʊ/ (3%), and /ʏ/ (3%). In only 1% of the cases, /o:/ is substituted with <au>. Apparently, the realization of /ow/ to [o:] in contemporary Farsi has no interfering effect on the perception of German /o:/.

Unconstricted /ʊ/ is identified correctly in 47% of its occurrences. Perceptual substitutions occur with /ɔ/ (14%), /u:/ (10%), /ʏ/ (10%), /y:/ (10%) and /o:/ (6%).

Preferred response categories for /u:/ (42% correct) are /ʊ/ (28%), /y:/ (15%), /ø:/ (8%) and /ʏ/ (6%).

All front rounded vowels have very low id_correct scores. Within this class, least problems are observed with /y:/ (57% correct), which is substituted most often with /ʏ/ (33%) and less often with /u:/ (4%) and /œ/ (3%). All other categories show considerable variation concerning the choice of response categories.

For /ʏ/ (49% correct), the most preferred categories are /y:/ (13%), /œ/ (10%), and /ø:/ (10%); other options are /ʊ/ (8%), /u:/ (4%) and <eu> (4%).

/œ/ (49% correct) is substituted with /ɔ/ (14%), /ø:/ (11%), /ɛ:/ (8%), /o:/ (7%), /ɛ/ (3%), and <eu> (3%). A process of perceptual unrounding of /œ/ and a substitution with /ɛ:/ (8%) and /ɛ/ (3%) is not observed with three of four subjects.

/ø:/ is the most difficult category for Farsi listeners with an id_correct score of only 17%. Instances of /ø:/ are substituted most often with /y:/ (38%), /œ/ (21%) and /ʏ/ (10%) and less often with /u:/ (6%), <eu> (6%) and /ʊ/ (3%).

Farsi listeners show two patterns of *asymmetry* with front rounded vowels: (1) /œ/ is not substituted with /ʏ/, while /ʏ/ is substituted with /œ/ or /ø:/ in 10% of the cases each, and (2) /y:/ is substituted only once with /ø:/, while /ø:/ is attracted to /y:/ in 38% of the cases.

A similar but weaker pattern of asymmetry is observed with /o:/ and /u:/ and with /e:/ and /i:/: German /i:/ is identified with /e:/ only 3 times (4%), whereas /e:/ is substituted with /i:/ in 12 cases (17%). While /u:/ is identified as /o:/ only once, /o:/ is identified as /u:/ in 7 cases (10%).

A very similar pattern is observed in the historical development from early New Persian to contemporary Iranian Farsi: The long qualities of early New Persian /i:/ and /e:/ as well as /u:/ and /o:/ merged to Farsi /i:/ and /u:/ respectively (Windfuhr 2011: 457). The use of graphemes for /i:/ and /u:/ but not for mid qualities in the Perso-Arabic writing system may also contribute to this substitution effect. However, these observations indicate an interesting asymmetry but have to be confirmed on a larger data base.

10.6.4.4 Similarity and distance

A comparison of similarity scores (see Table 10.33) shows that the distinction of /e:/ and /ɛ/ is least distinctive for Farsi learners. This contrast pair shows a strong pattern of asymmetry: /e:/ attracts 49% of the /ɛ:/-stimuli, whereas only 13% of the /e:/-stimuli are substituted by /ɛ:/ and responses for /e:/ are spread over e- and i-qualities and the diphthong <ei>.

The second highest sim_score is observed with the contrast /ø:/ - /y:/ (0.528), followed by /ɑ/ - /ɑ:/ (0.489). Very high sim_scores between 0.489 and 0.333 are moreover observed with vowel pairs that differ only in constriction degree and phonemic length, such as /ø:/ -/œ/ (0.489), /i:/ - /ɪ/ (0.465), /y:/ - /ʏ/ (0.434), /u:/ - /ʊ/ (0.422), /o:/ - /ɔ/ (0.356), /e:/ - /ɛ/ (0.333), for /ø:/ - /o:/ (0.306) and /ʏ/ - /ø:/ (0.298).

V	ɑ	ɑ:	ɛ:	ɛ	e:	ɪ	i:	ɔ	o:	ʊ	u:	œ	ø:	ʏ	y:
a	1.000														
ɑ:	0.489	1.000													
ɛ:	0.000	0.000	1.000												
ɛ	0.009	0.022	0.241	1.000											
e:	0.000	0.000	1.229	0.333	1.000										
ɪ	0.000	0.000	0.066	0.214	0.113	1.000									
i:	0.000	0.000	0.056	0.000	0.242	0.465	1.000								
ɔ	0.010	0.000	0.000	0.000	0.000	0.000	0.000	1.000							
o:	0.000	0.000	0.026	0.000	0.000	0.000	0.000	0.356	1.000						
ʊ	0.000	0.000	0.000	0.011	0.000	0.000	0.000	0.120	0.103	1.000					
u:	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.148	0.422	1.000				
œ	0.000	0.000	0.122	0.023	0.035	0.015	0.000	0.202	0.169	0.029	0.015	1.000			
ø:	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.066	0.306	0.065	0.238	0.489	1.000		
ʏ	0.000	0.000	0.000	0.000	0.000	0.015	0.000	0.012	0.034	0.188	0.108	0.100	0.298	1.000	
y:	0.000	0.000	0.018	0.000	0.000	0.000	0.025	0.000	0.046	0.107	0.197	0.039	0.528	0.434	1.000

Table 10.33: Similarity matrix for L1 Farsi listeners

The major difficulty for Farsi listeners in German is the differentiation of *front vs. back rounded* vowels as evident in the confusion matrix and the similarity matrix. A perceptual under-differentiation of front vs. back rounded vowels with respect to their constriction location is reflected in the high sim_scores for these vowel pairs (see Table 10.34), which range, however, after that of contrasts differing in constriction degree and phonemic length.

front vs. back contrasts	*sim_scores*
ø: - o:	.306
u: - ø:	.238
œ - ɔ	.202
u: - y:	.197
ʊ - ʏ	.188
o: - œ	.169
u: - ʏ	.108
ʊ - y:	.107
ɔ - ø:	.066
ʊ - ø:	.065
o: - y:	.046
o: - ʏ	.034
œ - ʊ	.029
o: - ɛ:	.026
ɛ - ɑ:	.022
u: - œ	.015
ʏ - ɔ	.012
ɛ - ɑ	.009

Table 10.34: Similarity scores of German front-back vowel contrasts for L1 Farsi listeners

Within the class of non-rounded vowels, palatal vs. pharyngeal contrasts receive very low sim_scores, whereas the differentiation of vowel categories that differ only in constriction degree and phonemic length (e.g. /i:/ - /ɪ/ (0.465) and /e:/ - /ɛ/ (0.333) of pre- vs. mid-palatal vowels is relatively difficult, especially for /i:/ - /e:/ (0.242) and /ɪ/ - /ɛ/ (0.214).

non-rounded Vs	*sim_scores*
e: - ɛ:	1.229
ɑ - ɑ:	.489
i: - ɪ	.465
e: - ɛ	.333
i: - e:	.242
ɛ - ɛ:	.241
ɪ - ɛ	.214
ɪ – e:	.113
ɪ - ɛ:	.066
i: - ɛ:	.056
ɛ - ɑ:	.022
ɛ - ɑ	.009

Table 10.35: Similarity scores of German [–round] vowels for L1 Farsi listeners

Confusion of vowels within the class of *front vowels* is reflected by high sim_scores (Table 10.35) for /e:/ - /ɛ:/ (1.229), /i:/ - /ɪ/ (0.465), /e:/ - /ɛ/ (0.333), /i:/ - /e:/ (0.242), /ɛ/ - /ɛ:/ (0.241), /ɪ/ - /ɛ/ (0.214), and lower values for /ɪ/ - /e:/ (0.113), /ɪ/ - /ɛ:/ (0.066), and /i:/ - /ɛ:/ (0.056).

front Vs	*sim_scores*
e: - ɛ:	1.229
ø: - y:	.528
œ – ø:	.489
i: - ɪ	.465
ʏ - y:	.434
e: - ɛ	.333
ʏ - ø:	.298
i: - e:	.242
ɛ - ɛ:	.241
ɪ - ɛ	.214
œ - ɛ:	.122
ɪ – e:	.113
œ - ʏ	.100
ɪ - ɛ:	.066
i: - ɛ:	.056
œ - y:	.039
œ – e:	.035
y: - i:	.025
œ - ɛ	.023
y: - ɛ:	.018
œ - ɪ	.015
ʏ - ɪ	.015

Table 10.36: Similarity scores of German front vowels for L1 Farsi listeners

High sim_scores for contrasts differing only in constriction location but not length or constriction degree are observed for contrast pairs of [+/–back] rounded vowels, e.g. /ø:/ - /o:/, /u:/ - /ø:/, /œ/ - /ɔ/. Contrasts involving front vowels vs. a-vowels receive very low sim_scores. Table 10.37 summarizes the sim_scores for back vowels and shows a relatively high degree of perceived similarity between velar and uvular qualities: /u:/ - /o:/ (0.148), /ʊ/ - /ɔ/ (0.120) and /ʊ/ - /o:/ (0.103).

back Vs	*sim_scores*
ɑ - ɑ:	.489
u: - ʊ	.422
o: - ɔ	.356
u: - o:	.148
ʊ - ɔ	.120
ʊ - o:	.103

Table 10.37: Similarity scores of German back vowels for L1 Farsi listeners

Contrasts of rounded vs. non-rounded vowels have relatively low sim_scores below 0.1 (see Table 10.38) with the only exception of /œ/ - /ɛ:/, where the sim_score of 0.122 may partly be due to the orthographic similarity of <ö> and <ä>.

round-non-rounded contrasts	*sim_scores*
œ - ɛ:	.122
œ – e:	.035
o: - ɛ:	.026
y: - i:	.025
œ - ɛ	.023
y: - ɛ:	.018
œ - ɪ	.015
ʏ - ɪ	.015

Table 10.38: Similarity scores of German [+/–round] contrasts for L1 Farsi listeners

rounded Vs	*sim_scores*
ø: - y:	.528
œ – ø:	.489
ʏ - y:	.434
u: - ʊ	.422
o: - ɔ	.356
ø: - o:	.306
ʏ - ø:	.298
u: - ø:	.238
œ - ɔ	.202
u: - y:	.197
ʊ - ʏ	.188
o: - œ	.169
u: - o:	.148
ʊ - ɔ	.120
u: - ʏ	.108
ʊ - y:	.107
ʊ - o:	.103
œ - ʏ	.100
ɔ - ø:	.066
ʊ - ø:	.065
o: - y:	.046
œ - y:	.039
o: - ʏ	.034
œ - ʊ	.029
u: - œ	.015
ʏ - ɔ	.012

Table 10.39: Similarity scores of German [+round] vowels for L1 Farsi listeners

Within the class of rounded vowels, the highest sim_scores are observed for /ø:/ - /y:/ (0.528) followed by contrasts that are based only on phonemic length and constriction degree, i.e. /œ/ - /ø:/ (0.489), /ʏ/ - /y:/ (0.434), /ʊ/ - /u:/ (0.422), /ɔ/ - /o:/ (0.356). Lower sim_scores are observed for rounded vowel contrasts that differ in constriction location; these are contrast of either front vs. back rounded vowels, pre- vs. mid-palatal vowels, or velar vs. uvular vowels (see Table 10.39).

To summarize, a high portion of inter-category similarity is observed for the class of rounded vowels and for front vowels at the pre- and mid-palatal constriction location. A visualization of the similarity relations as perceived by Farsi natives is presented in the next section.

10.6.4.5 The perceptual vowel map for L1 Farsi listeners

From the similarity scores calculated by Shepard's formula, distance scores and non-metric MDS solutions were derived. For the non-metric MDS solutions, RSQ values of 0.538 in dimension 1, 0.835 for the two-dimensional solution, and 0.884 for the three-dimensional solution are obtained.

In the two-dimensional MDS solution derived from distance scores, a spatial representation of perceived similarity relations is combined with a representation of preference patterns indicated by type size.

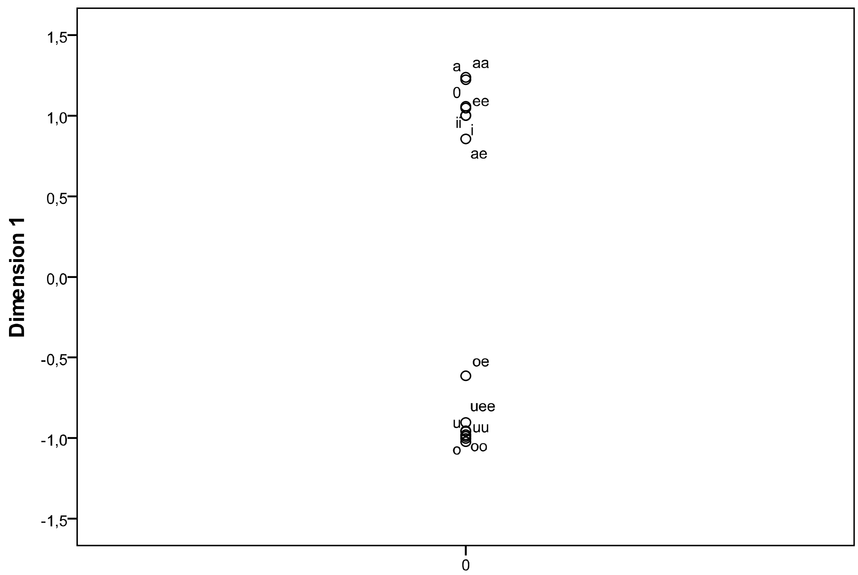

Figure 10.39: One-dimensional MDS representation of the perceptual map of German vowels for L1 Farsi listeners (N=4, 1079 tokens, RSQ .538)

The one-dimensional solution in Figure 10.39 and show a principal differentiation with respect to the feature [+/–round] by Farsi learners of German.

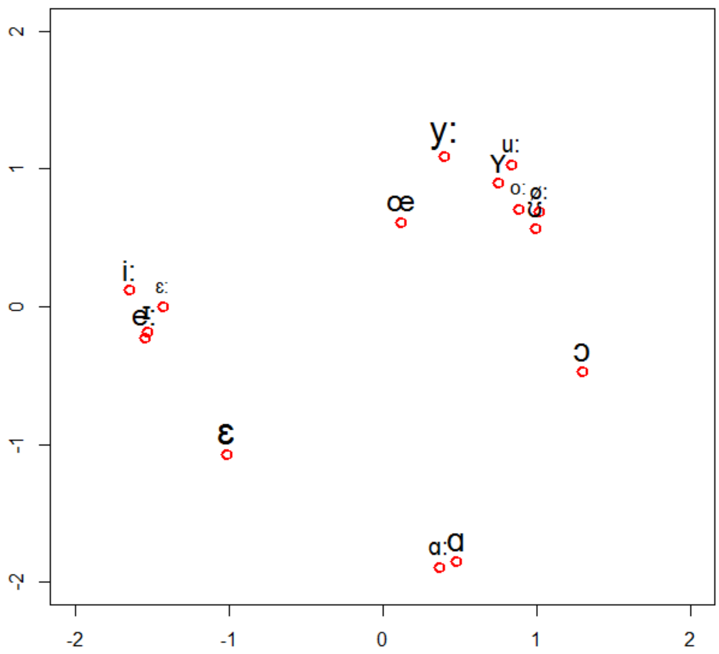

Figure 10.40: Two-dimensional MDS representation of the perceptual map of German vowels for L1 Farsi listeners (N=4, 1079 tokens, RSQ .835)

The two-dimensional MDS representation in Figure 10.40 shows moreover three major clusters in the perceptual vowel space: (1) front non-rounded vowels, (2) a-vowels and (3) the class of rounded vowels, where front and back rounded qualities are in close proximity.

Within the class of *rounded vowels*, /ɔ/, /y:/ and /œ/ are differentiated quite well from other rounded vowels, whereas for all other rounded qualities a great portion of inter-category confusion is observed (see the discussion of similarity scores above). Among rounded vowels, /y:/ and /ɔ/ show the highest preference scores (see Figure 10.37), attracting responses for other stimuli of this class.

/ɔ/ and /ɛ/ are clearly differentiated from other qualities which reflects the high percentage of correct identifications (/ɔ/ 68% correct) and /ɛ/ (74% correct).

For *palatal* non-rounded vowels, /ɛ/ is the most preferred response option, followed by /e:/ and /i:/. However, front and back rounded vowels are not easily distinguished in perception by Farsi listeners. There is considerable confusion between these front rounded and back rounded vowels, as the confusion matrix and the two-dimensional MDS solution show. Perceptual unrounding processes are observed only occasionally for /œ/.

The three-dimensional MDS solution shows a more differentiated representation for front non-rounded vowels and a differentiation of /ɔ/ and /œ/ from all other rounded vowels. It is interesting to note that the RSQ in the three-dimensional solution increases only slightly compared to the two-dimensional solution.

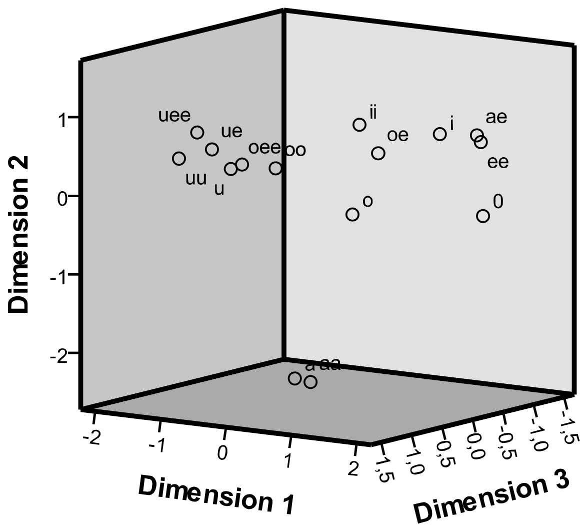

Figure 10.41: Three-dimensional MDS representation of the perceptual map of German vowels for L1 Farsi listeners (N=4, 1079 tokens, RSQ .884)

10.6.5 Summary

To summarize the results for the Farsi sample, most difficulties are observed with the differentiation of front vs. back rounded vowels with respect to the constriction location. Moreover, a differentiation with respect to phonemic length and constriction degree is difficult for Farsi listeners. There are several instances of asymmetrical perceptual substitutions. Patterns of asymmetry are strongest with /ø:/ vs. /y:/ (/ø:/ > /y:/ (38%) vs. /y:/ > /ø:/ (1%). /y:/ moreover exerts the highest attraction for all rounded vowel stimuli. A similar but less strong asymmetry effect is observed with /e:/ - /i:/ (/e:/ > /i:/ (17%) vs. /i:/ > /e:/ (4%)) and /o:/ - /u:/ (/o:/ > /u:/ (10%) vs. /u:/ >/o:/ (1%)).

Due to the small sample size, these observations have of course to be viewed with caution and verified on a larger data base. A special point of interest for future research would be the perceptual sensitivity of Farsi listeners to phonemic length and constriction degree for high vs. mid qualities. Since the question of the existence and phonemic relevance of differences in duration in Farsi is not completely solved yet, the influence of syllable structure and durational differences in Farsi and German would be another interesting field of research.

10.7 Hungarian

10.7.1 General characteristics

Hungarian is a Uralic language belonging to the group of Finno-Ugric languages. It is spoken by about 10 million people in Hungary and by around three million people in the neighbouring countries Romania, Slovakia, Serbia, Ukraine, Austria, Croatia and Slovenia as well as by emigrants who migrated to several other countries, e.g. the United States, Canada and Israel.

Typologically, as a representative of the Finno-Ugric language group, Hungarian is a language with agglutinating morphology, non-configurational syntax, and syllable-timed prosody (Siptár & Törkenczy 2000: 13). Hungarian does not show large dialectal differences, mainly for historical reasons (Kiss 2003: 907). However, on the phonetic-phonological level, regional variation and differences between the Standard variety and regional dialects can be observed. Referring to vowels, this concerns differences in the number and frequency of distinctive vowel phonemes. Most phonological and phonetic descriptions refer to a variety called Educated Colloquial Hungarian (ECH) that is spoken by the younger generations of urban speakers (e.g. Dressler & Siptár 1989; Szende 1994, 1999; Siptár & Törkenczy 2000).

10.7.2 Phonological and phonetic description

The Hungarian consonant inventory consists of 24 (25)[105] consonants (see Table 10.40). On the phonetic surface, consonants occur as short or long, consonant length is phonemically distinctive, though geminates are mostly derived by concatenation or assimilation (Siptár & Törkenczy 2000: 19).

Plosives	p b t d c ɟ k g
Fricatives	f v s z ʃ ʒ j h
Nasals	m n ɲ
Liquids	l r
Affricates	ts (dz) tʃ dʒ

Table 10.40: The consonant inventory of Hungarian (Siptár & Törkenczy 2000: 18f)

105 for discussion of /dz/, see Siptár & Törkenczy (2000: 87ff)

The Hungarian vowel inventory consists of 14 vowel sounds which can be described in terms of seven pairs of corresponding short and long vowels that are orthographically represented as <a á e é i í o ó ö ő ü ű> and can be described in terms of traditional phonological features as in Table 10.41 (Siptár & Törkenczy 2000: 15; Gósy 2004).
In Hungarian orthography, vocalic length is denoted with an acute accent on the respective vowel. Diphthongs only occur in some dialects (Kiss 2003: 908f).
The phonetic values of the short and long equivalents do not match exactly (see discussion below). This holds especially for <e> vs. <é> and <a> vs. <á>. The difference is particularly strong with /é/ [e:] vs. /e/ [ε].

	front					back	
high	i:	i	y:	y		u	u:
mid	e:		ø:	ø		o	o:
			ε			ɔ/ɒ	
low					a:		

Table 10.41: The vowel system of Hungarian in terms of traditional phonological features

Szende (1994: 92f) assumes three additional vowels: [ε:], [a], [ɑ:]. Vago (1980: 1) mentions some dialects distinguishing three mid vowels /e/, /ε/ and /e:/[106]. Szende (1999: 106) proposes two co-existent vowel systems, one of which – a "Regional Standard version" of Hungarian that is spoken by around 50% percent of Hungarian natives – does distinguish two different heights for short mid front non-roundeded vowels /e/ vs. /ε/. About half of the Hungarian speakers (Abondolo 2011: 484) distinguish a third e-quality that is similar to English *met*. This vowel is represented as <ë> in Hungarian dialectology. e.g. in *szëg* 'carpenter's nail' (like English *beg*) vs. *szeg*- 'break' (like English *set*, cf. Abondolo 2011).
Long [ε:] and [ɑ:] or [ɔ:] are very marginally occurring phonemes (see Dressler & Siptár 1989: 32). A short counterpart to long /a:/ occurs in some words like *Svájc* "Swiss", *spájz* "larder" or *advent* "advent" and in a few minimal pairs with short labial [ɔ] and long [a:] (Dressler & Siptár 1989: 32; Gósy 2004: 66f).
The exact phonetic quality of short <a> as a low or mid vowel is a matter of debate. Gósy (2004: 74; see Figure 10.42) describes the phonetic quality of short /a/ as [ɔ], while Szende (1994, 1999: 105) suggests a transcription of short /a/ as [ɑ] or [ɒ] due to its lower degree of

[106] for a brief description of dialectal variation, see Siptár & Törkenczy (2000: 20f).

rounding [107] (cf. also Mády 2001). Herok (2001: 83) describes realizations for short /a/ as [ɔ] or [ɒ] as instances of a more extreme articulation with less lip-rounding. However, as Herok (2001: 82f) points out, considering the respectable range of variation due to contextual factors (e.g. the influence of palatal consonants) but also to individual and gender-specific factors (see Kontra 1995), the question whether short /a/ and also /e/ are (open) mid vowels or rather (closed) low vowels is not to be answered definitely.

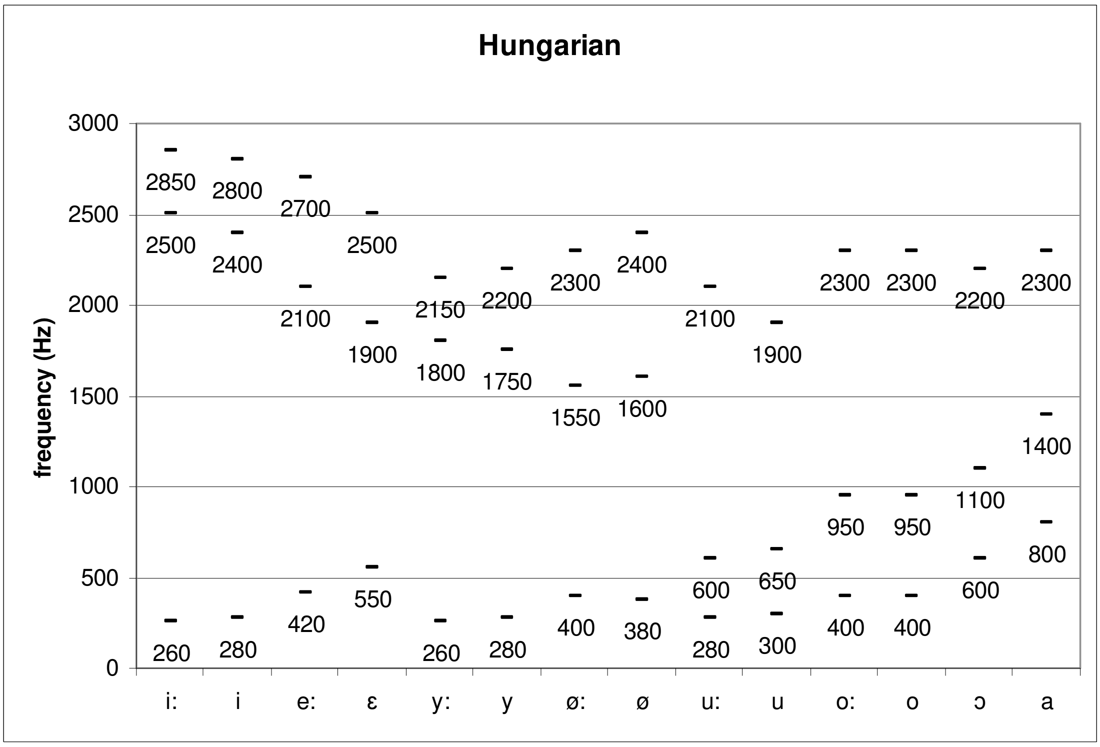

Figure 10.42: Formant frequencies F1, F2, F3 (in Hz) for Hungarian vowels (based on data from Gósy 2004: 78)

The relation of quantity and quality in the long-short vowel pairs is an interesting question for the phonetic and phonological analysis of Hungarian. Auditively and acoustically, there is a clear phonetic difference between long [e:] and short [ɛ] as well as between [a:] and [ɔ/ɒ]. However, the other long-short pairs /i:/ - /ɪ/, /y:/ - /y/, /u:/ - /u/, /ø:/ - /ø/, /o:/ - /o/ show only small though systematic differences in quality (e.g. Bolla 1995). These differences are auditively more salient with o- and ö-qualities than with high i-, u- and ü-qualities (cf. Vago 1980; Herok 2001; Mády 2010). Herok (2001) interprets the available articulatory-phonetic material with respect to differences in articulation that occur either alone or conjointly in terms of (1) a lower position of the tongue body, and/or (2) a more 'central' position of the

[107] see discussion here and in Dressler & Siptár (1989: 32) and Gósy (2004: 67f.).

whole tongue body, i.e. less fronted with long front vowels and more fronted with short back vowels, and/or (3) less lip spreading for front vowels and less lip rounding for back vowels. However, in the articulatory-based approach presented in section 4.7.5 and 4.7.6, (1) and (2) could also be interpreted as differences in constriction degree (see Figure 10.42). The question whether quantitative and/or qualitative features are actually distinctive in Hungarian "long-short" oppositions is of particular interest for a contrastive analysis of Hungarian and German and for the interpretation of the perceptual confusion data in the present study (see 10.7.3).

Though vowel quantity is generally considered to be distinctive in Hungarian, the quantity opposition appears to become unstable in colloquial speech especially for high vowels (Mády 2010). While there are substantial spectral differences between long and short /a/ and /e/, long high /i y u/ do not differ systematically from corresponding short vowels in spectral characteristics. Moreover, high /i y u/ differ in their restricted functional load from low /a/ and /e/ as well as in their morphonological behaviour (only low vowels are involved in phonological lengthening processes, but all qualities can be affected by systematic vowel shortening). Mid /o/ and /ø/-qualities show an intermediate behaviour and a higher functional load than high vowels: short vowels tend to be more centralized and the quantity distinction is connected to quality to some extent for mid qualities (Mády 2010).

To summarize, qualitative differences between long vs. short vowels are strongest with low vowels and least strong with high vowels that rather differ in quantity than in quality though this opposition is eventually becoming unstable in more colloquial speech (Mády 2010). An alternative account to the traditional long-short dichotomy that is not sufficient especially for low vowels which also differ in terms of spectral cues is proposed here to describe the traditional opposition of long vs. short Hungarian vowels as a difference in constriction degree together with – for front vowels – a distinction of a pre- vs. mid-palatal constriction location.

Typologically, Hungarian can be characterized as a syllable-timed language. Typical characteristics are a fixed accent on the word-initial syllable and the preference for a trochaic rhythm (Dressler & Siptár 1989: 48). The main word stress is assigned to the initial syllable of a word. Accented and unaccented syllables are similar in their phonological and phonetic shape. Each syllable contains only one vowel. Accented vowels are not significantly longer than unaccented vowels and syllable peaks in unaccented syllables are not significantly shorter than in accented syllables.

Vowel bleaching to schwa is extremely restricted in Hungarian, since it would obscure Hungarian vowel harmony. Thus, the syllable timing and the agglutinating character favour each other (Dressler & Siptár 1989: 48).

A specific phenomenon in Hungarian is *vowel harmony*. The relevant feature for Hungarian vowel harmony is [+/– back] regulating the occurrence of vowels in a word. In contemporary Hungarian, vowel harmony affects actively only suffixes, with important exceptions in the class of verb roots with /i/ or /i:/ that take back vowel suffix forms. In many Hungarian roots, vowels within a word are either all front ([ɛ eː i iː ø øː y yː]) or all back ([ɒ/ɔ aː o oː u uː]), while "mixed" words are rather uncommon and often of foreign origin (e.g. *telefon* 'telephone', *kosztüm* 'costume', *sofőr* 'chauffeur'). Since most 'mixed' word forms and morphemes contain front unrounded vowels, (/i/, /i:/, /e/ and /e:/) together with back vowels (*papír* 'paper', *patika* 'pharmacy'), these /i/- and /e/-qualities are considered to be 'neutral' (Kiss 2003: 908) or even 'transparent' (vs. 'opaque' vowels like e.g. /a/ or /ü/) letting harmony pass through them (Benus, Gafos & Goldstein 2003). This leads to a classification of 'front harmonic', 'back harmonic' and 'neutral' vowels (Siptár & Törkenczy 2000: 62f).

Hungarian vowel harmony is "stem controlled", this means that the harmonic status of the word stem controls that of affixes. Suffix harmony can be described as morphonological due to post-lexical rules and does not represent constraints on pronounceability or perceptibility (Dressler & Siptár 1989: 31). Suffixes containing a front vowel combine with stems containing front vowels, suffixes with back vowels are associated with back vowel roots. Several suffixes occur in two or three variants: The affixation of suffixes occurs in three variants, back (labial), front labial, front non-labial, e.g. the suffix for accusative occurs as -*ot/-öt/-et*, and is sensitive to labiality and therefore called *labial harmony*.

10.7.3 Contrastive analysis

As discussed above, the opposition of Hungarian /a/ - /a:/ is commonly described as an opposition in quality and quantity. Many studies transcribe Hungarian /a/ as [ɔ] (e.g. Dressler & Siptár 1999; Gósy 2004), reflecting the higher degree of aperture for /a/ than for /a:/, but it is important to note that its realization is not equivalent with German [ɔ] (Mády 2001). Other researchers prefer [ɑ] (e.g. Szende 1999) or [ɒ] (e.g. Mády 2001). Compared to German /ɔ/, Hungarian /a/ shows higher F1 values (Figure 10.42) due to the relative higher degree of aperture and less lip rounding for Hungarian /a/ than for German /ɔ/ suggesting a phonetic difference between German /ɔ/ and Hungarian /a/. Mády (2001: 38) describes wrong pronunciations of /a/ observed with German learners of Hungarian, who substitute Hungarian

/a/ [ɒ] with German [ɔ] in pronunciation, which is perceived as equivalent to Hungarian /o/ by Hungarian natives rather than to /a/ (Mády 2001: 37f).

In both languages, German and Hungarian, we find an opposition of vowel phonemes that is commonly described as opposition in phonemic length. In German this quantitative opposition is associated with other qualitative differences commonly referred to as tenseness/laxness or – as is argued in the present study (see section 4.7.6 and 5.4) – differences in constriction degree at least for non-low vowels.

Though vowel quantity is commonly considered to be distinctive in Hungarian, Herok (2001) posits the existence of a tense/lax contrast for Hungarian as well, noting one fundamental difference between German and Hungarian concerning the association of phonemic length and 'tenseness' in the opposition of long vs. short vowels: For Hungarian, Herok posits short /a/ as the lax counterpart (with "higher position of the tongue") of tense /a:/, whereas for German no "tense/lax" differentiation is assumed for a-vowels (Herok 2001: 85f). Moreover, he describes a higher degree of lip spreading for Hungarian /e:/ and /a:/ as opposed to /ɛ/ and /a/. If the contrast between short /e/ and /a/ vs. long /e:/ and /a:/ is viewed to be based on 'tenseness' ([e:] vs. [ɛ], [a:] vs. [ɔ/ɒ]), then in Hungarian the tense/lax contrast is considerably stronger with lower vowels than with higher ones, whereas in German it is stronger with higher vowels.

Herok (2001) moreover refers to regional differences in German concerning the quantitative and qualitative differentiation of long tense and short lax vowels. North German varieties show considerable qualitative differences between corresponding tense and lax vowels, whereas Southern German varieties exhibit less differentiation in "tenseness" with high and mid vowels. Referring to differences between "Standard German" and regional varieties of Standard German in Austria or Switzerland and the fact that a lower degree of qualitative difference of the long/short pairs is observed in Southern Standard varieties of German, Herok considers durational differences in these varieties to be more distinctive than differences in quality. Herok (2001) predicts that Hungarian learners will therefore rely on quantitative differences in vowel identification in Southern German varieties. Following this approach, it is expected that Hungarian learners of German may misidentify high long vowels as short or short vowels as long due to a misinterpretation of variability in duration, an expectation that has to be experimentally verified with controlled stimuli varying systematically in one of the expected parameters (duration or formant frequency F1, F2, F3).

Therefore, Herok expects Hungarian learners of German to be relatively insensitive in perception to [tense/lax] (or [+/–constricted]) differences with high vowels and argues that the "auditive filter" of L1 Hungarian will lead to an over-estimation of quantitative differences in L2 German, especially for high vowels. For high vowels, Herok (2001) expects differences in vowel duration to be more salient to Hungarian learners especially with high vowels, while for German listeners differences in duration are expected to be less relevant than differences in quality (Herok 2001: 87; Sendlmeier 1981; Mády 2001: 36). Perceptual problems with the differentiation of the long/short contrast in Hungarian and German are also reported by Mády (2001: 36f) and are also explained by the higher relevance of quality differences for high vowels, e.g. /iː/ vs. /ɪ/, for German listeners compared to Hungarians. However, as Mády (2010) describes, quantitative differences for high vowels are not maintained in all speaking styles and the long-short opposition of high vowels is becoming instable in colloquial speech and are possibly on the way to perceptual loss and neutralization of contrast. Experimental production and perception studies (e.g. Mády & Reichel 2007; Mády 2010) show that in production low /e/ and /a/ are clearly differentiated in quality and also quantity and that high /i y u/ are hardly differentiated in quality and inconsequently in duration. The results suggest that vowel quantity is perceived or weighted differently in high, mid and low vowels in Hungarian. While for low vowels spectral cues are more salient than durational ones in perception, the role of spectral cues in the perception of high vowels is very limited; mid vowels show an intermediate behaviour (Mády 2010).

Following Herok's (2001) hypothesis, we may expect that quantitative contrasts will be perceptually easier to Hungarian learners of German especially with high vowels (at least for stimuli from a Southern variety of German), whereas qualitative differences will be perceptually less salient with high vowels than with low vowels. The quantitative contrast of German low a-vowels may be difficult, since Hungarian listeners use more qualitative features to differentiate /a/ - /aː/. For high vowels we can expect more perceptual sensitivity to quantity than to quality distinctions, i.e. length should be recognized correctly, while qualitative differences in constriction degree may cause perceptual problems.

To summarize, the differential role of spectral and durational cues in vowel identification in Hungarian and German, which has a very similar vowel inventory but apparently another type of cue weighting, is an interesting issue, though the present study can give only limited insights into this question. To investigate this issue, stimuli that are controlled for the decisive acoustic cues are necessary to get insights into the role and relative perceptual weight of quantity and quality differences for German and Hungarian listeners.

Diphthongs do only occur in Hungarian dialects. In pronunciation of L2 German, Hungarians show a strong tendency towards monophthongization of German diphthongs. We may therefore expect the hypercorrect perceptual diphthongization of German vowels, i.e. the substitution of monophthongs with diphthongs in L2 perception.

However, due to the "high resolution" of the Hungarian vowel space and the many similarities in number and quality of vowels in German and Hungarian, relatively few problems especially with more experienced learners of German are expected. The differentiation of long constricted vs. short unconstricted high, mid and low vowels will be an issue of special interest in the interpretation of the experimental data.

10.7.4 Results and discussion

26 Hungarian subjects were tested, 11 beginners, 11 advanced and 4 very advanced learners. The beginners were tested in a private language school in Budapest. The advanced learners are second year students participating in a teacher education program for German at the ELTE Budapest. The very advanced learners are Hungarian native speakers with a very high level of fluency in German who had lived in Vienna for several years at the time the experiment took place.

The data set discussed here consists of 6136 valid responses (66 missing answers). For each vowel category, 410 to 414 valid responses were delivered.

10.7.4.1 Difficulties

The analysis shows that the id_wrong scores are as expected very low for L1 Hungarian listeners (see Figure 10.43) due to the highly resolved system of the listeners' native language and the numerous parallels in phonemic distinctions. However, as the boxplot diagram in Figure 10.44 shows there are several outliers mostly found with beginners. The only exception is /u:/ that is substituted with /ʊ/ by two advanced learners. The highest range of inter-speaker variation with respect to id_wrong scores is observed for /ɛ:/.

Least problems are observed with /i:/ and /u:/ followed by /a:/ and /y:/. The category with the lowest id_correct score is /ɛ:/ due to its marginal status in German. All other categories show considerably less difficulties. With front rounded qualities, less difficulties are observed with /y:/ and /œ/ than with /ø:/ and /ʏ/.

Within the class of long constricted vowels, we can observe that /i:/, /u:/, /a:/ and /y:/ have very high id_correct scores, whereas with /e:/, /o:/ and /ø:/ lower id_correct scores are observed. Interestingly, among short unconstricted vowels, /œ/ and /ɪ/ show slightly higher id_correct scores than /ɛ/ and /ʊ/. /ʏ/ and /ɑ/ have the lowest id_correct score within the class of short unconstricted vowels.

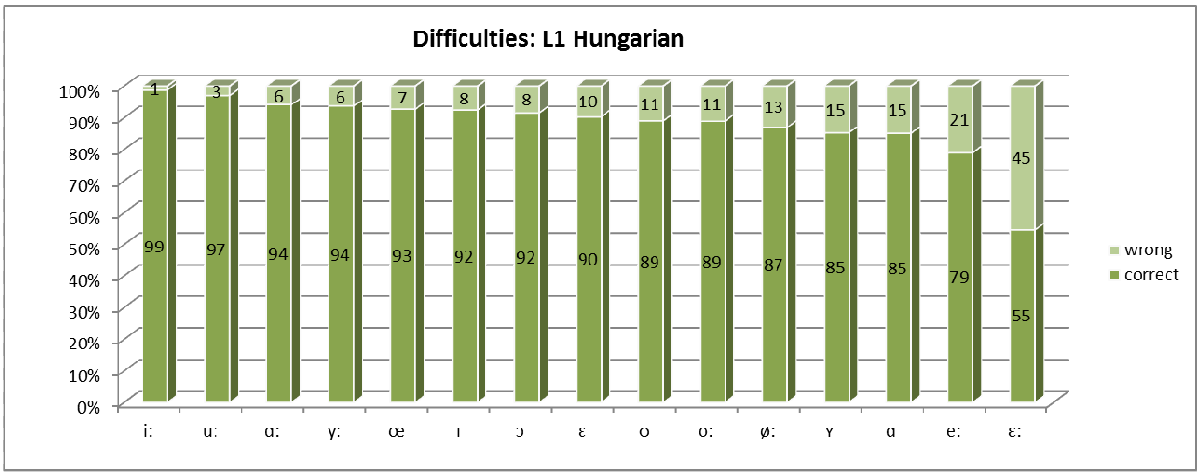

Figure 10.43: Percentage correct and wrong identifications for Hungarian listeners (N=26, 6136 responses)

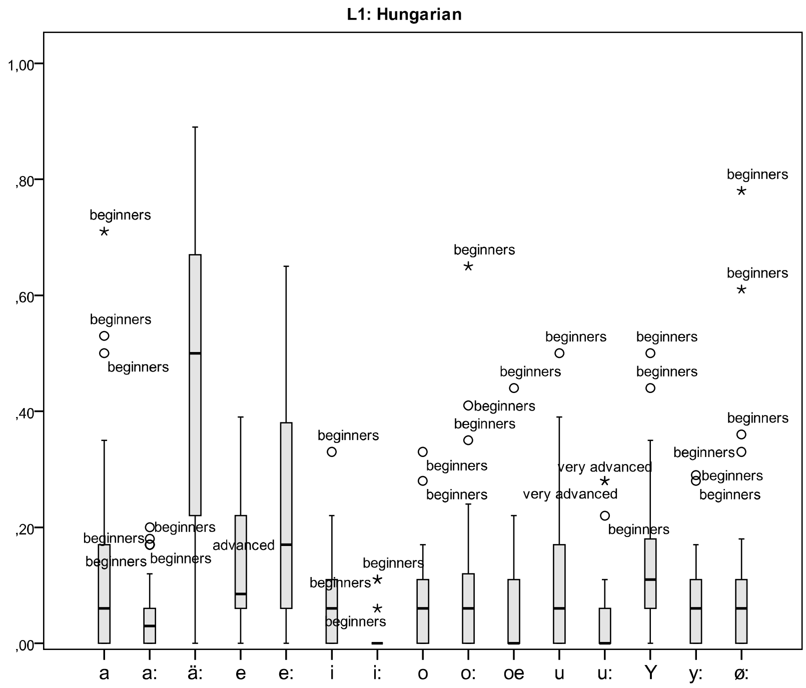

Figure 10.44: Id_wrong scores for Hungarian listeners (N=26, 6136 responses)

The listeners' mean id_wrong score was 12% (9% standard deviation), the median was 9%. The listener with the highest percentage of correct answers had an id_wrong of only 3%, the highest id_wrong score was 40% wrong responses.

10.7.4.2 Preferences

The preference patterns observed with L1 Hungarian learners differ from many other languages in this study. The preference scores for monophthong categories vary within a range of only 3.3% (min /ɛ:/ 4.6%, max /e:/ 7.9%). Hungarian listeners show very balanced preference scores between 6 and 7.3% for most monophthongal categories, with the only exception of /e:/ (7.9%) and /ɛ:/ (4.6%). /ɛ:/ is the least preferred category. 33% of /ɛ:/-stimuli are categorized as /e:/. Id_scores for diphthongs vary between 0.6% for <ei> and 0.1% for <eu>.

Long constricted /e: y: u: i: ɑ:/ are selected more frequently than their short unconstricted counterparts, while short unconstricted /œ ɔ/ are slightly more preferred as response category than long constricted /ø:/ and /o:/.

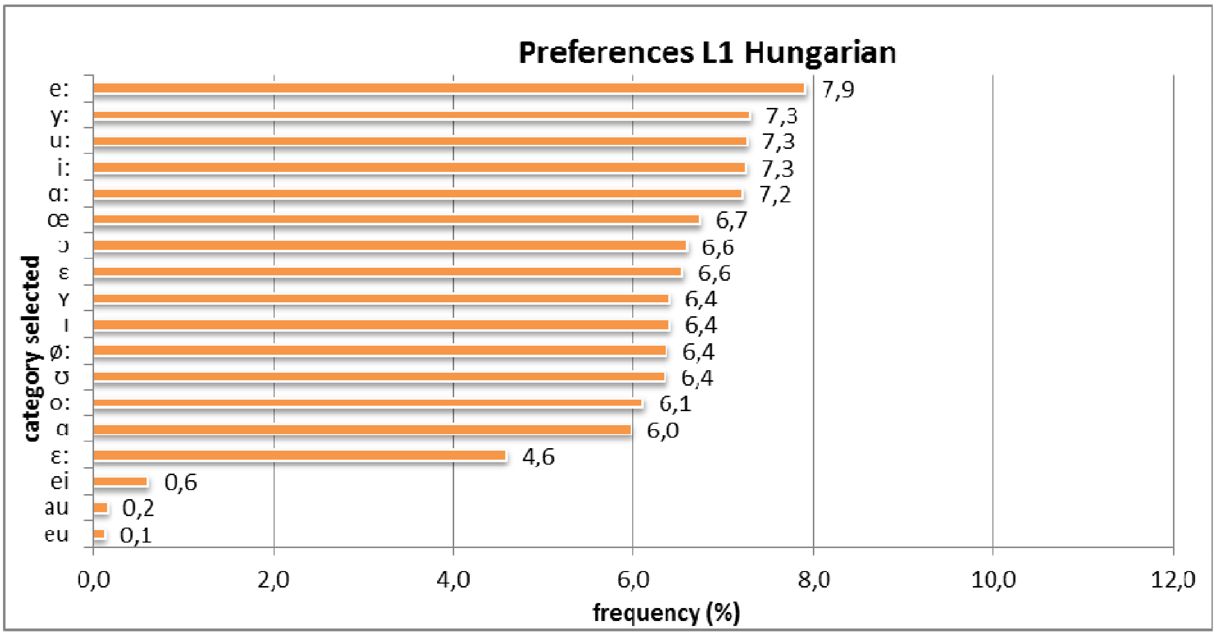

Figure 10.45: Preference scores for L1 Hungarian listeners

10.7.4.3 Patterns of confusion

The listeners' responses in the confusion matrix in Table 10.42 show that category confusion is rather limited for Hungarian learners of German. However, four major areas of confusion are observed: (1) *a-vowels* are confused with respect to phonemic length but not with respect to quality, (2) *e-vowels* are confused with respect to quality and quantity, (3) *back rounded vowels* are confused with respect to constriction location as well as phonemic length and constriction degree, (4) *front rounded vowels* are confused with respect to constriction location as well as phonemic length and constriction degree.

a-qualities are very stable: /ɑ/ (85% correct) is identified as /ɑ:/ in 14% of the cases and /ɑ:/ (94% correct) is identified as short /ɑ/ in only 4% of the cases. No substitutions of /ɑ/ with /ɔ/ are observed. However, in three cases /ɔ/ is incorrectly identified as /ɑ/ by two beginners (1%). This can be interpreted as an interference of the Hungarian writing system, where the grapheme <a> corresponds to [ɒ/ɔ].

%	ɑ	ɑ:	ɛ:	ɛ	e:	ɪ	i:	ɔ	o:	ʊ	u:	œ	ø:	ʏ	y:	‹ei›	‹eu›	‹au›	wrong	corr
a	**85**	14																	15	**85**
a:	4	**94**																1	6	**94**
ɛ:			**55**	5	33											6			45	**55**
ɛ			4	**90**	5											1			10	**90**
e:			10	1	**79**	2	7									1			21	**79**
ɪ					1	**92**	3							2					8	**92**
i:							**99**									1			1	**99**
ɔ	1							**92**	8										8	**92**
o:								3	**89**	3	4							1	11	**89**
ʊ								3		**89**	6								11	**89**
u:										3	**97**								3	**97**
œ												**93**	7						7	**93**
ø:												1	**87**	2	8				13	**87**
ʏ												6	1	**85**	7				15	**85**
y:														6	**94**				6	**94**
Σ	*6*	*7*	*5*	*7*	*8*	*6*	*7*	*7*	*6*	*6*	*7*	*7*	*6*	*6*	*7*	*.6*	*.1*	*.2*	*12*	*88*

Table 10.42: Confusion matrix for L1 Hungarian. Correct (in bold) and wrong identifications in % for L1 Hungarian listeners (N=26, 6136 responses) (presented stimuli in rows, selected categories in columns in semi-orthographic representation, values < 0.5 not indicated)

With *e-qualities*, more problems are observed. The most stable category is /ɛ/ (90% correct) which is incorrectly identified as /e:/ (5%), /ɛ:/ (4%) or <ei> (1%). Most difficulties are observed with /ɛ:/ (55% correct), which is incorrectly identified as /e:/ in 33% of the cases or as /ɛ/ (5%) or <ei> (6%). The confusion of /e:/ and /ɛ:/ is also observed with /e:/-stimuli (79% correct) that are identified less frequently as /ɛ/ (only 10%). Responses for /e:/ (and also for /o:/) show the highest dispersion of responses over different categories. /e:/ is also substituted with /i:/ (7%), /ɪ/ (2%), /ɛ/ (1%) and <ei> (1%).

German i-vowels show least problems. /i:/ is identified correctly in 99% of the cases and very occasionally substituted with <ei> by two beginners.

/ɪ/ (92% correct) is incorrectly identified with /i:/ (3%), /ʏ/ (2%) and /e:/ (1%).

While e-qualities are incorrectly identified as i-qualities, this pattern is very uncommon in the other direction (only 1% /e:/ categorizations for /ɪ/) and can be interpreted as a uni-directional attraction effect of e- to i-vowels.

For /ɔ/ (92% correct) most problems are due to a wrong categorization concerning phonemic length (/o:/ (8%). In 3 instances, /ɔ/ is identified as /ɑ/ (1%, two beginners).

/o:/ (89% correct) is one of the categories that shows more inter-category variation in the listeners' responses (/u:/ (4%), /ɔ/ (3%), /ʊ/ (3%)), which is mostly due to incorrect categorizations with respect to constriction location and/or constriction degree. Only one of the beginners chooses <au> as response option (1%).

German u-qualities are particularly easy for Hungarian learners. For /u:/ (97% correct) only misidentifications with /ʊ/ (3%) are observed. /ʊ/ (89% correct) is identified as /u:/ (6%) and as /ɔ/ in 3% of its occurrences.
The class of *front rounded vowels* causes no severe problems in perception for Hungarian learners. /y:/ (94% correct) is the most stable category and is incorrectly substituted only by /ʏ/ (6%), followed by /œ/ (93% correct) which is only confused with its long constricted counterpart /ø:/ (7%).
/ø:/ and /ʏ/ cause more difficulties in perception for Hungarian learners: /ʏ/ (85% correct) can be substituted by /y:/ (7%), /œ/ (6%) or /ø:/ (1%). With /ø:/ (87% correct) we observe that phonemic length is identified correctly in 95% of the cases where either /ø:/ or /y:/ (8%) is chosen as response option. A wrong identification of the constriction place is comparatively seldom (/ʏ/ (2%) and /œ/ (1%)). Again, we find here a unidirectional attraction of /y:/ for /ø:/-stimuli that is not observed in the opposite direction.
To summarize, we find patterns of asymmetric attraction with the higher constricted categories, a pattern that is strongest with /ø:/ > /y:/ (8 vs. 0%), followed by /e:/ > /i:/ (7% vs. 0%), and /o:/ > /u:/ (4% vs. 0%). On the other side, for short unconstricted qualities, an asymmetric pattern in the opposite direction is observed only for rounded vowels: /ʏ/ > /œ/ (6%) vs. /œ/ > /ʏ/ (0%) and /ʊ/ > /ɔ/ (3%) vs. /ɔ/ - /ʊ/ (0%), while for /ɛ/ and /ɪ/ substitutions in neither direction are observed. Another pattern of asymmetry that is also observed in most other languages subgroups is the attraction of /e:/ for /ɛ:/-stimuli. A higher percentage of substitutions with diphthongs is only observed for /e:/ > <ei> (6%).

10.7.4.4 Similarity and distance

Compared to all other languages in this study, the sim_scores for Hungarian listeners (see Table 10.43) are very low due to the restricted number of wrong categorizations, making an interpretation relatively easy.
A comparison of the sim_scores shows the highest value for the /e:/ - /ɛ:/ contrast (0.324). High sim_scores are also observed with contrast pairs that differ only in phonemic length (/ɑ/ - /ɑ:/ (0.100), /ɛ/ - /ɛ:/ (0.062)) or in phonemic length and constriction degree, where sim_scores vary between 0.072 and 0.019: /ʏ/ - /y:/ (0.072), /ɔ/ - /o:/ (0.062), /ʊ/ - /u:/ (0.047), /œ/ – /ø:/ (0.045), /ɛ/ - /e:/ (0.033).
Category confusion with respect to different constriction location is observed either with contrasts of mid- vs. pre-palatal vowels or of velar vs. uvular vowels: /ø:/ - /y:/ (0.043), /i:/ -

/e:/ (0.040), /œ/ - /ʏ/ (0.033), /ø:/ - /ʏ/ (0.021), /u:/ - /o:/ (0.021), /ʊ/ - /ɔ/ (0.019), /ɪ/ - /e:/ (0.018), /ʊ/ - /ɔ/ (0.017). Very low sim_scores below 0.01 are observed with other contrasts.

V	ɑ	ɑ:	ɛ:	ɛ	e:	ɪ	i:	ɔ	o:	ʊ	u:	œ	ø:	ʏ	y:
a	1.000														
ɑ:	0.100	1.000													
ɛ:	0.002	0.003	1.000												
ɛ	0.000	0.000	0.062	1.000											
e:	0.001	0.000	0.324	0.033	1.000										
ɪ	0.000	0.000	0.002	0.004	0.018	1.000									
i:	0.000	0.000	0.000	0.000	0.040	0.019	1.000								
ɔ	0.004	0.000	0.000	0.000	0.000	0.000	0.000	1.000							
o:	0.000	0.000	0.000	0.000	0.000	0.001	0.000	0.062	1.000						
ʊ	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.019	0.017	1.000					
u:	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.021	0.047	1.000				
œ	0.000	0.000	0.000	0.003	0.000	0.001	0.000	0.000	0.000	0.001	0.000	1.000			
ø:	0.000	0.000	0.002	0.000	0.000	0.001	0.000	0.000	0.001	0.000	0.003	0.045	1.000		
ʏ	0.000	0.000	0.000	0.000	0.000	0.012	0.000	0.000	0.000	0.003	0.000	0.033	0.021	1.000	
y:	0.000	0.000	0.002	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.003	0.000	0.043	0.072	1.000

Table 10.43: Similarity matrix for L1 Hungarian listeners

10.7.4.5 The perceptual vowel map for L1 Hungarian listeners

From the similarity scores calculated by Shepard's formula, distance scores and non-metric MDS solutions were derived. For the non-metric MDS solutions, RSQ values of only 0.148 in dimension 1, 0.671 for the two-dimensional solution, and 0.799 for the three-dimensional solution are obtained.

The one-dimensional solution has a very low RSQ value of only 0.148 and can therefore not be considered as reliable enough. The RSQ values given for the two-dimensional MDS solution are relatively low as well. This is due to the high number of correct identifications by Hungarian listeners and the high portion of zero values obtained in the similarity matrix that are transformed to very high values in the distance matrix. However, a comparison of the confusion matrix and the MDS visualization provides interesting insights in the perceived similarity of German vowels by Hungarian learners of German.

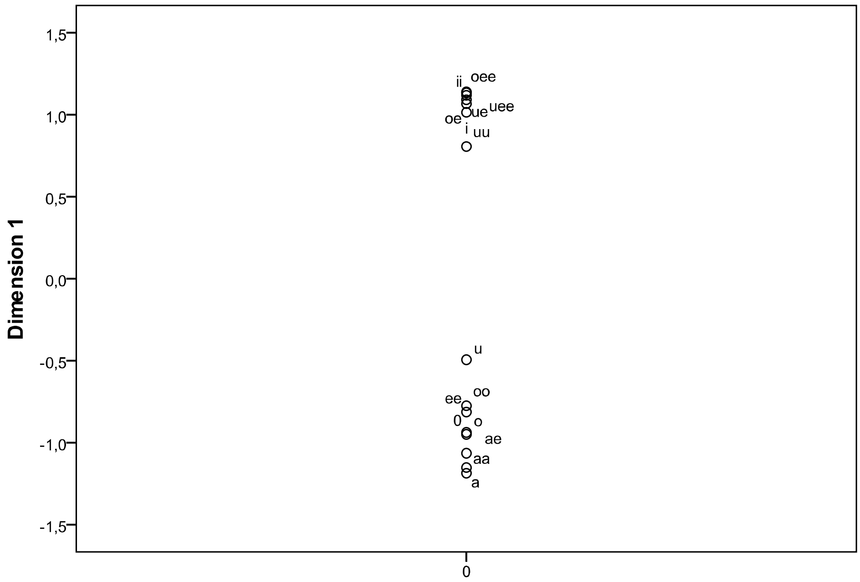

Figure 10.46: One-dimensional MDS representation of the perceptual map of German vowels for L1 Hungarian listeners (N= 26, 6136 responses, RSQ .148)

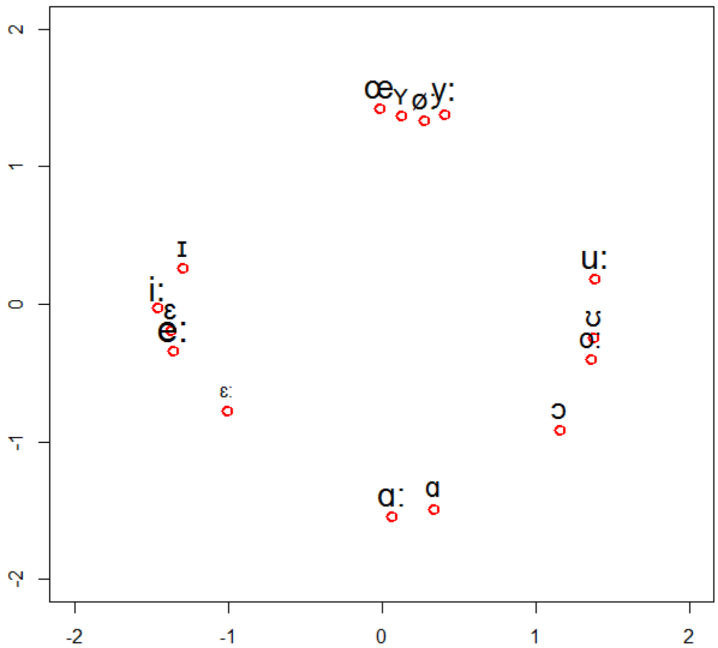

Figure 10.47: Two-dimensional MDS representation of the perceptual map of German vowels for L1 Hungarian listeners (N= 26, 6136 responses, RSQ .671)

The two-dimensional perceptual vowel map for Hungarian listeners in Figure 10.47 (RSQ .671) shows the four major areas of category confusion discussed in section 10.7.4.3. As the

two-dimensional MDS solution shows, the differentiation of front rounded vowels from front unrounded as well as back rounded vowels represents no problem to Hungarian learners of German. However, the differentiation of vowel sounds within these three vowel classes is difficult in some cases, especially for beginners. Least difficulties to differentiate qualities within a given class are observed for back rounded vowels. Here, /u:/ and /ɔ/ are differentiated best, i.e. least inter-category confusion is observed with these two categories. More difficulties are observed with /o:/ and /ʊ/. While short unconstricted /ʊ/ is substituted either with short unconstricted /ɔ/ (3%) or long constricted /u:/ (6%) but not with /o:/, we observe perceptual substitutions of /o:/ with /ɔ/ (3%) as well as with u-qualities (/u:/ 4%, /ʊ/ 3%).
Within the class of *front rounded* vowels, /y:/ and /œ/ are the least difficult as well as the most preferred categories. Both are substituted only with qualities at the same constriction location, i.e. /œ/ ~ /ø:/ and /y:/ - /ʏ/, whereas for /ø:/ and /ʏ/ perceptual substitutions with mid-palatal as well as pre-palatal vowels are observed. Since no substitutions with vowels articulated at other constriction locations are observed, the cluster of front rounded vowels is clearly set apart from other vowel classes.
Within the class of *front unrounded* vowels, stronger asymmetries with respect to response preferences are observed, especially for /ɛ:/. The most preferred response options are long constricted /e:/ and /i:/. /ɛ:/ is only substituted with other e-vowels, which is reflected in the smaller distance of /ɛ:/ to /e:/ and /ɛ/. The MDS solution reflects what has also been evident in the confusion matrix in Table 10.42 showing that the differentiation of mid-palatal constricted /e:/ from pre-palatal i-vowels as well as long unconstricted /ɛ:/ is difficult for Hungarian speakers and that /i:/ seems to function as a "perceptual attractor" for long constricted /e:/ (7%), while pre-palatal i-vowels are better differentiated from mid-palatal e-vowels.

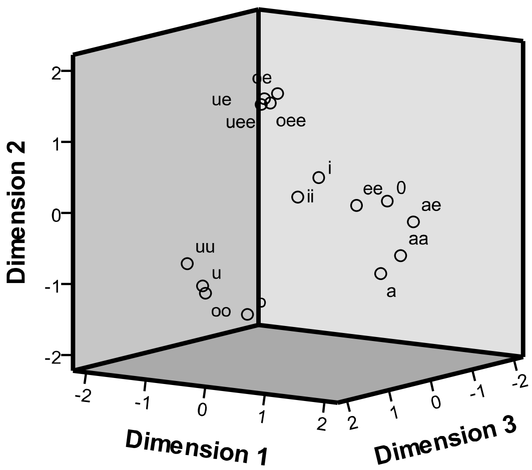

Figure 10.48: Three-dimensional MDS representation of the perceptual map of German vowels for L1 Hungarian listeners (N= 26, 6136 responses, RSQ .799)

The three-dimensional MDS solution gives a more differentiation reflection of similarity relationships between front non-rounded vowels, whereas the positions for rounded vowels are quite similar to that in the two-dimensional solution.

10.7.5 Summary

The number of wrong identifications as well as asymmetries in preference patterns are weaker in the Hungarian sample than in many other language sub-groups of this study with the only exception of /ɛ:/ that is considerably less often selected as response category (see also Figure 10.45). Lest difficulties are observed with /i: u: ɑ: y:/ and most difficulties are observed with /ɛ:/ (55% correct) and /e:/ (79% correct).

To summarize, Hungarian learners have no difficulties to differentiate the four major classes of pharyngeal a-vowels, palatal non-rounded vowels, palatal rounded vowels and back rounded. However, the differentiation of uvular o- and velar u-vowels and of mid-palatal vs. pre-palatal vowels causes some difficulties in perception. Asymmetry in perceptual substitutions is especially observed with pre- vs. mid-palatal long constricted qualities, i.e. with /e:/ ~ /i:/ and /ø:/ ~ /y:/, where perceptual substitutions of mid-palatal with pre-palatal vowels take place, but substitutions in the opposite direction are not observed. With long constricted qualities, the category that has a smaller degree of aperture and is more fronted seems to function as a perceptual attractor: /e:/ > /i:/ (7% vs. 0%), /o:/ > /u:/ (4% vs. 0%), and /ø:/ > /y:/ (8% vs. 0%). The opposite tendency is observed with short unconstricted rounded qualities /ʏ/ > /œ/ (6% vs. 0%) and /ʊ/ > /ɔ/ (3% vs. 0%) but not for front non-rounded vowels. The difficulties can be hypothesized to arise by the prior influence of [rounding] and differences in [constriction degree]. The differentiation of short unconstricted pre-palatal and velar vowels vs. long constricted mid-palatal and uvular vowels causes perceptual confusion with respect to the location of constriction.

Additional confusion among German e-vowels is provoked by long unconstricted /ɛ:/ which has a marginal status in the functional oppositions of German and is realized as [e:] in many parts of the German speaking countries and is moreover represented with the more marked grapheme <ä> in German. It is not surprising, that this category is the most difficult for Hungarian learners of German.

The differentiation of e-qualities in German is one of the major points of difficulties for Hungarian learners of German. The phonemic contrast of Hungarian long <é> vs. short <e> in Hungarian is very similar to the German contrast pair, however, the phonetic differentiation is somewhat different from German. The articulatory data presented in Bolla (1995: 146, 177,

199) does not provide clear evidence that the contrast of pre- vs. mid-palatal constriction location may be realized in Hungarian as in German, an assumption that seems more probable because of the rich inventory of front vowels in Hungarian, though it does not become clear whether this difference is used systematically.

As the data show, Hungarian learners have generally less difficulties to differentiate German vowel qualities correctly, which is largely due to the highly resolved vowel system of their native language Hungarian, which has the same functional oppositions as German. However, the phonetic correlates of Hungarian vowel contrasts differ in some respect from the German distribution, specifically in contrasts that are traditionally described as long vs. short oppositions, where the relative weight of duration and spectral cues is differently distributed in both languages for high, mid and low vowels. While in Hungarian high vowels rather differ in duration than in spectral cues and low vowels differ more in spectral than in durational cues, for German, specifically for North German, the opposite pattern is observed (Herok 2001, Mády & Reichelt 2007; Mády 2010). However, the data collected in the present study does not provide sufficient information on how these differences between the two languages affect the listeners' performance in an identification task in L2 German. Further research will have to concentrate on this question using stimuli that are systematically controlled for spectral and durational properties.

10.8 Mandarin/Standard Chinese

10.8.1 General characteristics

The Chinese language family is a major branch of the Sino-Tibetan language family. The varieties of Chinese are commonly grouped into seven major dialect groups: Mandarin, Wu, Yue, Min, Hakka, Xiang, and Gan. Each of these groups consists of several dialects which differ in phonology, syntax and vocabulary. Traditionally, different varieties of Chinese are called "dialects", even if mutual intelligibility is not given in many cases due to dialect-specific differences.

"Mandarin" actually refers either to (1) a *group of dialects* that are spoken by around 70% of the Chinese population in the Northern and South-West regions of China (Lin 2007: 1; Li & Thompson 1981), or 2) to *Standard Chinese*, the standard dialect (Duanmu 2005).

Standard Chinese (SC), also called "Mandarin" or "Pǔtōnghuà" in China and "Guóyǔ" in Taiwan, is the official standard language of China, Taiwan and Singapore. It is a standardized form based on the Northern dialects of Chinese, with the Beijing phonological system as its norm of pronunciation (Lin 2007: 3). Other terms for SC are Standard Mandarin, Mandarin Chinese or simply Mandarin.

During the language reform program in the mid-1950s, a codification and subsequent propagation of Standard Chinese took place. The Pinyin romanization system (approved in 1958) was created for disseminating standard pronunciation. SC is used in schools and universities as well as in national broadcast and official functions in China, while in regional broadcasts and everyday communication local dialects are used in addition to SC. Thus, diglossia is a common feature in China and Taiwan: people speak one or more dialects together with Standard Chinese.

In spite of the high range of variation within the Chinese dialects, all dialects can be typologically described as analytic and tonal in nature. Classical Chinese is mostly described as a monosyllabic language. However, Modern Mandarin has quiet a large number of polysyllabic words and can thus no longer be described as a monosyllabic language (Li & Thompson 1981). 95% of morphemes are monosyllabic in Modern Chinese, but about half or even more than half of the words consist of more than one morpheme and are thus polysyllabic (Lin 2007: 5).

Typologically, Mandarin Chinese is a syllable-timed language (Lin & Wang 2008). Each syllable is associated with one tone. However, not all tones are realized with all possible syllables. Distinctive *tones* can be described as contrastive patterns of voice pitch which are associated with a syllable.

The Chinese system uses four tones: high (55), high rising (35), dipping/falling-rising (214), high falling (51). The numbers in brackets refer to a scale of five tone levels, 1 being the lowest and 5 the highest tone (Li & Thompson 1981: 8).

Mandarin has a large inventory of homophone morphemes, since constraints on syllable structure restrict the number of possible syllables to around 400. These 400 syllables can be associated with one of the four different tones, resulting in around 1200 tonally differentiated syllables that can occur as morphemes.

An issue frequently discussed is the relationship of Chinese morphemes, syllables, tones and words. In Mandarin, each "word" equates with a character in the writing system and each character corresponds to one syllable in the spoken language. Syllables and morphemes are considered as equivalent units of speech. Most morphemes are monosyllabic and each syllable generally bears a tone. Each word may consist of one or more morphemes and can hence be monosyllabic or polysyllabic.

The writing system of Chinese is commonly described to be a logographic system. However, Mandarin could rather be characterized as a morpheme-graphic system rather than a logographic system (Třísková 2008: 511), since characters in the writing system represent morphemes and since words are not always monosyllabic.

10.8.2 Phonological and phonetic description

The phonological system of Standard Chinese (SC) is based on the Beijing dialect. However, the pronunciation of SC is by no means uniform but covers a range of different pronunciations.

The traditional approach to the phonological description of Chinese describes the phoneme inventory in terms of initials, finals and tones. *Initials* represent the consonantal beginning of a syllable. Since there are no consonant clusters, the consonantal beginning of a syllable consists of only one consonant. Mandarin allows no clusters in initial and final position. Syllables without an initial consonant are described as containing a "zero" initial.

The system of Mandarin initials can be described as a system of 22 initials[108]. The phonological or allophonic status of some of these initials is a matter of debate, various

[108] Duamnu (2000: 44ff) posits 24 underlying phonemes: 19 consonants and 5 vowels.

solutions have been suggested, e.g. for the group of alveo-palatals (cf. Duanmu 2002: 33, 44ff; Lin 2007: 47ff).

The syllable-initial position is quite unrestricted with only /ŋ/ prohibited, consonants in the coda position are highly restricted with only /n/, /r/ and /ŋ/ allowed. Traditional accounts describe only two consonants that can occur as finals in a Mandarin syllable, the alveolar nasal [n] and the velar nasal [ŋ].

	bilabial	labio-dental	dental	alveolar	post-alveolar	palatal	velar
Plosives	p pʰ			t tʰ			k kʰ
				ts tsʰ	t̠ʃ t̠ʃʰ	tɕ tɕʰ	
Nasals	m		n				ŋ
Fricative		f		s	ʃ̠	ɕ	x
Approximant	w				ɹ̠	j	
Lateral Approximant				l			

Table 10.44: The consonant system of Standard Chinese (Beijing) (Lee & Zee 2003: 109)

The number of vowel phonemes and their allophones in Standard Chinese, the number of diphthongs and the status of triphthongs are unanimously discussed (for a detailed overview of context-sensitive allophonic variation, see e.g. Marshall Howie (1976: 6ff), for a different account see Lin 2007 and Duanmu 2000).

Table 10.44 shows the consonant inventory of Standard Chinese (Beijing) in IPA symbols as described by Lee & Zee (2003: 109).

Finals are the part of the syllable without the initial. In a traditional account, finals correspond to vocalic elements and can consist of monophthongs, diphthongs or triphthongs.

Phonological descriptions of monophthongs differ considerably with respect to the number and quality of vowels and their allophones: An inventory of five vowel phonemes is assumed by Duanmu (2000, 2005) and Lin (2007), Lee & Zee (2003) assume seven vowels /i y u ɤ a ə ɚ/, and Hachenberg (2003) and Hunold (2009) posit nine vowels. Lee & Zee (2003: 111) present a phonetic description of the Standard Chinese vowel system consisting of seven vowels /i y u ɤ a ə ɚ/: /i y u ɤ a ə/ occur in meaningful (C)V monosyllables. Duanmu (2000, 2005) and Lin (2007) posit a 5-vowel system consisting of /i y u a ə/. All other vowel qualities are described as phonetic variants or allophones of these 5 phonemes governed by phonological and phonetic rules. While /ə/ can change in frontness and rounding depending on the phonetic environment, /a/ can change only in frontness but not in rounding (Duanmu 2005).

The vowels [i u a y] occur in open syllables as well as in syllables closed by a nasal and are described as near to the cardinal vowels in open syllables, where [a] is slightly retracted and [ɤ] is slightly diphthongized from [ɤ] to [ʌ] and only occurs in open syllables. In syllables closed by a nasal, [i y] are close to the corresponding cardinal vowels, [a] is slightly fronted, and [u] is realized as [ʊ] (Zee & Lee 2001).

Duanmu (2000: 25) and Lin (2007: 70f) assume three high Vs /i u y/ with their allophones [j w ɥ] which occur as non-syllabic counterparts of /i u y/. When followed by another vowel, these vowels behave like glides [j w ɥ]. /i/ and /u/ can also follow a non-high vowel to form a diphthong.

Especially for the number of allophones of mid vowels there is considerable disagreement in the literature. According to Zee & Lee (2003), [ə] is a plain-mid central schwa that occurs only in syllables closed by a nasal [n ŋ] (Zee & Lee 2003). [ɚ] is actually a sequence of two sounds, a plain [ə] followed by a rhotic [ɚ], more appropriately described as [əɚ] (Zee & Lee (2001: 643)). [ə] is rhotacized to [ɚ] in open syllables, which consists of a plain [ə] followed by a rhotic [ɚ], or realized as post-alveolar approximant [ɹ] in coda resulting in a sequence of [əɚ], i.e. the initial portion of [ɚ] is not rhotic, see Lee & Zee (2001: 643) and Zee & Lee (2003: 111); for a different analysis, see Duanmu (2000: 41f).

Duanmu (2000: 42) describes the retroflex vowel and a retroflex consonant as variants of the same phoneme that are context-sensitively distributed. Lin (2007: 73ff) assumes one mid vowel /ə/ and four allophones [ə e o ɤ] with specific context-sensitive distribution (in a diphthong or in syllable final position preceded by [j] or [ɥ] or [w] (for a description of the variants of /ə/, see also Duanmu (2000: 39f)).

For *low vowels*, Lin (2007: 77f) assumes one low vowel phoneme /a/ with three allophones [a, ɑ, ɛ]. When the low vowel is between [j]/[ɥ] and [n], the vowel becomes fronted and raised to [ɛ/æ] or [ɛ/a].

The diphthong [uo] after labial consonants [p p^h m f] has been treated as a monophthong [o] in some studies but is not included in modern descriptions of the vowel inventory of SC, see. Zee & Lee (2001: 643); see also Chiao & Kelz (1985: 53) and Hachenberg (2003: 112).

The so called *apical vowels* [ɹ ɻ] only occur after apical consonants. Traditionally, these apical vowels have been described as a plain apical vowel and a retroflex counterpart. In most modern accounts, apical vowels are considered to be syllabic consonants rather than vowels, which are a voiced extension of the preceding consonant into the syllabic nucleus position, in

Pinyin represented as *zi ci si* and *zhi chi shi ri*, i.e. they occupy the syllable nucleus position and function like vowels (Li 2007: 72; Duanmu 2000: 36). Zee & Lee (2001: 643) describe a syllabic apico-alveolar approximant [ɹ̩] and a syllabic apico-post-alveolar approximant [ɻ̩] which is apico-post-alveolar rather than retroflexed.

The so called *retroflex* or *rhotacized vowels* (Lee & Zee 2003) or rhotic vowels (Ladefoged & Maddieson 1996: 313), occur in the rime *-er* in words without a suffix frequently transcribed as [ɚ] and in the speech of Beijing speakers (for discussion see Lin 2007: 80ff, 62f; Duanmu 2000: 41f). Some accounts of the so called rhotacized vowels consider them to be a combination of the vowel [ə] and [a]. They are mostly dealt as vowels with added articulatory features similar to the post-alveolar [ɹ], and have a retracted tongue body (Lin 2007: 81).

The *duration* of vowels depends on the syllable structure: Standard Chinese has full syllables and weak syllables. Weak syllables are shorter than full syllables. Full syllables are similar in duration. In full open syllables a vowel is long (e.g. [ma:]), in full closed syllables it is short (e.g. [man] or [mai]). Vowel length is not contrastive neither in full syllables nor in weak syllables, since it is predictable from syllable structure (Duanmu 2000: 42).

Standard Chinese has a very rich system of diphthongs (falling vs. rising) and triphthongs (depending on the analysis). SC *diphthongs* are described in very different ways throughout the literature. Duanmu (2000) presents a comparison of four different transcriptions of the 34 different rhyming groups occurring in Standard Chinese. Apart from the monophthongs (V) there are combinations of glide + vowel (GV) and of glide + vowel + further element (GVX), which can be either a /n/ or nasal element or /ŋ/.

Lin (2007: 78ff) considers only four falling diphthongs to be real diphthongs: [ai ɑu ei ou]. They are interpreted as a complex vowel belonging to the syllable nucleus. High vowels before a mid/low vowel are considered to be glides in onset position ([ja je wo wa]). A mid or low vowel and its following high vowel are falling diphthongs which appear in the syllable nucleus. Thus, in Lin's account, rising diphthongs are not viewed as diphthongs but rather as a sequence of a glide + vowel (e.g. [ja]) associated to the syllable onset and a nuclear vowel. In this view, triphthongs do not have to be considered since they result from the association of a glide and a diphthong (e.g. [jau, wei]).

Diglossia is a common phenomenon in China and Taiwan. People speak a local dialect and Standard Chinese. The pronunciation of Standard Chinese differs regionally. Thus, people are generally confronted with a wide range of allophonic and regional variation in vowel realizations. Like with Arabic, a detailed contrastive analysis of the L2 perception of native speakers of Standard Chinese would therefore have to include a description of the local dialects spoken by the person together with an analysis of articulatory and acoustic data.

A typological analysis of a large sample of Chinese dialects has been carried out by Zee & Lee (2007) who investigated a regionally and genetically balanced representative sample of 86 Chinese dialects. The size of vowel systems in the sample ranges from three to eleven. The 7-vowel system is most frequent. 85 of the 86 sample dialects, have /i a u/, /i ɑ u/, or /i ɒ u/. The most frequent vowels are /i u a y/ followed by /o e ɔ ɛ/ in the mid range and by less common /ɿ ɤ ɚ œ ɑ ø æ ə ɯ/ and rare /ɒ ʮ ɵ ɨ ʏ/ in descending order of frequency. 33 of the 86 sample dialects have a contrast of oral and nasal vowels, the number of nasal vowels is ranging from 1 to 7.

To summarize, recent studies have described the phonological system of Standard Chinese or Mandarin to consist of five (Duanmu 2000; Lin 2007) or seven (Lee & Zee 2003) vowel phonemes. The vowel system of Standard Chinese can be described as consisting of the three basic qualities /i u a/, which occur in a majority of Chinese dialects (Zee & Lee 2007), a high front rounded vowel /y/ and one or two "mid" vowels [ɤ] and [ə] (and [ɚ]). Context-sensitive allophonic variation among high vowels /i y u/ is rather restricted but does occur for low [a] and especially for the "mid" vowels resulting in a wide range of different phonetic qualities and differing phonological accounts.

10.8.3 Contrastive analysis

Several studies on the pronunciation of Chinese learners in L2 German are available (Fluck, Zaiza & Qichang 1984; Yen 1992; Wang 1993; Hachenberg 2003; Hunold 2009) providing data for a contrastive analysis of German and Mandarin mostly from a didactically oriented viewpoint.

The Mandarin vowel system is considerably smaller than the German system. While Wang (1993: 146) expects no major difficulties with German vowels for Chinese learners, other contrastive studies of German and Standard Chinese describe vowels as one of the most difficult areas in the acquisition of German sound patterns (Yen 1992: 184; Hachenberg 2003; Hunold 2009: 166ff).

Studies on the German pronunciation of Chinese learners of German emphasize that vowel *quantity* is one of the most difficult problems (Yen 1992; Hachenberg 2003; Hunold 2009) since phonemic length is not distinctive in Mandarin. Hachenberg (2003) describes the differentiation of long and short German vowels as one of the major difficulties for Mandarin learners of German in perception as well as in production. In SC, vowels are long in full CV syllables. Kaden (1976: 194, cit. in Yen 1992: 162) mentions that Chinese monophthongs in closed syllables may appear to be short to German listeners, with the only exception of [a] in a /CVn/-context.

In an analysis of German pronunciation of 10 Chinese learners of German, Hunold (2009: 178) observed that short vowels seem to cause more problems than their long counterparts, i.e. vowel lengthening in pronunciation is frequently observed in syllables that are pronounced as stressed by the L2 learners (even in cases where word-stress is inappropriately applied). The insufficient differentiation of German short-long contrasts by L2 learners in production is described as having a strong negative influence on the rhythmic structure in L2 German and on the adequate realization of the contrast of stressed vs. unstressed syllables within word boundaries, with a strong impact on the general comprehensibility of Chinese learners' L2 speech (Hunold 2009).

i-, u- and y-qualities exist in both languages and are due to restricted context-sensitive variation with these vowel categories in SC. They are basically described as relatively stable categories in L2 perception of Chinese listeners, whereas in production severe problems with the correct realization of German i-, y- and u-vowels are observed (Hachenberg 2003; Hunold 2009).

Hachenberg (2003: 106) describes the realizations of German i-, y- and u-vowels by Chinese learners as undifferentiated with respect to German quality and quantity distinctions, i.e. no clear differentiation between 'tense' constricted vs. 'lax' unconstricted is realized. The phonetic realizations are described as more similar to the German "tense" vowels in quality and quantity by Hachenberg (2003: 106). In SC, /i/, /u/ and /y/ are pronounced more open before nasals (/ŋ n/) than in other contexts, i.e. as [ɪ], [ʊ] or [ʏ] respectively (Yen 1992: 192f).

The pronunciation of German /ɪ/ and /ʊ/ poses problems with respect to quality and quantity. Hunold (2009) describes 'lax', i.e. short unconstricted vowels to be frequently realized as long if they occur in stressed position.

Qualitative deviations are generally observed with long constricted vowels: /u:/ may be realized as [ʊ:]. /o:/ is rather pronounced as [ɔ:], while for /i:/ and the diphthongs, no major problems are observed. For the u-vowel in SC, according to Chiao & Kelz (1985: 92), a certain similarity between SC /u/ and German /o:/ is observed, i.e. for SC /u/ the tongue is reported to be more back similar to German /o:/, but lip aperture is smaller than for German /ʊ/ and lips are rounded and protruded in SC. Hachenberg (2003: 107) mentions moreover that in Hanyu Pinjin transcription, SC /u/ is transcribed as <o> in the context of post-vocalic

/ŋ/. Confusion with respect to the differentiation of German u- and o-qualities can therefore be expected that may be enhanced by the restricted status of /o/ in SC (see below).

Short German /ʏ/ is frequently realized too long as [y:], occasionally it may also be realized as [y] or [œ] (Hunold 2009: 167). Phonemic /y:/ did not occur in Hunold's material.

Hachenberg (2003: 109) reports severe problems of Chinese natives in differentiating German /e:/ - /i:/ and /o:/ - /u:/ in perception as well as in production. /e:/ is frequently substituted by a non-long i-quality between [i] and [ɪ]. Moreover, Hachenberg (2003) describes diphthongized [eɪ] and a range of e-qualities that are "somewhere between" [ɪ], [eɪ], [ɛɪ] and [ɛ] as substitutes in production for long constricted /e:/. Yen (1992: 228) considers only [ɛ] or [eɪ̯] as substitutes for /e:/. The diphthongization of /e:/ to [eɪ] in production is one of the major characteristics of Chinese learners of German in production (Hachenberg 2003: 109). The analysis of the perceptual data will show whether diphthongization is also observed in perception.

Hachenberg (2003: 110) observes a strong mono-directional tendency of e-vowels to be replaced by i-qualities in production. A similar mono-directional tendency is reported for ö-vowels being replaced by ü-qualities. However, this tendency is less strong with o-and u-vowels.

Similarly to front /e:/, the differentiation of /o:/ vs. /u:/ is described as very difficult for Chinese learners. In production, realizations between [u] and [ɔ] as well as diphthongizations are observed with a range of different realizations like [ʊ], [o], [oʊ], [ɔʊ], [ɔ] and [ɔʌ] (Hachenberg 2003). In Hunold's (2009) data, 9 of 10 speakers have problems with German /o:/ that is mostly realized as [ɔ:] (2009: 167).

Hachenberg (2003: 110f) explains difficulties with German /e:/ and /o:/ by the fact that these phonemes have no direct counterpart in SC and occur only as instable parts of diphthongs.

[e] occurs in SC only in one type of syllable before [ɪ], i.e. in the diphthong [eɪ] or in the triphthong [ʊeɪ], where it often merges with [ɛ] or [ɛɪ] in certain tonal constellations.

Similarly, [o] occurs only before [ʊ] in the diphthong [oʊ] and the triphthong [ɪoʊ], where it is not stable in certain tonal constellations and may merge with [ɔ] or a diphthong [ɔʊ]. Moreover, SC triphthongs containing e- and o-elements can be reduced to diphthongs (Hachenberg 2003: 111). Problems with /e:/ and /o:/ and the tendency towards diphthongization is one of the most persistent problems even with very advanced learners of German.

German /ɛ/ has a direct and stable SC equivalent [ɛ] which occurs in several syllable types, whereas the quantitative contrast of /ɛ/ - /ɛ:/ is more difficult for Chinese learners of German who often cope with a simple reduction in quantity (Hachenberg 2003: 112). In stressed position, German /ɛ/ is frequently lengthened and substituted by [ɛ:] or [e:] (Hunold 2009: 178). Occasionally, /ɛ/ is realized as [a] or [e] (Hunold 2009: 167).

In SC, /ɔ/ occurs as second part of a diphthong [ʊɔ] that can be associated with almost all initials. This diphthong seems to exhibit a strong tendency towards monophthongization after /b p m f/ (Hachenberg 2003: 112; Yin 1990: 77).

For /ɑ/, least problems are reported with respect to quality (Hachenberg 2003: 115; Hunold 2009: 167f). In Hunold's (2009) data, /ɑ/ is the only short vowel that is quantitatively correctly produced in most cases, while with long /ɑ:/ more quantitative deviations are observed. /ɑ:/ can be shortened in stressed position, whereas lengthening of /ɑ/ is usually not observed.

The differentiation of /ø:/ and /œ/ with respect to quantity and quality seems to be difficult, since both qualities do not occur in SC. /ø:/ can be realized as [œ] or [ɔ], e.g. in 'Söhnen' pronounced as [zɔnə], a pronunciation that is presumably influenced by the orthographic representation and morphological alternations (Hunold 2009: 124). Hachenberg (2003) describes considerable problems with the qualitative distinction of /y:/ - /ø:/, /ʏ/ - /œ/ and /ø:/ - /œ/. Yen (1992: 229f) describes a substitution of German /ø/ with SC unrounded [ɤ], [ɤə], and [əɪ] (arguing that SC [ɤ] is a good starting point for pronunciation training), while Hachenberg does not confirm these observations but describes phonetic realizations between [y/ʏ] and [ø/œ] as well as diphthongs like [øy] or [yø] that neither exist in German nor in Chinese. For /ø/ and /œ/, a tendency towards y-qualities is observed.

To summarize, problems with respect to quality and quantity are expected in L2 perception of German vowels. Based on the contrastive analysis of the L2 production of German vowels by Chinese learners, several instances of under-differentiation or confusion with long constricted vs. short unconstricted vowels with respect to quality and quantity are expected in L2 perception. Most problems are expected with e-, o- and ö-qualities due to their inexistence or restricted status in SC. These vowels are substituted in production by a range of qualities between /i/ and /ɛ/, /u/ and /ɔ/, and /y/ and /œ/, respectively. Diphthongization is another way of realizing German e- and o-qualities. Since Mandarin has a front rounded /y/, the perception of this quality is expected to be generally good, whereas for the other members of the class of front rounded vowels, especially for /ø:/ and /œ/, problems are expected due to their non-occurrence in the SC system. The analysis of the perceptual data of the present study will show, whether the observed phenomena also occur in perception.

10.8.4 Results and interpretation

Six Mandarin subjects were tested, one beginner and five advanced learners. Four persons were born in China, one in Taiwan and one in Singapore. All of them reported to have Mandarin as their native language, even if this information has to be viewed with great caution (see discussion above). All learners were tested in Vienna. Their id_wrong ranges from 21% to 63% wrong identifications. 1620 responses were collected from the six participants. The analysis presented here is based on 108 responses for each vowel category. The data set discussed here consists of 1580 valid responses (40 missing answers). For each vowel category, 103 to 107 valid responses were delivered.

10.8.4.1 Difficulties

In general, the data show a quite low percentage of correct identifications revealing the high portion of perceptual uncertainties of Mandarin learners of German: only 51.3% of all responses were correct.

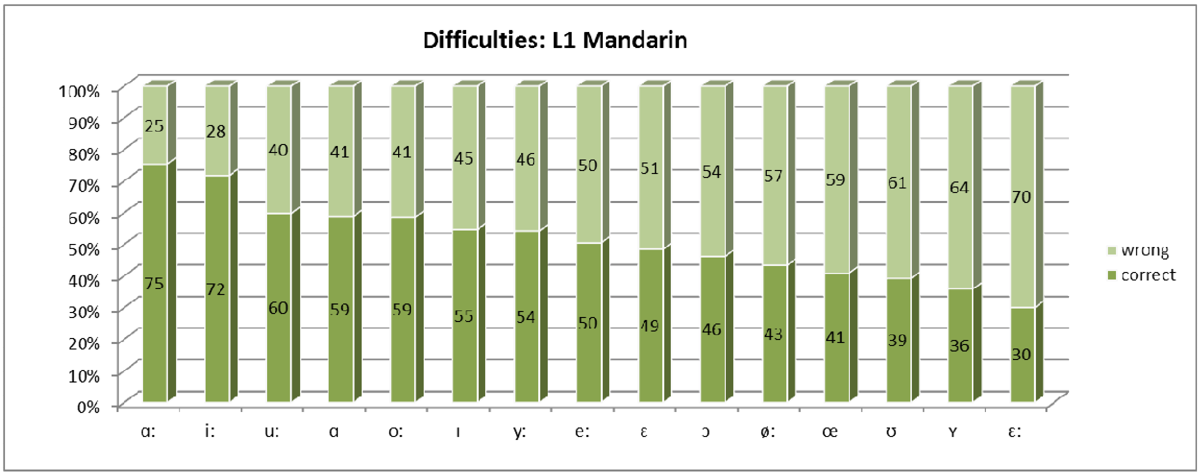

Figure 10.49: Percentage correct and wrong identifications for Mandarin listeners (N=6, 1620 responses)

Least problems are observed with /a:/, /i:/ and /u:/ (see Figure 10.49). As the boxplot diagram in Figure 10.50 shows, there are learners that identify these categories in all cases correctly. /ɛ:/, /ʏ/, /ʊ/ and /œ/ show the highest percentage of incorrect answers (Figure 10.49). The boxplot diagram shows that /ɛ:/, /ʏ/, /ø:/ but also /ɔ/ cause considerable problems to all participants.

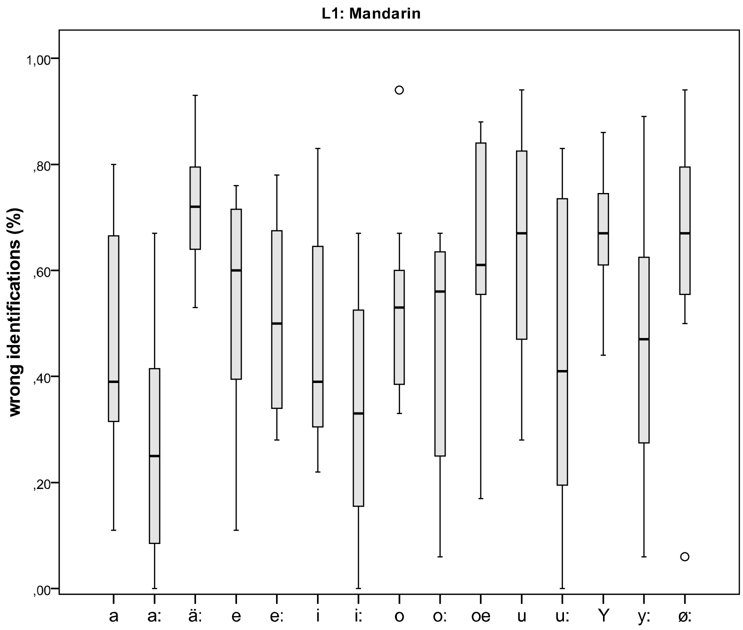

Figure 10.50: Id_wrong scores for Mandarin listeners

The listeners' mean id_wrong score was 52% (16% standard deviation), the median was 57%. The listener with the highest percentage of correct answers had an id_wrong of 21%, the highest id_wrong score was 68% wrong responses.

10.8.4.2 Preferences

Figure 10.51 compares the preference scores for L1 Mandarin listeners. The preference scores for monophthong categories vary within a range of 5.1% (min /ɛ:/ 3.6%, max /o:/ 8.7%). The least preferred response category is <ä:> followed by /ɔ/.

Surprisingly, even if /o:/ is expected to be one of the most difficult categories for Mandarin listeners (see section 10.8.3), it is also the most preferred response category. This effect is mainly due to a strong tendency to categorize /ɔ/ as /o:/ indicating the listeners awareness of difficulties with this phonemic category.

As Figure 10.51 shows, Mandarin listeners clearly prefer [+long] response categories. This may be due to their general awareness of this feature and to a kind of perceptual compensation of their problems to differentiate long vs. short contrasts in German. SC shows no phonemic length contrasts. However, there are differences with respect to vowel length in SC that are due to syllable structure: vowels in open syllables are longer than in closed syllables (Duanmu 2000: 42).

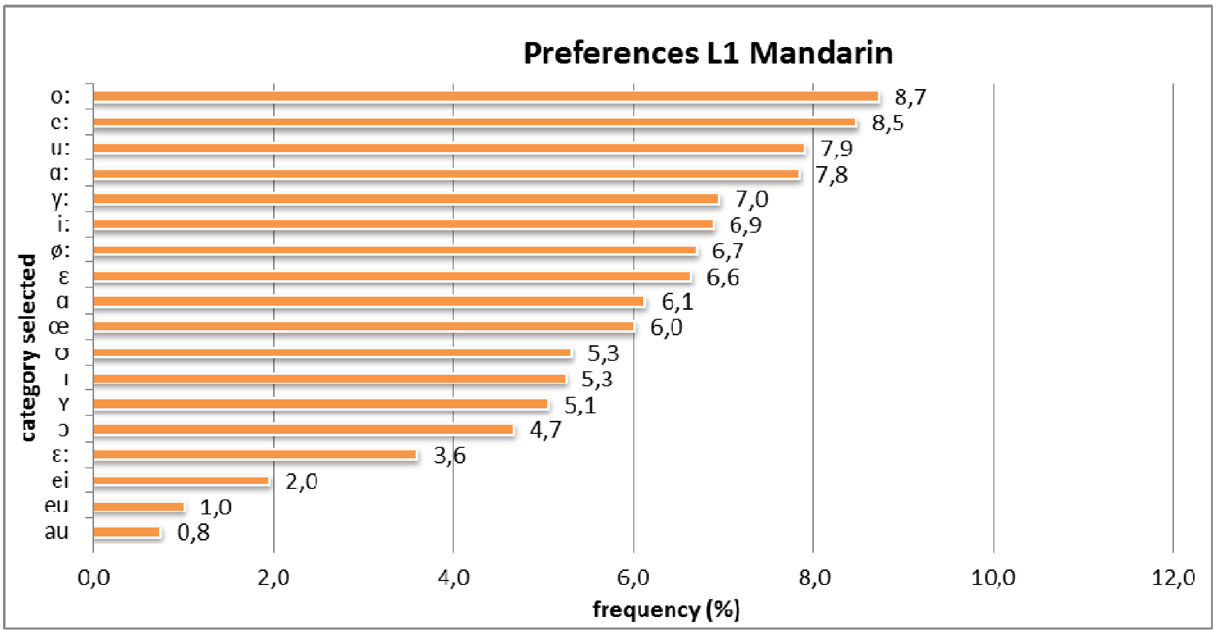

Figure 10.51: Preferences per category for Mandarin listeners

Standard Chinese has a fairly rich number of diphthongs. In the identification task, 5 of 6 participants considered the categories <ei>, <eu> and <au> to be appropriate response options for short as well as for long vowels in some cases. The most attractive diphthong was <ei> that functions as substitute for /ɛ/ (7%), /ɛ:/ (8%) and /e:/ (14%). <eu> and <au> are only used as substitutes for rounded qualities.

10.8.4.3 Patterns of confusion

The confusion matrix in Table 10.45 shows perceptual substitutions of Mandarin learners of German. The matrix shows three major confusion clusters: (1) *pharyngeal a-vowels* underdifferentiated mainly with respect to phonemic length, (2) *palatal non-rounded* vowels underdifferentiated in quality and quantity, or more precisely in constriction location, degree of aperture, and phonemic length and constriction degree, (3) *rounded vowels* which are underdifferentiated with respect to constriction location (front vs. back), degree of aperture, and phonemic length and constriction degree. Especially with the class of rounded vowels, most difficulties are observed and a wide range of intercategory confusion is given. As mentioned above, diphthongs seem to be an attractive response option for several German vowel categories, especially for e- and o-vowels and less frequently for ö-vowels and short unconstricted /ʊ/.

a-vowels are the most stable category. Most instances of wrong categorizations concern instances where phonemic length is identified incorrectly or hypercorrectly. /ɑ/ has an id_correct score of 59% and is identified as /ɑ:/ in 40% of the cases and once as /ɛ:/.

%	ɑ	ɑ:	ɛ:	ɛ	e:	ɪ	i:	ɔ	o:	ʊ	u:	œ	ø:	ʏ	y:	‹ei›	‹eu›	‹au›	wrong	**corr**
a	**59**	40	1																41	**59**
a:	23	**75**	1	1															25	**75**
ɛ:	4		**30**	15	42		2									8			70	**30**
ɛ	4	1	15	**49**	19	2	2		1		1					7			51	**49**
e:	2		7	10	**50**	3	13									14			50	**50**
ɪ				21	8	**55**	16									1			45	**55**
i:				3	7	18	**72**												28	**72**
ɔ								**46**	40		1	2	1				5	6	54	**46**
o:								4	**59**	8	18	1	5	1			3	2	41	**59**
ʊ								15	10	**39**	23	6		2	1		1	3	61	**39**
u:								2	3	21	**60**	7	5		3				40	**60**
œ				2	1			3	10	4	2	**41**	26	7	3		2	1	59	**41**
ø:									6	1	7	9	**43**	8	21		5		57	**43**
ʏ						1			4	5	4	21	6	**36**	23				64	**36**
y:					1					1	2	4	15	23	**54**				46	**54**
	6	8	4	7	8	5	7	5	9	5	8	6	7	5	7	2	1	.8	48.7	***51.3***

Table 10.45: Confusion matrix for L1 Mandarin listeners. Correct (in bold) and wrong identifications in % for L1 Mandarin listeners (N=7, 1580 responses), (presented stimuli in rows, selected categories in columns in semi-orthographic representation, values < 0.5 not indicated)

/ɑ:/ has the highest id_correct score (75%) and is substituted with /ɑ/ in 23% or with /ɛ:/ (one case) and /ɛ/ (one case). In SC, /a/ has three allophones [a ɑ ɛ] (Lin 2007: 77f; Duanmu (2000: 40) mentions 5 different allophones). The substitutions of German a-qualities by ɛ-qualities appear twice in the context after /j/ and could be caused by the transfer of the SC fronting process by which /a/ is fronted to [ɛ]/[æ] in positions between [j]/[ɥ] and [n] (Lin 2007: 77f). Moreover, the orthographic similarity of German <ä> and <a> and the interference of English orthography and pronunciation of English <a> may play a role in these instances of wrong categorization.

As is predicted by the contrastive analysis (CA) presented above, German *e-vowels* seem to cause considerable phonemic confusion with respect to quality and quantity, a fact that is presumably connected with allophonic variation of /e/ in Mandarin.

/ɛ:/ is identified correctly in only 30% of the cases. It is most frequently substituted by /e:/ (42 %) or /ɛ/ (15%), but also by /i:/ (2%) and <ei> (8%).

/ɛ/ is identified correctly in 49% of its occurrences and it is substituted with /e:/ (19%), /ɛ:/ (15%) and <ei> (7%), but also with /ɑ/ (4%) and /ɑ:/ (1%), and considerably less frequently with /i:/ (2%), /ɪ/ (2%), /o:/ (one case) and /u:/ (one case).

/e:/ has an id_correct of 50% and is incorrectly substituted by <ei> (14%), /i:/ (13%), /ɛ/ (10%), /ɛ:/ (7%), /ɪ/ (3%), and /ɑ/ (2%). The substitution of /e:/ to /i:/ (13%) could be interpreted as a kind of fronting. While /e:/ is attracted towards /i:/ in 13% of its occurrences, this is a very infrequent strategy for /ɛ/ and /ɛ:/. A strong preference for diphthongization of /e:/ (14% <ei>) is another interesting strategy observed.

i- vowels seem less difficult: Long constricted /i:/ is identified correctly in 72% of its occurrences. Its primary substitute is /ɪ/ (18%), other substitutes are /e:/ (7%), /ɛ/ (3%) and <ei> in only one case.

Unconstricted /ɪ/ is identified correctly in 55% of its occurrences. It is substituted by /ɛ/ (21%), /i:/ (16%), /e:/ (8%) and <ei> (1%).

Diphthongization occurs with all mid-palatal qualities but not with pre-palatal qualities. The strongest tendency towards diphthongization to <ei> is observed with /e:/ (14%), followed by /ɛ:/ (8%) and /ɛ/ (7%).

With /ɪ/ vs. /ɛ/, perceptual asymmetries are observed: While /ɪ/ > /ɛ/ occurs in 21%, /ɛ/ > /ɪ/ is considerably less frequently observed (only 2%). An inverse and less strong asymmetry is observed with the long constricted palatal vowels: /i:/ > /e:/ (7%) and /e:/ > /i:/ (13%). The confusion matrix shows moreover that the contrast of /e:/ vs. /ɪ/ apparently does not cause many problems. Rather, it is the contrast of /e:/ vs. /i:/ that seems to be difficult. We can therefore assume that the listeners have a certain sensitivity to quantity distinctions, whereas the exact quality of German palatal vowels appears to cause considerably more problems. While with long constricted vowels a tendency towards the pre-palatal quality is observed, the inverse trend is observed for short unconstricted vowels.

Considerable confusion is observed with *o- and u-vowels*: For /ɔ/, the qualitative identification of unconstricted /ɔ/ or rather the graphemic identification with <o> seems to be relatively clear, whereas difficulty appears to be more related to phonemic length and constriction degree. /ɔ/ has an id_correct score of 46% and is identified as /o:/ in 40% of its occurrences. Moreover, /ɔ/ is substituted by <eu> (5%), <au> (5%), /œ/ (2%) and in one case each by /ø:/ (1%) or /u:/ (1%).

Long constricted /o:/ is identified correctly in 59% of the cases or substituted with /u:/ (18%), /ʊ/ (8%), /ø:/ (5%) or /ɔ/ (4%), but also less frequently with <eu> (3%) or <au> (2%), and in one instance each with /œ/ (1%) or /ʏ/ (1%).

Unconstricted /ʊ/ (39% correct) is substituted by /u:/ (23%), /ɔ/ (15%), /o:/ (10%), /œ/ (6%) and less frequently with <au> (3%), /ʏ/ (2%), /y:/ (1%) and <eu> (1%).

Long constricted /u:/ has an id_correct score of 60% and is substituted by /ʊ/ (21%), /œ/ (7%), /ø:/ (5%), /y:/ (3%), /o:/ (3%), and /ɔ/ (2%).

A similar asymmetric pattern as with the mid- and prepalatal vowels is observed with back rounded vowels. While with long constricted vowels a stronger tendency towards velar qualities is given (/u:/ > /o:/ (3%) vs. /o:/ > /u:/ (18%)), short unconstricted vowels are more frequently categorized as uvular (/ʊ/ > /ɔ/ (15%) vs. /ɔ/ > /ʊ/ (0%)).

Front rounded vowels are expected to be difficult for Mandarin listeners, since only /y/ is phonemic in their L1.

For /œ/ (41% correct) considerable variation in the responses is observed. /œ/ is substituted by /ø:/ (26%), /o:/ (10%), /ʏ/ (7%), /ʊ/ (4%), /y:/ (3%), /ɔ/ (3%), /u:/ (2%), /ɛ/ (2%), <eu> (2%), /e:/ (1%) and <au> (1%).

Similarly, incorrect responses for /ø:/ (43% correct) spread over several categories. /ø:/ is substituted by /y:/ (21%), /œ/ (9%), /ʏ/ (8%), /u:/ (7%), /o:/ (6%), <eu> (5%) and /ʊ/ (1%).

/ʏ/ has a very low id_correct score of 36%. It is substituted by /y:/ (23%), /œ/ (21%), /ø:/ (6%), /ʊ/ (5%), /u:/ (4%), /o:/ (4%), and /ɪ/ (1%).

A higher id_correct score is given for /y:/ (54% correct). /y:/ is substituted with /ʏ/ (23%), /ø:/ (15%), /œ/ (4%), /u:/ (2%), /ʊ/ (one case) and /e:/ (one case).

With back vowels this effect is weaker for unconstricted /ʊ/ but however given: /o:/ > /ʊ/ (8%) vs. /o:/ > /u:/ (18%), and /ʊ/ > /o:/ (10%) vs. /ʊ/ > /ɔ/ (15%).

/u:/ is attracted towards /o:/ in only 3% of the cases, but is substituted by /ʊ/ in 21% of its occurrences. As with e- and i-qualities, these observations can be interpreted as evidence for a certain perceptual sensitivity to length distinctions, whereas categorization with respect to quality is not performed correctly.

While with long constricted vowels, pre-palatal /y:/ is the more attractive perceptual target (/ø:/ > /y:/ (21%) vs. /y:/ > /ø:/ (only 15%)), mid-palatal short unconstricted /œ/ is the more preferred response option for the unconstricted qualities (/ʏ/ > /œ/ (21%) vs. /œ/ > /ʏ/ (only 7%)).

Diphthongs are very frequently selected as response option. The most attractive diphthong is <ei> that functions as substitute for front /ɛ/ (7%), /ɛ:/ (8%) and /e:/ (14%), but never for /ɛ:/. Rounded vowel stimuli are only substituted by <eu> and <au>. Long constricted /i: u: y:/ and a- vowels are never identified as diphthongs. /e:/ is the category that is most frequently identified as a diphthong (14% /e:/), followed by /ɔ/ (5% <eu>, 6% <au>).

The preference for diphthongs in Mandarin listeners' responses is presumably due to the high number of diphthongs and even triphthongs in Chinese (cf. Lin 2007: 160f) but also due to general perceptual difficulties with German mid vowels. Perceptual categorization with diphthongs can be interpreted as one of the learners' strategies to cope with these difficulties.

10.8.4.4 Similarity and distance

A comparison of the sim_scores in Table 10.46 shows the highest sim_score for the contrast /e:/ - /ɛ:/ (0.611), followed by minimal pairs that differ only in terms of phonemic length and constriction degree: The highest sim_score for this class is given for /ʏ/ - /y:/ (0.512), followed by /ɑ/ - /ɑ:/ (0.468), /ʊ/ - /u:/ (0.453), /œ/ - /ø:/ (0.417), /ɔ/ - /o:/ (0.414), and /ɛ/ - /ɛ:/ (0.379). Lower sim_scores are given for /ɛ/ - /e:/ (0.296) and /ɪ/ - /i:/ (0.272). High sim_scores are moreover observed for front rounded /ø:/ - /y:/ (0.368) /œ/ - /ʏ/ (0.365).

As described for the confusion matrix data, Mandarin listeners show several instances of wrong categorizations that are due to wrong identifications of the constriction location. The sim_scores of minimal pairs that differ in constriction location generally are below 0.2 (except for /ɪ/ - /ɛ/ (0.219)). Such wrong categorizations do not only refer to the mid- vs. prepalatal contrasts but also to the contrast of velar vs. uvular vowels, both involving so-called 'mid' vowels, and to the contrast of front vs. back rounded vowels.

V	ɑ	ɑ:	ɛ:	ɛ	e:	ɪ	i:	ɔ	o:	ʊ	u:	œ	ø:	ʏ	y:
a	1.000														
ɑ:	.468	1.000													
ɛ:	.054	.009	1.000												
ɛ	.035	.015	.379	1.000											
e:	.017	.000	.611	.296	1.000										
ɪ	.000	.000	.000	.219	.098	1.000									
i:	.000	.000	.019	.040	.163	.272	1.000								
ɔ	.000	.000	.000	.000	.000	.000	.000	1.000							
o:	.000	.000	.000	.009	.000	.000	.000	.414	1.000						
ʊ	.000	.000	.000	.000	.000	.000	.000	.175	.184	1.000					
u:	.000	.000	.000	.009	.000	.000	.000	.027	.178	.453	1.000				
œ	.000	.000	.000	.021	.010	.000	.000	.054	.105	.117	.084	1.000			
ø:	.000	.000	.000	.000	.000	.000	.000	.011	.103	.011	.109	.417	1.000		
ʏ	.000	.000	.000	.000	.000	.011	.000	.000	.051	.089	.041	.365	.180	1.000	
y:	.000	.000	.000	.000	.009	.000	.000	.000	.000	.020	.041	.070	.368	.512	1.000

Table 10.46: Similarity matrix for L1 Mandarin listeners

Within the class of front vowels, the highest sim_scores are observed for /e:/ - /ɛ:/ (0.611), /ʏ/ - /y:/ (0.512), /œ/ - /ø:/ (0.417), /ɛ/ - /ɛ:/ (0.379), /ø:/ - /y:/ (0.368), and /œ/ - /ʏ/ (0.365). Sim_scores between 0.3 and 0.1 are obtained for /ɛ/ - /e:/, /ɪ/ - /i:/, /ɪ/ - /ɛ/, /ʏ/ - /ø:/, and /i:/ - /e:/. All other contrasts of front vowels have scores below 0.1. Rounded vs. non-rounded contrasts receive values below 0.03.

front Vs	*sim_scores*
e: - ɛ:	.611
ʏ - y:	.512
œ - ø:	.417
ɛ - ɛ:	.379
ø: - y:	.368
œ - ʏ	.365
ɛ - e:	.296
ɪ - i:	.272
ɪ - ɛ	.219
ʏ - ø:	.180
i: - e:	.163
ɪ – e:	.098
œ - y:	.070
i: - ɛ	.040
œ - ɛ	.021
i: - ɛ:	.019
œ – e:	.010
y: - ɛ:	.009

Table 10.47: Comparison of similarity scores for [+front] vowels for Mandarin listeners

Within the class of back vowels, high sim_scores are observed for vowels differing in phonemic length (/ɑ/ - /ɑ:/ 0.468) or in phonemic length and constriction degree (/ʊ/ - /u:/ 0.453, /ɔ/ - /o:/ 0.414). Lower scores are observed for /ʊ/ - /o:/ (0.184), /u:/ - /o:/ (0.178), and /ʊ/ - /ɔ/ (0.175), followed by /ʊ/ - /ʏ/ (0.089), /ɛ:/ - /ɑ/ (0.054) and /u:/ - /ɔ/ (0.027).

back Vs	*sim_scores*
ɑ - ɑ:	.468
ʊ - u:	.453
ɔ - o:	.414
ʊ - o:	.184
u: - o:	.178
ʊ - ɔ	.175
ʊ - ʏ	.089
ɛ: - ɑ	.054
u: - ɔ	.027

Table 10.48: Comparison of similarity scores for [+back] vowels for Mandarin listeners

There are several contrasts pairs that differ in constriction location in terms of the feature front vs. back and are confused by SC learners of German. The highest scores are obtained for front vs. back contrasts within the class of rounded vowels /ʏ/ - /ɔ/ (0.365), /œ/ - /ʊ/ (0.117), /u:/ - /ø:/ (0.109), /o:/ - /œ/ (0.105), and /ø:/ - /o:/ (0.103).

Vs	*sim_scores*
ʏ - ɔ	.365
œ - ʊ	.117
u: - ø:	.109
o: - œ	.105
ø: - o:	.103
u: - œ	.084
œ - ɔ	.054
o: - ʏ	.051
u: - ʏ	.041
u: - y:	.041
ɛ - ɑ	.035
ʊ - y:	.020
e: - ɑ	.017
ɛ - ɑ:	.015
ɔ - ø:	.011
ʊ - ø:	.011
ʏ - ɪ	.011
ɛ: - ɑ:	.009
o: - ɛ	.009
u: - ɛ	.009

Table 10.49: Comparison of similarity scores for [+/–front] contrasts for Mandarin listeners

10.8.4.5 The perceptual vowel map for L1 Mandarin listeners

From the similarity scores calculated by Shepard's formula, distance scores and non-metric MDS solutions were derived. For the non-metric MDS solutions, RSQ values of 0.635 in dimension 1, 0.865 for the two-dimensional solution, and 0.911 for the three-dimensional solution are obtained.

The one-dimensional MDS solution (see Figure 10.52) has a relatively high RSQ of 0.635 and shows that rounding is the feature that divides German vowels into two major classes. This is also very evident in the two-dimensional MDS solution in Figure 10.53. Hardly any cases of perceptual confusion by Mandarin listeners in L2 German are due to wrong identifications based on the non-differentiation of this feature (only a few instances of perceptual unrounding of front rounded vowels are observed: twice for /œ/ > /ɛ/ and once for /œ/ - /e:/. A look at the confusion matrix (Table 10.45) shows that wrong categorizations due to the insufficient differentiation of long/short contrasts and of contrasts based on different constriction locations are significantly higher.

Front /i:/, /ɛ:/ and /ɛ/ are differentiated considerably well in terms of constriction location, whereas for /ɪ/ and especially for /e:/ more category dispersion in the listeners' responses is observed in the confusion matrix. The perceptual attraction effect of /i:/ for /e:/-stimuli is reflected in the close position of /i:/ and /e:/ in the representation. Due to some instances of

/a/-responses for e-stimuli, an affinity of German a-vowels with /ɛ:/ and /ɛ/ is observed in the MDS solution.

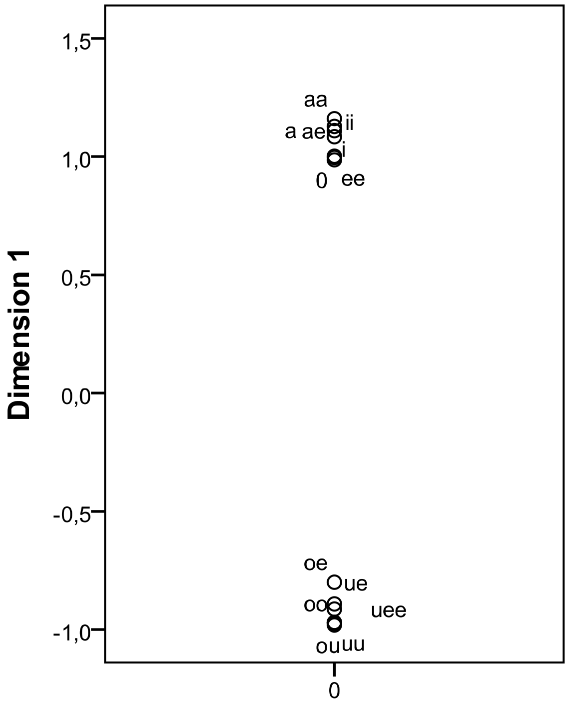

Figure 10.52: One-dimensional MDS representation of the perceptual map of German vowels for L1 Mandarin listeners (N=6, 1620 responses, RSQ .635)

The two-dimensional solution and the confusion matrix in Table 10.45 show that the perceptual differentiation within the class of rounded vowels is considerably more difficult with respect to constriction location (front vs. back, mid- vs. pre-palatal, velar vs. uvular) and phonemic length and constriction degree.

/y:/ and /ɔ/ are the categories that are most distinct from all other rounded vowels. In the learners' responses, long constricted /o:/ and /u:/, but also /y:/ and /ø:/ are more preferred than short unconstricted categories. For short unconstricted /ʊ œ ʏ/ but not for /ɔ/, a high percentage of inter-category confusion is observed, indicating the perceived ambiguity of the stimuli with respect to constriction location. Due to lip rounding, the relative degree of aperture is less distinctive with back and front rounded vowels than with front non-rounded vowels. These perceptual difficulties are enhanced by the short duration of the stimuli itself.

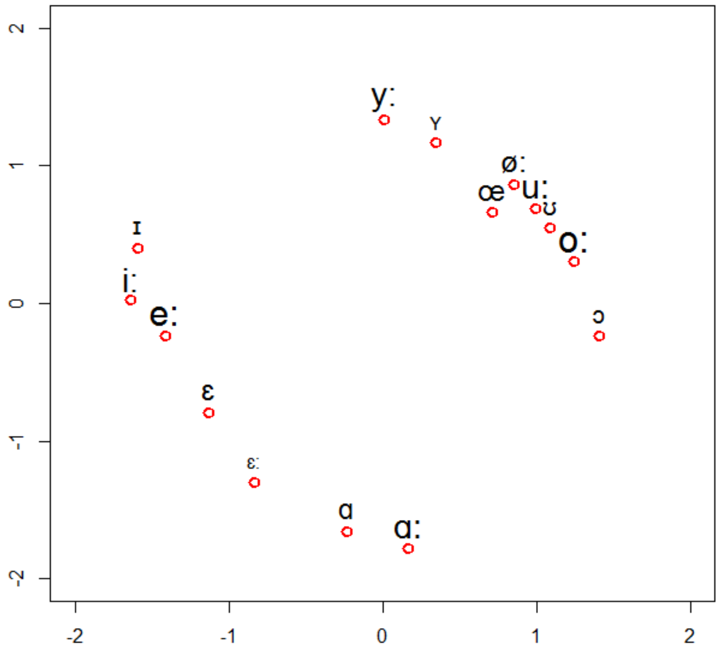

Figure 10.53: Two-dimensional MDS representation of the perceptual map of German vowels for L1 Mandarin listeners (N=6, 1620 responses, RSQ .865)

The three-dimensional MDS solution gives a similar impression as the two-dimensional representation, though the class of rounded vowels is better differentiated in the third dimension. The solution shows that back rounded vowels are quite well differentiated from front rounded vowels, though confusions of front and back rounded vowels are of course observed (see the confusion matrix in Table 10.45).

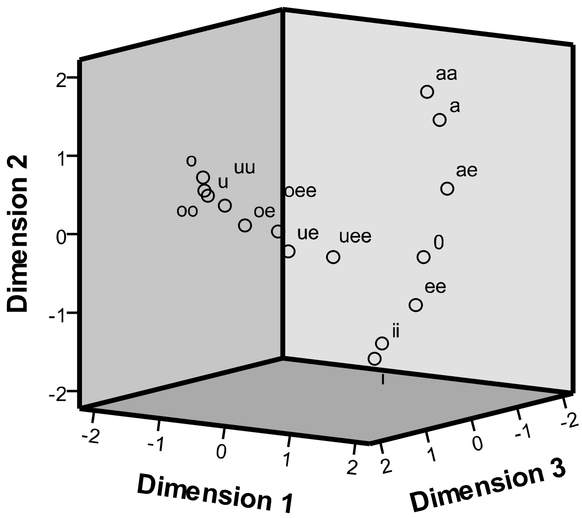

Figure 10.54: Three- dimensional MDS representation of the perceptual map of German vowels for L1 Mandarin listeners (N=6, 1620 responses, RSQ .911)

To conclude, the MDS solutions reflect the perceptual difficulties of Mandarin learners of German that are also evident in the confusion matrix in Table 10.45 indicating that lip rounding is in a certain sense masking the perceptual effects of constriction location, yielding considerable inter-category confusion among German rounded vowels.

10.8.5 Summary

To summarize, the low percentage of correct responses (51% overall id_correct) shows that Mandarin/Standard Chinese listeners have considerable difficulties with the correct perceptual differentiation and categorization of German vowels. Qualitative and quantitative distinctions cause considerable perceptual problems that are reflected by the wide spreading of responses in the confusion matrix. Difficulties with respect to quality occur with contrasts of front rounded vs. back rounded vowels, with mid- vs. pre-palatal and velar vs. uvular contrasts. These difficulties are specifically numerous with e-, o- and u-qualities and with the class of front rounded vowels. For /a:/, /i:/ and /u:/, least difficulties are observed.

For the class of front rounded vowels, significant problems are observed. The listeners' responses are widely spread over several front as well as back categories. Within this class, /y:/ is the category with the highest percentage of correct responses and least dispersion over other response categories, due to the existence of an equivalent vowel quality in SC.

Patterns of asymmetry are observed in a mono-directional attraction of /e:/ > /i:/ (13% vs. 7%) and /o:/ > /u:/ (18% vs. 3%). An inverse pattern is observed for short unconstricted /ɪ/ > /ɛ/ (21% vs. 2%), /ʊ/ > /ɔ/ (21% vs. 7%) and /ʏ/ > /œ/ (15% vs. 0%). While with /e:/ and with /o:/ a certain asymmetric attraction effect of long constricted /i:/ and /u:/ respectively is given, the unconstricted counterparts /ɪ/ and /ʊ/ exert less attraction. For /y:/ and /ø:/, this mono-directional attraction is less strong: /ø:/ > /y:/ (21%) vs. /y:/ > /ø:/ (15%). Moreover, for /e:/, /o:/ and /ø:/ several instances of perceptual diphthongization are observed.

10.9 Polish

10.9.1 General characteristics

Polish is a Slavonic language, belonging to the Lechitic sub-group of Western Slavonic languages. Polish is the largest of the West Slavonic languages in terms of the number of speakers and the second most widely spoken Slavonic language spoken by about 41 million native speakers in Poland, by Polish minorities in other countries like Lithuania, Belarus and Ukrainia and approximately 10 million emigrants who migrated to countries all over the world (Jassem 2003).

Differences between dialects or regional variants of Polish are moderate. The Polish writing system is based on the Latin alphabet using several additional diacritics and digraphs.

10.9.2 Phonological and phonetic description

A typical characteristic of Polish is its rich consonant system (see Table 10.50). Polish has been described as a "consonantal language" referring to two characteristic properties: (1) a very large system of consonant phonemes at the phonetic surface, and (2) the appearance of heavy consonant clusters in all word positions. In word-initial position, phonological words may contain clusters of up to five consonants, which are not simplified even in fluent spontaneous speech (Jassem 2003; Konopka 2003).

Descriptions of the Polish phoneme inventory differ with respect to some partly inter-related issues regarding the underlying segment inventory: (1) the status of [c ɟ] as phonemes or as fronted allophones of velar /k g/, (2) the treatment of palatalized labials preceding /i/ and /j/ as phonemes or as allophones (for discussion, e.g. Bethin 1992: 91; Konopka 2003: 661; Rubach 2006: 677), (3) the description of the nasal vowels <ą ę> as single phonemes or as sequences of oral plus nasal vocalic elements, and (4) the status of the vowel represented as <y> in orthography as a phoneme or as an allophone of /i/. Table 10.50 represents the Polish consonant system as it is described in most phonological descriptions assigning phonemic status to pre-palatal consonants but considering palatalized labials as context-sensitive allophones of labials.

A common classification of Polish consonants is based on the distinction of "hard" and "soft" consonants. Hard consonants are /p b m f v t d s z ʦ ʣ ʃ ʒ ʧ ʤ l r k g x/. The so-called soft consonants /ɕ ʑ ʨ ʥ c ɟ ɲ/ are palatalized and can be described as [-back], whereas hard consonants can be described as [+back] (Rubach 2006). In many studies, "underlying palatals” are differentiated from consonants that are palatalized on the phonetic surface before /i j/ within words as well as across word boundaries (see Rubach 1984: 25)

	labial	labio-dental	(post-) dental	(post-) alveolar	pre-palatal	palatal	velar
Plosive	p b		t d			c ɟ	k g
Fricative		f v	s z	ʃ ʒ	ɕ ʑ		x
Affricate			ʦ ʣ	ʧ ʤ	ʨ ʥ		
Nasal	m		n		ɲ		ŋ
Lateral			l				
Flap/Trill				r			
Approximant		ʋ				j	

Table 10.50: The consonant system of Polish

The contrast of sibilant fricatives and affricates at three different places of articulation, i.e. dental, (post)alveolar and pre-palatal, is a characteristic feature of Polish. Unique to Polish is the contrast between post-alveolar and alveolo-palatal/pre-palatal fricatives and affricates, e.g. in *kasza* /kaʃa/ ‘porridge’ – *Kasia* /kaɕa/ ‘Cathy’, *żar* /ʒar/ ‘heat’ – *ziarno* /ʑarnɔ/ ‘grain’ (Rubach 2006).

Labial /p b f v m/ preceding /i j/ can be described either as palatalized surface allophones /p^j b^j f^j v^j m^j/ (e.g. Rubach 1984) or as phonemes like in minimal pairs like *piasek* [p^jasek] ‘sand’ vs. *pasek* [pasek] ‘belt (dim.)’, where the contrast would only be based on the initial consonant (for discussion, see also Biedrzycki 1974: 42f; Bethin 1992: 91).

The affricates /ʦ ʣ ʧ ʤ ʨ ʥ/ are distinct from /ts dz tʃ tɕ dʑ/, e.g. *czy* /ʧɨ/ ‘whether’ – *trzy* /tʃɨ/ ‘three’, *wieczny* /ˈvjeʧnɨ/ ‘eternal’ - *wietrzny* /ˈvjetʃnɨ/ ‘windy’, *dźwig* /ʤvik/ ‘crane’ - *dzik* /ʤik/ ‘boar’ - *drzwi* /dʒvi/ ‘doors’ (Jassem 2003: 104; Rubach 2006: 676).

Polish has two glides /j/ and /w/ that can be derived from /i/ and /u/ by rules of gliding (Rubach 2006: 677). The front glide /j/ is in complementary distribution with front high /i/ (for discussion, see Rubach 1984; Bethin 1992: 86ff; Strutyński 1999).

The Polish vowel inventory consists of six vowel phonemes /i/ <i>, /ɛ/ <e>, /ɨ/ <y>, /a/ <a>, /ɔ/ <o>, and /u/ <u/ó>. Frequencies (in Hz) for the first four formants (Jassem 1968) are represented in Table 10.51 and graphically represented in Figure 10.56.

Generally, little contextual allophony is reported for Polish vowels. However, in the vicinity of palatal or palatalized consonants, all vowels are heavily fronted and raised, e.g. /ɛ/ > [e] in *pieśń* [pjeɕɲ] ‘song’; /a/ is fronted in this position, e.g. *dzisiaj* [ˈdʑiɕaj] ‘today’ (Sawicka

1995; Jassem 2003: 106). In rapid speech, this coarticulatory effect may lead to a neutralization of the contrast of /ɛ/ and /i/, e.g. *pieniądze* [pʲɛɲɔndzɛ] ~ [pʲiɲɔndzɛ] 'money'.

V	F1	F2	F3	F4
i	220	2300	3100	3500
ɨ	330	2000	2600	3500
ɛ	550	1850	2550	3500
a	800	1250	2500	3500
ɔ	530	850	2600	3250
u	270	620	2500	3250

Table 10.51: Formant frequencies for the first four formants (in Hz) for Polish vowels (cf. Jassem (1968) – average values based on five repetitions for each vowel realized by eight Polish speakers)

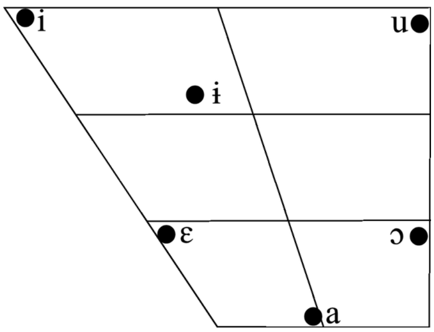

Figure 10.55: Polish vowel chart[109]

The status of the so-called *nasal vowels* <ę> and <ą> as vowel phonemes or as diphthongs is controversially discussed in the literature (cf. Bethin 1992: 56ff; Konopka 2003: 660; Gussmann 2007: 2; Stone 2011: 293). Traditionally <ę> and <ą> are described to be phonemically distinct, even though their phonemic status depends only on their occurrence in word-final position (e.g. in *głowa* 'head (nom.)' – *głową* 'head (instr.)' – *głowę* 'head (acc.)') or before fricatives (e.g. *mąż* /mɔ̃ʃ/ 'husband', *gęś* /gɛ̃ɕ/ 'goose'). In orthography, however, the nasal vowel letters appear before all types of consonants (Stone 2011: 293). Phonetically, before plosives and affricates, <ą> and <ę> are realized as a sequence of an oral vocalic part analogous to /ɔ/ and /ɛ/, respectively, and a nasal part, which is pronounced as [m] before a bilabial, [n] before a dental stop, and [ŋ] before a velar stop (e.g. *ząb* /zɔmb/ 'tooth', *mosiądz* /mɔɕɔnts/ 'brass', *ręka* /rɛŋka/ 'hand'). In some studies, <ą> and <ę> are therefore represented as diphthongs before continuants or in word-final position and as vowel plus homorganic

109 http://commons.wikimedia.org/wiki/File:Polish_vowel_chart.svg?uselang=de (accessed 2014-11-09)

nasal before non-continuant obstruents (Bethin 1992: 56f). In loanwords, all vowels before a nasal /m/ or /n/ followed by a fricative may be realized as nasal vowel, e.g. in *tramwaj* /trãvaj/ ‘tram’, *sens* /sẽs/ ‘sense’, *inspektor* /ĩspɛktɔr/ ‘inspector’, *triumf* /triũf/ ‘triumph’. In fluent speech, however, the so-called nasal vowels are rather realized as diphthongs or even triphthongs (Konopka 2003: 660).

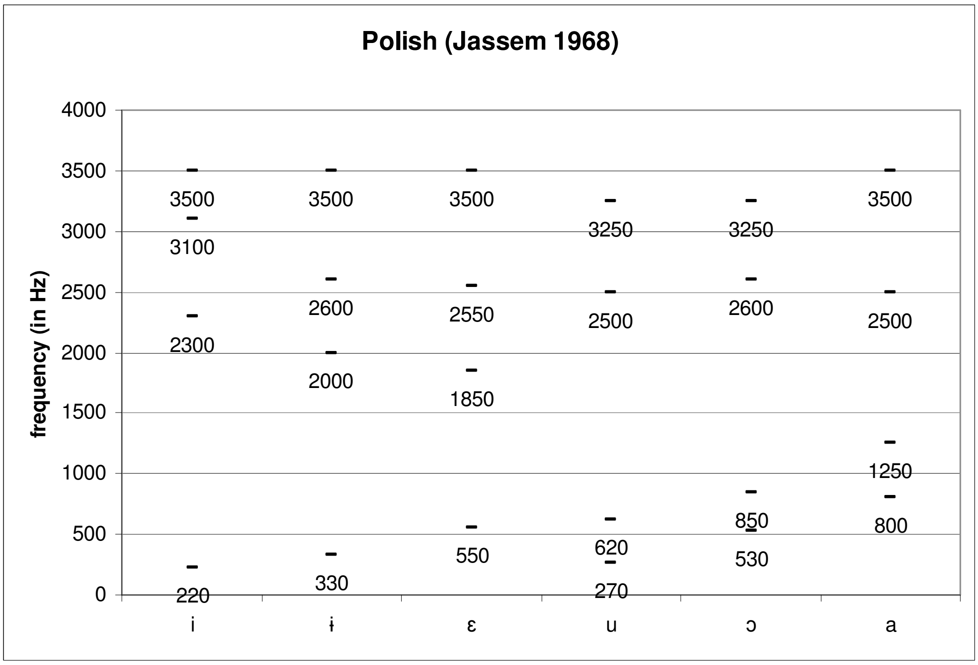

Figure 10.56: Formant frequencies for the first four formants (in Hz) for Polish vowels (Jassem 1968)

The phonemic status of */ɨ/* is controversially discussed with respect to (1) its status as a phoneme or as allophone of /i/ and (2) its description in terms of phonetic and phonological features. Several studies treat /ɨ/ as positional variant of /i/ after hard consonants (e.g. Bethin 1992: 32; Rothstein 1993; Rothstein 2006: 672; Stone 2011: 294), whereas others (e.g. Biedrzycki 1974; Rubach 1984; Savicka 1995; Jassem 2003) consider /ɨ/ as separate phoneme and argue for its phonemic status with the existence of some minimal pairs, e.g. *bić* ‘beat’ - *być* ‘be’ or *trik* ‘trick’ - *tryk* ‘ram’, though the functional load of the /i/ - /ɨ/ contrast is limited and in many other contexts /i/ and /ɨ/ are context-sensitively distributed[110]. Studies considering /ɨ/ as a separate phoneme do not agree on the exact description in terms of distinctive features:

[110] Stone (2011: 294) considers the vowel letters <i> and <y> to represent the same phoneme /i/, functioning as indicators for the quality of the preceding consonant, e.g. <ci dzi si zi pi bi mi fi wi ni> would correspond respectively to /ʨi ʥi ɕi ʑi p̦i ḅi ṃi f̣i ṿi ṇi/, whereas <cy dzy sy zy py by my fy wy ny> would represent /ʦi ʣi si zi pi bi mi fi vi ni/. The letter <i> is rarely written after the letters < d t cz dz sz ż ch ł r>, except in words of foreign origin and some onomatopoetic words. The letter <y> does not occur after <k g l>, with a few exceptions in words of foreign origin.

Biedrzycki (1974: 12, 61) describes /ɨ/ as pre-dorsal and “almost central”, Rubach (1984: 22) treats /ɨ/ as a high central non-rounded vowel, and many other studies describe /ɨ/ as a separate phoneme that is described as “central” or “centralized” (cf. Sawicka 1995; Nowak 2006). The commonly used description as /ɨ/ as “central” does mainly refer to the position of the vowel in an F1/F2 representation. However, as the formant frequency values for /ɨ/ given by Jassem (1968) indicate (see Figure 10.56 and Table 10.51), the low values for F1 (~330 Hz) together with a relatively high F2 (2000 Hz) and F3 (2600 Hz) clearly show that /ɨ/ is a front vowel. The affinity of F3 and F4 values for /i/ in Jassem’s data indicate that Polish /i/ is rather positioned in the very front part of the palatal constriction area.

Diphthongs have no phonemic status in Polish, but there are sequences of vowel plus glide /aj ɛj ɔj uj iw ɛw aw ɔw/, which are normally not followed by another consonant in the same syllable, suggesting that the glide occupies the coda position (cf. Bogacka 2009). /j/ only appears with a preceding or following syllabic vowel. /w/ for <u> is pronounced in some Greek- or Latin-derived words, e.g. *auto* /awtɔ/ ‘car’, *euphoria* /ɛwfɔrja/ ‘euphoria’ or for orthographic <ł>, e.g. *kłam* /kwam/ ‘falsehood, lie’.

Vowel length has no phonemic relevance in contemporary Standard Polish. Nowak (2006) presents a detailed study on the variability of Polish vowels under the influence of consonantal context in interaction with vowel duration. In his data, vowel duration and place of articulation of adjacent consonants prove to be good predictors for formant variability. The impact of preceding consonants was found to be stronger than that of post-vocalic ones.

Polish has no phonological vowel reduction, i.e. the inventory of vowels in stressed syllables is identical with the inventory of unstressed vowels. Phonetic vowel reduction in the sense of a qualitative reduction or “centralization” does hardly occur in Polish (Sawicka 1995; Jassem 2003; Konopka 2003; Rothstein 2006).

Word stress falls normally on the penultimate position and can fall on a preposition if the following noun or pronoun is monosyllabic, e.g. *ˈpod nim* ‘under it’, though there are exceptions with antepenultimate stress, mostly in words of Latin or Greek origin or some verb forms, e.g. *fonetyka* /fɔˈnetɨka/ ‘phonetics’. Secondary stress is usually placed in the initial syllable (Booij & Rubach 1984). Rothstein (2006: 673) describes a growing tendency, especially in emphatic speech, to shift stress to the initial syllable.

Polish is considered to be syllable-timed by some linguists (e.g. Hayes & Puppel 1985) and stress-timed by others (e.g. Rubach & Booij 1985; Ramus, Nespor & Mehler 1999), though Ramus, Nespor & Mehler (1999) stress the fact that Polish shows characteristics of both stress-timed as well as syllable-timed languages: Polish presents a great variety of syllable types and high syllabic complexity, a feature that is typical for stress-timed languages, but shows no vowel reduction at normal speech rates.

10.9.3 Contrastive Analysis

A comparison of the vowel inventories of Polish and German reveals the following systemic differences:

(1) The number of German vowel phonemes is considerably higher than in the Polish system.

(2) German uses a large number of vowel phonemes in the palatal constriction area. Palatal vowels in German can be [+/–round]. The series of German front rounded vowels has no counterpart in Polish. In other words, the combination of the features [+palatal] and [+round] is not found in Polish.

(3) Polish /ɨ/ has no counterpart in German.

(4) In contemporary Polish, duration is not used for contrastive purposes. There are neither long nor tense vowels in Polish.

(5) German diphthongs have not direct counterpart in Polish, where only sequences of vowel plus glide or vowel plus nasal exist.

(6) Nasality functions contrastively in Polish as in *kot* /kɔt/ 'cat' – *kąt* /kɔ̃t/ 'angle', but not in Standard German.

(7) The reduction of vowels in unstressed syllables does not occur to the same extent in Polish as in German. German e-schwa and a-schwa have no counterparts in Polish.

Several studies investigated the influence of L1 Polish on the acquisition of the German vowel system in production as well as perception. Evidence has been gathered either by empirical investigation (Górka 1973; Hentschel 1981, 1986; Kerschhofer 1995) or by a contrastive analysis and experience gained in teaching practice (Ortmann 1976; Morciniec & Prędota 1973; Prędota 1979). Table 10.52 summarizes the results of comparative studies on the acquisition of the German vowel system for those vowels that are commonly considered as difficult for Polish learners in L2 German.

Hentschel (1981) analyzed read speech of Polish learners of German and identified four major types of "phonological interference" in production: (1) *shortening* of German /a: i: u:/ (33%), /e: o:/ (25%) and /y: ø:/ (7%) (irrespective of other qualitative changes of the vowel), (2) *diphthongization* of German long vowels /e:/ > [ɛi], /o:/ > [ou] (27%) and /ø:/ > [øi] (10%), (3) *delabialization* of front rounded vowels /ø:/ > [e:], /ø/[111] > [ɛ] (17%), /y:/ > [i:], /y/ > [i]

[111] [sic!]. Hentschel (1981) denotes /ʏ/ and /œ/ as /y/ and /ø/.

(45%), and (4) *raising and diphthongization* of front labial vowels /øː/ > [yøː] (4%) and /ø/ > [yø] (21%). Substitutions of /øː ø/ > /yː ʏ/ are regarded as "exceptional". For /a ɛ ɪ ɔ ʊ/ no substantial difficulties are described.

German target V	/iː/	/ɪ/	/eː/	/yː/	/ʏ/	/øː/	/œ/	/uː/	/ʊ/	/oː/	/ɛː/
Górka (1973)[112]	i	i / ɨ	ɛ / ɛi / iɛ	i / iu	i / ɨ	ɛ / ɛi	ɛ	u	u[113]	ɔ / ɔu[114]	ɛ
Morciniec & Prędota (1973[115]/1982)	i	ɨ	ɛ	i	i	ɨ ɛ / ɛi (1982)[116]	ɨ ɛ / ɨ (1982)	u	u	ɔ	ɛ
Prędota (1979)[117]	i	i / ɨ	ɛ	i	i	ɨ	ɨ	u	u	ɔ	ɛ
Hentschel (1981)[118]	i	-	ɛ, ɛi	i, iː	i	œ, øi, eː, yː, yøː	ɛ, yø	u	-	ɔ, ou	-
Hentschel (1986)[119]	i	i ~ ɨ (context-sensitive)	i ~ ɨ (context-sensitive)	i ~ ɨ (context-sensitive)	ɨ	ɨ (88%) ɛ (6%)	ɛ	u	u / ʊ	u / ʊ	ɛ

Table 10.52: Comparison of previous studies on the interference of Polish vowels on the acquisition of German vowel categories (cf. Kerschhofer 1995: 108)[120].

Hentschel (1986) investigated the perception of German vowels by naïve Polish listeners in three different experiments: Test 1 was a perceptual assimilation task ("Basisidentifikationstest"), where Polish listeners had to identify German vowel stimuli as either belonging to one of the Polish vowel categories or as "foreign". Test 2 was an identification task where listeners had to identify German long vowels as either [V], [Vː], [Vi], or [Vu] to find out whether diphthongizations as observed in Hentschel (1981) are due to perceptual or to articulatory reasons. Test 3 was a discrimination task for vowel contrasts found to be difficult in production. The context-sensitive distribution of /i/ and /ɨ/ in Polish depending on the preceding consonants receives special interest in Hentschel's study and is considered to be of relevance for the vowel perception and realization in L2 German. Hentschel (1986: 80ff) discriminates "i-contexts", "ɨ-contexts" and "i- and ɨ-contexts" and

[112] analysis of read speech.
[113] and o-like realizations.
[114] and u-like realizations.
[115] contrastive comparison for teaching purposes, see also Morciniec & Prędota (1982).
[116] Morciniec & Prędota (1979) differ from Morciniec & Prędota (1982) in this respect.
[117] contrastive analysis and prediction of errors.
[118] auditive evaluation of read speech.
[119] perceptual classification in an identification task.
[120] For German /a/, /aː/, /ɛ/ and /ɔ/, all authors agree on substitutions with /a/, /ɛ/, and /ɔ/ respectively.

divides German consonants into three groups according to the context-sensitive distribution of /i/ and /ɨ/ before consonants: The group of "i- and ɨ-consonants" consists of German labials and nasals /p b f v m n/ that can be followed by /i/ as well as /ɨ/ in Polish. German /k g/, /l/ and /ç/ are assumed to be followed only by /i/ but not /ɨ/ (in Polish, velars and /l/ are palatalized before /i/). "Only ɨ"-contexts are dental-alveolar obstruents and German /r/ and /R/ that occur only before /ɨ/ but not /i/ in Polish (with the exception of some words of foreign origin).

From the percentage a given vowel phoneme was classified as "foreign". i.e. as "not Polish" in Test 1 (Basisidentifikationstest), Hentschel (1986) derived a measure for the "degree of foreigness" or acceptability of German vowels for speakers of Polish. Hentschel (1986) hypothesized that /y: ʏ ø: œ/ and /e: ɪ o: ʊ/ will be qualitatively "foreign" to Polish native listeners. As evident in Figure 10.57, front rounded vowels were classified most frequently as "foreign" by Polish listeners. Long vowels showed a higher degree of "foreigness" than their short counterparts. The higher and longer a front rounded vowel the rather it was classified as a "foreign sound". /œ/ was classified as less "foreign" than /ʏ ø: y:/.

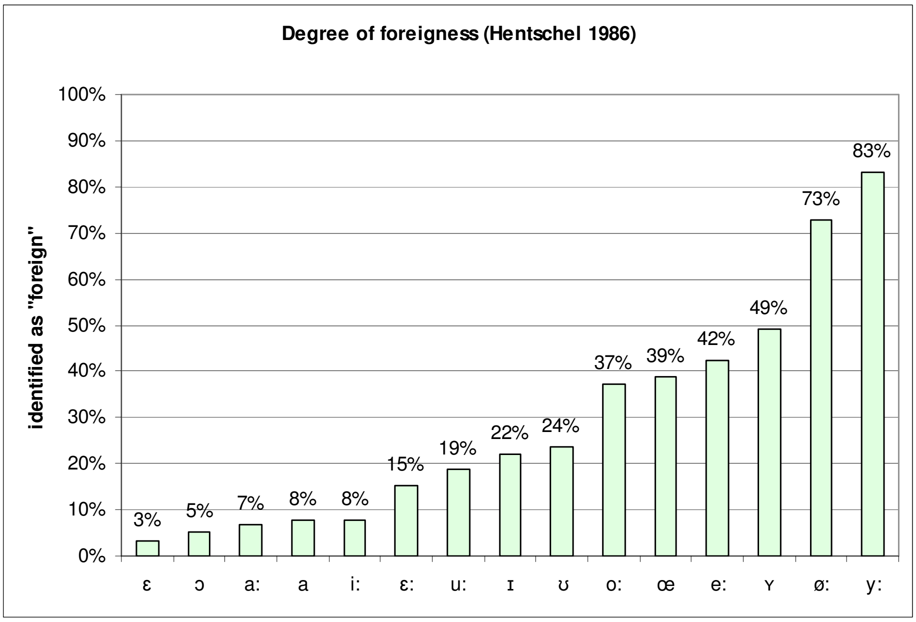

Figure 10.57: Hentschel's (1986) scale of foreigness/acceptability of German vowels for Polish listeners

In a discrimination task in Hentschel (1986), /e:/ - /ɛ/ were discriminated in 100% of the cases, /i:/ - /ɪ/ were discriminated in 98% of the times, but /e:/ -/ɪ/ were discriminated in only 70% of the cases. /i:/ was discriminated from /ɪ/ and /y:/ in 98%, the contrast /ɛ/ - /œ/ was perceived in only 50% of the times; The contrasts /ɪ/ - /ʏ/, /ø:/ - /e:/ and /i:/ - /y:/ were better discriminated, while the contrast of /ø:/ - /ʏ/ appeared to cause more difficulties.

Table 10.53 summarizes the perceptual substitutions observed by Hentschel (1986). German /ɑ ɑ: ɛ ɛ: ɔ i: u:/ were mostly identified with their Polish equivalents and were classified as "foreign" in less than 20% of the cases. Other vowel qualities with higher scores of "foreigness" showed either invariant or variable patterns of perceptual substitutions with Polish vowels: /ø:/ was identified with Polish /ɨ/ in 88% and with /ɛ/ in 6% of the cases, /ʏ/ was mostly identified with /ɨ/, and /œ/ was most frequently identified with Polish /ɛ/. For German /ɪ e: y:/ and /ʊ o:/, more than one substitute was selected by the listeners. For /ɪ e: y:/ the context-sensitive distribution of Polish /i/ - /ɨ/ appeared to be of relevance in perception (see Table 10.53), whereas German /i:/ and /ø: ʏ/ were not affected by this influence of consonantal context.

German	i:	e:	ɪ	y:	ø	ʏ	ɛ	ɛ:	œ	a:	a	ɔ	o:	ʊ	u:
Polish	i	i ~ ɨ			ɨ		ɛ			a		ɔ	ɔ ~ ʊ		u

Table 10.53: Perceptual substitutions of German vowels with Polish vowel categories by L1 Polish listeners (Hentschel 1986)

In Kerschhofer's (1995[121]) study on vowel substitutions by Polish learners of German, perception and production data were elicited in different types of tasks (discrimination, feature identification, production of single vowels, production in an imitation task, translation and spontaneous speech). The data obtained in the different sub-tests were analyzed and interpreted within the framework of Natural Phonology assuming that difficulties and errors in pronunciation as well as production are due to substitution processes. The elicited pronunciation data were transcribed in a close IPA transcription, from which the application of natural phonological substitution processes was derived. Vowel perception was tested in two types of tasks, a discrimination task and a feature identification task. Unlike many previous studies on the pronunciation of German vowels by Polish learners, the transcriptions in Kerschhofer (1995) considered that deviant pronunciations cannot only be described as direct realizations of corresponding L1 sounds but that there are sub-phonemic differences in vowel realizations in the learners' interlanguage that can be intermediate between L1- and L2-sounds in their phonetic quality. Deviations from the target language in perception or production are described as the outcome of one or more natural phonological fortition

[121] see also Kerschhofer (1998)

processes (cf. Donegan 1978; for discussion, see section 1.2.4, 4.7.2 and 11.2) which could affect vowel quality and/or vowel quantity. Some of the results will be reported here.

Wrong realizations can be due to either a lack of motor control in production or due to perception-based deviations and result in substitutions of the difficult vowel by another one. Within the framework of *Natural Phonology* (see section 1.2.4 and 4.7.2), fortition processes such as fronting, labialization and delabialization, raising and lowering, centralization, lengthening, shortening, or diphthongization are considered to result in *substitutions* of difficult sounds with other sounds that differ with respect to the difficult feature(s) while maintaining other characteristic features of the segment (Donegan 1978; Kerschhofer 1995, 1998a).

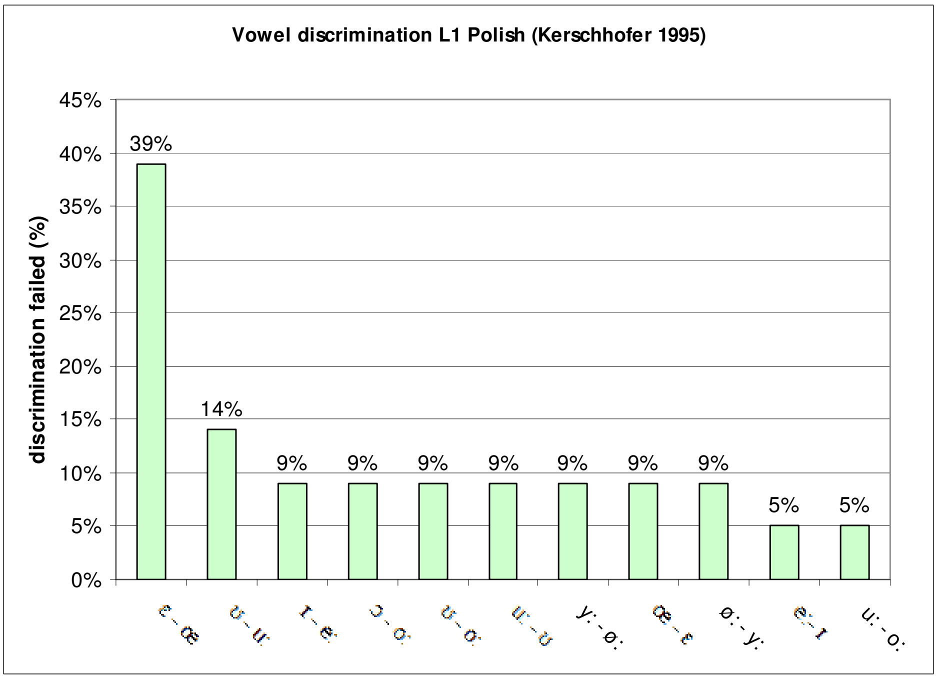

Figure 10.58: Discrimination of German vowel pairs by L1 Polish learners of German (11 subjects) (Kerschhofer 1995)

In a *discrimination task*, Kerschofer (1995) tested the perceptual discrimination ability for vowel contrasts that had been described as difficult to Polish learners in previous studies. The subjects had to decide whether two members of a presented vowel pair were "equal" or "different from each other". Interesting differences with respect to different vowel categories as well as with respect to the order of vowels in a contrast pair were observed (see Figure 10.58). The /ɛ/ - /œ/-contrast caused most difficulties in discrimination. In 39% of the cases, the listeners failed to discriminate the two qualities for /ɛ/ - /œ/. However, when these two vowels occurred in the inverse order (/œ/ - /ɛ/), qualitative differences were better discriminated (only 9% “equal”). A similar but weaker order effect was observed for /ɪ/ - /e:/ (9%) vs. /e:/ - /ɪ/ (5%), but not for /y:/ - /ø:/ and /ø:/ - /y:/ (Kerschhofer 1995: 122f; see also Barry 1975).

In a *feature identification* task, subjects were asked to identify test items as [+/–round] and [+/–long], assuming that features that are difficult in L2 German are not only difficult to realize in production but are also difficult in perception. The failure to perceive the feature [round] in rounded vowel qualities can be described as caused by a process of perceptual delabialization resulting in a substitution with equivalent non-rounded qualities. Figure 10.59 shows that the identification of the feature [round] in short unconstricted /œ/ and /ʏ/ is significantly more difficult than in long constricted /ø:/, /u:/ or /y:/. /i:/ and /ɪ/ were hypercorrectly categorized as [+round] in a few cases, whereas e-vowels were never perceived as [+round]. These results are highly consistent with the results in the production task and with the predictions of Natural Phonology. Donegan (1978: 85ff) states that delabialization is more likely to occur with mixed, i.e. labio-palatal, lower and laxer vowels. Therefore, short unconstricted /œ/ is the best candidate for delabialization in perception as well as in production, a fact that is very much compatible with the articulatory properties of front rounded vowels described in section 5.4.1 and 5.4.2.

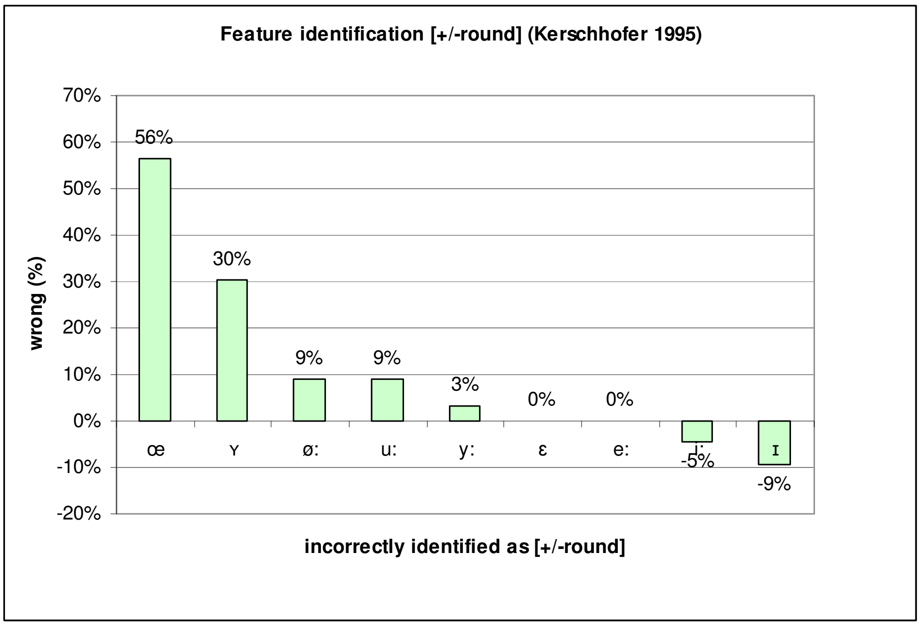

Figure 10.59: Results of the feature identification task (Kerschhofer 1995). Wrong and hypercorrect identifications of the feature [round] in %.

In the transcribed pronunciation data elicited in four types of imitation tasks, about 50% from a total number of 1627 vowel realizations deviated from the accepted standard. The data showed considerable variation between categories in the number of deviant realizations and in the type of processes that were applied by the learners. Long constricted qualities showed a higher percentage of deviant realizations than their short unconstricted counterparts. Interestingly, the highest number of deviant pronunciations was not found for the front rounded ü- and ö-vowels, but for long constricted /o:/, /e:/, and /u:/ (see Figure 10.60).

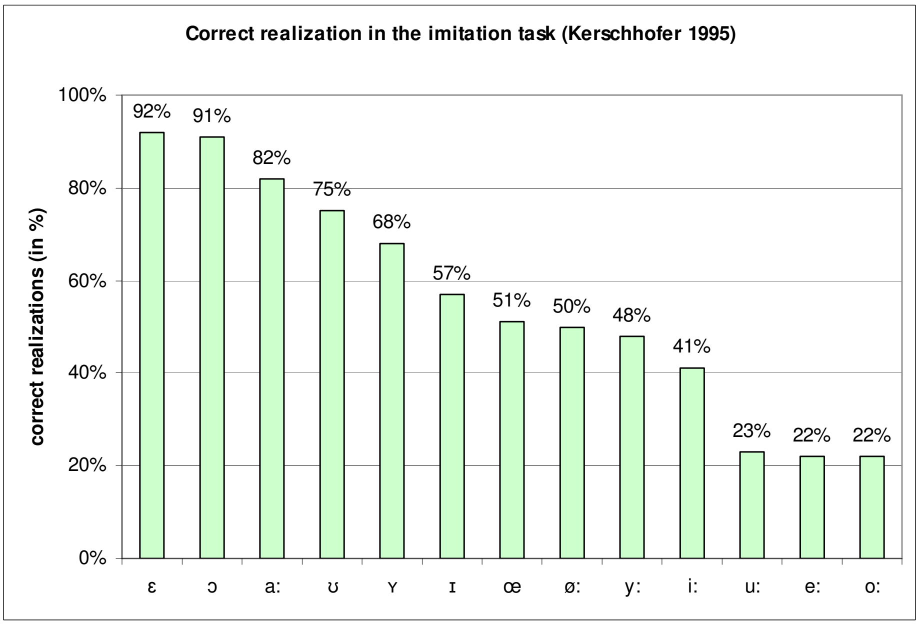

Figure 10.60: German vowel production percentage correct by 11 Polish speakers (imitation task) (Kerschhofer 1995).

The realization of phonemic length in production was found to be easier with lower (and more open) /ɛ:/ and /ɑ:/ than with higher and more constricted vowels /e:/, /i:/ and /y:/.

palatal			labiopalatal			labial		
	diphthongization	shortening		diphthongization	shortening		diphthongization	shortening
i:	24%	23%	y:	10%	27%	u:	25%	32%
ɪ	1%	-	ʏ	-	-	ʊ	-	-
e:	50%	15%	ø:	10%	18%	o:	21%	28%
ɛ	1%	-	œ	-	-	ɔ	-	-

Table 10.54: Diphthongization and shortening in vowel production (Kerschhofer 1995, 1998a)

One strategy to overcome articulatory difficulties with long vowels observed with Polish learners of German was diphthongization, i.e. a heterogeneous articulation, another strategy was vowel shortening. The process of diphthongization was easier suppressed with high /i:/ than with lower /e:/ and /o:/. With rounded vowels, a heterogeneous realization was more frequently observed with /u:/ and /o:/ than with /y:/ or /ø:/, the latter being of course subject to processes that aim at other difficult features of these qualities. A diphthongized realization of long constricted /e:/ and /o:/ or their substitutions with unconstricted [ɛ] and [ɔ], respectively, corresponding to the equivalent Polish vowel qualities was one of the most persistent characteristics of L1 Polish learners of German that was observed even with very advanced and near-native speakers.

Qualitative differences concern either combinations of the features [front/back] and [round] or the relative height and/or degree of constriction.

V target	delabialization
y:	15%
ʏ	20%
ø:	12%
œ	31%

Table 10.55: Delabialization of front rounded vowels (Kerschhofer 1995, 1998a)

Rounding in front rounded vowels was found to be easier in pronunciation (see Table 10.55) and perception with long constricted than with short unconstricted vowels, but also with higher /y:/ and /ʏ/ than lower /ø:/ and /œ/. /œ/ was found to be the most difficult category with respect to [rounding] both in perception and production. In other words, it was most frequently subject to a process of delabialization (see Table 10.56).

palatal			labiopalatal			labial		
	lowering	raising		lowering	raising		lowering	raising
i:	14%	-	y:	15%	-	u:	42%	-
ɪ	3%	20%	ʏ	8%	-	ʊ	11%	8%
e:	19%	2%	ø:	4%	17%	o:	44%	-
ɛ	-	4%	œ	2%	15%	ɔ	-	9%

Table 10.56: Raising and lowering in vowel production (Kerschhofer 1995, 1998a)

The results are very compatible with the predictions of Natural Phonology. Donegan (1978) posits processes of raising and lowering by which the height of a vowel is increased or decreased by one degree. As Table 10.56 shows, raising is rather observed with more chromatic and tense vowels, whereas lowering is rather observed with more achromatic, laxer, longer and mixed, i.e. labio-palatal vowels (Donegan 1978: 77f). However, these implicational conditions are not fully confirmed by the data, since not all vowels are affected to the same degree with a specific process. Rather, L2 learners seem to choose between different strategies, i.e. processes eliminating the feature that is difficult to them, to cope with difficulties they meet in L2 production.

To summarize, substitutions of L2 target sounds are induced either by difficulties in perception or in production (see also Kerschhofer 1998a). Difficulties observed in the pronunciation data concern contrasts in quality and/or quantity. The results of the perception studies in Hentschel (1986) and Kerschhofer (1995) suggest that Polish listeners have considerable difficulties with the correct discrimination and identification of several German vowel categories. Perceptual difficulties of Polish leaners of German will be investigated in more details in the present study.

10.9.4 Results and interpretation

31 Polish subjects were tested. All subjects were students at the University of Wrocław. The sample is divided into two groups, 22 beginners and 9 advanced learners. Beginners were participants of the general language courses for German in their first year of study. Advanced learners were studying German at the Department of German studies and were in the second year of their studies. Since the Polish sample is systematically controlled for the listeners' level of proficiency in German, a comparison of results for beginners vs. advanced learners will be provided for some issues.

The sample presented here consists of 9 advanced learners (with an average of 60% correct responses) and 22 beginners (with an average of 43% correct responses). The data set discussed here consists of 8260 valid responses (107 missing answers). For each vowel category, 556 to 558 valid responses were delivered.

Generally, the sample shows a very high percentage of wrong responses for Polish learners of German: The listeners' mean id_wrong score was 52% (12% standard deviation), the median was 53%. The listener with the highest percentage of correct answers had an id_wrong of 15%, the highest id_wrong score was 69% wrong responses. The high percentage of incorrect answers (mean of 52% id_wrong) is partly also due to the structure of the sample and the percentage of beginners vs. advanced learners and partly due to the structure of the German and the Polish vowel system.

10.9.4.1 Difficulties

Figure 10.61 presents the percentage of wrong and correct responses for L1 Polish listeners. The most difficult categories are /øː/ (16% id_correct), /ɛː/ (20% correct), /œ/ (21% correct) and /eː/ (24%). Least difficulties are observed for /ɑː/ (82.5% correct).

/iː/, /ɛ/, /ɪ/, /ɔ/, /ɑ/, and /uː/ show very similar id_correct scores between 65% and 61%. Lower scores are observed for /ʊ/ (55%), /yː/ (49%), /ʏ/ (40%), and /oː/ (35%). All front rounded vowels, show a low percentage of correct categorizations, especially /øː/ (16%) and /œ/ (21%). Major difficulties are also observed for /ɛː/ (20% correct) and /eː/ (24% correct), but not for /ɛ/ (64% correct). These results are consistent with results in Hentschel (1986) and Kerschhofer (1995).

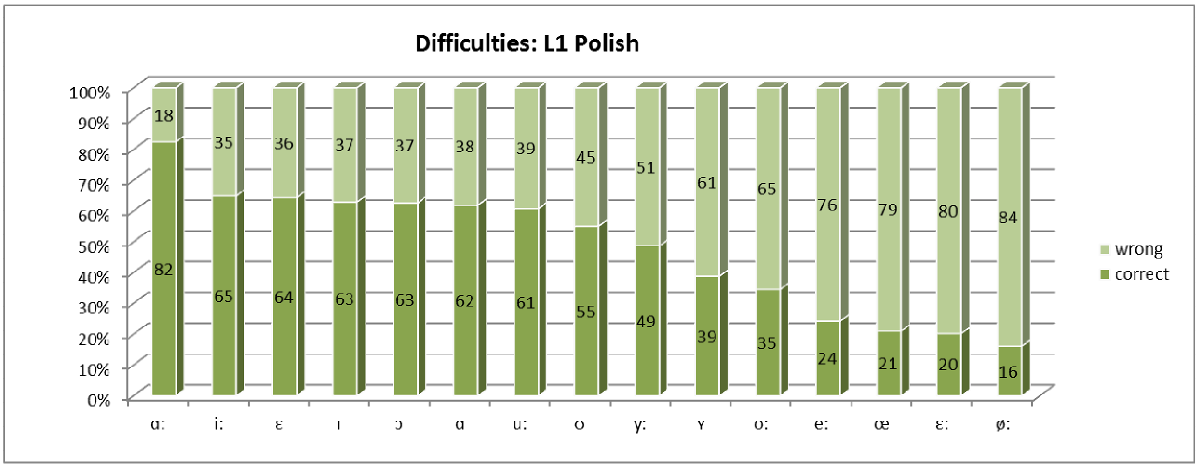

Figure 10.61: Correct and wrong identifications in % for L1 Polish listeners (N= 31, 8260 responses)

A comparison of beginners (Table 10.58) and more experienced language learners (Table 10.59) shows that the correct identification of i-vowels and of /oː/, /ʏ/ and /yː/ improves considerably with advanced learners, whereas for /ɛː/ no improvement at all is observed. While advanced learners have less difficulties with /yː/ and /ʏ/ than beginners, the ö-vowels still cause persistent problems even to advanced learners.

A comparison of id_wrong scores for all L1 Polish subjects (see the boxplots in Figure 10.62) shows interesting differences between categories.

For a few categories, /ɑː/, /ɛ/, /iː/, /ɔ/, /ʊ/ and /uː/, 100% id_correct scores are reached by at least one participant. A relatively low percentage of wrong responses is also achieved for /ɑ/, /ɪ/, /yː/, /ʏ/, and /œ/ by some subjects. Long constricted /øː/ is not only the most difficult category, it also shows a very high percentage of wrong responses even with advanced learners of German. A similar but less strong effect is observed for /ɛː/, /eː/, /œ/, and /oː/. For these vowels, considerable difficulties are observed with the establishment of stable perceptual categories, though there are a few participants who seem to cope better with the correct identification of these categories.

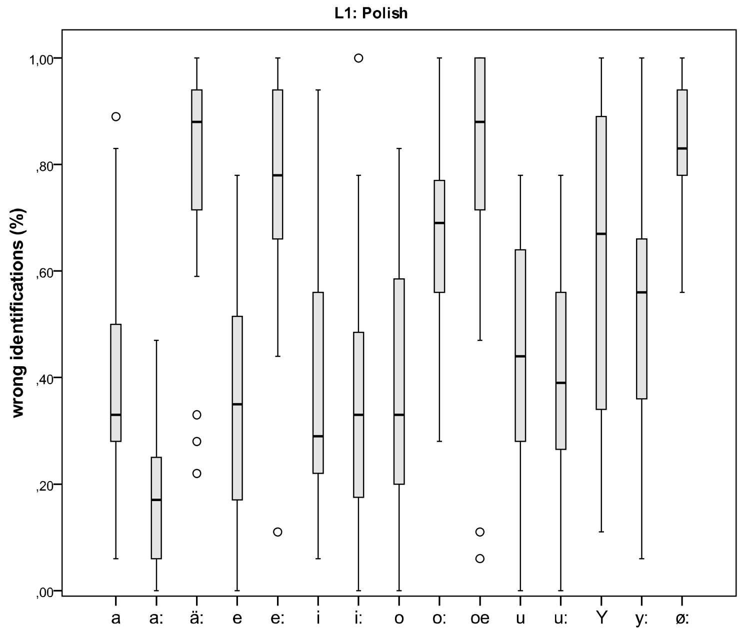

Figure 10.62: Id_wrong scores for L1 Polish listeners

10.9.4.2 Preferences

Considerable asymmetries for the pref_scores of German vowel categories are observed for L1 Polish learners (see Figure 10.63). The preference scores for monophthong categories vary considerably within a range of 6.1% (min /ɛ:/ 3.0%, max /y:/ 9.1%). Those categories that were described above as difficult are also least preferred in the responses of L1 Polish listeners. The lowest pref_scores are observed for /ɛ:/ (only 3%), /œ/ (3.4%), and /ø:/ (3.5%). Pref_scores below 5.5% are observed for /o:/ (5.1%), short /ɑ/ (5.2%), and /ɔ/ (5.4%). Higher pref_scores (≥ 6.9%) are observed for /ɛ/, /ʏ/, /ʊ/, /ɪ/, /e:/, and /ɑ:/. The most preferred categories are /y:/ (9.1%), /u:/ (9%) and /i:/ (8.9%).

Interestingly, long constricted categories seem to be more attractive response options than short unconstricted categories, with a few exceptions: /o:/, /ø:/, and /ɛ:/ are less frequently selected than their short counterparts /ɔ/, /œ/ and /ɛ/.

Diphthongs are quite frequently selected response options. The most preferred diphthong is <ei> with a pref_score of 2% (mostly substituted for e-vowels). <eu> is selected in 0.6% of the cases (mostly for rounded vowels), and <au> is hardly ever selected (0.1%, mostly for /o:/ (3 cases) and /ʏ/ in 2 cases).

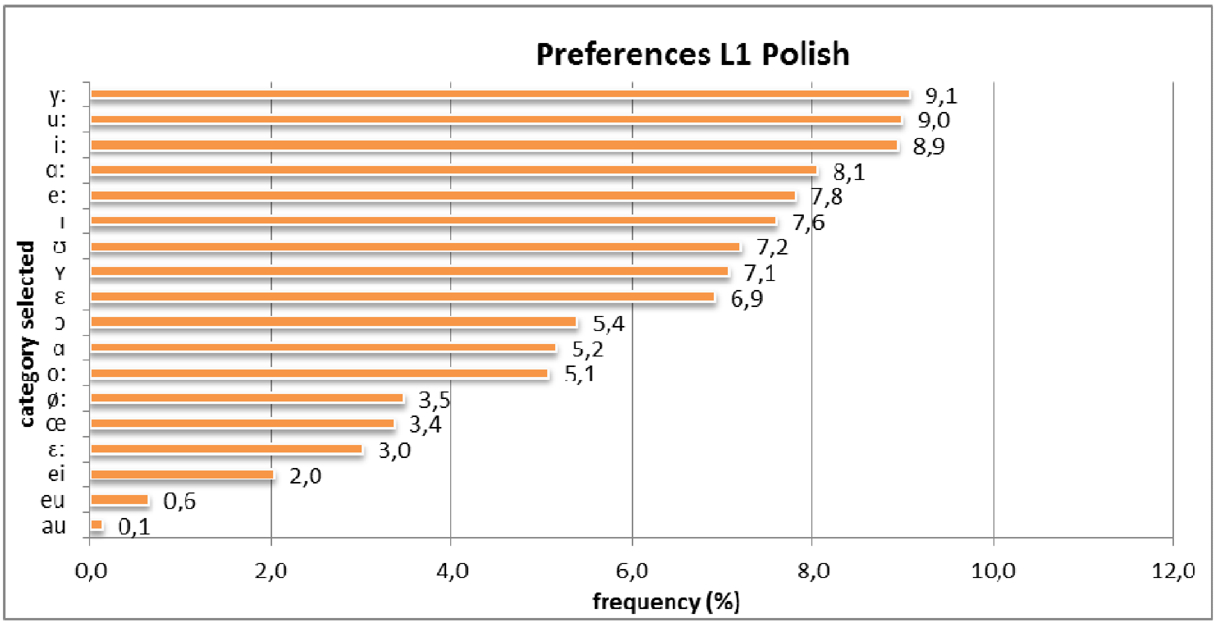

Figure 10.63: Preferences L1 Polish

10.9.4.3 Patterns of confusion

The confusion matrix in Table 10.57 represents the perceptual substitution patterns for L1 Polish learners of German. A first look at the matrix shows three major areas of intercategory confusion: (1) confusion within the class of *front non-rounded vowels*, especially for e-vowels and confusion of /e:/, <ei>, i- and ü-qualities, (2) considerable difficulties with all *front rounded vowels* resulting in substitutions with other front rounded vowels, u-vowels and front non-rounded qualities, and (3) confusion within the class of *back rounded vowels* resulting in substitutions back as well as front rounded vowels. The listeners' responses spread over several categories.

The confusion matrix in Table 10.57 does not show id_scores below 0.5%, id_scores between 0.4% and 1.4% are printed in grey. Table 10.58 and Table 10.59 represent the results for L1 Polish beginners and for advanced learners.

Difficulties for *a-vowels* concern almost exclusively length distinctions. While L1 Polish listeners are quite successful in identifying length in /ɑ:/-stimuli (82% correct, 15% /ɑ/), they identify short /ɑ/ (only 62% correct) in 36% of its occurrences hypercorrectly as long /ɑ:/. Occasionally, other categories are selected for a-vowels, e.g. <ei> for /ɑ:/ in 4 cases and /ɛ:/ for /ɑ/ in 3 cases.

German *e-vowels* can be divided into two subclasses: For ɛ-qualities, Polish listeners show a similar behaviour that clearly differs from responses for /e:/. Least problems are observed for /ɛ/ (64% correct), which is quite similar to the Polish e-vowel. /ɛ/ is identified incorrectly as /e:/ in 23% of its occurrences and /ɛ:/ in 7% of the cases.

Long unconstricted /ɛ:/ shows a very low percentage of correct answers (only 20%). Its most preferred substitute is long constricted /e:/ (46%). Less preferred response options for /ɛ:/ are <ei> (15%) and /ɛ/ (12%), occasionally, other categories are selected, e.g. /i:/ (2%), /ɪ/ (6 cases), /œ/ (6 cases) and /eu/ (5 cases).

%	ɑ	ɑ:	ɛ:	ɛ	e:	ɪ	i:	ɔ	o:	ʊ	u:	œ	ø:	ʏ	y:	‹ei›	‹eu›	‹au›	wrong	correct
a	**62**	36	1																38	**62**
a:	15	**82**														1			18	**82**
ɛ:			**20**	12	46	1	2					1	1		1	15	1		80	**20**
ɛ			7	**64**	23	1	1						1			1			36	**64**
e:			7	3	**24**	11	35						1	2	4	11			76	**24**
ɪ			1	3	3	**63**	16					1	1	7	3	1			37	**63**
i:					2	29	**65**							1	2				35	**65**
ɔ							1	**63**	31			1	1	1			1		37	**63**
o:							1	7	**35**	12	27	2	8	2	5		1	1	65	**35**
ʊ								8	4	**55**	21	3	2	4	1				45	**55**
u:									3	21	**61**	1	1	5	7		1		39	**61**
œ		1	8	20	12	1	1	2	1	1	1	**21**	8	13	6	1	3		79	**21**
ø:					1	1	1		1	4	13	6	**16**	15	41				84	**16**
ʏ					2	3	2			11	6	10	6	**39**	19		1		61	**39**
y:					1	2	8			4	6	3	6	19	**49**		1		51	**49**
Σ	*5*	*8*	*3*	*7*	*8*	*8*	*9*	*5*	*5*	*7*	*9*	*3*	*3*	*7*	*9*	*2*	*.6*	*.1*	*52*	***48***

Table 10.57: Confusion matrix for L1 Polish listeners (N=31, 8260 responses). Correct (in bold) and wrong identifications in % (presented stimuli in rows, selected categories in columns in semi-orthographic representation, values < .5 not indicated, values < 5.5 in grey)

Long constricted /e:/ (24% id_correct) seems to cause considerable difficulties to Polish subjects. Its most preferred substitute is /i:/ (35%), followed by <ei> (11%), /ɪ/ (11%), /ɛ:/ (7%), /y:/ (4%), /ɛ/ (3%), /ʏ/ (2%). Occasionally, substitutions with ö-vowels (6 cases) and <eu> (2 cases) are observed.

Within the class of e-vowels, a strong asymmetric pattern is observed: Mid-palatal /e:/ is substituted with pre-palatal /i:/ in 35% of its occurrences, whereas /i:/ stimuli are rarely substituted with /e:/ (only 2%). Long unconstricted /ɛ:/-stimuli are categorized as long constricted /e:/ in 46% of the cases, while the inverse pattern is significantly less frequently observed (/e:/ > /ɛ:/ 7%). These patterns together with the high id_wrong score for /e:/ reveal that the differentiation of the mid-palatal and the pre-palatal constriction location in German is specifically difficult for L1 Polish listeners, which is of course due to the fact that this contrast is not exploited in their native language, where palatal /i/ and /e/ [ɛ] differ mainly in degree of aperture.

/e:/ is significantly better identified by advanced learners (47% correct) than by beginners (only 15% correct) (see Table 10.58 and Table 10.59), but substitutions of /e:/ > /i:/

(beginners 38%, advanced 29%) are a significant perceptual pattern in both groups, whereas <ei> is a very non-attractive option for advanced learners.

German *i-vowels* show relatively good results with respect to id_correct scores. /i:/ is identified correctly in 65% of its occurrences and is mainly confused with short unconstricted /ɪ/ (29%). Only occasionally substitutions with other categories occur, especially with /e:/ (2%) and /y:/ (2%), and less frequently with /ʏ/ and ö-vowels. The differentiation of short unconstricted /ɪ/ from other palatal categories seems to cause more problems to Polish listeners. /ɪ/ (63% correct) is most frequently confused with /i:/ (16%), but also with /ʏ/ (7%), /e:/ (3%), /ɛ/ (3%), /y:/ (3%), and less frequently with ö-vowels, /ɛ:/, and <ei> (1% each).

Large differences in id_correct scores are observed for the *o-vowels* (/ɔ/ 63% correct vs. /o:/ 35% correct). Short unconstricted /ɔ/ (63% correct) is one of the most stable categories, only the contrast /ɔ/ - /o:/ and the differentiation with respect to phonemic length and constriction degree cause considerable problems to Polish listeners. The contrast /ɔ/ - /o:/ shows an interesting pattern of asymmetry: A substitution of /ɔ/ > /o:/ occurs in 31% of the cases, whereas /o:/-stimuli are categorized as /ɔ/ in only 7% of the cases.

While German /ɔ/ is very similar to its Polish equivalent, especially with respect to the degree of aperture, the more constricted German /o:/ cannot be sufficiently differentiated from u-vowels.

For /o:/ (only 35% correct), a high degree of instability in category formation is indicated by the wide range of different categories that are selected as response options for long constricted /o:/ (only 35% correct, /ɔ/ 7%). Almost 30% of the responses are substitutions with velar u-qualities: /u:/ (27%) and /ʊ/ (12%) are the most preferred substitutes, indicating the insufficient differentiation of /o:/ from u-vowels. The instable categorical representation of long constricted /o:/ is moreover reflected by substitutions with front rounded categories (/ø:/ (8%), /y:/ (5%), /œ/ (2%), /ʏ/ (2%)), and by occasional substitutions with other qualities such as <eu>, <au>, and i-vowels. A comparison of beginners and advanced learners shows that responses for /o:/ are less widespread for advanced learners and that sensitivity for length seems higher with advanced learners.

For German *u-vowels*, an inverse pattern is observed: While the perceptual representation for long constricted /u:/ (61% correct) seems quite stable, the response behaviour for short unconstricted /ʊ/ (55% correct) appears to be more diverse. Long constricted /u:/ is identified correctly in 61% of its occurrences and is categorized as /ʊ/ in 21% of the cases, as /y:/ in 7%, /ʏ/ in 5% and /o:/ in 3% of the cases, but as /ø:/ or /œ/ in 1% of the cases.

/ʊ/-stimuli are identified correctly in 55% of the cases and mainly substituted with /u:/ (21%), but also with /ɔ/ (8%) or /o:/ (4%). However, front rounded vowels also function as substitutes for short unconstricted /ʊ/: /ʏ/ (4%), /œ/ (3%), /ø:/ (2%) or /y:/ (1%), indicating difficulties in the differentiation of front rounded and back rounded vowels.

The correct identification of front rounded *ö- and ü-vowels* causes severe problems to Polish listeners. The instable perceptual representation of these categories becomes evident from the high percentage of wrong categorizations (id_correct scores between 16% and 49%) and from the wide range of different response categories that are selected for front rounded stimuli. Confusion of front rounded vowels with either (1) other front rounded vowels, or (2) back rounded vowels, or (3) front non-rounded vowels is observed. However, the response patterns differ considerably between categories. While front non-rounded categories are very frequently selected for /œ/-stimuli, and also attract responses for /ʏ/ and /y:/, they are infrequently selected for /ø:/-stimuli. Back rounded categories are attractive response options for /ø:/, /ʏ/ and also /y:/, but not for /œ/-stimuli.

/y:/ is the most stable category among front rounded vowels with an id_correct score of 49%. /y:/ is most frequently substituted with /ʏ/ (19%). Other response categories selected for /y:/-stimuli are ö-, u- and i-vowels, i.e. /i:/ (8%), /ø:/ (6%), /u:/ (6%), /ʊ/ (4%), /œ/ (3%), /ɪ/ (2%) and in a few cases also e-qualities, mostly /e:/ (6 cases), and <eu> (5 cases).

Responses for short unconstricted /ʏ/ are a little more dispersed than for /y:/, which is not evident in Table 10.57 due to several low scores below 0.5. /ʏ/ (39% correct) is most frequently substituted with /y:/ (19%), /ʊ/ (11%), /œ/ (10%), /u:/ (6%), /ø:/ (6%), /ɪ/ (3%), /i:/ (2%), /e:/ (2%), and in 3 cases with <eu>.

/œ/ shows a very low id_correct score of 21%. For /œ/ the widest range of selected response categories is observed; all categories except /ɑ/ and <au> are selected for /œ/-stimuli, indicating severe problems with the identification and perceptual representation for this vowel category. /œ/-stimuli are most frequently substituted with other front rounded categories (in total 27% of the responses) or with front non-rounded vowels (40% of the responses). This response behaviour shows that for this specific vowel quality the feature [rounding] is not sufficiently perceived by Polish learners, due to the generally low degree of lip rounding for /œ/, and that a differentiation with respect to the place of constriction (palatal vs. other) causes a little less difficulties. Incorrectly selected response categories for /œ/ are /ʏ/ (13%), /ɛ/ (20%), /e:/ (12%), /ø:/ (8%), /ɛ:/ (8%), /y:/ (6%), <eu> (3%), /ɔ/ (2%), and in single cases also other categories.

/ø:/ shows the lowest id_correct score, i.e. it is the most difficult quality for L1 Polish learners of German. Only 16% of the /ø:/-stimuli are correctly identified. Similar to /œ/, the listeners' responses spread over a wide range of different categories except a-vowels, and /ɛ/, /ɔ/, or <ei>, from which it is clearly differentiated. The most attractive candidates for perceptual substitutions are /y:/ (41%), /ʏ/ (15%), and /u:/ (13%), while short unconstricted /œ/ (6%) and /ʊ/ (4%) appear considerably less attractive. All other categories are selected only in single cases (between 1 and 6 cases). For the contrast of long constricted /ø:/ - /y:/, an asymmetric response behaviour is observed: A categorization of /ø:/ > /y:/ occurs in 41% of the cases whereas the inverse substitution /y:/ > /ø:/ is observed in only 6% of the cases.

<ei> is the most preferred diphthongal category, selected in 168 cases for different vowel qualities, especially for /ɛ:/ and /e:/. The frequent substitution of e-vowels with <ei> is interesting, since a diphthongization of long e-vowels as diphthongized [e^{i}] is also observed in production with Polish learners of German, e.g. [le^{i}bən] *leben* 'live'. Previous studies on the pronunciation of German by Polish learners could not clarify whether this diphthongization is motivated articulatorily or perceptually. Hentschel (1986) observed a general tendency to diphthongize German long vowels in perception (for discussion, see Hentschel 1986: 198ff), but his results also show that /ʊ ɪ ʏ/ were substituted by diphthongs in a considerable number of cases. In the data presented here, only /ʏ/ and /ɪ/ but not /ɔ/ are diphthongized in a few cases. The comparison of beginners (Table 10.58) and more experienced language learners (Table 10.59) shows that substitutions with diphthongs are observed only in very few cases.

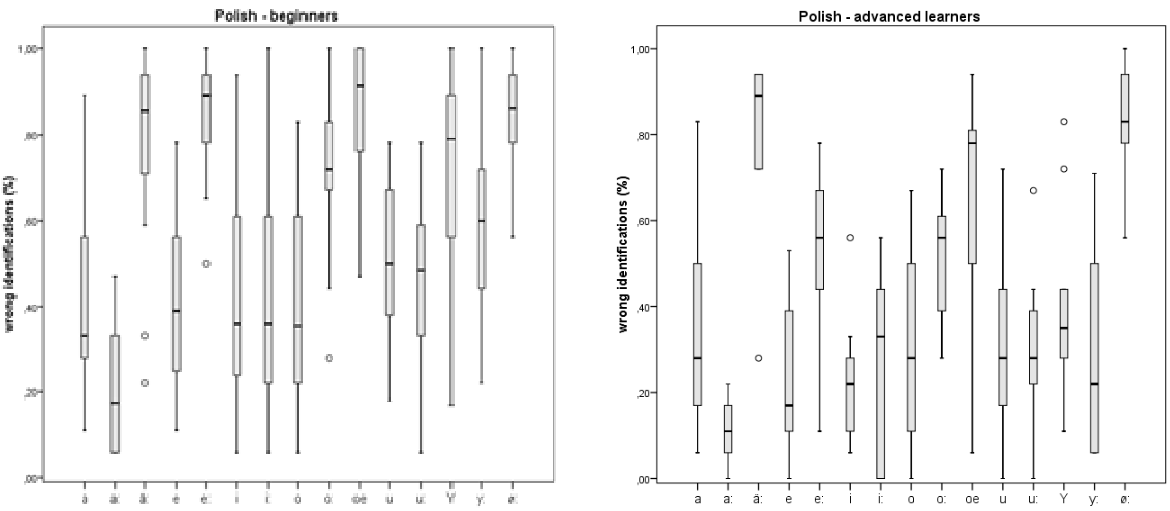

Figure 10.64: Comparison of id_wrong scores for Polish beginners and advanced learners of German

A comparison of category-specific id_wrong scores for beginners and advanced learners (see Figure 10.64) shows that even with advanced learners difficulties with /ɛ:/ and /ø:/ persist in more advanced stages of L2 German acquisition, whereas for /œ/ considerable variation is observed and substantial improvements are observed for /ʏ/ and /y:/.

%	ɑ	ɑ:	ɛ:	ɛ	e:	ɪ	i:	ɔ	o:	ʊ	u:	œ	ø:	ʏ	y:	‹ei›	‹eu›	‹au›	wrong	correct
a	**60**	38	1		1		1												40	**60**
a:	17	**80**		1	1											1			20	**80**
ɛ:			**20**	12	39	1	2					2	1		1	20	1		80	**20**
ɛ			7	**59**	28	1	1				1		1			2			41	**59**
e:			7	2	**15**	14	38	1					1	2	5	16	1		85	**15**
ɪ			2	3	4	**57**	19			1		1	2	7	3	1	1		43	**57**
i:					3	33	**61**						1	1	2				39	**61**
ɔ							1	**60**	32			1	2	1		1	2		40	**60**
o:							1	7	**29**	14	28	3	8	2	5		2	1	71	**29**
ʊ							1	10	5	**49**	24	3	2	4	1				51	**49**
u:									5	23	**56**	1	1	5	8		1		44	**56**
œ		1	10	27	16		2	2	2	1	1	**14**	8	8	4	1	4		86	**14**
ø:					2	1	2		1	5	18	7	**16**	13	35				84	**16**
ʏ					3	3	3		1	13	8	11	7	**30**	19	1	1	1	70	**30**
y:			1	1	2	3	10		1	5	8	4	7	18	**40**		1		60	**40**
Σ																			*57*	***43***

Table 10.58: Confusion matrix for L1 Polish beginners (N = 22, 5937 responses, values < .5 not indicated, values < 5.5 in grey)

%	ɑ	ɑ:	ɛ:	ɛ	e:	ɪ	i:	ɔ	o:	ʊ	u:	œ	ø:	ʏ	y:	‹ei›	‹eu›	‹au›	wrong	correct
a	**67**	33				1													33	**67**
a:	10	**89**				1													11	**89**
ɛ:			**20**	12	65	1	1								1	1			80	**20**
ɛ			9	**77**	13	1												1	23	**77**
e:			6	5	**47**	4	29					1	1	3	2	1			53	**47**
ɪ			1	4	2	**77**	10					1		5	1				23	**77**
i:					1	21	**76**					1		1	1				24	**76**
ɔ						1	1	**69**	30										31	**69**
o:						1		7	**49**	7	25		7	1	2				51	**49**
ʊ						1		2	1	**69**	15	3	2	5	2				31	**69**
u:						1				15	**72**	1	2	4	6				28	**72**
œ			4	3	3	1	1			3	1	**39**	9	25	11				61	**39**
ø:						1				1	2	6	**17**	18	56		1		83	**17**
ʏ						3	1			5	1	9	4	**60**	17				40	**60**
y:							3					2	4	22	**70**				30	**70**
Σ																			*40*	***60***

Table 10.59: Confusion matrix for L1 Polish advanced learners (N = 9, 2430 responses, values < .5 not indicated, values < 5.5 in grey)

To summarize the confusion data for Polish learners of German, least problems are observed with a-vowels, whereas considerable problems are observed for the class of front rounded vowels both with the differentiation within this class and the differentiation from other categories. While the perception of ü-vowels seems to improve with more experienced learners, ö-vowels cause persistent problems even to advanced learners. Moreover, the differentiation of /ɛ:/, /e:/ and /i:/ causes considerable difficulties. Substantial difficulties with e-qualities are observed with more advanced learners. Long constricted /o:/ and its differentiation from other rounded vowels is another difficult feature of the German vowel system for Polish listeners.

Back rounded categories are attractive response options for /ø:/, /ʏ/ and also /y:/, but not for /œ/-stimuli, which is more frequently substituted with front non-rounded e-qualities, especially by less experienced learners, most probably due to the low degree of lip rounding for /œ/. This perceptual unrounding was observed in previous studies (see section 10.9.3). Front rounded categories are attractive response options for /y:/ and /ʏ/, but are considerably less frequently selected for /ø:/-stimuli. Diphthongs, especially <ei>, are quite frequently selected by beginners but not by advanced learners.

10.9.4.4 Similarity and distance

Similarity scores for German vowels as perceived by L1 Polish listeners are summarized in Table 10.60. Compared to other languages, the sim_scores for Polish subjects are relatively high, a fact that is due to the high number of beginners in the sample as well as due to fundamental differences between the German and the Polish vowel system. Asymmetric effects are of course not evident in sim_scores.

By far the highest sim_score is given for the contrast /e:/ - /ɛ:/ (1.186), resulting from a very high percentage of /e:/-responses for /ɛ:/-stimuli (46%). Very high sim_scores are also obtained for those contrast pairs that show strong asymmetry patterns and for contrast pairs that are only based on a difference in constriction degree and phonemic length such as /ø:/ - /y:/ (0.729), /ʏ/ - /y:/ (0.434), /i:/ - /e:/ (0.420), /œ/ - /ø:/ (0.402), and /ɔ/ - /o:/ (0.397). Slightly lower sim_scores are obtained for contrasts within the class of [+round] vowels (/ʏ/ - /œ/ (0.381), /ʏ/ - /ø:/ (0.373), /u:/ - /o:/ (0.316)), and for other pairs that are differentiated with respect to constriction degree (and phonemic length): /ʊ/ - /u:/ (0.362), /ɪ/ - /i:/ (0.359), /ɑ/ - /ɑ:/ (0.355), /ɛ/ - /ɛ:/ (0.235). Sim_scores below 0.25 are found for contrast pairs that are differentiated by other features.

For *front vowel*-contrasts, three major types of difficulties are observed that are reflected in the sim_scores: (1) contrasts within the class of *front non-rounded* vowels (/e:/ - /ɛ:/ (1.186), /i:/ - /e:/ (0.420), /i:/ - /ɪ/ (0.359), /e:/ - /ɛ/ (0.297), /ɛ/ - /ɛ:/ (0.235)./ɪ/ - /e:/ (0.166) and others with lower scores), (2) contrasts within the class of *front rounded* vowels (/ø:/ - /y:/ (0.729), /ʏ/ - /y:/ (0.434), /œ/ - /ø:/ (0.402). /ʏ/ - /œ/ (0.381), /ʏ/ - /ø:/ (0.373), /y:/ - /œ/ (0.131) and others with lower scores), and (3) less frequently with sim_scores below 0.3, contrasts of *front rounded vs. front non-rounded* vowels: /œ/ - /e:/ (0.282), /œ/ - /ɛ/ (0.240), /œ/ - /ɛ:/ (0.222, orthographic influence), /ʏ/ - /ɪ/ (0.094), /y:/ - /i:/ (0.082), /y:/ - /e:/ (0.075), /ʏ/ - /e:/ (0.066), /ø:/ - /e:/ (0.054), /y:/ - /ɪ/ (0.042), /ʏ/ - /i:/ (0.030), /ø:/ - /ɪ/ (0.021), /ø:/ - /ɛ:/ (0.020) and others with lowe scores.

V	ɑ	ɑ:	ɛ:	ɛ	e:	ɪ	i:	ɔ	o:	ʊ	u:	œ	ø:	ʏ	y:
a	1.000														
ɑ:	.355	1.000													
ɛ:	.007	.004	1.000												
ɛ	.000	.004	.235	1.000											
e:	.004	.003	1.186	.297	1.000										
ɪ	.001	.001	.031	.031	.166	1.000									
i:	.003	.001	.019	.007	.420	.359	1.000								
ɔ	.000	.001	.000	.003	.004	.003	.006	1.000							
o:	.000	.000	.007	.000	.000	.004	.006	.397	1.000						
ʊ	.000	.001	.002	.000	.002	.006	.003	.071	.175	1.000					
u:	.000	.000	.000	.003	.000	.003	.001	.001	.316	.362	1.000				
œ	.000	.007	.222	.240	.282	.019	.019	.028	.066	.057	.018	1.000			
ø:	.000	.000	.020	.007	.054	.021	.018	.018	.173	.081	.191	.402	1.000		
ʏ	.002	.003	.003	.000	.066	.094	.030	.007	.030	.161	.108	.381	.373	1.000	
y:	.000	.000	.013	.003	.075	.042	.082	.002	.059	.050	.117	.131	.729	.434	1.000

Table 10.60: Similarity matrix for L1 Polish listeners

Within the class of *back vowels*, the highest sim_scores are obtained for the pairs /ɔ/ - /o:/ (0.397), /ʊ/ - /u:/ (0.362), and /ɑ/ - /ɑ:/ (0.355), which differ only in phonemic length (and constriction degree). Moreover, the contrast [velar] vs. [uvular] seems to be difficult, especially for /u:/ - /o:/ (0.316), but also for /ʊ/ - /o:/ (0.175) and /ʊ/ - /ɔ/ (0.071).

Contrasts of *front vs. back* (velar and uvular) vowels show sim_scores below 0.2. For front vs. back contrasts, a higher degree of similarity is observed for vowel pairs that share the feature [+round]: /u:/ - /ø:/ (0.191), /o:/ - /ø:/ (0.173), /ʊ/ - /ʏ/ (0.161), /u:/ - /y:/ (0.117), /u:/ - /y:/ (0.108), /u:/ - /ʏ/ (0.108). Lower scores are obtained for rounded vs. non-rounded (≤ 0.007) or non-rounded vs. non-rounded (≤ 0.007) contrasts. This pattern indicates that [round] is the more salient feature in these contrast pairs, but that the exact localization of the constriction location causes problems.

Within the class of *rounded vowels*, the highest sim_score is observed for front /ø:/ - /y:/ (0.729), followed by /ʏ/ - /y:/ (0.434) and /œ/ - /ø:/ (0.402) reflecting major difficulties with front rounded vowels. High sim_scores are also given for /ɔ/ - /o:/ (0.397), /ʏ/ - /œ/ (0.381), /ʏ/ - /ø:/ (0.373), /ʊ/ - /u:/ (0.362) and /u:/ - /o:/ (0.316), followed by several other contrasts that differ in front vs. back constriction location with sim_scores ≤ 0.191 (except /ʊ/ - /o:/ and /y:/ - /œ/).

10.9.4.5 The perceptual vowel map for L1 Polish listeners

From the similarity scores calculated by Shepard's formula, distance scores and non-metric MDS solutions were derived. For the non-metric MDS solutions, RSQ values of 0.689 in dimension 1, 0.844 for the two-dimensional solution, and 0.922 for the three-dimensional solution are obtained.

The one-dimensional MDS solution (Figure 10.65) has a relatively high RSQ value of 0.689 compared to other languages, but the representation clearly differs in that the categories are fairly wide-spread and do not cluster as in other language sub-groups. Viewing the one-dimensional representation, German u- and a- vowels can be considered as most different from each other, e-vowels are differentiated from other categories quite well, and rounded vowels do not show a very tight clustering.

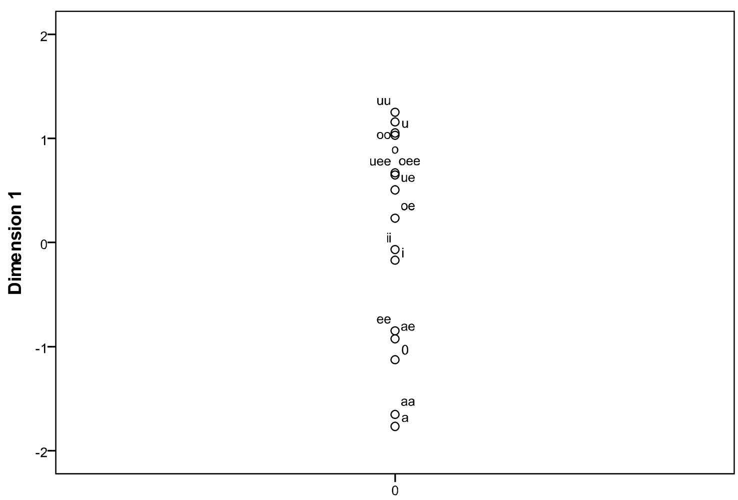

Figure 10.65: One-dimensional MDS representation of the perceptual map of German for L1 Polish listeners (N= 31, 8260 responses, RSQ .689)

The two-dimensional diagram MDS solution of the perceptual vowel map representing the relative similarity of German vowels as perceived by Polish listeners in Figure 10.66 that a-vowels are most distinctive from all other German categories. It shows moreover three major areas of difficulties in the differentiation of German vowel categories: (1) the differentiation *of front non-rounded i- and e*-vowels, (2) difficulties *with front rounded ü- and ö*-vowels, and (3) difficulties to differentiate *back rounded u- and o*-vowels.

Within the class of back vowels, /u:/ and /ɔ/ are differentiated best from each other, whereas a high degree of perceptual similarity is given for /ʊ/ and /o:/, which almost overlap in the MDS solution.

A high degree of perceptual overlap is also observed for German i-vowels (sim_score 0.359) that are positioned very close to e-qualities, especially to /e:/ due to a high percentage of inter-category confusion of e- and i-qualities.

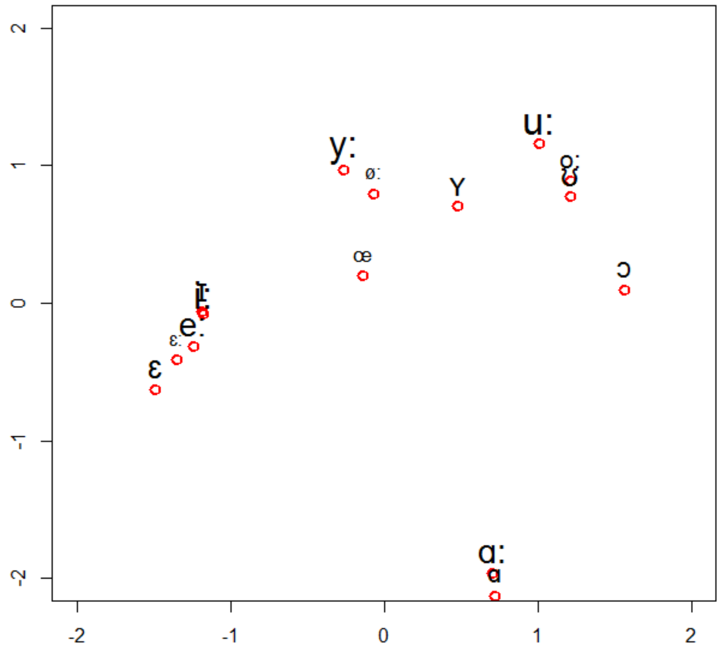

Figure 10.66: Two-dimensional MDS representation of the perceptual map of German for L1 Polish listeners (N= 31, 8260 responses, RSQ .844)

The intermediate status of front rounded ü- and ö-vowels is indicated by their position between the cluster of front non-rounded and back rounded vowels. The relative position of each of the front rounded vowels depends on the number of substitutions with either front or back categories.

/ʏ/ holds an intermediate position between front rounded /y:/, ö-vowels (/ø:/ and /œ/), and back rounded u- and o-vowels. The position of /œ/ is slightly closer to e- and i-qualities due to several substitutions of /œ/ with front categories. /y:/ and /ø:/ are located close to each other due to a high percentage of confusions for these two categories. However, as differences in type size indicate, /y:/ is the more preferred response category attracting responses for almost all other rounded categories.

Differences in type size show that /y:/, /u:/, /i:/ and /ɑ:/ are preferred response options that attract responses from similar sound categories. These asymmetries in the listeners' response behaviour are not represented in the MDS solution and the sim_scores, but become evident in the interpretation of the confusion matrix presented above.

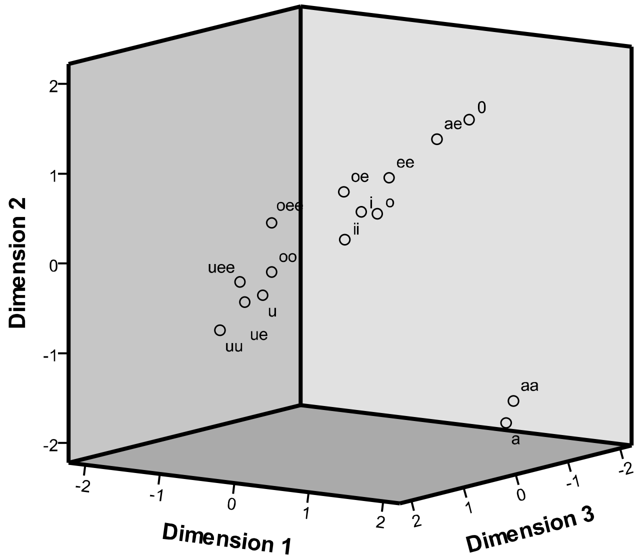

Figure 10.67: Three-dimensional MDS representation of the perceptual map of German for L1 Polish listeners (N= 31, 8260 responses, RSQ .922)

The three-dimensional solution shows again very clearly that a-vowels are very distinct from all other categories. Like in the two-dimensional solution, a kind of triangle of a-vowels, /u:/ and /ɛ/ is evident in the diagram. The position of /œ/ and /ɔ/, which is very close to i- and e-vowels, may be surprising at first sight but could be explained by the high percentage of substitutions of /œ/ to e-vowels, specifically short /ɛ/ (27%). For the position of /ɔ/, however, there is no clear-cut explanation found. A pattern that is found to be fundamentally different from the two-dimensional solution is the very differentiated location for front rounded vowels. While /y:/ and /ʏ/ hold a position that is quite close to u-vowels, /ø:/ is a little more distant from the u- and ü-vowels and /œ/ holds a remote position in the representation. A look at the confusion matrix (Table 10.57) shows that this reflects the substitution patterns observed especially with beginners who show a strong preference for front non-rounded response categories.

10.9.5 Summary

The data presented above demonstrate that Polish learners show the ability to localize basic differences in constriction location, which is indicated by sim_scores below 0.2 for contrast pairs that differ in constriction location (palatal, velar, uvular, pharyngeal). However, significant perceptual difficulties are observed for some vowel classes, especially for the class of front rounded vowels but also for contrasts of back vowels.

As previous studies on the pronunciation of German vowels by Polish learners have shown the realization of German ü- and ö-vowels is difficult and these sounds are mostly pronounced as front non-rounded vowels. However, the data presented in this study shows that the perceptual differentiation of front rounded vowels from front non-rounded as well as from back vowels causes considerable difficulties to Polish listeners and that German ü- and ö-vowels are not only substituted with front non-rounded vowels but can also identified with back vowels in perception, a pattern that does however not hold for /œ/ to the same extent. Confusion of front rounded vowels with (1) other front rounded vowels, (2) back rounded vowels, and (3) front non-rounded vowels is observed. However, substitution patterns differ considerably between categories. German ü-vowels (/y:/ 49% correct, /ʏ/ 39% correct) clearly show a higher percentage of correct responses than ö-vowels (/œ/ 21% correct, /ø:/ 16% correct). ö-vowels continue to cause difficulties also for advanced learners of German.

While a substitution with front non-rounded categories is frequently observed for /œ/, for which front non-rounded categories are selected in 42% of the cases, /ʏ/ (7% front non-rounded categories), and /y:/ (11% front non-rounded), this pattern is not dominant for long constricted /ø:/ (only 3% front non-rounded). While lip rounding in /ø:/-stimuli is clearly recognized, the identification of the place of constriction for /ø:/-stimuli causes considerable problems to Polish listeners.

Another major area of difficulties is the differentiation of pre- vs. mid-palatal vowel qualities. These difficulties are observed with roundedas well as with non-rounded vowels.

The most attractive response categories are /y:/, /u:/ and /i:/: /y:/ is the most favoured response option, especially for /y:/-, /ø:/- and /ʏ/-stimuli and less frequently for /u:/, /œ/, and /o:/, but also for non-rounded /e:/, /ɪ/, /i:/, a pattern that could be interpreted as hypercorrect reaction to difficulties with the differentiation of front roundedand front non-rounded vowels as described above.

/u:/ attracts a larger portion of responses for all rounded categories, but not for /ɔ/ and /œ/, probably due to their low degree of lip rounding. /i:/ mainly serves as perceptual substitute for /ɪ/ and /e:/-stimuli and less frequently also for front rounded/y:/ and /ʏ/, most probably due to a hypercorrect reaction to difficulties with the differentiation of front roundedand front non-rounded vowels.

Asymmetric patterns in the response behaviour of Polish subjects are particularly observed for the contrast pairs /e:/ - /i:/ (35% /e:/ > /i:/ vs. 2% /i:/ > /e:/), /o:/ - /u:/ (27% /o:/ > /u:/ vs. 3% /u:/ > /o:/) and /ø:/ - /y:/ (41% /ø:/ > /y:/ vs. 6% /y:/ > /ø:/). This pattern, which is not observed for the short unconstricted qualities, seems to indicate that the acoustic consequences of a higher degree of constriction in mid-high qualities is (mis)interpreted as a higher degree of [height] or [frontness].

Issues for further research would be the relative role of qualitative perceptual cues for high vowels, especially for front rounded vowels and their interference with Polish /ɨ/ and the role of duration – which is not phonemically functional in Polish – in vowel identification.

10.10Romanian

10.10.1 General characteristics

Romanian is a Romance language spoken by around 24 to 28 million people primarily in Romania and Moldova, by minor groups in neighbouring countries such as the Vojvodina region in Serbia and by migrants in several European countries, Canada and the United States. Balkan Romance is classified into four subgroups of which Daco-Romanian, the national language of Romania, is the main one. The wider term Balkan Romance comprises three other subgroups: (1) Arumanian, spoken in northern Greece, Albania, Serbia and Macedonia, (2) Megleno-Romanian, spoken in a small area to the north of Salonika, and (3) Istro-Romanian in the Istrian peninsula of Croatia (Mallinson 2011: 253). The following description will only refer to Daco-Romanian.

The exact taxonomy of Daco-Romanian sub-dialects is a matter of debate in the literature (e.g. Mallinson 1986; Iliescu 2003), though there is general agreement on the division of a Muntenian, or Wallachian, and a Moldavian sub-dialect. Other sub-dialects are the Banat sub-dialect that is spoken in Western Romania and parts of Serbia, and the Transylvanian dialects, which can be further divided into the variety of Crişana and the variety of Maramureş (for another kind of classification, see e.g. Iliescu 2003: 533). The differentiation of Daco-Romanian sub-dialects mainly refers to lexical and phonetic features such as the palatalization of affricates or vowel changes, but mutual understandability between speakers of different sub-dialects is basically given.

For political and geographical reasons, the Muntenian sub-dialect of Daco-Romanian became the standard language, which came into contact with French and Italian in the nineteenth century. Induced by this contact, several Western Romance innovations entered into Daco-Romanian as a whole, inducing a tendency to "reromance" in the last 150 years (Mallinson 2011: 265).

Romanian differs in several aspects from the Western Romance languages, many of which are common features of the Balkan Sprachbund, e.g. the use of suffixes to mark definiteness, a feature that Romanian shares with Albanian and Bulgarian.

With respect to morphology, the Latin declension system is better preserved in Balkan Romance than in the Western Romance languages, a fact that is partly due to the use of suffixed determiners (Mallinson 2011). Throughout the centuries, the mainly Latin-based lexical stock of Romanian vocabulary was enriched by borrowings from other Balkanic languages, mainly Slavonic languages, and was expanded by contact with French and Italian in the nineteenth century. Borrowed vocabulary from French, Italian, Greek, Turkish or German has been partly assimilated, whereas international vocabulary that entered the language, mainly from English in more recent times remains largely unassimilated. Some phonological processes are sensitive to this stratification of the lexicon (Chitoran 2002a).

Romanian is written in a Latin-based alphabet with three diacritics. It has been developed in the nineteenth century. Relatively little divergence of the spoken and written language is observed.

10.10.2 Phonological and phonetic description

Standard Romanian has an inventory of 32 phonemes.

The consonant inventory of Romanian consists of 20 consonants (see Table 10.61). On the phonetic surface, palatalization of consonants occurs if a morphological marker containing a front vowel is attached: -C#+ /i/ => [C^j]. Though Petrovic (1956) claims the existence of a very complex consonant inventory with four distinct consonant classes (neutral, palatalized, labialized, labio-palatalized[122]), this view is generally rejected in other descriptions (e.g. Chitoran 2002a; Iliescu 2003; Mallinson 2011). In most descriptions, the series of palatalized consonants is treated as a set of non-palatal consonants and a recurring palatal off-glide.

	labial	labio-dental	dental	alveolar	post-alveolar	palatal	velar	glottal
Plosives	p b		t d				k g	
Affricates			ts		tʃ dʒ			
Nasals	m		n					
Fricatives		f v	s z		ʃ ʒ			h
Liquids				l r[123]				
Semi-vowels	w					j		

Table 10.61: The Romanian consonant system

A characteristic that Romanian shares with other Romance languages is the fronting of velar plosives before front vowels, giving rise to post-alveolar affricates. Moreover, /k/ and /g/ appear as slightly fronted allophone when preceding a front vowel, e.g. *chema* ‘call’, *ghetou* ‘ghetto’. These two different strategies in treating velar plosives reflect the distinction of

122 By positing a palatalized and a labialized series, Petrovic (1956) is able to avoid the diphthongs /ea/ and /oa/ (for discussion, see Chitoran 2002: 11).

123 /r/ is generally realized as a flap [ɾ] and occasionally as a trill in word-inital position (Chitoran 2002: 10).

words inherited directly from Latin and those borrowed from other languages in more recent times (Mallinson 2011).

Commonly, the Romanian vowel system is described as a seven-vowel system (see Table 10.62). Three so-called “central” vowels are present in the vowel inventory: high /ɨ/, mid /ə/, and low /a/.

In addition to these seven vowel system, Romanian has three diphthongs /ea/, /oa/ and /eo/ and two glides /j/ and /w/. Though, at first sight, the description of the vowel inventory seems quite simple, several questions related to the status of the so-called “central” qualities, to diphthongs and the status of glides in the phonemic system are under discussion in the literature (for a detailed discussion, see Chitoran 2002a).

	front	central	back
close	i	ɨ <î> or <â>	u
mid	e	ə <ă>	o
open		a	

Table 10.62: The vowel system of Romanian in terms of traditional phonological features (Iliescu 2003: 533)

Vowel length is phonemically not distinct in Romanian, though length differences may occur in vowels as one feature of emphasis together with stress and intonation (Mallinson 1986: 342). Gregor-Chirită (1991: 16f) describes the duration of vowels in open stressed syllables as usually longer than in open non-stressed syllables, e.g. *'linje* ‘line’ vs. *li'man* ‘waterside’, '*vază* ‘vase’ vs. *va'por* ‘steamer’. In closed syllables, vowels are usually shorter in duration in stressed as well as unstressed syllables.

In word-initial position, basically all vowels may occur. By the influence of Slavonic languages, in a limited number of words, especially in forms of the copula, initial /e/ is affected by ioticization, e.g. *este* /jeste/ ‘is’, while in more recent loans from Western Romance languages, /e/ is not affected: *era* /'era/ ‘the era’ vs. *era* /je'ra/ ‘was’.

Initial /i/ before other vowels is also pronounced as /j/: *iute* /jute/ ‘quick’, *iar* /jar/ ‘again’. Final <-i> represents the palatalization of a preceding (non-liquid) consonant, e.g. *lup* [lup] ‘wolf’ vs. *lupi* [lup^j] ‘wolves’, but *tigri* [tigri] ‘tigers’. Final <-ii> represents the full vowel /i/.

It is important to note that /ɨ/ and /ə/ are not reduced vowels but surface in stressed syllables and are involved in alternations with /e/ and /o/ (Chitoran 2002a: 8).

Numerous examples are found where /ə/ (<ă>) is the result of a weakening process of unstressed Latin /a/, but in contemporary Romanian, /ə/ features as a stressed vowel, e.g. in *dă* ‘give’ or *cîntă* ‘he sings’. Other Latin vowels also developed into /ə/, e.g. *săptămînă* ‘week’.

/ɨ/ is presumed to be a borrowing from Slavonic that entered the language via words like *rîs* 'lynx' from Slavic *rys*, or *smîntînă* 'sour cream', but its functional load increased as Latin words were also affected (Mallinson 1990; for a detailed discussion on the historical roots of /ɨ/ in Romanian, see Renwick 2012: 24ff). In orthography, <î> and <â> are used as allographs for /ɨ/. These two allographs were originally used following etymological principles, whereas in contemporary orthography, <î> is used word-initially and –finally, and <â> word-medially (Mallinson 2011: 257).

The phonemic status of /ɨ/ and /ə/ is confirmed by a few widely cited minimal pairs such as *vîr* 'I thrust' vs. *văr* 'cousin' but is not found in any other Romance language. The distribution of both sounds is however largely complementary in that it is determined by preferences for specific positions and environments: /ɨ/ occurs mainly in stressed syllables and especially before nasals and /r/ as a result of historical processes, while /ə/ is typically unstressed. This distribution and the restricted functional load gives reason to assume a relationship of marginal contrastiveness (Renwick 2011; 2012: 80, 84ff).

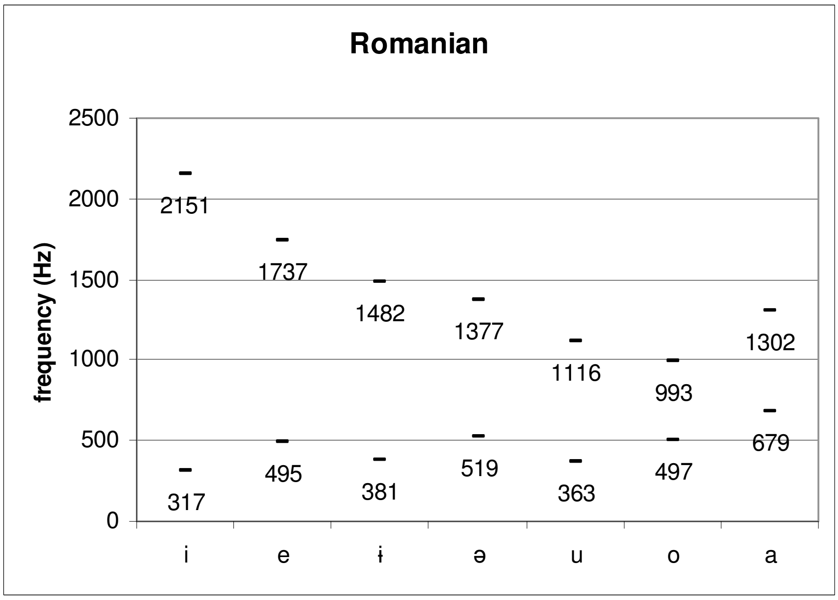

Figure 10.68: Vowel formant frequencies for the first two formants of Romanian vowels. Mean F1 and F2 frequency values of 3 male speakers for vowels in stressed syllables (Renwick 2012: 158)

Acoustic data for Romanian vowels are provided by Renwick (2012) and by Teodorescu, Pistol & Feraru (2010). Renwick (2012) analyzed vowel realizations from 3 male and 14 female native speakers of Romanian living in Rumania and the USA reading 60 lexical items embedded in a carrier sentence. Each of the seven Romanian vowels was tested and analyzed in at least four words in stressed and unstressed syllable position. Formant frequencies for the first two formants were measured at the midpoint of vowel tokens. Figure 10.68 represents the mean F1 and F2 values for three male speakers collected by Renwick (2012), no data for F3 are available in this study[124]. Instead of the standard transcription for <ă> as [ə], Renwick (2012: 18) proposes a transcription as [ʌ], which should reflect the full phonemic status of this vowel. Precendence for this transcription comes from Steriade (2008: 315).

Teodorescu, Pistol & Feraru (2010) present acoustic data (F0, F1, F2, F3) for Romanian vowels realized by male and female speakers. These data seem however not reliable enough for two reasons: (1) Some vowels were realized in a sentence context spoken under different emotional conditions while other vowel qualities were only realized in isolation. (2) The same tokens were analyzed with Praat, WASP, Klatt and Cold Wave. However, the results obtained by the different analytical programs differ very much from each other.

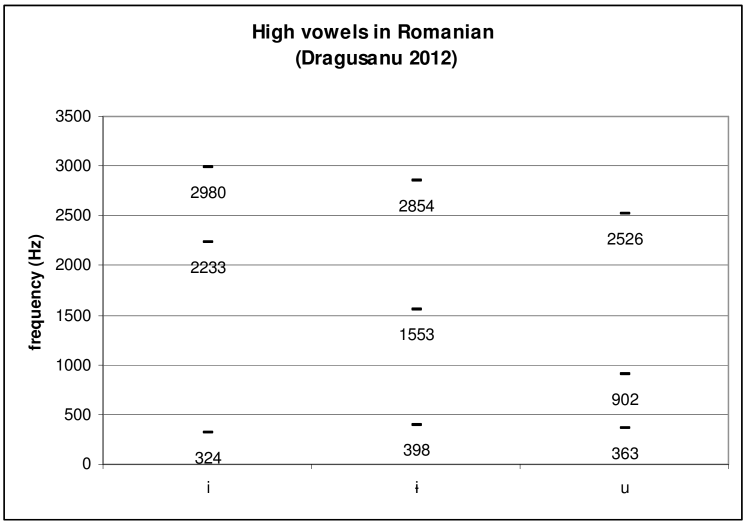

Figure 10.69: Formant frequency values (F1, F2, F3) for Romanian high vowels (Dragusanu 2012)

Vowel realizations by Romanian heritage speakers were studied by Dragusanu (2012). A word list of 99 words was read by the participants of the study. Figure 10.69 presents the mean formant frequency values (F1, F2, F3) for high vowels[125] for a 50 year old male Romanian speaker who spent the last 10 years in Canada. The high F3 values for /i/ and the presence of /ɨ/ in the Romanian vowel system suggest that Romanian uses the pre-palatal constriction location for /i/. The data in Figure 10.68 and Figure 10.69 suggest that /ɨ/ can be considered most likely as a palatal vowel [ɨ] (as usually assumed) or even as a palato-velar

[124] Renwick's F2 values appear very high, especially for female speakers (cf. Renwick 2012: 146), indicating a potential confusion of F2 and F3 frequencies in the analysis procedure. Values for F3 are not presented in Renwick's study.

[125] For other vowels, no measured values are presented in the paper.

vowel [ɯ], for which values for F1 around 350 Hz and for F2 around 1500 Hz together with a much higher F3 are characteristic (cf. Wood 1990: 200f). The assumption of /ɨ/ being realized as a palato-velar vowel [ɯ] rather than [ɨ] has however to be verified on a larger database with sufficient data for the higher formants F3 and F4 for all palatal vowels in the system. An investigation of the formant frequencies for <ă> would provide indirect evidence for the articulatory status of this sound to decide whether it is better transcribed as [ʌ] as was proposed by Steriade 2008 and Renwick 2012) than as [ə].

In addition to the seven vowel phonemes, front rounded vowels /y/ and /œ/ or /ø/ had been introduced into Romanian via a number of borrowings from other languages like French, German or Turkish. As Romanian has no front rounded vowels, these qualities had been replaced with /ju/, /u/, /i/, /iu/ or /o/, /jo/, /e/, respectively, and adjusted to the Romanian orthography (Chitoran 2002a: 29f; see Close 1974, cited in Chitoran 2002a)), but they are preserved in some of the more recent borrowings (see examples (d) and (g)).

Three forms of treating foreign front rounded vowels can be observed in Romanian: Usually, front rounded vowels lose one of their feature specifications, either rounding or frontness. A third possibility is to split the French vowel into a glide-vowel sequence [ju] or [jo] (see (a) and (e)) which may be interpreted as a perceptually-based strategy (Chitoran 2002a: 29f.).

(a) French *purée* – Romanian *pjuré* /pju're/ 'puree'
Turkish *kül* – Romanian *ghiul* /gjul/ 'large ring'
(b) French *bureau* – Romanian *birou* /bi'row/ 'desk, office'
(c) German *Düse* – Romanian *duză* /'duzə/ 'nozzle'
French *bulletin* – Romanian *buletin* /bule'tin/ 'bulletin'
(d) French *tul* – Romanian *tul* /tyl/ 'tulle'
French *ecru* – Romanian *ecru* /e'kry/ 'ecru'
(e) French *liqueur* – Romanian *likjor* /lik'jor/ 'liquor'
(f) French *portefeuille* – Romanian *portofel* 'wallet'
(g) French *bleu* – Romanian *bleu* /blø/ 'light blue'
French *pasteurizer* – Romanian *pasteuriza* /pastøri'za/ 'to pasteurize'

Further characteristics of the Romanian sound inventory are the two glides /j/ and /w/ and the two diphthongs /ea/ and /oa/ with mid glides as their non-syllabic element (for a detailed discussion see Chitoran 2002a: 8ff; 2002b; 2002c). Chitoran (2002c) presents a comparative acoustic and perceptual analysis of the two diphthongs [ea] and [oa] and the glide-vowel sequences [ja] and [wa]. Although these sequences seem to be auditorily very similar, they differ in their phonological patterning.

The glides /j/ and /w/ occur in a large number of /jV/, /Vj/, /wV/ and /Vw/ combinations which are often referred to as diphthongs (e.g. Iliescu 2003), though in a strict sense they have to be described as glide-vowel or vowel-glide sequences where the glide fills the syllable onset or coda (Chitoran 2001: 221ff). However, many descriptions of Romanian phonology refer to glide-vowel sequences as rising and to vowel-glide sequences as falling diphthongs (e.g. Gregor-Chiriță 1991; Iliescu 2003).

Triphthongs may result from a sequence of a diphthong /ea/ and /oa/ and a glide /j/ or /w/ before or after the diphthong or from a glide-vowel-glide sequence (e.g. /jaj/, /waw/, /jow/). There is no consensus in the literature concerning the exact number of these “triphthongs” (Iliescu 2003). The large number of Romanian diphthongs and triphthongs has led some researchers to describe Romanian as a “vocalic” language. The set of diphthongs has been recently enlarged by borrowings from English resulting in diphthongs like /jə/, /we/, /wi/, /wo/.

Common vowel alternations are /ea/ ~ /e/, /oa/ ~ /o/, and /a/ ~ /ə/ (for discussion, see Renwick 2012: 184f; see also Chitoran 2002b, describing diphthongization as vowel lowering under stress). Some alternations are due to so-called palatalization under the influence of front vowels (Iliescu 2003: 535): a ~ ə /_Ce, e.g. *masa* ‘table’ – *mese* ‘tables’, *fată* ‘girl’ – *fete* ‘girls’, where in noun and adjective singular/plural alternations the stem vowel assimilated historically to the stem vowel. Even though this is not a synchronic rule, Mallinson (1986: 342) observes that new loans may be accommodated within such a system, but would be a result of (conscious or unconscious) analogy rather than assimilation.

Romanian syllable structure is relatively complex allowing two- and three-consonantal clusters in the syllable-onset as well as in the -offset (see Chitoran 2002a: 12ff; for syllabification of word-medial clusters, see Mallinson 1986: 340f).

Word stress is largely unpredictable, but particular stress patterns are associated with certain word classes and word class members (Mallinson 1986: 342). Romanian has a stress accent, which is based on right edge prominence for primary stress and on feet for secondary stress (Chitoran 2002a: 51ff). Some homophonic words are distinguished by means of word stress, e.g. *módele* ‘the fashions’ vs. *modéle* ‘model’, *cằntă* ‘sings, sing’ vs. *căntắ* ‘sang’ (Mallinson 2011: 259). Stress is however not regularly represented in Romanian orthography. The stressed syllable is usually more intense, longer and higher pitched, but intensity is the main feature. Stressed and unstressed syllables are differentiated by intensity, length and pitch, though to a less extent than in German (Iliescu 2003: 537).

10.10.3 Contrastive Analysis

A contrastive analysis of Romanian and German has to deal primarily with the considerable differences in the number of phonemes in Romanian and German. Romanian learners of German have to deal with a large number of new phonemes, especially with the front rounded vowel qualities and the qualitative and quantitative distinctions of long and short German vowels. These differences cause problems in production and perception for Romanian native speakers. The orthographic similarity may be of help in some cases but may cause additional problems in others (Gregor-Chiriță 1991: 11).

A comparison of the two systems shows systemic differences between the two languages and possible areas of difficulty in the acquisition of L2 German:

(1) For Romanian /i u e o a/ two equivalents are given in German, for Romanian /e/ even a third phoneme /ɛ:/ has to be considered as a candidate for a possible equivalence classification.

(2) /ɨ/ and /ə/ have phonemic status, occurring in stressed syllables in Romanian, but have no corresponding phoneme in German. /ɨ/ does not occur at all in German. Romanian /ə/ occurs in unstressed and stressed syllables and therefore cannot be described as equivalent to German /ə/ which occurs only in unstressed syllables. Both vowels may be articulated with rounded lips, being perceptually close to German /ʏ/ and /œ/ respectively, at least for German listeners without learning experience in Romanian.

(3) Front rounded /y ʏ ø œ/ have no corresponding phonemes in Romanian, though they are familiar to Romanian speakers due to their occurrence in borrowings from other languages.

(4) Romanian /a/ is reported as not directly corresponding neither to German /a/ nor /ɑ/. Gregor-Chiriță (1991: 12) describes Romanian /a/ as rather "central".

(5) The correct perception and production of phonemic length, which co-varies with constriction degree in German, is expected to cause problems for Romanian learners.

(6) German and Romanian differ in number and structure of their diphthongs. Triphthongs do not exist in German.

The large inventory of diphthongs and triphthongs shows that the process of diphthongization is not restricted in Romanian. This may lead Romanian learners of German to realize long constricted /e:/ or /o:/ in open syllables as diphthongs (Gregor-Chiriță 1991: 20), e.g. in German 'froh' /fro:/ and 'Weh' /ve:/ are often realized as [frow] and [ve:i].

To realize German front rounded vowels, two strategies are reported for Romanian learners (Gregor-Chiriță 1991: 40): (1) either a substitution with a back vowel of the same height (e.g. Germ. böse /bøzə/ => [bo:zə]), or (2) the sequential realization of two corresponding vowels of the same height (/bø:zə/ => [biozə] or [bjozə]). Using the latter strategy, the contrast has apparently been recognized but it is not realized correctly by the learner. The graphemic similarity of <ü, ö> with <u, o> may of course have an influence, especially with respect to the first strategy. Unrounding as a strategy to deal with German front rounded vowels in production is not reported for L1 Romanian learners.

The influence of orthography can also be observed in the case of diphthongs: Romanian orthography provides a very surface-near transcription of diphthongs or vowel-glide sequences which is not given in German, where diphthongs are represented as <ei>, eu> and <au> are mostly pronounced as [ae/ai, ɔi, ao] in German, while in several varieties of German, an even stronger tendency towards monophthongization is given. When realizing German diphthongs, Romanian learners mostly transfer Romanian sequences [ei, ai, aw, ew] to German (Gregor-Chiriță 1991: 44).

To summarize, difficulties in perception of German vowels are expected (1) with front rounded vowels, (2) with the correct identification of vowels with respect to the phonemic length and constriction degree, and (3) potentially with the identification of diphthongs. For front rounded vowels, different strategies are expected (cf. the treatment of borrowings), a substitution with front unrounded vowels is less likely.

10.10.4 Results and interpretation

12 Romanian subjects were tested, 2 beginners, 9 advanced and 1 very advanced learners. All participants were tested in Vienna.

The data set discussed here consists of 3230 valid responses (10 missing answers). For each vowel category, 212 to 216 valid responses were delivered.

10.10.4.1 Difficulties

The diagrams in Figure 10.70 and Figure 10.71 show the relative difficulty of German vowel categories for L1 Romanian listeners.

Least difficulties in correct identification are observed with /ɑ:/, /i:/ and /u:/. /ɑ:/ shows the highest percentage of correct identifications (90% correct). Most problems are observed with /ɛ:/ (only 31% correct) and with /ɪ/ (46%) and /ø:/ (55%). Within the class of front rounded

vowels, /œ/ and /y:/ show the highest id_correct scores (both around 79%), while for /ø:/ most difficulties are observed (46% correct). /ʏ/ shows better results (68.5% correct).
For short unconstricted qualities, more problems than for long constricted qualities occur in the identification task (/ʏ/ 69%, /ɔ/ 67%, /ʊ/ 63% and /ɪ/ 46%). Only /œ/ and /ɛ/ show higher id_correct scores than their long constricted counterpart.

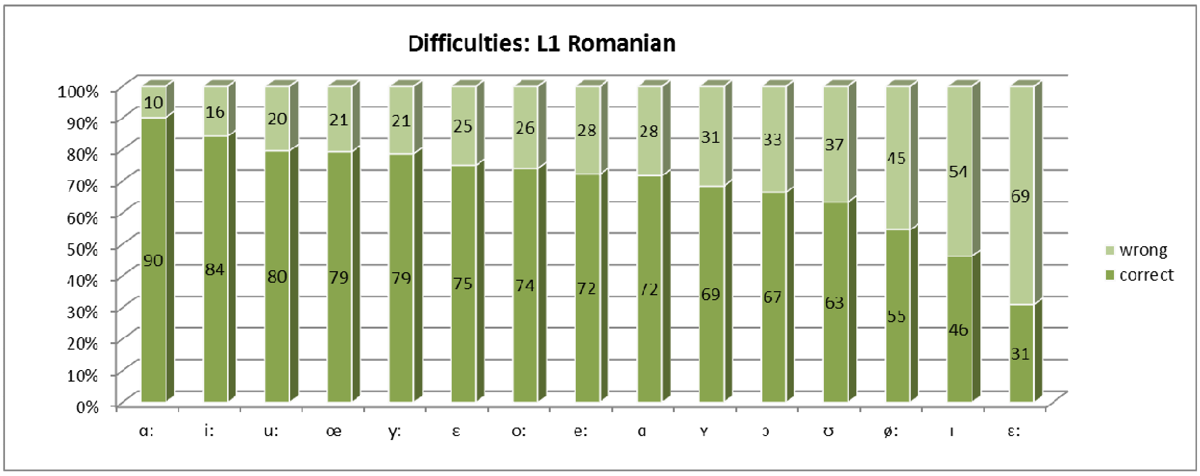

Figure 10.70: Correct and wrong identifications in % for L1 Romanian listeners

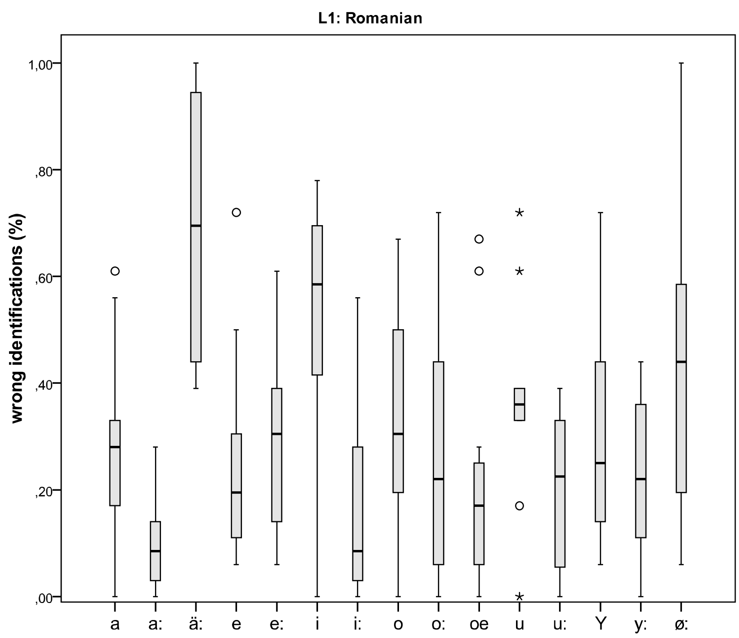

Figure 10.71: Id_wrong scores for L1 Romanian listeners

The boxplot diagram in Figure 10.71 represents inter-personal differences in id_correct scores of L1 Romanian subjects.

/ɛ:/ causes significant problems to all participants. A small range of inter-personal variation of id_wrong scores is observed for /ɑ:/, /œ/, /ʊ/, /u:/, /y:/, and /ɛ/. The highest range is observed for /ø:/, followed by /ɪ/, /o:/, /ɔ/, and /ʏ/.

The listeners' mean id_wrong score was 31% (11% standard deviation), the median was 33%. The listener with the highest percentage of correct answers had an id_wrong of 6%, the highest id_wrong score was 47% wrong responses.

10.10.4.2 Preferences

The preference scores for L1 Romanian listeners are represented in Figure 10.72. The preference scores for monophthong categories vary within a range of 7.1% (min /ɪ/ 3.7%, max /e:/ 10.8%).

The most preferred category is /e:/ (10.8% of the responses), an effect that is mostly due to the high percentage of /ɛ:/-stimuli categorized as /e:/ (/ɛ:/ > /e:/ 61%). Interestingly, it is /ɪ/ that has the lowest pref_score (3.7%) of all monophthongs in the Romanian sub-group, followed by /ɛ:/ (4.3%). For both categories, the low pref_scores coincide with very low id-correct scores. All other categories show pref_scores between 5.3% and 8.1%.

Diphthongs are very rarely selected as output category: <eu> 0.3% and <au> and <ei> 0.1%. The most preferred diphthong is <eu> which serves as substitute for /œ/, /ø:/ and /o:/.

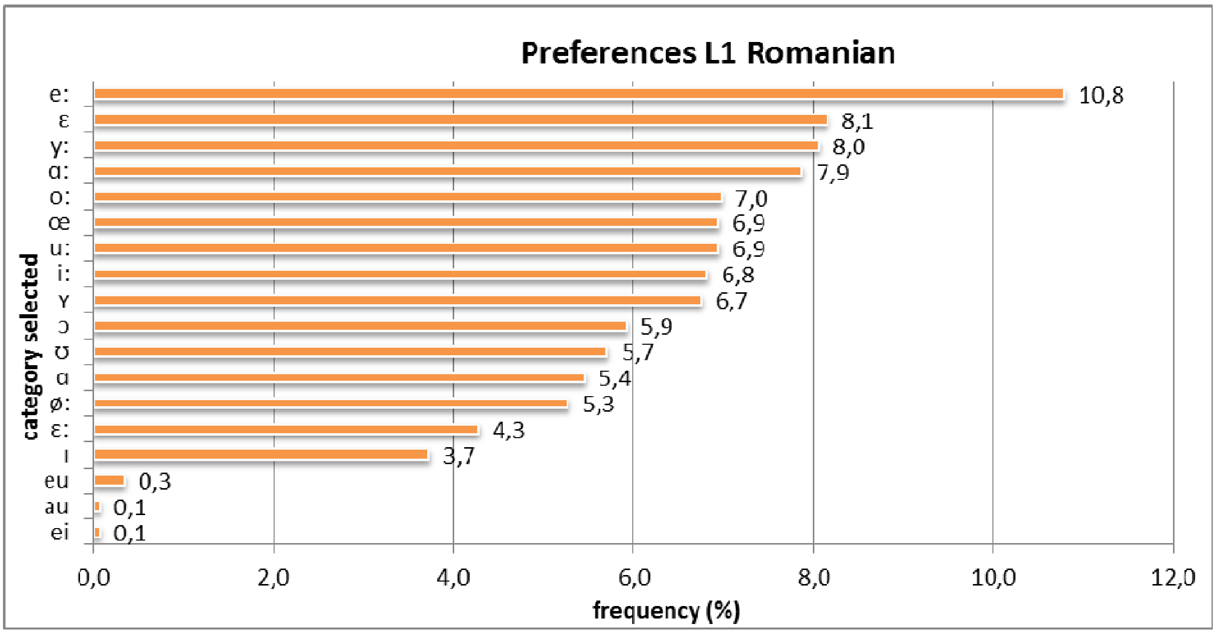

Figure 10.72: Preferred response categories for L1 Romanian listeners

10.10.4.3 Patterns of confusion

Patterns of category confusion for Romanian learners of German are represented in Table 10.63. A first look at the confusion matrix shows (1) difficulties with a-vowels that are only related to length, (2) perceptual confusion within the class of e-vowels, (3) inter-category confusion between e- and i-vowels, (4) inter-category confusion of o-and u-vowels and confusion of back vowels with front rounded vowels, (5) difficulties and inter-category confusion for front rounded vowels and confusion with u-vowels.

%	ɑ	ɑ:	ɛ:	ɛ	e:	ɪ	i:	ɔ	o:	ʊ	u:	œ	ø:	ʏ	y:	‹ei›	‹eu›	‹au›	wrong	**correct**
a	**72**	28																	28	**72**
a:	9	**90**																	10	**90**
ɛ:			**31**	6	61														69	**31**
ɛ			10	**75**	15														25	**75**
e:			18	6	**72**		3												28	**72**
ɪ			2	31	7	**46**	11							1					54	**46**
i:					6	8	**84**												16	**84**
ɔ						1		**67**	25			4	2						33	**67**
o:								5	**74**	3	10	1	5		1		1		26	**74**
ʊ								17	4	**63**	10	1		4	1				37	**63**
u:									2	13	**80**			1	4				20	**80**
œ			3	1								**79**	12				2		21	**79**
ø:										1	2	9	**55**	7	23		1		45	**55**
ʏ										4	2	10	3	**69**	12				31	**69**
y:													1	19	**79**				21	**79**
	5	*8*	*4*	*8*	*11*	*4*	*7*	*6*	*7*	*6*	*7*	*7*	*5*	*7*	*8*	*.1*	*.3*	*.1*	*31.0*	***69.0***

Table 10.63: Confusion matrix L1 Romanian. Correct (in bold) and wrong identifications in % for L1 Romanian listeners (N=12, 3240 responses) (presented stimuli in rows, selected categories in columns in semi-orthographic representation, values < .5 not indicated)

For a-vowels, problems relate to the correct perception of phonemic length: /ɑ/ is identified correctly in 72% of its occurrences and is substituted with long /ɑ:/ in 28% of the cases. /ɑ:/ is identified correctly in 90% of the cases and incorrectly substituted with /ɑ/ in 9% of its occurrences (and in one single case with /i:/).

For e-vowels, inter-category confusion relates in almost all cases to difficulties within the class of e-vowels. /e:/ attracts responses from all other e-vowels. Most difficulties are observed for /ɛ:/, which is correctly identified in only 31% of its occurrences; its most preferred substitute is /e:/ (61%), /ɛ/ is selected in only 6% of the cases.

/ɛ/ shows a much higher id_correct score of 75%, being substituted with /e:/ in 15% and with /ɛ:/ in 10% of its occurrences.

/e:/ is identified correctly in 72% of the cases and is substituted with /ɛ:/ (18%), /ɛ/ (6%), or /i:/ (3%). Romanian listeners show a strong tendency to identify phonemic length correctly, whereas the qualitative contrast of /e:/ and /ɛ:/ is harder to identify.

Long constricted /i:/ is correctly identified in 84% of its occurrences and substituted with /ɪ/ (8%) and /e:/ in 6% of the cases. Interestingly, with Romanian listeners /i:/ is identified more frequently as /e:/ (6%) than /e:/ as /i:/ (only 3%).

While /i:/ shows a very high id_correct score of 84%, a significantly lower score is observed for German /ɪ/ (only 46% id_correct).

Responses for /ɪ/ (46% correct) spread over several categories. Interestingly, the most preferred substitute for /ɪ/ is /ɛ:/ (31%), followed by /i:/ (11%), /e:/ (7%), /ɛ:/ (2%), and /ʏ/ (1%). /ɪ/ is moreover the least preferred response category. Difficulties with the correct identification of /ɪ/ could be related to /ɨ/ that exists as phonemic category in Romanian but has no direct equivalent in German.

Within the class of o- and u-vowels, more difficulties are observed with short unconstricted categories (/ʊ/ 63% correct, /ɔ/ 67% correct) than with long constricted vowels (/u:/ 80% correct, /o:/ 74% correct). Most inter-category confusion is observed for /ʊ/ and /o:/, for which responses spread over several categories. /u:/ is the most stable category. However, /o:/ is the most preferred response option within the class of back vowels. Occasionally, back roundedvowels are incorrectly categorized as front rounded vowels.

Short unconstricted /ɔ/ is identified correctly in 67% of the cases and substituted with long constricted /o:/ (25%), /œ/ (4%), /ø:/ (2%), and /ɪ/ (1%).

Long constricted /o:/ is identified correctly in 74% of the cases and incorrectly categorized as /u:/ (10%), /ɔ/ (5%), /ø:/ (5%), /ʊ/ (3%), or with /œ/, /y:/ or <eu> (1% each).

For unconstricted /ʊ/ (63% correct) a large range of different qualities serve as substitutes. It is incorrectly categorized as /ɔ/ (17%), /u:/ (10%), /o:/ (4%), /ʏ/ (4%), /œ/ (1%) or /y:/ (1%).

Long constricted /u:/ is correctly identified in 80% of the cases. Its most frequent substitute is /ʊ/ (13%), followed by /y:/ (4%), /o:/ (2%), and /ʏ/ (1%).

For the class of back roundedvowels, an asymmetric pattern in perceptual substitutions is observed: While /o:/ is identified as /u:/ in 10% of the cases, the inverse pattern /u:/ > /o:/ is less frequently observed (only 2%). For short unconstricted back vowels, an inverse pattern is observed: /ɔ/ > /ʊ/ (only one single case) vs. /ʊ/ > /ɔ/ (17%).

Front rounded vowels cause considerable problems in perception. While /œ/ and /y:/ both show id_correct scores of 79%, more difficulties are observed for /ø:/ (55% correct) and /ʏ/ (69% correct) more difficulties are observed.

/œ/ is identified correctly in 79% of its occurrences. It is most often substituted with /ø:/ (12%), followed by /ɛ:/ (3%), /eu/ (2%) and /ɛ/ (1%).
/ø:/ (55% id-correct) is incorrectly categorized as long constricted /y:/ in 23% of the cases. Other substitutes are /œ/ (9%), /ʏ/ (7%), /u:/ (2%), /ʊ/ (1%) and <eu> (1%).
/ʏ/ is identified correctly in 69% of the cases and mainly substituted by /y:/ (12%) and /œ/ (10%). Other substitutes are /ʊ/ (4%), /ø:/ (3%), and /u:/ (2%).
/y:/ (79%) shows less variation in the choice of response alternatives. It is mainly substituted with /ʏ/ (19%) but only in 1% of the cases with /ø:/.
As predicted, perceptual unrounding of labio-palatal vowels, i.e. the substitution of front rounded with front non-rounded qualities, is not observed in the data, except for /œ/ which is categorized as /ɛ/ in three cases (1%) or /ɛ:/ <ä> in 3% of the cases (presumably for reasons of perceptual and orthographic similarity).
Like with back roundedvowels, we find an asymmetric substitution pattern: Long constricted /ø:/ > /y:/ is observed in 23% of the cases, but /y:/ > /ø:/ in only 1%. On the other hand, short unconstricted /œ/ is never substituted with ü-qualities, but /ʏ/ is classified as /œ/ in 23% of the cases. As was mentioned above, the listeners' response behaviour shows a strong confusion effect with /ø:/ and /ʏ/, which are perceptually substituted by o-, u- and ü-vowels but never classified as o-categories.
To summarize, for front as well as back rounded vowels, asymmetric effects are observed in perception, such that long constricted /o:/ and /ø:/ are rather substituted with /u:/ and /y:/ respectively than the other way round. For short unconstricted /ɔ/ and /œ/ no evidence for perceptual substitution with /ʊ/ or /ʏ/ is found, while there are strong tendencies to categorize /ʊ/ as /ɔ/ (or /o:/) and /ʏ/ as /œ/ (or /ø:/). In other words, constricted roundedqualities are perceived as more fronted, i.e. palato-velar u-qualities, whereas unconstricted qualities are rather substituted with more back, i.e. uvular/pharyngovelar qualities.
However, this asymmetric effect does not become evident for front non-rounded vowels with Romanian learners of German (long constricted /e:/ > /i:/ (only 3%) vs. /i:/ > /e:/ in 6%). This pattern could be due to the fact that Romanian listeners make use of the pre-palatal constriction location for Romanian /i/. We can therefore posit that the German contrast of mid-palatal /e:/ vs. pre-palatal /i:/ does not cause many difficulties to Romanian listeners, who seem to be familiar with the differentiation of vowels articulated at the pre- and the mid-palatal constriction location in the front part of the vocal tract. The assumption of the use of the pre-palatal location for Romanian /i/ is moreover supported by the presence of /ɨ/ in the

Romanian system. However, as has been mentioned above, the acoustic data resources for Romanian vowels are still too scarce to fully verify such an assumption.

With German e-vowels, more confusion is observed, mostly due to categorization difficulties with /ɛ:/. Short unconstricted /ɪ/ is the most difficult category for Romanian listeners. Responses for this category spread over all e- and i-categories, in a few cases subjects even ü-vowels are considered as response option.

10.10.4.4 Similarity and distance

Table 10.64 summarizes the similarity scores for German vowels as perceived by L1 Romanian listeners.

V	ɑ	ɑ:	ɛ:	ɛ	e:	ɪ	i:	ɔ	o:	ʊ	u:	œ	ø:	ʏ	y:
a	1.000														
ɑ:	.229	1.000													
ɛ:	.000	.004	1.000												
ɛ	.000	.000	.153	1.000											
e:	.000	.000	.762	.145	1.000										
ɪ	.000	.000	.024	.260	.059	1.000									
i:	.000	.003	.008	.006	.059	.149	1.000								
ɔ	.000	.000	.000	.000	.000	.008	.003	1.000							
o:	.000	.000	.000	.000	.000	.000	.003	.212	1.000						
ʊ	.000	.000	.000	.000	.000	.004	.000	.132	.048	1.000					
u:	.000	.000	.000	.000	.000	.000	.003	.000	.080	.155	1.000				
œ	.003	.000	.030	.009	.000	.000	.003	.032	.009	.006	.000	1.000			
ø:	.000	.000	.000	.000	.004	.005	.003	.019	.040	.012	.017	.157	1.000		
ʏ	.000	.000	.000	.003	.000	.008	.000	.000	.000	.060	.019	.069	.083	1.000	
y:	.000	.000	.000	.000	.000	.004	.000	.000	.006	.010	.029	.000	.184	.211	1.000

Table 10.64: Similarity matrix for L1 Romanian listeners

Table 10.65, Table 10.66 and Table 10.67 summarize the sim_scores for back contrast pairs, front contrast pairs, and front vs. back contrast pairs separately. Sim_scores for contrasts involving front vowels range from 0.762 (for /e:/ - /ɛ:/) to 0.003. Sim_scores for back contrast pairs range between 0.229 (/ɑ/ - /ɑ:/) and 0.002. Contrast pairs involving front vs. back vowels show considerably lower values between 0.060 and 0.003.

As evident in the confusion matrix in Table 10.63, German vowel phonemes are perceptually quite easily differentiated by Romanian listeners regarding the front/back dimension (with the only exception of /œ/). On the other side, the perceptual differentiation of German vowels with respect to the round/non-rounded dimension is even clearer to Romanian listeners, as becomes also evident in the two-dimensional MDS solution in Figure 10.74.

The sim_scores suggest that for vowels in the *back* region of the vocal tract differences with respect to constriction location are differentiated quite clearly by L1 Romanian listeners. The sim_scores for back vowels reveal the highest perceptual similarity for contrasts in phonemic length and constriction degree: /ɑ/ - /ɑ:/ (0.229), /ɔ/ - /o:/ (0.212), and /ʊ/ - /u:/ (0.155).

back Vs	*sim_scores*
ɑ - ɑ:	.229
ɔ - o:	.212
ʊ - u:	.155
ʊ - ɔ	.132
u: - o:	.080
ʊ - o:	.048
ɔ - ø:	.019

Table 10.65: Similarity scores > .001 for L1 Romanian listeners for back vowel contrasts

Lower values are given for contrast pairs whose members differ with respect to the velar vs. uvular constriction location, i.e. for contrasts of o- vs. u-vowels: The sim_score for short unconstricted /ʊ/ - /ɔ/ is higher (0.132) than for long constricted /u:/ - /o:/ (0.048). Lower values are observed for /ʊ/ - /o:/ (0.048) and /ɔ/ - /ø:/ (0.019).

front/back Vs	*sim_scores*
ʊ - ʏ	.060
ø: - o:	.040
œ - ɔ	.032
u: - y:	.029
u: - ʏ	.019
u: - ø:	.017
ʊ - ø:	.012
ʊ - y:	.010

Table 10.66: Similarity scores > .01 for L1 Romanian listeners for front-back vowel contrasts

Confusion of vowels located in the *front vs. back* part of the vocal tract is comparatively seldom. In most instances, where confusion of front vs. back vowels occurs, rounding together with vowel height are the common features that cause perceptual confusion and wrong categorization, e.g. /ʊ/ - /ʏ/ (0.06), /ø:/ - /o:/ (0.04), /œ/ - /ɔ/ (0.32), /u:/ - /y:/ (0.029), /u:/ - /ʏ/ (0.019), /u:/ - /ø:/ (0.017), /ʊ/ - /ø:/ (0.012), /ʊ/ - /y:/ (0.010), or /o:/ - /œ/ (0.009). All other contrast pairs show sim_scores below 0.01.

A look at the sim_scores for *front vowel* contrast pairs shows two major areas of difficulty: (1) confusion within the class of e-vowels and contrasts involving /ɪ/, and (2) contrasts involving ö- and ü-vowels. The highest sim_score is obtained for /e:/ - /ɛ:/ (0.762). A strong effect of perceptual under-differentiation with respect to the pre- vs. mid-palatal constriction location is observed for /ɪ/ - /ɛ/ (0.260) and /ø:/ - /y:/ (0.211), but also for /ʏ/ - /ø:/ (0.083), /œ/ - /ʏ/ (0.069), /ɪ/ - /e:/ (0.059), /i:/ - /e:/ (0.059), /ɪ/ - /ɛ:/ (0.024). Other pre- vs. mid-palatal contrast pairs show scores below 0.01. Perceptual similarity due to under-differentiation of a contrast in phonemic length and constriction degree yields relatively high sim_scores for /ʏ/ -

/y:/ (0.211), /œ/ - /ø:/ (0.157), /ɛ/ - /ɛ:/ (0.153), /i:/ - /ɪ/ (0.149), and /e:/ - /ɛ/ (0.145). Similarity of roundedvs. non-rounded vowels is relatively limited (sim_scores below 0.01), only a few instances of perceptual confusion with respect to the feature [round] are observed resulting in very low sim_scores: /œ/ - /ɛ/ (0.009), /ʏ/ - /ɪ/ (0.008), /ɔ/ - /ɪ/ (0.008), /ø:/ - /ɪ/ (0.005), /ø:/ - /e:/ (0.004), /ø:/ - /i:/ (0.003), /ʏ/ - /ɛ/ (0.003), /œ/ - /i:/ (0.003).

front Vs	*sim_scores*
e: - ɛ:	.762
ɪ - ɛ	.260
ʏ - y:	.211
ø: - y:	.184
œ - ø:	.157
ɛ - ɛ:	.153
i: - ɪ	.149
e: - ɛ	.145
ʏ - ø:	.083
œ - ʏ	.069
ɪ – e:	.059
i: - e:	.059
œ - ɛ:	.030
ɪ - ɛ:	.024

Table 10.67: Similarity scores > .01 for L1 Romanian listeners for front vowel contrasts

10.10.4.5 The perceptual vowel map for L1 Romanian listeners

From the similarity scores calculated by Shepard's formula, distance scores and non-metric MDS solutions were derived. For the non-metric MDS solutions, RSQ values of 0.476 in dimension 1, 0.751 for the two-dimensional solution, and 0.830 for the three-dimensional solution are obtained.

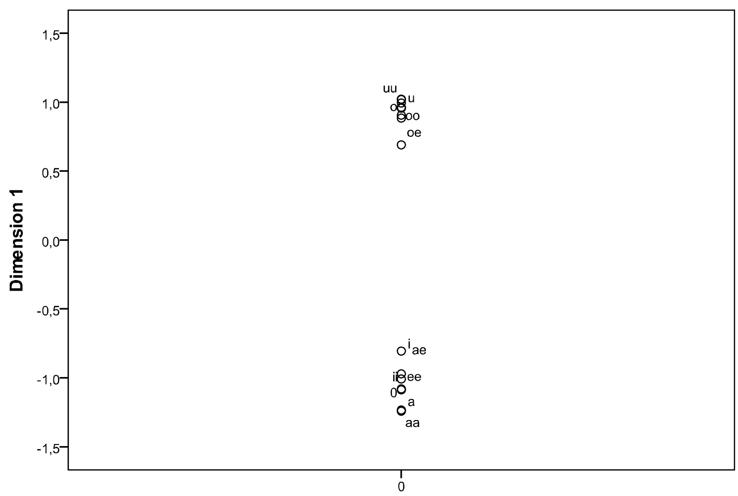

Figure 10.73: One-dimensional MDS representation of the perceptual map of German vowels for L1 Romanian listeners (N=12, 3230 responses, RSQ = .476)

The one-dimensional MDS solution (RSQ .476) shows that roundedand non-rounded vowels are perceived as clearly distinct (Figure 10.73).

The two-dimensional MDS solution in Figure 10.74 presents similarity relations between German vowels together with perceptual preferences for L1 Romanian learners of German indicated in type size. The two-dimensional MDS map shows that the distinction of roundedvs. non-rounded vowels is a basic differentiation for Romanian listeners. German a-vowels are clearly discerned from all other vowel categories. Front non-rounded vowels cluster tightly together. Within the cluster of palatal vowels, pre- vs. mid-palatal qualities are located closer together, an effect that is mainly due to substitutions involving i-qualities and /e:/ but not /ɛ:/ or /ɛ/.

Within the cluster of front non-rounded vowels, /e:/ is the most preferred category, attracting responses for e- as well as i-stimuli, followed by /ɛ/. The two-dimensional solution shows a strong affinity of /e:/ and /ɛ/, and of /e:/ and /i:/. /ɛ:/, a category with a very low pref_score, is closer to /e:/ and /ɛ/ than to i-vowels, reflecting the fact that i-vowels do not function as perceptual substitute for /ɛ:/. For /ɪ/, which is the most difficult category for L1 Romanian listeners, we observe a slightly closer position to front rounded vowels, which is due to a few instances of perceptual substitutions that are not completely represented in Table 10.63, because values below 0.5% are not indicated in the confusion matrix.

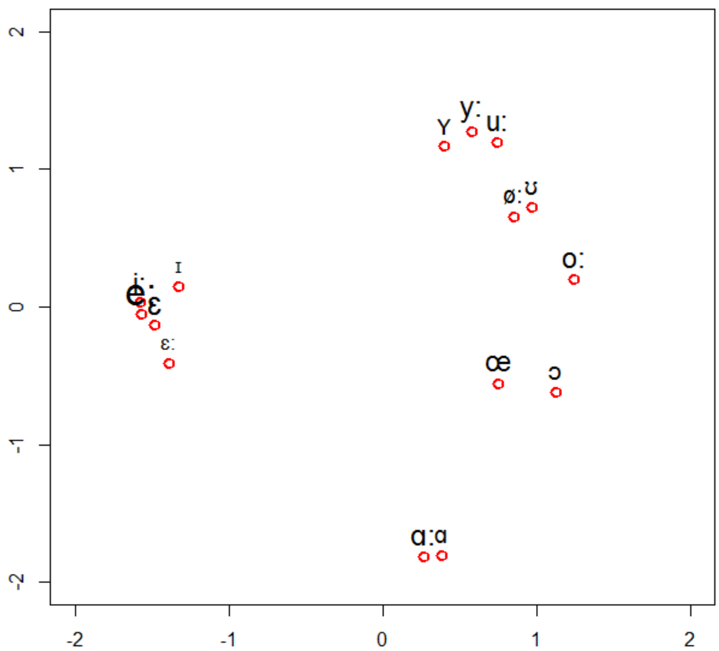

Figure 10.74: Two-dimensional MDS representation of the perceptual map of German vowels for L1 Romanian listeners (N=12, 3230 responses, RSQ = .751)

As observed above, substitutions of back roundedvowels with non-rounded qualities occur only in a few instances for /œ/. Rather, perceptual substitutions of front rounded vowels with back rounded vowels are observed. The substitution of front rounded vowels with back rounded vowels due to insufficient differentiation of constriction location is also observed, as the MDS map shows quite clearly. Moreover, for roundedvowels, a certain sensitivity with respect to vowel height is noticeable, that should however not be overestimated. As becomes evident in the confusion matrix in Table 10.63, responses for front as well as back roundedvowels spread over a considerable range of different categories. Least variation is observed for /y:/. This range of variation in the listeners' responses can of course not be directly represented in the two-dimensional MDS solution.

The three-dimensional MDS representation of the perceptual vowel space for Romanian listeners shows again that a-vowels are most distinct from all other categories and that the roundedvs. non-rounded dimension is well differentiated by Romanian listeners. Interestingly, ö- and o-vowels are located between front i- and e-qualities on the one side and u- and ü-vowels on the other side, a fact that can be explained by the wider spreading of responses for o- and ö-vowels and the occasional substitutions of /œ/ with non-rounded /ɛ/ and /ɛ:/.

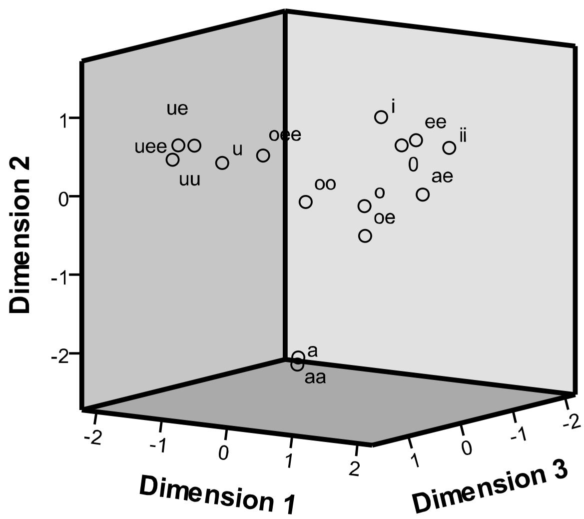

Figure 10.75: Three-dimensional MDS representation of the perceptual map of German vowels for L1 Romanian listeners (N=12, 3230 responses, RSQ = .830)

10.10.5 Summary

In summary, L1 Romanian learners show several areas of difficulty with the perceptual categorization of German vowels:

(1) The differentiation within the class of *e-vowels* is difficult, an effect that is mostly due to categorization problems observed for /ɛ:/, whereas difficulties to differentiate mid-palatal e- from pre-palatal i-vowels are less frequently observed for long constricted /e:/ (in 3% of the cases). This pattern could be due to the fact that Romanian apparently distinguishes the pre-palatal and mid-palatal constriction location for i- and e-vowels, a hypothesis that has been formulated in section 10.10.2 but has to be verified on a larger data base with formant frequency values (especially F2, F3, F4) for all Romanian vowels.

(2) Confusion within the class of *rounded vowels* is considerably strong. Substitutions in both directions ([front round] > [back round] and [back round] > [front round]) are observed in the data, whereas substitutions of front roundedwith non-rounded vowels hardly ever occur. The data show instances of inter-category confusion of o-and u-qualities, especially for /o:/ and /ʊ/, and confusion within the class of front rounded vowels, especially for /ø:/ and /ʏ/.

For both classes of roundedvowels, back roundedand front roundedvowels, asymmetric patterns are observed: Substitutions of long constricted /o:/ and /ø:/ with /u:/ and /y:/ respectively are considerably more frequent than substitutions of /u:/ and /y:/ with /o:/ and /ø:/ respectively. The asymmetry pattern is especially evident with the substitution of long constricted /ø:/ > /y:/ (23%) vs. /y:/ > /ø:/ (only 1.4%). For short unconstricted /ɔ/ and /œ/, the inverse pattern is observed: no evidence for perceptual substitutions of /ɔ/ and /œ/ with /ʊ/ or /ʏ/ is found, while there are strong tendencies to categorize /ʊ/ as /ɔ/ (or /o:/) and /ʏ/ as /œ/ (or /ø:/). In other words, constricted roundedqualities are perceived as more fronted, i.e. palato-velar u-qualities, whereas unconstricted qualities are rather substituted with qualities that are situated more back in the vocal tract.

However, this asymmetric effect does not become evident for front non-rounded vowels with Romanian learners of German.

(3) Perceptual difficulties with *front rounded vowels* mostly result in inter-category confusion of front roundedand back roundedvowels, occurring as substitutions of ü- and ö-qualities with u- and o-qualities. Front rounded vowels do not exist in the primary phoneme inventory of Romanian. However, Romanian native speakers are familiar with ö- and ü-qualities, since they entered into the language via loans from French, German and Turkish, where they are adapted in different manners and degrees. In loanword adaption (cf. Chitoran 2002a: 27ff) as

well as in L2 production (Gregor-Chiriță 1991), two strategies are observed in dealing with front rounded vowels: (a) the substitution with back vowels, or (b) the sequential realization of two corresponding vowels (mostly of the same height, cf. section 10.10.3). These results indicate that [round] is a feature of primary perceptual relevance for Romanian listeners, as becomes evident in the MDS vowel maps, while the differentiation of different constriction locations, i.e. of palatal (i- and e-vowels), palato-velar (u- and ɯ-vowels) and pharyngo-velar/uvular o-vowels appears to cause more difficulties to Romanian listeners.

For further research on the acquisition of German vowels by Romanian learners, a more detailed articulatory and acoustic analysis of Romanian vowel sounds would be necessary. Based on an acoustic analysis (with higher formant frequency values available) it would be possible to treat several questions, e.g. (1) Does Romanian use the pre-palatal constriction location for /i/ to differentiate it from mid-palatal /e/? (2) What is the acoustic-phonetic status of <ə> ([ɨ] or [ɯ]) and <ă>([ə] or [ʌ])? (3) In how far does the existence of these vowels influence the perception of "similar" German vowels? To answer these questions, a test design in the PAM framework would be appropriate. Additional data would provide more insight into the reasons for the high amount of perceptual difficulties with German /ɪ/ for Romanian learners and the role of interference effects of Romanian /i/, /ɨ/ and /ə/.

10.11SerBoCroatian

10.11.1 General characteristics

"Serbo-Croatian" is a term that is used to refer to several closely related dialectal varieties of a South Slavonic language continuum with at least two standardized forms spoken today as national standards in Croatia, Serbia, Montenegro and in Bosnia-Hercegovina, all situated in the countries of former Yugoslavia. Serbo-Croatian, together with Slovenian, forms a sub-group of South Slavonic languages, which is commonly referred to as Western South Slavonic, while Bulgarian and Macedonian are classified as Eastern South Slavonic languages. Sub-dialects and regional varieties within the Serbo-Croatian group are mutually intelligible, though some phonological, morphological and lexical differences exist.

For the branch of Western South Slavonic languages a certain mismatch between the linguistic differentiation of dialects and the definition of national languages can be observed that is strongly conditioned by the rise of nations and (new) political states and is related to the historical affiliation to religious communities (catholic and orthodox) and their centres of power as well as to literary traditions and the use of different writing systems (Glagolitic, Cyrillic, Latin, Arabic). In Croatia the Latin alphabet is used, in Serbia the Cyrillic and the Latin system are used.

The major language of the former republic of Yugoslavia was called Serbo-Croatian *srpsko(-)hrvatski* or *hrvatsko(-)srpski* 'Serbo-Croatian' or 'Croato-Serbian', and was spoken in the Yugoslav republics of Bosnia-Hercegovina, Croatia, Montenegro and Serbia by a total of 17 million people according to the microcensus from 1981 (Corbett & Browne 2011: 333). Several other languages, e.g. Slovenian, Macedonian or Albanian, were also spoken by larger groups of the population of Yugoslavia. A large part of the people from other language groups was proficient in Serbo-Croatian.

Outside of the countries of former Yugoslavia, larger Serbo-Croatian speaking communities developed due to historical migration movements. In the 17th and 18th century, groups of Serbs and Croats migrated to several regions located in present-day Romania, Hungary, Austria and Slovakia. In the last decades, larger groups of speakers of SerBoCroatian migrated to countries of Western Europe, Canada, the United States and Australia and other countries as labour migrants (since the 1960s) or refugees (since the early 1990s).

Throughout the centuries and the changeful political history of the region a large range of different terms were used to refer to the language: *srpsko(-)hrvatski* or *hrvatsko(-)srpski*

‘Serbo-Croatian’ or ‘Croato-Serbian’, *srpski ili hrvatski* and *hrvatski ili srpski* ‘Serbian or Croatian’ and ‘Croatian or Serbian’, *srpski* ‘Serbian’, *hrvatski* ‘Croatian’, *ilirski* ‘lllyrian’, and more colloquially also *naš jezik* or *naški* ‘our language’. In the 18th century, regional terms like *slavonski* “Slavonian” were also common. All these terms have been used with a number of meanings and connotations, and their exact significance depends on the writers and as well as on the dialect and the historical period they were writing in (Naylor 1980: 66), reflecting two contrary positions with respect to the status of Serbian and Croatian: (1) considering Serbo-Croatian as a polycentric standard language (like e.g. German) and describing Serbian and Croatian as varieties of the same language, or (2) considering Serbian and Croatian (and more recently also Bosnian and Montenegrin) as autonomous standard languages.

From the first half of the nineteenth century on attempts were initiated to converge the languages of Serbs and Croats to a common standardized language, but since the recent warfare in the countries of former Yugoslavia in the 1990ies of the last century and its political consequences they are mostly considered to be different national languages, which are commonly referred to as Serbian, Croatian, and – more recently – Bosnian and Montenegrin respectively. Croatian and Serbian, but also Bosnian and Montenegrin are thus frequently used to refer to the standard languages of countries in the region of former Yugoslavia. In the beginning of the 1990ies, Croatian and Serbian were constitutionally established as standard languages of Croatia and Serbia. Bosnian (*bosanski jezik*) is used for one of the standard versions based on the western variant of the Štokavian dialect and is used primarily by Bosniaks in Bosnia and Hercegovina. Montenegrin refers to the language of the republic of Montenegro. These languages are often considered to be autonomous languages by the people living in the countries involved, even though many scholars outside the region still prefer to use Serbo-Croatian or BCS (Bosnian-Croatian-Serbian) as a more general cover-term for all varieties in the region, acknowledging Bosnian, Croatian and Serbian as three standardized forms based on very similar linguistic material.

In the current study, the term “SerBoCroatian” or SBC will be used as a cover-term to refer to all standard varieties spoken in the region. This is not meant to deny differences between the language systems but to refer to the essential features that all languages and varieties have in common. Referring to scientific work concerning either SBC in general or more specifically a certain variety, the denomination chosen by the author(s) of the study will be used in the current study.

Although the languages mentioned are all situated in the Balkan region, Standard Serbian and Standard Croatian do not belong to the languages of the so-called Balkan Sprachbund, which share some common structural properties, e.g. the use of suffixes to mark definiteness or the loss of an infinitive, even though Serbian shares a few typological features with languages of the Balkan Sprachbund (mostly in Timok-Prizren dialects in the South and South-East of Serbia, cf. Kunzmann-Müller 2003: 704).

Dialects of the Western South Slavonic languages are customarily classified by the word used for 'what': *što/šta*, *kaj* or *ča* dividing the dialects into the Štokavian, the Kajkavian and the Čakavian group. These three main dialects differ not only in the word for 'what' but in other features like accent patterns, endings, the case and tense system and in vocabulary. Speakers are commonly conscious of local dialects and are able to name the one they are using (Browne & Alt 2004).

Štokavski is nowadays the most wide-spread dialect group which covers Montenegro and Bosnia, parts of Croatia and all of Serbia (except the so-called Timok-Prizren dialects in the southeast that show gradual typological proximity to Macedonian and Bulgarian). Štokavian spread over most parts of formerly Kajkavian and Čakavian areas in the context of the Ottoman conquest of the Balkans in the 15^{th} to 17^{th} century, when groups of Štokavian speakers fled northward and westward.

Kajkavski is spoken around Zagreb in the north-west of Croatia, bordering Slovenia and has some features in common with Slovenian.

Čakavski is spoken in parts of the Croatian coast and most of the Adriatic islands, in Istria and in a small part of northern Croatia. Čakavian preserved a specific accentual system and some morphological features that are more conservative than Štokavian. All the dialects are mutually intelligible, though the Čakavian of the northern Adriatic coast and islands is difficult for others because of its accentuation, which is more like that of Slovene and Russian.

Another division into sub-dialects is made according to a typical vowel or diphthong, the Common Slavonic vowel ě (/jat/), which developed into /i/, /e/ or /(i)je/ in the three different sub-dialects Ikavian, Ekavian and Ijekavian. This differentiation of dialects is mainly based on the pronunciation of the front vowel /e/ from earlier Slavonic ě ь, but there are lexical and morphological differences as well.

In most parts of Serbia, the sub-dialect is Ekavian, while in Croatia, Bosnia and Montenegro Ijekavian is spoken. Ijekavian is the most widespread variant, spoken in the western part of Serbia, Montenegro, the east of Bosnia-Hercegovina and all parts of Croatia which are not Ikavian. Ikavian is spoken in Dalmatia, the west of Bosnia-Hercegovina and parts of Lika and Slavonia.

Čakavian dialects are Ekavian, Ikavian or mixed Ikavian/Ekavian. Kajkavian dialects show varied vowel systems, where ě usually developed to /e/ (Browne & Alt 2004).

Standard Croatian is based on the Ijekavian variant (e.g. *lijepo* 'beautiful', *zvijezda* 'star', *mlijeko* 'milk', *vjetar* 'wind'), and Standard Serbian is Ekavian (*lepo, zvezda, mleko, vetar*). Bosnian generally refers to a western variant of the Štokavian dialect (see the grammar of Ridjanović 2012). However, in Bosnia-Hercegovina, Ekavian, Ikavian and Ijekavian dialects are spoken.

Spelling in SBC is very close to the spoken language, following the principle of "one letter – one sound", with a few exceptions in the Latin script that uses dž, lj, and nj for / ʥ ɲ ʎ/. In the initial process of language codification in the 19th century two competing principles, each connected with one of the two major codification efforts, influenced the relationship of written and spoken form in the modern period of the language: the *phonological* principle ("write as you speak" and "read as is written"), advocated by the language reformer and ethnographer Vuk Stefanović Karadžić's (1787-1864) language reform program, and the *morphophonemic* principle, preserving the integrity of words, favoured by Ljudevit Gaj (1809-1872), a linguist and protagonist of the Croatian national reformation called the 'Illyrian reform movement' (Ronelle 2006: 376ff). In the codified form of the standard language, the phonological approach by Vuk Karadžić was favoured in most issues, reproducing the speech facts of voicing assimilation and cluster simplification in spelling, though in some instances the principle of "write as you speak" was relaxed to allow for morphophonemic spelling instead of phonological spelling (for some peculiarities, see Gvozdanović 1980: 105).

10.11.2 Phonological and phonetic description

Serbo-Croatian has five vowels, 25 consonants and 4 so-called accents or tones. Vowels comprise 41.9% of all sounds in an analysis of lexical material by Kostić & Das (1969).

The SBC consonant system is represented in Table 10.68. A characteristic feature of the consonant system is a series of affricates /ts ʧ ʤ ʨ ʥ/ represented as <c č dž ć đ> in the writing system, and of palatal consonants /ʨ ʥ ɲ ʎ j/ written as <ć đ nj lj j>.

	labial	labio-dental	dental	alveolar	post-alveolar	Palatal	velar
Plosives	p b		t d				k g
Affricates			ʦ		ʧ ʤ	ʨ ʥ	
Nasals	m		n			ɲ	
Fricatives		f	s z		ʃ ʒ		x
Tap				r			
Approximant		ʋ				j	
Lateral Approximants					l	ʎ	

Table 10.68: The consonant system of Serbo-Croatian (Landau et al. 1999: 83)

The Serbo-Croatian vowel system consists of five phonemes /i e u o a/. Nasal vowels and diphthongs do not occur in SerBoCroatian.

For Ijekavian dialects, a diphthong /i̯e/ is assumed, that can also be pronounced [ije] (see Figure 10.76). Landau et al. (1999: 84) posit that [ije] functions as a single syllable, though in spoken language /ije/ occurs in a mono-syllabic and a bi-syllabic variant, e.g. /ʦvjet/ ~ /ʦvijet/ ‘flower’.

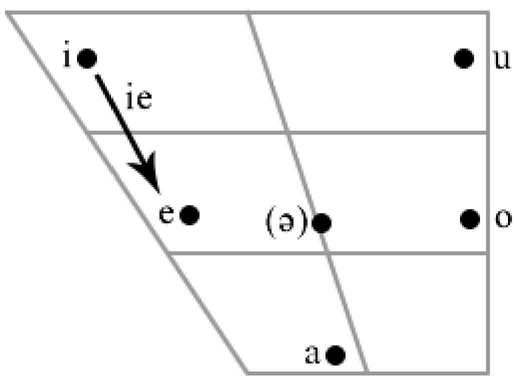

Figure 10.76: Vowels in Croatian (Landau et al. 1999: 84)[126]

Additionally to the five vowel phonemes, a syllabic trill [r̩] is found, that is realized either as long (4 to 5 contacts) or short (1 or 2 contacts) (Landau et al. 1999: 84). Between consonants it is sometimes pronounced with non-phonemic [ə], e.g. /vr̩t/ [ʋərt] ‘garden’. Syllabic /r̩/ is treated as vowel phoneme in some phonological descriptions (e.g. Browne & Alt 2004: 11; Corbett & Browne 2011: 336). /l/ does not function as syllable centre.

[126] http://upload.wikimedia.org/wikipedia/commons/9/91/Croatian_vowel_chart.png (accessed 2014-11-26)

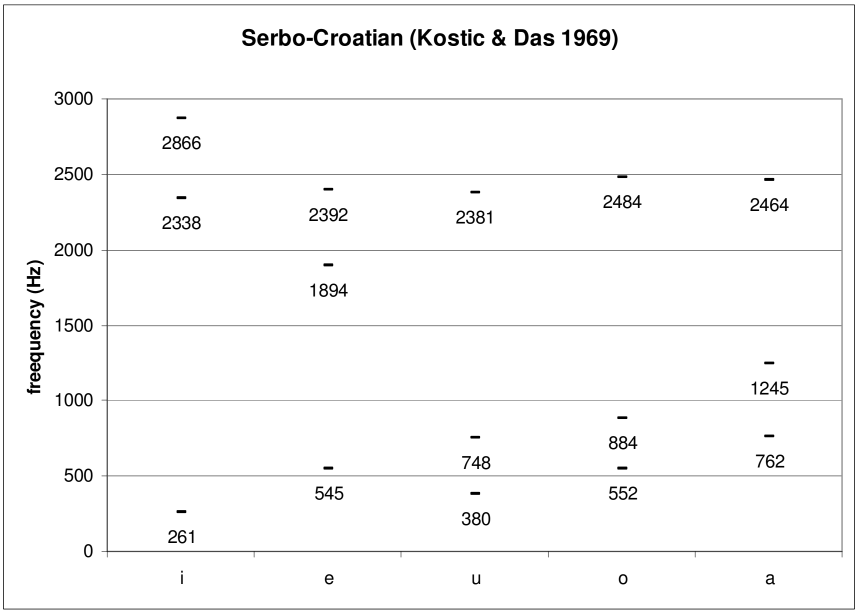

Figure 10.77: Vowel formant frequencies (F1, F2, F3) for vowels in Serbo-Croatian averaged over 10 male subjects and over four accents (Kostić & Das 1969)

Table 10.69 presents the mean formant frequencies (in Hz) for the first three formants (averaged over 10 male subjects and four different accents) as measured by Kostić & Das (1969). In an analysis of lexical material, 41.9% of the phonemes are vowels, where the five different vowel qualities are not equally distributed. While /i/ (10.9%) and /a/ (13.6%) occur very frequently, /e/ and /o/ and especially /u/ are considerably less often found in the data (see Table 10.69).

	F1	F2	F3	frequency of occurrence
/i/	261	2338	2866	10.89
/e/	545	1894	2392	7.05
/u/	380	748	2381	3.41
/o/	552	884	2484	6.95
/a/	762	1245	2464	13.58

Table 10.69: Serbo-Croatian vowels: Formant frequencies (in Hz) and percentage of occurrence in lexical material (Kostić & Das 1969)

A comparison with formants measured by Babić (1991) for Croatian (Figure 10.78) show no substantial differences in the acoustic characteristics of vowel sounds.

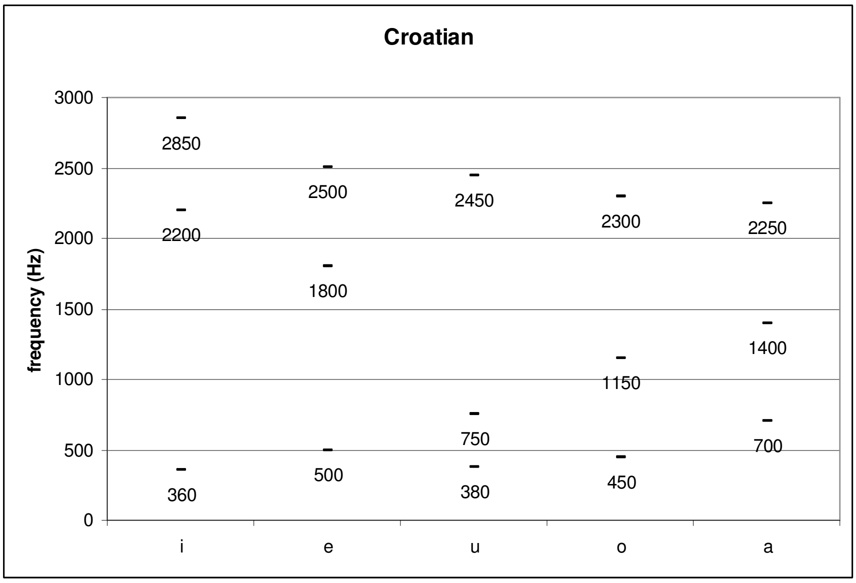

Figure 10.78: Vowel formant frequencies (F1, F2, F3) for vowels in Croatian (Babić 1991: 186)

All vowels can occur in initial, medial or final position of a word and may occur before or after any consonant. All vowels and syllabic /r̩/ can be realized under pitch accent or can be unaccented. Neither Standard Croatian nor Standard Serbian have qualitative vowel reduction (Kunzmann-Müller 2003).

All vowels can occur in quantitative long-short oppositions in SBC. Phonemic length functions to discriminate different words or word forms (e.g. genitive singular vs. plural). No substantial qualitative differences are claimed to exist in stressed or unstressed syllables (Kostić & Das 1969; Landau et al. 1999). However, in minimal pairs as cited e.g. in Landau et al. (1999: 84), native speakers also show qualitative differences connected with differences in quantity, at least for short /e/ [ɛ] and /o/ [ɔ] (see the transcription in phonetic brackets in examples (b) and (d) below).

In addition to the length contrast, the standard language uses a system of rising and falling pitch accents. Accented syllables either contain a vowel or a syllabic /r̩/ nucleus. The standard language uses a system of four different accents, two falling (a long and a short falling) and two rising (a long and a short rising) accent. In other words, the five vowels can each be accented in four different ways, increasing the system up to 20 different states.

In studies on SBC, "accent" is the preferred terminus for these distinctions, indicating that pitch and length rather than intensity are the features involved in contrast pairs. Vowel quality changes according to the pitch level during the stressed vowel itself, its relation to the pitch level of the next syllable, and the duration of the stressed syllable, which falls into two classes, long and short. Vowels with a short accent type are more opened in timbre than vowels stressed with a long accent type (Sovilj-Nikić et al. 2006). For the description of acoustic and perceptual correlates of pitch movement see Magner & Matejka (1971), Gvozdanović (1980) and Lehiste & Ivić (1986).

A small number of words have no accent-bearing syllable, e.g. proclitic *ne* 'not', enclitic *li* and certain pronoun or word forms (Browne & Alt 2004: 16). Apart from these, all words have one accented syllable. Within the paradigm of a word, placement or contour of accent may alternate.

Traditional grammars and dictionaries use diacritic signs for the four accents: long falling [ˆ], short falling [̏]), long rising [́], and short rising [̀]). Post-accentual length is notated with a macron, e.g. *prâvda* 'justice' – *prâvdā* 'he/she justifies'. Post-accentual length is shortened in most dialects in specific positions.

Accent is an essential feature of the word that functions distinctively. The following examples show some minimal pairs, where words differ with respect to length and/or accent.

(a) *vȉle* /ʋíle/ 'hayfork' – *víle* /ʋǐ:le/ 'fairies'

(b) *tȅk* /têk/ [tɛk] 'only' – *têk* /tê:k/ 'appetite'

(c) *pȁs* /pâs/ 'dog' – *pâs* /pâ:s/ 'belt'

(d) *kȍd* /kôd/ [kɔd] 'by, at' – *kôd* /kô:d/ 'code'

(e) *dùga* /dûga/ 'stave' – *dúga* /dǔ:ga/ 'rainbow' – dûg 'debt' – *dȕg* 'long (m.)' – dùga 'long (fem.)'

Vowel timbre may vary with accent type. In general, vowels with a short accent type have a more opened timbre than vowels with a long accent (Sovilj-Nikić et al. 2006: 387), as indicated in the phonetic transcription of example (b) and (d).

Vowel length is found in stressed and post-tonic positions (Corbett & Browne 2011: 337; for a detailed analysis, see Lehiste & Ivić 1963; Magner & Matejka 1971; Gvozdanović 1980). Length may occur in accented and non-accented syllables. Accent is not found in final syllables. In non-accented syllables, the inventory is thus smaller than in accented syllables, where length and accent can be realized (cf. Müller & Chudoba 1999). Post-accentual syllables are mostly short in many regional varieties. In the modern Štokavian system, falling tone is found only in initial syllables, since stress moved historically from all other syllables to produce rising tone. Monosyllables have falling tone because they were not involved in stress shifts. In Bosnian usage, the old distinctions of long and short as well as rising and falling accents survive well, whereas in Standard Croatian length distinctions are preserved but rising and falling accents are not reliably distinguished and post-accentual vowels are shortened. Speakers of Standard Serbian preserve the contrast of short rising and short falling in most words, long rising and long falling tend to be distinguished, and most older post-accentual lengths are lost (Browne & Alt 2004: 17).

Sovilj-Nikić et al. (2006) present a statistical analysis of vowel duration (based on an analysis of two hours continuous speech from a female Serbian speaker), analyzing the influence of neighbouring phonemes and their voicing on vowel duration. They moreover present a more detailed analysis of stressed and unstressed vowels and of stressed vowels with closed and open timbre (see Table 10.70, Sovilj-Nikić et al. 2006: 387f).

vowel duration (ms)	i	e	u	o	a	mean
stressed V in voiced context	127.84	146.58	142.49	149.39	170.60	147.38
stressed V in voiceless context	109.07	143.90	122.23	129.33	144.92	129.89
unstressed V in voiced context	77.30	75.74	77.92	75.34	80.05	77.27
unstressed V in voiceless context	55.98	57.30	63.88	65.41	70.44	62.60

Table 10.70: Vowel duration (in ms) in voiced and unvoiced context (Sovilj-Nikić et al. 2006)

The results show that vowels in voiced context are longer than vowels in unvoiced context and that stressed vowels differ substantially in duration from unstressed vowels in voiced as well as in voiceless context. Moreover, the vowel position relative to the stressed vowel within a word has an impact on the duration of unstressed vowels. Qualitative differences due to accent are associated with differences in duration. Interestingly, stressed vowels with opened timbre are reported to be shorter in duration than stressed vowels with closed timbre.

10.11.3 Contrastive Analysis

The vowel inventory of Serbo-Croatian is significantly smaller than the German system. However, the five basic qualities vary considerably when associated with accent and vowel length, causing intra-category phonetic variation with respect to degree of aperture, duration and pitch contour especially for /e/ and /o/. Vowels with rising accent show a lower degree of aperture than vowels with falling accent. Long vowels can be pronounced as very long in duration (Müller & Chudoba 1999; Sovilj-Nikić et al. 2006).

In very general terms, SBC /e o a/ are described as more similar in quality to their German short unconstricted equivalents, whereas /i/ and /u/ rather correspond to German constricted

/i:/ and /u:/ respectively. SBC short vowels are slightly more closed than their German counterparts (Müller & Chudoba 1999: 5).

Difficulties of SBC speakers learning German are described more generally by Müller & Chudoba (1999) and in detail by Gehrmann (1995: 67ff) and will be summarized in this section. Gehrmann's (1995) detailed description of typical difficulties for all German vowel categories is mainly based on his experience as teacher of German at the University of Zagreb and will be summarized here:

German *a-vowels* are described as not sufficiently differentiated with respect to length by SBC learners of German. For /ɑ:/, interference of rising or falling accent is occasionally observed (Gehrmann 1995: 67).

German *o-vowels* differ in degree of lip aperture and lip protrusion, length and constriction degree, whereas these differences are not distinctive in SBC. German /ɔ/ is very similar to the corresponding sound in SBC. However, Gehrmann (1995) and Müller & Chudoba (1999) observe that German /ɔ/ is pronounced as "too close" and "too tense" by SBC learners. On the other side, for long constricted /o:/, lips are not sufficiently protruded, the sound is pronounced as "too open" and "too lax" and similar to [ɔ]. A clear differentiation in pronunciation between German /o:/ and /ɔ/ is therefore difficult for SBC learners of German.

Similar problems are observed with German *u-vowels*. The pronunciation of short unconstricted /ʊ/ is described as articulatorily "similar" to German /ɔ/ and as "too open". Lips are not sufficiently protruded and rounded (Gehrmann 1995: 67). The pronunciation of long constricted /u:/ is described as "too open and lax", lips are not sufficiently protruded and rounded. According to Gehrmann, the inadequate realization is due to the fact that the mouth is not narrow enough and remains in a position similar to German [ɔ] (Gehrmann 1995: 67). In other words, the phonemic distinction of /u:/ and /ʊ/ is not sufficiently realized by SBC learners of German, both vowels are considered to be similar to SBC /u/. However, no problems with the differentiation of o- and u-vowels in German are mentioned by the authors.

German *i-vowels* are not sufficiently distinguished either. Long constricted /i:/ is described as "too close and lax" [sic] with insufficient lip spreading and similar to SBC *zima* 'cold' with occasional pitch change by influence of rising or falling accent, while short unconstricted /ɪ/ is described as "too tense and too close" with a lip position that comes close to e-vowels (Gehrmann 1995: 67f).

Similar effects are observed for German *e-vowels*. Short unconstricted /ɛ/ as pronounced by SBC learners is described as "too tense and too close", but very similar to the short e-vowel in SBC as in *selo* 'village'. Articulatory problems are therefore not observed. Long constricted

/e:/ is realized as “too open” and “too lax” compared to German /e:/, lips are not sufficiently spread and SBC learners’ realizations of /e:/ are described as “too open” (Gehrmann 1995: 68). /e:/ and /ɛ:/ are not sufficiently differentiated. Long unconstricted /ɛ:/ is described as similar to a long e-vowel in SBC (e.g. *petak* ‘Friday’) but a little “more open”, with occasional influence of rising or falling pitch accent (Gehrmann 1995: 68).

Considerable problems with German *ü- and ö-vowels*, which have no equivalents in SBC, are observed. Front rounded vowels as realized by SBC natives are described as very similar in production to the corresponding non-rounded vowels at the same place of articulation with lips not sufficiently rounded and slightly spread (Gehrmann 1995: 68), resulting in insufficient differentiation of German words like *können* ‘can’ – *kennen* ‘know’.

For the class of front rounded vowels, in the contrastive analyses of Gehrmann (1995) and Müller & Chudoba (1999) no differentiation is made between realizations of long constricted and short unconstricted front rounded German vowels. However, the results in Kerschhofer (1995) and Kerschhofer-Puhalo (1998a) for realizations of German vowels by Polish learners show that differences between long constricted and short unconstricted vowels were observed and that short unconstricted /œ/ showed the highest percentage of wrong realizations due to its high degree of aperture and low degree of lip rounding. Similar difficulties are expected for learners of SBC since they lack front rounded vowels, though the learners’ strategies to cope with these differences may of course vary language-specifically.

None of the contrastive descriptions mentioned above assumes that German front rounded ü- and ö-vowels can be confused with u- and o-vowels in pronunciation or perception. Only substitutions of front rounded with front unrounded vowels are reported. However, as the experience of the author of the present study shows, intercategory confusion of ü- and ö-vowels with front non-rounded as well as back rounded vowels does occur in the production of SBC learners of German (e.g. insufficient differentiation of front-back *fördern* ‘foster’ – *fordern* ‘demand’ or pre-/mid-palatal *dösen* ‘doze’ – *düsen* ‘to jet’), especially with learners who have rich experience with German in its written form. Confusion of front rounded with front unrounded vowels can be due to (1) articulatory difficulties, (2) to interference of the German writing system, which uses < ü ö ä> for sounds that do not occur in SBC, and (3) due to the occurrence of the so-called Umlaut-vowels in paradigmatic alternations in German word derivation and inflection.

For the so-called schwa-vowel /ə/ in unstressed syllables in German, no equivalent sound exists in SBC. /ə/ is usually realized as an e-vowel by SBC speakers. In general, the phenomenon of qualitative reduction of unstressed vowels does not exist in SBC, whereas reduction in terms of duration can clearly be observed for SBC vowels in unstressed syllables (Sovilj-Nikić et al. 2006: 388).

To summarize, the vowel systems of German and SBC differ with respect to system size and the contrastive function of length and qualitative differences and with respect to the phonetic realization of vowels that can be considered as equivalents in German and SBC.

In SBC, qualitative variation is mostly associated with pitch contours in accented syllables. Processes of metric or emphatic lengthening in L2 German or English are unlikely with SBC native speakers, since in their native language vocalic length (combined with accent and pitch contour) is a feature that determines the lexical identity of words. SBC natives would therefore not resort to processes of vowel lengthening because it would affect the meaning of words and is therefore suppressed by SBC natives (Josipovic Smojver 2010).

Perceptual problems resulting in incorrect vowel categorizations and difficulties in establishing categories in the target language German are not reported at all in the contrastive studies described above. The results of the present study will reveal whether perceptual problems are found for L1 SBC learners and whether insufficient perceptual differentiation may account for category confusion and articulatory difficulties in the L2 acquisition of German.

10.11.4 Results and interpretation

33 subjects were tested, 17 beginners, 14 advanced and 2 very advanced learners.

The data comprises a group of Croatian language students that had been tested in two groups (11 beginners and 6 advanced learners) in Rijeka (Croatia) and of 16 subjects of Serbian, Croatian and Bosnian origin who were tested in Vienna.

The data set discussed here consists of 8842 valid responses (67 missing answers). For each vowel category, 571 to 594 valid responses were delivered.

10.11.4.1 Difficulties

The most difficult category are /ɛ:/ (only 31.5 % correct) and /ø:/ (31.8% correct). All other categories show significantly better results. Least difficulties are observed with /ɑ/ (88.2% correct), /ɑ:/ (87.5% correct), /i:/ (85.4% correct), and /ɔ/ (84.7%). Id_wrong scores between 80% and 70% are given for /ɛ/, /ɪ/, /ʊ/, /y:/, and /u:/. For /ʏ/, /œ/, /o:/, and /e:/, id_correct scores below 70% are obtained. The lowest id_scores are observed for the long constricted mid-high qualities /o:/, /e:/, /ø:/, and /ɛ:/. Difficulties with these categories are due to several factors which will be discussed below.

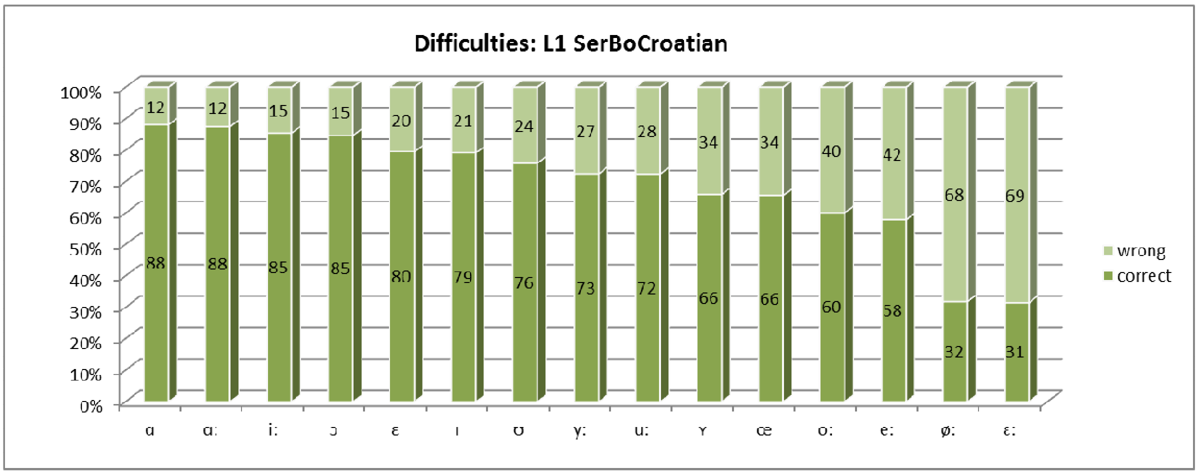

Figure 10.79: Correct and wrong identifications in % for L1 SerBoCroatian listeners

The data for SBC subjects show considerable variation with respect to individual differences in the percentage of wrong identifications (see Figure 10.80). Apart from /ɛ:/ and /ø:/, all vowels are 100-percent correctly identified by at least one participant. In other words, for /ɛ:/ and /ø:/ most categorization problems are observed and even very advanced learners show persistent problems with the correct identification of these two categories.

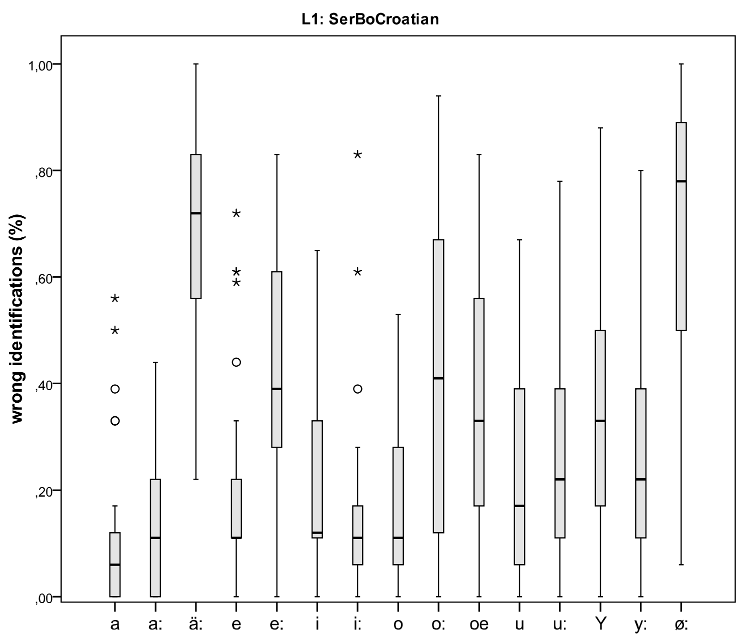

Figure 10.80: Boxplot diagram with id_wrong scores for L1 Serbo-Croatian listeners

The listeners' mean id_wrong score was 31% (14% standard deviation), the median was 29%. The listener with the highest percentage of correct answers had an id_wrong of 9%, the highest id_wrong score was 56% wrong responses.

10.11.4.2 Preferences

Figure 10.81 represents the preference scores for L1 SBC learners of German. The preference scores for monophthong categories vary within a range of 5.9% (min /ɛ:/ 4.0%, max /e:/ 9.9%). Interestingly, /y:/ is the most preferred response category by SBC listeners (9.9% pref_score), followed by /e:/ (8.8%) and /ʏ/ (8.2%). All other categories show pref_scores of 7% or lower. Least preferred are /ɛ:/ (4%), /ø:/ (4.4%) and /o:/ (4.8%), which also show very low id_correct scores. Diphthongs are only occasionally selected by the participants with a percentage of 0.2 % or less. <ei> is the most preferred diphthong with a pref_score of 0.2%, functioning as substitute for e-vowels (and for i-vowels in two cases). <eu> and <au> are selected only in single instances.

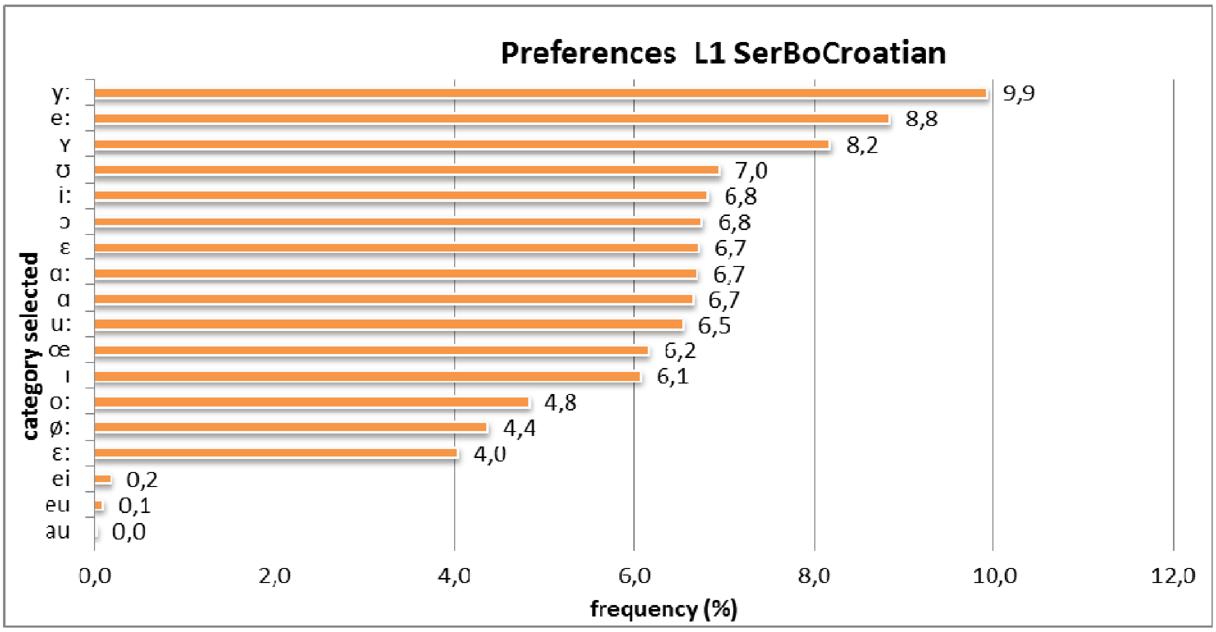

Figure 10.81: Preferences L1 SerBoCroatian

10.11.4.3 Patterns of confusion

Table 10.71 shows the confusion patterns for L1 SerBoCroatian learners of German. A first look at the confusion matrix shows difficulties with the classes of e- and i-vowels, but especially difficulties with the correct categorization of all rounded vowels resulting in a kind of “perceptual overlap” of front rounded and back rounded categories.

%	ɑ	ɑ:	ɛ:	ɛ	e:	ɪ	i:	ɔ	o:	ʊ	u:	œ	ø:	ʏ	y:	‹ei›	‹eu›	‹au›	wrong	**corr**
ɑ	**88**	12																	12	**88**
ɑ:	11	**88**	1																12	**88**
ɛ:			**31**	6	58							1	2			1			69	**31**
ɛ			4	**80**	13							1				1			20	**80**
e:			21	6	**58**	1	8					1	2		1	1			42	**58**
ɪ			1	6	1	**79**	8					1		4					21	**79**
i:					1	10	**85**							2	1				15	**85**
ɔ								**85**	10	1		3	1						15	**85**
o:								6	**60**	3	9	4	13	2	3				40	**60**
ʊ								4	1	**76**	7	2		7	2				24	**76**
u:									1	10	**72**	1	1	5	11				28	**72**
Œ			1	2	2			5	2	2	1	**66**	11	8	2				34	**66**
ø:								1	1	1	5	6	**32**	12	42				68	**32**
ʏ						1				11	2	6	1	**66**	13				34	**66**
y:			1				1			2	2	1	3	18	**73**				27	**73**
	7	*7*	*4*	*7*	*9*	*6*	*7*	*7*	*5*	*7*	*7*	*6*	*4*	*8*	*10*	*0*	*0*	*0*	*3.8*	***69.2***

Table 10.71: Correct (in bold) and wrong identifications in % for L1 SerBoCroatian listeners (N=33, 8842 responses) (presented stimuli in rows, selected categories in columns in semi-orthographic representation, values < .5 not indicated).

German *a-vowels* are quite stable in perception. For short /ɑ/ (88% correct, 12% /ɑ:/), only confusions with respect to phonemic length do occur, while for long /ɑ:/ (88% correct, 11% short /ɑ/) a few instances of wrong categorization with <ä:> (4 cases) are observed.

For German *e-vowels* considerable variation is observed in the listeners' response behaviour. Within the class of e-vowels, short unconstricted /ɛ/ is the most stable category (80% correct). It is most frequently substituted with other e-vowels (/e:/ 13% and /ɛ:/ 4%) and only occasionally with qualities such as /œ/ and <ei> (each 1%). Long constricted /e:/ is correctly identified in only 58% of its occurrences. Substitutions of /e:/ with <ä:> form the greatest part of wrong responses (21%), but considerable confusion with other categories is also observed: /i:/ (8%), /ɛ/ (6%), /ø:/ (2%), and /ɪ/, /œ/, /y:/ and <ei> (1% each).

Long unconstricted /ɛ:/ shows a very low id_correct score of 31%. The most preferred substitute is /e:/ (58%), followed by /ɛ/ (6%), /ø:/ (2%), /œ/ (1%), and <ei> (1%). Interestingly, for /ɛ:/ and /e:/ (and less frequently also for /ɛ/) instances of hypercorrect categorization with [round] qualities such as /œ/, /ø:/ or /y:/ are observed, a pattern that is not evident in most other languages studied here (except Polish).

German *i-vowels* show higher id_correct scores than e-vowels. Long constricted /i:/ (85% correct) is most frequently substituted with /ɪ/ (10%) and occasionally with other qualities: /ʏ/ (2%), /y:/ (1%), and /e:/ (1%). Short unconstricted /ɪ/ is correctly identified in 79% of its occurrences and incorrectly categorized as /i:/ (8%), /ɛ/ (6%), /ʏ/ (4%), or <ä:> (1%), /e:/, and

/œ/ (1% each). Similarly to e-vowels, instances of hypercorrect substitution with front rounded vowels are observed for i-vowels: short unconstricted /ɪ/ is substituted with /œ/ (1%) or /ʏ/ (4%), and /i:/ is hypercorrectly identified as /ʏ/ (2%) and /y:/ (1%). These categorizations are interpreted here as hypercorrect reactions to perceptual uncertainties with front rounded categories.

For *e- and i-vowels* asymmetric response patterns are observed: While /i:/ functions as substitute for /e:/ as well as /ɪ/ (8% each), /e:/ does not seem to be an attractive response category for i-vowels (only 1% each for /i:/ and /ɪ/). This pattern can be explained by the existence of /e:/ and the contrast of pre-palatal i- and mid-palatal e-vowels in German which is not exploited in SBC. The inverse pattern is observed for unconstricted /ɪ/ - /ɛ/: /ɪ/ > /ɛ/ (6%) vs. /ɛ/ > /ɪ/ (0%). /ɛ:/ is substituted with /e:/ in 58% of the cases, whereas substitutions of /e:/ > /ɛ:/ occur only in 21%.

Back rounded vowels show difficulties in categorization that result mainly in confusion within the class of back vowels but also in substitutions with front rounded vowels. Least difficulties are observed for /ɔ/ (85% correct), followed by /ʊ/ (76%) and /u:/ (72%). A very high percentage of wrong categorizations is observed with /o:/, which shows an id_correct score of only 60% and a wide range of different substitution categories.

Long constricted /u:/ (72% id_correct) is most frequently confused with short unconstricted /ʊ/ (10%), as could be expected by the contrastive analysis presented in section 10.11.3, but also with /y:/ (11%) and /ʏ/ (5%) and occasionally also with /o:/, /œ/, and /ø:/ (1% each). Interestingly, in cases of incorrect categorization, /y:/ is the most frequently selected category (11%), though this pattern is not predicted at all by contrastive studies presented above.

Short unconstricted /ʊ/ (76% correct) shows similar patterns: incorrectly selected response categories for /ʊ/ are long /u:/ (7%), short /ʏ/ (7%), and /ɔ/ (4%), but also /y:/ (2%), /œ/ (2%), and /o:/ (1%).

Short unconstricted /ɔ/, the most stable category of all back rounded vowels (85% correct), is incorrectly categorized as /o:/ (10%), /œ/ (3%), and occasionally as /ʊ/ or /ø:/ (1% each).

Long constricted /o:/ shows the highest percentage of wrong responses (only 60% correct) and is substituted with a wide range of different categories. /o:/ is moreover the least preferred back rounded quality. Among the incorrectly selected response categories for /o:/, the most frequently selected one is /ø:/ (13%), followed by /u:/ (9%), /ɔ/ (6%), /œ/ (4%), /ʊ/ (3%), /y:/ (3%), and /ʏ/ (2%).

A striking pattern is the high percentage of substitutions of /o:/ with front rounded /ø:/ as well as substitutions of /u:/ with /y:/, indicating that the features of length and rounding are better detected in the input signal than the constriction location.

Within the class of *front rounded vowels*, large differences in id_correct scores are observed. For front rounded vowels three basic strategies of perceptual categorization can be observed: They are substituted either with (1) other front rounded vowels (indicating that rounding and the front constriction location are sufficiently identified), or (2) with back rounded vowels (only rounding is identified correctly), or (3) (least frequently) with front non-rounded vowels (constriction location but not rounding is identified correctly). However, the perceptual substitution patterns differ considerably between categories.

Long constricted /y:/ has the highest id_correct score (73%), followed by short /ʏ/ (65.9% correct) and /œ/ (65.6% correct), whereas /ø:/ shows the highest number of wrong identifications (only 32% correct). /y:/ is moreover the most preferred response options, while /ø:/ is the least preferred response category.

Long constricted /y:/ (73% correct) is the most stable and also the most preferred category within the class of front rounded vowels, attracting responses especially for /ø:/ (42%) and /ʏ/, but only occasionally for /œ/ (2%). /y:/ is incorrectly categorized as /ʏ/ in 18% of its occurrences and considerably less frequently with other categories: long /ø:/ (3%), /u:/ (2%), short /ʊ/ (2%) or /œ/ (7%), and less frequently with front unrounded /i:/, /ɛ:/ and /ɛ/ (id_scores below 1%).

Short unconstricted /ʏ/ is identified correctly in 66% of the cases and incorrectly categorized as /y:/ (13%), /ʊ/ (11%), /œ/ (6%), /u:/ (2%), /ø:/ (1%), and /ɪ/ (1%).

Short unconstricted /œ/ (66% correct) shows the highest number of different substitutes. It is most often substituted with front rounded /ø:/ (11%), short unconstricted /ʏ/ (8%) and with long constricted /y:/ (2%), but also with back rounded /ɔ/ (5%), /o:/ (2%), /ʊ/ (2%) and /u:/ (1%). Especially for /œ/, several instances of substitutions with front unrounded vowels are observed (/ɛ/ (2%), /e:/ (2%), and <ä:> (< 1%)). Substitutions with front non-rounded vowels are mainly observed for /œ/, which has the lowest degree of lip rounding and lip protrusion and a higher degree of aperture compared to all other vowels of the same class. This effect is considerably weaker for all other front rounded vowels. The fact that <ä:> is selected for /œ/ only in four cases (below 1%) suggests that orthographic factors play a less significant role, but that the perceptual similarity of e-vowels and /œ/ causes the confusion between front rounded and front unrounded vowels (see also the results for L1 Polish subjects in section 10.9.4.3). Substitutions of the type /œ/ > /ɛ/ have to be considered as response patterns that are due to articulatory and acoustic properties in the input signal rather than to orthographic factors.

Long constricted /ø:/ causes most difficulties to SBC learners of German indicated by a very low id_correct score of 32%. In the responses of L1 SBC learners, /ø:/ – together with /ɛ:/ – is the least preferred category. In 42% of its occurrences /ø:/ is categorized as /y:/, indicating that these two categories are not sufficiently differentiated by L1 SBC learners. However, a strong effect of asymmetry is observed here: while 42% of the /ø:/ stimuli are categorized as /y:/, only 3% of the /y:/ stimuli are categorized as /ø:/. Substitutions with long /y:/ indicate that length but not quality is recognized correctly. /ø:/ is also categorized as /ʏ/ in 12% and /œ/ in 6% of the cases, reflecting difficulties with respect to quality differences and the pre- vs. mid-palatal constriction location. Substitutions with back rounded vowels are less frequently observed: most common is /u:/ (5%), followed by /o:/, /ʊ/, and /ɔ/ (1% each). The id_scores for /ø:/ show very clearly, how vague the perceptual representations of front rounded vowels in general are for L1 SBC learners of German. This is also reflected by the position of these qualities between front non-rounded and back rounded vowels in the MDS map in Figure 10.83.

For all rounded vowels, a considerable number of wrong categorizations are observed that seem to be due to a hyper-corrective response behaviour resulting in the selection of a [+round, +front] category for back rounded input stimuli, e.g. /ʊ/ > /ʏ/ (7%) and /ʏ/ > /ʊ/ (11%). However, this effect is not equally strong for all categories and is moreover asymmetric in some cases. The effect is least effective for /ɔ/, probably due to its high percentage of correct identifications, but is particularly strong and asymmetric for /u:/ and /o:/: [u:] > /y:/ 11% (but /y:/ > /u:/ only 2%), [o:] > /ø:/ 13% (but /ø:/ > /o:/ only 1%).

While short unconstricted categories show a slightly higher tendency to be substituted with back vowels, long constricted stimuli are more frequently categorized as [+front] than short unconstricted qualities.

10.11.4.4 Similarity and distance

The similarity scores for L1 SBC learners of German are summarized in Table 10.72. The highest sim_scores are obtained for the contrast /e:/ - /ɛ:/ (0.889) and for the contrast /y:/ - /ø:/ (sim_score 0.431), reflecting the high percentage of substitutions of /ø:/ > /y:/ (42%). High sim_scores are also obtained for /ʏ/ - /y:/ (0.219) and /œ/ - /ø:/ (0.177) differing only with respect to phonemic length and constriction degree.

V	ɑ	ɑ:	ɛ:	ɛ	e:	ɪ	i:	ɔ	o:	ʊ	u:	œ	ø:	ʏ	y:
ɑ	1.000														
ɑ:	.129	1.000													
ɛ:	.000	.008	1.000												
ɛ	.002	.003	.090	1.000											
e:	.000	.001	.889	.142	1.000										
ɪ	.000	.000	.006	.037	.014	1.000									
i:	.001	.000	.003	.001	.057	.108	1.000								
ɔ	.000	.000	.000	.001	.001	.000	.002	1.000							
o:	.000	.000	.002	.000	.000	.000	.000	.116	1.000						
ʊ	.000	.000	.000	.000	.000	.001	.000	.030	.027	1.000					
u:	.000	.001	.002	.000	.000	.000	.000	.001	.069	.113	1.000				
œ	.000	.001	.016	.026	.022	.007	.003	.049	.043	.028	.007	1.000			
ø:	.000	.000	.032	.003	.025	.002	.000	.010	.156	.009	.057	.177	1.000		
ʏ	.000	.000	.002	.001	.003	.032	.011	.005	.015	.125	.047	.105	.133	1.000	
y:	.000	.000	.006	.002	.005	.002	.013	.000	.020	.026	.095	.026	.431	.219	1.000

Table 10.72: Similarity matrix for L1 SerBoCroatian listeners

The results will be discussed here separately for roundedand non-rounded vowels and for the class of front vowels, back vowels and front vs. back contrasts.

For contrasts of *non-rounded* vowels, the highest sim_score is observed for /e:/ - /ɛ:/ (0.889), followed by contrast pairs with significantly lower scores differing in phonemic length (and constriction degree): /e:/ - /ɛ/ (0.142), /ɑ/ - /ɑ:/ (0.129), /i:/ - /ɪ/ (0.108), /ɛ/ - /ɛ:/ (0.090).

The highest sim_scores for *roundedvowel* contrasts are obtained for /ø:/ - /y:/ (0.431), /ʏ/ - /y:/ (0.219), /œ/ - /ø:/ (0.177), /ø:/ - /o:/ (0.156), /ʏ/ - /ø:/ (0.133), /ʊ/ - /ʏ/ (0.125), /ɔ/ - /o:/ (0.116), /ʊ/ - /u:/ (0.113), and /œ/ - /ʏ/ (0.105), followed by several other contrast pairs that receive sim_scores below 0.1.

For the class of *back vowels*, difficulties are mainly observed for contrasts in phonemic length and constriction degree: /ɑ/ - /ɑ:/ (0.129), /ɔ/ - /o:/ (0.116), and /ʊ/ - /u:/ (0.113). Lower sim_scores (< 0.1) are obtained for contrasts that are based on differences in constriction location (velar vs. uvular): /u:/ - /o:/ (0.069), /ʊ/ - /ɔ/ (0.030), and /ʊ/ - /o:/ (0.027).

Major difficulties within the class of *front vowels* are observed for the contrast /e:/ - /ɛ:/ with a very high sim_score of 0.889 due to numerous instances of substitutions of /ɛ:/ > /e:/ (58%) and /e:/ > /ɛ:/ (21%). High sim_scores are especially obtained for front rounded vowels, i.e. for vowel pairs differing in pre- vs. mid-palatal constriction location, such as /ø:/ - /y:/ (0.431), /ʏ/ - /ø:/ (0.133), /œ/ - /ʏ/ (0.105), or in phonemic length and constriction degree like /ʏ/ - /y:/ (0.219), /œ/ - /ø:/ (0.177), /ɛ/ - /e:/ (0.142), and /ɪ/ - /i:/ (0.108). Sim_scores below 0.1 are found for non-rounded pre- vs. mid-palatal vowels (e.g. /i:/ - /e:/ (0.057), /ɪ/ - /ɛ/ (0.037), and for roundedvs. non-rounded contrast pairs. Contrast pairs for front vowels that differ only with respect to the feature [rounding] show sim_scores below 0.04, e.g. /ʏ/ - /ɪ/ (0.032), /œ/ - /ɛ/ (0.026), /ø:/ - /e:/ (0.025), or /y:/ - /i:/ (0.013).

For contrast pairs of front vs. back vowels, sim_scores below 0.16 are obtained. Within this group of contrast pairs, higher sim_scores are obtained for front roundedvs. back roundedvowels that differ only with respect to constriction location (i.e. palatal vs. uvular or palatal vs. velar) but share the features [rounding], [length] and [height], e.g. /ø:/ - /o:/ (0.156), /ʊ/ - /ʏ/ (0.125), /u:/ - /y:/ (0.095), /œ/ - /ɔ/ (0.049). Lower sim_scores are obtained for pairs that differ in more than one feature, e.g. /u:/ - /ø:/ (0.057), /u:/- /ʏ/ (0.047), /o:/ - /œ/ (0.043). Contrast pairs that differ in [rounding], [height] and [length & constriction degree] show even lower sim_score values.

To summarize, the sim_scores reveal two major clusters of similar vowels: e-vowels and roundedvowels. Sim-scores for round-non-rounded contrasts are substantially lower than for contrasts where both members share the feature [round]. This observation is especially relevant for the class of front rounded vowels. Contrastive studies usually report substitutions of front rounded with front unrounded vowels. However, the data of the present study show that perceptual substitutions result considerably more frequently in substitutions with back vowels.

10.11.4.5 The perceptual vowel map for L1 SerBoCroatian listeners

From the similarity scores calculated by Shepard's formula, distance scores and non-metric MDS solutions were derived. For the non-metric MDS solutions, RSQ values of 0.302 in dimension 1, 0.782 for the two-dimensional solution, and 0.878 for the three-dimensional solution are obtained.

The one-dimensional MDS representation differs fundamentally from that for other language sub-groups in that no clear-cut distinction is observed for rounded vs. non-rounded vowels. Front rounded vowels are rather located near front non-rounded vowels. a-vowels are most distant from all other categories and u-and o-vowels occupy a separate position. However, the RSQ-value for the one-dimensional solution is very low, indicating the limited descriptive significance for this solution.

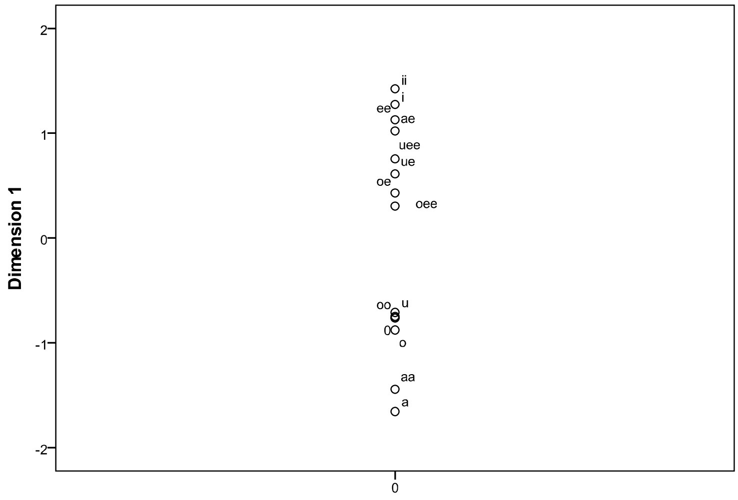

Figure 10.82: One-dimensional MDS representation of the perceptual map of German vowels for L1 SerBoCroatian listeners (N=33, 8842 responses, RSQ .302)

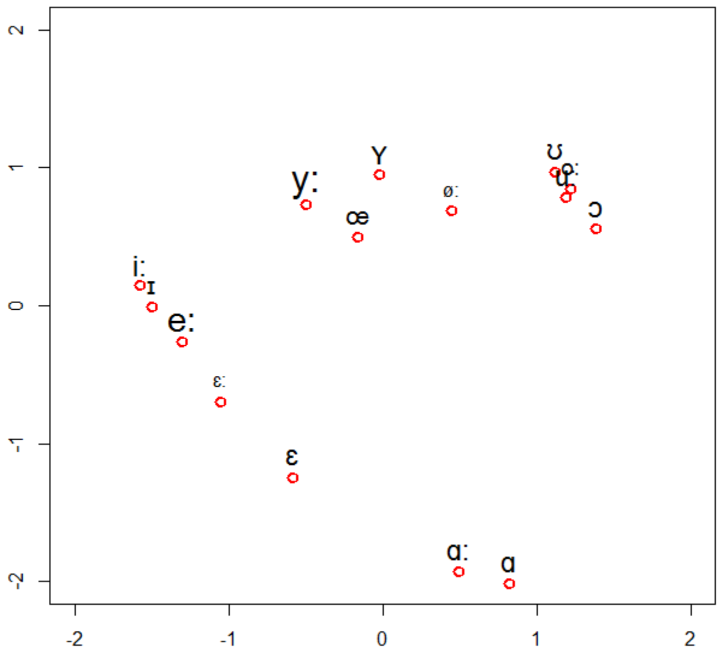

Figure 10.83: Two-dimensional MDS representation of the perceptual map of German vowels for L1 SerBoCroatian listeners (N=33, 8842 responses, RSQ .782)

Figure 10.83 represents the two-dimensional MDS solution for the perceptual vowel map of L1 SerBoCroatian listeners. Differences in type size indicate the relative preferences for specific response categories: While /y:/ and /e:/ are the most preferred categories, /ɛ:/ and /ø:/ are very much dispreferred.

The diagram shows that a-vowels are kept well apart from all other qualities. Similarly, wrong categorizations of i-vowels are mainly due to incorrect identification of length and constriction degree, resulting in the close position of /i:/ and /ɪ/ in the MDS solution.

The diagram shows a cluster of back rounded vowels that are insufficiently distinguished from each other.

The position of front rounded ü- and ö-vowels between the cluster of back rounded and front non-rounded vowels reflects the intermediate status of labio-palatal vowels between these two vowel classes in the perception of L1 SBC learners. Front rounded vowels are substituted by back as well as front vowels, though back rounded categories are more attractive response categories. This is reflected by the position of these categories closer to the back vowel cluster.

Within the group of e-vowels, /ɛ/ (80% correct) is kept quite well apart from all other categories. The position of /e:/ close to i-vowels is conditioned by substitutions of /e:/ with /i:/ (8%). /ɛ:/, which is incorrectly substituted with /e:/ in 58% of its occurrences but only in 6% with /ɛ/, is located between /ɛ:/ and i-vowels.

The region occupied by o- and u-vowels in the diagram is considerably smaller compared to that of front (roundedor non-rounded) vowels, due to a high amount of inter-category confusion for back roundedvowels. /ɔ/ is kept apart best from all other back rounded vowels. The highest percentage of incorrect identifications and a wide-spread range of different substitution candidates are observed for /o:/, which is also the least preferred category in this vowel class.

The MDS solution reflects well the fact that [rounding] can be considered as a feature of primary relevance for L1 SBC learners for front rounded vowels. Moreover, [length] is considerably well identified by L1 SBC learners, whereas several difficulties occur that are related to an incorrect identification of the constriction location. This pattern is particularly evident with the high number of substitutions of front rounded vowels with back rounded vowels. Substitutions with front non-rounded qualities only occur for /œ/ in a larger number of cases.

The three-dimensional solution gives an impression that is very similar to the two-dimensional representation, though the cluster of u- and o-vowels is more differentiated. The intermediate position of ö- and ü-vowels reflects the confusion patterns described above. In this representation, /i:/ is more closely situated to front rounded vowels probably reflecting the occasional hypercorrect substitutions of i-vowels to ü-qualities.

All MDS representations for SBC show that an intermediate position of front rounded vowels between back and front non-rounded vowels must not be understand in the way that these qualities are kept well apart from other classes but rather that can be perceptually substituted with back but also with front vowels. The exact preferences for back or front are vowel-specific and influenced by asymmetric preference patterns.

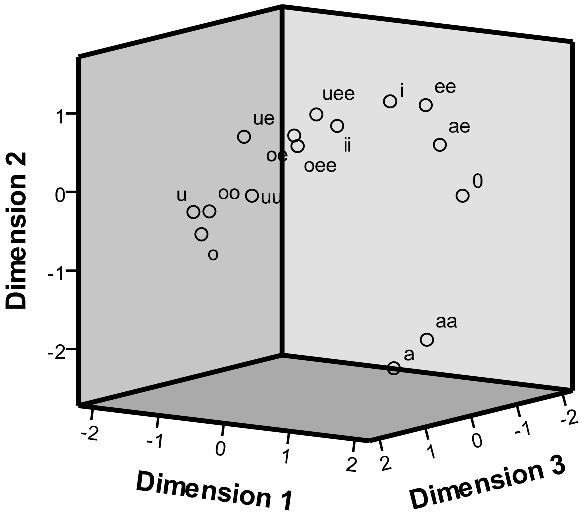

Figure 10.84: Three-dimensional MDS representation of the perceptual map of German vowels for L1 SerBoCroatian listeners (N=33, 8842 responses, RSQ .878)

10.11.5 Summary

To summarize, most problems for SBC learners of German in perceptual categorization are observed with /ɛ:/ (31% id_correct), /ø:/ (32% (id_correct) and /o:/ (60% id_correct). Least difficulties are observed for /a ɑ: i: ɔ ɛ/ (id_correct > 80%). In general, SBC listeners seem to have considerable difficulties with the perceptual differentiation of front rounded from back rounded vowels, a pattern that has not been described in previous contrastive studies. Moreover, the differentiation of e-vowels appears to be difficult, whereas for a- and i-vowels less problems are observed.

Generally, the substitution patterns for long vowels show that *length* is quite well identified, as expected from the contrastive analysis, since SBC does use phonemic length distinctions, and that it is rather quality, i.e. the correct identification of the constriction location, which causes difficulties. This pattern is observed for all long categories where substitutes that differ in quality and quantity are found, i.e. for /ɛ:/ (31% correct, 91% [long] vs. 7% [short]), /i:/

(85% correct, 87% [long] vs. 12% [short]), /o:/ (60% correct, 85% [long] vs. 15% [short]), /u:/ (72% correct, 85% [long] vs. 16% [short]), /ø:/ (32% correct, 80% [long] vs. 20% [short]), /y:/ (73% correct, 80% [long] vs. 21% [short]). A striking pattern in the listeners' responses is the high percentage of substitutions of /o:/ with front rounded /ø:/ and substitutions of /u:/ and /ø:/ with front rounded /y:/, indicating that the features [length] and [rounding] are better detected in the input signal than the identification of the location of constriction.

A strong effect of asymmetry, where categories with a higher degree of aperture serve as substitutes for more open qualities, is found for the contrast of /y:/ - /ø:/ (/ø:/ > /y:/ (42%) vs. /y:/ > /ø:/ (3%)), this effect is weaker for /i:/ - /e:/ (/e:/ > /i:/ (8%) vs. /i:/ > /e:/ (1%)) and for /u:/ - /o:/ (/o:/ > /u:/ (9%) vs. /u:/ > /o:/ (1%)). While short unconstricted categories show a slightly higher tendency to be substituted with back vowels, long constricted stimuli are more frequently categorized as [+front] than short unconstricted qualities.

For the class of rounded vowels, rounding and also length seem to cause less problems than the correct identification of the constriction location. While /ɔ/, /ʊ/, /u:/ and /y:/ show quite good results with respect to qualitative and quantitative differentiation, other categories cause considerably more difficulties and show a wide range of substitution candidates (see the confusion matrix in Table 10.71), indicating that front rounded and back rounded vowels are perceptually not sufficiently differentiated with respect to the location of constriction. Only the contrast of ü- and o-vowels is differentiated well by all participants. This under-differentiation of [+round] categories results in perceptual substitutions of front rounded vowels with back vowels and of back vowels with front rounded categories. This under-differentiation is also evident in the similarity scores and is assumed to be due to (1) articulatory and acoustic similarity, (2) orthographic similarity, and (3) by morphophonemic alternations of front roundedUmlaut sounds and back roundedsounds are involved.

To conclude, the features [rounding] and [length] are well identified by SBC learners of German, whereas the identification of the constriction location causes considerable difficulties resulting in instable categories and in confusion of front rounded vowels, which do not exist in SBC, and back rounded vowels and for some categories also with front non-rounded vowels.

10.12Turkish

10.12.1 General characteristics

Turkish belongs to the South-Western branch of the Turkic language family. It is the official language in Turkey and Northern Cyprus and is a recognized minority language in Kosovo, Macedonia and Romania. It is further spoken by smaller groups of speakers in several other countries like Iraq, Syria, Greece, Bulgaria, Albania, Montenegro, Turkmenistan, Azerbedjan and Uzbekistan. Via migration it spread to Western Europe, the United Kingdom and the United States. Larger immigrant communities of Turkish origin live in Germany, Belgium, the Netherlands, Sweden and Austria.

Common features of the Turkic language family are vowel harmony, verb-final order and agglutinative morphology. As an agglutinating language, Turkish has an extremely productive inflectional and derivational morphology and lexicon. Turkish words consist of morphemes concatenated to a root morpheme. Morphotactics can be quite complex when multiple derivations are involved. Morphophonemic processes such as vowel harmony, consonant assimilation and elision determine the surface realization of the root morpheme and the derivational and inflectional suffixes (Oflazer & Inkelas 2003). A considerable part of the Turkish lexicon consists of loanwords mostly from Arabic, Persian, French, and English (see Yavaş 1982).

Since 1928, Turkish is written in a Latin alphabet using some diacritics, e.g. <ş ç ğ> for /ʃ tʃ ɣ/. Until then, a writing system which was based on the Arabic alphabet was in use. The Turkish alphabet consists of 29 graphemes, 21 for consonants and 8 for vowels and is conceived as a one letter-one sound-system. At first glance, Turkish seems to have an almost one-to-one mapping between phonemes and graphemes. However, there are quite a number of subtle phenomena not represented in orthography, mainly in loan words and in morphonology due to coarticulation, vowel harmony and resyllabification, inducing exceptional vowel lengthening and palatalization in the surface realizations of certain suffixes (Oflazar & Inkelas 2003).

10.12.2 Phonological and phonetic description

Turkish has an 8-vowel inventory consisting of /i, y e œ u ɯ o a / corresponding to <i ü e ö u ı o a> in Turkish orthography. Traditionally, the inventory has been described in terms of two phonologically distinct degrees of height or aperture, contrastive backness and contrastive rounding/labiality and can be grouped in sets of four vowels each: four front vs. four back vowels[127], four close (high) vs. four open (often termed ‘low’[128]) vowels, and four roundedvs. four non-rounded vowels. These divisions are useful to understand the application of Turkish vowel harmony.

	front		back	
	non-rounded	round/labial	non-rounded	round/labial
close	/i/ <i>	/y/ <ü>	/ɯ/ <ı>	/u/ <u>
open	/ɛ/ <e>	/œ/ <ö>	/ɑ/ <a>	/ɔ/ <o>

Figure 10.85: Vowel phonemes of Turkish

Labial activity in so-called “round” vowels is not homogenous but shows a range of coarticulatory variation. A look at MR images by M.A. Kiliç (in Arısoy et al. 2010: 7) for Turkish vowels shows that labial activity for Turkish /y/ and /u/ is very strong, while for /œ/ and /o/ less rounding and a higher degree of lip aperture can be observed. Özen (1985: 82ff) posits that the common terminology concerning “round” vs. “non-rounded” vowels is too undifferentiated to describe the articulatory realization of Turkish vowels and suggests the feature “labial” as more appropriate to characterize those vowels that are commonly described as “round”.

	labial	labio-dental	dental	alveolar	post-alveolar	palatal	velar	glottal
Plosives	p b		t d			c ɟ	k g	
Affricates					tʃ dʒ			
Nasals	m		n					
Fricatives		f v	s z		ʃ ʒ		ɣ	h
Tap				ɾ				
Approximant						j		
Lateral Approximants			ɫ		l			

Figure 10.86: The consonant inventory of Turkish (Zimmer & Orgun 1999: 154)

Syllable-final plosives are typically devoiced which can be reflected in Turkish orthography: *kitap* /ci-tap/ - ‘book’ (from Arabic *kitab*), *kitab+a* /ci-ta-ba/ - ‘book-dative’, *kitap-lar* /ci-tap-lar/ - ‘book-plural’. Final devoicing does not operate on voiced fricatives or sonorants. Suffix-

127 that are in “naïve” terms often described as “light” or “thin” (front vowels) and “dark” (back vowels).

128 So-called “low” vowels have to be further differentiated in mid and low vowels.

initial plosives assimilate in voice to the preceding segment *kitap-ta* (‘book-locative’) vs. *araba-da* (‘car-locative’).

The velar consonants /g/ and /k/ are reduced to a voiced velar fricative /ɣ/ in most root-suffix boundaries. Between front vowels, /ɣ/ is pronounced as a weak front-velar or palatal approximant. In word-final position or when followed by a consonant, /ɣ/ can be realized phonetically as a lengthening of the preceding vowel. Elsewhere in intervocalic position, /ɣ/ is often not realized phonetically (Zimmer & Orgun 1999: 155).

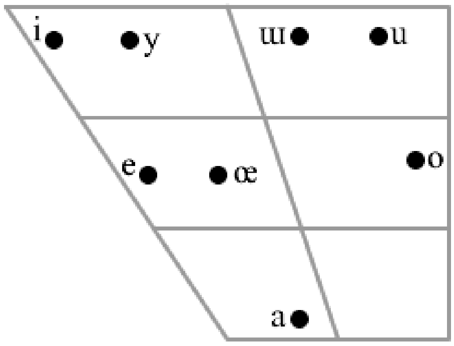

Figure 10.87: Schematic representation of the vowels of Turkish (Zimmer & Orgun 1999: 155)

All Turkish vowels are oral (see Figure 10.88 for a comparison of vowel formants presented in Türk et al. 2004). In a few lexical items a nasal triggers deletion with compensatory lengthening and vowel nasalization, e.g. *sonra* > [sõra] ‘later’ (Kornfilt 1997: 489).

Turkish vowels occur word-initially, word-finally and between consonants. All vowels are allowed in the first syllable of a lexical word. In words of native origin, the occurrence of /o/ and /œ/ is restricted to the first syllable (Göksel & Kerslake 2005: 10). In stem-final position, principally all vowels can occur although there is a restriction concerning roundednon-high vowels. The two non-high roundedvowels /o/ and /ø/ occur only in the first syllable of native words but may occur in other positions in borrowed stems. While there are numerous borrowed words ending in /o/, only very few are found with final /ø/, e.g. *tiyatro* ‘theater’, *koro* ‘choir’; *mösyö* ‘mister’, *banliyö* ‘suburb’ (Kornfilt 1997: 494).

In phrase-final open syllables, vowels except /a/ and /o/ have a ‘lower’ variant (Zimmer & Orgun 1999: 155), resulting in allophonic variants [ɪ ʏ ɛ ʊ] in phrase-final open syllables, e.g. *kel* [kel] ‘bald’ vs. *kale* [ka’lɛ] ‘castle’ (Göksel & Kerslake 2005).

/i/ is realized as [i] or [ɪ], [ɪ] usually occurs in word-final position and [i] elsewhere (Göksel & Kerslake 2005: 19).

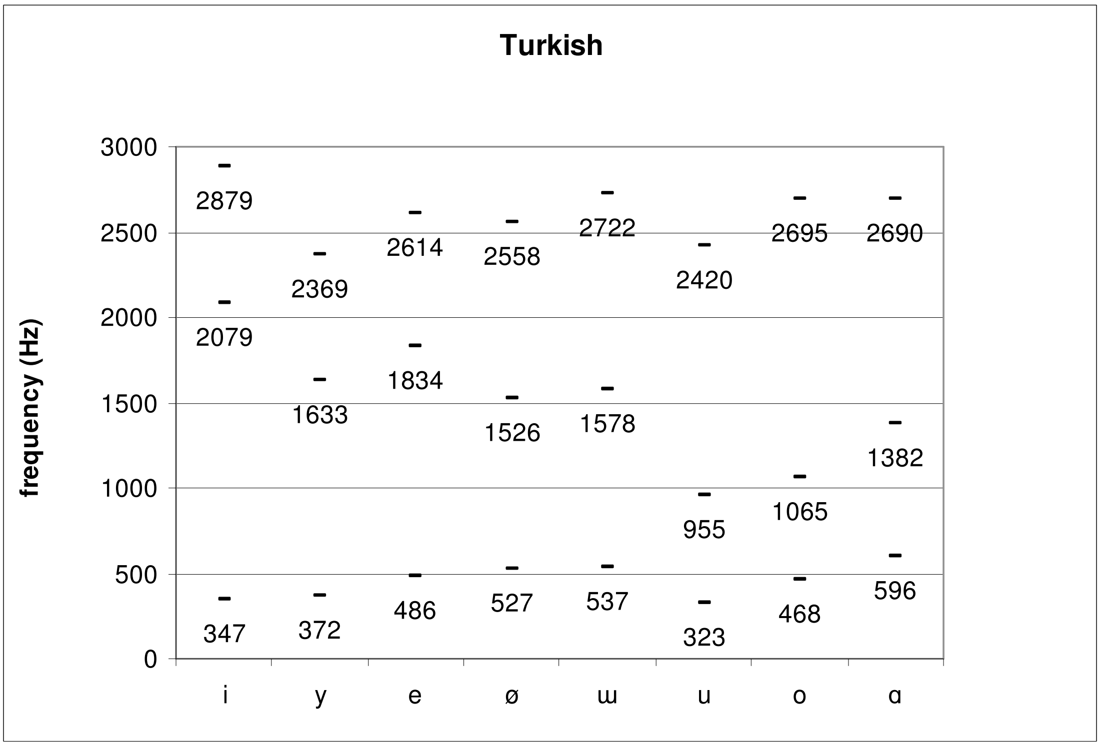

Figure 10.88: Formant frequencies for Turkish vowels measured by Türk et al. (2004). The values represent the mean formant frequencies for vowels in connected speech (words and sentences) read by 105 male speakers

Allophones for Turkish /e/ are [e], [ɛ] and [æ]. All three allophones occur in words such as *gezegende* [gezegændɛ] 'on the planet', *perende* [perændɛ] 'somersault' (Göksel & Kerslake 2005: 10). [ɛ] occurs in word-final position. [æ] occurs before /l m n r/ in instances where the sequences /*-er-, -em-, -en-, -el-*/ are not followed by a vowel, as in *her* 'each, all', *gerdi* 's/he stretched', *kent* 'town', *pergel* 'pair of compasses'. [e] occurs in all other positions.

In a limited number of words, /e/ may be pronounced either as [e] or [æ] before /l m n/, as e.g. in *elbise* 'dress', *kendi* 'self' or *hem* 'both' with [e], whereas in others /e/ is realized with [æ] (Göksel & Kerslake 2005: 10).

/ɑ/ is described as a low, back and unrounded vowel. For /ɑ/ two allophones are observed: a more back [ɑ] and a front unrounded variant in syllables containing /c ɟ/ and /l/ in loanwords (Göksel & Kerslake 2005: 10).

/o/ occurs only in the initial syllable in Turkish words, e.g. o*tur-* 'sit', but can appear in all positions in loanwords, e.g. *protokol* 'protocol'. Back /o/ is palatalized if adjacent to /l/ in loan words, e.g. *lokum* 'Turkish delight', *rol* 'role' (Göksel & Kerslake 2005: 11).

/u/ is palatalized when it occurs adjacent to the palatal consonant /c ɟ/ and /l/ in loan words, e.g. *blucin* 'denims', *lugat* 'dictionary', and is realized as [ʊ] in word final positions, e.g. *bu* 'this', kutu 'box' (Göksel & Kerslake 2005: 11).

/y/ occurs in two allophones, [y] and [ʏ]. [ʏ] occurs in word-final position, while [y] occurs elsewhere, e.g. *üzüntü* [yzyntʏ] 'sadness'.
/œ/ occurs only in the first syllable in native words, but may appear in any position in loanwords.
There is considerable disagreement about the status of the high unrounded vowel represented as <ı> in Turkish orthography. <ı> is the least stable and the shortest vowel of Turkish (Kiliç & Öğüt 2004: 151) and is difficult to acquire both for Turkish children in L1 acquisition as well as for L2 learners of Turkish. Some linguists claim that *ı* is a central vowel corresponding to IPA /ɨ/ (e.g. Selen 1979). This is also suggested in several phonological analyses (e.g. van der Hulst & van de Weijer 1991; Pöchtrager 2010), where the sound is transcribed as [ɨ], while in phonetic descriptions this sound is transcribed as a high back unrounded [ɯ] in IPA (Zimmer & Orgun 1999; Göksel & Kerslake 2005). The assumption that *ı* represents a central vowel has to be doubted if we look at the phonological behaviour of *ı* which strongly suggests that *ı* is a back vowel, since it harmonizes with other back vowels in the so-called backness harmony (see below).
Kiliç & Öğüt (2004) investigated the articulatory, acoustic and auditory characteristics of *ı*. Their articulatory analysis of *ı* spoken by 5 male Turkish native speakers, using midsagittal magnetic resonance images, showed no statistically significant differences between *ı* and *ε* and *ı* and *u*. Kiliç & Öğüt (2004: 152) interpret the configuration of the vocal tract area functions of their subjects suggesting that the constriction place of *ı* is between mid and back (see also MRI by Kiliç in Arısoy et al. 2010: 7). The acoustic analysis (mean F1 355 Hz, mean F2 1482 Hz) revealed no statistically significant differences between *ı* and ε, and between phoneticians' realizations of the IPA vowels [ɨ], [ɯ], and [ɤ]. For auditory analysis, an identification task using synthetic stimuli with 26 IPA vowels was carried out, in which tokens of [ɨ], [ɯ], and [ɤ] vowels were identified as *ı*. The auditory analysis showed that, auditorily, *ı* was more back than acoustically, and that this particular part of the auditory space was wide enough to involve [ɨ], [ɯ], and [ɤ] (Kiliç & Öğüt 2004: 152f). Kiliç & Öğüt (2004) conclude that the relative position of *ı* in the vowel space is in an area between [ε] and [u] and that in a broad phonetic transcription /ɯ/ is the appropriate IPA symbol. However, due to the high variability or instability of this vowel, the range of realizations may vary between [ɯ], [ɨ], [ɤ] or [ə].
The use of vowel length as distinctive feature in Modern Turkish is an issue of debate, though differences in vowel duration can be observed in several loanwords and in words that are affected by phonological processes such as deletion of *ğ* and *v* with compensatory vowel

lengthening. Long vowels occur in loanwords from Persian and Arabic and are of two separate sources of origin: Either they are originally long vowels or they occur in words that originally contained a glottal stop, e.g. *saat* 'time, clock' where the two a-vowels of the Arabic loan are originally heterosyllabic.

In spite of phonetic differences in vowel duration (long vs. short) and the existence of some minimal pairs which are differentiated only by vowel quantity (e.g. in words like *ali - âli, alem - âlem, hala 'father's sister' - halâ 'till now' - hâlâ 'still'*), vowel length is generally not considered to have a contrastive function in Turkish but is a feature of foreign origin (cf. Kramsky 1990: 305). Length is generally not marked orthographically with some exceptions in borrowed words, e.g. in some derivational forms like the Arabic adjectival suffix */–î/*, which is originally a long vowel, where length is indicated by a circumflex, e.g. *ilim* 'science' vs. *ilmî* [ilmi:] 'scientific'.

While long /a: e: i: u:/ have been introduced by Arabic and Persian loans into Turkish, e.g. *zaman* /za'ma:n/ 'time' there are also native words, where long vowels arise through the loss of a voiced velar fricative /ɣ/, the so-called 'soft g' represented by <ğ> in Turkish orthography, which is deleted in some dialects and preserved in others (Underhill 1986). In intervocalic position /ğ/ is deleted resulting in a bi-syllabic 2-vowel sequence that may merge into a long vowel in fast speech, e.g. *agaç* [aac] 'tree', *eger* [eer] 'if'. In syllable-final position, it is deleted causing compensatory lengthening of the preceding vowel: *dag* [da:] 'mountain', *tug* [tu:] 'banner', *igne* [i:ne] 'needle'.

Lexically underlying long vowels in open syllables in loanwords may be shortened in closed syllables resulting from suffixation (*/zama:na/* 'time-dative' vs. */zamanda/* 'time-locative'). Thus, long vowels in open syllables are in complementary distribution with short vowels in closed syllables with coda consonant.

In loanwords, long vowels are often retained in open syllables (e.g. Arabic *imza* [imza:] 'signature'), or they preserve length in the final syllable of the root when a vowel is added but are shortened otherwise (e.g. [za'man] 'time' vs. [zama:ni] 'time-3Sg. Poss.'). In other cases, vowels in open syllables are shortened (e.g. [frikik] 'free-kick', [tabu] 'taboo').

In non-formal speech styles, /h/, /j/ and /v/ may be deleted causing *compensatory lengthening* of the neighbouring segments. /h/-deletion occurs syllable-finally if a continuant or nasal follows (*kahya* > *ka:ya* 'steward', *mahsus* > *ma:sus* 'special to') and in syllable-initial position after a vowel or a voiceless consonant (*tohum* > *toum* 'seed', *ishal* > *isal* 'diarrhea').

Arısoy et al. (2004) investigated the durational properties of Turkish vowels. Vowels in 111 words realized by 6 female speakers were subject to phonetic analysis. For each vowel 5 to 7 occurrences in word-initial syllable and 5 to 7 occurrences in the final syllable were selected, each in words where they were preceded or followed by a voiceless stop. Since /o/ and /œ/ are restricted to first syllables in Turkish (see below), they were tested in non-sense words in final position. The results show that the position in the word is a determining factor in the duration of vowels. Vowels in the first syllables have significantly lower mean durations than vowels in final syllables. The duration range is greater in final syllables and mainly due to inter-speaker variation. Şayli (2002) showed that the mean duration of vowels decreases with the number of syllables in a word and the number of words in a sentence. The mean durations of vowels were greatest in word-final position and smallest in word-initial positions.

/j/ (<y>) is deleted after a sequence of a front vowel and a sonorant or after /i/, e.g. *öyle* > *ö:le* 'thus', *seyret* > *se:ret* 'watch', *iyi* > *ii* 'good', *deyil* > *deil* 'is not'. /v/ is deleted only after a labial vowel and before either a labial consonant or a vowel: *övmek* > *ö:mek* 'praise' (inf.), *över* > *öer* '(s)he praises'. In all these cases, deletion leads to vowel lengthening in syllable-final position, while no such effect is observed in an onset position (Sezer 1986; van der Hulst & van de Weijer 1991).

To summarize, Turkish has long vowels at the phonetic surface which have their origin in loans, allophony, diachronic change and phono-stylistic variation. Turkish exhibits processes of vowel shortening of long vowels and compensatory lengthening of short vowels. We can conclude that long vowels in Turkish are either underlying or derived by consonantal deletion and compensatory lengthening, or by vowel assimilation. The relative length of vowels depends on the syllable structure and its position in a word. While long vowels are preserved in open syllables, they are shortened in closed syllables.

The *syllable structure* in Turkish is less complex than in German. It generally allows open and closed syllables but no onset clusters. A subset of consonant clusters occurs in coda position. Across syllable boundaries a larger set of consonant clusters can occur. Word-final clusters in loanwords that do not correspond to the allowed structures may be reduced by deleting the final consonant, e.g. 'direct' – *direk*. Another way of treating such structures is the shift of syllable boundaries by attachment of a vowel, e.g. 'protest' – *[pu-ro-tes-to]*.

There are, however, syllable-initial consonant clusters in some words, mainly in loan words of Western origin, which are split in some cases, especially in casual speech styles, while preserved in more "sophisticated" pronunciation (Kornfilt 1997: 493): *kral* > *kıral [kuıral]* 'king', *grup* > *[gurup]* 'group', *klüp* > *[kulüp]* 'club', *prince* > *[pirens]* 'prince', *flöt* > *[fülöt]* 'flute'. The epenthetic vowel harmonizes in frontness with the following root vowel after

labial and dental consonants. If initial clusters are preserved, as in more careful styles of speech, the velar vowels are invariably back in quality (Clements & Sezer 1982: 248).
Other word-initial clusters, especially /sp-, st-, sk-/, are avoided by the prothesis of /i/ or /ɯ/ before the onset: *ıspanak* 'spinach', *ıstatistik* 'statistics', *iskeleton* 'skeleton'.
In casual speech styles, vowels in unstressed syllables are not stable and can be reduced or even deleted, e.g. *gidecek* [gidicek] 'he will go', *nerede* ['nerde] 'where?'. Syllable-reduction in Turkish is a controversial issue. Especially /i/ and /ɯ/-sounds are reduced in some consonantal and vocalic contexts, which may be perceived at the phonetic surface as resulting in CCV- or CCVC-syllables (Özen 1985: 43). The occurrence of /u/ or /y/ in the following syllable may reduce this reduction. This would explain why vowel epenthesis in loan words such as *klor* 'chlorine' or *gram* 'gram' is less likely than in words like *külüp* 'club' or *gurup* 'group', even if the structure of the cluster is the same in both cases (Özen 1985: 43).
In many originally monosyllabic Arabic loan words vowels can be deleted. An epenthetic vowel, which was inserted to split the original cluster, can be deleted if a suffix with vocalic onset is added, e.g. Arabic *ism* > Turkish *isim* 'name' > *ismim* 'my name'. This is coherent with the deletion of the second vowel in bisyllabic Turkish words when a suffix with vocalic onset is added and the second syllable becomes unaccented, e.g. *alın* 'forehead' > *alnım* 'my forehead'.
Sequences of monosyllabic vowels are generally very rare in Turkish, since the basic syllable structure CVC(C) disprefers such sequences. Though diphthongs do not occur as such in the Turkish system, vowel-glide sequences can be represented with <ay> , <oy> and <av>. When preceded by a vowel, e.g. /–av/, /v/ is frequently pronounced as a bilabial fricative or approximant.
In some loanwords from other languages sequences of vowels survive in both writing and pronunciation (Kornfilt 1997: 495): e.g. *dua* 'prayer', *boa* 'boa', *kuaför* (French) 'hairdresser', *studyo* (Italian) 'studio', *kamyon* (French) 'truck', *ayna* (Persian) 'mirror'. Moreover, the deletion of the intervocalic voiced velar fricative /ğ/ produces vowel sequences, e.g. *yoğurt* [jourt], *ağır* [aɯɾ̥]. Deletion of /h/, /v/, /j/ in syllable final position also leads to sequences of vowels in colloquial styles (e.g. *över* > *öer* '(s)he praises', *davul* > [daul̥] 'drum').
A characteristic property of Turkish is *vowel harmony*. In very general terms, vowel harmony rules require that all vowels within a given domain agree with respect to at least one property. Turkish vowel harmony operates from left to right. It refers to two features, frontness/backness and rounding. Turkish vowel harmony applies within morphemes as well

as across morpheme boundaries. Its function can be described as binding together interdependent units of meaning, signalizing the beginning and the end of a unit of meaning (Krámský 1990: 313). It does not apply between the components of compounds, in some loanwords and with some invariant affixes.

Word stems may contain any of the eight vowels in the first syllable. Any following vowel assimilates to the vowel in the root stem. Vowels in non-first syllables always have the same feature ([front] or [back]) as the vowel in the preceding syllables. Suffix vowels typically harmonize in *frontness/backness* with the preceding vowel. This type of harmony is commonly referred to as 'backness harmony' (e.g. Levi 2001) or '*palatal harmony*' (e.g. van der Hulst & van de Weijer 1991) and is found in almost all Turkic languages (Kaun 1995). Many roots are internally harmonic but many others are not; these include loanwords (e.g. *kitap* 'book', *otobüs* 'bus') as well as some native words (e.g. */anne/* 'mother'). With some loanwords, suffixes do not harmonize with the stem, e.g. *harf+ler* 'letter, grapheme'.

The vowels in suffixes always belong to the same natural class ([front] or [back]) as the preceding vowel. Suffixes generally appear in two forms, containing a front vowel /e/ or a back vowel /a/, e.g. *ev+ler* - 'house-plural', *at+lar* - 'horse-plural'. The last vowel before the suffix triggers the relevant suffix: e.g. *evde* 'house-Loc.', *Türkiyede* 'Turkey-Loc.', *büroda* 'office-Loc.', *sinemada* – 'cinema-Loc.'. Palatal harmony is especially relevant for declination.

The second type of vowel harmony in Turkish is '*rounding harmony*' or 'labial attraction' or '*big vowel harmony*', which is not as common in other Turkic languages (Kaun 1995) and makes reference to the dimensions of *height and rounding*. In Turkish, the feature [round] is spread from left to right from vowel to vowel but only if the target is [+high] (Kaun 1995: 2, 11). The suffixes occur in four different forms containing /i/ (if trigger is front non-rounded, e.g. *sekreter+im*, *ingiliz+im*) or /u/ (if back and round, e.g. *doktor+um*, *Rus+um*) or /y/ (if front round, e.g. *Türk+üm, şoför+üm*) or /ɯ/ (if back and non-rounded, e.g. *avukat+ım*, *fransiz+ım*). /o/ and /ö/ as [-high] vowels do not occur in the suffixes. This type of vowel harmony is especially relevant for possessive suffixes and the attachment of the interrogative particle *-mi, -mı, -mü, -mu*). After a suffix with only two vocalic modifications these four-fold suffixes are reduced to two, e.g. if the plural suffix –lar/ler is inserted between the stem and the possessive pronoun.

To summarize, the full vocalic repertoire is only represented in the first syllable of a word. In all other syllables, phonological oppositions are neutralized.

Kaun's (1995) comparative analysis of several Turkic languages shows that in many languages with labial harmony lower rounded vowels are avoided. In some languages sequences of roundedvowels of different height are generally avoided. In Turkish, non-high suffix vowels will not be roundedafter a roundedvowel, e.g. *gül* 'rose', *gülün* 'rose-Gen.', *güller* 'roses Pl'.

Stems which do not correspond to vowel harmony restrictions, mostly loan words, trigger vowel harmony in a specific manner: the exceptional vowel (or the last exceptional vowel, if more than one) determines the harmony of the following vowels: e.g. *dekor+u* 'stage design+Acc.', *buket+i* 'bouquet+Acc.', *otobüs+ü* 'bus+Acc' (Kornfelt 1997: 499).

The phonological constraints for sequences of roundedor non-rounded vowels in Turkish apparently have an effect on the consonants between roundedvowels. Labial coarticulation for sequences of roundedvowels and non-labial consonants has been observed in Turkish. Boyce (1990) examined the presence or absence of labial coarticulation with American English and Turkish speakers and observed language-specific differences in the way English and Turkish speakers organize their (co-)articulation at least concerning the use of lip protrusion for rounded segments. While English speakers show a reduction of lip protrusion of consonants in intervocalic position, in Turkish lips are protruded during consonant production over the all articulatory string of a consonant between two rounded vowels.

Coarticulatory effects in Turkish also account for the so-called *consonant harmony* which causes the palatalization of oral velar stops /g k/ and the lateral /ʟ/ in the environment of front vowels (with some exceptions) resulting in palatal allophones /ɟ c/ and /l/. The lateral /l/ is a velarized [ʟ] when it occurs with a tautosyllabic back vowel, e.g. *hala* [haʟa] 'father's sister'. With a tautosyllabic front vowel it is palatalized, e.g. *bile* [bil̡e] 'even'. However, in loan words, the palatal lateral can occur with tautosyllabic back vowels, e.g. *lâle* [l̡a:l̡e] 'tulip' (Kornfilt 1997: 486).

Moreover, we observe a harmony concerning the voicing of consonants referring to the tendency of stops within a word and separated by a morpheme boundary to belong to the same voice class (Krámský 1990: 316f): A voiceless consonant becomes voiced in intervocalic position through suffixation (e.g. *çorap* – *çorabi* 'sock(s)', *kanat* – *kanadi* 'wing(s)', *çoçuk* – *çoçuğa* 'child(ren)'), although there are many exceptions mainly in monosyllabic words and in words with final /t/.

Word stress which is expressed as pitch accent rather than dynamic stress (Underhill 1986) generally falls on the word-final syllable in suffixed word forms moving to the right-most syllable as suffixes are added. Differences in syllable structure play no role for regular stress assignment, its location is predictable. Turkish has therefore been described as a quantity-insensitive, fixed stress system (van der Hulst & van de Weijer 1991).

Exceptions where word stress appears on the non-final syllable fall into three regular classes (Underhill 1986): (1) Many lexical words, mostly loan words of foreign origin but also native words, have inherent stress on a non-final syllable, which does not move if suffixes are added, e.g. *lo'kanta* 'restaurant' – *lo'kantalar* 'restaurants'; (2) certain suffixes, e.g. the negative *–me*, do not bear stress, and stress falls on the syllable preceding the left-most unstressable syllable, e.g. *git'me* 'going' vs. *'gitme* 'don't go'; (3) certain lexical and syntactic classes such as place names, vocatives and some types of adverbs call for initial stress, e.g. *be'bek* 'baby' vs. *'Bebek* 'suburb of Istanbul', *ba'ba* 'father' vs. *'baba* 'Father!' (Underhill 1986; Zimmer & Orgun 1999: 155; Kabak & Vogel 2001).

The phonetic properties of stressed syllables in Turkish and the status of Turkish as a pitch-accent language have been discussed by many phonologists (e.g. Kabak & Vogel 2001 vs. Inkelas & Orgun 2003). Levi (2005) presents an analysis of the acoustic correlates of lexical accent in Turkish. Her results show significant differences in duration, peak amplitude and peak F0 between words with final vs. non-final lexical accent. Duration differences do not seem to be sufficiently perceptible to function as a reliable cue to accent placement while F0 peaks differ dramatically between lexically accented and unaccented syllables. A discriminant analysis confirms that F0 is the most robust cue to accent placement followed by intensity and then duration. In words where accent placement has been disputed, a marked drop in F0 reveals the difference between the unaccented syllable and the preceding accented syllable (Levi 2005).

10.12.3 Contrastive Analysis

The Turkish vowel system consisting of only eight different vowel qualities is considerably smaller than the German system. In German, vowels at a given constriction location are further differentiated by degree of constriction and phonemic length, whereas in Turkish both features are not used to distinguish vowels phonemically.

The pronunciation of all Turkish vowels is described in general as more "lax" or "centralized" (e.g. Kornfilt 1997: 490) or in other terms "less constricted" than their counterparts in German. Turkish /i e a/ are acoustically somewhat closer to German short unconstricted vowels but have no constricted counterpart. Özen (1985: 26f) observes moreover weaker lip-rounding for Turkish "round" vowels than for German vowels.

German distinguishes two high front unrounded vowels /i/ and /ɪ/, whereas Turkish has only one /i/, which is generally described to be more lax than the German long tense /i/. In word-final position, [ɪ] occurs as an allophone for /i/ in Turkish (Göksel & Kerslake 2005).
A similar pattern is observed for the German high front rounded vowels /y/ and /ʏ/ and Turkish /y/ and its allophone [ʏ].
For mid front vowels, we observe a similar pattern. While German distinguishes long tense /e/ and short lax /ɛ/, Turkish has only one mid vowel /e/ which is in close proximity to German /ɛ/. Allophones of Turkish /e/ are [ɛ] and [æ] (Göksel & Kerslake 2005).
German distinguishes mid front rounded /ø/ and /œ/, while in Turkish only one ö-quality is used. In native words, this vowel can occur only in word-initial position.
In the back area of the vowel space, four vowel qualities are used in Turkish: /ɯ u o ɑ/. While for /u o ɑ/ similar vowels in German are given, no equivalent for Turkish /ɯ/ is found in German. /o/ does not occur in non-initial syllables in Turkish native words.
The MR-images provided by Kiliç (in Arısoy et al. 2010: 7) show a clear difference in degree of lip aperture and tongue height for Turkish /i/ vs. /ɛ/ and /u/ vs. /ɔ/, while differences in tongue height are smaller for /y/ - /œ/ but differences in degree of lip aperture and lip protrusion are very clear for /u/ [-open] vs. /ɔ/ [+ open] and /y/ [-open] vs. /œ/ [+open]. Differences in degree of aperture are weaker though clearly evident for minimal pairs of non-rounded vowels, i.e. /ɯ/ - /ɑ/ and /i/ -/ɛ/. For non-rounded vowels, differences in tongue height are more evident in the MR-images[129].
A look at the major distinctive dimensions describing German vowel phonemes and their phonetic realization (see Figure 10.85) and a comparison with the German system reveals that both languages distinguish front vowels in two major dimensions: (1) front roundedö- and ü-vowels vs. non-rounded e- and i-vowels and (2) front high i- and ü-vowels vs. non-high e- and ö-vowels. It seems reasonable to describe Turkish vowel oppositions in terms of (1) frontness/backness, (2) lip rounding and (3) degree of aperture.
Turkish vowel harmony is a phonological process that determines which vowel may appear in non-initial syllables. *Fronting harmony* is an assimilation in terms of frontness/backness of a vowel with the vowel in the preceding syllable: front vowels are followed by front vowels, back vowels are followed by back vowels. The assimilation of a vowel with the vowel in the preceding syllable in terms of roundedness is called *rounding harmony*, affecting only suffixes and clitics containing high vowels (see Göksel & Kerslake 2005: 21ff). Figure 10.89

[129] the degree of jaw aperture is not visible in the images provided by Kiliç.

summarizes permissible vowel sequences in native Turkish words. However, there are several exceptions to these restrictions occurring in loan words from other languages and even in a few native words (see discussion in 10.12.2).

front vowels		back vowels	
i – e	e – e	ı – a	u – a
i – i	e – i	ı – ı	u – u
ü – e	ö – e	a – a	o – a
ü – ü	ö – ü	a – ı	o – u

Figure 10.89: Permissible vowel sequences in native Turkish words

It could be presumed that Turkish vowel harmony may have some influence on the perceived vowel quality in the test items. This influence could only occur in bisyllabic words, since Turkish allows all vowels to occur in monosyllabic words. In more-syllabic words, the vowel in the first syllable determines the following vowels. In the bisyllabic test items /pVtə/, the perception of German schwa in the unstressed final syllable of more-syllabic words (cf. also the test items /pVtə/) could be influenced by Turkish vowel harmony, resulting in an equivalence classification of German final unstressed /ə/ with either /ı/ or /e/130, whereas the perception of the vowel quality in the initial syllable is not expected to be affected by rules of vowel harmony. Turkish native speakers have moreover sufficient experience with exceptions to harmony in non-native words borrowed from other languages. It is therefore expected here, that in the experimental setting of the present study Turkish vowel harmony will not influence the listeners' categorizations.

Considerable differences are observed with respect to word stress and its phonetic realization in Turkish and German. Most Turkish roots are stressable on the final syllable. If a stressable suffix is added to a final-stressed root, then the position of the word stress moves to the final suffix (Göksel & Kerslake 2005: 29). In Turkish, vowels in final syllables are significantly longer compared to vowels in the first syllable of a multisyllabic word (Arısoy et al. 2004) and word stress is realized as pitch accent rather than dynamic stress. A possible influence of these differences in word stress patterns on the listeners' L2 performance may be expected in production concerning the general timing of phrases in L2 speech, but is less likely to influence perception.

[130] see also example (1) below.

Texts written by L2 learners of German may give us some hints to the fact that there are however, at least for some categories, similarity relationships in the learners' interlanguage system that cause category confusion even for more experienced learners. The examples (1) and (2) are single words or texts (in italics) written by Turkish learners of German who did not participate in the study. The examples show considerable problems with the orthographic representation of some German vowels and diphthongs.

The author of example (1) has been literalized in Turkish and learned German as an adult. She uses several orthographic strings that are directly transferred from Turkish (e.g. <ay> for /ai/, <av> for /au/). Interestingly, the unstressed schwa in word-final position is represented with Turkish <ı>: *keine* <kaynı> – 'no(ne)'. /ɛ/ is not clearly differentiated from i-vowels as is indicated by <f*i*rtik> *fertig* 'ready' and <ös*i*ray> *Österreich(er)* 'Austrian'.

(1) adult learner of German (L1 Turkish)

Alle kinder östray gebren ösiray
Ick nik arbayet haus firaV
Alles kinder sulhe 3 kinder mitte sulle eine kinde harteL dize yare firtik
kaynı telefon

alle Kinder Österreicher geboren in Österreich
ich nicht arbeite Hausfrau
alle Kinder Schule drei Kinder Mittelschule ein Kind HTL dieses Jahr fertig
kein Telefon

(2) pupil of the 3rd grade primary school (L1 Turkish)

Der Papa sucht Nikolausmütze Er sucht im Kinderzmer dan hort gerosche dan feschtet im die Kasten raein dan komt iere son komtr im Kinderzmer dan romt ire Splseuge Karta raein dan schaut was Nikolaus was bint zu im Splseuge aber kaeinen Splseuge gebint dan schut ire Pa... im iere Papa sakt dan mach iere son Kasten zu. da was nich ire zon im Kasten ire Papa sit dirn ...

Der Papa sucht die Nikolausmütze. Er sucht im Kinderzimmer. Dann hört er Geräusche und versteckt sich im Kasten. Sein Sohn kommt ins Kinderzimmer und räumt sein Spielzeug in einen Karton hinein. Er schaut, ob ihm der Nikolaus Spielzeug gebracht hat, aber er hat ihm keines gebracht. Papa schaut aus dem Kasten heraus zu. Der Sohn weiß nicht, dass Papa im Kasten sitzt und sperrt den Kasten zu.

Example (2) is written by a boy who attended primary school in Austria from the first grade on and wrote the text in the third grade. The text shows considerable difficulties on the phonemic as well as the syllabic level that will not be discussed in detail here (see Kerschhofer-Puhalo 2010b). With respect to category formation for German vowels category confusion is observed for o- and ö-vowels <h*o*rt> *hört* 'hears' and for the diphthongs /au/ and

/oi/ and their orthographic representation: <schut> *schaut* 'looks', <gerosche> *Geräusche* 'noise', <romt> *räumt* 'put away, clear'.

Oturan (2002) investigated the perception of Turkish and German vowels by "naïve" Turkish natives in an experiment using the perceptual assimilation paradigm. 31 native speakers of Turkish with no experience with German were asked to listen to /bV/-syllables cut from German words containing /i:/, /ɪ/, /e:/, /ɛ/, /ɑ/ and /ɑ:/ and to label them with one of the Turkish vowel phonemes. The participants were moreover asked to rate the relative similarity between German vowels and the selected Turkish category along a rating scale from "1 = identical", "2 = very similar", "3 = quite similar", "4 = different", "5 = very different", "6 = other vowel". Only i-, e- and a-vowels were tested in this study.

The results show that only German /ɛ/ was mapped onto Turkish /e/ (2.03 average goodness rating, i.e. "very similar"), whereas the three other front vowels /i:/, /ɪ/ and /e:/ were mapped onto Turkish /i:/, though the average goodness ratings varied for these three categories: /ɪ/ was found to be "very similar" to Turkish /i/ (1.97), while long constricted pre-palatal /i:/ (3.0), and mid-palatal /e:/ (2.87) were found to be "quite similar" to Turkish /i/. German /a/ (2.15 goodness rating) was perceived to be more similar to Turkish /a/ than German long /a:/ (2.80). These results indicate that there is a closer similarity relationship between Turkish /i/ and German /ɪ/ than between Turkish /i/ and long constricted German /i:/ or /e:/. However, these results are obtained by listeners without experience in L2 German and do not provide evidence for the perceived similarity for more advanced learners of German since Oturan's study focussed on the perceptual assimilation of L2 vowels to L1 categories by naïve listeners. Oturan's study does not inform us about the perceived similarity of vowel categories *within* the L2 system of more *advanced* Turkish learners of German, an aspect that is focussed on in the present investigation.

Van Heuven (1985) reports on experiments describing pronunciation errors in Dutch by Turkish natives and their perception by native Dutch listeners to describe the nature and perceptual consequences of wrong vowel pronunciation. The results of this study are reported here, since the basic phonological features for Dutch and German are similar in many respects. Realizations of the 15 Dutch vowels produced by 3 Turkish and 3 Dutch speakers were presented to Dutch native listeners for identification. The results show that Turkish speakers have no problem with the distinction of front vs. back vowels in production. While the front-back dimension is correctly preserved in the Turkish realizations and the Dutch listeners' identifications, vowel height is often incorrectly identified. With respect to vowel *height*, Turkish speakers seem to have difficulties to realize the appropriate Dutch distinctions. Moreover, since Turkish speakers do not reproduce the quality differences of Dutch short-long oppositions appropriately, the perceptual contrast between these pairs of vowels is shifted entirely onto the duration parameter. However, the duration difference between long and short vowels is not properly realized either. This holds especially for the oppositions of Dutch /i/ - /ɪ/, /ɪ/ - /e/, and /ɔ/ - /a/. Van Heuven describes the vowel realizations of Turkish natives to be intermediate to these qualities. He describes the natural Turkish vowel duration to be closer to the Dutch short vowels than to Dutch long vowels and describes the difference between short vs. long realizations of the Turkish speakers as much smaller than for the Dutch speakers with substantial overlap of the two categories in the data of Turkish speakers. In a perceptual labelling task with synthetically modified stimuli, Turkish native speakers showed that for front vowels only two vowel heights could be distinguished (corresponding to Dutch ɪ/i: vs. e/e:). While Dutch natives distinguished three vowel heights and used duration only to differentiate between short /ɪ/ and long /e:/, Turkish listeners used duration to discriminate between the members of each height class. For the contrast between Dutch short back /ɑ/ and long front /a:/, Turkish listeners seem to rely exclusively on the durational cue while disregarding the spectral parameter, which results in a wrong labelling in around half the cases, a fact that fits with the production data for the /ɑ/ vs. /a:/ contrast which differs more in duration than in quality. These data suggest that difficulties with the correct perceptual identification of vowels produced at a given constriction location will also be difficult in German with respect to aspects of quality as well as quantity. The use of only two vowel heights by Turkish natives, however, as reported by Van Heuven (1985) is assumed here to be better described in terms of a distinction [+/– open] as discussed above.

Darcy & Krüger (2012) investigated German vowel perception and production with Turkish children who acquired L2 German early in life. Fourteen ten-year old early sequential bilingual Turkish-German children[131] who were first exposed to German at an age between two and four years were tested in an oddity discrimination task and a production task. The control group consisted of fourteen monolingual children. The test items for the ABX oddity categorization task consisted of the contrast pairs /a:/ ~ /i:/, /i:/ ~ /ɪ/, /e:/ ~ /ɛ/, /i:/ ~ /e:/ recorded from three speakers. Early bilinguals were found to detect /i:/ ~ /ɪ/ and /i:/ ~ /e:/ significantly less accurately than monolinguals, while the results of both groups are equally good for the contrast pairs /a:/ ~ /i:/ and /e:/ ~ /ɛ/.

[131] The bilingual children were exposed to L2 since Kindergarten at an age of 2 - 4 years and are schooled in a bilingual elementary school. They are reported to speak only Turkish at home. The children can thus be described as early successive bilinguals. No information is given on the dominance of one or the other language.

In the production task, a picture naming task embedded in a "memory" game to elicit near-minimal pairs for the contrasts /i:/ ~ /ɪ/, /e:/ ~ /ɛ/ and /a:/ ~ /a/, monolinguals and bilinguals were globally indistinguishable, i.e. non significant differences between groups were found. Spectral and durational differences to realize vowel contrasts were implemented adequately by both groups. However, subtle differences in phonetic realization between groups were found: In the bilingual group, /e:/ was found to be slightly closer to /i:/ ($p < .006$), /a:/ was found to be more front but not higher ($p < .001$; $p = .2$) and /a/ was more front and higher ($p < .005$; $p < .003$). Significant differences in duration were found for all short-long vowel pairs. The short-long ratio was found to be shorter for bilinguals ($p < .05$). The study shows that category "merger" for some L2 contrasts can even be observed with bilingual children who were exposed to L2 very early in life. Early bilinguals were found to realize phonemic distinctions adequately in production but could not necessarily discriminate them successfully in perception, especially /i:/ ~ /ɪ/ and /i:/ ~ /e:/.

To conclude, evidence from previous studies on the perception of German vowels and from written texts suggests that difficulties with respect to German short unconstricted /ɛ/, /ɪ/ but also ö-qualities as well as difficulties with German diphthongs may be expected. Difficulties with phonemic categories of the German vowel system are expected with respect to the following features: (1) the use of duration together with degree of constriction to express phonemic contrasts, e.g. for /ɪ/ - /i:/, and (2) differences in distinctions referring to constriction location, degree of jaw aperture and vowel height, e.g. /i:/ - /e:/ or /e:/ - /ɪ/.

To summarize the contrastive analysis of German and Turkish, the use of [degree of aperture] which is at least as relevant as [height], is expected to play a significant role in the perceptual categorization of German vowel phonemes. This aspect is especially relevant for the distinction of German /i ɪ e: ɛ/, /y: ʏ ø: œ/ and /u: ʊ o: ɔ/. German /i:/ vs. /ɛ/, /y:/ vs. /œ/ and /u:/ vs. /ɔ/ are expected to be clearly differentiated from each other, since these vowel contrasts are distinct in terms of more than one feature and are moreover considered to be very similar or even "equivalent" to Turkish contrast pairs, whereas the differentiation of German /e:/ from i-vowels, /ø:/ from ü-vowels, and /o:/ from u-vowels is expected to cause more problems to Turkish learners of German due to smaller differences in degree of aperture. Moreover, the differentiation of /ɪ ʏ ʊ/ from other qualities at the same constriction location is expected to be difficult mainly due to their short intrinsic duration.

10.12.4 Results and interpretation

24 Turkish subjects were tested, 12 beginners, 10 advanced and 2 very advanced learners. All participants were tested in Vienna.

The data set discussed here consists of 6240 valid responses (238 missing answers). For each vowel category, 431 to 432 valid responses were delivered.

10.12.4.1 Difficulties

The diagram in Figure 10.90 shows the relative difficulty of German vowel categories for L1 Turkish listeners.

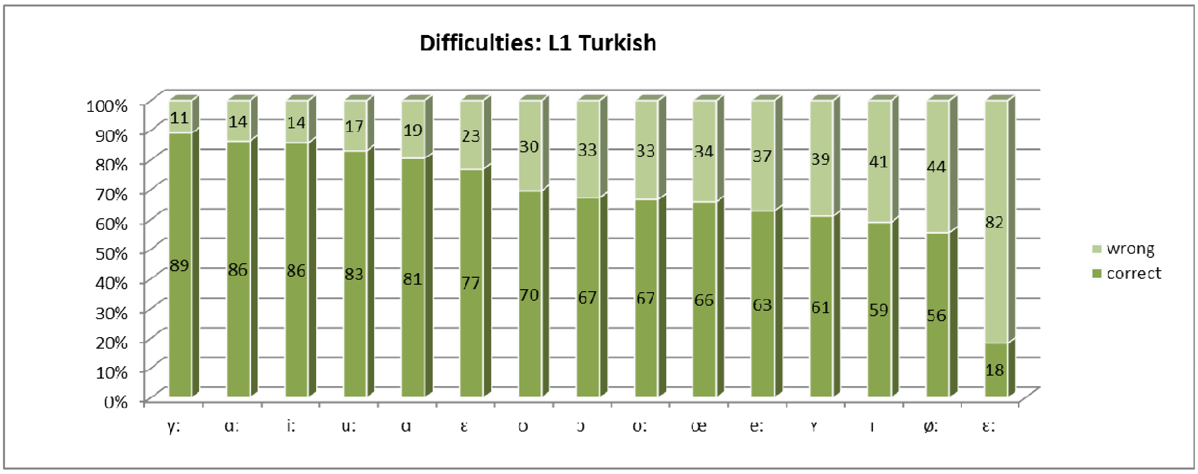

Figure 10.90: Percentage of correct and wrong identifications for Turkish learners (N=24, 6240 responses)

The most difficult German category is /ɛː/ with a very low id_correct score of 18%. The category that is most frequently identified correctly is /yː/ followed by /ɑː iː uː/, all with id_correct scores above 80%, and by /ɑ/ (81%) and /ɛ/ (77%). Within the class of front rounded vowels, /yː/ is identified correctly in 89% of its occurrences, while all other front rounded vowels show considerably lower id_correct scores: /œ/ 66%, /ʏ/ 61% and /øː/ only 56%. For /oː/ (66%) and /eː/ (63%) more difficulties are observed than with /ɑː/ (86%), /iː/ (86%) and /uː/ (83%).

The boxplot diagram in Figure 10.91 shows least interpersonal variation with respect to id_correct scores for /yː iː ɑː uː/. None of the subjects identified all instances of /ɛː/, /ʏ/ and /ɔ/ correctly, indicating persistent difficulties with these categories even for very advanced learners of German, while all other categories were identified correctly by at least one subject of the Turkish sample. Most interpersonal variation is observed for /øː/ and /œ/ as well as for /ʊ/ and /ɪ/.

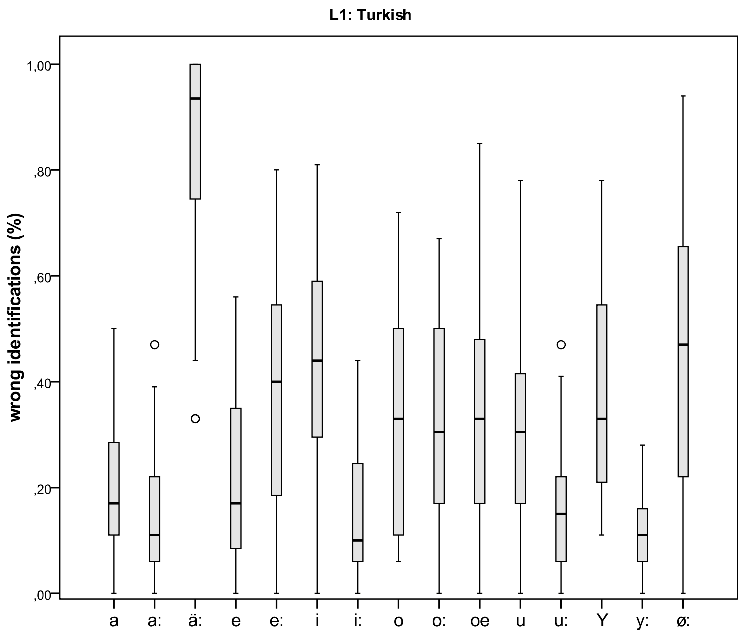

Figure 10.91: Boxplot diagram with id_wrong scores L1 Turkish listeners

The Turkish listeners' mean id_wrong score was 32% (12% standard deviation), the median was 32%. The listener with the highest percentage of correct answers had an id_wrong of 11%, the highest id_wrong score was 51% wrong responses.

10.12.4.2 Preferences

The preference scores (Figure 10.92) show strong asymmetries concerning the preference for specific response categories. The preference scores for monophthong categories vary within a range of 7.9% (min /ɛ:/ 2.5%, max /e:/ 10.4%). The most preferred categories are /e:/ (selected in 10.4% of the cases) and /y:/ (9.8%). /ɛ:/ (2.5%) is the least preferred response category, followed by the short unconstricted qualities /ɪ/ (4.7%), /œ/ (4.9%), /ʏ/ (4.9%) and /ɔ/ (5.4%). Long vowel categories (with the only exception of /ɛ:/ with a pref_score of only 2.5%) are clearly more preferred as response option (pref_scores ranging from 10.4 to 6.2 percent) than short categories that show scores between 4.7 and 6.2 percent (with the only exception of /ɛ/ 6.9%). In other words, German long constricted vowel qualities except /ɛ:/ seem to work as perceptual attractors for short unconstricted qualities.

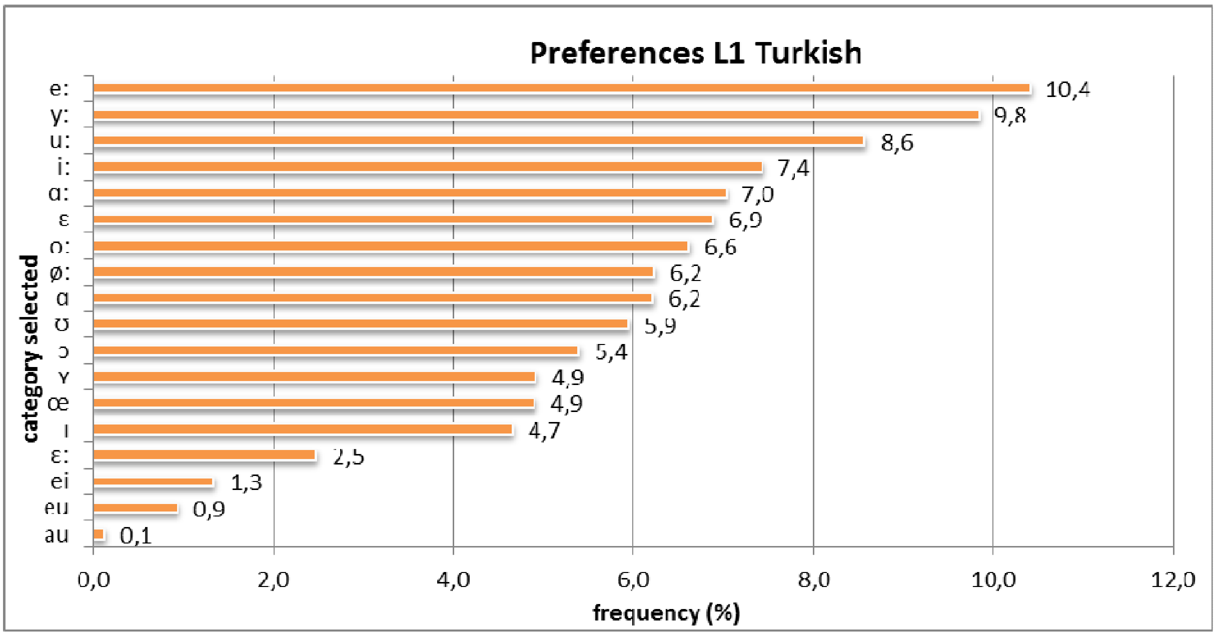

Figure 10.92: Preference scores for L1 Turkish listeners

This pattern may be determined by the distribution of [+/–constricted] allophonic variants in Turkish and by stress patterns in Turkish words. In Turkish, the position in the word is a determining factor for the quality and duration of vowels. The unconstricted allophones /ɪ ʏ ɛ ʊ/ can occur for /i y e u/ in Turkish, but only in final syllables. Most Turkish roots are stressable on the final syllable. Word stress is realized as pitch accent rather than dynamic stress. In Turkish, vowels in final syllables are significantly longer compared to vowels in the first syllable of a multisyllabic word (Arısoy et al. 2004). The test items in the present study consist of mono- and bisyllabic logatomes. The observed preference patterns for long constricted categories could be due to the fact that listeners interpret the short vowel in stressed syllable in bi-syllabic items as [long] since the duration of these vowel stimuli in non-final stressed syllables exceeds the expectation of Turkish listeners with respect to length. The observed preference patterns may be influenced by a hypercorrect reaction to differences in word stress and allophonic variants in Turkish and German. However, the interference of L1 patterns for allophonic variation, stress placement and syllable position cannot be sufficiently described by the present data but is an issue for further research.

Diphthongs are selected as response option by several subjects. While <au> is the least attractive category which is selected only in 0.1% of the cases, <ei> (1.3%) and <eu> (0.9%) are selected more frequently (see the discussion below).

10.12.4.3 Patterns of confusion

Patterns of category confusion for Turkish learners of German are presented in Table 10.73. The confusion matrix shows four areas of category confusion and a tendency to select diphthongs as response category, especially for non-high front unrounded qualities. While difficulties with a-vowels only arise with respect to phonemic length, other areas of category confusion reflect difficulties with respect to quality and quantity.

%	ɑ	ɑ:	ɛ:	ɛ	e:	ɪ	i:	ɔ	o:	ʊ	u:	œ	ø:	ʏ	y:	‹ei›	‹eu›	‹au›	wrong	**corr**
a	**81**	19																	19	**81**
a:	12	**86**																	14	**86**
ɛ:			**18**	6	69											7			82	**18**
ɛ			6	**77**	15											1			23	**77**
e:			10	4	**63**	2	11									10			37	**63**
ɪ			2	17	6	**59**	15									1			41	**59**
i:				1	4	9	**86**									1			14	**86**
ɔ								**67**	27								4		33	**67**
o:								6	**67**	2	20						4		33	**67**
ʊ								7	3	**70**	19						1		30	**70**
u:									1	15	**83**								17	**83**
œ												**66**	31				2		34	**66**
ø:											3	1	**56**	5	32		2		44	**56**
ʏ										1	1	6	5	**61**	26				39	**61**
y:											1		2	7	**89**				11	**89**
	6	7	2	7	10	5	7	5	7	6	9	5	6	5	10	1.3	.9	.1	31.5	***68.5***

Table 10.73: Confusion matrix L1 Turkish. Correct (in bold) and wrong identifications in % for L1 Turkish listeners (N=24, 6240 responses) (presented stimuli in rows, selected categories in columns in semi-orthographic representation, values < .5 not indicated, values <5.5 in grey)

/ɑ/ is identified correctly in 81% of its occurrences, in 19% it is identified as long /ɑ:/.

/ɑ:/ is identified correctly in 86% of the cases and identified as short /ɑ/ in 12% of its occurrences.

Short unconstricted /ɛ/ is identified correctly in 77% of the cases and incorrectly identified as /e:/ (15%), <ä:> (6%) or <ei> (1%).

Long constricted /e:/ seems to cause more difficulties to Turkish learners of German. It is correctly identified as /e:/ in 63% of the cases. Wrong responses spread over several categories: /i:/ (11%), /ɛ:/ (10%), <ei> (10%), /ɛ/ (4%) and /ɪ/ (2%).

/ɛ:/ is the most difficult category for Turkish learners of German (only 18% correct answers). The most preferred response category is /e:/ (69%), followed by <ei> (7%) and /ɛ/ (6%).

<ei> is selected as response option only for e- and i-vowels, and within this group mostly for the long categories /e:/ (10%) and /ɛ:/ (7%). In other words, vowels that exceed the expectations of Turkish listeners with respect to a certain duration in non-final syllables may be reinterpreted as diphthongs. This pattern is stronger with <ei> and front non-rounded e-

and i-vowels than with <eu> and <au> and roundedvowels. However, it has to be considered that the test design in a certain sense could encourage the participants to select diphthongs as response categories even if the test material does not contain any diphthongal stimuli.

German /i:/ is identified correctly in 86% of the cases, while the id_correct score for /ɪ/ is only 59%. /i:/ is substituted perceptually with /ɪ/ in 9% of the cases, with /e:/ in 4% of its occurrences and only occasionally with /ɛ/ (1%) and <ei> (1%).

Responses for short unconstricted /ɪ/ (59% correct) spread over several categories: /ɛ/ (17%), /i:/ (15%), /e:/ (6%), /ɛ:/ (2%), and <ei> (1%) indicating strong uncertainties with respect to the exact quality of this category.

/ɔ/ is identified correctly in 67% of the cases. In 27% of its occurrences it is identified as long constricted /o:/ and in 4% of the cases as <eu>.

/o:/ has an id_correct score of 67% and is mainly substituted by /u:/ (20%), and less frequently by /ɔ/ (6%), <eu> (4%), and /ʊ/ (2%).

Interestingly, the confusion of o-vowels and <eu> is also observed in the text in example (2) in <gerosche> *Geräusche* 'noise', <romt> *räumt* 'puts away'. The phonetic realization of German <eu> seems generally only vaguely represented for many Turkish learners of German, and these difficulties are possibly enhanced by the existence of <oy> as orthographic representation for a Turkish vowel-glide sequence /oi/. Therefore, problems with the phoneme-grapheme correspondence for German /oi/ - <eu> rather than perceptual difficulties with the contrast /ɔ/ - /oi/ are presumed here.

Long constricted /u:/ is identified correctly in 83% of its occurrences and incorrectly categorized as short unconstricted /ʊ/ in 15% of the cases and only occasionally as /o:/ (1%).

The responses for short unconstricted /ʊ/ (70% correct) spread over more categories: /u:/ (19%), followed by /ɔ/ (7%) and /o:/ (3%).

For /o:/ and /u:/ a strong asymmetric pattern is observed. /o:/ is substituted with /u:/ in 20% of its occurrences, whereas /u:/ is substituted only in 1% with /o:/. An inverse pattern is observed for short constricted /ʊ/ and /ɔ/: /ɔ/ is never substituted by /ʊ/, but /ʊ/ is substituted in 7% of the cases by /ɔ/.

German /y:/ is identified correctly in 89% of its occurrences and is substituted by /ʏ/ in 7% of the cases and only occasionally categorized as /ø:/ (2%) and /u:/ (1%).

The responses for short unconstricted /ʏ/ (61% correct) are more wide-spread: /ʏ/ is perceptually substituted by /y:/ in 26% of its occurrences, whereas /y:/ is significantly less often substituted by /ʏ/ (only 7%). This pattern of asymmetry suggests that /y:/ is the more stable category. /ʏ/ is moreover categorized as /œ/ (6%), /øː/ (5%), and occasionally as /ʊ/ or /uː/ (1% each).

German /œ/ is incorrectly categorized as /øː/ in 31% of the cases and occasionally as <eu> (2%). No other substitutes are observed, indicating a clear mental representation for this category.

Responses for long constricted /øː/ (56% correct) spread over several categories: /yː/ (32%), /ʏ/ (5%), /ʊ/ (3%), <eu> (2%) and /œ/ (only 1%).

Interestingly, the id_correct score for mid-palatal /œ/ (66% correct) is higher than for /øː/ (56% correct), a pattern that is found only with Hungarian and Turkish L2 learners and with German natives. All other language sub-groups show more perceptual confusion for short unconstricted /œ/ than for constricted /øː/.

For ö- and ü-vowels, a similar asymmetry like for u- and o-vowels and for e-and i-vowels is observed: /øː/ is substituted with /yː/ in 32%, whereas /yː/ > /øː/ occurs only in 2%. The opposite pattern is observed for /œ/ - /ʏ/, where /ʏ/ > /œ/ occurs in 6% but substitutions of /œ/ > /ʏ/ are not observed.

With ö-vowels, we observe another interesting pattern of asymmetry: While /œ/-stimuli are identified in 31% of the cases incorrectly as /øː/, /øː/ is substituted only in 1% of the cases as /œ/ and is rather categorized as /yː/ (32%). This pattern can be explained by two facts: (1) the Turkish ö-vowel is rather unconstricted and more similar to German /œ/ than German /øː/, and (2) the contrast of Turkish /y/ - /œ/ does not seem to be primarily based on the place of constriction as in German, where mid-palatal ö-vowels contrast with pre-palatal ü-vowels, but rather on the degree of aperture and degree of constriction. The MR-images presented by Kiliç (in Arısoy et al. 2010: 7) show that the degree of lip aperture and the degree of constriction in the oral cave is considerably higher for /y/ than for /œ/ in Turkish, which is also indicated by the high F1 value for <ö> in Figure 10.88.

A high percentage of difficulties in correct categorization is observed for short unconstricted /ɪ/, /ʏ/ and /ʊ/. Especially the responses for /ɪ/ and /ʏ/ spread over several categories indicating persistent problems also for more advanced learners, which are due to difficulties to differentiate German vowel qualities at the mid- vs. pre-palatal constriction location as well as to the generally shorter duration of these vowels (section 5.4.5.1 and 5.4.5.3).

10.12.4.4 Similarity and distance

A comparison of the similarity scores (see Table 10.74) shows that there are three major areas of category confusion for (1) front unrounded i- and e-vowels, (2) back rounded o- and u-vowels, and (3) front rounded o- and u-vowels, especially when long constricted categories are involved.

The highest sim_scores are given for the contrast /e:/ - /ɛ:/ (0.971). All other sim_scores are significantly lower. Values above 0.2 are given for /œ/ - /ø:/ (0.270), /ɔ/ - /o:/ (0.249), /ø:/ - /y:/ (0.234), /ʏ/ - /y:/ (0.225), and /ʊ/ - /u:/ (0.223). Of course, these sim_scores ignore the asymmetries described above.

V	ɑ	ɑ:	ɛ:	ɛ	e:	ɪ	i:	ɔ	o:	ʊ	u:	œ	ø:	ʏ	y:
a	1.000														
ɑ:	.188	1.000													
ɛ:	.000	.002	1.000												
ɛ	.003	.000	.124	1.000											
e:	.000	.002	.971	.136	1.000										
ɪ	.000	.000	.031	.121	.061	1.000									
i:	.000	.000	.000	.004	.095	.160	1.000								
ɔ	.000	.002	.003	.000	.000	.000	.002	1.000							
o:	.002	.003	.000	.002	.002	.000	.002	.249	1.000						
ʊ	.002	.000	.000	.000	.000	.000	.000	.052	.037	1.000					
u:	.000	.000	.000	.000	.000	.000	.000	.000	.142	.223	1.000				
œ	.000	.000	.000	.002	.000	.002	.000	.004	.000	.000	.000	1.000			
ø:	.000	.000	.000	.000	.000	.000	.000	.000	.000	.004	.026	.270	1.000		
ʏ	.000	.000	.000	.000	.000	.002	.000	.000	.000	.005	.005	.049	.082	1.000	
y:	.000	.001	.000	.001	.002	.000	.001	.000	.000	.000	.008	.000	.234	.225	1.000

Table 10.74: Similarity matrix for L1 Turkish listeners

Table 10.75 and Table 10.76 summarize the sim_scores for back contrast pairs, front contrast pairs, and front vs. back contrast pairs separately. Sim_scores for contrasts involving front vowels (except /e:/ - /ɛ:/) range from 0.27 to 0.001. Sim_scores for back contrast pairs range between 0.249 and 0.002. Contrast pairs involving front vs. back vowels are considerably lower between 0.026 and 0.001. In other words, German vowel phonemes are perceptually clearly distinguished by Turkish listeners regarding the front/back dimension and especially with respect to the round/non-rounded dimension, as becomes also evident in the two-dimensional MDS solution in Figure 10.94.

back Vs	*sim_scores*
ɔ - o:	.249
ʊ - u:	.223
ɑ - ɑ:	.188
u: - o:	.142
ʊ - ɔ	.052
ʊ - o:	.037

Table 10.75: Similarity scores > .03 for L1 Turkish listeners for back vowel contrasts

The sim_scores for *back* vowels reveal the highest perceptual similarity for contrasts in phonemic length and constriction degree: /ɔ/ - /o:/ (0.249), /ʊ/ - /u:/ (0.223), /ɑ/ - /ɑ:/ (0.188). Slightly lower values are given for contrast pairs whose members differ with respect to the velar vs. uvular constriction location, i.e. for contrasts of o- vs. u-vowels. Lower values are given for contrasts between pharyngeal a-vowels and uvular of velar vowels.

front Vs	*sim_scores*
e: - ɛ:	.971
œ - ø:	.270
ø: - y:	.234
ʏ - y:	.225
i: - ɪ	.160
e: - ɛ	.136
ɛ - ɛ:	.124
ɪ - ɛ	.121
i: - e:	.095
ʏ - ø:	.082
ɪ – e:	.061
œ - ʏ	.049
ɪ - ɛ:	.031

Table 10.76: Similarity scores > .03 for L1 Turkish listeners for front vowel contrasts

A look at the sim_scores for *front* vowel contrast pairs shows that contrasts involving /ø:/ or /y:/ and other palatal rounded vowels receive quite high scores.

Sim_scores for contrast pairs involving a *front/back* contrast are very low, i.e. below 0.01, with the only exception of /u:/ - /ø:/, indicating that the front/back dimension is quite distinctive for Turkish listeners but that lip rounding may decrease the perceptual distinctivity of front rounded vs. back rounded vowels in very few cases, which are not represented in the confusion matrix in Table 10.73 since they are below the 0.5% mark.

10.12.4.5 The perceptual vowel map for L1 Turkish listeners

From the similarity scores calculated by Shepard's formula, distance scores and non-metric MDS solutions were derived. For the non-metric MDS solutions, RSQ values of 0.405 in dimension 1, 0.693 for the two-dimensional solution, and 0.801 for the three-dimensional solution are obtained.

The one-dimensional MDS solution (RSQ .405) shows a clear-cut differentiation of roundedvs. non-rounded vowels and considerable overlap within the rounded class.

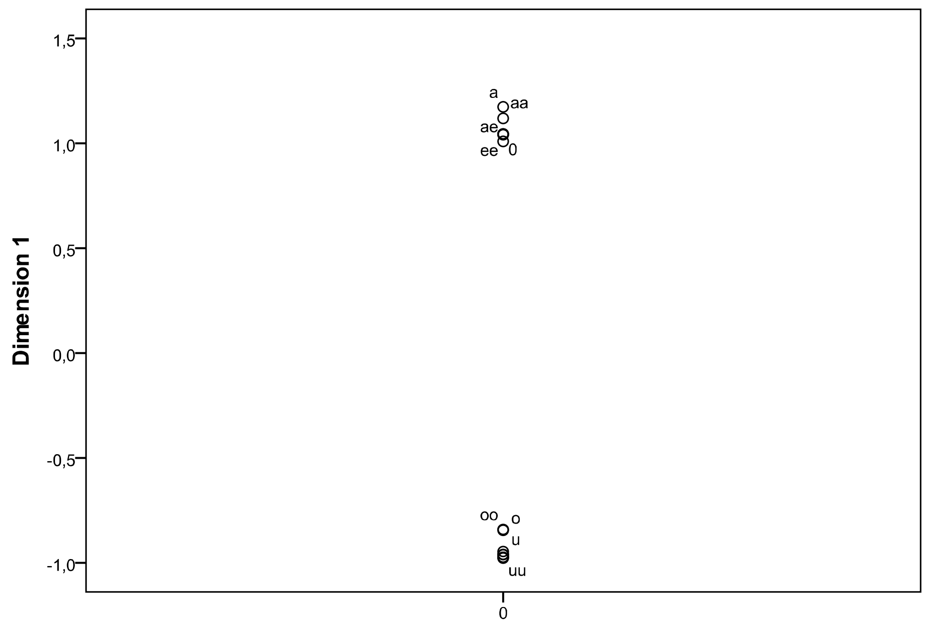

Figure 10.93: One-dimensional MDS representation of the perceptual map of German vowels for L1 Turkish listeners (N=24, 6240 responses, RSQ .405)

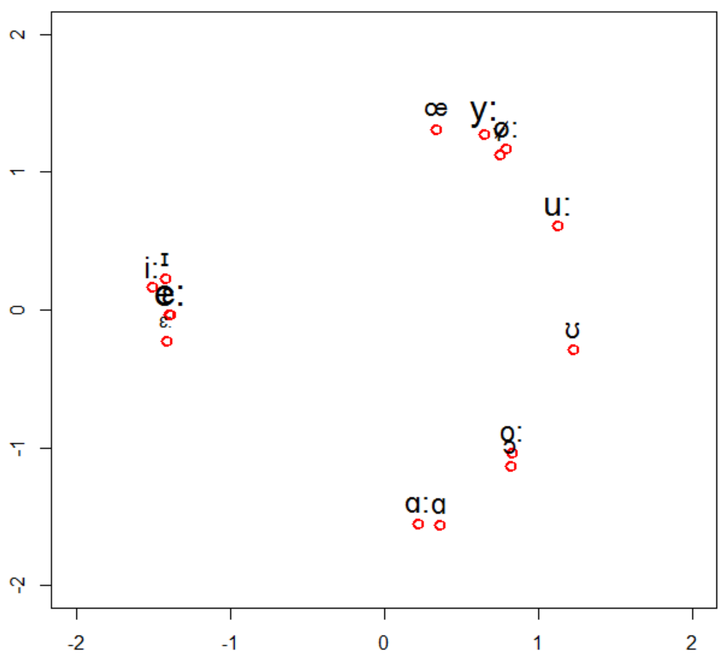

Figure 10.94: Two-dimensional MDS representation of the perceptual map of German vowels for L1 Turkish listeners (N=24, 6240 responses, RSQ .693)

The two-dimensional MDS solution in Figure 10.94 presents similarity relations between German vowels together with perceptual preferences for L1 Turkish listeners. It shows that front unrounded i- and e-vowels are clearly differentiated from all other categories. The MDS vowel map moreover shows that front unrounded vowels are not differentiated sufficiently from each other, whereas velar and uvular vowels appear to be perceived as less similar to each other.

a-vowels are clearly differentiated from all other German vowel categories, whereas the differentiation of long vs. short a-vowels is difficult for Turkish learners. Long /ɑ:/ is slightly more preferred than short /ɑ/.

Front rounded vowels cluster tightly together in the two-dimensional MDS representation. The preference scores in Figure 10.92 and the confusion matrix in Table 10.73 shows an insufficient differentiation within this class but also clear asymmetries with respect to preferences for specific categories: While /e:/ (pref_score 10.4%) is the most attractive response category for front non-rounded vowels, followed by /i:/ (7.4%), all other palatal categories are less frequently selected. The lowest preference score is observed for /ɛ:/ (2.5%). Within the category of front roundedvowels, /ʏ/ and /ø:/ show a strong overlap in the MDS representation, while /œ/ and /y:/ are better kept apart from other roundedpalatal qualities. In this class, the highest pref_score is obtained for /y:/ (9.8%), followed by /ø:/ (6.2%). Short unconstricted /ʏ/ and /œ/ have considerably lower pref_scores (both 4.9%). As the confusion matrix and the MDS representation shows, hardly any substitutions of front roundedvowels with non-rounded categories occur. Substitutions of front roundedvowels with back u-vowels do not occur frequently either and are observed mostly for /ʏ/ and /ø:/ (/ø:/ > /u:/ 3.4%, /ø:/ > /ʊ/ 0.2%, /y:/ > /u:/ 1.2%, /ʏ/ > /u:/ 0.7%, /ʏ/ > /ʊ/ 0.7%, but only one single occurrence of /œ/ > /ɔ/). /œ/ is never substituted with u-vowels, which explains the relatively larger distance of /œ/ to back u- and o-vowels. The position of /u:/ being nearer to the cluster of front rounded vowels than o-vowels is explained by these substitutions of front rounded vowels with u-qualities. /ʊ/ is differentiated quite well from all other roundedvowels, even if some substitutions with o-vowels are observed in the data.

Interestingly, the MDS representation in Figure 10.94 shows an affinity of German o-vowels and a-vowels. A similar affinity is also found in the data presented by Terbeek (1977: 248f). A look at Terbeek's (1977) vowel dissimilarity values as visualized by Johnson (2012: 147) reveals a smaller distance of /ɑ/ and /o/ for Turkish listeners than for English, German or Thai listeners. Johnson hypothesizes that this affinity is influenced by the presence of backness harmony in Turkish emphasizing a linguistic affinity to Turkish listeners of [ɑ] to other back vowels. However, to understand the exact position of a vowel in the MDS representation it is necessary to consider all distances between vowels in the two-dimensional space. The position of a vowel phoneme in the MDS solution is primarily determined by the number of instances of inter-category confusion which is of course indirectly conditioned by phonological regularities in the listeners' L1. In the data of the present study, substitutions of a-vowels with o-vowels occur only very sporadically: Substitutions of /ɑ/ > /o:/, /ɑ/ > /ʊ/, /ɑ:/ > /ɔ/, /ɑ:/ > /o:/, and /o:/ > /ɑ:/ are observed each only once in the data. It is true, however, that substitutions of a-vowels by o-vowels only occur in the Turkish data (with one single exception in the Farsi data), but never in the data of listeners from other languages. These substitutions reveal a certain psychological affinity of German a- and o-vowels that may be due to a certain influence of Turkish backness harmony together with the articulatory proximity of pharyngeal and uvular vowels and the specific vowel articulation in Turkish.

The three-dimensional solution shows a better differentiation of distances between front non-rounded vowels, a clear-cut differentation of /ʊ/, /ɔ/ and /o:/, and a cluster of front roundedvowels and /u:/, where the position of /u:/ may be surprising at first sight but is explained by single occurrences of /u:/ > /y:/ and /u:/ > /ø:/ that are below < 0.5% and are therefore not indicated in Table 10.73.

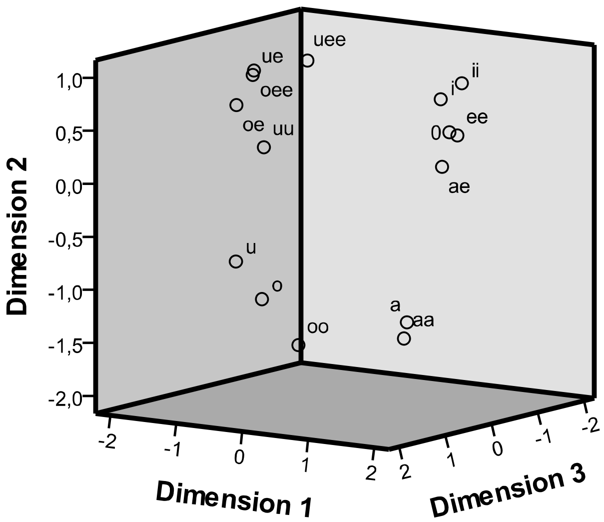

Figure 10.95: Three-dimensional MDS representation of the perceptual map of German vowels for L1 Turkish listeners (N=24, 6240 responses, RSQ .801)

10.12.5 Summary

To summarize, problems in perceptual categorization are observed for three major classes of German vowels: (1) mid- vs. pre-palatal non-rounded vowels, (2) mid- vs. pre-palatal roundedvowels, and (3) velar vs. uvular vowels.

/y: ɑ: i: u:/ the categories that are most stable with id_correct scores between 80 and 90%, /ɛ:/ is the most difficult category (18% correct). Within the class of front rounded vowels, /ø: ʏ œ/ are specifically difficult, while /y:/ has the highest id_correct score (89%).

Asymmetry in categorization patterns is observed in all three classes for long constricted vowels. The data for long constricted categories show a strong preference to substitute mid-palatal qualities with pre-palatal qualities and uvular qualities with velar qualities, or more general, long constricted qualities are perceived as more [fronted] than the original input stimuli: (1) /e:/ > /i:/ (11%) vs. /i:/ > /e:/ (only 4%), /ɛ:/ > /e:/ (69%) vs. /e:/ > /ɛ:/ (only 10%), (2) /ø:/ > /y:/ (32%) and /ʏ/ (5%) vs. /y:/ > /ø:/ (only 2%), and (3) /o:/ > /u:/ (20%) vs. /u:/ > /o:/ only /1%).

For short unconstricted vowels, an inverse pattern of asymmetry is observed: (1) /ɪ/ > /ɛ/ (17%) or /e:/ (6%) vs. /ɛ/ > /ɪ/ or /i:/ (both 0%), (2) /ʏ/ > /œ/ (6%) or /ø:/ (5%) vs. /œ/ > /ʏ/ or /y:/ (both 0%), and (3) /ʊ/ > /ɔ/ (7%) or /o:/ (3%) vs. /ɔ/ > /ʊ/ or /u:/ (both 0%). These observations indicate that short unconstricted categories show a strong tendency to be identified as more [back] than the original input stimuli. A general preference for long constricted vowels is observed.

In general, the extent of inter-category confusion is rather moderate for Turkish listeners. Especially front rounded vowels, which are particularly difficult for listeners from many other languages, are quite well differentiated from other vowel classes. Within the class of front rounded vowels, /y:/ and /œ/ are well differentiated, while /ø:/ and /ʏ/ cause more difficulties.

Difficulties in differentiating i-vowels and e-vowels are found as described in previous studies specifically for /e:/ and /ɪ/.

11 Vowel-specific Analysis and Cross-language Comparison

The perception of an actual vowel stimulus in an identification experiment is determined by the available acoustic information in the input signal as well as by the listeners' mental category representations. Mental representations of L2 categories are guided by language knowledge and hypotheses about the L2 vowel system, its categories and contrasts that refer to articulatory and acoustic characteristics as well as to orthographic representations.

In an experimental context, the rate of correct vowel identification in L2 is expected to be determined by (1) *intrinsic properties* of the vowel stimuli such as formant frequency values, dynamic spectral change, duration, pitch variation, and formant trajectories to adjacent sounds and (2) by the listeners' *mental representations* of L2-categories, influenced by aspects of language-specific contrastiveness and psychological relationships of *inter-language similarity* of L1 and L2 sounds as well as *intra-language similarity* within the emerging L2 system, and (3) by the experimental design and the set of response options.

Spectral patterns of vowel signals mainly depend on the position of the tongue body, mandibular aperture and lip configuration. Within a vowel theory that focusses on specific *constriction locations* rather than assuming a continuous vowel space, it is hypothesized that vowel categories can be described in terms of qualitatively distinct classes each of which refers to typical articulatory constellations and their acoustic effects (see section 4.7.5, 4.7.6, and chapter 5). Following the classification of the major narrowing areas in the vocal tract presented in chapter 5, four major types of discrete constriction locations are differentiated: *palatal, velar, uvular and pharyngeal vowels* (Wood 1979). For German, the class of palatal vowels is further divided into a pre-palatal and a mid-palatal sub-class.

This chapter will focus on a *vowel-specific analysis* comparing patterns of perceptual identification for each constriction location and describing patterns of confusion within and between constriction locations. The following sections will analyse major trends in perceptual substitution patterns in a cross-language comparison for each of the major vowel classes and search for possible reasons for between-category asymmetries in difficulty and preferences.

11.1 Patterns of confusion

Patterns of confusion between L2 German categories were summarized in confusion matrices that provide evidence for the number of wrong and correct identifications as well as for the direction of perceptual substitutions. To compare cross-linguistic trends and language-specific differences in vowel identification, the data will be discussed separately for each of the major constriction areas, i.e. the class of *pharyngeal a*-vowels, *velar u*- and *uvular o*-vowels and for the class of *palatal non-rounded i*- and *e*-vowels and *palatal rounded ü*- and *ö*-vowels. The discussion will consist of a short quantitative description of the relative difficulty of the respective vowel class as reflected by id_wrong scores and is followed by a more qualitative description of patterns of confusion and perceptual substitutions observed for vowel categories at a given constriction.

For each of the vowel classes under consideration, the data will be presented in a confusion table ("comparison of identification scores") summarizing id_V scores and substitution patterns from all L1 sub-samples for each vowel category. For reasons of clarity and conspicuity, the percentage of correct identifications (id_correct) is represented in bold numbers, the percentage of wrong identifications below the 5.5% mark is printed in grey, id_scores below 0.5% are not indicated. The tables' last column indicates the token number of responses (#) for a specific input category obtained from listeners of a given language sub-group. The last row "all L2" summarizes the results for all L2 listeners. For reasons of comparability, the results of the German native control group are presented in the tables, but are not included in the percentages in the tables' last row.

11.1.1 Pharyngeal vowels

Compared to other German vowel categories, pharyngeal ɑ-vowels cause relatively few difficulties in L2 perception, even though the id_wrong scores for German ɑ-vowels show some variation across and within the language sub-groups (see Table 11.1 and Figure 11.1).

% wrong	Alb	Arab	Engl	Farsi	Hung	Mand	Pol	Rom	SBC	Turk	**all L2**	German
ɑ	38	15	27	25	15	41	38	2	12	20	**24**	3
ɑ:	15	25	36	47	6	25	18	10	13	14	**16**	2

Table 11.1: Comparison of percentage wrong identifications for pharyngeal vowels

For short /ɑ/, least difficulties are observed with Serbo-Croatian (11.8% id_wrong), Arabic (14.5% id_wrong) and Hungarian (14.9% id_wrong) listeners, whereas the id_wrong scores for Mandarin (41.1%), Polish (38.1%) and Romanian (28.2%) listeners are much higher. The lowest id_wrong scores for long /ɑ:/ are observed with listeners from L1 Hungarian (5.9%), Romanian (9.8%) and SerBoCroatian (12.5%), while Farsi (47.2%) and English (36.2%) listeners show very high id_wrong scores.

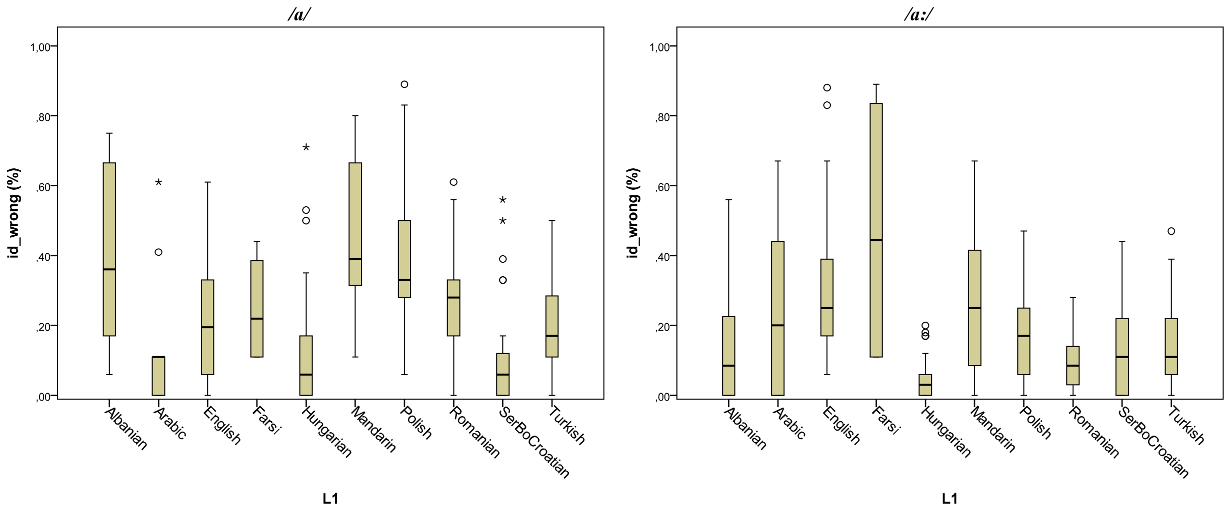

Figure 11.1: Cross-language comparison of identification scores (id_wrong) for German pharyngeal vowels

/ɑ/ (%)	ɑ	ɑː	ɛː	ɛ	eː	ɪ	iː	ɔ	oː	ʊ	uː	œ	øː	ʏ	yː	‹ei›	‹eu›	‹au›	#
Albanian	**62**	37	1																205
Arabic	**85**	12	2							1									179
English	**73**	22	2	1						2	1								265
Farsi	**75**	21						1										3	72
Hungarian	**85**	14																	409
Mandarin	**59**	40	1																107
Polish	**62**	36	1																554
Romanian	**72**	28																	216
SBC	**88**	12																	592
Turkish	**81**	19																	416
German	**98**	2																	324
all L2	**76**	24																	3069

Table 11.2: Comparison of identification scores for German /ɑ/

/ɑː/ (%)	ɑ	ɑː	ɛː	ɛ	eː	ɪ	iː	ɔ	oː	ʊ	uː	œ	øː	ʏ	yː	‹ei›	‹eu›	‹au›	#
Albanian	15	**85**																	206
Arabic	23	**75**	2																175
English	34	**64**	1															1	260
Farsi	42	**53**		1														4	72
Hungarian	4	**94**																1	409
Mandarin	23	**75**	1	1															106
Polish	15	**82**														1			553
Romanian	9	**90**																	214
SBC	11	**88**	1																594
Turkish	12	**86**																	412
German	2	**98**			1														324
all L2	16	**84**																	3055

Table 11.3: Comparison of identification scores for German /ɑː/

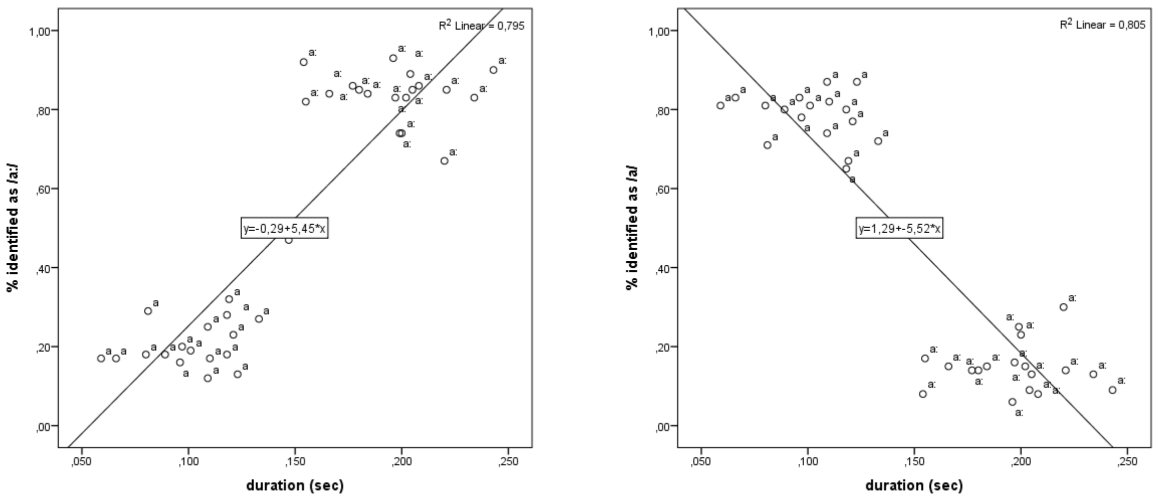

Figure 11.2: Stimuli duration (in sec) for /ɑ:/ and /ɑ/ and the percentage of identifications as long /ɑ:/ (left) or short /ɑ/ (right)

Long /ɑ:/ is identified more often correctly (84% correct vs. 15% identified as /ɑ/) than short /ɑ/ (76% vs. 24% /ɑ:/). <ä:> was selected in less than 0.5% of all responses for /ɑ/ and /ɑ:/ respectively. Other vowel categories do not seem to be attractive response options for a-vowels.

Difficulties with pharyngeal a-qualities are mainly related to uncertainties regarding phonemic length. The mean duration of short /ɑ/-stimuli is 104 ms (min 59 ms, max 147 ms), the mean duration for long /ɑ:/ is 197 ms (min 154 ms, max 243 ms). Figure 11.2 represents the impact of intrinsic duration on identification of ɑ-stimuli as long or short vowel. The results indicate that the critical value is around 150 ms. For a-stimuli with a duration of more than 150 ms the percentage of [+long] identifications increases significantly.

11.1.2 Velar and uvular vowels

German back rounded vowels can be divided into two classes, velar u-vowels and uvular o-vowels, and are further differentiated into long constricted vs. short unconstricted qualities. The perceptual differentiation of German back rounded vowels is commonly reported to cause problems in the acquisition of German to L2 learners with different native languages. Difficulties in category discrimination and correct identification of o- and u-vowels become evident in the box-and-whisker plots in Figure 11.3 representing the id_wrong scores for o- and u-vowels for each language sub-group. The relative degree of difficulty varies considerably across vowel categories as well as across and within the language sub-samples depending on the learners' L2 proficiency as well as on specific characteristics of their L1 (see Table 11.4 representing the id_wrong values for all L2 learners and for each language sub-group separately). Listeners in the German native control group show id_correct scores of 98% or more for back vowels and very limited variation in the selection of substitutes, whereas the results for non-native listeners are much more wide-spread.

In the "all L2" sample least difficulties are observed with /ɔ/ (26.3% id_wrong), whereas /o:/ seems to cause most difficulties in correct identification (39.3% id_wrong), followed by /ʊ/ (34.3%) and /u:/ (29.3%).

% wrong	Alb	Arab	Engl	Farsi	Hung	Mand	Pol	Rom	SBC	Turk	**all L2**	German
ɔ	35	14	33	32	9	54	37	34	15	33	**26**	2
o:	26	51	47	67	11	41	65	26	40	33	**39**	2
ʊ	37	41	60	53	11	61	45	37	24	30	**34**	2
u:	22	51	65	58	3	40	39	20	28	17	**29**	1

Table 11.4: Comparison of wrong identifications for velar and uvular vowels (%)

The generally high percentage of wrong identifications for velar and uvular vowels are mostly due to a high percentage of inter-category confusion within the class of back vowels. These difficulties are not surprising because of the relative acoustic similarity of German o- and u-vowels in terms of formant frequencies (see section 5.7 and 5.8). While F1- and F2-frequencies for German /u:/ and especially for /ɔ/[132] are clearly distinct from other back vowels, there is considerable overlap of F1- and F2-frequencies observed especially for /o:/ and /ʊ/. This overlap can be disambiguated by German native listeners by using duration cues (together with lexical and morphological information in everyday language processing) but causes severe difficulties to non-native listeners.

For L2 listeners, considerable confusion within the class of back vowels as well as within the larger class of rounded vowels is observed. L2 learners in some language sub-samples show not only difficulties in differentiating velar vs. uvular vowels but also considerable difficulties to differentiate front rounded and back rounded vowels, whereas German native listeners never substitute back vowels with front rounded vowels. Moreover, German natives show no substitutions of /o:/ > /ʊ/ or vice versa and only occasionally substitutions of /u:/ > /ʊ/ (3 cases), /ʊ/ > /u:/ (5 cases), and /o:/ > /ɔ/ (3 cases). Moreover, an asymmetry in substitutions of /o:/ > /u:/ is observed in 4 cases (1%) with L1 German listeners that never occur vice versa.

For /ɔ/ (26.3% id_wrong), least difficulties are observed with Hungarian listeners (8.5% id_wrong), whereas Mandarin listeners show a very high id_wrong score of 53.8%.

[132] This effect is mainly due to the higher degree of aperture in articulation for /ɔ/ and the resulting higher F1-frequencies compared to other back vowels.

The highest percentage of wrong identification of /o:/ (39.3% id_wrong for all L2 listeners) is observed with Polish (65.4%) and Farsi listeners (66.7%), followed by Arabic (51.2%) and English listeners (47.3%), while Hungarian listeners identify /o:/ incorrectly in only 11% of the cases.

Very high id_wrong scores are also observed for /ʊ/ (34.3% id_wrong for all L2 listeners) in the sub-samples of Mandarin (60.7%), English (60.2%) or Farsi (52.8%), but also in the Polish (44.9%) and the Arabic (40.6%) sub-sample. The highest id_wrong scores for /u:/ (29.3% id_wrong) are found in the sub-sample of English (64.7%), Farsi (58.3%), Mandarin (40.2%) and Polish (39.2%).

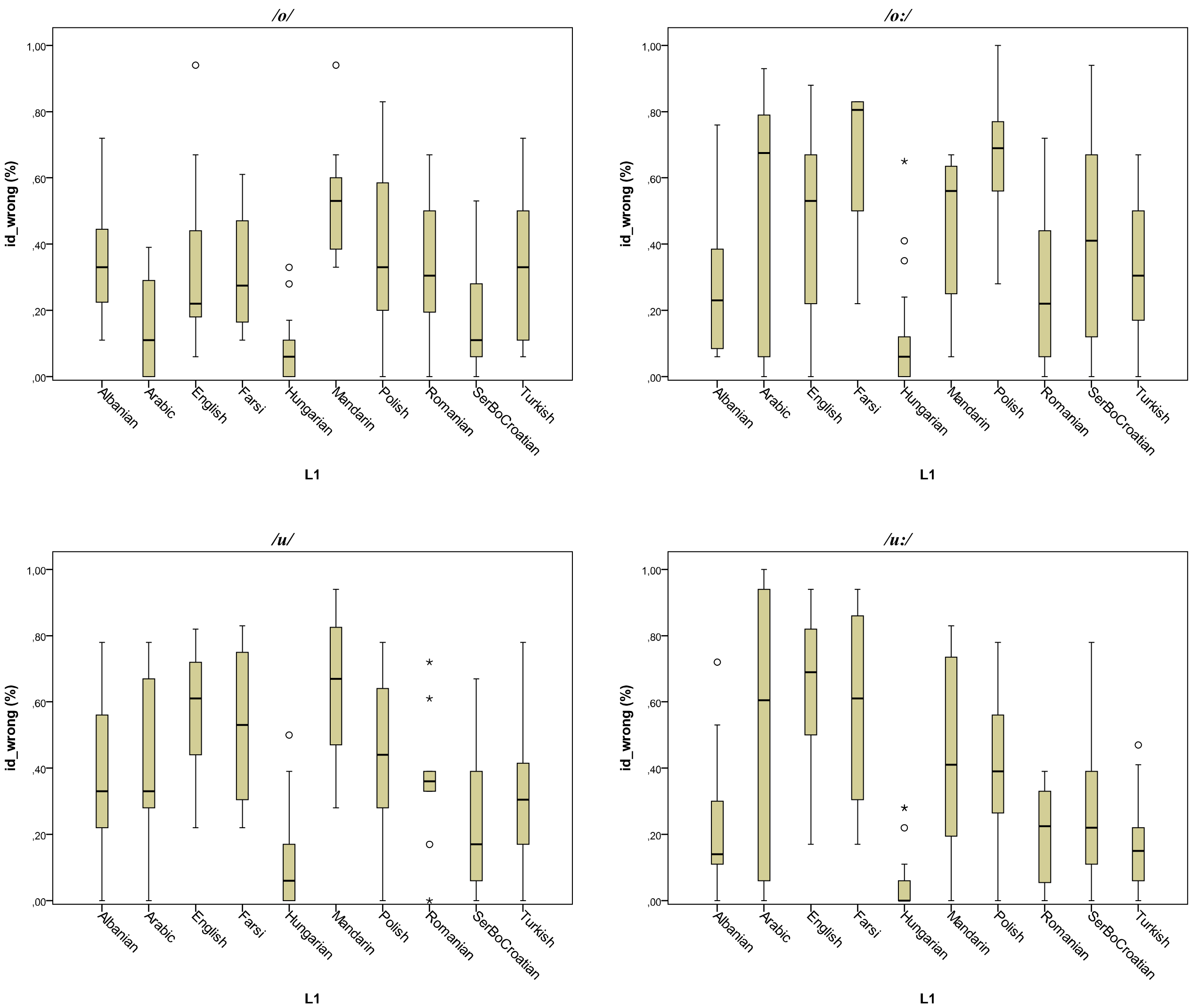

Figure 11.3: Cross-language comparison of identification scores (id_wrong) for German back rounded vowels

While Hungarian and Turkish listeners never substitute back vowels with front rounded vowels, such substitutions are observed in all other language sub-groups, though this effect is limited with Albanian listeners. These results can be explained by the existence of front rounded vowels in the learners' native language, i.e. in Hungarian, Turkish and Albanian. However, Mandarin listeners, who are supposed to have at least some perceptual sensitivity to the front-back distinction within the class of rounded vowels due to the existence of /y/ in their native inventory, have considerable problems with the differentiation of back vs. front rounded vowels.

The under-differentiation of back vs. front rounded vowels becomes evident in a comparison of the two-dimensional MDS solutions summarized in Figure 9.18, where the phenomenon is particularly conspicuous for English, Arabic and Mandarin, but is also observed for other languages such as Polish[133] or SerBoCroatian, whereas Albanian and Romanian listeners show better abilities to discriminate back from front rounded vowels. For Farsi listeners we observe an interesting pattern: while /o:/, /ʊ/ and /u:/-tokens are incorrectly identified with ü- as well as ö-vowels, /ɔ/ items are only substituted with /œ/ or /ø:/ but never with ü-vowels.

In some of the language sub-samples, substitutions of back rounded vowels with <eu> and <au> are observed. These occur more often with uvular o-vowels than with velar u-vowels. Substitutions of /ɔ/ or /o:/ with <eu> and <au> are observed relatively often with learners from Mandarin, Turkish or Farsi, but are considerably less common in other language sub-groups. A similar though weaker effect becomes evident for /o:/, but is not as common for velar u-vowels.

The preferred substitution option for /ɔ/ is long constricted /o:/. Several participants show no clear cut differentiation of /ɔ/ and ö-vowels, whereas the differentiation of /ɔ/ from u-vowels seems somewhat easier; only Arabic and English listeners show higher percentages of substitutions with u-vowels. Id_scores above 5.5% are only observed in the Farsi and the Mandarin samples, an effect that is of limited significance because of the small size of the Mandarin and Farsi sample and the higher number of less experienced L2 German learners in the sample, who have generally more difficulties than more proficient learners in other sub-samples.

---

[133] Polish learners also show a strong bias to substitute front rounded vowels with either back rounded or front non-rounded vowels. Substitutions with front non-rounded vowels can be described as perceptual delabializations (see section 11.2.3), which explain the position of front rounded vowels „between“ front non-rounded and back rounded vowels in the MDS solution for Polish. A similar but much weaker effect is observed for SerBoCroatian: SBC listeners (as well as listeners from many other languages) show a strong bias to substitute /œ/ with front non-rounded e-qualities; with SBC listeners, this pattern occurs only very occasionally with /y:/ and /ʏ/, but never with /ø:/ (a single case of /ø:/ identified as <ä:> is most probably due to orthographic similarity).

/ɔ/ (%)	ɑ	ɑ:	ɛ:	ɛ	e:	ɪ	i:	**ɔ**	o:	ʊ	u:	œ	ø:	ʏ	y:	‹ei›	‹eu›	‹au›	#
Albanian								**66**	31			3							206
Arabic								**86**	5	3	1	4					1		175
English	1							**67**	18	3	2	4	1	2			1		262
Farsi								**68**	13			10	6			1	3		72
Hungarian	1							**92**	8										413
Mandarin								**46**	40		1	2	1				5	6	106
Polish							1	**63**	31			1	1	1			1		554
Romanian						1		**67**	25			4	2						215
SBC								**85**	10	1		3	1						589
Turkish								**67**	27								4		416
German								**98**	2										324
all L2								**74**	20	1		2	1				1		3062

Table 11.5: Comparison of identification scores for German /ɔ/

The results for /o:/ (see Table 11.6) are considerably more wide-spread. Substitutions of /o:/ with u-vowels are observed in all L2 sub-samples, whereas German native listeners substitute /o:/ with /u:/ in only 4 of 324 cases. Several language sub-samples show a considerable number of substitutions of /o:/ with /ø:/ or /œ/, while the frequency of substitutions with /y:/ and /ʏ/ lies below the 5.5% mark in all groups. Confusion with /o:/ and front rounded vowels occur in all language groups except Hungarian and Turkish. Moreover, there are several instances of incorrect identifications of /o:/ with either <eu> or <au> in all sub-samples except Albanian, SerBoCroatian and German.

/o:/ (%)	ɑ	ɑ:	ɛ:	ɛ	e:	ɪ	i:	ɔ	**o:**	ʊ	u:	œ	ø:	ʏ	y:	‹ei›	‹eu›	‹au›	#
Albanian								9	**74**	2	4	1	6		1				203
Arabic								16	**49**	3	6	5	10	5	4		2		168
English								20	**53**	3	4	6	7	4	1		1	2	260
Farsi			1					24	**33**	3	10	7	14	3	4		1		72
Hungarian								3	**89**	3	4							1	382
Mandarin								4	**59**	8	18	1	5	1			3	2	104
Polish							1	7	**35**	12	27	2	8	2	5		1	1	547
Romanian								5	**74**	3	10	1	5		1		1		212
SBC								6	**60**	3	9	4	13	2	3				571
Turkish								6	**67**	2	20						4		418
German								1	**98**		1								324
all L2								8	**61**	4	12	2	6	1	2		1	1	2988

Table 11.6: Comparison of identification scores for German /o:/

The results for /ʊ/ (see Table 11.7) are similarly wide-spread as for /o:/. Major substitution options for /ʊ/ are either long constricted /u:/ (except for Arabic and German listeners) or short unconstricted /ɔ/ (in all language groups), /ʏ/ (mainly found with Arabic, English, Farsi and SBC listeners), /o:/ (mainly found in the Mandarin, English and Farsi sample[134]) and /œ/ in the Arabic and Mandarin sample. These data reflect the general observation that short unconstricted vowels are easily confusable due to their duration and the nearly inexistent quasi-stable phase in the acoustic signal. Substitutions of short /ʊ/ with diphthongs are rarely occurring.

/ʊ/ (%)	ɑ	ɑ:	ɛ:	ɛ	e:	ɪ	i:	ɔ	o:	**ʊ**	u:	œ	ø:	ʏ	y:	‹ei›	‹eu›	‹au›	#
Albanian								10	4	**63**	20	1	1	1					206
Arabic							1	12	3	**59**	1	7	1	16	1		1		180
English								19	7	**40**	12	4	4	10	4				264
Farsi				1				14	6	**47**	10	1	1	10	10				72
Hungarian								3		**89**	6								413
Mandarin								15	10	**39**	23	6		2	1		1	3	107
Polish								8	4	**55**	21	3	2	4	1				554
Romanian								17	4	**63**	10	1		4	1				216
SBC								4	1	**76**	7	2		7	2				594
Turkish								7	3	**70**	19						1		421
German										**98**	2								324
all L2								9	3	**66**	13	2	1	5	1				3081

Table 11.7: Comparison of identification scores for German /ʊ/

/u:/ (%)	ɑ	ɑ:	ɛ:	ɛ	e:	ɪ	i:	ɔ	o:	ʊ	**u:**	œ	ø:	ʏ	y:	‹ei›	‹eu›	‹au›	#
Albanian									1	20	**78**				1				205
Arabic								1	3	18	**49**	1	5	12	10		1		172
English								2	6	22	**35**	1	3	15	14		2		258
Farsi									1	28	**42**		8	6	15				72
Hungarian										3	**97**								414
Mandarin								2	3	21	**60**	7	5		3				107
Polish									3	21	**61**	1	1	5	7		1		548
Romanian									2	13	**80**			1	4				216
SBC									1	10	**72**	1	1	5	11				589
Turkish									1	15	**83**								419
German										1	**99**								324
all L2									2	15	**71**	1	1	4	6				3054

Table 11.8: Comparison of identification scores for German /u:/

Id_wrong scores for long constricted /u:/ (Table 11.8) show a large range of variation across the language sub-samples. /u:/ is particularly difficult for English, Farsi and Arabic participants but very stable in the Hungarian, Turkish, Romanian and Albanian sample. /u:/ is most frequently confused with its short counterpart /ʊ/, a strong pattern that occurs in all language groups. Moreover, a considerable percentage of substitutions with ü-vowels is observed in all language sub-samples except Hungarian and Turkish. This pattern of substitutions of u-vowels with ü-vowels can be explained by mechanisms of hypercorrect

[134] The percentage rate for Mandarin and Farsi has to be viewed with caution due to the small sample size.

identification behaviour (see section 1.3.2) that is inhibited by the higher sensitivity of Hungarian and Turkish learners to contrasts of front rounded vs. back vowels due to the existence of front rounded vowels in their native languages.

To summarize, the relative acoustic similarity of velar u-vowels and uvular o-vowels causes considerable category confusion to non-native listeners of German. Three major factors are assumed to explain these difficulties: (1) the location of u- and o-vowels in the back oral cavity and the "obscuring" effect of lip rounding together with the lower amplitude of disambiguating higher formant frequencies, causing (2) considerable overlap of F1-, F2- and F3-frequencies; specifically the similarity of F1 for /ʊ/ and /o:/ can only be disambiguated by duration and higher formant frequencies, and (3) the similarity of orthographic representations of o- and u-vowels and front rounded ö- and ü-vowels.

While /ɔ/ appears to be the most stable category, where mainly substitutions with its long constricted counterpart /o:/ are observed, considerably more confusion is observed for /o:/, /ʊ/ and /u:/. For /u:/, confusion with vowels of the same tongue height or degree of aperture, i.e. /ʊ/, /y:/ and /ʏ/, is observed, whereas the responses obtained for /o:/ and /ʊ/ show a wide range of different candidates for substitutions. However, substitutions of back rounded vowels with front non-rounded vowels are rather rare, whereas substitutions in the opposite direction are observed quite frequently (see the discussion of labio-palatal vowels in section 11.1.4).

A comparison of substitutions of /o:/ > /u:/ and /u:/ > /o:/ shows a clear pattern of asymmetry: Substitutions of /o:/ > /u:/ (see Table 11.6) are much more common than substitutions in the opposite direction. Only English natives show a significant tendency to substitute /u:/ by /o:/ (6%, possibly due to orthographic interference).

To conclude, the differentiation of German velar u- and uvular o-vowels is difficult to L2 learners for several reasons:

(1) the existence of four back rounded vowel phonemes in the German system that are differentiated not only in terms of constriction location but also constriction degree and phonemic length, mostly not used in the listeners' native language.
(2) the "damping effect" of lip rounding decreasing the intensity of higher formant frequencies,
(3) orthographic factors that may (mis)lead L2 learners causing confusion of front ü- and ö-vowels and back u- and o vowels (see section 11.2).

These characteristics of the German system in interaction with the specific constellation of the vowel inventory and the typical realization of u- and o-vowels in the listeners' L1 explain the high portion of difficulties to differentiate German front vowels correctly. The results for the class of front rounded vowels show an even higher percentage of category confusion and are discussed in the following section.

11.1.3 Pre- and mid-palatal vowels

In the front part of the oral cavity, German differentiates a set of front non-rounded vs. front rounded vowel phonemes. These are further divided by the exact location of constriction in the front part of the oral cavity into a pre- and a mid-palatal constriction location. A further differentiation can be described in terms of differences in constriction degree and phonemic length. Long unconstricted /ɛ:/ does not correspond to the common differentiation of long constricted vs. short unconstricted phonemes and has a specifically marked status in the German system (see section 5.2 and 5.3).

The specific acoustic quality of palatal vowels is determined by a low F1 and the so-called "center of gravity", a complex formed by higher formant frequencies for F2, F3 and F4 (Stevens 1997: 503). While front vowels can clearly be differentiated from back vowels by their low F1 and the complex of higher formants, considerable difficulties *within* the class of front non-rounded vowels are observed in the data of the present study that are strongly connected with the the differentiation of pre-and mid-palatal qualities and the marginal status of /ɛ:/ (see Table 11.9).

%wrong	Alb	Arab	Engl	Farsi	Hung	Mand	Pol	Rom	SBC	Turk	**all L2**	German
ɛ:	64	54	54	80	46	70	80	69	69	82	**67**	21
ɛ	41	21	28	26	10	51	36	25	21	23	**26**	6
e:	32	56	57	69	21	50	76	28	42	37	**45**	7
ɪ	36	31	40	57	8	45	37	54	21	41	**32**	2
i:	18	27	36	44	1	28	35	16	15	14	**20**	1

Table 11.9: Comparison of wrong identifications for palatal non-rounded vowels (%)

As with back vowels, the results in the full L2 listeners' sample show considerable variation across vowel categories and across and within the language sub-samples (see Table 11.9 for scores of the full L2 sample). High id_wrong scores are obtained in all language sub-groups for /ɛ:/ (66.9% id_wrong), while /ɪ/ (31.6%) and /e:/ (45.4%) show lower error rates. Least difficulties are observed for /i:/ (20.2% id_wrong) and /ɛ/ (25.8%).

Figure 11.4 shows box-and-whisker plots representing the id_scores for each language sub-group. While Arabic, English and Farsi listeners seem to have least difficulties with /ɛ/, in all other language sub-groups the more stable category is /i:/. Perceptual stability of /i:/ and /ɛ/ is also indicated by the relatively limited range of selected response categories for both categories, especially for /i:/, whereas the responses for /ɪ/ and /e:/ spread considerably over several categories. A heavy portion of wrong identifications is observed for /e:/, especially in the Arabic, English, Farsi and Polish sample, a fact that can be explained by difficulties in differentiating /e:/ from /ɪ/ as well as from /ɛ:/ but also by the specific composition of these sub-samples.

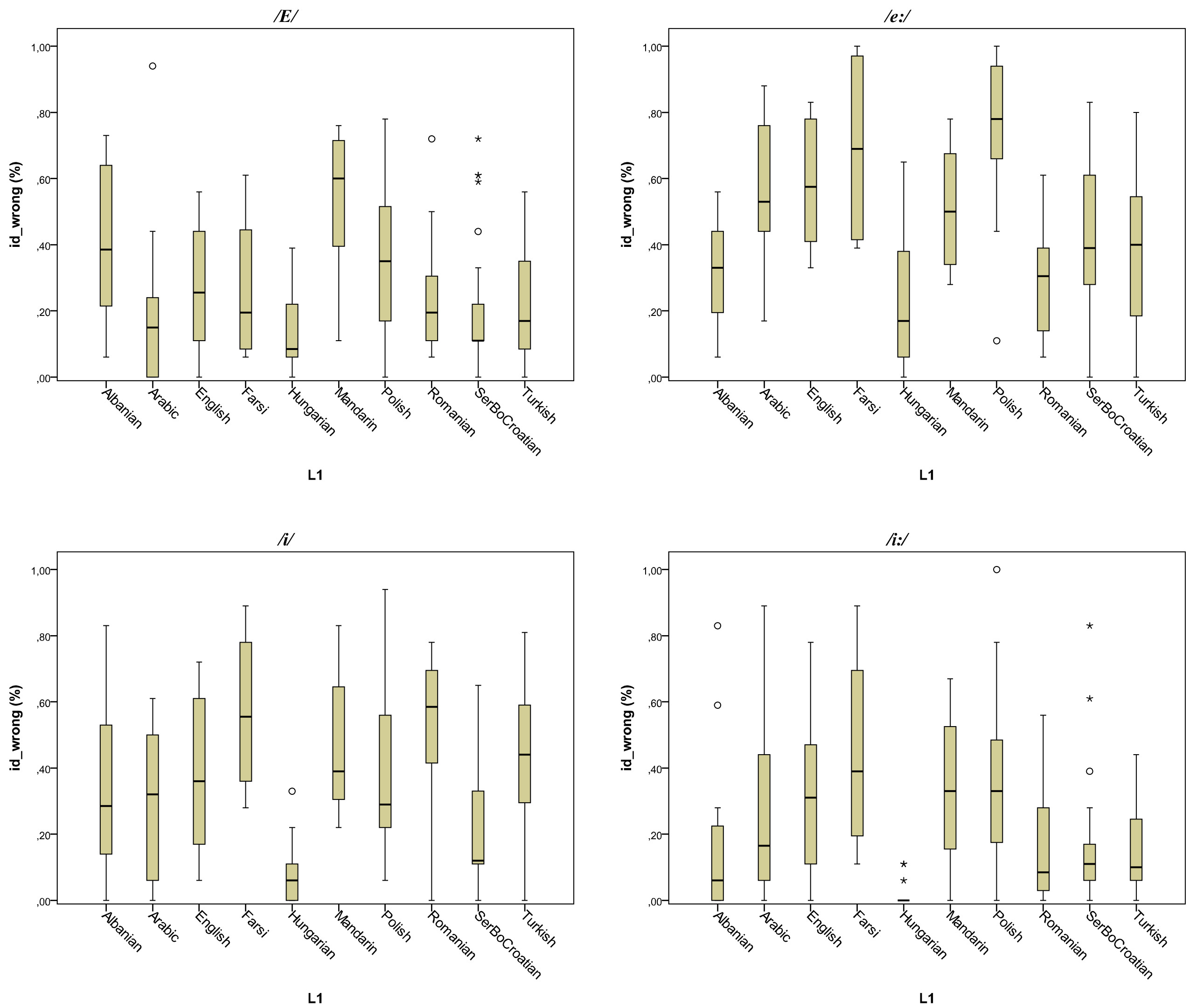

Figure 11.4: Cross-language comparison of identification scores (id_wrong) for German front non-rounded vowels /ɛ/, /e:/, /ɪ/, and /i:/

Table 11.10, Table 11.11, Table 11.12 and Table 11.13 show cross-language comparisons of confusion patterns for German front non-rounded vowels. Confusion patterns for /ɛ:/ are represented in Table 11.14.

/i:/ shows least spread in the choice of response options. It is most frequently confused with its short unconstricted counterpart /ɪ/. Arabic, English, Mandarin and Romanian listeners also show a higher portion of substitutions with /e:/ that is uncommon in other samples.

For /ɪ/, significantly more difficulties are observed. In all samples except Hungarian, substitutions of /ɪ/ with /ɛ/ are observed, whereas this pattern is not observed in the opposite direction, i.e. substitutions of /ɛ/> /ɪ/ occur rarely in the data (the highest score of 5.1% is found in the Arabic sample, see Table 11.12).

Within the class of front non-rounded vowels, the greatest spread of responses over several categories is observed in the data for German /e:/, demonstrating significant difficulties in vowel differentiation within the class of front non-rounded vowels. Non-native subjects identified 14.9% of the /e:/-tokens incorrectly or hypercorrectly as /ɛ:/ and 6.8% as /ɛ/. In all sub-samples, a high percentage of wrong identifications for /e:/ is due to substitutions of /e:/ > /ɛ/ and /e:/ > /ɛ:/, where /e:/ > /ɛ:/ seems to be the more common pattern. The substitution of /ɛ:/ > /e:/ (and less often /ɛ:/ > /ɛ/) is observed in the opposite direction, though for /ɛ:/ generally less spread over different response categories is observed.

Substantial portions of wrong identifications for /e:/ are due to substitutions of /e:/ > /i:/ that are observed in all sub-samples except Romanian and German. Attractive substitutes for /e:/ were /i:/ (12.3%), /ɪ/ (3.6%) and <ei> (4.7%). Id_scores for other categories (/œ/, /ø:/, /ʏ/, /y:/ and <eu>) are below the 0.5% mark.

Farsi, Mandarin, Polish and Turkish listeners moreover show a high percentage of substitutions of /e:/ with a diphthong <ei>. Listeners from these sub-samples seem to reinterpret the phonetic duration of /e:/ as iphthongization (though the results for Farsi and Mandarin are of limited significance because of the small sample size).

Within the class of palatal non-rounded vowels, incorrect identifications of monophthongs with <ei> occur with all – long or short – members of the class but are most frequently observed for /e:/ (particularly in the Mandarin, Farsi, Polish and Turkish sample) and /ɛ:/ (particularly in the Polish, Mandarin, Turkish and Hungarian sample).

Arabic, English and Mandarin listeners show occasional confusion between German a-vowels and e-vowels, where the wrong identification of mid-palatal e-qualities with a-vowels occurs more frequently than in the other direction that may be largely due to allophonic spread of a-vowels in the listeners' L1s.

Several hypercorrect identifications of front non-rounded vowels i- and –e qualities as front rounded categories ü- and ö-vowels are observed, though these hypercorrect substitutions do not occur in all language sub-groups. While Albanian, Turkish or Mandarin listeners never show substitutions with front rounded vowels in any of the five front categories, indicating a high portion of perceptual sensitivity to [+/–round] distinctions in the front part of the oral tract, this sensitivity is weaker with listeners from other languages.

/ɪ/ (%)	ɑ	ɑː	ɛː	ɛ	eː	**ɪ**	iː	ɔ	oː	ʊ	uː	œ	øː	ʏ	yː	‹ei›	‹eu›	‹au›	#
Albanian			1	9	5	**64**	19												202
Arabic			2	17	3	**69**	7							1			1		175
English			1	17	6	**60**	12			1						2			262
Farsi			4	22	3	**43**	17					1				10			72
Hungarian					1	**92**	3							2					411
Mandarin				21	8	**55**	16									1			106
Polish			1	3	3	**63**	16					1	1	7	3	1			553
Romanian			2	31	7	**46**	11							1					216
SBC			1	6	1	**79**	8					1		4					591
Turkish			2	17	6	**59**	15									1			418
German				1		**98**	1							1					324
all L2			1	10	4	**68**	11							2	1	1			3060

Table 11.10: Comparison of identification scores for German /ɪ/

/iː/ (%)	ɑ	ɑː	ɛː	ɛ	eː	ɪ	**iː**	ɔ	oː	ʊ	uː	œ	øː	ʏ	yː	‹ei›	‹eu›	‹au›	#
Albanian						17	**82**												205
Arabic					6	19	**73**							1	1				176
English				4	10	17	**64**									4			260
Farsi					4	29	**56**								3	8			72
Hungarian							**99**									1			408
Mandarin				3	7	18	**72**												103
Polish					2	29	**65**							1	2				554
Romanian					6	8	**84**												216
SBC					1	10	**85**							2	1				591
Turkish				1	4	9	**86**									1			417
German						1	**99**												324
all L2				1	3	14	**80**								1	1			3056

Table 11.11: Comparison of identification scores for German /iː/

/ɛ/ (%)	ɑ	ɑː	ɛː	**ɛ**	eː	ɪ	iː	ɔ	oː	ʊ	uː	œ	øː	ʏ	yː	‹ei›	‹eu›	‹au›	#
Albanian	1		14	**59**	25														201
Arabic		1	7	**79**	5	5	3												175
English	3	2	6	**72**	13	2	1												261
Farsi	1	1	4	**74**	14	3										1	1		72
Hungarian			4	**90**	5											1			414
Mandarin	4	1	15	**49**	19	2	2		1		1					7			105
Polish			7	**64**	23	1	1						1			1			550
Romanian			10	**75**	15														216
SBC			4	**80**	13							1				1			591
Turkish			6	**77**	15											1			408
German			2	**94**	2							1							324
all L2	1		7	**74**	15	1										1			3047

Table 11.12: Comparison of identification scores for German /ɛ/

The category that is most prone for confusion involving the feature [+/–round] is /ɪ/, whereas /ɛ/ and /ɛː/ are most resistant to substitutions with [+round] categories, a fact that may be connected with the high degree of aperture for ɛ-qualities. The high susceptibility of /ɪ/ is most probably due to its intrinsic duration (for discussion of duration, see section 5.4.5), but could also be due to the listeners' experience with allophonic variation of /ɪ/ in some varieties of German.

/e:/ (%)	ɑ	ɑ:	ɛ:	ɛ	e:	ɪ	i:	ɔ	o:	ʊ	u:	œ	ø:	ʏ	y:	‹ei›	‹eu›	‹au›	#
Albanian			15	14	**68**	1	1												206
Arabic	1	3	24	15	**44**	4	10												176
English	1	2	30	15	**43**	3	5									1			261
Farsi			13	21	**31**	6	17					1				13			72
Hungarian			10	1	**79**	2	7									1			410
Mandarin	2		7	10	**50**	3	13									14			107
Polish			7	3	**24**	11	35						1	2	4	11			552
Romanian			18	6	**72**		3												216
SBC			21	6	**58**	1	8					1	2		1	1			591
Turkish			10	4	**63**	2	11									10			418
German			7		**93**														324
all L2			15	7	**55**	4	12						1		1	5			3063

Table 11.13: Comparison of identification scores for German /e:/

Within the class of palatal non-rounded vowels, /ɛ:/ causes most difficulties in the identification task (33% correct, 49% /e:/ and 9% /ɛ/). Figure 11.5 represents a cross-language comparison of id_wrong scores for /ɛ:/, which are above 45% in all language sub-samples. This effect was also observed in the German native control group (79.3% correct, 19.8% /e:/) and is most probably due to the fact that /ɛ:/ is not phonemically functional in most parts of the German speaking countries (see chapter 4).

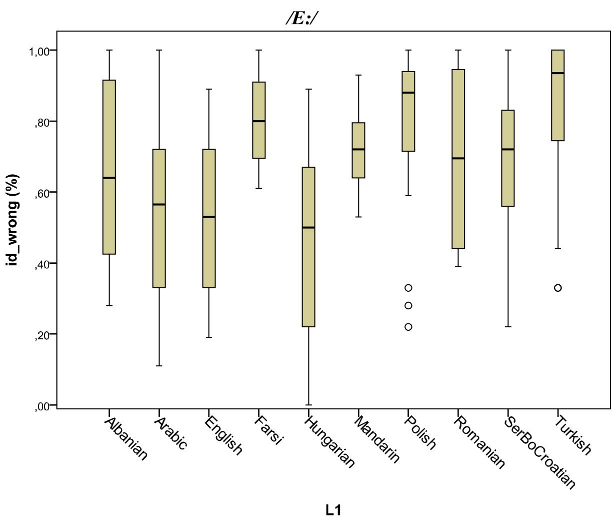

Figure 11.5: Cross-language comparison of identification scores (id_wrong) for German /ɛ:/

The phonemic status of German /ɛ:/ is a matter of debate in many works on the German phoneme inventory (see section 5.2 and 5.3). Maas (1999) calls the opposition a

“Phantomopposition” (cf. Maas 1999: 174f). While /ɛ:/ may be used to differentiate homonymes that differ in the orthographic representation of <e> vs. <ä>, like *Beeren* ‘berrys’ – *Bären* ‘bears’, *Ehre* ‘honour’ – *Ähre* ‘(grain) spike’, it is commonly not used in many parts of the German-speaking countries (Kleiner & Knöbel 2011). Interestingly, while most native speakers of German can realize this distinction in pronunciation and even do in some cases to avoid e.g. homonymy, the results for the German native group show that many participants do not differentiate [e:] and [ɛ:] in perception (for similar results, see Sendlmeier 1981; Hentschel 1986).

/ɛ:/ (%)	ɑ	ɑ:	**ɛ:**	ɛ	e:	ɪ	i:	ɔ	o:	ʊ	u:	œ	ø:	ʏ	y:	‹ei›	‹eu›	‹au›	#
Albanian			**36**	11	52											1			205
Arabic	1	5	**46**	17	27	1	3									1			175
English	2	6	**46**	12	30	1	1									2			263
Farsi			**20**	18	49		4								1	4	3		71
Hungarian			**55**	5	33											6			409
Mandarin	4		**30**	15	42		2									8			103
Polish			**20**	12	46	1	2					1	1		1	15	1		549
Romanian			**31**	6	61														216
SBC			**31**	6	58							1	2			1			594
Turkish			**18**	6	69											7			418
German			**79**		20														324
all L2		1	**33**	9	49		1									5			3057

Table 11.14: Comparison of identification scores for German /ɛ:/

To summarize, the differentiation of German front non-rounded vowels is difficult for L2 listeners for a number of reasons:

(1) the high number of vowel phonemes located in the front part of the oral cavity and the differentiation of pre-palatal vs. mid-palatal rounded vowels,

(2) the differentiation of front rounded and front non-rounded vowels at the pre- and mid-palatal constriction location in the German system,

(3) the functional load of variation in duration and constriction degree that differentiate long constricted vs. short unconstricted vowels that is not used in many of the listeners’ L1s, and

(4) the existence of a grapheme <ä> that can be realized as /e:/ but also as /ɛ:/ at least in some regional and phonostilistic varieties of German.

These characteristics of the German system in interaction with the specific constellation of the vowel inventory of the listeners’ L1 explain the high portion of difficulties to differentiate German front vowels correctly. The results for the class of front rounded vowels show an even higher percentage of category confusion and are discussed in the following section.

11.1.4 Labio-palatal vowels

Beside a set of front non-rounded vowels, German uses a set of front rounded vowel phonemes /y: ʏ ø: œ/ that are characterized by a combination of articulatory gestures in the front part of the oral cavity and gestures of lip rounding/protrusion. As their non-rounded counterparts, front rounded or labio-palatal vowels are further differentiated by the exact location of constriction in the front part of the oral cavity into a pre- and a mid-palatal sub-class. A further differentiation refers to differences in constriction degree and phonemic length.

The specific acoustic quality of front rounded vowels is determined by a low F1 and the so-called "center of gravity" (Stevens 1997: 503) formed by higher formant frequencies for F2, F3 and F4, showing lower frequencies for labio-palatals than for their non-rounded counterparts. A comparison of the acoustic data of labio-palatal vowels shows considerable similarities for /ø:/ and /ʏ/ in terms of F1 and F3 but not F2 (see Figure 5.13), whereas the F1 and F2 values for /y:/ and /œ/ show no overlap with other categories.

%wrong	Alb	Arab	Engl	Farsi	Hung	Mand	Pol	Rom	SBC	Turk	**all L2**	German
œ	39	49	51	51	7	59	79	21	34	34	**41**	2
ø:	42	55	72	83	13	57	84	45	68	44	**55**	2
ʏ	53	36	63	51	15	64	61	32	34	39	**42**	4
y:	25	38	53	43	6	46	51	21	27	11	**30**	2

Table 11.15: Comparison of wrong identifications for labio-palatal vowels (%)

The data show that the correct identification of German front rounded vowels causes considerable difficulties to L2 learners of German (see Table 8.6). These difficulties seem to persist even with very advanced learners of German who show a mean of 11% wrong identifications of labio-palatals (beginners 51% id_wrong, advanced learners 38% id_wrong).

A comparison of the overall id_wrong scores obtained in the perception task (see Table 11.15) shows that a relatively low percentage of 29.4% wrong responses is obtained for /y:/, whereas /ø:/ receives the highest id_wrong score (55.4%) that is considerably higher than for /ʏ/ (41.7%) and /œ/ (41%). In the German control group, /ʏ/ shows the highest id_wrong scores (4%) within the class of front vowels.

The box-and-whisker plots in Figure 11.6 compare the id_wrong scores of German labio-palatals obtained for each language sub-group. The results show considerable variation across vowel categories and across and within the language sub-samples. Within the class of front rounded vowels, /y:/ receives the lowest id_wrong scores in almost all sub-samples except the

Arabic sample, where /ʏ/ receives the best scores, and the English and Romanian sub-sample, where /œ/ receives the highest id_correct scores.

Interestingly, the highest percentage of wrong identifications is obtained for /ʏ/ in the Albanian, Hungarian and Mandarin sample. In the Arabic, English, Farsi, Mandarin, Polish, Romanian, SBC and Turkish sample, on the other hand, /ø:/ is the category with the highest id_wrong scores. However, as evident in the boxplots in Figure 11.6, the listeners' performance varies considerably within each sub-sample.

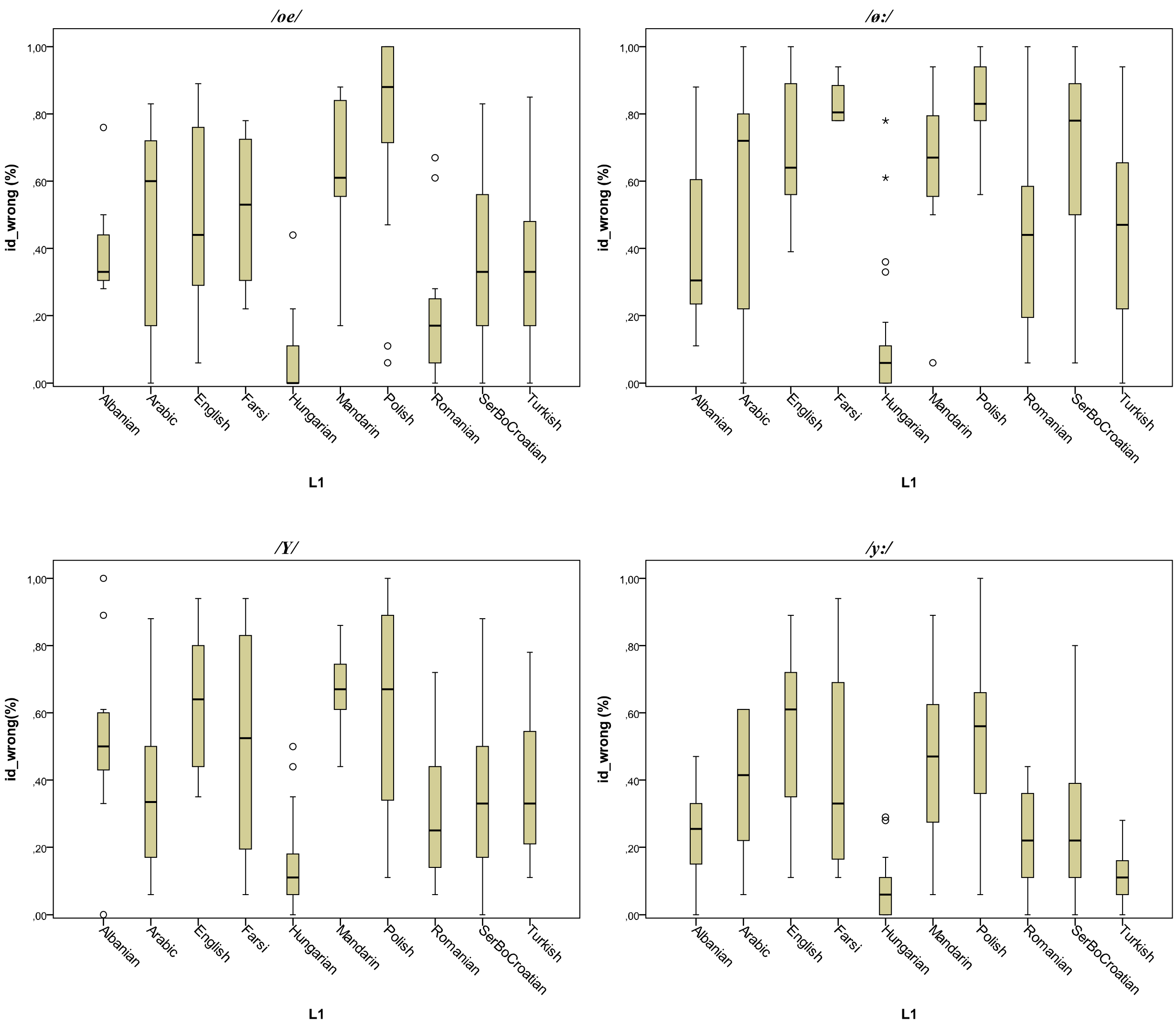

Figure 11.6: Cross-language comparison of identification scores (id_wrong) for German front rounded vowels /œ/, /ø:/, /ʏ/, and /y:/

The data from the perception task show that front rounded vowels are confused with front non-rounded vowels as well as with back vowels. However, the extent to which the responses spread over categories varies considerably within the class of front rounded vowels. Responses spread least over categories for /y:/ (see Table 11.17) and most for /œ/ (see Table

11.18) due to the selection of front non-rounded categories for /œ/-stimuli in several sub-samples.

A cross-language comparison shows that *substitutions with front rounded vowels* are most common for /œ/ (except with Hungarian and Turkish listeners) and least common for /ø:/. /ø:/ shows the lowest percentage of substitutions with front categories in all sub-samples except the Polish and occasionally the Albanian, Arabic and English sample. Romanian listeners substitute only /œ/ with front rounded <ä:> and /ɛ/, Albanian listeners substitute only /œ/ and /ø:/ with <ä:>, most probably due to similarity of <ö> and <ä> in orthography. The highest portion of substitutions with front non-rounded vowels is observed in the Polish sample.

The diphthong <eu> is selected as response option for all labio-palatals, but slightly more often for /œ/ and /ø:/, whereas the selection of <ei> and <au> can be considered as marginal phenomenon.

The most frequently selected substitute for /ʏ/ is /y:/ in almost all language sub-samples except Arabic and English, where /ʏ/ is most frequently substituted with its velar counterpart /ʊ/. The results show moreover a preference for short unconstricted /œ/, /ʊ/ and /ɪ/ rather than with their long constricted counterparts. Front unconstricted /ɪ/ is selected only in a few cases by Arabic, English, Farsi, Mandarin, Polish and SBC listeners. Moreover, a higher percentage of substitutions (> 5.5%) with /ø:/ for /ʏ/ is observed in the Albanian, Farsi, Hungarian, Mandarin and Polish sample.

/ʏ/ (%)	ɑ	ɑ:	ɛ:	ɛ	e:	ɪ	i:	ɔ	o:	**ʊ**	u:	œ	ø:	**ʏ**	y:	‹ei›	‹eu›	‹au›	#
Albanian									1			19	8	**47**	22	1			204
Arabic						1		2	1	16	1	8	2	**64**	5		1		177
English						1		1		22	13	8	5	**37**	10		3		262
Farsi						1		1		8	4	10	10	**49**	13		4		72
Hungarian												6	1	**85**	7				412
Mandarin						1			4	5	4	21	6	**36**	23				103
Polish					2	3	2			11	6	10	6	**39**	19		1		549
Romanian										4	2	10	3	**69**	12				216
SBC						1				11	2	6	1	**66**	13				589
Turkish										1	1	6	5	**61**	26				419
German												1		**96**	2				324
all L2						1				8	3	9	4	**58**	15		1		3057

Table 11.16: Comparison of identification scores for German /ʏ/

Long constricted /y:/ seems to be the most stable category within the class of labio-palatals. It is most frequently substituted with its short unconstricted counterpart /ʏ/. /u:/ is the second most frequently selected category in the Arabic, English and Farsi sample, whereas /ø:/ is the second most frequently selected category with Albanian, Mandarin, Polish, Romanian, SBC and Turkish listeners. Short unconstricted qualities do not seem to be attractive response categories for long constricted /y:/.

/y:/ (%)	ɑ	ɑ:	ɛ:	ɛ	e:	ɪ	i:	ɔ	o:	ʊ	u:	œ	ø:	ʏ	**y:**	‹ei›	‹eu›	‹au›	#
Albanian												1	4	18	**75**				203
Arabic										3	4		1	26	**62**	1	4		178
English									1	4	13	4	6	24	**47**		2		263
Farsi										1	4	3	1	33	**57**				72
Hungarian														6	**94**				412
Mandarin					1					1	2	4	15	23	**54**				105
Polish					1	2	8			4	6	3	6	19	**49**		1		542
Romanian													1	19	**79**				216
SBC			1				1			2	2	1	3	18	**73**				591
Turkish											1		2	7	**89**				412
German													1	1	**98**				324
all L2							2			2	3	1	3	16	**71**		1		3048

Table 11.17: Comparison of identification scores for German /y:/

/œ/ (%)	ɑ	ɑ:	ɛ:	ɛ	e:	ɪ	i:	ɔ	o:	ʊ	u:	**œ**	ø:	ʏ	y:	‹ei›	‹eu›	‹au›	#
Albanian			4					2	1		1	**61**	27	1	1				205
Arabic			2	5	1		1	6	1	8	1	**51**	6	15	2	1	2	1	177
English				1				1	2	10	4	**49**	15	9	3		4		255
Farsi			8	3	1			14	7	1	1	**49**	11		1		3		72
Hungarian												**93**	7						411
Mandarin				2	1			3	10	4	2	**41**	26	7	3		2	1	105
Polish		1	8	20	12	1	1	2	1	1	1	**21**	8	13	6	1	3		547
Romanian			3	1								**79**	12				2		213
SBC			1	2	2			5	2	2	1	**66**	11	8	2				584
Turkish												**66**	31				2		414
German												**98**	2						324
all L2			2	5	3			2	1	2	1	**59**	14	6	2		1		3037

Table 11.18: Comparison of identification scores for German /œ/

The most wide-spreading range of perceptual substitutions is observed for /œ/ though the preferences for specific response categories vary considerably across the language-subgroups: The most attractive response option for Albanian, English, Hungarian, Mandarin, Romanian, SBC and Turkish listeners is /ø:/, whereas Arabic and Polish listeners choose /ʏ/ most frequently, and Farsi listeners prefer /ɔ/ over all other categories. Hungarian, Romanian and Turkish listeners show only substitutions with /ø:/ but never confusions with front categories or back u- or o-qualities (with a few exceptions of substitutions with <ä:> and /ɛ/ in the Romanian sample). Substitutions with front /ɛ/, /e:/ and /ɛ:/ are most frequently observed in the Polish sample, but Polish hardly ever selects i-vowels for /œ/-stimuli. In the Albanian, the Arabic and the Farsi sample, substitutions with /ɛ/ or /ɛ:/ <ä:> are occasionally selected, whereas i-vowels do not seem to be attractive response options for /œ/. Id_scores above 5.5%

are moreover obtained for /ɔ/ in the Arabic and Farsi samples and for /ʊ/ in the Arabic and the English samples. However, the wide spread of responses over different categories observed in all samples except Hungarian, Romanian and Turkish shows that /œ/ is one of the perceptually least stable categories in German due to the combination of palatal and labial features together with a rather low degree of labiality and its relatively short intrinsic duration.

/ø:/ (%)	ɑ	ɑ:	ɛ:	ɛ	e:	ɪ	i:	ɔ	o:	**ʊ**	u:	œ	**ø:**	ʏ	y:	‹ei›	‹eu›	‹au›	#
Albanian			1					1	3			14	**58**	5	17				204
Arabic			1							2	6	9	**45**	19	17		2		176
English				1					1	9	7	19	**28**	12	19		4		264
Farsi									1	3	6	21	**17**	10	38		6		72
Hungarian												1	**87**	2	8				409
Mandarin									6	1	7	9	**43**	8	21		5		106
Polish					1	1	1		1	4	13	6	**16**	15	41				554
Romanian										1	2	9	**55**	7	23		1		216
SBC								1	1	1	5	6	**32**	12	42				591
Turkish											3	1	**56**	5	32		2		414
German												1	**98**		1				324
all L2									1	2	5	7	**45**	9	28		1		3060

Table 11.19: Comparison of identification scores for German /ø:/

/ø:/ is the category that causes most difficulties to Arabic, English, Farsi, Mandarin, Polish, Romanian, SBC and Turkish listeners. The high percentage of wrong responses for /ø:/ is due to confusion within the class of labio-palatal vowels as well as to substitutions with back vowels /u:/, /ʊ/ and /o:/ but hardly ever /ɔ/. Two substitution patterns are specifically prominent: substitutions with short unconstricted /œ/ or /ʏ/ and substitutions with /y:/. Long constricted /y:/ exerts a high attraction effect for /ø:/-stimuli (see Table 11.19), an asymmetric pattern that is considerably less common in the opposite direction (see Table 11.17), where /y:/ is mainly substituted with /ʏ/. Here, we find a strong pattern of asymmetry in all language sub-samples. The Arabic, English, Farsi, Mandarin and Polish sample show moreover substitutions with u- and o-qualities above the 5.5% mark. Occasional substitutions with front vowels occur in the Albanian, Arabic, English, and Polish sample, where responses may partly be induced by orthographic influence of <ä> in the Albanian and Arabic data.

To summarize, front rounded vowels show a high percentage of wrong identifications and considerable perceptual instability for L2 listeners in most language samples. Perceptual substitutions within the class of front rounded vowels, with back rounded vowels but also with front non-rounded vowels are observed in the data. The occurrence of front rounded vowels in the listeners' native languages reduces this instability and the percentage of wrong identifications to a certain degree. The data moreover provides evidence that listeners are to some extent sensitive to intrinsic duration as an indicator of phonemic length causing a higher portion of substitutions with a long or short equivalent respectively (for discussion of intrinsic duration and its influence on error rate see section 5.4.5).

11.2 Substitution processes and perceptual bias

In L2 acquisition, the learners' set of L2 sound intentions as well as perceptions is defined by their mental representations of the L2 phonological system. Some of these representations might be incomplete or wrong due to the influence of the L1 system, the lack of L2 language experience, the listeners' (temporary) hypotheses, and last but not least by misperceptions of L2 sounds. In all these cases, perceptual substitutions described in terms of phonological processes may result in an incorrect reconstruction of the original sound intention.

Perceptual substitutions can be described in terms of *phonological processes* defined as responses to phonetic difficulties (Donegan & Stampe 1979: 136) resulting in regular sound substitutions. "Native speakers" of the same language or language variety are expected to be consistent in their speech behaviour and the application or suppression of phonological processes. Processes affect speaker's perception as well as their production. In L2, processes account for deviations in L2 production as well as for misperceptions, i.e. wrong identifications of L2 sounds. Processes are influenced by inter-lingual and intra-lingual similarity of specific L1 or L2 sounds resulting in sound substitutions. The complex interactions of processes in production and perception are particularly relevant to explain the eventual development of representations for vowel contrasts in language acquisition.

Processes are considered to be "innate", "natural" and thus "universal" in the sense that they are motivated by characteristics of the human articulatory and auditory system. They are therefore cross-linguistically recurrent in the structures of the world's languages, in first language acquisition and in language loss (Jakobson 1968, 1971; Greenberg 1966, 1978; Maddieson 1984, 1997, 1999, 2008a, b; Ladefoged & Maddieson 1996) and are also expected to apply in foreign language acquisition. Processes are mental in that they apply subconsciously as responses to physiological limitations (e.g. limitations of vowel articulation due to the form of the oral cavity) or perceptual limitations (e.g. when certain contrasts are more difficult to perceive). These limitations exert "pressure" towards the occurrence of specific sounds in the phonemic system of a given language. The response to such pressure is the application of (fortition) processes "*to limit the universe of intendable, perceivable sounds to a phoneme inventory, for each language – i.e. a set of perceptually idealized (prototypical) sounds (…), an instance of prototypicality effects*" (Nathan & Donegan forthc: 5).

Processes may be limited in *feature-specific* ways, applying to *classes of sounds*. Classes are conceived as mental categories limited in feature-specific ways. As processes are conceived to be "universal" in the sense that they are motivated by physiology, e.g. by capacities and needs of the vocal tract or the human auditory system, common tendencies or biases are expected in L2 perception that operate along implicational conditions. In other words, Natural Phonology attributes perceptual and/or articulatory motivation to *biases* in the application of processes.

The discussion of confusion patterns and asymmetries in perceptual substitutions observed in the present data is partly inspired by Donegan's (1978) classification of vocalic fortition processes. It is, however, adapted to the articulatory-acoustic approach that has been discussed in section 4.7.5, 4.7.6, and 5.4 and visualized in Figure 4.3 and Figure 5.24 (schematic representation of the German vowel system). This model classifies German vowels along four to five constriction locations in the vocal tract (see Wood 1979, 1982, 1986): pharyngeal /ɑ ɑ:/, uvular /ɔ o:/, velar /ʊ u:/ and palatal vowels, which are further divided into mid-palatal /ɛ e: ɛ/ and /œ ø:/ and pre-palatal /ɪ i:/ and /ʏ y:/ (see section 5.4 and 5.6).

A full model describing the conditions for process application for vowels is presented by Donegan (1978) in her work "On the Natural Phonology of Vowels". Donegan's model is based on the assumption of implicational conditionsfor processes applying within a given class of vowels, e.g. rounded vowels or front high vowels. She posits two major dimensions of vowel classification: (1) *sonority* or vowel *height* and (2) *colour*, which is further divided into *labiality* and *palatality*. Donegan (1978) moreover describes three pairs of vocalic fortition processes (see section 4.7.2):

(1) Tensing & Laxing

(2) Lowering & Raising

(3) Bleaching and Colouring

However, as argued in section 4.7.5, 4.7.6, 5.4, 5.6 and 5.8, this terminology has some shortcomings with respect to the phonological labels and their articulatory. This refers especially to the process types Raising & Lowering and Tensing & Laxing.

Within the articulatory-acoustic approach of the present study, the terms [high]/[low] and [tense]/[lax] are avoided. [tense] and [lax] are replaced by a differentiation of [long, constricted] vs. [short, unconstricted] vowels when referring to contrasts of the type /i:/ - /ɪ/, /e:/ - /ɛ/, /u:/ - /ʊ/, /o:/ - /ɔ/, /y:/ - /ʏ/ and /ø:/ - /œ/. For the contrast pairs /ɑ:/ - /ɑ/ and /ɛ:/ - /ɛ/ only a differentitation in terms of phonemic length is assumed here, whereas the former contrasts are described as differing in terms of constriction degree and phonemic length.

Substitutions involving changes in constriction degree and/or phonemic length will be referred to as *Lengthening* and *Shortening* processes.

Substitutions involving changes in *Labiality* and/or *Palatality* (section 11.2.3) will be referred to as *(De-)Labialization & (De-)*Palatalization alternatively to Donegan's terms of Bleaching and Colouring. In Donegan's approach, Laxing & Tensing and Bleaching and Colouring are considered to be functionally parallel to each other and to share many implicational conditions (see section 4.7.2).

In Donegan's framework, Raising and the opposed process of Lowering refer to changes in "*vowel height by one degree*". While "Raising" "*increases vowel height by one degree* (…) *[n + 1 high]*" (Donegan 1978: 77), "Lowering" "*makes high vowels mid and mid vowels low*" and "*increases sonority by decreasing the height of vowels by one degree*" (Donegan 1978: 68). Donegan (1978: 77ff) posits implicational conditions of [+chromatic], [+tense] and [!lower] for Raising, whereas vowels which are [!achromatic], [!lax], [!mixed] (i.e. labial and palatal), and [!long] are more susceptible to Lowering. Raising appears to affect "chromatic" vowels, i.e. palatal, labial or labiopalatal vowels, but not plain ones.

Laxing & Tensing and Bleaching and Colouring are considered to be functionally parallel to each other and to share many implicational conditions (see section 4.7.2)

In the Natural Phonology framework, the traditional feature "Height" is conceived as a scalar rather than a binary feature (Donegan 1978: 82). Height, – whether binary or scalar – is very much influenced by the representation of the vocalic space in terms of the quadrilateral IPA vowel chart. The traditional IPA quadrilateral, however, is not a precise model of articulatory constellations in the vocal tract especially for the class of back vowels, where Height cannot be considered as a precise construct for articulatory differences between u-, o- and a-vowels (as discussed in section 4.7.5 and 4.7.6). Rather, it is the place of narrowing or constriction in the vocal tract that differentiates these qualities and determines acoustic resonances, i.e. velar u-vowels are located at a constriction location that is more front than o- or a-vowels.

Alternatively, in the present study, and in accordance with the proposed description of the German system in section 5.6 in terms of constriction locations as visualized in Figure 5.24 and Figure 11.7, the process types of *Fronting and Backing* are proposed here for changes in *constriction location* (see section 11.2.1).

The present approach differs from more traditional accounts specifically with respect to the differentiation of a pre-palatal and a mid-palatal qualities constriction location in German. In this sense, contrasts like /e:/ - /ɪ/ or /ø:/ - /ʏ/ strongly refer to a difference in Frontness (pre-palatal vs. mid-palatal) together with constriction degree & phonemic length and less to differences in “tongue height” (see section 5.4.1 and 5.4.4; for a critical discussion of “tongue height”, see also Ladefoged 1993: 221).

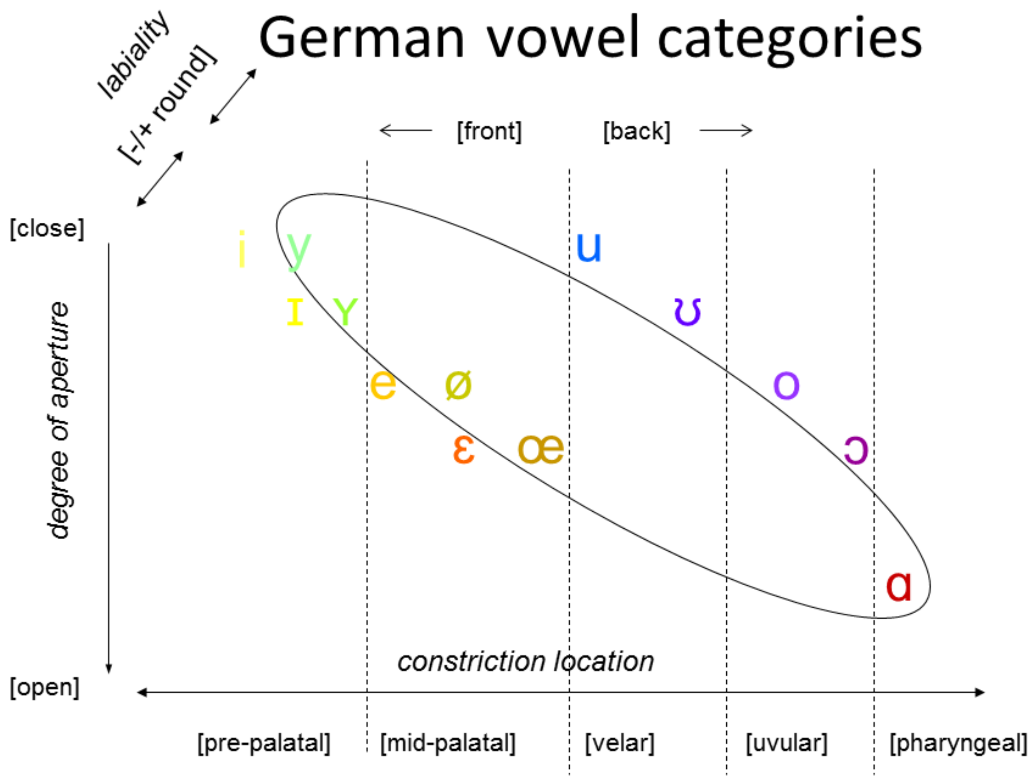

Figure 11.7: Schematic representation of the German vowel system (Figure 5.24)

Also, contrast pairs such as /y:/ - /u:/ can be treated in analogy with /u:/ - /o:/ as differing in Frontness. The schematic diagram Figure 11.7 reflects the articulatory proximity of i-/ü-vowels and u-vowels that is not evident in the IPA quadrilateral. The IPA chart suggests maximal distance of front i-/ü- and back u-vowels, though in fact velar u-vowels are nearest to the front palatal constriction location. The proposed representation also accounts much better for substitutions of the type /y:/ > /u:/ but also for the context-specific occurrence of [y]-like allophones for /u/ e.g. in several varieties of English, than the traditional quadrilateral IPA chart.

The analysis of confusion patterns in the current data will search for common tendencies in L2 perception in the full sample of all L2 listeners and will describe them in terms of perceptual substitution processes. The full L2-sample consisting of 46,694 responses by 173 non-native listeners from ten languages will serve here as data base to search for “universal” tendencies in vowel perception. For each vowel category around 3,110 responses are available.

In the following sections, three major types of perceptual substitution processes observed in the L2 data will be discussed in terms of

(1) substitutions involving changes in constriction degree and phonemic length (section 11.2.1 and Figure 11.10) referred to here as *Lenghtening & Shortening*,

(2) substitutions involving changes of constriction location (section 11.2.2 and Figure 11.11) referred to here as *Fronting & Backing*, and

(3) substitutions involving changes in labiality and/or palatality (section 11.2.3 and Figure 11.12) referred to as *(De-)Labialization & (De-)Palatalization*[135].

Figure 11.10, Figure 11.11 and Figure 11.12 summarize substitutions for each of these three types. The percentages correspond to the identification scores (id_V) in the confusion matrix for all L2 listeners (cf. Table 9.5) and inform about the percentage of cases where input stimuli of a particular category were identified as belonging to a specific category by non-native listeners. To differentiate more frequent from less frequent substitutions, id_scores below 5.5% are printed in grey, id_scores below 0.5% are not indicated.

11.2.1 Constriction degree and phonemic length

The largest portion of substitutions by L2 listeners concerns changes in constriction degree and phonemic length. Figure 11.10 summarizes all perceptual substitutions where input stimuli and response category differ only in phonemic length (and constriction degree). This set of substitutions refers to cases where listeners show difficulties in the differentiation of phonemic length even if they are able to identify the place of constriction correctly.

A one-way ANOVA shows that there are significant differences in percentage correct identifications for long vowels as well as for short vowels ($p < 0.001$) and that the mean sum of squares is 0.549 for long vowels but only 0.098 for short vowels indicating that vowel category explains 54.9% (for long vowels) and 9.8% (for short vowels) of the variance between vowel categories. A comparison of long vs. short vowel categories in the identification data shows strong differences in id_wrong patterns within the group of long vowels, where /ɛ:/ and /ø:/ show the highest percentage wrong identifications (see Figure 11.8).

In German, phonemic length is closely associated with differences in constriction degree for most contrasts except /ɑ/ - /ɑ:/ and /ɛ/ - /ɛ:/ (see section 5.4). *Duration* is an important cue to German vowel identification, especially for qualities with similar spectral shape, where

---

135 These process types correspond to Donegan's Colouring & Bleaching.
Processes of Palatalization & Depalatalization are equivalent with Fronting & Backing processes (section 11.2.1).

duration can have a disambiguating effect (Weiss 1972, 1976, 1978; Wängler & Weiss 1975; Sendlmeier 1981).

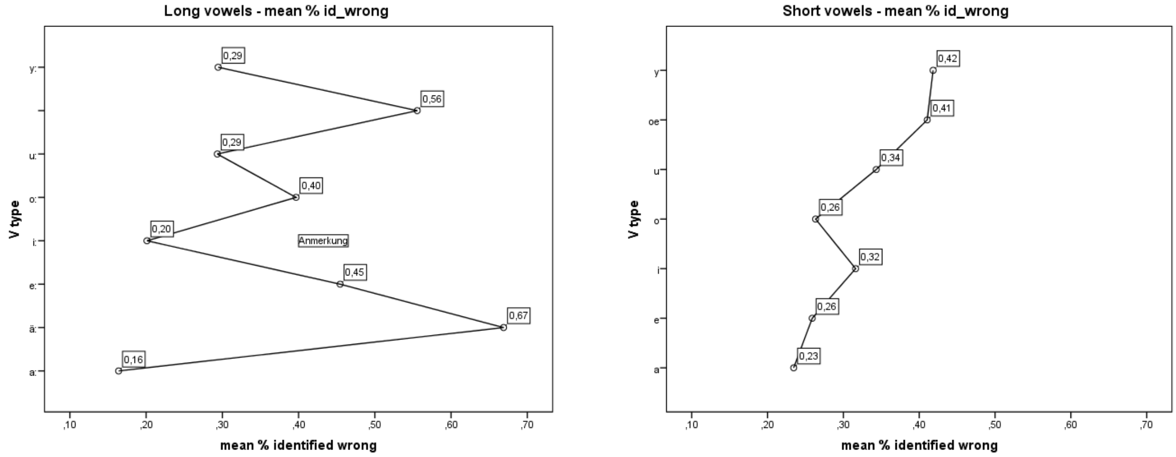

Figure 11.8: Percentage wrong identifications of long and short German vowels (173 non-native listeners)

Bohn (1994) could show that for German native listeners vowel quantity is more influential for differentiating vowels differing along a phonetic continuum than for English native listeners. German listeners use durational cues also in foreign language perception. But the use of durational cues is apparently not limited to languages that have phonemic length contrasts. Several L2 studies (e.g. Bennett 1968; Bohn 1994, 1995; Escudero 2002, 2005; Flege, Bohn & Jung 1997; Ćavar, Rydzewski & Oštarić 2012) show that non-native speakers of different language backgrounds show a strong tendency to rely on durational cues to differentiate vowels that are spectrally similar, especially in earlier stages of acquisition, even if their native language has no phonemic length distinction. Bohn (1995) suggests a kind of language-universal strategy that makes listeners use durational cues when spectral cues are not sufficient: "*[W]henever spectral differences are insufficient to differentiate vowel contrast because previous linguistic experience does not sensitize listeners to those spectral differences, durational differences will be used to differentiate the non-native vowel contrast*" (Bohn 1995: 294f).

In other words, intrinsic duration may be used to differentiate some spectrally similar vowels, e.g. German /e:/ - /ɪ/, /o:/ - /ʊ/ or /ø:/ - /ʏ/. On the other side, the use of phonemic distinctions in length and constriction degree in the German system causes considerable intra-language confusions between short and long categories that are located at the same constriction location but differing in mean duration and degree of constriction. Constricting at a given location in the oral cavity causes specific acoustic changes and makes vowels apparently "higher" in

perception. This effect underlying substitutions of the type /e:/ > /ɪ/ or /e:/ > /i:/ will be discussed in the following section.
Substitutions of short/long stimuli by their long/short counterpart at a given constriction location will be referred to as Lengthening or Shortening respectively. In most cases – except for the contrasts /ɑ:/ > /ɑ/ and of /ɛ/ > /ɛ:/ - substitutions involve not only changes in phonemic length but also in constriction degree.

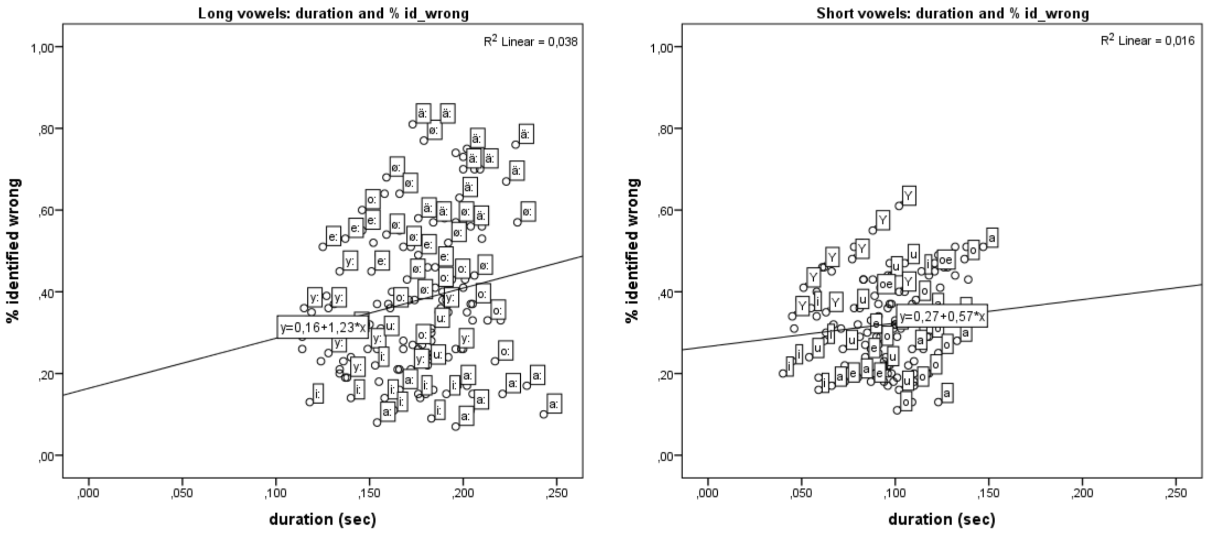

Figure 11.9: Id_wrong scores for long vs. short categories (173 L2 listeners)

Figure 11.9 represents the id_wrong scores for long categories (left) and short categories (right) showing that categories differ considerably in terms of duration as well as id_wrong scores.
Figure 11.10 summarizes all substitutions where input and output differ only in phonemic length (and constriction degree), whereas substitutions of the type /e:/ > /ɪ/ or /o:/ > /u:/ involving changes in constriction location are summarized in Figure 11.11.
The results show that categories differ in their susceptibility for the processes of Shortening and Lengthening. The data show (1) quantitative differences with respect to vowel type and (2) asymmetries in the direction of substitution.
Differences in the susceptibility of categories to Lengthening and Shortening seem to be related to degree of aperture (in traditional teminology “height”). Lengthening seems to occur more frequently with lower than with higher vowels. This is particularly evident with back categories: Short [ɑ] is more frequently identified as [ɑ:] (24%) than [ɑ:] > [ɑ] (only 16%). Lengthening of non-high /ɔ/ occurs in 20% of its occurrences, but in only 13% of the cases for high /ʊ/. Perceptual Shortening occurs less frequently with back open vowels (/ɑ:/ > /ɑ/ 15%,

/o:/ > /ɔ/ 8%), but is equally frequent for close high u-qualities: /u:/ > /ʊ/ (15%) vs. /ʊ/ > /u:/ (13%).

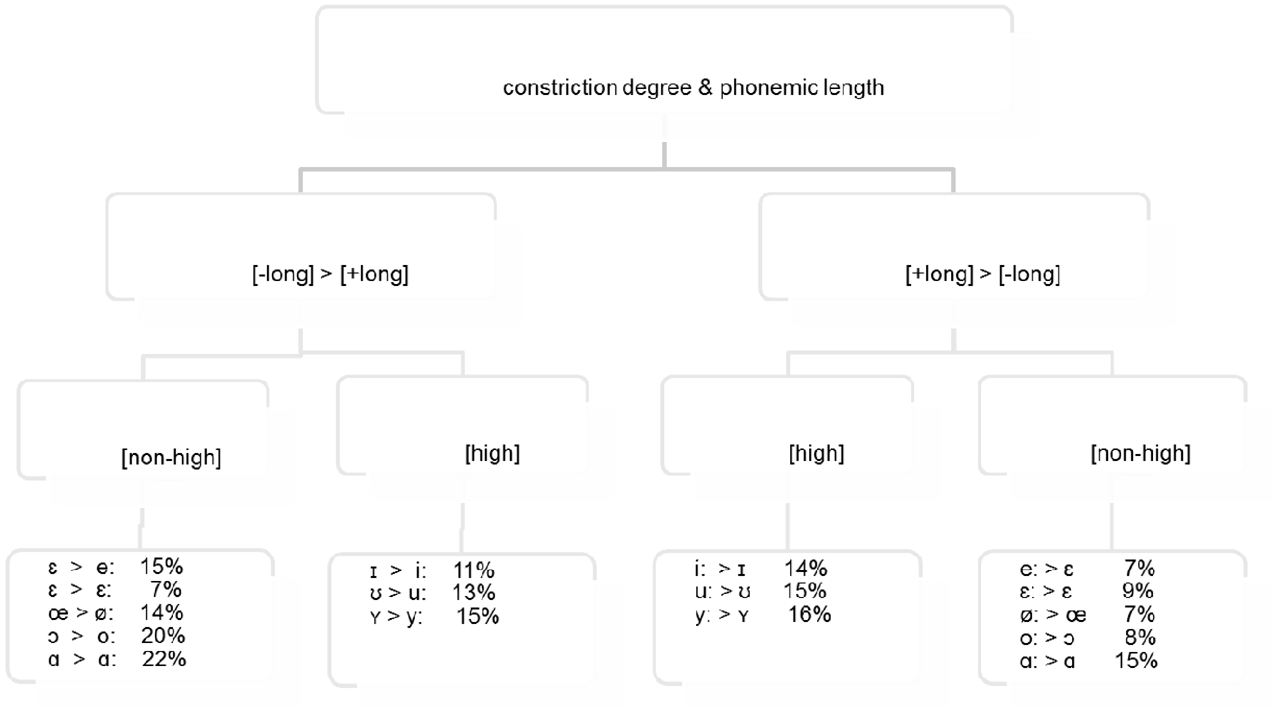

Figure 11.10: Perceptual substitutions involving changes in constriction degree and phonemic length (% of occurrence in the "all L2" sample)

This difference between categories with respect to perceived length and degree of aperture is weaker for the class of front vowels, presumably due to the existence of front rounded vowels and the special status of /ɛ:/. /ɛ:/ is identified as /ɛ/ in 9% of its occurrences, whereas the opposite substitution type of /ɛ/ > /ɛ:/ occurs in only 7% of the cases.

Lengthening of high short unconstricted /ɪ ʊ ʏ/ occurs only slightly less frequently than Shortening of their respective long constricted counterparts /i: u: y:/, though these differences are far from significant. A stronger asymmetry with respect to direction of substitutions is observed for non-high e-, o- and ö-vowels. Substitutions of short unconstricted non-high /ɛ/, /œ/ and /ɔ/ with their long counterparts occur more frequently than substitutions in the opposite direction: /ɛ/ > /e:/ (15%) vs. /e:/ > /ɛ/ (only 7%), /ɔ/ > /o:/ (20%) vs. /o:/ > /ɔ/ (only 8%), and /œ/ > /ø:/ (14%) vs. /ø:/ > /œ/ (only 7%). In other words, L2 listeners show a tendency to identify non-high short unconstricted /ɛ ɔ œ/ more frequently as [+long] and [+constricted] than they would identify long constricted /e: o: ø:/ as [short] and [-constricted]. This tendency is not observed for high i-, ü- and u-vowels.

To account for these asymmetries, phonetic characteristics as well as more general cognitive factors such as hypercorrection (see section 1.3.2) in response behaviour and a preference for “marked” [+long] categories may be assumed.

High i-, ü- and u-vowels are intrinsically more than 20 ms shorter than e-, o- and ö-qualities (see Table 11.20 and Table 5.6). The higher intrinsic duration of short /ɑ/, /ɔ/, /ɛ/ and /œ/ due to their higher degree of aperture may explain why stimuli belonging to these categories tend to be more susceptible to perceptual Lengthening than /ɪ/, /ʏ/ and /ʊ/ respectively. However, difficulties with /ɪ/, /ʏ/ and /ʊ/ are also related to their acoustic similarity with /e:/, /ø:/ and /o:/, respectively. The contrast pairs /e:/ - /ɪ/, /ø:/ - /ʏ/ and /o:/ - /ʊ/ show considerable overlap in spectral cues that are differentiated only by duration or by lexical or morphological context information.

duration (ms)	short			long		
high	ɪ	ʏ	ʊ	i:	y:	u:
	74	83	87	150	151	157
non-high	ɛ	œ	ɔ	e:	ø:	o:
	100	109	108	175	184	186
			ɑ	ɛ:		ɑ:
			104	192		197

Table 11.20: Mean duration (ms) of short and long stimuli

In the present data, long constricted /e:/, /o:/ and /ø:/ show a strong tendency to be substituted by “higher” or more “fronted” /i:/, /u:/ and /y:/ rather than being substituted by their short, less similar counterpart, as discussed in the following section (see Figure 11.11).

The listeners’ knowledge about the existence of “long” vs. “short” categories interferes with these acoustic similarities – of which (phonetically non-trained) learners of German are not aware – causing substantial variation in the listeners’ responses (see the vowel-specific and language-specific analyses in the previous chapters).

The differentiation of long constricted vs. short unconstricted vowels in German and perceptual “boundaries” between acoustically similar long constricted and short unconstricted vowels at different locations in the vocal tract are a research field that lends itself to the experimental paradigm of *rectangular variation* to identify *perceptual boundaries* between categories. In this type of study, synthetic or synthetically manipulated natural stimuli are varied systematically along a continuum of two acoustic dimensions by systematic manipulation of e.g.parameters such as duration and F1 or F1-F0. The listeners’ responses can be subject to regression analyses (cf. Figure 11.2) revealing the boundaries between stimuli that are perceived as "long" vs. "short" or "more open" vs. "closed" as a function of specific acoustic properties (e.g. F1, F0, F1-F0, duration) varying along a continuum (e.g. Escudero 2002; Ćavar et al. 2012).

It is however necessary to be aware of the fact that "boundaries" between categories are a theoretical construct (see section 12.1.3). Phonetically, such perceptual "boundaries" between categories correlate with a bundle of spectral and durational cues due to variation in constriction location, constriction degree and the relative duration and degree of aperture. The interaction of constriction location, duration and intrinsic pitch was discussed in section 5.4.4 and 5.4.5. In L2, the interpretation of acoustic cues may not only vary by category distribution in the listeners' native language or language variety but even by the L2 variety to which L2 learners have been exposed to (cf. Escudero 2002; Williams 2013). The data of the present study provides precious evidence for hypotheses on the language-specific and category-specific interplay of intrinsic duration, intrinsic pitch and degree of aperture in further research (as discussed in section 5.5).

To conclude, while *duration* may be used to differentiate some spectrally similar vowels (by native as well as non-native listeners, cf. Bohn 1995), e.g. German /eː/ - /ɪ/, /oː/ - /ʊ/ or /øː/ - /ʏ/, the existence of phonemic distinctions in length and constriction degree in the German system also contributes to inter-category confusions between long vowels that will be discussed in the next section. While vowels differentiated by length and constriction degree at a given constriction location in German, e.g. /eː/ - /ɛ/, /yː/ - /ʏ/ or /oː/ - /ɔ/ hardly overlap in acoustic dimensions (see the F1xF2 representations in Figure 12.7), contrasts of "neighbouring" /eː/ - /ɪ/, /øː/ - /ʏ/ or /oː/ - /ʊ/ show considerable overlap between categories articulated in adjacent constriction locations. Substitutions of the latter type are discussed in the following section.

11.2.2 Constriction location

Figure 11.11 provides an overview of perceptual substitutions involving changes in constriction location, e.g. mid-palatal /eː/ perceived as pre-palatal /iː/ or /ɪ/, or uvular /oː/ perceived as velar /uː/, and velar /uː/ perceived as front labial /yː/, but also inverse processes like /ɪ/ > /ɛ/ or /ʏ/ > /œ/ for the "all L2" sample. In the present framework, these substitution types will be referred to as *Fronting* and *Backing* involving substitutions where targets of perceptual substitutions are located at a constriction location that is either more front or more back, respectively, than the original input.

In general, the perceptual identification of constriction location seems more difficult for short unconstricted qualities, especially if round. Considerable difficulties are observed in differentiating /ʏ/, /œ/ and /ʊ/. Difficulties with these categories may be accounted for by the shortness of the quasi-stable part in the signal in relation to CV- and VC-transitions in the signal due to lower intrinsic duration particularly for short unconstricted /ʏ/ and /ʊ/ (see Figure 5.16) and the additional "damping" effect of lip rounding.

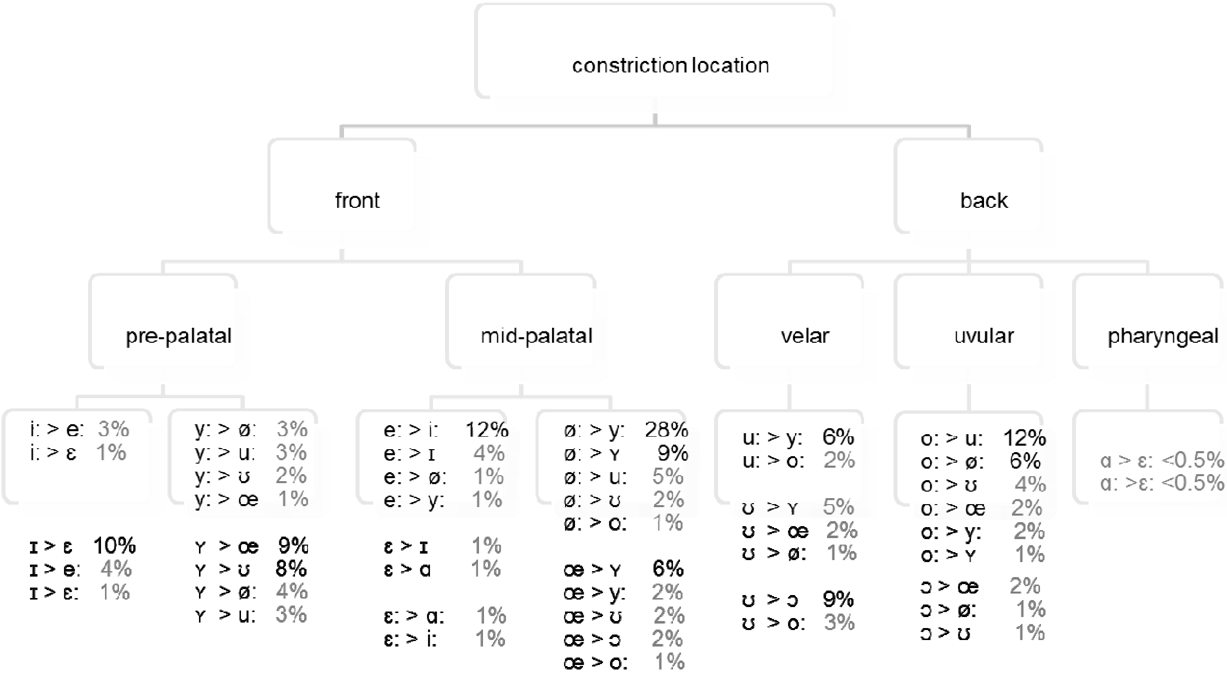

Figure 11.11: Perceptual substitutions involving changes in constriction location (% of occurrence in the "all L2" sample)

The perception data show category-specific differences and asymmetries with respect to Fronting and Backing: In general, the data show that mid-palatal *e*- and *ö*-qualities and uvular *o*-qualities are more frequently subject to changes in constriction location than pre-palatal *i*- and *ü*-qualities or velar *u*-qualities. Substitutions of *i*-stimuli with *e*-qualities are rare as are substitutions of *a*-stimuli with front [ɛ:][136]. Generally, *i*-qualities and *a*-qualities are least frequently affected by changes in constriction location. These observations conform with the assumption that peripheral vowels are easier to identify than non-peripheral ones (for discussion of peripherality see section 12.3.3.1).

[136] These cases are presumably co-conditioned by orthographic similarity of <a> and <ä> and considerable allophonic variation in some languages, e.g. Arabic, Farsi or English (see Table 11.2 and Table 11.3).

Fronting – in traditional terminology Raising – refers to substitutions where the target category is situated at a more fronted constriction location than the input, such as /e:/ > /i:/, /e:/ > /ɪ/, /ɛ/ > /ɪ/, /ø:/ > /y:/ or /o:/ > /u:/ ~ /ʊ/. In all these cases the target is more fronted than the input (see Figure 11.7).

Backing is the opposed process – "Lowering" in traditional accounts – and refers to substitutions of e.g. /u:/ > /o:/, /ʊ/ > /ɔ/, /ɪ/ > /ɛ/ or /ʏ/ > /œ/, because their targets are located at a constriction location that is more back than their input.

The advantages of a description in terms of the process types Fronting and Backing are their *larger scope* and *articulatory grounding*, without having to refer to the notion of "tongue height"[137].

Fronting can account for a larger set of processes involving changes in constriction location of e.g. mid-palatal /e:/ > pre-palatal /i:/ or /ɪ/ and of uvular /o:/ > velar /u:/ or /ʊ/ but also for /ø:/ > /ʏ/ or /u:/ > /y:/, where /y:/ is cannot be describes as "higher" with respect to tongue height but is however more fronted. Substitutions of the type /u:/ > /y:/ (6% in the current data) are a very good example for Fronting as observed e.g. in several English varieties where [y]-qualities are allophonic variants for /u/ in alveolar contexts (see Harrington et al. 2008; Harrington et al. 2011).

Another general tendency observed in the data is a higher preference for Fronting with *long constricted* qualities, i.e. substitutions by a category articulated in a more fronted area of the vocal tract, whereas short unconstricted vowels are rather subject to Backing, i.e. to substitutions with "less fronted" qualities. This type of asymmetry is found to a variable extent in all language sub-groups. While Fronting is observed for [e:] > [i:] in 12% of the cases, [ø:] > [y:] in 28%, and [o:] > [u:] in 12%, it is less common with short unconstricted qualities ([ɛ] > [ɪ] 1%, [ɔ] > [ʊ] 1%, [œ] > [ʏ] 6%). The inverse pattern is also rather uncommon: Backing of [i:] > [e:] occurs in 3%, [y:] > [ø:] in 3%, and [u:] > [o:] in 2% of the cases.

Short unconstricted qualities, on the other hand, are rather frequently substituted by less fronted qualities: e.g. pre-palatal [ɪ] > mid-palatal [ɛ] or [e:] (4% both). This tendency is particularly common with rounded vowels, for which stronger patterns of asymmetry are

[137] The IPA vowel chart suggests that points in the vowel quadrilateral represent the position of the highest point of the tongue. However, a representation of vowels in terms of the highest point of the tongue would require a different form of diagram. Moreover, the position of the highest point of the tongue cannot be considered to indicate vowel quality in an adequate way. Ladefoged (1993) therefore avoids the term "tongue height" and uses instead the term "vowel height" by which an "*auditory quality that can be specified in acoustic terms rather than in articulatory terms*" is meant (Ladefoged 1993: 221).

observed: e.g. [ʊ] > [ɔ] (9%) vs. fronting of [ɔ] > [ʊ] (only 1%); [ʏ] > [œ] (9%) vs. [œ] > [ʏ] (6%).

However, the asymmetry observed with respect to constriction location, length and the occurrence of Fronting vs. Backing is an interesting pattern that cannot be explained by intrinsic duration alone. Other implicational conditions formulated by Donegan (1978: 77ff) for Raising as [+chromatic], [!tense/constricted] and [!lower] prove to be relevant for the occurrence of Fronting in the present data.

[!long] as condition for Lowering (Donegan 1978: 71) does not apply to the data in the sense of phonemic length, which would contradict Donegan's [!tense/constricted] condition for Fronting, but rather [!long] can be understood in the sense of "intrinsically longer" (for discussion, see Donegan 1978: 71ff).

The process of *Backing* works in the opposite direction and accounts for substitutions like /ɪ/ > /ɛ/ or /ɪ/ > /e:/ and /ʊ/ > /ɔ/ or /ʊ/ > /o:/. Donegan's implicational conditions for Lowering [!lax], [!mixed] (i.e. [+labial, +palatal]) are confirmed by the present data, see e.g. various substitutions of /ʏ/, whereas her condition [!achromatic] referring to plain vowels like [ʌ] or [ɨ] (Donegan 1978: 68) is less relevant, since these categories do not exist in the German vowel inventory. One could however argue that categories with short intrinsic duration and a relative high portion of CV- and VC-transition may be perceived as "achromatic" by less experienced listeners, specifically in non-word conditions.

11.2.3 Labiality and palatality

A specific characteristic of the German vowel inventory is the set of *front rounded* vowels. In Donegan's approach they are described in terms of *palatality* referring to frontness and *labiality* referring to lip rounding and/or lip protrusion. Since there is no physiological opposition between palatality and labiality, palatal vowels can occur with or without lip rounding (for discussion, see section 5.4.2), where lip rounding is an additional rather than a primary articulatory gesture (Stevens & Keyser 2010[138]). By combining palatality and labiality in front vowel realizations, the acoustic effects of lip rounding as well as of constriction location in the front part of the vocal tract are attenuated. In other words, labio-palatals are less palatal than their front non-rounded counterparts and less labial than their back rounded

[138] Stevens & Keyser (2010) refer to lip rounding and lip spreading not as defining gestures but rather as enhancing gestures for a distinctive contrast in a given language. Lip *rounding* in non-low back vowels is posited to enhance the contrast to non-low non-rounded front vowels. Non-low front vowels are often produced with lip *spreading* to strengthen the contrast to rounded back vowels.

counterparts. Front rounded vowels are commonly called labio-palatals, bichromatic or mixed (Donegan 1978: 37ff, 47).

Donegan (1978: 83ff) describes two process types, *Colouring* and *Bleaching* as processes that remove or add colour – palatality or labiality – directly, without changing vowel height. As Colouring and Bleaching are functionally parallel to Donegan's Tensing and Laxing processes, they share many of their implicational conditions: Bleaching, i.e. the removing of Labiality or Palatality by processes of Delabialization or Depalatalization, refers to the conditions [!lower], [!lax], [!mixed], whereas the applicability of Colouring, i.e. processes of Palatalization or Labialization, is mainly determined by the condition [!higher].

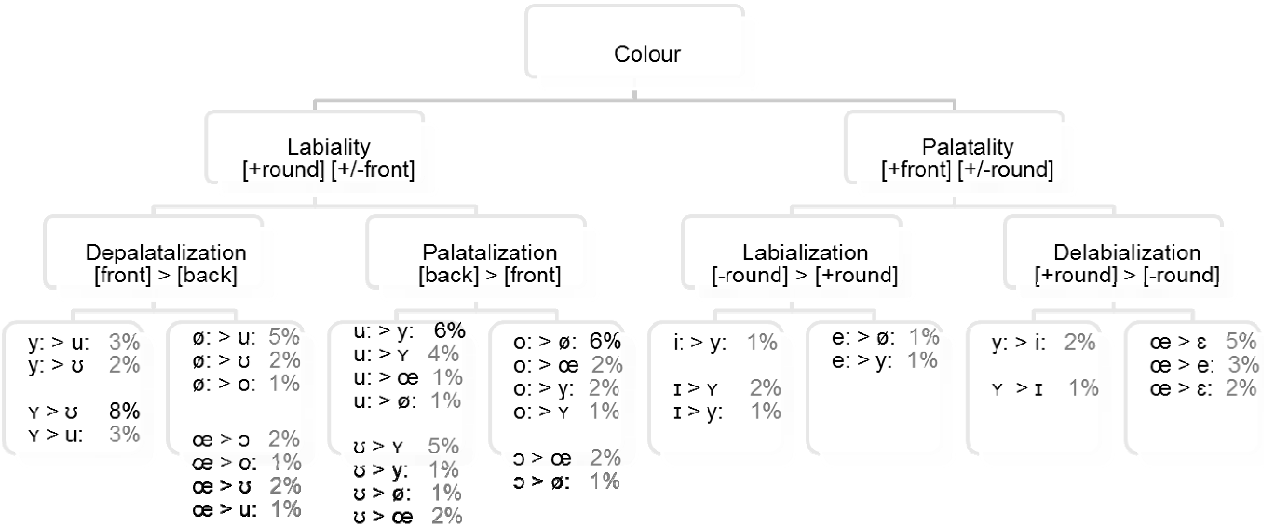

Figure 11.12: Perceptual substitutions involving changes in palatality or labiality (% of occurrence in the "all L2" sample)

Palatalization is considered here as parallel to the process of Fronting, whereas Depalatalization equals Backing. Labialization is conceived as addition of lip rounding or protrusion (see section 4.3.3 and 5.4.2). An inverse process of loss of Labialization is called Delabialization.

Figure 11.12 summarizes all perceptual substitutions in the L2 sample (> 0.5%), where a change either in palatality or labiality is involved. The left part of the diagram shows Fronting/Backing substitutions that are also represented in Figure 11.11 involving changes in constriction location, e.g. Backing of /y:/ > /u:/ or Fronting of /u:/ > /y:/.

Fronting, however, comprises also substitutions of the type /o:/ > /u:/ and /e:/ > /i:/, whereas Backing refers to substitutions of the type /i/ > /e:/, /ɪ/ > /ɛ/, /ʏ/ > /œ/, /ʏ/ > /ʊ/ or /ʊ/ > /ɔ/, where no changes in Labiality are given (for discussion, see section 11.2.2).

Perceptual *Delabialization* occurs mainly with /œ/, which is perceived as either /ɛ/, /e:/ or /ɛ:/ in 10%, and only occasionally fronted to /y:/ and /ʏ/.

Very few cases of perceptual *Labialization* of palatal vowels are observed (/i:/ > /y:/ 1%, /ɪ/ > /ʏ/ 2%), which are presumably mainly due to hypercorrection.

Instances of perceptual *Backing* or *Depalatalization*, i.e. substitutions of front (rounded) with back vowels, are observed for all four front rounded qualities /y: ʏ ø: œ/ and occur most frequently for /ʏ/ > /ʊ/ (8%) and /ø:/ > /u:/ (5%). In total, /y:/ is categorized as back vowel /u:/ or /ʊ/ in 5%, /ʏ/ in 11% (/u:/ or /ʊ/), /ø:/ in 8% (/u:/, /ʊ/, or /o:/) and /œ/ in 6% (/u:/, /ʊ/, /o:/ or /ɔ/).

Fronting or *Palatalization*, i.e. substitutions of back qualities with front rounded categories, are most frequently observed with /u:/ > /y:/ (6%), /o:/ > /ø:/ (6%), /ʊ/ > /ʏ/ (5%) and /u:/ > /ʏ/ (4%).

To account for these confusion patterns involving [rounded] vowels, phonetic as well as orthographic factors have to be considered. At first glance, these substitutions may largely be orthographically motivated by the graphemic similarity of <u> and <ü> or <o> and <ö>. However, phonetic reasons may influence the listeners' responses as well.

In the first place, the effects of lip rounding and/or lip protrusion decrease with an increase of degree of aperture. Within the class of front rounded vowels, /œ/ shows the lowest degree of lip rounding together with the highest degree of aperture, which explains the higher percentage of Delabialization of /œ/.

Another factor that has to be considered in the description of difficulties with front rounded vowels is the articulatory proximity of velar u-vowels and palatal ü- and ö-vowels, which is not obvious in the traditional IPA chart but evident in the articulation-oriented approach as represented in Figure 11.7. The IPA suggests a maximal distance between front i-/ü-vowels and back u-vowels. In fact, the maximal distance as assumed in the IPA quadrilateral cannot be confirmed by articulatory data (see Figure 11.7), where u-vowels are located nearest to front vowels situated in an intermediate position between front palatal and back uvular and pharyngeal vowels. This articulatory vicinity of u-vowels and palatal vowels is of relevance for the perception of velar u-vowels as well as palatal rounded vowels, particularly ü-vowels, and also accounts for articulatory variation in speech production. In many structurally diverse languages e.g. in Swedish, Albanian or different varieties of English, allophonic spreading in terms of Fronting of back vowels, particularly for u-vowels, is observed (for details, see section 10.5 for English).

For German due to the existence of /ʊ/ and /u:/ as well as /y:/ and /ʏ/ in the system and for other languages like French or Swedish that do have front rounded qualities, a higher degree of "peripherality" for /u:/ is assumed, whereas – as Harrington et al. (2011) suggest – /u/ may be more fronted in languages without phonemic contrasts like /y ʏ u ʊ/. Harrington et al. (2011) observed perceptual confusion by German native listeners between /ʊ/ and /ʏ/ spoken at fast rate. In their identification experiment, confusion of /ʊ/ and /ʏ/ occurred more frequently in the direction of /ʊ/ > /ʏ/ than /ʏ/ > /ʊ/ but the results varied context-specifically. In *tVt*-contexts substantially more /ʊ/ >/ʏ/ identification occurred, in *pVp*-contexts more /ʏ/ > /ʊ/ occurred and in *kVk*-contexts hardly any differences in direction were observed (for a detailed discussion of vowel perception in an English-German comparison, see section 10.5.3 and Table 10.25 and Table 10.26).

In L2 perception, *context effects* of L1 and L2 may interact. An analysis of the effect of context on the occurrence and direction of perceptual substitutions would require a study of context-specific effects on vowel identification in the listeners' L1 (which is not available for many of the languages analysed here) before testing the influence of context variation on the perception of specific German vowel categories. A detailed discussion of substitution patterns observed in the present data for each of the 14 different context conditions in each of the 10 language sub-samples is far beyond the scope of the present chapter. A context-sensitive general discussion of the data is presented in section 9.1.3. Where relevant studies for a given language are available, a context-sensitive discussion is presented in the language-specific analyses (see section 10.5 for native English listeners and 10.9 for native Polish listeners).

To conclude, substitutions or confusions of front rounded with back rounded vowels and back vowels with front rounded vowels are considered as instances of Backing and Fronting, i.e. of substitutions of front palatal vowels with back velar and uvular vowels and vice versa. It is moreover hypothesized that this type of substitutions mostly applies in cases where phonetic similarity and orthographic influences (e.g. the similarity of <o> and <ö>) interact. Further studies will have to test context-sensitivity with respect to place of articulation of adjacent consonants.

11.3 Asymmetries and perceptual bias

The data presented above show systematic differences in (1) category-specific perceptual *difficulty* of input categories, (2) listeners' *preferences* for specific response categories and (3) targets and direction of perceptual substitutions. Difficulties, preferences and directions of asymmetries were found to vary systematically.

Difficulties are indicated by the number of wrong identifications and refer to the category-specific probability of wrong identification (id_wrong scores). *Preferences* for specific vowel categories are indicated by the listeners' selection of specific response categories (pref_scores). The direction of *perceptual substitutions* in the listeners' identifications is indicated by perceptual substitutions (id_V scores) observed.

Asymmetries in *difficulties* are conceived as the greater probability of a given *input* category to be identified incorrectly. Asymmetries in *preferences* are observed when listeners select a specific *response* option more frequently than others. In other words, difficulty refers to input stimuli and preference to response options.

Asymmetries can partly be explained by *language-specific* factors as the occurrence and distribution of similar sounds in the specific region of the phonetic space in the listeners' L1 and in L2. However, similar patterns of asymmetry occur in several language sub-samples. Along the argumentation of the present study, these patterns are considered to be strongly motivated by phonetic and phonological factors referring to signal-inherent properties of the stimuli and/or characteristics of the stimulus category or the response category. Acoustic-phonetic properties of the stimuli such as spectral characteristics, vowel-intrinsic duration, intrinsic pitch or context-sensitve variation due to coarticulation can account for a considerable part of the listeners' responses. However, to account for *asymmetry* effects like substitutions of the type /ɛ:/ > /e:/ (49% in the "all L2"-sample) vs. /e:/ > /ɛ:/ (15%) a purely phonetic account is not sufficient and other more general factors have to be considered.

The following sections will briefly present evidence for asymmetries in differences, preferences and direction of substitutions.

11.3.1 Difficulties

Asymmetries in *difficulty* are conceived as systematic and significant inter-category differences in probability of a given category to be identified correctly, acquired and mentally represented according to language-specific norms and distributions. An indicator for the relative difficulty of German vowel categories in L2 perception and acquisition is the number of wrong and correct identifications (id_wrong/id_correct) within a given sample.

An inter-category comparison of id_wrong scores for each of the 15 input categories (18 tokens for each of the 15 vowel types) shows between-category differences in means as well as in standard deviation. Figure 11.13 compares the mean id_wrong scores of all German vowel categories for the sample of all L2 listeners (see also Figure 9.1, Figure 9.2, Figure 9.3 and Figure 9.4 in section 9.1). For short vowels, the lowest id_wrong scores (<30%) are obtained for short /ɑ/, /ɛ/ and /ɔ/, intermediate scores for short /ʊ/ (34%) and /ɪ/ (32%), and higher id_wrong scores above 40% are given for /œ/ (41%) and /ʏ/ (42%). In the class of long vowels, low id_wrong scores below 30% are observed for /ɑ:/, /i:/, /u:/ and /y:/, whereas high id_wrong scores above 50% are obtained for /ø:/ (56%) and /ɛ:/ (67%).

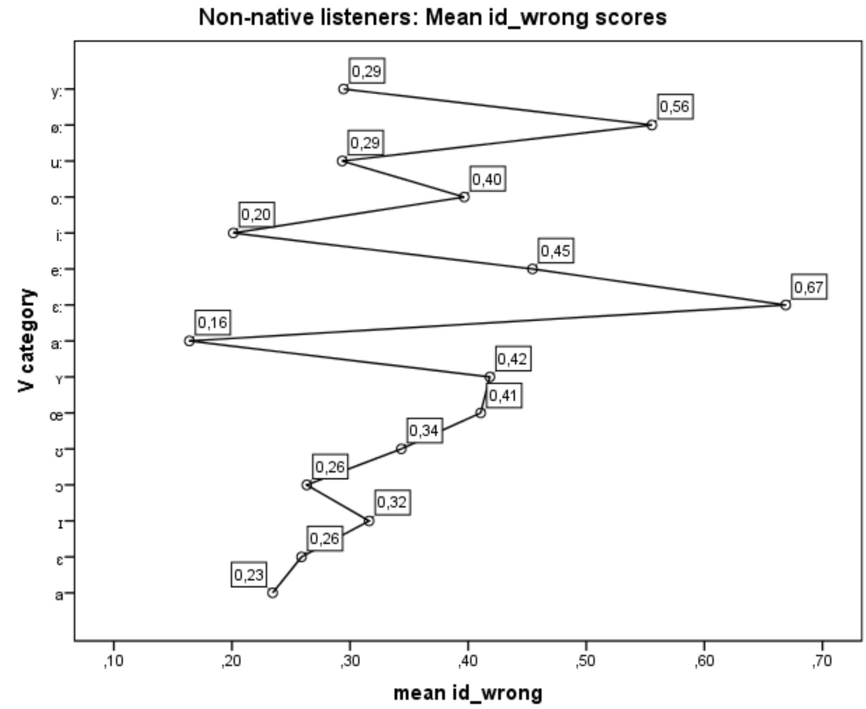

Figure 11.13: Mean id_wrong scores for long and short vowel categories (270 stimuli, 173 non-native listeners)

A one-way ANOVA shows that there are significant differences between categories in percentage of wrong identifications. The tests of between-category effects show a main effect of *vowel category* with an *F*-value of 47.879 that is statistically significant ($p < 0.001$) and a partial eta squared of 0.724, indicating that around 72% of the variability in id_wrong scores can be accounted for by the independent variable of vowel category, a very high score for accounted variability between categories. The homogeneity of variance assumption is tested in a Levene's test of equality of error variances and shows no statistically significant differences (0.226), indicating that the error variances between vowel categories can be assumed to be homogeneous.

Table 11.21 represents the results of a post-hoc Scheffé procedure and compares the mean id_wrong scores by vowel category. Categories that differ significantly in their mean percentage of correct or wrong identifications are marked by *. The table is to be read as e.g. “/y:/ differs significantly in percentage wrong identifications from /ɛ:/, /e:/ and /ø:/ but not from all other categories, i.e. for all other categories difficulties are not significantly more common than for /y:/.”

V	ɑ	ɑ:	ɛ:	ɛ	e:	ɪ	i:	ɔ	o:	ʊ	u:	œ	ø:	ʏ	y:
a															
ɑ:	-														
ɛ:	*	*													
ɛ	-	-	*												
e:	*	*	*	*											
ɪ	-	*	*	-	*										
i:	-	-	*	-	*	-									
ɔ	-	-	*	*	*	-	-								
o:	*	*	*	-	-	-	*	-							
ʊ	-	*	*	-	-	-	*	-	-						
u:	-	-	*	-	*	-	-	-	-	-					
œ	*	*	*	*	-	-	*	*	-	-	-				
ø:	*	*	-	*	-	*	*	*	*	*	-	*			
ʏ	*	*	*	*	-	-	*	*	-	-	-	-	*		
y:	-	-	*	-	*	-	-	-	-	-	-	-	*	-	

Table 11.21: Comparison of mean id_wrong scores (173 L2 listeners) in a post-hoc Scheffé procedure. Significant inter-category differences in id_wrong scores (*p* < 0.05) are marked with *.

To conclude, significant inter-category differences in relative difficulty (indicated by percentage wrong identifications) are observed for several but not for all categories (see Table 11.21). Some vowel contrasts can be considered as more “difficult” to discriminate by L2 learners of German than others.

However, single speech sounds are not difficult in perception as such. While articulatory difficulty can be regarded as a property of an individual sound (in a particular context), perceptual difficulty rather refers to difficulty of contrasts a category is involved in. Perceptual difficulty lies in the correct categorization onto the set of contrasting categories of a given language. The perceptual difficulty of sounds has been hypothesized as strongly corresponding to perceptual markedness, which depends on the salience and stability of the contrasts a particular category is involved in (Lindblom 1986; Flemming 2004). Contrasts that are less confusable and more stable, and as such “unmarked” are preferred over more confusable “marked” ones. The concept of markedness may also explain why some L2 vowels are more difficult to categorize than others, especially if there is no equivalent category in L1.

Perceptual difficulty is considered to be due to several set-dependent as well as set-independent factors (cf. Nosofsky 1991: 134). Biases resulting from differential densities in a vowel system are *set-dependent* biases. *Density*, however, is an aspect of individual stimuli or categories. The density associated with stimuli *i* and *j* depends on the relative similarity to other members of a given set or – in a spatial metaphor – on the relative proximity or distance from other stimuli in the perceptual space. Density is therefore defined as a bias, which emerges from *similarity* relations between a stimulus and other items in the set (Nosofsky 1991: 134).

Frequency of occurrence would be a *set-independent* bias, which is independent of other items in the stimulus set (Nosofsky 1991; Bybee 2007). Frequency of occurrence is considered to affect the strength with which an item is stored in memory due to its frequency of presentation (for a detailed discussion of frequency and experimental results, see section 12.3.3.2).

There may be other stimuli-specific properties that affect the listeners' bias to perceive or remember specific stimuli. A stimulus may be highly "salient" in perception so that it can be easily decoded or encoded. It is important to underline that these properties refer to *auditory* salience andpertain to individual stimuli and not to responses. Within a set of response categories, however, response categories may differ in *psychological* salience, affecting the listeners' disposition to select one response option rather than another (see section 3.3). A detailed discussion and statistical testing of factors affecting asymmetries and listeners' biases will follow in section 12.3.3.

11.3.2 Preferences

Asymmetries in *preferences,* i.e. in the choice of response categories, are observed in all language sub-samples (for a language comparison, see Figure 9.12 and Table 9.4). Figure 11.14 summarizes the preference scores for all non-native listeners. In the "all L2" sample, /e:/ (8.57%) and /y:/ (8.51%) are the most frequently selected categories. /u:/, /i:/ and /ɑ:/ are also very attractive response targets (more than 7%) while their short unconstricted counterparts are less frequently selected. /ɛ:/ (and diphthongs[139]) are dispreferred response options, followed by the ö-qualities /œ/ (5.67%) and /ø:/ (5.13%), and by /o:/ (5.81%) and /ɪ/ (5.95%). A high preference for /e:/ is strongly connected with a dispreference for /ɛ:/ (only 4.01%). Another interesting fact is that /y:/ is a highly preferred response category: In almost

[139] Remember that the set of input stimuli did not contain diphthongs but that diphthongs were offered (and also selected) as response options.

all language sub-samples, /e:/ and /y:/ receive fairly high preference scores. In some language sub-groups (Polish, Farsi, and SBC), /y:/ is even the most preferred category. The qualitative interpretation of the response data shows that the preference for /y:/ is strongly connected with a general dispreference of ö-qualities (/ø:/ and /œ/).

Figure 11.14 presents the preference scores for the “all L2” sample (see also Figure 9.11). Interestingly, L2 participants show a general preference for phonemically long over short response categories. A language-specific comparison of preferences scores (see Figure 9.12) shows that in all L1 sub-samples a high preference is observed for /y:/, /ɑ:/, /i:/, /u:/ and /e:/, whereas /ɛ:/ and /ø:/ are the least preferred categories in most language-sub-groups (except Albanian). To generalize more largely, high long vowels seem more preferred than short non-high vowels and non-high or mid vowels are rather dispreferred especially when mid and front-rounded as /ø:/ (see Figure 11.14, see also section 12.3.3 and 12.5).

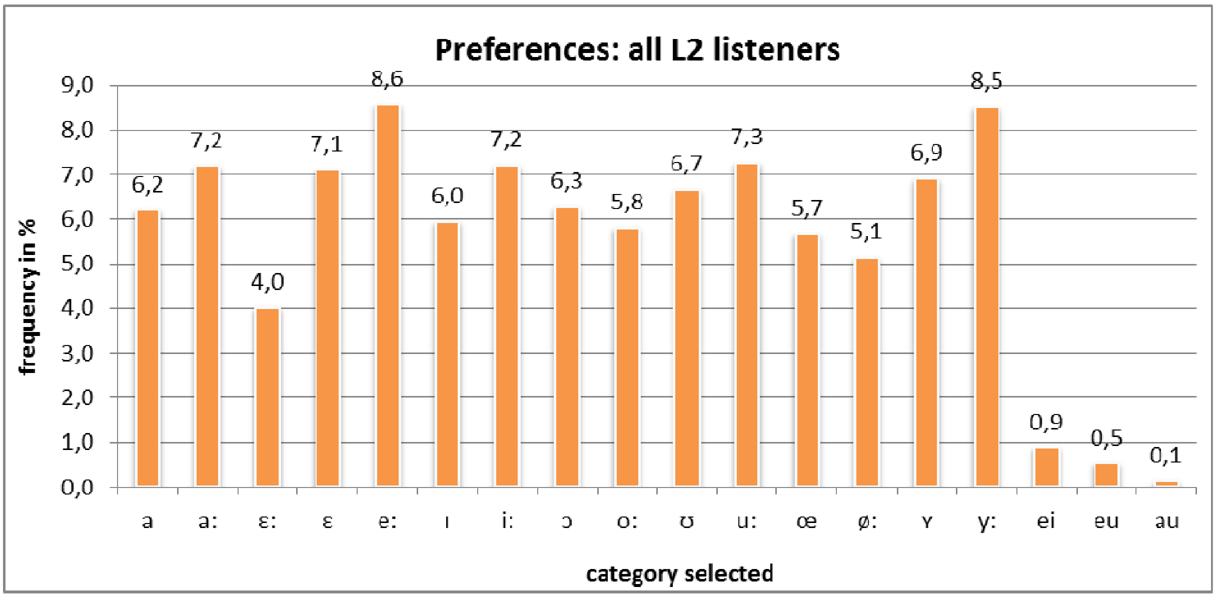

Figure 11.14: Preferred response categories (270 stimuli, 173 non-native listeners)

11.3.3 Direction of substitutions

According to the principles of Natural Phonology, regular perceptual substitutions are assumed to reflect the operation of one or more phonological processes following the principle of minimal substitutions (see section 4.7.2). Difficult sounds are assumed to be substituted by other less difficult sounds usually lacking the particular difficult feature. In this sense, the candidate for substitution (input) and its substitute (output) can be considered as *similar* to each other.

The choice of a specific substitute may vary by the listeners' native language, the phoneme categories of the target language, the relative difficulty of the input category and its phonetic realisation in the input signal and the relative similarity between input signal and selected response option as perceived by L2 listeners of a specific native language.

Processes apply categorically to classes of sounds. The applicability of substitution processes is based on universal "pressures", i.e. phonetic motivations, and implicational conditions. Implicational hierarchies on the application of processes are considered to determine the relative *probability* of specific substitution patterns within a given class. Implicational conditions for each of the observed process types suggest that there are preferred *directions* in substitution patterns. Such implicational conditions as formulated by Donegan (1978) are supposed to account for most of the observed patterns of asymmetry. For example, implicational conditions such as [!lower], [+chromatic], [!tense] for processes of Fronting (in Donegan's terminology Raising) can explain why /e:/ > /i:/, /ø:/ > /y:/ or /o:/ > /u:/ occur more frequently than substitutions in the opposite direction. Implicational hierarchies can make predictions about the probability of process application *within* a given sound-*class*, but they cannot predict the specific choice of processes, since there are no implicational relations between processes, only within processes (Donegan 1978: 103; Kerschhofer-Puhalo 1998a).

To gain more insights into the different strategies or processes that L2 learners use to deal with difficulties related to L2 perception and production, relationships of distance and similarity between L2 categories have to be identified. Several approaches to the issue of "distance" and "similarity" are discussed in chapter 12.

11.4 Summary and conclusions

This chapter investigated common trends and differences in perceptual vowel substitutions across the ten L1 sub-samples. A category-specific comparison of confusion patterns was presented in section 11.1 for each of the constriction locations. In section 11.2 perceptual processes occurring in the full sample of L2 listeners were summarized and discussed to reveal relevant aspects of perceptual difficulty and preferences for vowel classes or specific vowel categories in order to identify major patterns of asymmetry. Asymmetric patterns and perceptual biases are discussed in section 11.3 with respect to the number of wrong identifications (difficulties) and the L2 listeners preferences for specific classes in their responses.

To summarize, the analyses presented here show that perceptual substitutions in L2 do not occur randomly but pattern systematically. Not all categories are equally prone to specific substitution processes. Specific vowel categories are – as the previous sections have demonstrated – susceptible to different types of substitution processes.

According to the articulatory-acoustic classification of German vowels presented in section 5.4 and 5.6 (cf. Figure 5.24) and in analogy to Donegan's (1978) typology of vowel processes, three major types of *perceptual substitution processes* were identified in the data involving changes in (1) phonemic length and constriction degree, (2) constriction location and (3) changes in labiality and/or palatality.

Major types of perceptual substitution processes were discussed in terms of (1) Lengthening and Shortening processes causing changes in *constriction degree and phonemic length* (section 11.2.1), (2) Fronting and Backing processes involving changes in *constriction location* (section 11.2.2) and (3) changes in *labiality and palatality*, where processes of Fronting and Backing combine with Labialization and Delabialization (section 11.2.3).

The data reveal considerable *asymmetries* in difficulties, preferences and direction of perceptual substitutions that are discussed in section 11.3. In other words, a vowel's susceptibility for a specific type of substitution processes varies inter-categorically within a given vowel class. These are specifically conspicous with the contrast pairs /i:/ - /e:/, /y:/ - /ø:/ and /u:/ - /o:/, where a strong tendency to substitute /e: ø: o:/ with more fronted /i: y: u:/ respectively is found, whereas substitutions in the opposite direction are considerably less frequently observed. While these substitution types are largely motivated by phonetic factors and language-specific differences in distribution in the respective region of the articulatory and acoustic vowel space, substitutions of the type /ɛ:/ > /e:/ (49% in the "all L2"-sample) vs. /e:/ > /ɛ:/ (15%) are not only phonetically motivated. Rather, asymmetries in relative psychological prominence of /e:/ vs. /ɛ:/ have to be considered to understand the L2 listeners' behaviour and their response biases for specific categories in the identification task.

In terms of a more cognitive approach to L2 vowel perception that will be discussed in the following chapter, a vowel's propensity to undergo a specific substitution process in L2 perception as well as its probability to be selected as response option can be modelled as *bias*. In terms of approaches that assume *phonetic grounding* of specific constellations of vowel systems and the occurrence of substitution processes in perception and production referring to concepts of markedness and phonetic universals, the observed *perceptual biases* for one rather than another category are assumed to be phonetically motivated. *Implicational conditions* on the application of substitution processes as formulated by Donegan (1978) provide explanations for a considerable portion of error types and error rates in identification (id_wrong), for the listeners' choices for specific response options (pref_scores) and the direction of perceptual substitutions. However, in L2 acquisition, these "universal biases" have to be considered to act along with language-specific biases and individual cognitive biases accounting for differences in relative prominence of one or the other vowel category. It is suggested here that stimulus bias and category bias are related to the signal as well as to the listener – or referring to L2 acquistion – to the language learner and his/her emergent interlanguage system.

The rate of wrong identifications for a specific stimulus category in the current perception data is considered to reflect the degree of *difficulty* of a given category. The higher the percentage of wrong identifications, the lower the perceptual salience of a specific stimulus or contrast (for similarity, salience and difficulty, see section 3.3) and the higher the degree of difficulty in decoding the given input signal in L2. Degree of difficulty or perceptual salience concern properties of individual stimuli and categories in contrast. *Difficulty* – referring to a category's propensity to undergo perceptual substitutions – can therefore be characterized as *stimulus biases*, whereas *preferences* are rather viewed as *response biases* referring to a category's probability to function as target in perceptual substitution processes (see section 12.4).

The relative frequency of a given substitution process in the data is considered to reflect the relative *similarity* of the two members of a given vowel contrast. Processes that occur more frequently indicate a higher degree of similarity and a lower degree of contrastiveness for a given contrast pair. Similarity may refer to objective *stimulus similarity* (e.g. phonetic properties) as well as subjective *response similarity* (for a detailed discussion cf. Nosofsky 1991; see also chapter 3 and section 12.4). The fundamental question, however, concerns the cognitive and/or physiological status of the observed patterns in *difficulty* (percentage wrong identification), *preferences* for specific response options and *asymmetries* in direction of perceptual substitutions in an empirical explanation or a theoretic model.

The next chapter presents cognitive approaches to similarity, contrastiveness and perceptual bias and discusses their relevance for non-native perception and foreign language acquisition. Chapter 12 will discuss the differentiation of stimulus similarity vs. response similarity and of stimulus bias vs. response bias (Nosofski 1991) in more details. The aim of chapter 12 will be to operationalize those forces or pressures that (1) explain the susceptibility of vowel categories for perceptual substitutions (*stimulus bias*) due to characteristics of the input signal accounting for *perceptual difficulty*, and (2) to account for *perceptual preferences* in the output in selecting one category rather than another as response option (*response bias*). The discussion will consider the interaction of vowel-specific, language-specific, individual and more general factors, and will discuss the role of vowel-inherent phonetic motivation and implicational conditions for process application and bias as well as factors outside the speech signal.

12 Modelling Similarity, Difficulty and Preferences in Non-Native Vowel Perception

The correct perception of vowels is a crucial component in the acquisition of an L2 sound system. Difficulties in L2 sound perception due to cross-linguistic influence may manifest themselves in many ways in L2 performance, in perception as well as in production, in the development of phonemic categories, in the rate of acquisition of specific phonemic contrasts, the route of development, and the frequency of errors. A well-funded theoretical description of the acquisition of non-native vowel contrasts therefore requires an integrative conception of language knowledge, language processing and language learning, in particular (1) a theory of *mental representations* for knowledge-related questions on the mental organization of L1 and L2 sound categories and sound inventories, (2) a theory of language *performance* for the description of (non-native) speech processing, and (3) a theory of *learning* for an understanding of acquisition processes in the development of L2 knowledge and L2 speech processing. This dissertation is intended to be a contribution to an integrated view of performance-related, knowledge-related and learning-related aspects of the development of mental representations for the L2 vowel system in L2 acquisition.

This final chapter will summarize central aspects of mental representations of vowel sounds and sound-related linguistic knowledge, speech processing and language learning and link these aspects with fundamental psychological and linguistic concepts such as *categorization* and category formation, perceptual *similarity* of categories and *selective attention*, in order to describe the role of cross-linguistic influence, similarity and preferences in the acquisition and mental organization of an L2 vowel system.

One of the central questions in this chapter is the discussion and operationalization of the very flexible and rather vague notion of *similarity* that is a central component in many L2 acquisition models (see section 1.1.2, 2.4 and chapter 3). The aim of this chapter is (1) to review notions of *perceptual distance* and *similarity* between L2 categories, (2) to discuss the relevance of specific phonetic and cognitive correlates to define *dimensions of similarity* and (3) to estimate their impact on cross-linguistic influence, difficulties and preferences in the acquisition of an L2 vowel system. It moreover provides a theoretical and empirical discussion of determining factors of *asymmetry* and *bias* in vowel identification and the role of *selective attention* and *attentional learning* in L2 acquisition.

12.1 Vowel systems, vowel categories and the vowel space

12.1.1 The multi-dimensional vowel space

Vowel sounds vary across several phonetic parameters or dimensions. In this sense, the *phonetic vowel space* is *multi-dimensional*. Phonetic multi-dimensionality refers to articulatory, acoustic and perceptual dimensions of the vowel space that vary across languages. The multi-dimensional perceptual vowel space as well as the articulatory or acoustic vowel space varies language-specifically in terms of *n* parameters or dimensions that differentiate vowel categories within a given language system. Languages differ not only in the number of vowel categories but also in the differential weighting of the component dimensions of the *perceptual vowel space* (see section 12.5). Therefore, patterns of perceptual confusion in L2 are neither predictable from acoustic or articulatory properties of the input signal nor by a traditional contrastive analysis of L1 and L2.

Many phonetically-oriented studies on vowels and vowel systems and their perception in L2 are implicitly guided by the idea that nearly all information necessary to specify vowels in a given language system can acoustically be summarized in a two-dimensional F1xF2 representation based on measured values from a cross-section of the vowel spectrum at one steady point of the vowel continuum. However, for a number of reasons that have been discussed in section 4.3.4, 4.3.5 and 4.3.6, this two-dimensional view is too static and insufficient for a complete acoustic-phonetic description of vowel categories. Dynamic spectral change, the transitions to adjacent consonants, duration and fundamental frequency provide substantial cues for correct vowel perception (Strange, Jenkins & Johnson 1983; Nearey & Assmann 1986; Strange 1989; Andruski & Nearey 1992; Jenkins et al. 1994; Hillenbrand et al. 1995; Strange & Bohn 1998; for an overview see Hillenbrand 2013).

Likely, models of vowel articulation are implicitly influenced by this static view as expressed by a set of distinctive articulatory features to differentiate vowel categories and vowel contrasts abstracting from parametric phonetic variation within and between categories. Acoustic analysis and the use of more advanced methods for articulatory tracking (e.g. Magnet Resonance Imaging or Electro-Magnetic Midsagittal Articulography) challenge this static view and reveal the highly dynamic character of vowel production providing ample evidence for the considerable amount of context-dependent and free articulatory variability and for the complex interplay of the many articulators involved in vowel articulation. Due to the articulatory and acoustic complexity of vowels, a two- or three-dimensional representation will never be able to reflect the many dimensions in which vowel categories differ.

In speech perception, despite the high dynamics and variability of the speech signal, listeners show an astounding capacity to discriminate and identify vowel sounds correctly in a language they are familiar with. This is due to stable categorical representations in a given language together with lexically and morphologically driven top-down processes in speech perception, even if this capacity may not be equally good for all signals. The capacity or rather the particular way to discriminate and identify the input signal as belonging to one or the other category is considered to be language-specific and may therefore vary from language to language and from listener to listener according to his/her experience in one or more languages.

12.1.2 Spectral information and multidimensional variation

The phonological description of vowel sounds commonly refers to three major *dimensions* that are viewed as referring to basic modifications of the vocal tract in speech production: (1) frontness/backness, (2) height and (3) rounding. The systematic variation of articulatory and acoustic patterns is the basis of human categorical perception. Observable articulatory or acoustic properties of sounds are the phonetic basis of distinctive features in more abstract phonological representations of vowel sounds. Traditionally, a phonological feature is assumed to have a phonetic correlate: e.g. vowel height is frequently stated to correspond to the first formant frequency F1 or to the distance F1-F0 (Hoemeke & Diehl 1994; Syrdal & Gopal 1986). However, a direct relationship of phonetic substance and phonological features is difficult to establish, since in most cases there are no one-to-one relationships between phonological differentiations and phonetic correlates. Therefore, traditional phonological accounts commonly use a set of hybrid features and feature geometries that often result in considerable confusion about the roles of phonological features and/or articulatory, acoustic or perceptual features (for discussion, see Boersma 1998: 16ff).

A large number of acoustic-phonetic studies are based on the simplest way of extracting formants by local peak picking from the spectrum. However, the peak picking extraction is difficult when two adjacent formant frequencies are very close, e.g. F1 and F2 for back vowels or F2 and F3 for high vowels, or when fundamental frequency F0 is higher as in a woman's voice (Ito, Tsuchida & Yano 2001).

It is commonly accepted that spectral resonances called *formants* are the acoustic basis for human speech perception, though the exact nature of their representation is a matter of intense debate. The source signal generated by vibrations of the vocal fold is modified as it passes through the vocal tract which is shaped dynamically by the articulators. The acoustic signal shows systematic variation in accordance to typical modifications of the vocal tract and the speaker's intention. Formant frequencies result from the specific constellation of the articulators as the source signal passes through the vocal tract. Stevens & House (1955) considered formant frequencies as basic cues to vowel quality (see also Klatt 1982), whereas others stated that the relationships between formants are crucial for vowel identification (Peterson & Barney 1952; Syrdal & Gopal 1986). A fundamental question in this context is the issue of how formant frequencies are *perceptually* extracted from the signal. While some authors assume a mapping of formant frequencies in the input signals on representations of formants (formant hypothesis), this view is challenged by others assuming that the whole *spectral shape* is decisive in vowel perception (e.g. Bladon & Lindblom 1981).

Another crucial issue refers to the question whether formant frequencies per se, i.e. spectral *peaks* observed in the signal, or the overall *spectral shape* of vowels are the basis of vowel perception. Several other vowel-inherent characteristics additional to formant frequencies and their relationships to each other have been found to play a role in vowel perception (Kiefte, Nearey & Assmann 2013), such as duration (Hillenbrand, Clark & Houde 2000), fundamental frequency F0 (Nearey 1989), dynamic spectral change (Nearey & Assmann 1986; Hillenbrand 2013), and the transitions to adjacent sounds (Hillenbrand, Clark & Nearey 2001; Pierrehumbert 2003) that provide relevant cues for vowel identification. A factor that is often neglected is information from *bandwidth* (frequency region around a formant peak in which the amplification differs less than 3 dB) and formant *amplitudes* (see Ito et al. 2001; Kiefte, Nearey & Assmann 2010). Chistovich & Lublinskaya (1979) proposed a kind of averaging process of closely adjacent formants and called its effect "*center of gravity*". The center of gravity cannot simply be viewed as average of F2-F3 or F3-F4. Its exact location depends on the relative distance of F2-F1, F3-F2 or F4-F3 and varies between different vowel classes, e.g. back vs. front vowels, or front rounded vs. back rounded vowels as shown in the diagrams in section 5.4. The elimination of this "center of gravity" effect and a change in perceived vowel quality was observed when the *amplitude* ratio of two adjacent formants was below a critical distance of 3-3.5 Bark. For vowels with a higher than this critical distance, formant amplitudes showed a minor effect and the vowel quality was determined by formant peaks in the spectrum. Moreover, these cues interact in a complex way (e.g. Ito, Tsuchida & Yano 2001).

In other words, the role of specific acoustic cues may differ by vowel category or vowel class (e.g. back vowels where F2-F1 distance is minimal compared to front vowels where F3-F2 distance is minimal) and – an aspect that has to be explicitly stressed – is that the relative *weighting* of the essential cues may vary across languages: for example, naturally produced English vowels are rarely spectrally static and the importance of vowel-inherent *spectral change* for vowel identification may be stronger in English than in German particularly for long vowels.

Vowel identification is further complicated by external faactors such as variation between speakers, their sociolinguistic background and speech style, by speaking rate, by pre- and post-vocalic context, syllable and word structure, speech situation, but also by lexical structure and word token frequency, altering the spectral characteristics of vowel sounds.

To summarize, no direct relationship of acoustic cues in the input and perceptual response (output) can be assumed. The relationships of signal-inherent properties, mental representations of categories and perceptual responses depend on multiple factors that vary language-specifically. These relationships are difficult to isolate experimentally but have to be incorporated into a model of L2 perception.

Moreover, in L2 acquisition, the relative *weighting* of relevant cues and their interaction in the target language and the listeners' native language differ language-specifically and eventually develop or change in the course of L2 acquisition. Cue weighting may moreover differ across vowel classes or even single categories (see 12.4). For all these reasons, a simple comparison of formant frequency values in L1 and L2 will always be insufficient to predict or operationalize *perceptual ease or difficulty* in L2 and to cover the relevant factors that affect the listeners' perception and their decisions for specific response categories.

12.1.3 The continuous vowel space vs. categorical discreteness

One of the most fundamental concepts in linguistics is the assumption that linguistic elements can be classed into *categories* and that a system of linguistic elements can be separated into discrete *classes*, i.e. mutually exclusive sets of elements. According to the *categorical view*, a number of phones that share specific properties or features can be summarized as tokens of the same "category" or "type" or "phoneme" that is sufficiently distinct from other neighbouring "categories" or "phonemes" (Labov 1994: 349). The vowel system of a given language comprises a specific number of such discrete categories.

Basic assumptions about linguistic categories are largely influenced by set theory and Venn diagrams, such as the view that (1) category membership is a matter of "all-or-nothing", (2) that categories are expected to have a boundary, and (3) that all members of a category are considered to have an equal cognitive status within a category. Though these assumptions on the nature of categories have been applied in many linguistic frameworks and have influenced *phonological* descriptions of phoneme systems, they are however difficult to be upheld in this form for the majority of linguistic and cognitive categories (Janda 2010) (see also chapter 2 and 4). However, the construct of discrete categories has strongly influenced phonological descriptions of phoneme systems in the sense that (1) a given sound always belongs to one of the phoneme categories of a given language, which are (2) clearly differentiated from another category, and (3) intra-category variation and phonetic differences between tokens of a given phoneme type are considered to be phonologically irrelevant.

In a more *phonetically-oriented* view, by contrast, we find the assumption of a *continuous phonetic space* without any discrete borders. The representation of a vowel system in an F1xF2 plane supports this concept of a continuous articulatory-acoustic vowel space. The F1xF2 representation of acoustic formant frequencies is frequently discussed and interpreted in terms of two basic articulatory dimensions, height and frontness/backness. Moreover, referring to more or less peripheral or central vowels, the F1xF2 chart is described in terms of peripherality and centrality (cf. e.g. Polka & Bohn 2003, 2011; Labov 1995: 458; Labov 2010: 114; Schwartz et al. 2005; see section 4.7.4 and 12.3.3.1). However, vowels vary articulatorily and acoustically in many more than only these two dimensions and are therefore to be described along a continuum of several different acoustic parameters as well as articulatory parameters such as front or back articulation, degree of aperture, or degree of lip rounding, etc. Acoustic characteristics of the speech signal result from these articulatory constellations and vary in terms of several acoustic parameters beside the first two formant frequencies (as discussed in chapter 4 and 5).

From the coexistence of the categorical paradigm and the assumption of a continuous phonetic space in studies on vowel perception in L2 acquisition, several theoretical questions and problems arise with respect to the conception of the mental organization of the articulatory, acoustic and perceptual vowel space in a given language and their acquisition of a second language phonology.

12.1.3.1 The categorical view – basic assumptions

The common understanding of the notion "category" in many fields of linguistics is largely influenced by notions of set theory and Venn diagrams that have been adopted in experimental phonetics and sociolinguistics, where "category" is a well-established construct. Labov (1994 (I): 351) summarizes six characteristic components of the "*categorical view*" in phonetics and phonology that implicitly guide the majority of studies on phonetics and phonology in language variation, language acquisition and sound change: (1) the concept of *contrast*, (2) the concept of *distinctiveness*, (3) the *irrelevance of phonetics*, (4) the concept of *discrete category boundaries*, (5) the assumption of *symmetry in perception and production*, and (6) the reliability of intuitions of the *native speaker*. Labov (1994 (I): 351f) critically reviews these basic concepts that rely on the following assumptions:

(1) The concept of *contrast*: Contrast exists between *types*, which are exemplified by particular forms. A phonemic contrast of two types usually involves differences in meaning. The occurrence of two types is not predictable by rules of contexts. Native speakers are sensitive to differences between the two types but usually do not perceive consciously differences between tokens of the same type.

(2) The concept of *distinctiveness*: Members of a contrastive category are considered to share one or more distinctive features. Non-distinctive features are commonly considered as redundant and irrelevant for categorization.

(3) The assumed *"irrelevance" of phonetics* (Labov 1994 (I): 351): The phonological analysis of distinctive categories implies a certain amount of data reduction and a limitation to an abstract set of features that are taken as distinctive, while ignoring all phonetic details that are not distinctive and/or common to all members of the category such as detailed information on formant trajectories, amplitude contours or energy distributions. "*The amount of phonetic information provided in most phonological analyses is limited to the abstract set of features that are taken as distinctive. Vowels are described as high, low, or mid, and finer differences in fronting, backing, raising, lowering, and rounding are ignored.*" (Labov 1994 (I): 351f)

However, the successful acquisition of the phonological system of another language implies raised attention to specific *phonetic details* that may not be relevant or "distinctive" in one's native language. Within one's own native language, non-distinctive sub-phonemic differences between realizations of a given category are irrelevant in categorization.

Orthography favours this kind of data reduction supporting the illusion that all tokens of a given orthographic type are phonetically equal, an assumption that is not always correct, e.g. in German words where final devoicing applies in pronunciation but not in writing resulting in homonymes like *Jagd* 'hunt' (noun) - *jagt* 'hunts' (3rd sg.); *Tod* 'death' (noun) – *tot* 'dead' (adj.).

(4) The assumed discreteness of *boundaries*: Boundaries between categories are implicitly considered to be sharp in the sense that there are no intermediate forms, a view that is challenged by the concept of a continuous vowel space, the considerable within- and between-category variation and the overlap in articulatory gestures and acoustic properties in phonetic reality.

While the categorical view favours discreteness of phonetic parameters and phonological categories, we have to consider that there are several cases of overlap in the acoustic signal where single phones cannot unambiguously be classified as elements of one or the other discrete category and where the boundaries between phonetic categories are not discretely separated along articulatory or acoustic parameters (e.g. duration, acoustic correlates of frontness/backness, degree of aperture or lip rounding, for discussion see section 5.4).

(5) The assumed *symmetry of perception and production* and the implicit assumption that phonologies are neutral with regard to the perspective of the speaker or the hearer: Here, the underlying assumption refers to the vowel map as based on instrumental analyses of phoneme production that should show the same structure as the output of perception experiments such as discrimination or categorization tasks. This view of parallel monotonic changes in articulatory, acoustic and auditory parameters is not sustainable (cf. the contribution of Quantal Theory of Speech (Stevens 1972, 1989; Stevens & Keyser 2010); see also section 4.7.3).

(6) The assumed *reliability* of intuitions of the *native speaker*: The categorical view is strongly based on the assumption that native speakers have free and conscious access to their intuitions about what is correct or wrong within a linguistic system. Most experimental studies on L2 perception are based on the implicit working assumption that the listeners' behaviour in the experiment provides reliable evidence for the structure of the listeners' linguistic system, though the reliability and validity of this methodological approach is of course limited (see section 6.5).

Even if several experimental studies in phonetics and language acquisition provide evidence that native speakers do think categorically about their language or at least that they behave in the experimental setting as if they would think categorically, we have to be aware of the risk of circularity of the concept of categorical perception relying solely on the native speakers' intuitions about their language system and on their categorical behaviour in experiments to make assumptions about the underlying principles of a given language system without considering the fact that these findings are only based on what listeners think about their system (Labov 1994 (I): 352). Behavioural data will never be able to render directly what listeners really do or think and how language really is processed in the listeners' brains. Complementary evidence from other empirical fields, e.g. provided by brain-based methods (see section 2.3.7), is necessary for a more complete picture of phonetic processing.

A crucial issue related to the concept of the *native speaker* is the implicit assumption that native speakers of a given language will behave "equally" or at least similarly because they

share a common native language. However, as will be discussed in section 12.5 considerable *inter-individual variation* in the listeners' responses can be observed (for a detailed discussion of learner-related differences and factors influencing L2 performance, see also chapter 8).
The effect of deliberate decisions of an individual in a given experimental setting and the fact that the "native" listeners' linguistic competence as well as performance could be changed precisely by the experience of learning another language (Cook 2003; Escudero & Sharwood-Smith 2001; Sharwood-Smith 2011), is commonly not sufficiently considered or even disregarded in most experimental studies.
The notion of the "native speaker" has been challenged in more recent studies (cf. e.g. Davies 2003; Paikeday 2003; Sharwood-Smith 2011) that are of great relevance for studies on L2 and L3 acquisition, particularly when language learning is viewed in the sense of Vivian Cook's multicompetence approach (see Cook 1999, 2003).
Another issue not sufficiently controlled for in many L2 studies – even in this study – is the influence of *intra-lingual variation* in L1 as well as L2, not only with languages such as German, English or Arabic, due to regional as well as socio-linguistic variation. The influence of variety-specific sub-variants on phonetic boundaries between phonemic categories is hardly controlled for in experimental studies, but must be considered in the interpretation of empirical results (see e.g. Evans & Iverson 2004; Escudero 2002; Escudero & Williams 2012; Williams 2013). This issue may not be as relevant in the present study, as the study is intended as a survey study rather than as a study on the impact of specific sub-phonemic details on the perceived quality of single categories. The current study's major aim is to describe major areas of difficulty, preferences and directions of perceptual substitutions for the whole German system.
To conclude, the "native speaker" is a convenient construct, but experiments that ask for the impact of specific fine-grained phonetic variation on the perception of vowel contrasts will have to control not only for the influence of the listeners' regional native variety and accent but also for language proficiency and the L1 and L2 varieties the language learners have been exposed to and for the longitudinal changes that experience with specific language varieties involve.

12.1.3.2 The internal structure of sound categories

One of the central questions in cognitive psychology concerns the *internal structure* of sound categories and the way humans represent categories in memory. Models of the internal structure and organization of categories describe the relation of items within a category, the boundaries between categories and the way how stimuli are categorized and categories are represented in memory and learned in acquisition. Two major types of models, *exemplar-based models* and *prototype models* have been influential in studies on L2 phonetics and phonology.

Prototype models claim that categories are mentally represented by forming a summary representation reflecting central tendencies of experienced members of a category. In a *prototype* approach to categorization, it is assumed that humans learn characteristic features of categories or central tendencies in the data and use them to represent the category. Classification decisions are based on the similarity of an item to the stored prototypes (see Kuhl 1991, 1992, 1993, for non-native sound perception). In this approach, the *similarity* of the input and the prototype is relevant for categorization (Rips, Smith & Medin 2012: 183). Abstract prototypes do not necessarily have to correspond to any experienced example. Experimental phonetic studies show that not all members of a category have an equal cognitive status, i.e. some members are considered to be more *(proto)typical* than others. The prototype of a category can be viewed as an item with special salience that can, however, not be objectively defined in terms of e.g. one specific feature shared by most members. Even if members at the extreme ends have no feature in common, they might be linked by a chain of other members sharing common features. While the prototype has the densest structure of relationships to other members, peripheral members are less representative of a category than the prototype (Janda 2010: 12). In brief, a given category is motivated by and organized around a prototypical member to which all other members are in some relationship. *Boundaries* are not a necessary element of such models: "*Rather than having a defining boundary and no internal structure, human categories tend to have a defining internal structure and no boundary.*" (Janda 2010: 11)

Exemplar-based models (for discussion, see section 2.4.4) assume that categories are represented by storing all experienced individual members or exemplars of a category as separate traces. Classifications are performed on the similarity of items and stored exemplars (see e.g. Nosofsky 1986; Johnson 1997; Pierrehumbert 2001; Ettlinger & Johnson 2009; Bybee 2010).

An alternative account would be a "mixed exemplar-prototype" model of categorization involving a prototype representation together with all-or-none memories for specific exemplars (for discussion, see Nosofsky & Zaki 2002).

According to an *exemplar view* on categorization, the memory of already experienced exemplars can influence the classification of new exemplars by an assessment of their *similarity* to already stored examples and by assigning the exemplar to the category which has the most similar exemplars, even if these exemplars are not most similar to the prototype (if a prototype is assumed) (Rips, Smith & Medin 2012: 183). However, similarity to the category prototype and similarity to stored exemplars is expected to be highly correlated. On an exemplar-based view, *typical features* influence classification less than high *similarity to specific examples*, whereas a prototype model would rather expect that specific characteristics of the prototype are prior to properties of single items.

For speech perception, the exemplar-based view assumes a complex storing of many individual exemplars but a more direct way of mapping the input to the stored category without a procedure of feature extraction (Johnson 1997; Pierrehumbert 2001; Ettlinger & Johnson 2009), whereas the prototype view is based on the assumption of a less complex representation but a more complex mapping procedure.

In exemplar-based models, the process of sound *categorization* is described as a process of comparing auditory properties of a sound with the auditory properties of the stored exemplars; the *similarity* between an item and each exemplar determines the activation level of the exemplar (Johnson 1997). A good match of auditory properties involves a high activation level of a given exemplar. Auditory properties are the output from the peripheral auditory system. The set of category labels includes classifications that are important to the listener as available from the listener's knowledge at the time the exemplar was stored. "*The sum of activations over all of the exemplars of a category is taken as evidence that the unknown sound should be categorized as an instance of that category.*" (Johnson 1997: 147)

More recently, new kinds of classification theories have been formulated, such as rational approaches, decision-bound models, and neural network or connectionist models (for an overview, see Rips, Smith & Medin 2012: 184ff). *Decision-bound models* (e.g. Maddox & Ashby 1993) posit that category learning consists of developing bounds around a category. The closer an item to the decision bound, the harder it will be to categorize. This concept is compatible with observations that listeners have difficulty with vowel items that are not within the "(proto-)typical" array in their L1 in terms of acoustics and are therefore difficult to classify (see also Kuhl's Native Language Magnet Model, section 2.4.1). Applied to vowel identification, this view strongly refers to the notion of a continuous vowel space. It also allows for assuming language-specific decision bounds for vowels and context-dependent variation in vowel recognition.

Mixed models propose that ideas from more than one theory could be combined in a model, such as exemplars and prototypes, rules and examples, or rules and decision bounds, though the question whether prototypes are necessary at all remains controversial (Rips, Smith & Medin 2012).

However, predictions of alternative models tend to highly correlate in several respects. From these concepts, several relevant questions for the present study arise such as how decision bounds could compare with what is "optimal" or "prototypical" and which kinds of decision functions are easy or hard to acquire (Rips, Smith & Medin 2012).

12.1.4 The acquisition of vowel categories and vowel systems

In the present study, vowel categories refer to the phonetic substance of the speech signal as well as to the listeners' mental representations. The incoming acoustic signal is compared to mental representations of sound categories. A definition that considers variation in the signal and statistical probability in sound identification is provided by Pierrehumbert:

> "*A category is a mental construct which relates two levels of representation, a discrete level and a parametric level. Specifically, a category defines a density distribution over the parametric level, and a category system defines a set of such distributions. Using the density distributions for categories in a category system, incoming signals may be recognized, identified, and discriminated through statistical choice rules. (...) An example is provided by the standard representation of a vowel space as a set of density distributions in F1-F2 space (...). Each of these has phonetic correlates in its own right. Since a category is a statistical relationship between a discrete level and a parametric level, it follows that two categories are identical only if they are identical at both levels.*" (Pierrehumbert 2003: 119f)

Vowel systems are viewed as sets of mental categories that are conceived as sets of parametric *density distributions* for phonetic parameters. According to this view – which is strongly favoured in the present study – categories do not have exact bounds. The listeners select a specific category from the set of density distributions by a kind of calculation of *probabilities* driven by context, previous knowledge and expectations and calculated similarity relations. Languages vary in their set of categories, their phonetic correlates and their statistical distributions as well as in the listeners' probability theory. This view appropriately takes into account the active role of the individual listener or learner of a second language. Therefore, similarity is considered here as a crucial element in the conception of L2 vowel perception that strongly affects the listeners' decision bias in an experimental identification task.

In acquiring the sound categories in a second language, differences in L1 and L2 may cause problems if

> "*... the relationship of the categorical label to the parametric level is not actually the same in the two languages, even if a certain similarity can be discerned. If two categories are not identical, they may still be equivalent (if an appropriate*

equivalence relation can be defined, in the mathematical sense), or analogous (if they are comparable in some looser sense)." (Pierrehumbert 2003: 120)

In other words, categories may or may not be identical in phonetic parameters in an objective sense, but what is actually of relevance in L2 acquisition is the phonological, functional aspect. Equivalence and similarity relationships are established by the listener, here the language learner, and may vary according to the listeners' language knowledge and experience. *Equivalence classification* (Flege 1988) is understood here as the interpretation of the relation between the discrete and the parametric level; it is not existent in the signal itself but is established by the *listeners*. It is exactly this part of the speech processing chain (see Figure 3.1), where phonetics can be differentiated from phonology. The focus of phonetics is the signal and its properties rather than the individual active part that a listener has in the perception of speech. This question is a matter of language-specific and/or listener-specific phonology. The phonetic form itself allows only limited conclusions on the phonological function. In speech perception, the listener has to add information that the signal is missing.

Once a human has acquired his/her native language and knows which sounds and combinations of sounds to expect in the phonetic string, it is easy for him/her to decide on the phonological identity of sounds in the perceived speech signal. These decisions are implicitly based on the experience that *n* phonetic forms correspond to one phonological category. Decisions are induced not only by the phonological system the listener has acquired but also by a mental lexicon for words and their sound structure.

Pierrehumbert (2003: 116) describes phonological acquisition as "*a cognitive architecture with multiple levels of representation*". On her view, these levels minimally include: (1) *parametric phonetics* in terms of a quantitative map of the acoustic and articulatory space, in which proximity in multiple dimensions can be defined, (2) *phonetic encoding* involving low-level categorization of the phonetic space, (3) the *lexicon* containing lexical representations of word-forms, which provide a locus for association between form and meaning, (4) the *phonological grammar* with general constraints on word-forms in the lexicon such as constraints on metrical structure or segmental sequencing, and (5) *morphophonological* correspondences, where phonological relationships amongst morphologically related words are predictable from constraints on word form.

To conclude, vowel identification is a process of comparing and matching acoustic properties of the input signal with mental representations. Sound identification involves a comparison of a given acoustic item to be categorized and the stored exemplars or prototypes (or their memories or traces). This comparing and matching involves processes of identifying the relative "*similarity*" or "*distance*" of auditory sensations and stored representations, which are linked with a "probability theory" by the listener. The notions of distance and similarity are therefore crucial concepts in speech processing and language acquisition and will be discussed in details in the following sections with special consideration of vowel perception and phonological acquisition in L2. The following sections will (1) introduce types of cognitive *models* for *similarity*, (2) show how *perceived similarity* or distance can be "measured" or operationalized for L2 vowel perception, and (3) discuss how the *knowledge* of language(s) and the relevance of specific *properties* of the input signal influences processes of speech processing, mental category representations and the acquisition of L2 sound categories.

12.2 Similarity

12.2.1 Perceptual similarity and categorization

The construct of *similarity* is fundamental for an understanding of psychological and psychophysical tasks such as categorization, memory retrieval, learning, and other cognitive processes. Similarity plays a crucial role in humans' predictions on objects and their behaviour. Similar objects are commonly assumed to *behave* similarly and they are expected to behave similarly *because* they *are* similar (Goldstone & Yun Son 2005: 14). If an unknown object is similar enough to a known object, the category label of the known object may be applied to the unknown (Nosofsky 1986; Goldstone & Yun Son 2005: 14).

A central and widespread though controversial notion refers to *similarity* as an underlying principle for the structure and organization of *categories*. This basic assumption refers to the idea that items are grouped together into categories because they are similar to each other. Similarity of items is related by their status as members of the same or different categories. Commonly, we consider things to be similar *because* they belong to the same category and not vice versa (Medin & Aguilar 1999). On this view of similarity, speech sounds are organized for us and our mental categories map onto reality in the speech signals. In this sense, similarity could be explained in terms of *categories* rather than vice versa.

Similarity may also be defined in terms of shared properties, though the selection of the relevant properties in category definition and identification is a very flexible matter due to the multidimensionality of mental representation and the almost unlimited number of possible properties that may be shared: e.g. /a/ and /i/ are obviously assumed to be more similar to each other than /a/ and /p/, but even /i/ and /p/ could be described as similar to each other, since both are speech sounds, both are articulated in the front part of the vocal tract, both have a shorter intrinsic duration than /a/, etc.

In speech perception and sound categorization, we commonly refer to the concept of similarity of items belonging to the same category because they are similar to each other even if it is not precisely known *in what respect* they are actually similar.

> *"We tend to rely on similarity to generate inferences and categorize objects into kinds when we do not know exactly what properties are relevant, or when we cannot easily separate an object into separate properties. Similarity is an excellent example of a domain-general source of information. Even when we do not have specific knowledge about a domain, we can use similarity as a default method to reason about it." (Goldstone & Yun Son 2012: 156)*[140]

On the other hand, if we do *know* about a specific property, then this knowledge guides our decisions:

> *"The contravening limitation of this domain generality is that when specific knowledge is available, then a generic assessment of similarity is no longer as relevant" (Goldstone & Yun Son 2012: 156).*

Referring to perception in L2 learning, it is this (language-specific) knowledge about the existence of specific categories and particular distinctive features or properties of a given category in L1 and L2 that will guide the learners' decisions in L2 sound categorization.

Referring to the *confusion data* in the present study, we are confronted with two major aspects: (1) the listeners' *knowledge* about vowel systems and vowel categories, their knowledge about specific properties within a given language system and their theory of probability in a given context that leads or misleads L2 learners in vowel identification and guides their decisions in the identification task, and (2) the L2 listeners' *behaviour* in the experiment that is guided by implicit rather than explicit individual *assumptions on similarity* of German vowel categories but is not consciously related to specific acoustic parameters by the listener. The methodological problem is to identify those perceptual dimensions that are used as essential cues in L2 vowel identification.

The present study aims at the description of *perceived similarity* guiding L2 learners' decisions in an identification task rather than at a mere description of phonetic similarity in terms of acoustic or articulatory vowel parameters. The non-linear relationship of acoustic variation and perceived similarity relationships will therefore be discussed in more details in the following sections.

[140] emphasis mine.

12.2.2 Modelling perceptual similarity

In the last decades of psychological research, several models of *similarity* have been proposed to describe cognitive aspects of similarity and categorization. The following sections will discuss different models of perceptual similarity and cognitive approaches to account for the problem of symmetry, asymmetry and bias.

Four major types of models have been proposed: geometric, featural, alignment-based and transformational models (Goldstone & Yun Son 2012). These models have been more or less influential in cognitive psychology and are more or less relevant or applicable to the identification and categorization of speech sounds.

12.2.2.1 Geometric models of similarity

Geometric models (Torgersen 1965; Shepard 1957, 1972) are among the most influential approaches, using the "spatial metaphor" for statements about similarity of objects in an *n*-dimensional space. A classic model for predicting identification-confusion data is the *choice model* (Shepard 1957; Luce 1963). In his original formulation of the model, Shepard (1957) assigned an explicit interpretation to similarity parameters in terms of *distances d* between item *i* and *j* in a *psychological* space assuming

$$\eta_{ij} = f(d_{ij})$$

where *f* is some monotonically decreasing function and distances are metric.

Shepard (1957) proposed that (1) stimuli can be represented as points on a low-dimensional psychological space, (2) that d_{ij} can be derived by computing the distances between points in that space and (3) that similarity parameters are functionally related to distances in a multi-dimensional space. In a procedure of *multidimensional scaling* (MDS) a representation of the multi-dimensional perceptual space is sought that corresponds best to the identification data of a given stimuli set. According to this MDS-choice model (cf. Nosofsky 1986), the *probability* that stimulus *i* is identified as response *j* is described by

$$P(R_j|S_i) = \frac{b_j \eta_{ij}}{\sum_{k=1}^{n} b_k \eta_{ik}}$$

where $0 \leq b_j \leq 1$, $\Sigma b_j = 1$, $\eta_{ij} = \eta_{ji}$, $\eta_{ii} = 1$. The b_j parameters are interpreted as *response bias* parameters associated with item *j*, and η_{ij} parameters are interpreted as *similarity measures* on the stimuli items *i* and *j* ($0 \leq \eta_{ij}$, $\eta_{ij} = \eta_{ji}$) (Nosofsky 1986).

There are however two basic problems with the assumption of a direct monotonic relationship between distance and confusion in a geometric approach: (1) it holds only under the condition of *symmetry* if there is no bias either for one or the other response category, and (2) it holds only if the two items vary along only *one dimension* (cf. Boersma 1998: 111f). Criticism of geometric models regards the assumptions of minimality, symmetry and triangle inequality (see section 12.3.2) and has led Tversky (1977) to formulate an alternative set-theoretic approach, the contrast model based on *feature matching* (see 12.2.2.2).

Partly in response to problems with these two approaches, researchers have developed other types of models. In response to Tversky (1977), Krumhansl (1978) proposed the *distance density model*. Moreover, *alignment-based models* of similarity, wherein – additionally to feature mapping – correspondences or alignment of features to one another are considered and matching features are aligned to the extent that they have a similar function within their entities (cf. Goldstone 1994; Goldstone & Yun Son 2005: 24f; Markman & Gentner 1993). In *transformational models* (e.g. Garner 1978), similarity between two items is considered to be inversely proportional to the number of transformational operations from one representation into another. This approach has been recently revived (for a more detailed discussion of the different approaches, see Goldstone & Yun Son 2012). The following sections will discuss briefly major assumptions and problems for *feature-based* and *geometrical models* that are relevant for the modelling of similarity, perceptual distance and asymmetry. Geometrical models are of particular interest in the present study because they provide a spatial representation model of the perceptual vowel space for L2 learners. The technique of Multidimensional Scaling to represent the perceptual vowel maps in chapter 9 and 10 are based on the theoretical assumptions of geometric models (see section 12.2.3).

12.2.2.2 Feature-based models of similarity

Tversky (1977) questioned the metric and dimensional assumptions of geometric approaches to similarity and proposed a *contrast model*, where similarity is determined by matching *features* of compared entities. There are no restrictions on what may form a feature. Features may be concrete or abstract and may refer to any property, characteristic or aspect of a stimulus (Goldstone & Yun Son 2012). There are common features as well as distinctive features.

Describing elements in terms of feature sets has a long tradition in linguistics, especially in phonology and semantics.

In a feature-based approach, similarity S is expressed as a function of common and distinctive features (Tversky 1977: 332, 345) and is computed by

$$S(a,b) = \theta f(A \cap B) - \alpha f(A-B) - \beta f(B-A)$$

where *(A∩B)* refers to common features, *(A-B)* to features that *A* has but *B* does not, and *(B-A)* to features that *b* has but *a* has not. θ, α, and β are weights for the common and distinctive components. Since α is not constrained to equal β, contrast models can predict *asymmetry*, addressing one of the major problems with geometric models.

A widely accepted relation exists between *features* or properties of objects and *classes* to which objects belong. This correspondence of features and classes links feature-theory with the clustering approaches to proximity data. *Hierarchical clusters* (hcluster) and trees provide useful graphic representations of different classes to which objects belong to and can illustrate generalizations of similarity data. Subjective interpretations may be given post-hoc to the branches of an hcluster tree-diagram (see the hclusters in section 5.7).

An important contribution of Tversky (1977) refers to the issue that *similarity* – though basic for determining identification and categorization performance – should not be treated as a primitive element and is not invariant but varies across different *contexts* in a constrained and systematic way. Similarity largely depends on context and frame of reference. The frame of reference may be explicitly specified or inferred from the general context (Tversky 1977: 340). Applied to L2 vowel perception, when asked to assess the similarity of two given vowel sounds, subjects may infer some features they may think of as relevant, e.g. rounding or length or constriction degree, though the relevant frame may include more than only those features that are consciously available to the subjects. Moreover, the relative *weights* assigned to these features may differ inter-individually as well as language-specifically and depend on situation, context and stimuli set: "*In the present theory, changes in context or frame of reference correspond to changes in the measure of the feature space. In addition (...) the salience of features and hence the similarity of objects are also influenced by the effective context (i.e., the set of objects under consideration).*" (Tversky 1977: 340f).

Tversky (1977) moreover refers to the relevance of relative *salience* of features in a given stimuli set of objects under consideration. The salience of features has two components: intensity and diagnosticity. While the *intensity* of a feature is determined by perceptual and cognitive factors, *diagnosticity* changes with the reduction or enlargement of the item set. Within a set, where all objects share the same features, these features have no diagnostic value. When the set of objects is enlarged by objects not sharing these features, such features gain diagnostic value and increase the similarity of all those objects, which share the same feature, while increasing the dissimilarity of objects (for the diagnosticity principle and the extension effect, see Tversky 1977: 343f).

Tversky's *extension effect* is especially relevant for studies using identification tasks with respect to the selection of input stimuli and response categories. While a larger set of response categories is rather uncommon in perceptual studies, in the present study the full set of 15 German vowels and three diphthongs were presented as response labels. Including or excluding the response label <ä:> (/ɛ:/) or the three diphthongs does have an influence on the listeners' response behaviour, as is especially evident in the responses for e-stimuli /e:/, /ɛ/ and /ɛ:/. Responses for German *e*-vowels would differ considerably when /ɛ:/-stimuli or the response label <ä:> would be excluded. Likely, the fact that the full set of German vowels was included in the study as input stimuli and as well as response label rather than presenting only a sub-set of German vowels (e.g. only front vowels) does have an effect on the listeners' response behaviour, specifically for the class of front vowels that are frequently identified as back rounded or front non-rounded vowels in the data.

Referring to the dynamic interplay between similarity and classification, Tversky states that it is

> *"generally assumed that classifications are determined by similarities among the objects. The preceding discussion supports the converse hypothesis: that the similarity of objects is modified by the manner in which they are classified. Thus, similarity has two faces: causal and derivative. It serves as a basis for the classification of objects, but it is also influenced by the adopted classification."* Tversky (1977: 344) [141]

To summarize, a feature-based account for similarity makes essential contributions to an understanding of similarity relationships between vowel categories in a given phonemic system. It is strongly compatible with traditional phonological descriptions in terms of sets of distinctive features. However, the set of distinctive qualities in the vowel inventory and the set of common and distinctive features and their relative salience (weight) can vary language-specifically. Therefore, a theory or model of vowel similarity has to include context-dependent, and particularly *set-dependent* changes in performance (see also Nosofsky 1986). In L1 perception, the language-specific set of distinctive vowel qualities determines the salience and diagnosticity of specific features and the similarity of categories in the system. In L2, it is the specific set of distinctive categories as established by the L2 learner based on his/her knowledge and experience in L2 at a given point in time that determines the relative salience or *weight* of particular features and the similarity relationships between categories of the target language.

[141] emphasis mine.

In L2 vowel perception, the selection of features that are of relevance to the listeners in L2 perception is a crucial question. However, we do not know a priori and can only guess which features will be distinctive in L2 perception. we also do not know about the intensity and diagnosticity of specific features as perceived by L2 learners. Therefore, a consideration of different approaches to similarity seems rewarding to gain more insights into the particular set-specific constellations of similarity in L2 perception.

12.2.3 Multidimensional Scaling – Mapping the perceptual space

In geometric models of similarity (cf. the classical studies by Torgersen 1965; Shepard 1957, 1972), entities of a set of objects are considered to be situated in an *n*-dimensional space. Several objects, concepts or situations in human life seem to be psychologically structured in terms of dimensions. Geometric representations and interpretations are designed to capture some part of the structure and dimensionality of mental organization (Goldstone & Yun 2005: 17). Distances between items in the perceptual space can be visualized by *Multidimensional Scaling* (MDS). The similarity of two items *i* and *j* is taken to be inversely related to their distance. The most widely assumed metric in MDS is the Euclidean distance ($r = 2$) or the Minkowski-*r*-metric distance defined as

$$d_{ij} = \left[\sum_{k=1}^{n} \left| x_{ik} - x_{jk} \right|^{r} \right]^{1/r}$$

where n is the number of dimensions, X_{ik} is the value of dimension k for item i, and r is a parameter that allows different spatial metrics to be used[142].

Similarity scaling techniques such as Multidimensional Scaling or Hierarchical Clustering (for discussion, see Shepard 1980) provide insights into the underlying basis of similarity data reflecting item properties of individual items as well as inter-item similarities. *Multidimensional Scaling* (Shepard 1957, 1972, 1980; Torgersen 1965; Nosofsky 1986, 1987) is a technique of reducing the number of parameters in which stimuli may vary to a low-dimensional representation of the psychological space (Shepard 1957, 1972, 1980). MDS seeks for an *n*-dimensional solution that characterizes best the distances between items of a given stimuli set. The position of an item's place is relative to other items in this space. The MDS technique is used to visualize similarities or dissimilarities observed in the data. The

[142] The value of r depends on the type of dimensions that compose the stimuli (Nosofsky 1986). $r = 2$ invokes the standard Euclidean distance with the underlying assumption that at least some stimulus dimensions are not separately perceivable, considering integral-dimension stimuli. If $r = 1$, then distance involves a city-block metric, i.e. the sum of the distances on each dimension for stimulus dimensions that are assumed to be perceptually separable (Nosofsky 1986; Wilson 2006).

representation is constructed either from different kinds of *proximity data* such as perceived similarities or from *distances*. Proximity data comprise a wide variety of similarity data, such as similarity ratings, identification confusions, same-different errors, or latencies of discriminative responses.

A crucial advantage of MS is a reduction of the many parameters or dimensions in phonetic reality to a *low*-dimensional representation of the perceptual space. MDS seeks for the configuration of items in a low-dimensional space that accounts best for the given identification data. The number of dimensions in MDS solutions is not limited, the most common is a two-dimensional representation (for discussion, see section 7.3). The dimensions may be continuous and multivalued. Similarity parameters, it is argued, can be assumed to be functionally related to the *distances* in a multidimensional space: “*Once a multidimensional scaling solution for the stimulus set is derived, the similarity between any two stimuli will be a function of their distance in the psychological space.*” (Nosofsky 1986: 43). A strong connection of distances between items in the psychological space and their probability to be confused is moreover assumed in this paradigm (Shepard 1957; Nosofsky 1986, 1987).

Shepard (1957, 1980) emphasizes the advantages of a *psychological* compared to a psychophysical approach to similarity between items, which is especially useful to model the way humans process complex naturalistic stimuli such as colours, speech sounds, words or faces, for which the relevant physical dimensions are as yet not sufficiently characterized or known or where the physical stimulus does not even contain the relevant semantic dimensions of symbolic stimuli such as words. A psychological approach based on proximity data is considered to provide relevant information on the processing of complex naturalistic stimuli without any quantitative information about the physical properties of speech sounds necessary (Shepard 1980).

It is important to underline that geometric representations like the MDS solutions in section 9.5.2 based on confusion data do not provide a representation of the physical characteristics of vowel sounds but refer to the *psychological representation* of listeners from different language background. The present study introduces Multidimensional Scaling for a low-dimensional representation of the listeners’ *psychological vowel space*. Similarities and perceptual distances between vowel categories are assumed to directly reflect the language-specific selective attention patterns of native listeners of a given language (Nosofsky 1991). However, an essential distinction between stimulus-related vs. response-related aspects of similarity is necessary (see section 12.4).

The psychological representation is a listener- or language-specific form of the multi-dimensional *phonetic vowel space* that is defined in terms of several articulatory and acoustic phonetic parameters. MDS allows a reduction of the many dimensions of the phonetic vowel space of a given language to a low-dimensional psychological vowel space (Shepard 1957, 1980). MDS solutions for the perceptual vowel space of L2 learners of German are presented for each of the ten language sub-groups (see section 9.5). The position of a vowel type and its distance to all other categories in the system is determined by the response behaviour of L2 learners in a given language sub-sample and varies systematically with respect to the listeners' native language background.

In a post-hoc interpretation, the listeners' perceptual vowel space as modelled by the MDS solutions for each of the language sub-groups can be described in terms of two major psychological dimensions: (1) front vs. back and (2) rounded vs. non-rounded vowel categories. These phonological dimensions, however, do not directly correspond to the axes of abscissas and ordinates in a two-dimensional MDS solution. It has to be emphasized that no direct correlation of the two or three axes of an MDS representation and specific phonetic parameters can be established (Shepard 1972), an aspect that has not been sufficiently considered in several previous studies using MDS (e.g. Fox, Flege & Munro 1995; Bradlow, Clopper & Smiljanic 2007; for discussion, see section 7.3).

While the idea of a spatial geometric representation of the perceptual space is very appealing, it has been challenged by Tversky and colleagues (Tversky 1977; Tversky & Gati 1982) in feature-based accounts (see above). Criticism to geometric approaches referred to three major characteristics for which violations are empirically observed: Geometric models assume (1) *minimality* ($d_{ij} \geq d_{ji} = 0$), (2) *symmetry* ($d_{ij} = d_{ji}$), and (3) *triangle inequality* ($d_{ij} + d_{jk} \geq d_{ik}$). These three assumptions of geometric models are challenged by data from several empirical studies, which – as the data in the present study – show (1) that not all "identical" objects seem to be equally similar to each other because of acoustic variation, (2) that *i* may be judged as less similar to *j* than *j* to *i* (cf. asymmetries in substitutions observed in the data), and (3) that by distances derived from similarity scores (see section 7.2.5) may violate the condition of triangle inequality or – in terms of a feature-based approach – *i* and *j* may share a feature, and *j* and *k* may share a feature, though *i* and *k* have no feature in common. These challenges to geometric approaches – specifically the observed asymmetry in difficulties and preferences for some categories – will be discussed in the following sections in order to provide necessary complementary elements for a theory of similarity in L2 vowel perception.

12.2.4 Measures of distance and similarity

Theories of vowel contrasts and vowel inventory typology in the languages of the world have two central tenets in common that refer to (1) a preference for vowel qualities that are maximally distinct and (2) the assumption that the constellation of vowel systems is considered as motivated by maximal contrast (at minimal articulatory cost) (see Liljencrants & Lindblom 1972; Lindblom 1986; de Boer 2000, 2001; Flemming 2004, 2008; Schwartz et al. 2005).

From these assumptions two questions arise: (1) How can perceptual *distinctiveness* be defined? (2) How can perceptual distance and optimal or maximal *contrast* be measured and compared? The second question is related to inter-category *distances* of categories in terms of phonetic similarity and psychological distance.

A common assumption underlying most models of similarity is that there is an inversely functional correspondence between *distance d* and *similarity* $\eta_{ij} = f(d_{ij})$ (cf. Shepard 1957). It is also assumed that confusion of categories in perception experiments can be predicted from distance measures, i.e. the greater the similarity of the two items in the system the more confusable a pair of sounds and vice versa.

Referring to distance, the smaller the distance between two items and the smaller the number of distinctive features that differentiate them, the higher their similarity and the expected *probability for confusion*. Statements about *distances* between two given items and their relative *similarity* in a psychological space can be expressed in two basic ways, irrespective of what dimensions or parameters are considered for description:

(a) the *Euclidean distance*, which is the length of the direct line between two given points,

(b) the *rectilinear distance* or "*city block*"- or "taxicab"- or L_1-distance between two points, which is the sum of the absolute differences of their coordinates and can be computed by simply counting the number of features in which two items differ.

Both approaches attempt to define a discrete measure of contrast or distance between two items.

The *Euclidean distance*, i.e. the length of the straight line connecting two points in an *n*-dimensional space, is calculated by the Pythagorean formula:

$$d_{(i,j)} = \sqrt{\sum_{k=1}^{n} \left| i_k - j_k \right|^2}$$

For a vowel contrast such as /i:/ - /ɛ/, the Euclidean distance could be calculated e.g. by their distance in a two-dimensional F1xF2 representation of a given vowel system, as is done in many studies (despite the limited scope of a two-dimensional F1xF2-representation, see above).

Rectilinear, L_1, city-block distance or taxicab metric commonly refers to *feature-based* descriptions, where the dimensions of description are conceived as perceptually separable and the distance of *i* and *j* is computed by the formula

$$d_{(i,j)} = \|i - j\|_1 = \sum_{k=1}^{n} |i_k - j_k|$$

Figure 12.1 compares the two approaches. The Euclidean distance corresponds to the length of the direct line between /i:/ and /ɛ/ (dashed line). The rectilinear distance between /i:/ and /ɛ/ (dotted lines) can be computed by counting the sum of the line segments between the points along the coordinate axes in a two-dimensional representation from /i:/ to /ɛ/ via /e:/ or from /i:/ to /ɛ/ via /ɪ/, which is longer than the distance between /i:/ and /e:/ or /i:/ and /ɪ/. The distance $d_{(i:\ \varepsilon)} = 2$ corresponds to the number of distinctive features.

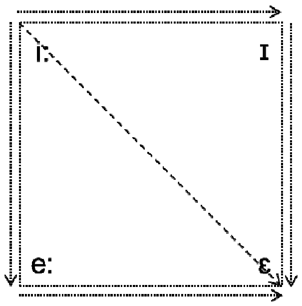

Figure 12.1: Schematic representation of rectilinear distance (dotted lines) vs. Euclidean distance (dashed line)

In a feature-based approach, *rectilinear distances* between stimuli in a three-dimensional representation of a set of items differing in the presence or absence of specific features are illustrated in Figure 12.2. The right, top and back faces of the cube represent the features under consideration, *1* indicates the presence of a given feature and *0* its absence. The right side of the cube represents all those items where feature 1 is present, the upper side those where feature 2 is present, and the back side those objects that share feature 3.

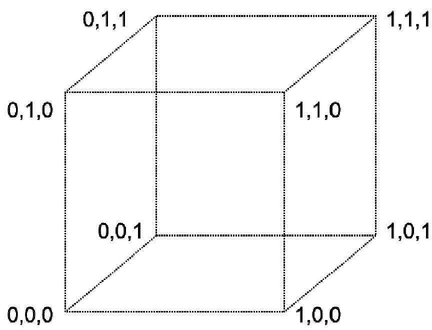

Figure 12.2: Three-dimensional cube illustrating city-block distances in a feature-based description (inspired by Nosofsky 1991: 105)

This scheme can be transferred, for example, to a representation of the eight vowel categories of Turkish (see Figure 12.3). Here, vowels at the front side of the cube share the feature [front], vowels on the top side share the feature [high] or [closed] and vowels at the right side share the feature [round].

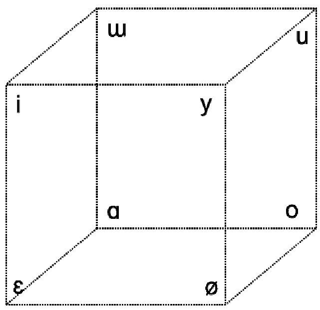

Figure 12.3: Schematic representation of eight basic vowel qualities of Turkish

This scheme is expanded to the set of German front vowels [+/–round] (left) and rounded vowels [+/–back] (right). Figure 12.4 visualizes differences between vowel categories in terms of distinctive features or rectilinear distances for front rounded and non-rounded vowels on the diagram's left side and rounded front and back rounded vowels on the right side of the diagram.

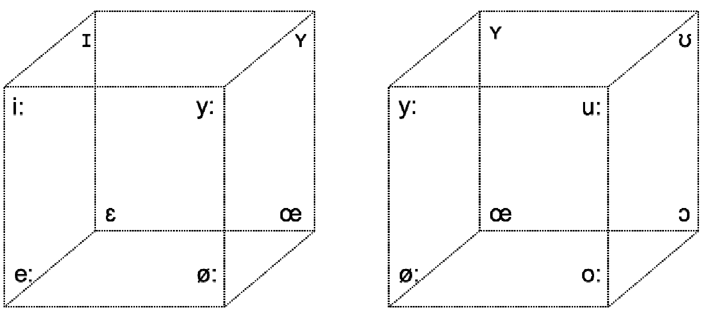

Figure 12.4: Schematic three-dimensional representation of German front (left) and rounded vowels (right)

Differences in number of distinctive features between members of a contrast pair is assumed to reflect the relative probability of perceptual confusion and substitution processes.
It is hypothesized here that less distant, i.e. more similar categories are more susceptible to mutual confusion than more distant categories. Moreover, less distant, i.e. more similar categories are more preferred response option. In other words, they are more likely to substitute a given vowel than less similar categories differing in more *n+1* features.
An analysis of patterns of perceptual confusion and substitution processes in the data of the present study (see section 11.2) shows that confusions and perceptual substitutions occur more frequently between less distant categories, which differ only in one feature (d = 1) than between categories that differ in more than one feature. German /o:/, for example, is substituted with different vowel types by non-native listeners (see Table 9.5). The target categories of these substitution processes differ in percentage as well as in number of differentiating features: /o:/ > /u:/ (12%, $d_{(\text{o: u:})}$ = 1), /o:/ > /ø:/ (6%, $d_{(\text{o: ø:})}$ = 1), /o:/ > /ʊ/ (4%, $d_{(\text{o: ʊ})}$ = 2), /o:/ > /œ/ (2%, $d_{(\text{o: œ})}$ = 2), /o:/ > /y:/ (2%, $d_{(\text{o: y:})}$ = 2), /o:/ > /ʏ/ (1%, $d_{(\text{o: ʏ})}$ = 3). Less distant categories serve more frequently as substiution targets than more distant ones.
The analysis shows that measures for distances in terms of number of distinctive features between items of a given set, here between categories of a language-specific vowel system, can contribute essentially to predict and/or explain the *probability of category confusion* and the possible targets of perceptual substitutions.
Rectilinear distances can serve to predict the relative probability of category confusion. However, there are two major aspects that have to be considered in a rectilinear, feature-based approach: (1) The *choice of descriptive features* that adequately reflect the physical characteristics of the input may differ from the set of features that are actually used by listeners in perception, and (2) the relative *weight or salience of a specific feature* in the listeners' perception may vary language-specifically, from category to category, or even inter-individually according to the listeners' experiences and hypotheses.
Euclidean distances between German vowels referring to *acoustic* properties can be calculated e.g. in a two-dimensional representation of the formant frequencies for the German vowel system as in Figure 12.7. For a critical discussion of Euclidean distances in the acoustic vowel space, see section 12.3.3.3.
Euclidean distance could also be used as distance measure of vowels in the perceptual space of L2 learners in the MDS configurations in Figure 9.18, which are based on the L2 listeners' wrong and correct identifications and on derived sim_scores and distance scores calculated by Shepard's fomula (see section 7.3).

Since natural vowel stimuli differ in a large number of phonetic parameters or dimensions (see section 12.1.2), measures of *acoustic similarity* cannot directly account for *perceived similarity* in terms of presence or absence of specific features or phonetic properties.

Basically, there are two major approaches to identify relationships or similarity between vowel categories: (1) An *a priori* approach, which identifies distances in terms of Euclidean or rectilinear distances to predict similarity and confusion patterns and (2) an analysis of confusion data, i.e. the listeners' behaviour in an identification task, from which perceived distances and similarities a derived in an *a posteriori* analysis to derive information on the learners' mental representations of L2 vowel categories. The perceptual space can be derived from confusion and similarity data. The advantage of geometric models is a low-dimensional spatial representation. Any distance measure can be adapted to any dimensionality, even if it is difficult to visualize a space of more than three dimensions (Shepard 1957, 1972; Tversky 1977; Krumhansl 1978; Nosofsky 1986, 1987).

In other words, *Confusion probability* can be considered as a more global contrast measure that combines various perceptual dimensions. The position of categories in MDS configurations as in section 9.5 and the distances between them are based on confusion data and reflect the *perceived* distances between categories in a vowel set by L2 listeners. While discrete measures of distance refer to specific properties of *stimuli*, confusion probability rather refers to listeners' reactions, i.e. behavioural *responses* to stimuli. However, as discussed above, the MDS dimensions in a two- or three-dimensional MDS representation must not be interpreted as correlating directly with dimensions of phonetic or phonological description due to the multidimensional character of the vowel space (see section 7.3).

The issues of perceived similarity and confusion probability, the advantages of the spatial metaphor and the interrelation between stimuli and responses will be addressed in the following sections.

12.3 Asymmetry

12.3.1 Identification and categorization

Studies on similarity in cognitive psychology use data from different types of experimental tasks. Several studies on similarity are based on data from similarity judgments of e.g. forms, faces, colours, or letters. Other approaches to make statements about distance and similarity with emphasis on perceived similarity are based on data from same-different judgments or on confusion data from identification tasks like in the present investigation.

In the experimental part of this study behavioural data were collected in an identification task to gain evidence for the perception of German vowels by listeners from different language backgrounds. The identification task consists of a forced-choice experiment in which *n* distinct stimuli each have to be assigned to one of the given response options. From the listeners' responses in the identification task, confusion data were derived. The data allow conclusions on L2 learners' language-specific *confusion probability* and on the relative perceptual distance and perceived similarity between categories within the German vowel system in L2 acquisition. Moreover, they provide evidence for the relative *difficulty* of German vowel categories and category confusion (confusion probability) and *preferences* for specific response categories.

The basic elements of a choice experiment in the *identification paradigm* (see Nosofksy 1986) are a set of *n* distinct stimuli (here $n_i = 270$), a set of response categories (here $n_j = 18$[143]) and the assignment of each of the *n* stimuli to a unique "correct" response. The listeners' responses are tabulated in an identification matrix and summarized in an *i* x *j* confusion matrix, where cell (*i, j*) gives the frequency by which a specific stimulus *i* was identified as *j*. In a *categorization paradigm*, *n* stimuli are partitioned into $m < n$ groups and each group is assigned a distinct response. The data are summarized in an *n* x *m* confusion matrix (here m = 15), where cell (*i, j*) gives the frequency a given stimulus *i* was classified as belonging to category *j* (Nosofsky 1986, 1987).

The *i* x *j identification matrix* represents a one-to-one-mapping of stimuli onto responses in identification, and is transformed into a *confusion matrix* that represents a many-to-one mapping of stimuli onto responses in categorization. "Within-class" identifications (of stimulus *i* with response *i*) result in correct identifications in the diagonal axis of the matrix, whereas "between-class" identifications or confusions (of stimulus *i* to response *j*) result in categorization errors.

It is assumed that similar principles underlie the identification and categorization of stimuli varying in *n* dimensions and that performance in these two task types is highly related (Nosofsky 1986, 1987). The underlying assumption relates to mental representations of categories (e.g. in form of exemplars and/or prototypes), where categorization decisions are based on the *similarity* of stimuli to stored representations.

[143] 15 German full vowels and 3 diphthongs were offered as response options.

Nosofsky (1986) proposes a categorization model which is a generalization of the context theory of classification developed by Medin & Schaffer (1978). The context model response rule defines the *probability* that Stimulus S_i is classified in category C_J, $P(R_J|S_i)$ by

$$P(R_J \mid S_i) = \frac{b_J \sum_{j \in C_J} \eta_{ij}}{\sum_{K=1}^{m} \left(b_K \sum_{k \in C_K} \eta_{ik} \right)}$$

where the parameter b_J represents the bias for making category response R_J, η_{ij} denotes the similarity between Stimuli S_i and S_j, and the index $_{j \in C_J}$ indicates that "*all j such that S_j is a member of C_J*" (Nosofsky 1986: 40)[144].

To predict categorization probability from identification data, cumulated stimulus-response results from the identification matrix are mapped onto a categorization matrix. Within-class confusions (S_i as R_I) result in correct categorization responses, whereas between-class confusions result in wrong categorizations (mapping hypothesis).

No direct relation between identification and categorization performance has to be assumed (Nosofsky 1986). The similarity parameters η_{ij} are not assumed to be invariant in identification and categorization paradigms. Though assuming the same basic multidimensional perceptual representation underlying performance in identification and categorization, a *selective attention process* is assumed to operate on the multidimensional representation of the perceptual space that underlies performance in identification and categorization yielding systematic changes in inter-stimulus similarity relations: "*Selective attention is modeled by differential weighting of the component dimensions in the psychological space, (...). In geometric terms, these weights act to stretch or shrink the psychological space along its coordinate axes.*" (Nosofksy 1986: 41)

Nosofsky (1986) formalizes the *selective attention* process by augmenting the distance formula and adding attentional weight parameters w_k [145]as

$$d_{ij} = \left[\sum_{k=1}^{N} w_k \left| x_{ik} - x_{jk} \right|^r \right]^{1/r}$$

where $0 \leq c\ \infty$, $0 \leq w_k \leq 1$, and $\sum w_k = 1$. Parameter *c* is a scale parameter reflecting overall discriminability in the psychological space; it is expected to increase with e.g. an increase in exposure duration or as subjects gain increased experience with the stimuli (Nosofsky 1986:

[144] Uppercase letters refer to categories (C_I) and categorization processes, lower case letters index individual stimuli (S_i) and identification processes.

[145] weight parameters in *k* dimensions

41). The scale parameter and the notion of selective attention is a useful construct to describe bias in language acquisition and will be discussed in more details in section 12.5.

12.3.2 Asymmetry, self-similarity and distance inequality

Standard geometric models of similarity are based on three basic assumptions: (1) *minimality* ($d_{ij} \geq d_{ji} = 0$), (2) *symmetry* ($d_{ij} = d_{ji}$), and (3) *triangle inequality* ($d_{ij} + d_{jk} \geq d_{ik}$). However, as similarity data from several studies show, these principles are violated by data gained in different experimental contexts:

(1) *Minimality* ($d_{ij} \geq d_{ii} = 0$) refers to the assumption that all objects are equally (dis)similar to each other and that in a metric space the distance between a point and itself must be zero and must be smaller than the distance between any two distinct points in the space. In terms of similarity, the similarity of an object and itself must be smaller than the similarity between this object and any other object in the set. Referring to confusion data, *self-similarity* becomes apparent by diagonal entries in the confusion matrix (Nosofski 1991: 101; Krumhansl 1978). The principle of minimality is violated in confusion data if the number of off-diagonal entries (wrong identifications) exceeds the number of some diagonal entries (correct identifications). In the present L2 data, this is the case for items of German /ɛ:/. /ɛ:/ is identified (incorrectly) as /e:/ in 49% of the cases, but identified as /ɛ:/ in only 33%. This pattern is also observed in several L1 sub-samples (see Table 11.14).

(2) *Symmetry* ($d_{ij} = d_{ji}$) refers to the assumption that the (dis)similarity of *i* and *j* equals the (dis)similarity of *j* and *i* because the distance between *i* and *j* is equal to the distance of *j* and *i*. However, as evidence from experiments using similarity judgments (see Tversky 1977) as well as confusion data show, an equality of distance and similarity is not observed in many studies. Asymmetries in the present data as described category-specifically in chapter 11 show that there are several cases of substitution types that occur more often in one direction (e.g. /e:/ > /i:/ 12%) than in the opposite direction (/i:/ > /e:/ 3%). The problem of *(a)symmetry* is specifically relevant for the present study to describe these directional differences in substitution patterns and will be discussed in more details in the following sections.

(3) *Triangle inequality* refers to the assumption that $d_{ij} + d_{jk} \geq d_{ik}$, but (in terms of a feature-based account) *i* and *j* may share a feature, and *j* and *k* may share a feature, though *i* and *k* have no feature in common.

The issue of *asymmetry* $d_{ij} \neq d_{ji}$ in confusion data is a specific challenge for models of similarity but is of central interest in the present study. Asymmetry is a challenge to accounts for differences in confusion probability and for the relative difficulty of vowels in L2, for the susceptibility to perceptual substitution processes, the direction of substitutions and for the observed preferences for specific response categories in the listeners' responses.

Numerous studies provided evidence for asymmetric similarity patterns showing that the proximity between *i* and *j* cannot be expected to be the same in both directions. That is, *i* may be confused with *j* far more often than *j* with *i*. In other words, subjects are expected to behave asymmetrically in responding to stimuli *i* and *j*; d_{ij} and d_{ji} will therefore not be equal for the same as well as for different stimuli. Referring to the cuboids in Figure 12.3 or Figure 12.4 this would imply that the distance from one corner to the other in one direction would not be equivalent to the distance between the two corners in the opposite direction. The data of the present study shows that this is the case for several vowel contrasts, where substitutions in one direction (e.g. /e:/ > /i:/, /ø:/ > /y:/) occur more frequently than in the opposite direction (/i:/ > /e:/, /y:/ > /ø:/) at least with some contrast pairs (see section 11.2). Such patterns can vary language-specifically, though there are also common tendencies observed.

Different reasons have been proposed to account for patterns of asymmetry in perception data. In many linguistic studies, patterns of asymmetry in L2 perceptual confusion data have been related to *theoretical constructs* such as markedness, naturalness, salience, robustness, maximal contrast, or prototypicality and/or to specific *acoustic characteristics* in the signal to account for biases for specific vowel qualities (see chapter 1). The underlying assumption common to all these concepts is that one category has specific characteristics that the other one has not. Several authors refer to "*salience*" or *perceptual bias*" towards specific sounds that are in some respect "better", "clearer" or more "prominent" in perception, e.g.:

> "*Natural phonology attributes such biases to processes with both perceptual and articulatory motivations, and it thus constitutes a full-blown theory of 'perceptual bias'. 'Perceptual bias' may in fact be bias toward segments with <u>better perceptual properties</u> ([i] is more clearly palatal than [y]) <u>or less demanding articulations</u> (stops require less precision than fricatives), or toward sequences with less demanding articulations …*" (Donegan & Nathan, forthc: 21)[146].

12.3.3 Asymmetry and bias

To account for asymmetry in perception, psychological models of similarity have proposed additional elements such as *bias*, *weight* or *density* (Shepard 1957; Rosch 1973, 1975; Tversky 1977; Goldstein 1977; Krumhansl 1978; Holman 1979; Nosofsky 1991) to account for differences in probability of category selection.

[146] emphasis mine.

In a feature-based approach, Tversky (1977) assumes that the direction of *asymmetry* is determined by the *relative prominence* of the stimuli and that prominence is related to factors such as salience, intensity, frequency, familiarity, or goodness in form and information content (see section 12.2.2.2). Less prominent objects are more frequently perceived as similar to more prominent ones than vice versa. This fact may be determined by the relative salience of the stimuli, where the variant is found to be more similar to the prototype than vice versa. In a feature-based approach, asymmetries in similarity s_{ij} and s_{ji} can be accounted for by differences in weight α and β, reflecting a greater *focus* on one of the objects compared to the other (cf. section 12.2.2.2; see also Krumhansl 1978: 452), which could be easily tested in a full similarity matrix by comparing the variance in row sums and column sums. The focusing hypothesis ($\alpha > \beta$) holds if more variance is found in the column sums, indicating that some features of *i* not shared by *j* detract more from the similarity η_{ij} of *i* to *j* than those features of *j* not shared by *i*. If $\alpha > \beta$, it follows that η_{ij} will be greater than η_{ji} if and only if *f*(B)[147] is larger than *f*(A) (cf. Tversky 1977; Krumhansl 1978: 452). In other words, in a feature-based account, *asymmetries* due to *unequal weight* are expected when one object has more features or more *salient* features than the other (cf. Krumhansl 1978: 452).

The *choice model* (Luce 1963; Shepard 1957; Nosofsky 1986, 1991: 112) defines the *probability* that stimulus *i* is identified as response *j* by

$$P(R_j|S_i) = \frac{b_j \eta_{ij}}{\sum_{k=1}^{n} b_k \eta_{ik}}$$

where the b_j parameters ($0 \leq b_j \leq 1$, $\sum b_j = 1$) is the *bias* associated with item *j*, and the η_{ij} parameters ($0 \leq \eta_{ij}$, $\eta_{ij} = \eta_{ji}$, $\eta_{ii} = 1$) are similarity measures on the stimuli S_i and S_j. Basically, similarity in the choice model is assumed to be symmetric ($\eta_{ij} = \eta_{ji}$). However, sufficient evidence is given that confusion data often are not symmetrical. The essence of the choice model is that the probability P_{ij} can be represented as the product of a symmetric function of *i* and *j* and a *bias* function on *j*. The model assumes that any observed asymmetry between P_{ij} and P_{ji} can be accounted for by a difference in the relative *response bias*. Nosofsky (1991) argues that symmetric choice models were found to work due to a higher number of parameters (for *n* stimuli, there are $n - 1$ freely varying *bias* parameters *b* that are part of the model (Nosofsky 1991: 95).

[147] *f* refers here to the measure or weight assigned to the features, *A* and *B* are the features of *i* and *j* respectively (cf. Tversky 1977; Krumhansl 1978).

To summarize, in order to cope with specific problems such as the violation of symmetry, minimality and self-similarity (for discussion, see Tversky 1977; Krumhansl 1978; Goldstone & Yun Son 2005: 18ff), additional constructs such as *bias*, *weight, prominence* or *salience* have been introduced to models of similarity in perception, e.g. by Krumhansl (1978) proposing the *distance-density* model or by Nosofsky (1986) by adding *bias* to the MDS choice model or by Nosofsky (1991) *combining inter-item distances and biases* towards particular items.

Several studies have proposed the idea that some objects within a given set of objects may be more prominent, more salient, more (proto)typical, more extreme or more distinctive than others. These are assumed to occur more frequently in the language systems of the world. Prominent items may function as "reference points" or "perceptual anchors" to which others are compared or set in relation to. Moreover, higher self-similarity may be assumed for prominent items. Less prominent or less salient items are more frequently perceived as similar to more prominent ones than vice versa (cf. Shepard 1957; Rosch 1975; Tversky 1977; Krumhansl 1978).

Referring to the acquisition of vowels and vowel systems in L2, several concepts have been proposed to account for difficulty and preferences for specific categories within a vowel system as well as across language systems. Some of the basic notions and constructs have been discussed in previous chapters, some shall be reviewed in the following sections. A discussion of major theoretical constructs in L2 acquisition – transfer, cross-linguistic influence, similarity, markedness and universals – has been offered in chapter 1 and chapter 3. Similarity as one of the central constructs in studies and models of L2 speech perception and the acquisition of a foreign language phonology has been discussed in section 2.4 and chapter 3. Theories on the organization of vowel systems and the preference for specific categories were discussed in chaper 4. A contrastive description of differences between the learner's native language and the target language German is provided in chapter 10 for each of the listeners' native languages. The following sections will discuss four theoretical constructs – *peripherality*, *frequency*, *density* and *self-similarity* – that have been used in phonetic and phonological studies and/or studies in cognitive psychology on asymmetry in similarity data in order to account for asymmetry and bias in perceptual data, patterns in the organization of vowel systems and in language acquisition.

The functionof these constructs as *weighting factors* and their contribution to account for bias in L2 vowel identification will be discussed referring to the current L2 confusion data.

12.3.3.1 Peripherality

For vowels, a universal prominent status of /i u a/ has been proposed in several phonetic and phonological studies. /i u a/ are described as “quantal” by Stevens (1972, 1989, 1998) or “peripheral” or “focal” (e.g. Polka & Bohn 2003; Schwartz et al. 2005; Vaissière 2011) located at phonetic “hot spots” (Stevens 1989). The three so-called corner vowels or point vowels /i u a/ are described as exploiting the vowel space to the maximum and are therefore also termed *peripheral* vowels located at the “extremes” or “margins” of the vowel space. They are moreover described as acoustically maximally distinct, perceptually easiest and most resistant to misperception. As a set, these vowel qualities can constitute a “typical”, “complete” vowel inventory (Maddieson 1984; Lindblom 1986, 1989; Crosswhite 2004). They are clearly allocated to specific regions of the articulatory and acoustic vowel space, so called “hot spots”, where acoustic stability is given despite considerable articulatory variation. In terms of their production, they all show quantal effects (Stevens 1989) allowing for considerable variation by slight articulatory displacements that do not produce correspondingly strong acoustic or perceptual changes. /i u a/ show typical patterns of formant merging creating sharp spectral energy in a well-defined frequency range by which articulatory and auditory stability is preserved (Stevens 1989; Vaissière 2011).

Vaissière (2011) proposes an acoustic definition of a set of “focal” reference vowels that show formant merging. Formant merging it is argued enhances spectral energy in specific areas of the frequency scale and guarantees articulatory and perceptual stability.While for /u/ the lowest possible concentration of energy is given by F1F2 merger at 400 Hz, the highest possible clustering of the two first formants is stated for /ɑ/ at around 1000 Hz; /i/, articulated in the prepalatal region, shows F3F4 merging at about 3200 Hz, whereas prepalatal /y/ shows F2F3 convergence around 1900 Hz due to the high sensitivity of F3 to lip rounding.

The acoustic stability of these vowel types seems to be reflected by their occurrence in the vowel systems of the world’s languages. In general, peripheral vowel qualities situated along the “margins” of the vowel space are typologically more frequent than non-peripheral qualities. Among 3-vowel systems, systems with three corner vowels /i u a/ occur most frequently: In the UPSID material analysed by Maddieson (1984: 125), /i/ is counted in 91.5%, /a/ in 88% and /u/ in 83.9% of the analysed systems (for a detailed discussion of frequency and distribution in the world’s languages, see chapter 4).

Theories of vowel systems and typology address the prominent status of /i u a/ in different ways. Jakobson (1941 [1968]) posits /a i u/ universally as first vowels in language acquisition. Dressler (1975) refers to the most basic classification of vowels as related to a distinction of central and peripheral vowels and divides basic vowel processes in centrifugal and centripetal processes. Peripheral vowels appear to constitute a natural class enrolled in diachronic changes in vowel systems by processes striving after maximal distances between phonemes, which are best achieved by centrifugal processes.

In a very different theoretical framework, Harris & Lindsey (1995, 2000) address the notion that /i u a/ form a natural class vis-á-vis mid vowels. /i u a/ are described as 'primitive' segments or autonomous realizations of the primes [I U A] with stand-alone phonetic interpretability not requiring support from other primes. Non-primitive or more complex vowels such as /e/ or /o/ are represented as compounds of such primes, e.g. /e/ as [A, I], /o/ as [A, U]. The description of these primes as "elements" is most usually associated with Government Phonology (cf. Kaye, Lowenstam & Vergnaud 1985; Rennison 1990), though the notion of autonomously existing *primes* exists in several other approaches as well (for discussion, see Harris & Lindsey 1995). /i u a/ are also associated with the assumption of primacy and simplicity as opposed to compositionality and complexity (Krumhansl 1978; Harris & Lindsey 1995). Therefore, *peripheral vowels* have been attributed to be *unmarked*, whereas vowels in acoustically less stable regions are considered to be more marked.

To generalize, peripheral /i u a/ can be considered as forming a natural class of universally preferred and perceptually stable qualities and primary contrasts that are assumed as articulatorily and acoustically more stable and less "difficult" than non-peripheral vowels. They are moreover expected to be psychologically more prominent. Since /i u a/ occupy areas of acoustic stability, their perception is expected to be more or less consistent along a certain range of different articulations. They are expected to be more or even most resistant to misperception, due to their acoustic stability and perceptual salience. Language-specific (sub-phonemic) variation in articulation and acoustics between L1 and L2 are not expected to have much influence on their correct identification, whereas the allocation of non-peripheral vowels and their acoustic and perceptual quality may show more language-specific variation. In other words, peripheral vowels are considered to be least "difficult" in L2 perception. This assumption is tested on the present L2 confusion data.

With respect to *correctness* in identification, the analysis of the present data shows high id_correct scores for peripheral *a*-, *i*- and *u*-vowels (see Table 9.1 and Figure 9.1), even if high id_correct scores are also observed for other categories. The highest percentage of correct identifications (id_correct) is given for /ɑ:/ (84%), /i:/ (80%) and /ɑ/ (76%), followed by /ɛ/ and /ɔ/ (both 74% wrong) and /u:/ (71%), /y:/ (71%), /ɪ/ (68%) and /ʊ/ (66%) (see Figure 9.1). However, the high id_correct scores for /ɛ/, /ɔ/ or /y:/ show that it is not (only) peripherality in a narrow sense referring to the "corner vowels" /iː/, /uː/ /ɑ/ and /ɑː/ that accounts for the observed id_correct scores. Therefore, either a wider scope of the notion "peripherality" or additional factors have to be considered.

Concerning the scope of the notion "peripherality" in a wider or narrower sense, questions arise with respect to (1) whether /ɛ/ and /ɔ/ should also be included in the set of peripheral vowels, (2) whether /yː/ should be included, and (3) whether /ɪ/, /ʊ/ and even /ʏ/ could be referred to as "peripheral". If peripherality is understood as referring to *articulatory extremeness* along the margins of the articulatory vowel space, German /yː/ could also be considered as peripheral, since there is no vowel more front and more rounded than /y/. Referring to articulation, short non-constricted /ɪ/, /ʊ/ and /ʏ/ may count as articulatorily peripheral as well. If, however, peripherality refers to *acoustic extremeness*, the definition would not hold for /ʊ/ or /ʏ/, since they show a high portion of acoustic *overlap* with other categories and a less peripheral position in a formant plot (see Figure 12.7 and Figure 12.8).

In the L2 perception data, peripheral vowels are not only less difficult but also more *preferred* as response options by non-native listeners. The *preference* ranking for the results of all non-native listeners (see Figure 9.11) shows a strong preference for /uː/ (7.3% of all responses) as well as for /iː/ and /ɑː/ (both 7.2%). However, the most frequently selected categories in the "all L2"-sample in the current data are /eː/ (8.6%) and /yː/ (8.5%) (see Figure 9.11). These results are due to the listeners' bias to select /yː/ for other front rounded vowels and /eː/ for /ɛː/-stimuli in 49% of the cases (see section 11.1.3 and 11.1.4). In other words, /eː/ and /yː/ apparently function as a kind of "perceptual magnet" *within* the class of front non-rounded or front rounded vowels, respectively.

To sum up, asymmetries observed in difficulty (id_wrong scores) and preferences (pref_scores) confirm the assumption of the specific status of some vowels as perceptually more stable and/or psychologically more salient. With respect to *difficulty*, it is stated that peripheral /ɑː/ and /iː/ show the highest id_correct scores above 80%, /ɑ/ occupies the third place and /uː/ and /yː/ the sixth and seventh place in an id_correct ranking. In many L1 sub-samples, *a*-, *i*- and *u*-vowels show fairly high id_correct scores (see section 9.1, for a language-specific comparison, see Table 9.1). Peripheral vowels are however not the most *preferred* categories in the L2 listeners' preference ranking (see section 9.2). While for peripherality of /i u a/ phonetic correlates are identifiable, the relative psychological prominence of /eː/ or /yː/ indicated by high id_correct and high pref_scores cannot be ascribed solely to specific acoustic cues.

To conclude, peripheral vowels are considered as both perceptually *stable categories* due to signal-inherent characteristics (stimulus-related) as well as attractive *perceptual anchors* or points of reference in vowel identification. They may act as “perceptual magnets” (response-related) within a given class of vowels. To account for the observed difficulties in identification and the listeners’ preferences in category selection, a differentiation of properties of the signal vs. properties of response categories is necessary. But even if there are particular signal-inherent or response-related properties that enhance the listener’s ability to perceive them correctly as well as the listener’s bias to select a category (hyper)correctly, the response patterns strongly manifest the individual part of the individual listener or language learner.

12.3.3.2 Frequency

A considerable part of language acquisition is determined by the learners’ input, i.e. the specific data-set an individual is exposed to. Each language learner is exposed to a slightly different data-set. Therefore, in acquisition, each individual learner in a sense has to “reinvent” the phonological system of a language, its phonetic categories and phonological contrasts in his/her particular individual way (Blevins 2004: 230). To accomplish this task, detailed phonetic knowledge, including patterns of free and context-specific variation has to be learned in L2 acquisition.

> „*The structure of the categories is such that nondiscrete members may be more central or more marginal, frequency in experience influences the importance of particular features, and features may be more or less important to defining the category, but no feature is redundant and thus dispensable.*“ (Bybee 2001: 33)

In L2, the frequency with which a particular vowel occurs in a given language and more specifically in the learner’s input is assumed to have an influence on the learnability of a specific vowel category in L2. Frequency of occurrence in the input is assumed to affect the rate and success in acquisition and the status and relative strength in mental representation (cf. Pierrehumbert 2000, 2001, 2003; Bybee 2001: 33).

For L2 acquisition, it has been hypothesized that frequency and salience of structures in the source language (L1) may promote transfer to L2, whereas transfer may be inhibited by frequency and salience of structures in the recipient language (L2) that are incompatible with transfer (Andersen 1983; Jarvis & Pavlenko 2008). The relative frequency of occurrence may affect ease and rate of acquisition in the sense that L2 learners will be more acquainted with frequently occurring structures.

From a typological point of view, general distributional patterns in the languages involved together with the presence or absence of a vowel category in the learners' L1 may be of relevance for the acquisition of the target language.

Perceptual *category formation*, i.e. the establishment of long-term representations for vowel categories in L2, is an important pre-condition for correct vowel identification. For less frequently occurring vowels, the process of category formation may be more difficult. Higher frequency in the input increases the learner's evidence available for the particular phonetic characteristics of a category and strengthens its mental representation. Stable phonemic categories in L2 will be established earlier for more common German vowel categories than for less frequently occurring sound structures. In other words, one can assume that "*frequency can only help*" (Jarvis & Pavlenko 2008: 23) to establish appropriate L2 categories. Therefore, more frequently occurring categories are expected to show better results in L2 perception experiments.

Frequency of occurrence, i.e. token frequency and/or number of exemplars for a specific vowel type, is moreover closely related to the role of language-specific *expectations* of the listeners in perception. In language acquisition, in speech processing and in experimental tasks, frequency may have an effect on *probabilistic hypothesizing* as well as on *strength* in representation. Considering the acquisition of a particular target language, the relative frequency of certain vowel qualities in the target language and the frequency of occurrence in the learners' L2 input are important factors to account for difficulties in the acquisition of the German vowel system.

The relative frequency of a category may also affect the listeners' expectations in an *experimental* situation. Frequency in occurrence and strength in representation have an influence on the listeners' *probabilistic weighting* of particular structures and affect listeners' expectations about the identity of vowel stimuli. In other words, the relative frequency of occurrence in the learners' input is expected to have an impact on the learners' *bias* to select a given category as response option in a perception task.

To conclude, the effects of frequency can be viewed as *knowledge*-related, *learning*-related and *performance*-related. More frequently occurring items are expected to be easier and faster acquired, better established as defined category and generally stronger and more prominent in in mental representation. Frequency of occurrence in the learner's input is expected to increase the probability that a given structure will be integrated in the learner's knowledge of L2 and the probability that a given structure will be activated in L2 performance or in an experimental task. To summarize, we can hypothesize an effect of frequency on (1) familiarity, (2) category formation, (3) learning, and (4) salience in perception affecting ease and correctness as well as expectation and probabilistic weigthing. Of course, it is difficult to isolate the frequency effect of L2 structures from the influence of vowel categories in the listeners' L1. Referring to data of the current study, an effect of frequency on difficulties and preferences is hypothesized.

As discussed in section 5.5, German vowel categories differ considerably in frequency of occurrence. Frequency scores for the present analysis were taken from Meier's (1967) 100,000 sounds corpus, a text corpus from poetry and prose texts (see section 5.5, Figure 5.20, Table 5.8 and Table 12.1 summarizing results of Meier 1967: 250ff). In Meier (1967), 38,71% of occurring sounds were vowels. The average score of frequency for all German vowel categories lies around 3.6%. The most common categories are /ə/ (22.4%, not occurring in the present data), and /ɪ/ (10.2% due to frequent words like *in* or *ist*). Least frequently occurring are /œ/ (0.3%, /øː/ (0.6%) and /ʏ/ (0.8%).

%	ɑ	ɑː	ɛː	ɛ	eː	ɪ	iː	ɔ	oː	ʊ	uː	œ	øː	ʏ	yː
frequency	7.4	4.1	0.6	6.4	3.0	10.2	6.3	3.3	2.5	5.5	2.5	0.3	0.6	0.8	1.1
id_wrong	24	16	67	26	45	32	20	26	39	34	29	41	55	42	29

Table 12.1: Token frequency in % (occurrence in the corpus of Meier 1967: 250ff, 38,7% were vowels) and id_wrong scores for 15 German vowel categories (all 173 L2 listeners)

The hypothesis "*token frequency influences stimulus difficulty*" was tested on the present data. The scatter plot in Figure 12.5 demonstrated the relationship of the two continuous variables frequency (M = 3.6%, SD = 2.9%) and id_wrong scores for all 270 stimuli (M = 35.1%, SD = 16%) of the 15 German input categories.

A Pearson product-moment correlation was conducted to evaluate the null hypothesis that there is no relationship between token frequency (Meier 1967) and correctness in identification (173 participants, 270 stimuli). Preliminary analysis shows that the data are subpar with respect to the assumptions of linearity and homoscedasticity (see Figure 12.5) and that the results of a Pearson correlation analysis have to be viewed with caution. The analysis showed that a significant correlation of frequency and id_correct scores can be assumed. A medium, negative association of frequency (M = 3.6%, SD = 2.9%) and percentage wrong identifications (M = 35.1%, SD = 16%) was found, r (269) = -.49, $p < 0.01$. Higher frequency of occurence appears to be associated with lower id_wrong scores. Frequency helps to explain 24% of the variance in the rate of wrong stimuli identifications (N = 270). However, frequency and difficulty seem to be not equally related for all categories. A strong relation is observed for the so-called Umlaut-categories /ɛ:/, /ø:/, /ʏ/, /œ/ and /y:/. Within this group the weakest relationship is given for /y:/ and the strongest for /ɛ:/ (see Figure 12.5).

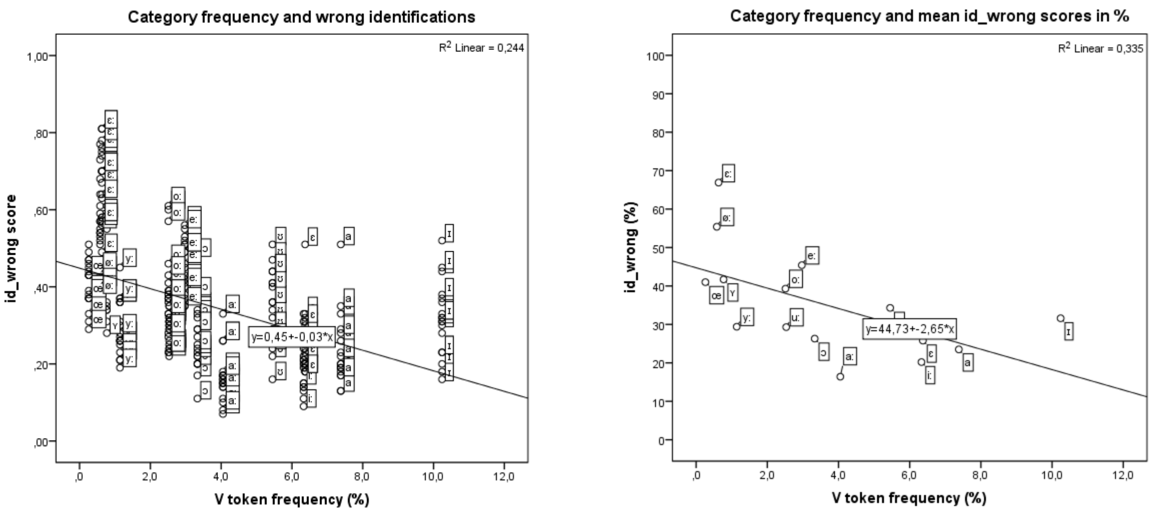

Figure 12.5: Frequency of occurrence (Meier 1967) and wrong identifications by 173 non-native listeners (id_wrong scores for 270 stimuli left, mean id_wrong scores right)

Moreover, the hypothesis "*token frequency influences response bias*" was tested on the data (see Figure 12.6). To evaluate the null hypothesis that there is no relationship between frequency of occurrence and *preference* scores for the input categories (N = 15), a Pearson product-moment correlation was conducted. Preliminary analysis shows that the assumption of homoscedasticity is violated (see the scatterplot in Figure 12.6). No significant correlation of frequency (M = 3.6%, SD = 2.9%) and preference (M = 6.6%, SD = 1.1%) was found in the data for non-native listeners, r (15) =.13, p = .654. Category frequency can explain less than 2% of the variance between categories (R^2 = .016, see Figure 12.6) indicating that other factors have to be considered to account for the listeners' preferences and bias to select a given response category in the identification task.

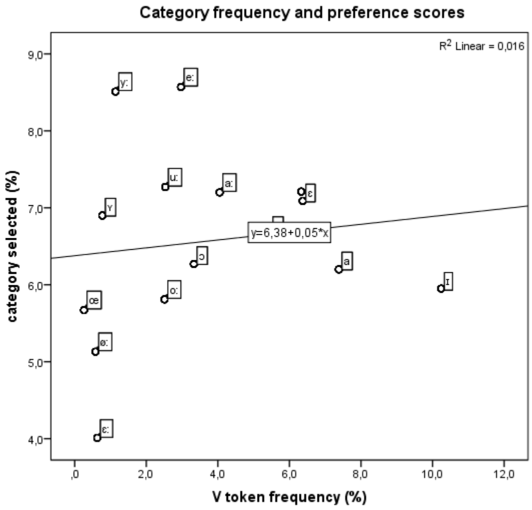

Figure 12.6: Frequency of occurrence and preference scores for all L2 listeners (N=173) (R^2 = 0.016)

A closer look at the relation between frequency and category *preference* (see Figure 12.6) shows that there are strong inter-category differences in preference for some vowel qualities but not for others, especially within the class of the so-called *Umlaut*-vowels: The data suggest a strong connection between frequency and preference for /ɛ:/, which occurs very rarely (0.6%) and is selected in only 4% of the responses. Here, the hypothesis “frequency influences response bias” would be confirmed. For /e:/ on the other side, no correlation of category frequency and category preference is given, since /e:/ attracts a considerable portion of /ɛ:/-stimuli in the listeners’ responses and therefore receives a high preference score of 8.6% in the data for all L2 listeners. Similarly, the low frequency of occurrence of 1.1% of /y:/ does not correlate with the high percentage of cases where /y:/ was selected as response category. As discussed above, L2 listeners in many sub-samples show a strong tendency to identify front rounded /ʏ ø: œ/-stimuli as /y:/ (see section 11.1.4). This preference in perceptual substitutions is particularly observed with /ʏ/- and /ø:/-stimuli and explains the very high preference scores for /y:/.

It follows that statements about the listeners’ preferences to select a given category must not be made on choice options across the full system but can – if at all – only be stated for a specific class of vowel contrasts, e.g. for front rounded vowels. Within a given vowel class, one category may act as attraction point or “perceptual magnet” (Iverson & Kuhl 2000) for other members of the same class. It is hypothesized here that these attraction effects do not work across the full perceptual vowel space but rather act locally between “neighbouring” categories in the vowel space as visualized in Figure 12.7. This view is strongly convergent with the assumption of phonological processes in production and perception that result in minimal substitutions of sounds differing in only one distinctive feature (see section 11.2 and section 12.2.4).

To summarize, a significant but moderate negative correlation of *frequency* of occurrence (token frequency) and their *difficulty* can be assumed in L2 perception of German vowels, particularly for front rounded vowels and /ɛ:/. On the other hand, no significant correlation of frequency and the listeners' *preferences* or overall bias to select a specific category as response option was found. Rather, local preferences within a given sub-set have to be assumed. The construct of peripherality may at least partly account for these local preferences. The crucial question that follows from these observations concerns the nature of these local preferences for a specific vowel category. Are they language-specific and simply happen to parallel in different languages, and – if yes – why do these tendencies occur in several languages? Are there vowel-inherent phonetic properties that lend themselves to explain the observed effects? Or are the reasons rather described in terms of phonology referring to specific phonemic contrasts? Toanswer these questions, additional concepts have to be included in the discussion.

12.3.3.3 Density

In order to account for asymmetric proximities and related phenomena with respect to "neighbouring" categories in a metric space, Krumhansl (1978) proposed the *distance-density model*, in which similarity is assumed to be a function of both interpoint distance and the *density* of stimulus points in a dimensionally organized metric space. She neither rejects traditional geometric assumptions nor set-theoretic feature-based approaches but proposes an additional assumption for similarity relationships between items in a space. Krumhansl refers to experimental similarity data of different types showing that items typically are not equally dispersed within the metric space, but that there are sub-regions in the space that are more or less dense compared to others. The central assumption of the distance-density model is that perceptual sensitivity to inter-item distinctions in *dense regions* of the space will differ from sensitivity in less dense regions. Krumhansl explicitly interprets the *bias* terms psychologically as a reflection of density in a psychological space[148]. A relationship between similarity and spatial density is assumed and operationalized by augmenting the basic similarity-distance function

$$s_{ij} = f(d_{ij})$$

[148] for a discussion of ways to identify density, see Krumhansl (1978: 446f).

by a second distance function $\bar{d}$, which depends on both interpoint distance in the configuration and some measure of spatial density in the regions surrounding *i* and *j*

$$\bar{d}_{ij} = d_{ij} + \alpha\delta_i + \beta\delta_j$$

where $d_{(ij)}$ is the inter-point distance, δ_i and δ_j are measures of spatial density in neighbourhoods associated with items *i* and *j*, and α and β are constants that reflect the relative weight given to densities δ_i and δ_j. Each point in a spatial configuration is associated with a density measure δ_α that measures the density of points within the surrounding region. Asymmetries arise when the conditions $\alpha \neq \beta$ and $\delta_i \neq \delta_j$ are simultaneously satisfied.

The model moreover assumes a monotonic decreasing function $\bar{f}$ of similarity s_{ij} to the *modified distance function* $\bar{d}_{ij}$

$$s_{ij} = \bar{f}(\bar{d}_{ij}).$$

Experienced listeners are expected to adjust their perception to smaller distances in a dense sub-region of the object space by higher sensitivity to discriminations in that sub-region compared to less dense sub-regions. With respect to similarity data, the central assumption of the distance-density model refers to the case that two points in a relatively dense region of the space would have a smaller similarity measure than two points of equal interpoint distance but located in a less dense sub-region due to finer perceptual discriminations within dense sub-regions (Krumhansl 1978: 446).

Krumhansl (1978: 446f) defines three ways to determine the value of the *density* function δ_i in a metric dimensional space (1) by *self-similarity* s_{ii} between an object and itself, where self-similarity is assumed to be related to spatial density in a region surrounding point *i* in a multidimensional space; evidence suggests that it may be possible to estimate the spatial density function from the diagonal entries in the similarity matrix (see Krumhansl 1978: 446f, 450) or (2) by direct computation of a measure of spatial density by the *location* in the configuration, where points near *i* add heavily to the measure of density relative to more distant points, or (3) based on the number of points within a fixed *radius* in the space.

The construct of *density* is especially appealing for the description of L2 perception data, where the number of vowel categories and therefore also the density of regions in the phonetic vowel space of the listeners' L1 may differ from the density of the respective region in L2. However, the crucial question is how to *operationalize* density in the phonetic and/or perceptual vowel space and how to integrate it in a model of L2 vowel perception. To deal with the problem of operationalization we have to consider *stimulus-related* aspects, i.e. physical properties of each input category and distances between stimuli in the acoustic vowel space as well as *response-related* aspects that are conditioned by properties of the respective categories and the listeners' knowledge and hypotheses.

The starting point in the present analysis is a description of the dispersion of vowel types and individual tokens in the *phonetic* space, more specifically in the *acoustic* space, on which assumptions about the psycho-acoustic structure of the *perceptual* vowel space can be based.

Different hypotheses have been formulated about acoustic dimensions that are relevant in vowel perception. However, a commonly accepted categorical definition exists neither for measures of *distance* between vowel categories in the phonetic space nor for *weight* or prominence/salience of signal-inherent characteristics varying along specific acoustic parameters (see section 12.2.4).

Figure 12.7 compares different ways to represent the *acoustic vowel space* for German vowels by (1) an F1xF2-plot in Hertz (left) and (2) an F1xF2-plot in Bark (right). Moreover, an F1-F0xF2-F1-plot in Bark and an F1-FxF3-F2-plot in Bark are provided in Figure 12.8. A comparison of these four scatter plots shows that there are regions in the acoustic vowel space of higher local density than others, but it also illustrates that dispersion, distances and density in a particular region of the space depend very much on the measures and dimensions selected for description. For example, the F1xF2-plot in Hertz (Figure 12.7 left), the most common representation, shows higher density, i.e. smaller inter-category distances, for back vowels and larger distances between front vowels than the Bark-scaled F1xF2-representation (Figure 12.7 right), where high and front vowels seem to be less dispersed compared to back vowels than in the Hertz-scaled representation. This effect is due to the different resolution of higher frequencies in the Bark-scaled representation. The effect of different measures and scales in description has to be considered when arguing about phonetic distances, relative density and acoustic similarity.

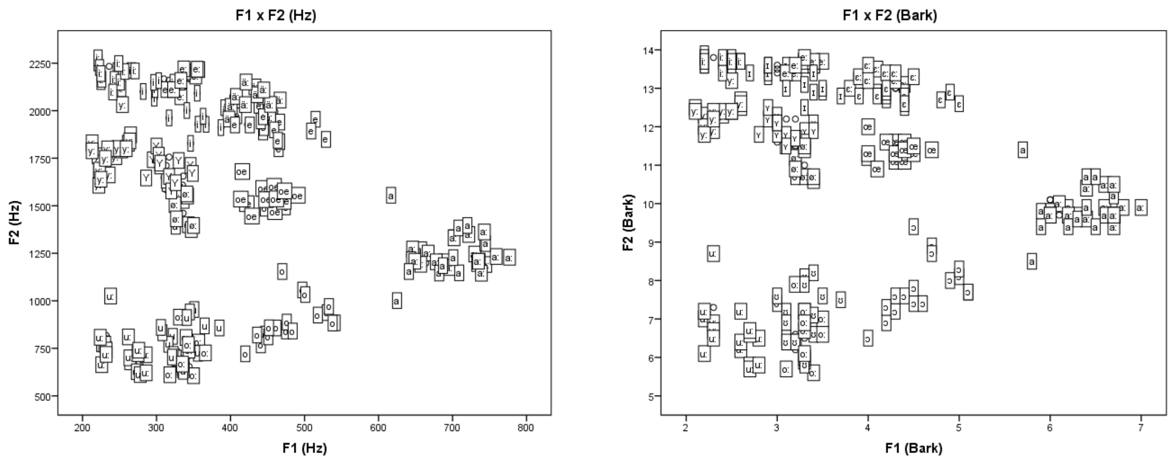

Figure 12.7: Dispersion, distance and density in the acoustic vowel space: Comparison of a Hertz-scaled and a Bark-scaled F1xF2 representation for German vowels (270 input stimuli read by one male speaker)

Syrdal & Gopal (1986) proposed a binary classification of vowels on the basis of Chistovich's and colleagues' concept of a "*spectral center of gravity*"-effect (Chistovich et al. 1979; Chistovich & Lubljinskaja 1979), which refers to the above mentioned effect of concentration of energy by the clustering of two adjacent formants. Syrdal & Gopal's (1986) proposal was embedded in a two-level perceptual model assuming an intermediate auditory processing followed by a higher phonetic stage of processing. Chistovich and colleagues assumed a critical distance between formant frequencies within a range of about 3 to 3.5 Bark between neighbouring formants referring to *F2-F1* distance. The high concentration of energy of *F1* and *F2* in the lower frequency range due to *F2-F1* distance < 3.5 Bark was found to be sufficient to create the quality of a back vowel so that vowels synthesized by only one formant in the low frequency range were perceived as back vowels. Syrdal & Gopal (1986) extended the concept of a *3-Bark threshold* and included formant distance measures *F2-F1*, *F3-F2*, *F4-F3*, and *F4-F2* in Bark as well as the distance of *F1* and fundamental frequency *F0* (visualized in Figure 12.8), since *F1* and *F0* vary systematically across vowels such that *F1* is relatively low for high vowels and *F0* higher for low vowels. A division of the acoustic vowel space in terms of Bark-scaled formant frequency distances is provided in Figure 12.9.

The 3-Bark threshold for *F3-F2* shows a clear distinction of *front vs. back* vowels. The low F2 frequencies for front rounded vowels also decrease the F2-F1 distances (see also Figure 12.8 and Figure 12.9). The 3-Bark level for *F1-F0*-distances in Bark provides a measure for *degree of aperture* or – in traditional terms – for vowel height (Syrdal & Gopal 1986). *F1-F0* values are lower for close vowels and higher for open vowels. The condition *F1-F0* > 3 Bark is referring particularly to *a*-vowels as [+open] but also to some of the /ɔ/-stimuli as well as to single tokens of /ɛ/, /ɛ:/,/œ/, and even /ɪ/ in one case, which are above the critical level of 3 Bark-difference in the *F1-F0* dimension[149].

Table 12.2 summarizes the mean formant frequencies in Bark and Bark-distances *F1-F0*, *F2-F1*, *F3-F2* for the input stimuli (distances smaller than the critical distance of 3 Bark are printed in grey), by which three major category types can be discerned: (1) open *a-vowels* (*F3-F2* < 3 Bark, *F1-F0* > 3 Bark), (2) *front i-*, *e-*, *ö-* and *ü*-vowels (*F3-F2* > 3 Bark, *F1-F0* < 3 Bark), where front rounded vowels show a lower F2 due to lip rounding, and (3) *back* vowels (*F3-F2* > 3 Bark, *F1-F0* < 3 Bark). Categories within a given sub-region do not only differ in formant frequencies but also in duration.

(Bark)	mean F1	mean F2	mean F3	mean F0	F1-F0	F2-F1	F3-F2
ɑ	6.3	9.9	14.5	1.5	4.7	3.7	4.5
ɑ:	6.4	9.9	14.5	1.5	4.9	3.5	4.6
ɔ	4.3	11.4	14	1.7	2.9	3.2	6.8
o:	3.2	11.3	13.6	1.6	1.6	3.2	8.2
ʊ	3.1	12	13.8	1.8	1.4	4.1	7.1
u:	2.3	12.3	13.4	1.8	0.6	4.2	7.8
ɛ:	4.1	13.3	14.8	1.6	2.5	9.2	1.5
ɛ	4.3	12.8	14.5	1.6	2.7	8.5	1.7
e:	3.2	13.6	14.9	1.7	1.6	10.3	1.4
ɪ	3.2	13.3	14.7	1.7	1.5	10.1	1.4
i:	2.4	13.7	15.5	1.8	0.6	11.3	1.8
œ	4.6	7.8	14.6	1.6	2.6	7	2.6
ø:	3.3	6.5	14.7	1.7	1.6	8.1	2.3
ʏ	3.3	7.3	14.4	1.8	1.3	8.9	1.9
y:	2.5	6.7	14.5	1.8	0.5	10	1.1

Table 12.2: Formant frequencies F1, F2, F3 (in Bark) of input stimuli and classification by critical distance (mean Bark-differences F1-F0, F2-F1, F3-F2, distances < 3 Bark in grey)

In order to operationalize *acoustic density*, the German vowel system is divided – on the basis of Syrdal & Gopal's concept of 3-Bark critical distances – into three major category types (see Figure 12.9), which differ with respect to basic acoustic characteristics and density in the acoustic space and may therefore receive different acoustic *density weights* (d_weight). As Figure 12.8 and Figure 12.9 show, the relative density of the acoustic vowel space of German differs between these three category types. Density in terms of number of categories in a specific region is much higher in the front region of the acoustic vowel space than in the back region. This distribution requires higher perceptual sensitivity for front than for back vowels to differentiate German front rounded vs. front unrounded vowels.

[149] Syrdal & Gopal (1986) use end-corrected Bark values (cf. Traunmüller 1981) that are slightly lower than the original Bark distances used here (cf. Zwicker & Terhardt 1980).

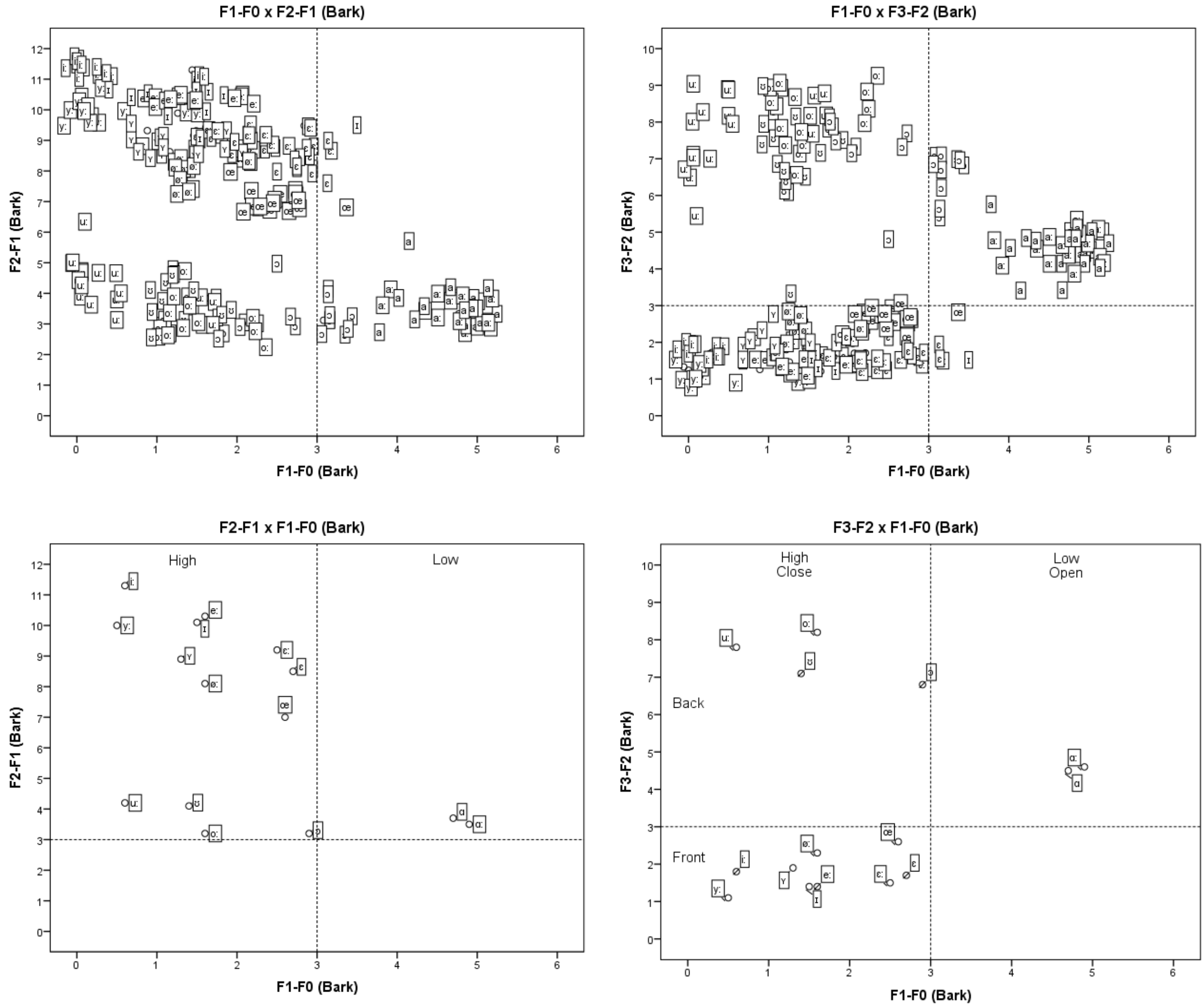

Figure 12.8: Dividing the acoustic vowel space: Mean formant distances in Bark F1-F0 x F2-F1 (left) and F1-F0 x F3-F2 (right) for German vowel categories

Three major category types can be differentiated in terms of the critical distances of F1-F0 and F3-F2 differing in the number of "neighboured" vowels in a sub-region of the German vowel space: For each category a density weight can be assumed that refers to the number of competing categories in the acoustic space:

(1) *Front* vowels: Nine categories, /i: ɪ e: ɛ ɛ: y: ʏ ø: œ/, occur in the front part of the acoustic vowel space with F1-F0 means and F3-F2 means below 3 Bark.

(2) *a-vowels*: Two categories show F1-F0 values above 3 Bark and F3-F2-values between 4-5 Bark.

(3) *Back u- and o-vowels*: Four categories /u: ʊ o: ɔ/ show F1-F0 values below 3 Bark and F3-F2 values that exceed 3 Bark, typically in a range from 6-9 Bark.

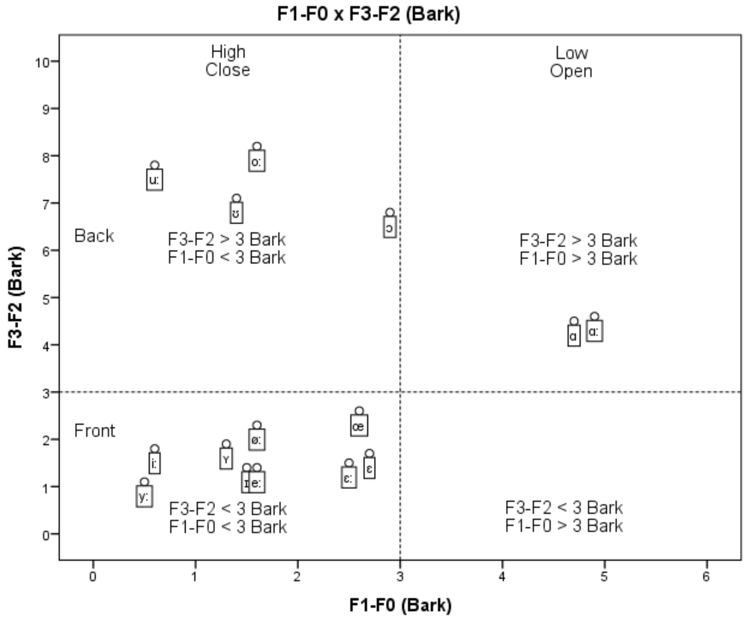

Figure 12.9: Four basic category types of German vowel categories based on mean F1-F0 and F3-F2 Bark-differences above or below 3 Bark (cf. Syrdal & Gopal 1986)

Native German listeners are assumed to show increased sensitivity in dense regions of the vowel space, whereas non-native listeners from languages that have a lower number of categories in a specific sub-region are expected to show less sensitivity to contrasts in that part of the vowel space. Most difficulties are expected in the front region that shows the highest density due to the existence of front rounded vowels and /ɛ:/ in German.

However, it is obvious that - partly due to the effects of peripherality and frequency, and partly due to the presence of absence of equivalent vowels in the listeners' L1 – higher density does not automatically imply higher difficulty. Rather, within a given sub-region there are more of less difficult categories there are some categories that are more difficult and some that are more preferred than others. Thus, the construct of density as such is of relative weak explanatory power. Rather, as proposed by Kurmhansl (1978), it functions as weighting factor that contributes to the relative difficulty or preferability of stimuli of a specific vowel category in a particular sub-region or vowel class. The relative density in a given sub-region interferes with the number and quality of categories in the listener's L1 and influences perceptual sensitivity and perceived similarities in L2, which may differ considerably between learners from different language background. Therefore, substantial differences in id_correct scores and id_V scores are found for categories within a specific sub-region that vary moreover by the listeners' L1 background.

To summarize, measures of distance in the acoustic space depend on dimensions and scales (Bark vs. Hertz) used for description. *Distances* in the acoustic space are not necessarily equivalent to distances in the perceptual space. They moreover vary language-specifically between categories and sub-regions of the vowel space. *Density* may be defined as the number of competing categories in a specific region of the vowel space. *Difficulty*, however, does not directly correlate with density, as e.g. the example of /iː/ will show (see 12.3.3.4). Difficulty may vary considerably between categories of a specific sub-region in the *perceptual* vowel space. To account for inter-category differences in difficulty, language-specific, stimuli-specific as well as more general cognitive factors have to be considered.

12.3.3.4 Self-similarity

Krumhansl's distance-density model accounts for asymmetries by assuming that the weights α and β may differ if the spatial density surrounding one point affects the distance measure $\bar{d}$ more than the density surrounding a second point. Dealing with stimulus-response data, it is stated that stimuli that are frequently associated with the correct response will infrequently be associated with incorrect responses. Or, as Shepard (1957) puts it in terms of similarity, high *self-similarity* of an object seems to be opposed to similarity to other objects.

Referring to *self-similarity*, Krumhansl (1978) hypothesizes that objects with large measures of self-similarity are located in relatively less dense regions in the geometric representation of the perceptual space. In the present data, this is the case for *a*-vowels. *a*-vowels show high id_wrong scores and are located in a region with low density in the MDS configurations for nearly all language sub-groups (see Figure 9.14). *u*-vowels, on the other hand, are found in a "dense" region of the perceptual space in close neighbourhood to *o*-, *ü*-, and *ö*-vowels in the two-dimensional MDS configurations for all language sub-groups except Hungarian and German (see Figure 9.14). The higher density of vowels occurring in the region of u-vowels is due to considerable confusion within the class of back vowels as well as to the relatively high amount of confusion of front rounded and back rounded vowels (see Table 9.5).

On the other hand, due to other constraints in the similarity data, stimuli with high self-similarity may not necessarily be located in less dense regions of the perceptual space (Krumhansl 1978: 450). In the present data, this is the case for /iː/-stimuli in most of the MDS representations for the language-specific perceptual vowel space (see Figure 9.14). In all language sub-groups, /iː/ is located relatively close to other vowel categories (mostly *e*-qualities and /ɪ/) in the language-specific perceptual maps, despite the fact that /iː/ is identified correctly in 80% of the cases and occupies an 'extreme' or peripheral area of the acoustic space. The reasons for a higher percentage of correct responses for /iː/ are found in several facts such as its peripheral character or the existence of an equivalent in L1, the frequency of occurrence in L1 and L2 and familiarity with L2 categories (see above). The small distances to other categories in the two-dimensional MDS representations despite the high percentage of correct responses for /i:/, is connected with the fact that /i:/ functions as substitution target for other vowels, especially for /ɪ/ and /e:/.

To conclude, there are different types of evidence for distance and density in the vowel space: Density can be determined by the number of correct or wrong identifications (self-similarity), the location in the MDS configurations or by the number of categories within a given sub-region of the vowel space. The *perceived distance* of two items in the perceptual vowel space is composed by a combination of the *density* function and of *differences in responses* to two items. Differences in responses to two items, i.e. differences in id_scores, can be taken as a measure of *perceived* inter-category distance, taking the form of a linear function of the distance between items on a psychological continuum and the density of stimuli lying between these two in the continuum (Krumhansl 1978: 447). The one-dimensional MDS representations (e.g. Figure 9.13) provide a good illustration of this conception of distance and density between categories in non-native vowel perception.

However, as the id_wrong scores for non-native listeners show, the percentage of wrong or correct identifications of categories within a given sub-region of the space may differ significantly. Especially in the set of German front vowels substantial differences between categories in id_correct scores are observed, which are not directly accountable by differences in density.

To summarize, within the sub-regions of the acoustic and/or perceptual vowel space there are some vowel qualities that are more prominent and more distinctive than others. These categories seem to be less difficult to L2 learners from different language background. Factors that have an impact on the listeners' rate of correct identification at different levels of language knowledge and speech processing are numerous, such as the relative strength of acoustic cues for a given vowel contrast, the "magnet effect" of more salient, peripheral and/or more frequent categories, the density of competing categories in a given sub-region of the vowel space, the existence of a cognate in the native language, the frequency of occurrence in the input together with differences in the functional load of contrasts in L1 and L2 as well as the graphemic representation(s) for a category. Moreover, instantaneous effects such as recency in the experimental task and the listeners' (temporary) hypotheses may affect the listeners' rate of correct identifications.

12.4 Differential bias of stimuli and responses

The exemplification of influencing factors presented in the previous sections has shown that reasons for asymmetries and bias in perception data may be of very different origin. Theoretical constructs such as density, frequency, salience or prototypicality, robustness and many others have been proposed to account for the listeners' difficulties and preferences in perceptual confusion data. The aim of the following sections is to discuss the status of similarity and *bias*, its impact on similarity relationships and its role in accounting for asymmetries of difficulties and preferences in L2.

Nosofski (1991: 97ff) emphasizes the importance of bias in similarity models and suggests "*differential bias*" as terminological alternative to "asymmetric similarity". He moreover argues for a differentiation of the notions of bias and similarity: While the fundamental construct of *similarity* is conceived as a relation between two objects, which is formally expressed as a function of two variables, e.g. s_{ij}, *bias* is commonly defined as a function of a single object i, expressed as b_i (see Figure 12.10). This crucial differentiation will be implemented in the discussion of vowel identification to gain more detailed insights into the mechanisms of asymmetry and the complex interaction of factors affecting the listeners' behaviour in an L2 perception task.

12.4.1 Stimulus-related and response related aspects of similarity and bias

Most perceptual classification paradigms refer to the notions of *stimulus similarity* and *response bias*. Stimuli are considered as confusable and responses as highly distinctive. While similarity is viewed as a relation between stimuli, bias is considered as a property of responses. However, this view may be reconsidered.

Nosofsky (1991) reviews several examples that – as he claims – are more easily interpretable in terms of *stimulus bias* than in terms of response bias. He therefore proposes a reconceptualization of the dichotomy suggesting a description of similarity and bias as orthogonal to stimuli and responses in order to differentiate stimulus similarity vs. response similarity and stimulus bias vs. response bias (see Figure 12.10): Just as stimuli may differ in degree of similarity, so do responses. This conceptualization is essential for the interpretation of experimental data on relationships of similarity visualized in Figure 12.10.

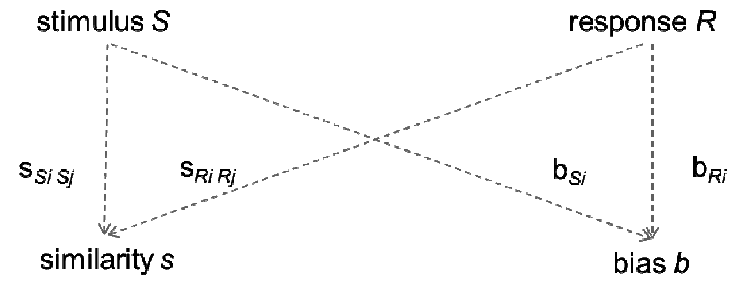

Figure 12.10: Orthogonal relationship of stimulus – response and similarity – bias

Following Nosofsky's argument, one can refer to the *phonetic similiarity* s_{SiSj} as relative articulatory or acoustic similarity between stimuli S_i and S_j, which is understood here as the "objective" phonetic similarity of vowel tokens in terms of articulatory or acoustic measures. *Similarity of response categories* s_{RiRj}, on the other hand, is established by the individual listener: Relationships of perceived similarity between response categories refer primarily to vowel types and the listeners' mental representation of L2 categories and are visualized in MDS configurations of the learners' perceptual space (see Figure 9.14).

Likewise, Nosofsky (1991: 99) differentiates *stimulus bias* and *response bias. Stimulus bias* can be observed when prior biases to perceive or remember certain stimuli exist. Garner (1978) suggested that individual stimuli can vary in goodness. "Good" stimuli are considered to be processed more efficiently than poor ones due to more salient phonetic cues. A particular stimulus may be highly salient in perception or memory in one or more dimensions and can therefore easily be encoded or decoded, e.g. palatal /i/ has a maximal distance of F1 and F2 and a characteristic concentration of energy in higher frequencies due to a small F3-F2 distance. Since these properties pertain to individual stimuli and not to responses, and since properties of the stimulus itself rather than relations between stimuli are concerned, they may be better characterized as *stimulus bias* than as stimulus similarity.

On the other hand, the *bias* to select certain *responses* may be independent of the presented stimuli. Independently of characteristics of the stimuli, listeners may have prior biases to choose specific responses. These biases are due to characteristics of the response categories, prior knowledge or the listeners' hypotheses about probabilities of occurrence conditioned by higher frequency and distribution in the input, previous occurrence in a specific context, orthographic familiarity, and so forth: "*it may be that apparent violations of distance symmetry can always be traced to some factor, like familiarity, which pertains to individual* *(rather than to pairs of) stimuli*" (Shepard 1957: 336). These biases are referred to as *response biases*[150].

Another important distinction refers to the differentiation of set-dependent and set-independent biases (see Nosofsky 1991: 134). The strength with which an item is stored in memory due to its *frequency* of presentation is an example of a *set-independent bias,* since the relative representational prominence of *i* and *j* due to frequency is interpreted as independent of other items in the stimulus set. Stimulus *i* may be stronger than stimulus *j* because *i* occurred more often than *j* in the listener's input (cf. 12.3.3.2).

The *density* associated with stimuli *i* and *j* depends on the perceived relative similarity to other members of the stimulus set. Both, density (see Krumhansl 1978 and section 12.3.3.3) and frequency (see section 12.3.3.2) can be considered as properties of individual stimuli and are therefore characterized as *stimulus bias*. The density bias itself emerges from relations between a given stimulus and other items in the set and is therefore set-dependent. Biases resulting from differential densities in the vowel space can be described as *set-dependent biases*. In an experimental context, the set of stimuli and the set of responses is considered to have a high impact on the results.

To summarize, the interpretation of (asymmetric) proximities in L2 confusion data in terms of differential bias has to differentiate between *response-related* and *stimulus-related* biases. A further differentiation refers to biases in the listeners' decisions that may be due to *set-related* factors, e.g. the set of vowel categories in L1 or L2 or the set of response categories offered in the experimental task, but also to *set-independent biases of stimuli or responses*.

To conclude, a comprehensive model of L2 vowel identification has to incorporate both, *intra*-lingual set-dependent similarities between items as well as properties of individual stimuli. A differentiation of stimulus bias vs. response bias and of similarity of stimuli vs. similarity of responses is a valuable distinction to account for asymmetries in difficulty of stimuli as well as in preferences for specific response categories. In the data of the present study, *stimulus bias* refers to qualitative differences between input stimuli and input categories and is reflected by id_correct scores and id_V scores. Id_V scores indicate the percentage stimuli from a given vowel category are identified as belonging to category *i* or *j*. *Response bias* refers to the listeners' preferences indicated by pref_scores, i.e. the percentage a given response option is selected.

---

150 Shepard (1957) refers to response bias as "stimulus weights".

Similarity of stimuli refers to variance of articulatory and acoustic properties of stimuli or to proximity in the multidimensional *phonetic space*, whereas *similarity of responses* describes the proximity of categories in the listeners' perceptual space, which is represented here by perceptual vowel maps in the MDS representations. In speech processing, biases arise on the input/substance side or at the decision/response side, which is strongly related determined by the listeners' mental representation and established relationships of similarity.

12.4.2 Bias in the signal or bias in the listener?

Many studies on L2 acquisition consider the learners' *proficiency* in L2 as major factor to account for their success in the acquisition of a second language sound system. Several *learner-related* variables have been proposed to explain the learners' proficiency and overall success in experimental perception tasks (for discussion, see chapter 8). Another type of more phonetically oriented studies on L2 perception focuses on *input-related* phonetic properties of the speech signal to explain experimental results in vowel discrimination or identification tasks. However, the phonetic analysis of the input signal allows only limited conclusions on psycholinguistic processes of speech processing, categorization and language acquisition. There is one factor that is commonly neglected in most phonetically-oriented experimental studies on L2 perception: It is the specific part of the *listener* in a given experimental setting together with the complex interactions of signal, task, setting and the listener at a given moment in time. In experimental settings, the listeners' behaviour and their choice for a specific response option is never completely predictable or explainable by the input signal. Perceptual identification tasks always contain a *moment of active decision* by the listener, which is affected by his/her *bias* for specific response options, influenced by the listener's language knowledge in L1 and L2 (or L3, Ln), but also by prior stimuli and response decisions in the experiment. The crucial question is whether and to what extent this bias is temporary or long-term or even constant in nature and whether it is rather related to the signal or to the listener.

Bias in a perception task, it is assumed here, can be of different origin such as (1) inherent properties of the input signal, (2) the specific experimental design, (3) the category system of the target language and the listeners' implicit and/or explicit knowledge or hypotheses about the category system of the target language, including temporary or long-term assumptions about relationships of contrast and similarity within the L2 system, (4) the stimuli set and reactive decisions influenced by previous stimuli and the listeners' previous response behaviour, i.e. the listeners' "learning" within the test situation, a factor that is difficult to control in an experimental setting, and (5) the set of response options and the listeners' temporary or long-term hypotheses about the response set.

To summarize, the listeners' response behaviour and their preferences for specific response categories are due to a close interaction of signal-inherent and external factors affecting the relative difficulty or similarity of given contrasts and the listeners' bias to select a specific response option. In other words, we have to differentiate between *bias in the signal* and *bias in the listener*.

12.4.3 Analysis of stimuli and responses

To conclude from what has been discussed so far, the theoretical constructs of similarity, distance and probability of identification or confusion are central notions for a model of L2 vowel perception. To address the issue of non-linear relationships between physical similarity of input signals and perceived similarity of categories in L2, the present study combines two analytical approaches:

(1) A *stimuli analysis* provides a detailed description of characteristics of the input stimuli in terms of articulatory characteristics, acoustic properties and phonological features for the German vowel system and a discussion of other influencing factors such as orthography or overall token frequency (see chapter 5). The stimuli analysis allows for statements about *objective similarity* by comparisons in terms of measurable *distances*, such as differences in a specific acoustic value, e.g. duration or formant frequency differences in Hz or Bark. Moreover, an analysis in terms of articulatory-acoustic features is provided (see Table 5.9). A predictive hierarchical cluster analysis (see section 5.7 and 5.8) illustrates different possibilities of phonetic similarity relationships between categories to simulate different forms of weighting acoustic parameters for predicting possible areas of perceptual confusion.

(2) A *response analysis* provides a detailed description of L1 and L2 listeners' performance in the identification experiment and an analysis of patterns of vowel *confusion* by native and non-native listeners from different language background under two major aspects: (a) *a language-specific* analysis (see chapter 10) and (b) a *category-specific* analysis for the full L2 sample and each of the language sub-samples (see chapter 11) revealing perceived inter-category similarities and major tendencies across language sub-samples.

The response analysis considers (I) *difficulties*, i.e. the percentage of wrong identifications and category confusion, (II) overall *preferences* for specific vowel qualities, (III) specific areas of *inter-category-confusion* due to perceived similarity, (IV) possible *asymmetric patterns* in identification, (V) *similarity scores* and *distances* derived from perceptual confusion data, and (VI) a MDS visualization as a *perceptual vowel map* for L2 learners from a given language background, transferring nominal data (identification matrix) into higher-level ordinal and metric data (see section 7.2 and 7.3). The data analysis was performed according to this scheme for the full L2 sample (chapter 9) and for each of the ten language sub-groups and the German native control group (chapter 10).

The *stimuli analysis* and the description of measurable differences and similarities in phonetic substance is the basis for an analysis of perceived similarities between items in the category set. The analysis of the input served thus as a basis for set-dependent and set-independent biases in L2 vowel identification and categorization.

The *response analysis* is expected to reveal the listeners' perception of *subjective similarity* and psychological distances between L2 vowel categories. In the interpretation of behavioural data, i.e. the response analysis, no direct monotonic correlations between specific physical properties of vowel sounds (the outcome of the stimuli analysis) and the subjects' performance in L2 perception are expected. Rather, the response analysis seeks for *perceived* similarity relationships between sound categories as established by L2 learners.

A comprehensive model of L2 vowel perception requires the differentiation of (1) phonetic distances vs. perceived distances and similarities, and (2) a consideration of differential bias in perception and its provenance as signal-related vs. listener-related, as stimulus-related vs. response-related and as set-dependent vs. set-independent biases. The major interest of this more holistic approach to L2 perception is to gain insights into the *psychological* representation of vowel categories and the L2 listeners' *perceptual vowel space* for a better understanding of difficulties, confusion probability and preferences in L2 perception.

12.4.4 $s_{ij} = p_{ij} * b_i * b_j$ – The formula for perceived similarity, phonetic proximity and biases

In a vowel identification experiment, a set of stimuli and a set of responses is given. These sets are associated by the listener with his/her mental representations of the vowel system. The listener's decision for a specific response option is driven by these three types of sets and by the biases for particular members of these sets. Biases are assumed for stimuli b_S as well as for responses b_R. In the following, stimulus bias will be referred to as b_i and response bias as b_j.

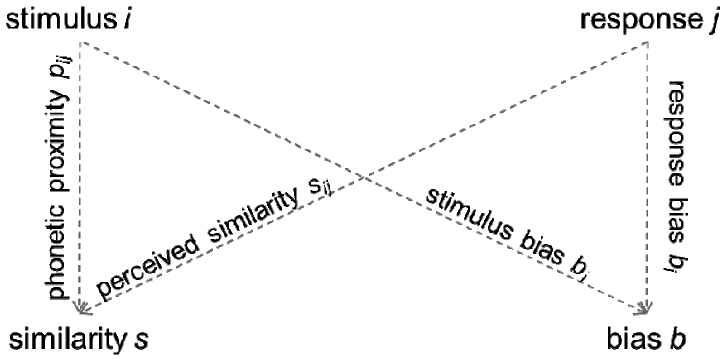

Figure 12.11: Stimulus- and reponse-related similarity and bias

Figure 12.11 represents the relationships of similarity and bias between stimuli and responses. *i* refers to the input signal, *j* refers to the response category in the listener's output.

Similarity is a relationship between two items and is formally expressed as a function of two items *i* and *j*, whereas bias is defined as a function of a single item and expressed as b_i or b_j (Nosofsky 1991: 97ff). Similarity between stimuli, i.e. between actually occurring signals, refers to articulatory and/or acoustic properties, i.e. *phonetic proximity* p_{ij}. Similarity between responses refers to *perceived similarity* s_{ij} between mental representations for categories.

Stimulus bias b_i refers to differences in goodness of the signal, i.e. in the relative strength of specific phonetic cues that have an impact on the listener's identification of a vowel category. In this sense, stimulus bias is most directly associated with the listeners' *difficulties* in vowel identification. Relevant phonetic cues may be formant peaks F1, F2, F3 or formant distances F3-F2 and F1-F0, or the overall spectral shape together with sonority, which is most closely related to degree of aperture (F1-F0), intrinsic duration, intensity and differential formant amplitudes that are influenced by degree of aperture and lip rounding. Phonetic cues may vary in intensity as well as in diagnosticity (Tversky 1978, see section 12.2.2.2).

Response bias b_j is closely related to the *listeners' knowledge of language(s)* and their mental representations of categories guiding their expectations in a given context. Response categories differ in prominence and stability of representation and are therefore expected to differ also in their bias or probability to be selected as response categories in an identification task. In the identification task of the present study, response biases are associated most closely with the listeners' *preferences* for specific response options.

It is however necessary to differentiate between *reasons vs. effects* of perceived similarity and response bias. Reasons for differences in *stimulus bias* may be viewed as differences in the "goodness" of input signals. Acoustic signals differ in the salience of specific cues relevant for identification. Stronger or more intense cues are expected to be better perceivable. Mental representations for categories that are signaled by stronger or rather by less ambiguous cues are assumed to be more stable in representation. However, there is no generally accepted definition of what may constitute a "strong" phonetic cue (see section 12.5). Moreover, phonological contrasts may be realized by a combination of two or more acoustic cues providing redundancy in the signal, e.g. [long] vs. [short] vowel pairs in German are cued in terms of quantitative and qualitative cues (see section 5.4.5) that guarantee perceptual robustness of the contrast. However, the combination of co-occurring cues for a specific contrast as well as their relative strength may vary language-specifically. Co-occurring cues may vary in prominence according to a given speaker and or a particular language. Language learners have to "accommodate" to the appropriate use of cues in a given language system. Therefore, the relative *strength* of specific cues is argued to be rather an *attentional* matter than a merely acoustic one (for selective attention and attentional learning, see section 12.5)

The constructs such as peripherality, frequency, density or self-similarity discussed above (see section 12.3.3) functioning as general weighting factors are assumed to account for a higher bias of more peripheral and/or more frequent stimuli or responses that show a higher density factor and a higher degree of self-similarity. *Frequency* of occurrence is one of the most unanimously discussed impact factors that affect the perceptual stability of specific categories in a given language (see section 12.3.3.2) and may contribute to stimulus bias as well as response bias, even if it was found to have no significant statistical effect on preferences in the present data. Frequency refers to the use of a category or a phonological contrast with respect to the language-specific lexicon and/or grammar (type frequency) or to language performance (token frequency). Higher frequency in the input implies higher probability of occurrence and thus also a higher degree of expectation. Higher frequency of occurrence also implies stronger and more stable mental representations of a specific sound category and as a result will also affect the listener's *expectations* that are an effect of statistical learning (Pierrehumbert 2003; Hume 2005, 2008; Rebuschat & Williams 2012). In this sense, response bias (effect) is closely connected with stimulus bias (reason) in the sense of goodness of the signal. However, response biases rather refer to differences in presumed *probability* than to differences in quality.

In the current data, the listeners' preference for /e:/ as response category for /ɛ:/-stimuli provides a good example for response bias: /e:/ is selected as response category for 49% of the /ɛ:/-stimuli (vs. /e:/ > /ɛ:/ for only 15% of the /e:/-stimuli), despite the fact that in an F1xF2 representation /e:/- and /ɛ:/-realizations do not overlap. In other words, /e:/- and /ɛ:/ are not "similar" in terms of formant frequencies, but they share the feature "phonemic length", which may be the reason for the considerable though asymmetric pattern of confusion. The reasons for the observed asymmetry in substitution patterns can be found in the marginal status and the low frequency of occurrence of /ɛ:/ in German and its “marked” orthographic status, which are clearly response-related.

General cognitive biases in mental representations as well as acquired long-term biases and temporary biases are assumed to have an impact on the listener’s behaviour in a perception task. In an experimental context, the listener’s expectations and current hypotheses about the probability of occurrence and the actual decision for a particular category may be affected not only by general but also by temporary factors. *Temporary* response biases are influenced by specific constellations in a given experimental setting such as stimuli set, response set, task type, recency in the input or the previous occurrence of other more or less similar items in the experiment, but also to the listener’s current strategies and hypotheses during task performance (cf. Strange 1992). *Long-term* response biases rather refer to the listeners’ preferences for a specific response option due to language experiences and proficiency in L2 together with the individual theories, assumptions and strategies that are a result of the learners’ experiences in one or more language. Both, long-term as well as temporary biases may lead the listeners to behave hypercorrectly in a given experimental situation.

To summarize, a model of difficulty and preferences in L2 vowel perception has to include the constructs of both similarity *and* bias. *Similarity* either refers to *physical* similarity p_{ij}, i.e. phonetic characteristics of the input signal or to *psychological* similarity s_{ij} of items as perceived by the language learner. Perceived similarity can be represented as spatial proximity in geometric MDS models. Moreover, stimulus bias and response bias have to be differentiated. While *stimulus bias* mainly refers to qualitative differences in goodness, *response bias* is both qualitative and quantitative in nature. In L2 perception, response bias is directly affected by differences in the category set of L1 and L2 and differences in frequency and probability of occurrence of a particular category in a given context. Languages may differ not only in number and quality of vowel categories but also in terms of frequency and distribution of categories.

The aim of the present study is to offer means for operationalizing similarity of vowels as perceived by learners in L2 vowel identification.

To estimate the *perceived similarity* s_{ij} of input *i* and response *j* the expression

$$s_{ij} = p_{ij} * b_i * b_j$$

is proposed, where p_{ij} refers to *phonetic proximity* of the input signal and the mental representation of the selected response category, b_i refers to *stimulus bias* and b_j to *response* *bias* that are conceived as weighting factors. The differentiation of stimulus bias b_i and response bias b_j allows for a better estimation of *perceived similarity* of vowel categories in the target language. Relationships of similarity can be established between stimuli (*phonetic proximity*) as well as between responses.

When performing a forced choice identification task, listeners are forced to establish relationships of similarity between stimuli (input) and response options in order to decide for a specific response (output). The listeners' choice of a particular response option, however, is influenced by a large set of influencing signal-related, language-related and learner-related factors at different levels of description.

In the present data analysis, *perceived similarity* s_{ij} is indicated by sim_scores derived from the listeners' confusion patterns and the position of categories in the MDS solutions, p_{ij} refers to measurable differences in phonetic properties, b_i to the relative *difficulty* (id_wrong scores) of the input category for listeners from a particular language background due to the presence or absence of specific cues for identification, and b_j to *preferences* in category selection given by pref_scores indicating the probability that listeners from a particular language background at a specific level of proficiency would select category *j*.

To conclude from what has been discussed in the previous sections, modelling the complex interplay of mental representations for vowel categories and the listeners' performance in an L2 identification task requires the consideration of several closely interacting dimensions:

(1) phonetic characteristics of the *input signal* and the presence and relative salience of specific acoustic cues,

(2) characteristics of the language-specific *set of vowel categories*, their statistical distribution and their mental representation in L1 and L2,

(3) the listener's *mental representations* of vowel categories and the established relationships of *contrast and similarity* in the interlanguage system influenced by quantity and quality of received input and state of L2 acquisition,

(4) the *set of response options* offered in a specific experimental design, and

(5) the listener's *expectations* and individual theories, hypotheses and strategies in a given experimental situation.

To summarize, differential bias is a crucial component for a model of L2 vowel perception. *Stimulus bias* refers to the relative frequency or probability of a given stimulus from category *i* to be identified as response *j*, whereas *response bias* refers to the relative frequency or probability a response *j* would be selected for a stimulus from category *i*. The reasons for these biases are partly contained in the signal and partly found in the listener, even though, due to the high adaptivity of the human auditory systems no clear-cut separation of signal-related, listener-related and system-related biases may be possible. Therefore, the theoretical discussion of similarity and bias in L2 acquisition requires another concept, the mechanism of *selective attention* as an element of a theory of speech processing and language learning.

12.5 Bias in language learning – Selective attention and attentional learning

Languages differ in number and phonetic realization of vowel categories. Vowel categories differ language-specifically with respect to acoustic cues for phonological contrasts and the relevance of specific phonetic dimensions. Listeners from different language backgrounds differ in their attention to particular acoustic *cues* that indicate the identity of an input signal as belonging to a specific category. In other words, not only the measurable "intensity" or "strength" of acoustic cues but also the attentional *weight* associated with them can differ from listener to listener and from language to language. The fact that the same acoustic input can be interpreted differently by listeners according to their language experience and individual or language-specific attentional weights is due to the mechanism of *selective attention*. Selective attention can be understood as referring to differential amplification or attenuation of a *subset* of components of a given stimulus (Kruschke 2011).

In terms of Nosofsky's (1986) *attention-optimization hypothesis*, subjects distribute their attention among the component dimensions of the perceptual space so as to optimize performance in a given categorization paradigm, aiming at a maximization of the average percentage of correct categorizations. The distribution of *attentional weights* optimizing performance is assumed to depend on the particular category structures of interest. Applied to the study of L2 vowel perception, the distribution of attentional weights enables listeners to allocate vowel stimuli to phoneme categories of a given language according to the density distributions he/she has experienced within a given category system. This explains why the same acoustic signals can be interpreted differently by listeners from different linguistic backgrounds.

Vowel systems differ in the set of categories and the perceptual weighting of acoustic parameters. Figure 12.12 shows a very simple two-dimensional space with vowel categories that differ in corresponding values in two dimensions, here frontness/backness and/or lip position (dimension 1) and degree of aperture (dimension 2). In the two-element space on the left (A), the correct category label can be inferred only by using dimension 1, whereas in the four-element space in (B) (right side) attention to differences in both dimensions is required to differentiate stimuli appropriately. While in (A) variation in dimension 2 is irrelevant and attention is primarily allocated to dimension 1, in (B) both dimensions are relevant.

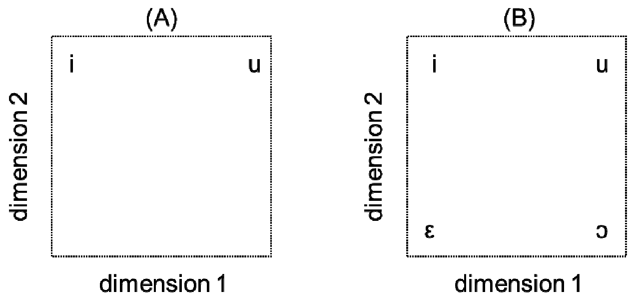

Figure 12.12: Selective attention to relevant and irrelevant dimensions in vowel categorization

In "phonetic reality", however, phonological contrasts are encoded not only in one dimenison buta are differentiated by several co-occurring acoustic cues. To differentiate /i/ and /u/, for example, we find competing and partly interfering cues indicating frontness/backness and/or lip rounding/protrusion (see section 5.4 for acoustic correlates of phonological contrasts). Selective attention enhances specific cues, which are relevant for category identification, and suppresses irrelevant ones. Individual listeners and language systems may therefore allocate differential *attentional weight* to these cues according to the set of categories and relevant dimensions.

The concept of selective attention is strongly related to the notion of "perceptual salience". Though there is no commonly accepted definition of the term *salience* in phonetic terms, salience could be defined as "*the relatively long-lived ability of a cue to attract attention*" (Kruschke 2011: 122). Co-occurring acoustic cues may however compete in salience: "*The salience of a cue is always relative to the saliences of other cues that are present at the same time*" (Kruschke 2011: 122).

However, due to within-category variation and acoustic overlap (see Figure 12.13), signals may be ambiguous. In a perceptual identification task, vowel identification is driven by bottom-up processing, whereas every-day speech perception usually involves top-down processing and the use of higher-level information from lexicon, grammar, non-linguistic contextual knowledge and the listeners' expectations to disambiguate the input (cf. section 12.1.4).

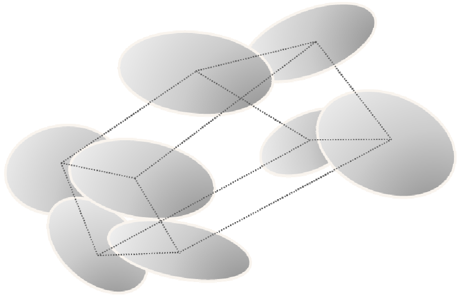

Figure 12.13: Schematic illustration of within-category variation, acoustic overlap and inter-category distances

The assumption of selective attention and differential attentional weight is highly compatible with exemplar-based theories of perception that assume that language users represent categories by storing individual category exemplars in memory (see section 2.4.4) as well as with studies describing the language-specific development of sound categories in L1 acquisition (see section 2.2).

According to the attention-optimization hypothesis the relative distances between items in the perceptual space are determined by attentional weights associated with the component dimensions. For modelling selective attention in categorization performance, a weighting factor w_j is introduced as additional parameter of geometric models (cf. Nosofski 1986). Deviations of the *attentional weight parameter* away from $w = ½$ (50:50) are considered as evidence for differential selective attention involving changes in inter-stimulus similarity relationships.

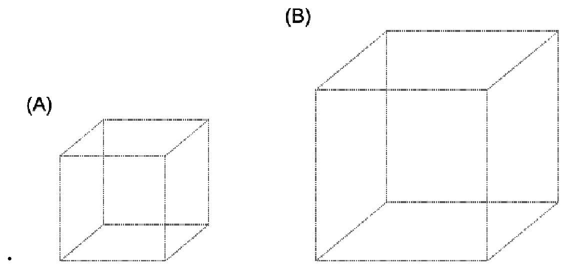

Figure 12.14: Increasing discriminability in the psychological space by increase of experience and frequency of occurrence

Moreover, referring to overall discriminability in the psychological space, Nosofsky (1986) considers a *scale parameter c*. This scale parameter is expected to increase with stimulus exposure duration or the subjects' increasing experience with the stimuli. In L2 perception, the scale parameter is assumed to increase with increased experience in the target language or with increased number of repetitions of a given input, i.e. by frequency of occurrence in the input. Increase of experience or frequency of occurrence of a given input is expected to increase the overall discriminability within the perceptual vowel space of L2. Overall sensitivity to contrasts within the L2 system – visualized schematically in Figure 12.14 by the size of the cuboids and the length of lines linking the corners – will increase as L2 learners gain experience with L2 by overall length of exposure or by stimuli repetition, i.e. frequency of occurrence (for discussion of the impact of language experience, see section 8.7). On the other hand, sensitivity for a given dimension may increase as an effect of increased experience in language learning and awareness of phonetic details.

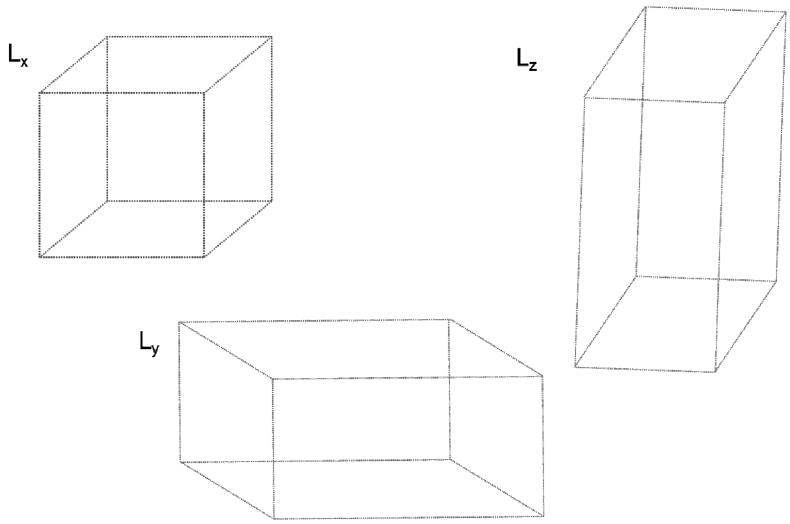

Figure 12.15: Schematic illustration of the attention-optimization hypothesis (Nosofsky 1986)

The influence of these weights and their stretching or shrinking effect along the coordinate axes of the psychological space is visualized in Figure 12.15, where L_x differs from L_y in the specific weight of one dimension that is stretched along this specific axis. The perceptual space is visualized here schematically as a three-dimensional cuboid where corners of the cuboid represent the categories and component dimensions *I, J, K* correspond to the lines between these corners. Listeners from language L_x and language L_y would differ in their selective attention to the component dimensions of the perceptual space. The effect of language-specific selective attention and differential attentional weight is visualized in Figure 12.15: Listeners of language L_y differ from listeners from language L_x in that they allocate higher attentional weight to one dimension, for instance dimension *K*, than to dimension *I* and *J*. In other words, listeners of language L_y are more sensitive to variation in dimension *K* than listeners of language L_x, whereas in L_z strong attentional weight is allocated to dimension *J*. The acquisition of L_y as L2 by learners from L_x or L_z would require increased attention to dimension *K* to perceive contrasts of L_y correctly, whereas in L_z variation in dimension *K* seems of comparatively less importance. When learning language L_z, L_y-listeners may inappropriately allocate more attention to variation in dimension *K* and may therefore fail to differentiate contrasts of L_z correctly.

However, distances between categories cannot be assumed as symmetric and equidistant as Figure 12.15 and Figure 12.14 might suggest, since attentional weight associated with specific cues may differ for contrasts *within* a given language. The language-specific mental representation of a given vowel system results from patterns of differential weighting and selective attention to the dimensions in the perceptual vowel space. The relative attentional weight of spectral and durational characteristics for specific contrasts may also vary between varieties of a given language (cf. Escudero 2002; Brandstätter & Moosmüller (in press)) which may have an effect on L2 acquisition (e.g. Escudero & Williams 2012; Williams 2013). Figure 12.16 visualizes the differential attentional weight allocated to dimensions which may vary language-specifically as well as contrast-specifically, illustrated here by non-symmetric distances between corners of the cuboid.

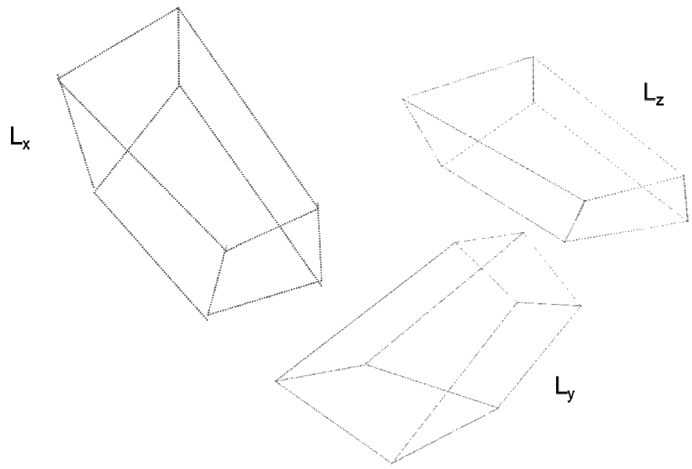

Figure 12.16: Language-specific and contrast-specific setting of selective attention along perceptual dimensions

A comparison of the Hierarchical Cluster (hcluster) analyses presented in section 5.7.2 (see Figure 12.17) illustrates the effects of selective attention. The four solutions differ in the input information, i.e. in formant peaks (F1, F2, F3) and/or formant distances (F3-F2, F2-F1) and the integration of durational information (for details, see section 5.7.2). The output of the clustering procedure is displayed on the left side of the dendrograms in Figure 12.17 showing different patterns of clusters that indicate the relative similarity of specific categories. Earlier branching indicates larger distance and less similarity in terms of physical properties of the speech signal. Minimally distant items are merged in a cluster in the leaves of the tree, reflecting maximal acoustic similarity (for a detailed discussion and comparison of the dendrograms, see section 5.7.2.).

In their output, the four solutions show very different patterns of clusters of ‘similar’ vowels (see the leaves of the dendrograms) illustrating the importance of selected criteria in the input to measure or assess “similarity” of categories (cf. section 12.2.4). This effect will be exemplified by the contrasts of German pre- and mid-palatal *i*- and *e*-vowels.

While solutions 2 and 3 may represent simulations of cue weighting for listeners who have integrated durational information in their categorical paradigm, solutions 1 and 4 are only based on spectral information. The hcluster analyses show that solutions with integrated durational information (solution 2 and 3) show a high degree of similarity for /e:/ - /i:/, while solution 1 and 4 suggest a higher degree of similarity for /e:/ - /ɪ/, an effect that is due to spectral similarity of /e:/ and /ɪ/, and a high degree of dissimilarity of pre-palatal /i:/ from all other front non-rounded categories.

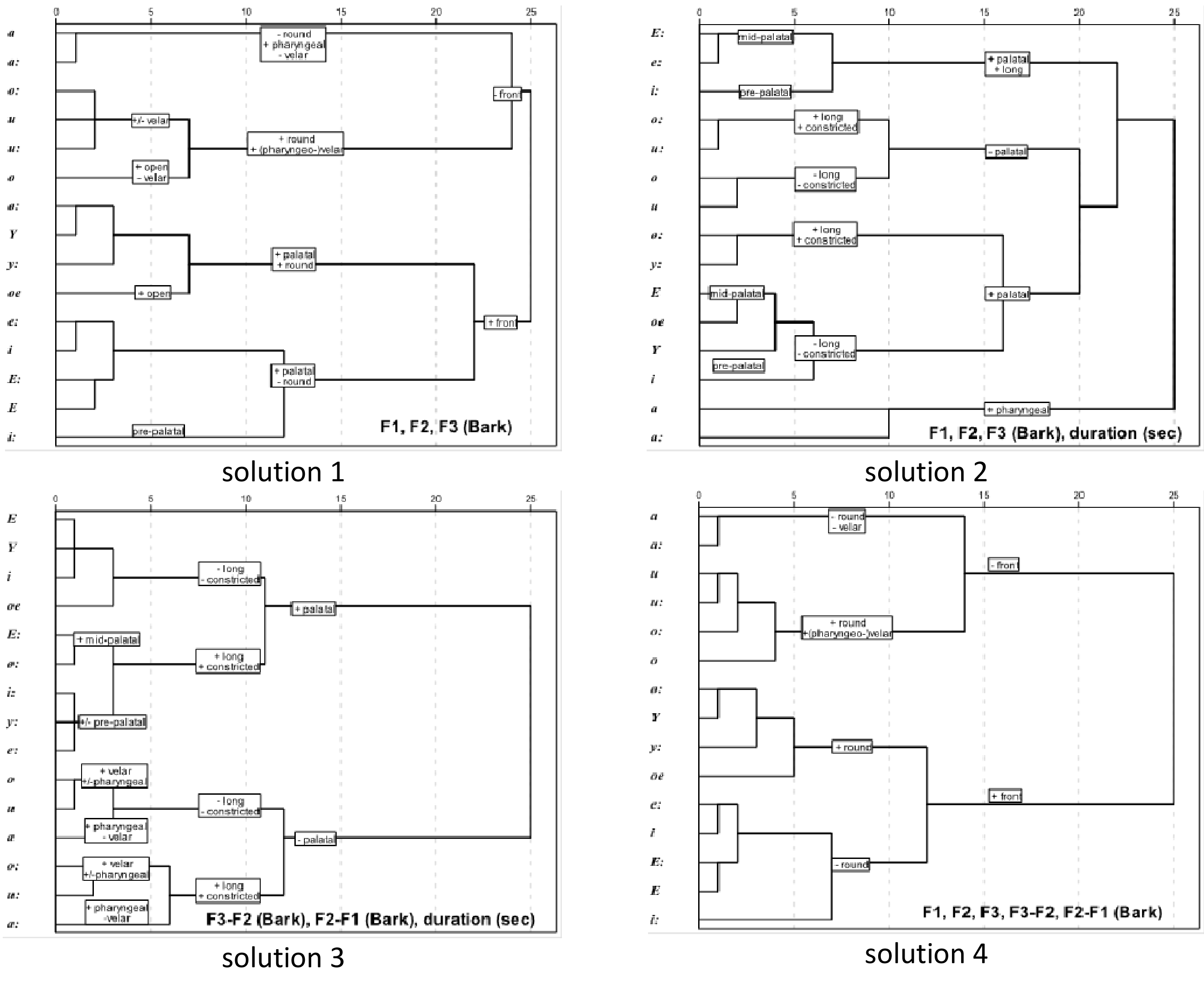

Figure 12.17: Comparison of Hierarchical Cluster Analyses illustrating differential strength of acoustic cues (for a detailed discussion of the four solutions, see section 5.7.2)

Listeners, who do not use durational cues for phonological length contrasts in their native language, are not expected to be successful in the use of durational information and the correct identification of the contrasts /i:/ - /ɪ/, at least in initial stages of learning. The effect of this lack of sensitivity to vowel-intrinsic duration as cue to length contrasts is an insufficient perceptual differentiation of /e:/ - /ɛ/ and /i:/ - /ɪ/ but also an insufficient differentiation of /e:/ and /ɪ/ that show considerable spectral similarity and are differentiated by native German listeners mainly by durational cues. In more advanced stages of L2 German acquisition, when listeners develop a certain sensitivity to durational cues (as in solutions 2 and 3), the increased attential weight in the durational dimension is expected to decrease the listeners' attention to spectral differences, causing less confusion for the /e:/ - /ɪ/ contrast but more confusion for the /e:/ - /i:/ contrast.

The appropriate weighting of specific cues in speech perception according to language-specific probability distributions in L1 is acquired by infants in their first year of life (see section 2.2), but is considered to be changeable by experience in a second language, even with "late" bilinguals (e.g. Flege & MacKay 2004, 2010; see section 2.4 and 8.7). Experiments on infants' speech perception suggest that young infants can differentiate contrasts for which older children and adults do not seem to be sensitive any more. Early studies on infants' speech perception (e.g. Eimas et al. 1971) have argued that this may be due to processes of neural atrophy by which initial perceptual sensitivity gets lost. However, other studies rather support the assumption of perceptual *attunement* by which the ability to differentiate specific contrasts is "tuned" or "warped" by exposure to linguistic stimuli in a given language (cf. a review by Aslin & Pisoni 1980; Werker & Tees 1984a, b; Best, McRoberts & Sithole 1988; see section 2.2 and 2.4). This *warping* of the perceptual space (Kuhl 1991, 2000b; Iverson et al. 2003) is due to attentional learning. *Attentional learning* involves a kind of selective attention of specific informative cues while ignoring uninformative or irrelevant ones; this re-allocation of selective attention is retained for future use (Kruschke 2011).

Several models for *attentional learning* have been proposed (e.g. Medin & Schaffer 1978; Nosofsky 1986; Kruschke 1992, 2011). Selective attention serves as explanatory construct in the learning process by which attention is rapidly *re-allocated* across cues, dimensions or representations:

> "*The attentional values are shifted in response to feedback regarding correct outcomes. Attention is shifted away from stimulus components that generate error, toward stimulus components that reduce error. The re-allocated attention values are then learned, as associations from the stimuli. The outcomes are learned from the attentionally filtered stimuli.*" (Kruschke 2011: 124).

Learning the phonetic patterns of a language moreover involves the learning of the language-specific *probability* distributions over the parametric phonetic space that can be understood as a high-dimensional cognitive map on which a metric of proximity and similarity in terms of acoustic and articulatory properties is defined (Pierrehumbert 2003: 128).

Since languages differ in number as well as in quality of phonological categories and contrasts, attentional learning does not only refer to a re-allocation of attentional weights for specific acoustic cues and probability distributions but sometimes also requires the integration of new dimensions and representations, specifically in the case when the category system of the target language is larger than the L1 system. A higher number of categories in a given L2 system requires more attention to variation of the component dimensions of the perceptual space than in L1. Learning difficulty increases by $n + 1$ dimensions by which L1 categories differ from L2 categories.

To acquire the appropriate weighting in L2, learners have to learn to allocate attention to the relevant dimension and to ignore those dimensions or parameters or acoustic cues that are not relevant for category identity. Discriminability along the relevant dimension is enhanced by selective attention. Selective attention on the other side also enhances *generalization* processes across values of irrelevant or less relevant dimensions (Kruschke 2011). As a result of increased sensitivity in a specific dimension *K* due to attentional learning, dimension *I* and *J* become comparatively less important. As a consequence of increased sensitivity in dimension *K*, the perceived similarity of items within a given class of categories sharing a feature in dimension *J or I* may decrease due to higher attentional weight of *K*. This pattern may particularly occur in dense regions of the vowel space, where several categories belonging to one class share features along a given dimension (for the relation of similarity and density, see section 12.3.3.3).

Selective attention enhances discriminability along the relevant dimension(s) and greatly enhances the possibility to generalize across values of the irrelevant dimension according to the distributional patterns of a given category system. By selective attention to one specific feature dimension, subjects would maximize within-category similarity and minimize between-category similarity (Nosofsky 1986). In terms of the spatial metaphor and geometric models of the perceptual space, the distance between these two items will be smaller compared to other items of the set, causing a higher probability of confusion of these two items.

By attending selectively to one particular dimension (e.g. duration), stimuli with equal or similar values in this specific dimension are rendered more similar to each other, but less similar to stimuli that differ in this specific feature. In an experimental context, stimuli with the same or similar feature values are expected to be rather confused than less similar ones.

> *"The key role for attention is how much each stimulus dimension is used in the calculation of similarity."* (Kruschke 2011: 124)

Attending selectively to one specific dimension (e.g. duration or frontness) maximizes *within-category* similarity as well as *within-class* similarity (e.g. [+long], [+front]) and minimizes between-category similarity with the effect of optimizing similarity relations and categorization performance in a given categorization paradigm (Nosofsky 1986: 41).

In L2 acquisition, attentional weight results from the interaction of language-specific and individual cue weighting. Language-specific weight allocated to the *n* dimensions of the perceptual space is determined by the influence of L1and L2 (and potentially also by L3, L*n*) and is moreover influenced by the learner's stage of acquisition and level of awareness for phonetic details.

> *"The tension between what is perceptually and cognitively foregrounded and what is backgrounded can be resolved in a variety of ways, and can even be resolved differently by the same person at different moments."* (Janda 2010: 9)

To conclude, as an effect of attentional learning, the perceived distance between categories that have been perceived as very similar in beginning stages of L2 acquisition, may increase because the learner allocate more attention to this specific feature, e.g. duration as cue to phonemic length distinctions in L2 German. However, by the increased sensitivity to intrinsic duration – in terms of the spatial metaphor – by stretching the perceptual space in this dimensions, distances in other dimensions, e.g. degree of aperture, may decrease.

Figure 12.15 also serves as a schematic visualization of the way how attentional learning can lead to differential weighting of component dimensions in a three-dimensional space in the course of learning processes. L_x represents the hypothetic (though quite commonly assumed) situation that stimuli vary "equally" along three binary-valued dimensions, e.g. in terms of "height", "backness" and "rounding" as assumed in most traditional phonological descriptions of vowel systems. L_y and L_z shows the situation when listeners attend selectively to one particular dimension.

The schematic representation (A) in Figure 12.18 shows a hypothetical "neutral" vowel system, in which all three dimensions (1) frontness/backness, (2) length and constriction degree and (3) aperture/height would receive "equal" attentional weight. This refers to the standard assumption in feature-based descriptions of phonological contrasts in terms of [+/– feature]. (B) represents the perceptual space of listeners who show sufficient perceptual sensitivity to differences in constriction location but make comparatively less use of durational cues due to lack of length contrasts in their L1 and insufficient experience in L2.

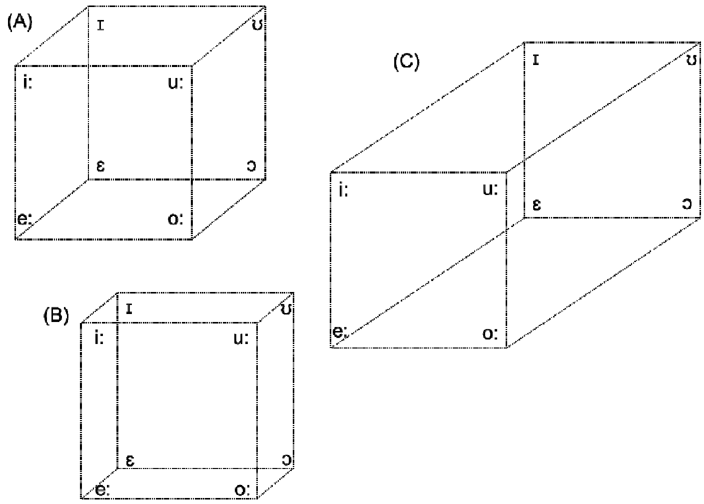

Figure 12.18: Attentional learning and the reallocation of attentional weight

In German, the so-called long-short vowel contrasts /i:/-/ɪ/, /e:/-/ɛ/, /u:/-/ʊ/, /o:/-/ɔ/, /y:/-/ʏ/, /ø:/-/œ/ do not only differ in quantity (duration) but also in quality due to differences in constriction degree. The listeners' attention in the schematical representation (B) is focused on differences in frontness/backness or more precisely in constriction location and degree of aperture, whereas differences in phonemic length and constriction degree are rather neglected. Stronger selective attention is allocated to the dimensions of constriction location/height and rounding than to the dimension of constriction degree and phonemic length. This pattern results in substitutions and confusions of vowels that differ only in constriction degree and phonemic length (as discussed in section 11.2.1 and Figure 11.10).

Eventually, by increase of experience in L2 German, listeners learn to attend more to the "duration" dimension, "stretching" their psychological vowel space by selective attention to differences in duration and constriction degree as represented in (C) in Figure 12.18. The perceptual space of listeners who have re-allocated their attention to cues in this dimension of the space is represented schematically in (C), though the complex interaction of physical characteristics and psychological moments is difficult to visualize. However, (C) may illustrate the effects of attentional learning in advanced stages of L2 German acquisition: As a consequence of the "stretching" of the perceptual space by increased sensitivity in the dimension "length and constriction degree", less attention would be allocated to acoustic cues for constriction location and degree of aperture. This may have the concomitant effect of less attention to cues that indicate the presence or absence of lip rounding and protrusion that are strongly related to differences in degree of aperture. Selective attention to a particular dimension may however enhance *generalization* processes across values of irrelevant or less relevant dimensions. As an effect of attentional learning, the perceptual space is shrunk along these dimensions whereas it is stretched in the "duration" dimension.

The selective attention to differences in duration and relative degree of constriction together with the overlap of formant frequencies for /e:/-/ɪ/, /o:/-/ʊ/ and /ø:/-/ʏ/ may result in confusion of *e*- and *i*-, *o*- and *u*- and *ö*- and *ü*-vowels and more specifically in substitutions of the type /e:/ > /i:/, /o:/ > /u:/ and /ø:/ > /y:/, which all refer to a foregrounding of durational cues at costs of the adequate identification of constriction location. Substitutions of the type /e:/ > /ɪ/, /o:/ > /ʊ/ and /ø:/ > /ʏ/, on the other hand, are rather assumed to be due to spectral similarity together with a lack of appropriate attention to durational cues that would help to disambiguate the signal. Less attentional weight may moreover be given to cues of lip rounding, since in many languages rounding is a redundant feature of back vowels, when systems differentiate front non-rounded vs. back rounded vowels. (For a detailed discussion of similarities in acoustic characteristics, see section 5.8).

To summarize, shifts in selective attention to one dimension, i.e. *stretching* the psychological space in one particular dimension, may cause shifts in another dimension, resulting in a *shrinking* of distances in a dimension that originally had received more attention. These shifts of attention as an effect of learning may cause shifts in confusion patterns and in preferences for specific response categories with more advanced learners before reaching the ultimate level of correct identification.

A comparison of the confusion data for Polish beginners (*N* = 22, 5937 responses, see Table 10.58) and Polish advanced learners (*N* = 9, 2430 responses, Table 10.59) confirms the assumed patterns of attentional learning. The Polish sample is referred to here, because it consists of two sub-samples that are relatvie homogeneous with respect to German language proficiency. The sample of beginners was recruted in university German courses after two months of German instruction. The sample of advanced learners consisted of 9 students in their second year of study who were enrolled in a teacher education program for German.

Polish beginners show a very high percentage of wrong identifications for all five front non-rounded vowels (see Table 12.3). The highest percentage of wrong identifications is given for /e:/ (only 15% correct vs. 47% correct with advanced learners). The responses for /e:/ in the sub-sample of Polish beginners are spread over a wide range of categories: /ɛ:/ 7%, /ɛ/ 2%, /ɪ/ 14%, /i:/ 38%, /y:/ 5%, <ei> 16%. With advanced learners, the percentage of wrong identifications as well as the spread over categories decreases but is still quite high (see Table 12.4). However, not only quantitative but also qualitative changes of substitution patterns are observed from initial to advanced stages of German language acquisition.

%	ɑ	ɑ:	ɛ:	ɛ	e:	ɪ	i:	ɔ	o:	ʊ	u:	œ	ø:	ʏ	y:	‹ei›	‹eu›	‹au›	wrong	corr
ɛ:			**20**	12	39	1	2					2	1		1	20	1		80	**20**
ɛ			7	**59**	28	1	1				1		1			2			41	**59**
e:			7	2	**15**	14	38	1					1	2	5	16	1		85	**15**
ɪ			2	3	4	**57**	19			1		1	2	7	3	1	1		43	**57**
i:					3	33	**61**						1	1	2				39	**61**

Table 12.3: Confusion matrix for front non-rounded categories by Polish beginners

%	ɑ	ɑ:	ɛ:	ɛ	e:	ɪ	i:	ɔ	o:	ʊ	u:	œ	ø:	ʏ	y:	‹ei›	‹eu›	‹au›	wrong	corr
ɛ:			**20**	12	65	1	1								1	1			80	**20**
ɛ			9	**77**	13	1												1	23	**77**
e:			6	5	**47**	4	29					1	1	3	2	1			53	**47**
ɪ			1	4	2	**77**	10					1		5	1				23	**77**
i:					1	21	**76**					1		1	1				24	**76**

Table 12.4: Confusion matrix for front non-rounded categories by Polish advanced learners

Beginners tend to interpret long *e*-vowels as diphthongs (or what they think <ei> may stand for): /e:/ > <ei> 16%, /ɛ:/ > <ei> 20%. With advanced learners, hardly any perceptual diphthongizations are observed, i.e. duration is correctly interpreted as length (see Table 12.4). The percentage of substitutions /e:/ > /ɪ/ decreases considerably with higher language proficiency (14% beginners, only 4% advanced), presumably due to the integration of durational cues in the learners' mental representations. As an effect of increased language experience, higher attentional weight is assigned to durational differences which enables listeners to identify the contrast /e:/ - /ɪ/ correctly at cost of attention or sensitivity to spectral cues that differentiate /e:/ and /i:/ (higher F1 and lower F3 for /e:/ than for /i:/). However, one type of substitution is still very frequent with Polish advanced learners of German: /e:/ is identified as /i:/ by advanced learners in 29% of the cases (vs. 38% beginners). In other words, advanced learners have acquired a sensitivity for length contrasts but still have difficulties with the differentiation of mid-palatal /e:/ and pre-palatal /i:/.

Similar effects are observed for the class of German front rounded vowels (see Table 12.5 and Table 12.6). Beginners show a very high percentage of wrong identification (> 60%) for this vowel class. With advanced learners, considerable improvement is observed for all front rounded vowels except /ø:/. Increased L2 experience does not only cause improvements with respect to the number of wrong identifications but also shows shifts in substitution patterns. Moreover, front rounded vowels seem to be less attractive targets for "hypercorrect" substitutions of front non-rounded with front rounded vowels.

%	ɑ	ɑ:	ɛ:	ɛ	e:	ɪ	i:	ɔ	o:	ʊ	u:	œ	ø:	ʏ	y:	‹ei›	‹eu›	‹au›	wrong	corr
œ		1	10	27	16		2	2	2	1	1	**14**	8	8	4	1	4		86	**14**
ø:					2	1	2		1	5	18	7	**16**	13	35				84	**16**
ʏ					3	3	3		1	13	8	11	7	**30**	19	1	1	1	70	**30**
y:			1	1	2	3	10		1	5	8	4	7	18	**40**		1		60	**40**

Table 12.5: Confusion matrix for front rounded categories by Polish beginners

%	ɑ	ɑ:	ɛ:	ɛ	e:	ɪ	i:	ɔ	o:	ʊ	u:	œ	ø:	ʏ	y:	‹ei›	‹eu›	‹au›	wrong	corr
œ			4	3	3	1	1			3	1	**39**	9	25	11				61	**39**
ø:						1				1	2	6	**17**	18	56		1		83	**17**
ʏ						3	1			5	1	9	4	**60**	17				40	**60**
y:							3					2	4	22	**70**				30	**70**

Table 12.6: Confusion matrix for front rounded categories by Polish advanced learners

In the sample of Polish beginners, /œ/ is identified correctly in only 14% of its occurrences (advanced learners 39% correct). The beginners' responses for /œ/ spread over a wide range of different categories: /ɛ/ 27%, /e:/ 16%, /ɛ:/ 10%, /ø:/ 8%, /ʏ/ 8%, /y:/ 4%, <eu> 4%, /i:/ 2%, /ɔ/ 2%, /o:/ 2%, /ʊ/ 1%, /u:/ 1% and <ei> 1%. Advanced learners show clearly different patterns in perceptual substitutions for /œ/. /œ/ is identified correctly in 39% of the cases and substituted by /ʏ/ 25%, /y:/ 11%, /ø:/ 9%, /ɛ:/ 4%, /ɛ/ 3%, /ʊ/ 3%, /ɪ/ 1%, /i:/ 1% and /u:/ 1%. Delabialization is substantially less frequent than with beginners. On the other hand, substitutions of /œ/ and /ø:/ with *ü*-vowels are more frequently observed than with beginners and no perceptual "diphthongizations" occur. While /œ/ is subject to substitution with corresponding non-rounded vowels in the beginners' sample, advanced learners have developed a sensitivity for [+/– round] contrasts for front vowels. However, in the sample of advanced learners, a shift of attention towards [+/– round] front contrasts goes along with a higher preference for *ü*-vowels that is even stronger with /ø:/-stimuli (see Table 12.6).

For /ø:/, no quantitative effects of language experience on id_correct scores are observed: /ø:/ is identified correctly by beginners in only 16% and by advanced learners in only 17% of the cases. However, considerable qualitative shifts are observed in the responses: While with beginners, major substitutes for /ø:/ are /y:/ (35%), /u:/ (18%) and /ʏ/ (13%), advanced learners prefer /y:/ (56%) and /ʏ/ (18%). These patterns clearly show a reallocation of selective attention as a result of attentional learning.

A rewarding issue for further research would be a longitudinal study of the development of perceptual discrimination ability for German vowels with listeners of a given language from the initial stages of learning in order to identify processes of attentional learning, the eventual development of category representations and the listeners' strategies in a long-term experiment wher L2 learners are tested at different stages of their acquisition process.

To summarize, attentional learning involves processes of reallocation of selective attention to specific cues in the signal. Reallocation of selective attention causes increased sensitivity in specific dimensions but may be accompanied by a decreased sensitivity to characteristics along other dimensions, specifically within a given class of vowels. Discriminability of L2 categories in the perceptual vowel space may vary *language-specifically* due to differential selective attention along perceptual *dimensions*. It may moreover vary between vowel classes as well as between specific category *contrasts* in a given natural class (see Figure 12.16) rather than developing in a symmetric and equidistant way as suggested in a feature-based view of phonological contrasts (cf. Figure 12.14). Attentional weight may moreover vary *inter-individually* as an effect of the listeners' native language and their language variety as well as by experience and level of proficiency in L2 acquisition.

12.6 Conclusions

This chapter provided a detailed discussion of the notions *similarity*, *bias* and *asymmetry* in L2 vowel perception and their effects in an experimental identification task. The articulatory, acoustic or perceptual vowel space is conceived as a multi-dimensional space varying in several phonetic parameters. Signals as well as mental representations vary in several dimensions. Rather than one specific distinctive feature, multiple co-occurring (and partly context-specific) cues in the signal that receive differential attentional weight in speech perception yield accurate identification. The attentional weight is determined by language-specific, signal-related and listener-related aspects that influence the listeners' decision for a specific response category in an experimental identification task.

Vowel identification can be viewed as a process of comparing and matching acoustic features of the input with mental representations. In these processes of comparing and matching, "similarity" is crucial. Several models have been proposed for the structure of mental vowel categories and their role in speech perception (see section 12.1.3.2). In the present study, a theoretical view is favoured that defines vowel categories in terms of statistical density distributions of phonetic correlates associated with category labels. According to this view, a vowel inventory is a set of statistical relationships between category labels and density distributions over the parametric level (Pierrehumbert 2003). The set of vowel categories, the relevant phonetic correlates and statistical distributions as well as the listeners' calculation of probability may vary from language to language and even from listener to listener due to their language experience. Auditory sensations and stored representations are linked by the listener in a kind of calculation of *similarity* and probability. *Similarity* is one of the fundamental constructs to account for cognitive tasks such as categorization and category learning. Similarity as a cognitive strategy is particularly involved when humans – e.g. learners acquiring a foreign language – do not exactly *know* about relevant cues in the input but use similarity to infer assumptions about stimuli and stored representations.

Several models of similarity have been proposed in cognitive psychology, among which feature-based and geometric models are most influential (see section 12.2). Feature-based models have a long tradition in linguistics, particularly in phonology and semantics, whereas assumptions of geometric models have been received rather implicitly in linguistics, e.g. by referring to the spatial metaphor of the "vowel space" in phonetic descriptions of a given vowel inventory. Geometric models assume that objects are located in an n-dimensional space and that similarity s can be viewed as a monotonous decreasing distance function d in a metric space. Distances between items may be identified in terms of Euclidean distances (e.g. in the metric space) or as rectilinear distances in terms of number of distinctive features (in feature-based accounts). Multi-dimensional Scaling (MDS) provides a low-dimensional metric representation of the n-dimensional perceptual space. The language-specific configurations, i.e. the "perceptual vowel maps" for each of the L2 learners' native languages (see Figure 9.18), are derived from the listeners' perceptual confusions in L2 vowel identification and provide a low-dimensional representation of the multi-dimensional vowel space by means of MDS (see section 7.3).

A comprehensive model of L2 vowel perception requires the differentiation of (1) phonetic distances p_{ij} vs. perceived similarities s_{ij}, and (2) a consideration of differential *bias* in perception and its provenance as (a) stimulus-related bias b_i vs. response-related bias b_j, (b) signal-related vs. listener-related, and (c) set-dependent vs. set-independent biases. This approach allows for differentiated insights into intra-lingual relationships of *perceived similarity* between L2 categories, their *mental representation* of L2 categories and the listeners' *perceptual vowel space* in L2. Relationships of perceived similarity as established by L2 learners allows for a better understanding of the learners' *difficulties* in acquiring the German vowel system, the probability of perceptual *confusions*, the listeners' *preferences* for vowel categories and directions of perceptual substitution processes in a perceptual identification task.

Referring to L2 confusion data from a forced choice identification task, the perceived similarity between two items can be defined by the formula $s_{ij} = p_{ij} * b_i * b_j$ to refer to relationships of perceived similarity s_{ij}, phonetic proximity p_{ij} and biases of stimuli b_i and responses b_j.

The probability that stimulus *i* is categorized as response *j* is determined by a function of distance and similarity in *n* dimensions and a *bias* parameter for the relevant dimensions of a metric perceptual space. *Bias* parameters are a crucial element of geometric similarity models to compensate for problems resulting from violations of the three principles (1) *minimality* and self-similarity ($d_{ij} \geq d_{ji} = 0$), (2) *symmetry* ($d_{ij} = d_{ji}$), and (3) *triangle inequality* ($d_{ij} + d_{jk} \geq d_{ik}$) (see section 12.3). In the present L2 confusion data, several instances of *perceptual bias* and *asymmetries* in substitution patterns are observed (see section 11.3 and 12.3). The nature of these biases and their influence on relationships of intra-lingual similarity between L2 vowel categories is considered in more details in section 12.4.

The acquisition of L2 vowel categories is determined by processes of selective attention and attentional learning. Stimulus-related bias b_i is modelled as *attentional weight* of specific cues in the signal. Attentional weight is the result of selective attention to specific cues assigned by the listener. Selective attention is minimizing the sensitivity to within-category differences and is maximizing the sensitivity to between-category differences within the category system of a given language. Perceptual enhancing of relevant cues brings benefits in speech processing in a given language and is a substantial element of language learning.

The acquisition of a second language phonology involves processes of attentional learning and the reallocation of patterns of language-specific selective attention to specific cues in the L2 signal. As exemplified in section 12.5, the data of the present study show that selective attention varies not only language-specifically, but also category-specifically and inter-individually according to the learner's stage of acquisition. The reallocation of selective attention to specific dimensions in L2 involves not only quantitative changes in the percentage of wrong and correct vowel identifications but also qualitative changes in the L2 learners' patterns of perceptual confusions and the direction of perceptual substitutions.

General Conclusions

Vowels are complex linguistic structures that vary within and across languages according to several articulatory and acoustic phonetic as well as phonological parameters. In this sense, they can be understood as conventionalized pairings of variable phonetic substance (form) and discrete phonological categories (function). Due to language-specific conventions, no one-to-one mapping can be assumed between the speech signal and the vowel category perceived by the listener. Rather, the interpretation of the same acoustic signal may vary with the listeners' native language background. In other words, vocalic form-function pairings vary language-specifically as a result of a listener's experience in one or more languages. This study has provided substantial evidence for language-specific pairings of vowel tokens and phonological categories.

Phonetically, vowels can be described as sounds varying along several *continuous* articulatory or acoustic parameters, whereas vowel perception apparently is determined by *discrete* category boundaries. A description in terms of distinctive phonological features such as Frontness, Height and Rounding is – apart from difficulties to identify their distinct acoustic correlates – not sufficient for an understanding of the complex interplay of articulatory, acoustic and auditory aspects of phonemic structures in native and non-native vowel perception. Rather than one specific distinctive feature, *multiple co-occurring cues* in the signal that are partly context-specific contribute to accurate vowel identification. The contrast /i/ - /y/, for example, is defined by several articulatory parameters such as degree of lip rounding, tongue location, larynx lowering, degree of palatal narrowing or tongue blade elevation and their acoustic effects (Wood 1986). Therefore, for an understanding of auditory effects in the listener's perception in L1 or L2, it is argued here, a phonetic description of vowel sounds in a given language in terms of articulatory gestures and their acoustic effects in the signal is necessary together with language-specific and listener-specific aspects that also involve knowledge from external components other than phonetics and phonology such as grammar and the lexicon and – as has been claimed in several studies – the influence of so called "universal" forces in speech perception. In the current study universals are understood in the weak sense of statistical tendencies for which motivations in terms of phonetic substance and physiological and cognitive conditions of human speech perception and production are given.

Research on vowel sounds, vowel inventories and vowel perception is strongly determined by the co-existence of two theoretical paradigms: the *categorization* paradigm and the paradigm of the *continuous vowel space*. Experimental phonetic studies on vowel identification are strongly influenced by the *categorical view*. Central notions of the categorization paradigm such as contrast and distinctiveness, category boundaries, the irrelevance of within-category phonetic variation, symmetry of perception and production and the consistency and reliability of intuitions of the so-called "native speaker" (for discussion, see section 12.1.3.1) are implicitly assumed in most works in experimental phonetics, socio-linguistics and second language acquisition research, though – as discussed in detail in this study – these implicit assumptions have to be scrutinized, especially when dealing with non-native speech perception.

Categorical perception seems responsible for the listeners' ability to distinguish exemplars of one category from those of other categories despite some shared properties, which have no differentiating function as such. Thus, categorical perception seems to create perceptual invariance, irrespectively of some auditorily detectable inter-stimuli differences, i.e. it implies the ability to identify a wide range of different phones as being "the same" accounting for the differentiation of between-category but not within-category differences. Signal-inherent as well as external factors such as contrast type, use and frequency of occurrence in the listener's native language contribute to the listener's capacity to reconstruct the intended form-function pairing in perception.

When speaking of vowel systems, reference is often made to the construct of a *system of discrete vowel categories*, whereas phonetic-oriented approaches rather focus on the construct of a *continuous vowel space*. The articulatory-acoustic vowel space can be understood as an abstraction of the total set of physiologically possible articulatory constellations and resulting resonances in the vocal tract that are traditionally described in terms of formant frequency values (F1, F2, F3, ...). The range of possible vowel sounds is delimited by articulatory, acoustic and perceptual coordinates of the vowel space that have been discussed in chapter 4 and 5.

Vowel categories are considered as related to each other by relations of contrast and perceptual distance. A number of typological and phonetic studies on vowel systems and vowel contrasts refer to a general preference for qualities that are maximally distinct from each other and to the assumption that constellations of vowel systems are motivated by maximal contrast, i.e. perceptual distinctiveness (at minimal articulatory cost) (cf. Liljencrants & Lindblom 1972; Lindblom 1986; Donegan 1978; de Boer 2000, 2001; Flemming 2004; Schwartz et al. 2005). Accordingly, a number of studies have argued that acoustically more *salient* and more *robust* categories are favoured in the languages of the world. Cross-linguistically, these preferences are manifested in frequency of occurrence in the languages of the world. Moreover, cross-linguistic distribution patterns have been argued to correlate with token frequency within an individual language and developmental sequences of language learners (see section 5.5). Perceptual distinctness and phonetic distance are however difficult to define in absolute terms.

Acoustically, vowels can be described in terms of several continuous metric variables such as formants, formant distances, overall spectral shape, bandwidth, amplitude and dynamic spectral change. The relevance of specific acoustic dimensions and cues in the signal for vowel identification and the mental representation has been subject of intense scientific investigation. It is necessary to be aware of the fact that phonetic analysis always involves specific underlying assumptions of the researcher that determine his/her choice of specific phonetic parameters such as spectral peaks or rather the whole spectral shape, formant frequencies or rather formant distances, measurements in Hertz or in Bark etc. In other words, when referring to acoustic properties but also to articulatory characteristics of vowel sounds there is no unanimously accepted set of defining parameters for a calculation of phonetic distances and similarities of vowels within a given system or between the system of L1 and L2 (see Figure 12.7 and Figure 12.8). A comparison of vowels and vowel systems in terms of a conventional F1xF2 chart is insufficient as it does not reflect the many parameters that determine vowel sounds and their categorization in perception.

In speech perception, sensory events are matched with previously experienced items (or their memories or traces). Perceptual stimuli are categorized by their degree of *similarity* to stored items. In this sense, categorization may vary as a function of previous experiences. Objectively different vowel sounds are perceived at some cognitive level as more or less similar or distinct from others, belonging either to the same or to different categories. In other words, the categorization of sounds is based on perceptual "identity" and relationships of "similarity". Shifts in listeners‘ perception of acoustically different sounds as belonging to the "same" or a "different" category are conditioned by signal-inherent as well as external factors from the listeners' language knowledge, context and expectations.

Phonetic *distance* and perceptual *similarity* can be understood as *gradient* concepts varying along a continuum, whereas category boundaries are rather conceived as *disrupt* changes in the listeners' perception. An inversely functional correspondence $\eta_{ij} = f(d_{ij})$ between distance d and similarity η is assumed in many cognitive models on similarity (e.g. Shepard 1957; Nosofsky 1986, 1988). Increasing distance in phonetic substance (form) is considered to increase distinctiveness and guarantees phonemic contrast (function). By a decrease in distance, on the other hand, which is in inverse functional correspondence to similarity, the form-function correspondence may be at risk. However – as outlined above – no linear correlation of gradient variation in specific phonetic parameters and perceived phonemic identity can be observed. Variation exceeding a specific threshold in a given parameter causes disrupt changes in perception. Rather than measurable *phonetic distance* and similarity, it is *perceived similarity* between stimuli and categories – which may vary language- and listener-specifically – determining a listener's matching of perceptual stimuli and mental categories. Therefore and for the above mentioned reasons of difficulty to select the perceptually "relevant" phonetic parameters, the current study offers an approach that is focused on *perceptual similarity*, i.e. the listeners' language-specific perception of vowel sounds rather than similarity in terms of a priori specified phonetic parameters.

The assumption that similarities and differences between L1 and L2 determine the learners' success in L2 has a long tradition from early works in the Contrastive Analysis paradigm to current models of perception and the acquisition of second language phonology. These approaches show a strong focus on *between-language differences* in category distribution and within-category variation. The operationalization of perceptual "similarity" in L2 has however so far not received sufficient attention.

In learner-oriented and usage-based approaches to second language acquisition that consider interlanguages as dynamic emergent systems based on the learners' experiences in L1 and other acquired languages, I have argued, *intra-lingual similarity* relations between L2 categories as perceived by the learners are crucial for an understanding of the specific way that L2 learners develop phonological categories of the target language. While a contrastive analysis focusses on *inter*-lingual differences between L1 and L2, intra-lingual relations are of equal relevance particularly for L2 learners that already have L2 experiences and hypotheses about the target language.

To conclude, in the present study, the *multi-dimensional perceptual vowel space* as well as the articulatory or acoustic vowel space is conceived as varying language-specifically in terms of a number of *n* articulatory and acoustic parameters or dimensions that differentiate vowel categories within a given language system. Vowel categories (types) and individual phones (tokens) belonging to a given category differ in a number of articulatory and acoustic parameters, for which it is difficult to determine their actual role in perception and categorization in absolute terms. Languages differ not only in the number of vowel categories but also in category distribution and range of phonetic variation within categories and the *differential weighting* of the component dimensions of the perceptual vowel space (see section

12.5). Listeners – according to their experience with one or more languages – differ with respect to their mental representations for categories of the language under consideration and their *selective attention* to specific cues in the acoustic signal. In other words, mental representations of a vowel system and its categories are determined by the listeners' perceptual experiences in one or more languages. Interlanguages conceived as dynamic systems are understood as emerging from the L2 learners' input and the individual way they perceive and conceive the vowel system of the target language. The present study therefore has aimed at the description of *perceived similarity* guiding L2 learners' decisions in an identification task rather than at a mere description of phonetic similarity in terms of acoustic or articulatory vowel parameters.

As a starting point, an articulatory-acoustic model based on a classification of vowels in terms of *constriction locations* (see Wood 1979; 1982) was presented in section 4.7.6, 5.4 and 5.6. In this account, four discrete constriction locations are defined: (1) a constriction in the *lower pharyngeal* region for [ɑ-a-æ]-like "pharyngeal" vowels, (2) a constriction in the *upper pharyngeal* region for [o-ɔ] and [ɤ]-like "uvular" vowels, (3) a constriction in the vicinity of the *soft palate* for [u-ʊ] and [ɨ]-like "velar" vowels, and (4) a constriction location along the *hard palate* for [i-ɛ]-like and [y-ø]-like "palatal" vowels (Wood 1979: 30f). For palatal vowels, language-specific tendencies to either pre-palatal or mid-palatal tongue positions are reported (Wood 1982a: 43). Languages contrasting [i] and [y] such as German seem to prefer the pre-palatal position for both vowels (Wood 1979, 1982, 1986).

As an alternative to the conventional IPA vowel triangular or quadrilateral representation, I propose an elliptic representation (see Figure 4.3. and for German Figure 5.24) that reflects the differentiation of vowel sounds in terms of front vs. back, the major constriction locations (pre- or mid-palatal, velar, uvular, pharyngeal), and additional parameters of constriction degree and rounding. This representation serves as a basic model for the classification of German vowels and for the identification of perceptual substitution processes that are manifested in the listeners' responses in the experimental part of this study.

For the German vowel inventory, I have proposed a classification in terms of the basic distinctive dimensions and parameters of constriction location, degree of mandibular aperture, constriction degree and phonemic length, and lip rounding and/or protrusion (see Table 5.9). German vowel categories are classified here into five major classes according to the location of narrowing in the vocal tract (constriction location): (1) pharyngeal /ɑ ɑ:/, (2) uvular /ɔ o:/, (3) velar /ʊ u:/, (4) mid-palatal /ɛ ɛ: e: œ ø:/ and (5) pre-palatal /ɪ i: ʏ y:/ qualities (cf. Wood 1979), which are further differentiated by mandibular aperture, rounding, constriction degree and phonemic length. On the basis of this classification into these five major vowel classes and based on acoustic measurements of the experiment's input stimuli, German vowel categories were compared and predictions on the relative similarity and difficulty of categories and contrasts were derived from a hierarchical cluster analysis (see section 5.7).

The experimental part of this study consists of a large-scaled cross-language comparison of German vowel identification. 173 non-native listeners from ten different native languages and a native German control group performed a forced-choice category identification task: 15 German vowel qualities were presented in varying consonantal contexts and had to be identified by the listeners as belonging to one of the presented response categories. The set of response options consisted of the full set of 15 vowels occurring in stressed syllables and three diphthongs. The results were analyzed with respect to three different aspects: (1) a listener-related analysis (chapter 8), (2) a language-specific analysis (chapter 10), and (3) a vowel-specific analysis (chapter 11). A general overview of the major results of the study and a cross-language comparison is provided in chapter 9 for all L2 listeners and in chapter 11 for each of the vowel categories.

The *listener-related analysis* showed a significant main effect of the listeners' native language and their level of proficiency on rate of correct vowel identification, while the very general assumption of "the earlier the better" (referring to age of learning onset) and "the longer the better" (referring to length of learning and/or length of residence) could not be fully proved by the data. These common though very general findings are further differentiated by a language-specific and a category-specific analysis in chapters 10 and 11.

The experimental data clearly illustrate that the same acoustic input is perceived differently by listeners from diverse native language backgrounds. A detailed quantitative and qualitative data analysis of observed correct and wrong identifications (*difficulties*), the listeners' *preferences* for specific response options, and type and direction of perceptual substitution processes (*patterns of confusion*) was performed for each of the L1 sub-samples. Both the category-specific analysis as well as the language-specific analysis reveal a number of language-specific differences as well as common patterns of perceptual substitutions that are found in several of the language sub-samples.

The listeners' perceptual substitutions were analysed in terms of substitutions involving changes in (1) constriction degree and phonemic length, (2) constriction location (by Fronting vs. Backing processes), and (3) labiality and/or frontness/backness. A cross-language analysis of the experimental results in chapter 9 and 11 has demonstrated that a major part of perceptual substitutions were due to insufficient differentiation of German vowels in terms of constriction degree and phonemic length, whereas less substitutions are observed with respect to under-differentiation of constriction location and labiality and/or frontness/backness.

With respect to substitutions due to Fronting vs. Backing and Labialization vs. Delabialization considerable evidence for *asymmetries*, i.e. differential propensity for wrong identification and directionality of perceptual substitutions are found.

While *a*-vowels, /i:/, /u:/, /ɛ/ and /ɔ/ are particularly stable, most difficulties are observed with contrasts of front rounded /y: ʏ ø: œ/ vowels, /e:/ and /ɛ:/. Several sub-samples show a strong tendency of /i:/, /u:/ and /y:/ to attract responses for /e:/-, /o:/- and /ø:/-stimuli respectively, whereas perceptual substitutions in the opposite direction are rather uncommon. In other words, substitutions of /e:/ > /i:/, /ø:/ > /y:/ and /o:/ > /u:/ are significantly more common than perceptual substitutions in the opposite direction. For front rounded qualities, the data reveal language-specific tendencies with respect to substitutions of front rounded vowels either with back rounded or front non-rounded categories.

The data analysis is not confined to a mere descriptive analysis of the L2 listeners' identifications and the observed substitution patterns and preferences for specific target categories. On the basis of perceptual confusion data (confusion matrices), quantifiable relations of similarity between vowel categories (similarity scores) as well as geometric representations of the language-specific perceptual vowel space of German for L2 learners from different language backgrounds were derived by the statistical technique of Multidimensional Scaling (MDS) (Shepard 1957, 1980). By means of MDS, the multi-dimensional vowel space, which is determined by a large number of articulatory and acoustic cues, is transformed to a low-dimensional metric representation of inter-category relationships of similarity for each of the L1 sub-groups. The MDS procedure offers an approach to perceptual similarity and distance as perceived by the listeners, which is independent of a priori selected phonetic parameters but can be interpreted in articulatory or acoustic terms in a post-hoc analysis. This approach to "psychological similarity" avoids the difficulty of selecting specific phonetic parameters outlined above, which may not correspond to the listeners' attentional weighting of cues in the signal.

Referring to the listeners' experience-driven matching of phonetic form and phonological function, it is essential to differentiate *phonetic similarity* in terms of measurable characteristics in the acoustic signal may be identified from *perceived similarity*. A qualitative analysis of the major *patterns of confusion* in L2 vowel identification, a category's propensity for wrong identification (*difficulty*), its susceptibility to specific perceptual substitution processes and its probability to be selected as response option (*preferences*) provides evidence not only for discrepancies between phonetic distances and perceived similarity but also for *differential perceptual biases*.

For perceived similarity between two items *i* an *j* the formula

$$s_{ij} = p_{ij} * b_i * b_j$$

was proposed referring to perceived similarity s_{ij}, that is conditioned by the interaction of phonetic distance or proximity p_{ij}, stimuli bias b_i and response bias b_j. A crucial distinction is made here between stimulus biases and response biases. Differential *biases* in perception must be differentiated in terms of (a) stimulus-related biases b_i vs. response-related biases b_j, (b) signal-related vs. listener-related biases, and (c) set-dependent vs. set-independent biases.

To conclude, the probability that stimulus *i* is categorized as response *j* is determined by a function of distance and similarity in *n* dimensions and by *bias* parameters for relevant dimensions in a metric perceptual space. The operationalization of language-specific perceived similarity in terms of (1) phonetic similarity, (2) stimulus biases and (3) response biases is a substantial outcome of the present study contributing to a better understanding of *intra*-lingual relationships of *perceived similarity* between L2 categories, their *mental representation* of L2 categories and the listeners' *perceptual vowel space* in L2 (see section 12.2.2 and 12.4). The operationalization of relationships of *perceived similarity* as established by L2 learners allow for a better understanding of the learners' difficulties in acquiring the German vowel system, the probability of perceptual confusions, the listeners' preferences for vowel categories and directions of perceptual substitution processes in a perceptual identification task. The MDS representation provides a way to visualize the language-specific constellation of categories in the perceptual vowel space.

Referring to an *integrated articulatory-acoustic model* as introduced in this dissertation (section 5.4 and 5.6), specific acoustic cues or structures cannot be argued to be auditorily or perceptually salient or distinct as such. Perceptual *salience* is conceived to be relative rather than absolute referring to the listeners' *selective attention* to specific cues in the signal. Selective attention varies according to the listeners' linguistic experience and individual strategies in an experimental task. Learning happens when shifts in selective attention to specific cues in the signal occur. In this sense, the learning of a second language phonological system is modelled here as a process of attentional reallocation, i.e. of "re-tuning" selective attention along specific acoustic parameters. Therefore, *perceptual learning* must involve a "continuous mode" of perception and a shift of attention towards fine phonetic details in the L2 input (that may be less relevant in L1) rather than perception in a "categorical mode"

(Wode 1990) in order to develop adequate mental representations and appropriate articulatory constellations for L2 vowel categories.

The study's results are of particular significance for *pedagogical purposes and teacher education*. The confusion matrices and perceptual vowel maps for each of the listeners' native languages provide valuable evidence for the language-specific way that vowel sounds are categorized according to the listeners' native system and his/her language experience in German and other languages. The data provides ample evidence for the fact that the same acoustic input may be perceived quite differently by listeners from diverse language backgrounds. The MDS representations are a valuable mean to visualize the language-specific tuning of the perceptual vowel space. Perceptual vowel maps represent perceived similarity between categories and small "distances" between similar categories. These perceived similarities refer primarily to three "similarity clusters" in the German vowel system (1) *i*- and *e*-vowels, (2) *ü*- and *ö*-vowels and (3) *u*- and *o*-vowels, accounting for a number of difficulties that are observed with L2 learners of German.

For teacher education as well as for pedagogical practice, I recommend that greater attention is given to the following two types of illusions: (1) "perceptual illusions" of phonemic identity despite substantial variation between tokens belonging to the same category; this perceptual illusion is strongly driven by (2) "orthographic illusions" that sounds that are represented by the same grapheme in German orthography were "the same" (cf. Kerschhofer-Puhalo 2012, in press a). Central issues in teacher education should refer to the phenomenon of language-specific perception as well as to consequences of category confusion in acquisitional processes. Perceptual category confusion between German vowel categories due to insufficiently perceived contrasts and language-specifically perceived similarity is at the origin of so many difficulties in the acquisition of adequate category representations and their function in grammar and lexicon. Misperceptions and misconceptions of German vowel categories are directly manifested not only in "non-native" pronunciation but also in written production and the listeners' conceptions of lexical items and morphological patterns in German.

Despite the significant achievements of research encouraged by technological advances in experimental methods of acoustic-phonetic research, several theoretical, typological, methodological and pedagogical issues have to be addressed in future research. Future studies will have to focus not only on the impact of specific signal-inherent acoustic cues in language-specific vowel identification as done in a number of experimental-phonetic laboratory studies that usually focus only on a small number of vowel contrasts in controlled contexts. More lab-based studies have to be done on the specific impact of consonantal contexts on vowel perception beyond the typical experimental contexts of /p t k/. As a matter of fact, place and manner of articulation have an effect not only on the perception of vocalic constriction location as well as on perceived duration and constriction degree. With respect to pedagogical purposes, contextual effects of other consonants such as fricatives, liquids, nasals and glides should receive more scientific attention even if they are "less optimal" in an experimental context. Another interesting aspect for further research refers to syllable structure and to the perception of phonemic length and constriction degree in CV- vs. CVC- or more complex syllables.

In addition to lab-based studies on the perception of fine phonetic details, acquisitional studies that focus on the listeners' strategies to cope with a large vowel inventory such as the German system might reveal new facts for pedagogical purposes as well as for linguistic theory. As the analysis of the two Polish sub-samples of beginners and advanced learners (section 12.5) has shown, perceptual substitution strategies not only vary with respect to the listeners' native language, but also by stage of language acquisition. In the course of L2 acquisition, learners develop not only perceptual categories but also hypotheses about the L2 system and strategies to cope with difficulties so far experienced. Longitudinal studies of L2 learners' difficulties in vowel perception and pronunciation and the development of mental representations for German vowel categories observing learners from initial stages of learning to more advanced stages are of course time-consuming and costly but provide valuable information on developmental sequences in language acquisition that are of relevance for pedagogical practice as well as for linguistic theory for a better understanding of constraints, preferences and biases in the course of language acquisition.

Furthermore, the role of orthography in the development of mental categories for German vowel sounds is an issue for future investigations as well. Experimental investigations may also address possible differences between learners that experience L2 German mainly by oral input vs. learners that learn German in language programs largely based on written input (cf. Kerschhofer-Puhalo, in press a). More empirical research would also be needed with respect to category formation and system organization in "the bilingual mind", i.e. the mind of learners who acquire two languages simultaneously from early childhood on. Here, achievements of first language acquisition and experimental phonetics may converge with the slowly growing field of third language phonological research.

A major desideratum is more empirical research on less studied languages as done in the present study. The phonetic analysis of vowel sounds and vowel systems of languages that are the native languages of immigrants living in German-speaking countries deserves considerably more scientific attention than it has received until now. The exploration of phonetic characteristics of less studied and typologically diverse languages will not only enlarge the range of evidence for typological comparisons but will also contribute to an improvement of language instruction and an understanding of the learners' difficulties in acquiring the complex German vowel system.

For a better understanding of the learners' difficulties and the way they develop a foreign language's phonological system, a *user*-centered view rather than a view focussing on language *systems* is recommended here. Only if second language acquisition is viewed in the more general context of real usage by language learners, i.e. by individuals who construct their specific interlanguage systems from the available input, an understanding of the many factors that determine the learners' language behaviour and their difficulties will be possible revealing not only differences but also common patterns and *similarities* between language systems and individual language learners. In this sense, universal preferences are understood not as referring to specific patterns or structures of language systems but rather to the direction of those physiological and cognitive *biases* that guide native and non-native listeners' perception and the development of mental representations of (inter-)language systems.

List of Figures

List of Tables

References

Abbreviations:
ICPhS = Proceedings of the International Congress of Phonetics
JASA = Journal of the Acoustical Society of America
JPhon = Journal of Phonetics

[URLs accessed on 2014-11-07]

Abd-El-Jawad, H. 1987. Cross-dialectal variation in Arabic: Competing prestigious forms. *Language in Society*, 16 (3), 359-367.

Abondolo, D. 2011. Hungarian. In: Comrie, B. (ed.). *The World's Major Languages*. London New York: Routledge, 484-496.

Abou Haydar, L. 1994. Norme linguistique et variabilité dialectale: analyse formantique du systéme vocalique de la langue arabe. *Revue de Phonétique Appliquée*, 110, 1-15.

Abramson, A. & Lisker, L. 1970. Discriminability along the voicing continuum: Cross-language tests. *ICPhS,* V, Münster, 569-573.

Ahmad, F. 1996. *Kontrastive Linguistik Deutsch/Arabisch. Zur Relevanz der kontrastiven Untersuchungen*. Heidelberg: Groos.

Aichert, I., Marquardt, C. & Ziegler, W. 2005. Frequenzen sublexikalischer Einheiten des Deutschen: CELEX-basierte Datenbanken. *Neurolinguistik*, 19, 55-81.

Al Ani, S. 1970. *Arabic Phonology: An Acoustical and Physiological Investigation*. The Hague: Mouton.

Alghamdi, M. 1998. A spectrographic analysis of Arabic vowels: a cross-dialect study. *Journal King Saud University*, 10 (1), 3-24.

Allen, J. & Miller, J. 2001. Effects of syllable-initial voicing and speaking rate on the temporal characteristics of monosyllabic words. *JASA*, 106, 2031-2039.

Alotaibi, Y.A. & A. Husain 2009. Formant Based Analysis of Spoken Arabic Vowels. ---: Fierrez , J., Ortega-Garcia, J., Esposito, A., Drygajlo, A., Faundez-Zanuy, M. (eds.). *Biometric ID Management and Multimodal Communication* (BioID_MulitComm2009 Proceedings), Berlin Heidelberg: Springer, 162-169.

Andersen, H. 1989. Markedness theory: the first 150 years. In: Tomić, O. (ed.). *Markedness in Synchrony and Diachrony*, Berlin: Mouton de Gruyter, 11-46.

Andersen, H. 2001. Markedness and the theory of linguistic change. In: Andersen, H. (ed.). *Actualization: Linguistic Change in Progress.* Amsterdam: Benjamins, 21-57.

Anderson, S. 1981. Why phonology isn't 'natural'. *Linguistic Inquiry*, 12, 493-539.

Andersen R. 1983. Transfer to somewhere. In: Gass, S. & Selinker, L. (eds.). *Language Transfer in Language Learning*. Rowley, MA: Newbury House, 177-201.

Andruski, J. & Nearey, T. 1992. On the sufficiency of compound target specification of isolated vowels and vowels in /bVb/ syllables. *JASA*, 91, 390-410.

Aoyama, K. & Guion, S. 2007. Prosody in second language acquisition. Acoustic analyses of duration and F0 range. In: Bohn, O.-S. & Munro, M. (eds.). *Language Experience in Second Language Speech Learning. In Honor of James Emil Flege*. Amsterdam Philadelphia: John Benjamins, 281-297.

Aoyama, K., Flege, J., Guion, S., Akahane-Yamada, R., Yamada, T. 2004. Perceived phonetic dissimilarity and L2 speech learning: The case of Japanese /r/ and English /l/ and /r/. *JPhon*, 32, 233-250.

Archangeli, D. & Pulleyblank, D. 1994. *Grounded Phonology*. MIT Press.

Archibald, J. 1995. Phonological competence. In: Archibald, J. (ed.). *Phonological Acquisition and Phonological Theory*. Hillsdale, NJ: Lawrence Erlbaum, XIII-XXI.

Archibald, J. 1998. *Second Language Phonology.* Philadelphia: John Benjamins.

Arısoy, E., Arslan, L.M., Demiralp, M.N., Ekenel, H.K., Kelepir, M., Meral, H.M., Özsoy, A.S., Şayli, Ö., Türk, O., Yolcus, B.C. 2004. Duration of Turkish Vowels Revisited. *Proceedings of the12th International Conference on Turkish* (ICTL 2004, Izmir, Turkey).

[http://www.busim.ee.boun.edu.tr/~speech/publications/Turkish_phonetics/ICTL_2004_duration_revisited.pdf]

Arısoy, E., Arslan, L.M., Demiralp, M.N., Ekenel, H.K., Kelepir, M., Meral, H.M., Özsoy, A.S., Şayli, Ö., Türk, O., Yolcus, B.C. 2010. *Turkish_phonetics_sunum.ppt* (with MR images by M.A. Kiliç). [http://www.busim.ee.boun.edu.tr/~speech/Turkish_phonetics.html]

Aslin, R. & Pisoni, D. 1980. Some developmental processes in speech perception. In: Yeni-Komshian, G., Kavanaugh, J., Ferguson, C. (eds.). *Child Phonology, Vol. 2: Perception*. New York: Academic Press, 67-96.

Aslin, R., Werker, J., Morgan, J. 2002. Innate phonetic boundaries revisited. *JASA*, 112 (4), 1257-1260.

Avram, A. 1990. Sur les voyelles neutres en Roumain, en Albanais et dans les Langues Romanes Occidentales. *Revue Roumaine de Linguistique,* 35 (1), 19-27.

Baayen, R.H., Piepenbrock, R., & Gulikers, L. 1995. *The CELEX Lexical Database* (CD-ROM). Philadelphia: Linguistic Data Consortium, University of Pennsylvania, PA.

Backhaus, K., Erichson, B., Plinke, W., Weiber, R. 2011a. *Multivariate Analysemethoden*. Heidelberg et al.: Springer.

Backhaus, K., Erichson, B., Weiber, R. 2011b. *Fortgeschrittene Multivariate Analysemethoden*. Heidelberg et al.: Springer.

Baker, W., Trofimovich, P., Flege, J., Mack, M., Halter, R. 2008. Child-adult differences in second-language phonological learning: The role of cross-language similarity. *Language and Speech*, 51, 317-342.

Baker, W., Trofimovich, P., Mack, M., Flege, J. 2002. The effect of perceived phonetic similarity on non-native sound learning by children and adults. In: Skarabela, B. et al. (eds.). *BUCLD Proceedings*, 26. Sommerville, MA: Cascadilla Press, 36-47. [http://education.concordia.ca/~pavel/pdf/baker2002.pdf]

Barry, W. 1975. Untersuchungen zu fremdsprachlichen Perzeptionsaufgaben. *Arbeitsberichte des Instituts für Phonetik der Universität Kiel,* 4, 36-56.

Barton, D. 1978. *The Role of Perception in the Acquisition of Phonology*. Bloomington, IN: Indiana University Linguistics Club.

Bassetti, B. 2009. Orthographic Input and Second Language Phonology. In: Piske, Th. & Young-Scholten, M. (eds.). *Input Matters in SLA*. Bristol et al.: Multilingual Matters, 191-206.

Battistella, E. 1990. *Markedness: The Evaluative Superstructure of Language.* New York: State University of New York.

Battistella, E. 1996. *The Logic of Markedness.* New York: Oxford University Press.

Beci, B. 1995. *Të folmet veriperëndimore të Shqipërisë dhe sistemimi fonetik i së folmes së Shkodrës*. Tiranë: Akad. e Shkencave e RSH, Inst. i Gjuhësisë dhe i Letërsisë.

Becker, T. 1998. *Das Vokalsystem der deutschen Standardsprache*. Frankfurt am Main: Peter Lang.

Becker-Kristal, R. 2010. *Acoustic Typology of Vowel Inventories and Dispersion Theory: Insights from a large Cross-Linguistic Corpus.* Dissertation. Los Angeles: University of California. http://www.linguistics.ucla.edu/faciliti/research/RBecker_diss.pdf

Beckman, M.E., Jung, T.-P., Lee, S.-H., de Jong, K., Krishnamurthy, A.K., Ahalt, S.C., Cohen, K.B., Collins, M.J. 1995. Variability in the production of quantal vowels revisited. *JASA*, 97 (1), 471-490.

Beddor, P. & Gottfried, T. 1995. Methodological issues in cross-language speech perception research with adults. In: Strange, W. (ed.). *Speech Perception and Linguistic Experience*. Timonium, MD: York Press, 207-232.

Belkaid, Y. 1984. Les voyelles de l'arabe littéraire moderne. Analyse spectographique. *Travaux de l'Institut de Phonétique de Strasbourg, Rapport, 16*, 217-240.

Bennett, D. 1968. Spectral form and duration as cues in the recognition of English and German vowels. *Language and Speech*, 11 (2), 65-85.

Bent, T., Bradlow, A., Smith, B. 2007. Segmental errors by position-in-word and intelligibility. In: Bohn, O.-S. & Munro, M. (eds.). *Language Experience in Second Language Speech Learning. In Honor of James Emil Flege*. Amsterdam Philadelphia: John Benjamins, 331-347.

Beňuš, Š., Gafos, A., Goldstein, L. 2004. Phonetics and phonology of transparent vowel in Hungarian. In: Nowak, P., Yoquelet, C., Mortensen, D. (eds.) *Proceedings of the 29th Annual Meeting of the Berkeley Linguistic Society*, 486-497. [http://www1.cs.columbia.edu/~sbenus/Research/BLS-03-final.pdf]

Bernard, J. & Mannell, R. 1986. A study of /h_d/ words in Australian English. *Working Papers of the Speech, Hearing and Language Research Centre*. Macquarie University, 1-106.

Bernard, J. 1970. Toward the acoustic specification of Australian English. *Zeitschrift für Phonetik, Sprachwissenschaft und Kommunikationsforschung*, 23 (2/3), 113-128.

Best, C. 1994. The emergence of native-language phonological influences in infants: A perceptual assimilation model. In: Goodman, D. & Nusbaum, H.C. (eds.). *The Development of Speech Perception: The Transition from Speech Sounds to spoken Words*. Cambridge, MA: MIT Press, 167-224.

Best, C. 1995. A direct realist perspective on cross-language speech perception. In: Strange, W. (ed.). *Speech Perception and Linguistic Experience: Issues in Cross-Language Research*. Timonium, MD: York Press, 171-204. [http://www.haskins.yale.edu/Reprints/HL0996.pdf]

Best, C., Faber, A., Levitt, A. 1996. Assimilation of non-native vowel contrasts to the American English vowel system. *JASA*, 99, 2602.

Best, C., Halle, P., Bohn, O.-S., Faber, A. 2003. Cross-language perception of non-native vowels: Phonological and phonetic effects of listeners' native languages. *ICPhS*, XV, Barcelona, 2889-2892.

Best, C. & McRoberts, G. 2003. Infant perception of non-native consonant contrasts that adults assimilate in different ways. *Language and Speech*, 46, 183-216.

Best, C., McRoberts, G., Goodel, E. 2001. Discrimination of non-native consonant contrasts varying in perceptual assimilation to the listener's native phonological system. *JASA*, 109 (2), 775-794.

Best, C., McRoberts, G., Sithole, N. 1988. Examination of the perceptual re-organization for speech contrasts: Zulu click discrimination by English-speaking adults and infants. *Journal of Experimental Psychology: Human Perception and Performance*, 14, 345-360.

Best, C. & Strange, W.1992. Effects of phonological and phonetic factors on cross-language perception of approximants. *JPhon*, 20, 305-330.

Best, C. & Tyler, M. 2007. Nonnative and second-language speech perception: Commonalities and complementarities. In: Bohn, O.-S. & Munro, M. (eds.). *Language Experience in Second Language Speech Learning. In Honor of James Emil Flege*. Amsterdam Philadelphia: John Benjamins, 13-34.

Bethin, C. 1992. *Polish Syllables. The Role of Prosody in Phonology and Morphology*. Columbus: Slavica Publishers.

Bialystok, E. 2001. *Bilingualism in Development: Language, Literacy, and Cognition*. New York, NY: Cambridge University Press.

Biedrzycki, L. 1974. *Abriß der polnischen Phonetik*. Wiedza Powszechna: Warszawa. (translated by Gontarczyk, S.) Leipzig: VEB Verlag Enzyklopädie.

Bijankhan, M. 2000. Farsi vowel compensatory lengthening: an experimental approach. In: *Proceedings of the 6th International Conference of Spoken Language Processing* (INTERSPEECH 2000), vol. 2, 507-510. [http://www.isca-speech.org/archive/icslp_2000/i00_2507.html]

Birdsong, D. 1999. *Second Language Acquisition and the Critical Period Hypothesis*. Mahwah Hillsdale: Erlbaum.

Birdsong, D. 2005. Nativelikeness and non-nativelikeness in L2 research. *International Review of Applied Linguistics in Language Teaching*, 43, 319-328.

Bladon, R. & Lindblom, B. 1981. Modelling the judgmenent of vowel qualitiy differences. *JASA*, 69, 1414-1422.

Blevins, J. 2004. *Evolutionary Phonology: The Emergence of Sound Patterns*. Cambridge: Cambridge University Press.

Bley-Vroman, R. 1991. The logical problem of foreign language learning. *Linguistic Analysis*, 20 (1-2), 3-49.

Bley-Vroman, R. 2009. The evolving context of the Fundamental Difference Hypothesis. *Studies in Second Language Acquisition*, 31, 175-198.

Boberg, C. 2008. Regional phonetic differentiation in Standard Canadian English. *Journal of English Linguistics*, 36 (2), 129-154.

Boë, L.-J., Perrier, P., Bailly, G. 1992. The geometric vocal tract variables controlled for vowel production: proposals for constraining acoustic-to-articulatory inversion. *JPhon*, 20 (1), 27-38.

Boersma, P. 1993. Accurate Short-Term Analysis of the Fundamental Frequency and the Harmonics-To-Noise Ratio of a Sampled Sound. *Proceedings of the Institute of Phonetic Sciences Amsterdam (IFA),* 17, 97-110.

Boersma, P. 1998. *Functional Phonology: Formalizing the Interactions between Articulatory and Perceptual Drives*. PhD Dissertation. University of Amsterdam. LOT International Series 11. The Hague: Holland Academic Graphics. [http://www.fon.hum.uva.nl/paul/diss/]

Boersma, P. 2000. Learning a grammar in functional phonology. In: Dekkers, J., van der Leeuw, F., van de Weijer, J. (eds.). *Optimality Theory: Phonology, Syntax, and Acquisition*. Oxford: University Press, 465-523.

Bogacka, A. 2004. On the perception of English high vowels by Polish learners of English. In: Dakalaki, E., Katsos, N., Mavrogiorgos, M., Reeve, M. (eds.). *CamLing: Proceedings of the University of Cambridge Postgraduate Conference on Language Research*. Cambridge: Cambridge Institute for Language Research, 43-50.

Bogacka, A. 2007. *Repopulating Vowel Space: English Diphthong Production by Polish Learners of English*. PhD Dissertation. Adam Mickiewicz University, Poznan.

Bogacka, A. 2009. Why can Poles perceive Sprite but not Coca Cola? A Natural Phonology account. In: Boersma, P. & Hamann, S. (eds.). *Phonology in Perception*. Berlin: de Gruyter, 25-54.

Bohn, O.-S. 1994. Hinweise auf ein sprachübergreifendes allgemeines Prinzip bei der Wahrnehmung fremdsprachlicher Vokalkontraste. In: Ramers, K.H., Vater, H., Wode, H. (eds.). *Universale phonologische Strukturen und Prozesse*. (Linguistische Arbeiten 310). Tübingen: Niemeyer, 189-204.

Bohn, O.-S. 1995. Cross-language speech perception in adults. In: Strange, W. (ed.). *Speech Perception and Linguistic Experience*. Timonium, MD: York Press, 275-300.

Bohn, O.-S. 1998. Wahrnehmung fremdsprachlicher Laute: Wo ist das Problem? In: Wegener, H. (ed.). *Eine zweite Sprache lernen: Empirische Untersuchungen zum Zweitspracherwerb*. Tübingen: Narr, 1-20.

Bohn, O.-S. 2002. On phonetic similarity. In: Burmeister, P., Piske, T., Rohde, A. (eds.). *An Integrated View of Language Development*. Trier: Wissenschaftlicher Verlag, 191-216.

Bohn, O.-S. & Flege, J. 1990. Interlingual identification and the role of foreign language experience in L2 vowel perception. *Applied Psycholinguistics,* 11, 303-328.

Bohn, O.-S. & Flege, J. 1992. The production of new and similar vowels by adult German learners of English. *Studies in Second Language Acquisition,* 14, 131-158.

Bohn, O.-S., Flege, J., Dagenais, P.A., Fletcher, S.G. 1992. Differenzierung und Variabilität der Zungenpositionen bei der Artikulation deutscher Vokale. *Zeitschrift für Dialektologie und Linguistik*, Beihefte 72, 1-26.

Bohn, O.-S. & Polka, L. 2001. Target spectral, dynamic spectral, and temporal cues in infant perception of German vowels. *JASA*, 110, 504-515.

Bohn, O.-S. & Polka, L. 2009. Accounting for the accented perception of vowels: Universal preferences and language-specific biases. *JASA*, 125, 2758.

Bohn, O.-S. & Steinlen, A. K. 2003. Consonantal context affects cross-language perception of vowels. *ICPhS*, XV, Barcelona, 2289-2292.

Bolinger, D. 1998. Intonation in American English. In: Hirst, D. & Di Cristo, A. (eds.). *Intonation Systems. A Survey of twenty Languages*. Cambridge: Cambridge University Press, 45-55.

Bolla, K. 1995. *Magyar fonetikai atlasz. A szegmentális hangszerkezet elemei*. Budapest: Nemzeti Tankönyvkiadó.

Bongaerts T., Planken, B., Schils, E. 1995. Can late starters attain a native accent in foreign language? In: Singleton, D. & Lengyel, Z. (eds.). *The Age Factor in Second Language Acquisition.* Clevedon: Multilingual Matters, 30-50.

Bongaerts, T. 1999. Ultimate attainment in L2 pronunciation: The case of very advanced late L2 learners. In: Birdsong, D. (ed.). *Second Language Acquisition and the Critical Period Hypothesis*. Mahwah, NJ: Erlbaum, 135-159.

Bongaerts, T., van Summeren, C., Planken, B., Schils, E. 1997. Age and ultimate attainment in the pronunciation of a foreign language. *Studies in Second Language Acquisition*, 19, 447-465.

Booij, G. & Rubach, J. 1984. Morphological and prosodic domains in Lexical Phonology. *Phonology Yearbook 1*, 1-27.]

Boomershine, A., Currie Hall, K., Hume, E. Johnson, K. 2008. The impact of allophony vs. contrast on speech perception. In: Avery, P., Dresher. E., Rice, K. (eds.). *Contrast in Phonology: Perception and Acquisition*. New York: Mouton de Gruyter, 143-172.

Boyce, S.E. 1990. Coarticulatory organization for lip rounding in Turkish and English. *JASA*, 88 (6), 2584-2595.

Bradlow, A. & Bent, T. 2003. Listener adaptation to foreign accented English. *ICPhS*, XV, Barcelona, 2881-2884.

Bradlow, A., Clopper, C., Smiljanic, R. 2007. A Perceptual Similarity Space for Languages. *ICPhS,* XVI, Saarbrücken, 1373-1376.

Bradlow, A. & Pisoni, D. 1999. Recognition of spoken words by native and non-native listeners: Talker-, listener- and item-related factors. *JASA*, 106 (4), 2074-2085.

Brandstätter, J., Kaseß, C., Moosmüller, S. (to appear). Quality and Quantity in high vowels in Standard Austrian German, in: Leeman, A., Kolly, M.-J., Dellwo, V., Schmid, S. (eds.), *Trends in Phonetics and Phonology in German speaking Europe* (working title).

Brandstätter, J., Moosmüller, S. (in press). *Neutralisierung der hohen ungerundeten Vokale in der Wiener Standardsprache*, in: Glauninger, M. & Lenz, A. (eds.). *Standarddeutsch in Österreich – Theoretische und empirische Ansätze* (Wiener Arbeiten zur Linguistik). Wien: Vandenhoeck & Ruprecht.

Brewer, M. 2000. Research Design and Issues of Validity. In: Reis, H. & Judd,C. (eds.). *Handbook of Research Methods in Social and Personality Psychology*. Cambridge: Cambridge University Press, 3-16.

Broselow, E. 1983. Nonobvious transfer: On predicting epenthesis errors. In: Selinker, L. & Gass, S. (eds.). *Language Transfer in Language Learning*. Rowley, MA: Newbury House, 269-280. Reprinted in: Ioup, G. & Weinberger, S. (eds.). 1987. *Interlanguage Phonology*. Cambridge, MA: Newbury House, 292-304.

Broselow, E., Chen, S., Wang, C. 1998. The emergence of the unmarked in second language acquisition. *Studies in Second Language Acquisition*, 20, 261-280.

Browman, C. 1978. Tip of the tongue and slip of the ear. *UCLA Working Papers in Phonetics*, 42. University of California, Los Angeles.

Browman, C. 1980. Perceptual processing: Evidence from slips of the ear. In: Fromkin, V. (ed.). *Errors in Linguistic Performance. Slips of the Tongue, Ear, Pen, and Hand*. New York: Academic Press, 213-230.

Browman, C. & Goldstein, L. 1989. Articulatory gestures as phonological units. *Phonology*, 6, 201-251.

Browman, C. & Goldstein, L. 1992. Articulatory phonology: An overview. *Phonetica*, 49, 155-180.

Brown, C. & Matthews, J. 1997. The role of feature geometry in the development of phonemic contrasts. In: Hannahs, S.J. & Young-Scholten, M. (eds.). *Generative Studies in the Acquisition of Phonology*. Amsterdam: John Benjamins, 67-112.

Brown, C. 1998. The role of the L1 grammar in the L2 acquisition of segmental structure. *Second Language Research,* 14, 136-193.

Brown, C. 2000. The interrelation between speech perception and phonological acquisition from infant to adult. In: Archibald, J. (ed.). *Second Language Acquisition and Linguistic Theory*. Malden, MA: Blackwell, 4-63.

Brown, W. & Alt, Th. 2004. *A Handbook of Bosnian, Serbian and Croatian*. Slavic and Eurasian Language Resource Center SEELRC. [http://seelrc.org:8080/grammar/pdf/compgrammar_bcs.pdf]

Brunswik, E. 1956. *Perception and the Representative Design of Psychological Experiments*, 2nd edition. Berkeley, CA: University of California Press.

Bühl, A. & Zöfel, P. 2002. *Erweiterte Datenanalyse mit SPSS. Statistik und Data Mining*. Wiesbaden: Westdeutscher Verlag.

Bürki-Cohen J., Miller J., Eimas, P. 2001. Perceiving non-native speech. *Language and Speech*, 44, 149-169.

Burnham, D. & Mattock, K. 2007. The perception of tones and phones. In: Bohn, O.-S. & Munro, M. (eds.). *Language Experience in Second Language Speech Learning. In Honor of James Emil Flege*. Amsterdam Philadelphia: John Benjamins, 259-280.

Burnham, D. 1986. Developmental loss of speech perception: Exposure to and experience with a first language. *Applied Psycholinguistics*, 7, 207-239.

Burnham, D. 2003. Language specific speech perception and the onset of reading. *Reading and Writing*, 16 (6), 573-609.

Butcher, A. 1976. *The Influence of the Native Language on the Perception of Vowel Quality*. Arbeitsberichte 6, Institut für Phonetik der Universität Kiel. [http://www.linguistik.uni-kiel.de/arbeitsberichte/aipuk/Aipuk6.pdf]

Bybee, J. 1999. Usage based phonology. In Darnell, M., Moravcsik, E., Newmeyer, F., Noonan, M. & Wheatley, K. (eds.) *Functionalism and formalism in linguistics. Volume I: General papers.* Amsterdam: John Benjamins. 211-242. [https://www.unm.edu/~jbybee/downloads/Bybee1999UsageBasedPhonology.pdf]

Bybee, J. 2001. *Phonology and Language Use*. New York: Cambridge University Press.

Bybee. J. 2007. *Frequency of Use and the Organization of Language*. New York: Oxford University Press.

Bybee, J. 2010. *Language, Usage and Cognition*. New York: Cambridge University Press.
Cabrelli Amaro, J . 2012. L3 phonology. An understudied domain. In: Cabrelli Amaro, J., Flynn, S. & Rothman J. (eds.). *Third Language Acquisition in Adulthood.* Amsterdam Philadelphia: John Benjamins. 33-60.
Cardoso, W. & Trofimovich, P. 2014. *Proceedings of the International Symposium on the Acquisition of Second Language Speech* (New Sounds 2013), *Concordia Working Papers in Applied Linguistics*, 5. [http://doe.concordia.ca/copal/volumes].
Carlson, R., Fant, G., Granström, B. 1975. Two-formant models, pitch, and vowel perception. In: Fant, G. & Tatham, M. (eds.). *Auditory Analysis and Perception of Speech*. New York: Academic, 55-82.
Carlson, R., Granström, B., Fant, G. 1970. Some studies concerning perception of isolated vowels. *Speech Transmission Laboratory: Q. Prof. Stat. Report*, 2-3, 19-35.
Carré, R. 2004. From an acoustic tube to speech production. *Speech Communication*, 42, 227-240.
Cauneau, I. 1992. *Hören, Brummen, Sprechen. Angewandte Phonetik im Unterrichtsfach Deutsch als Fremdsprache*. Handbuch. München: Klett.
Ćavar, M., Rydzewski, P., Oštarić, A. 2012. *Cue Weighting in the Perception of English Tenseness Contrast by Polish and Croatian Subjects*. Paper read at the Conference on Second Language Acquisition of Phonology, University of York, UK, July 2012. [http://www-users.york.ac.uk/~sh581/ORAL%208%20Cavar%20Rydzewski%20&%20Ostaric.pdf]
Chang, S., Plauché, M., Ohala, J. 2001. Markedness and consonant confusion asymmetries. In: Hume, E. & Johnson, K. (eds.). *The Role of Speech Perception in Phonology*. San Diego: Academic Press, 79-101.
Chistovich, L.A. 1985. Central auditory processing of peripheral vowel spectra. *JASA*, 77, 789-805.
Chistovich, L. & Lublinskaya, V. 1979. The "center of gravity" effect in vowel spectra and critical distance between formants: Psychoacoustical study of the perception of vowel-like stimuli. *Hearing Research*, 1, 185-195.
Chistovich, L.A., Sheikin, R., Lublinskaja, V. 1979. 'Centers of gravity' and spectral peaks as the determinants of vowel quality. In: Lindblom, B. & Ohman, S. (eds.). *Frontiers of Speech Communication Research*. London: Academic, 143-157.
Chitoran, I. 2002a. *The Phonology of Romanian: A Constraint-based Approach*. Berlin New York: Mouton de Gruyter.
Chitoran, I. 2002b. The phonology and morphology of Romanian diphthongization. *Probus*, 14, 205-246.
Chitoran, I. 2002c. A perception-production study of Romanian diphthongs and glide-vowel sequences. *Journal of the International Phonetic Association,* 32 (2), 203-222.
Chomsky, N. 1957. *Syntactic Structures*. The Hague: Mouton.
Chomksy, N. 1959. A Review of B. F. Skinner's Verbal Behavior. *Language*, 35 (1): 26-57
Chomsky, N. 1965. *Aspects of the Theory of Syntax.* Cambridge, MA: MIT Press.
Chomsky, N. 1975. *Reflections on language*. New York: Pantheon Books.
Chomsky, N. 1981. *Lectures on Government and Binding: The Pisa Lectures*. Holland: Foris Publications.
Chomsky, N. & Halle, M. 1968. *The Sound Pattern of English*. New York: Harper and Row.
Clements, G.N. & Sezer, E. 1982. Vowel and consonant disharmony in Turkish. In: van der Hulst, H. & Smith, N. (eds.). *The Structure of Phonological Representations. Part II.* Dordrecht: Foris Publications, 213-255.
Clements, G.N. 2006. *Does sonority have a phonetic basis?* Comments on the chapter by Bert Vaux, Ms. [http://ed268.univ-paris3.fr/lpp/publications/2006_Clements_Does_Sonority.pdf]
Clermont, F. 1996. Multi-speaker formant data on the Australian English vowels. A tribute to J.R.L. Bernard's Pioneering Research. *Proceedings of the 6th Australian International Conference on Speech Science and Technology*, Adelaide, 145-150. [http://www.assta.org/sst/SST-96/cache/SST-96-Chapter7-p3.pdf]
Clopper, C., Pisoni, D., de Jong, K. 2005. Acoustic characteristics of the vowel systems of six regional varieties of American English. *JASA*, 118 (3), 1661-1676.
Close, E. 1974. *The development of Modern Romanian: Linguistic theory and practice in Muntenia 1821-1838.* (Oxford Modern Languages and Literatures Monograph). Oxford University Press.
Collins, B. & Mees, I.M. 2008. *Practical Phonetics and Phonology. A Resource Book for Students*, 2nd edition (Original work published 2003). London: Routledge.
Connell, B. 2002. Tone languages and the universality of intrinsic F0: evidence from Africa. *JPhon*, 30, 101-129.
Cook, V. 1988. *Chomsky's Universal Grammar. An Introduction*. Oxford, New York: Basil Blackwell.

Cook, V. 1999. Going beyond the native speaker in language teaching. *TESOL Quarterly*, 33 (2), 185-209. [http://static.aston.ac.uk/lis/tesol/Quarterly-1967-2003/Vol_33_2.pdf#page=10]

Cook, V. (ed.). 2003. *Effects of the L2 on the L1*. Clevedon: Multilingual Matters.

Corbett, G. & Browne, W. 2011. Serbo-Croat. In: Comrie, B. (ed.). *The World's Major Languages*. London New York: Routledge, 330-346.

Coteanu, I. 1981. Structura şi evoluţia limbii române (de la origini pînă la 1860), Bukarest: Ed. Academiei.

Council of Europe 2001. *Common European Framework of Reference for Languages: Learning, Teaching, Assessment (CEFR).* Strassburg: Council of Europe.

Cox, F. 1999. Vowel change in Australian English. *Phonetica*, 56, 1-27.

Cox, F. & Palethorpe, S. 2001. The Changing Face of Australian English Vowels. In: Blair, D. & Collins, P. (eds.). *Varieties of English around the World: English in Australia.* Amsterdam: John Benjamins, 17-44.

Croft, W. 2003. *Typology and Universals*, 2nd edition. Cambridge: Cambridge University Press.

Crosswhite, K. 2004. Vowel reduction. In: Hayes, B., Kirchner, R., Steriade, D. (eds.). *Phonetically based Phonology*. Cambridge: Cambridge University Press, 191-231.

Crothers, J. 1978. Typology and universals of vowel systems. In: Greenberg, J. (ed.). *Universals of Human Language, Vol. 2: Phonology*. Stanford: Stanford University Press, 93-152.

Cumming, R. 2011. The effect of dynamic fundamental frequency on the perception of duration. *JPhon*, 39 (3), 375-387.

Cutler, A. 2005. Lexical Stress. In: Pisoni, D. & Remez, R. (eds.). *The Handbook of Speech Perception*. Malden, MA: Blackwell, 264-289.

Cutler, A. & Weber, A. 2007. Listening experience and phonetic-to-lexical mapping in L2. *ICPhS,* XVI, Saarbrücken, 43-48.

Damper, R. & Harnad, S. 2000. Neural network models of categorical perception. *Perception & Psychophysics*, 62 (4), 843-867.

Darcy, I. & Krüger, F. 2012. Vowel perception and production in Turkish children acquiring L2 German. *JPhon*, 40, 568-581.

Davies, A. 2003. *The Native Speaker. Myth and Reality*. Clevedon: Multilingual Matters.

de Boer, B. 2000. Self-organization of vowel systems. *JPhon*, 28 (4), 441-465.

de Boer, B. 2001. *The Origins of Vowel Systems*. Oxford: Oxford University Press.

de Bot, K. & Larsen-Freeman, D. 2011. Researching second language development from a Dynamic Systems Theory perspective. In: Verspoor, M., de Bot, K. & Lowie, W. (eds). *A Dynamic Approach to Second Language Development. Methods and techniques*. Amsterdam Philadelphia: John Benjamins, 5-23.

Dehaene-Lambertz, G. 1997. Electrophysical correlates of categorical phoneme perception in adults. *NeuroReport*, 8, 919-924.

Delattre, P., Liberman, A., Cooper, F., Gerstman, L. 1952. An experimental study of the acoustic determinants of vowel colour: Observations on one- and two-formant vowels synthesized from spectrographic patterns. *Word*, 8, 195-210.

Deterding, David. 1997. The formants of monophthong vowels in Standard Southern British English pronunciation. *Journal of the International Phonetic Association,* 27, 47-55.

Diehl, R. 1989. Remarks on Stevens' quantal theory of speech. *JPhon*, 17, 71-78.

Diehl, R. 1991. The role of phonetics within the study of language. *Phonetica*, 48, 120-134.

Diehl, R. 2007. Acoustic and auditory phonetics: the adaptive design of speech sound systems. *Philosophical Transactions of the Royal Society*, B, vol. 363, no. 1493, 965-978. [http://rstb.royalsocietypublishing.org/content/363/1493/965.full]

Diehl, R. 2011. On the Robustness of Speech Perception. *ICPhS,* XVII, Hong Kong, 1-8.

Diehl, R. & Kluender, K.R. 1989. On the objects of speech perception. *Ecological Psychology*, 1, 121-144.

Diehl, R., Lindblom, B., Creeger, C. 2003. Increasing realism of auditory representations yields further insights into vowel phonetics. *ICPhS,* XV, Barcelona, vol. 2, 1381-1384. Adelaide: Causal Publications.

Disner, S. F. 1983. Vowel quality: the relation between universal and language specific factors. Dissertation. UCLA, Los Angeles, CA.*UCLA Working Papers in Phonetics,* 58, 1-158.

Disner, S.F.1984. Insights on vowel spacing. In: Maddieson, I. (ed.) *Patterns of sounds*. Cambridge: Cambridge University Press, 136-155.

Dodi, A. & Gjinari, J. 1983. *Fonetika*, *Vol. 2. Fonetika dhe Gramatika e gjuhës së sotme letrare Shqipe*. Akademia e RPS të Shqipërisë.

Dodi, A. 2004. *Fonetika dhe fonologia e gjuhës shqipe*. Tiranë: Akademia e Shkencave e Shqiperisë, Instituti i Gjuhësisë dhe i Letërsisë.

Dogil, G. & Reiterer, S. 2009. *Language Talent and Brain Activity*. Berlin New York: de Gruyter.

Donegan, P. 1978. *On the Natural Phonology of Vowels*. Dissertation. Ohio State University.

Donegan, P. 1985. How learnable is phonology? In: Dressler, W. & Tonelli, L. (eds.). *Natural Phonology from Eisenstadt*. Padova: CLESP, 19-31.

Donegan, P. 1993. On the phonetic basis of phonological change. In: Jones, Ch. (ed.). *Historical Linguistics: Problems and Perspectives*. London: Longman, 98-130.

Donegan, P. 1995. The innateness of phonemic perception. In: Samiian, V. (ed.). *Proceedings of the 24th Western Conference on Linguistics (WECOL 7*). Fresno, CA: Western Conference on Linguistics, 59-69.

Donegan, P. 2001. Constraints and processes in phonological perception. In: Dziubalska-Kołaczyk, K. (ed.). *Constraints and Preferences*, Berlin New York: Mouton de Gruyter, 43-68.

Donegan, P. 2002a. Normal vowel development. In: Ball, M.J. & Gibbon, F. (eds.). *Vowel Disorders*. Boston: Butterworth-Heinemann, 1-36.

Donegan, P. 2002b. Phonological processes and phonetic rules. In: Dziubalska-Kołaczyk, K. & Weckwerth, J. (eds.), *Future Challenges for Natural Linguistics*. München: Lincom Europa, 57-82.

Donegan, P. & Stampe, D. 1978. The syllable in phonological and prosodic structure. In: Bell, A. & Bybee Hooper, J. (eds.). *Syllables and Segments*. Amsterdam: North-Holland, 25-34.

Donegan, P. & Stampe, D. 1979. The study of natural phonology. In: Dinnsen, D. (ed.). *Current Approaches to Phonological Theory*. Bloomington: Indiana University Press, 126-173.

Donegan, P. & Nathan, G. (forthcoming). Natural Phonology and Sound Change. In: Honeybone, P. & Salmons, J. (ed.). *The Handbook of Historical Phonology*. Oxford: Oxford University Press. [http://www.ling.hawaii.edu/faculty/donegan/Papers/201Xhistphon.pdf]

Dörnyei, Z. & Skehan, P. 2003. Individual differences in second language learning. In: Doughty, C. & Long, M.H. (eds.). *The Handbook of Second Language Acquisition*. Malden, MA: Blackwell, 589-630.

Dörnyei, Z. 2005. *The Psychology of the Language Learner*. Mahwah, NJ: Lawrence Erlbaum.

Dörnyei, Z. 2008. *The Psychology of Second Language Acquisition*. Oxford: Oxford University Press.

Dow, F. 1972. *An Introduction to the Pronunciation of Chinese*. Canberra: Australian National University Press.

Dragusanu, S. 2012. Vowels in contact: A case study in Romanian. *Working Papers of the Linguistic Circle of the University of Victoria*, 22 (1), 41-53. [http://journals.uvic.ca/index.php/WPLC/article/view/10848/3198]

Dressler, W. U. 1984. Explaining Natural Phonology. *Phonology Yearbook*, 1, 29-51.

Dressler, W. U. 1985. *Morphonology*. Ann Arbor: Karoma Press.

Dressler, W. U. 1989. Markedness and naturalness in phonology; the case of natural phonology. In: Tomić, O. (ed.). *Markedness in Synchrony and Diachrony*. Berlin New York: Mouton de Gruyter, 111-120.

Dressler, W.U. 1990. The cognitive perspective of "naturalist" linguistic models. *Cognitive Linguistics* 1 (1), 75-98.

Dressler, W. U. 1995. Wealth and poverty of functional analyses, with special consideration for functional deficiencies. In: Miller, Sh. & Mey, J. (eds.). *Form and Function in Language. Proceedings of the 1st Rasmus Rask Colloquium*. Odense: Odense University Press, 11-39.

Dressler, W. U. 1996. Principles of naturalness in phonology and across components. In: Hurch, B. & Rhodes, R. (eds.). *Natural Phonology: The State of the Art*. Berlin: Mouton de Gruyter, 41-52.

Dressler, W. U. 1999a. On a semiotic theory of preferences in language. In: Haley, M. & Shapiro, M. (eds.). *The Peirce Seminar Papers. Essays in Semiotic Analysis. Proceedings of the International Colloquium on Language and Peircean Sign Theory 1997*, vol. 4. New York: Berghahn Books, 389-415.

Dressler, W. U. 1999b. What is Natural in Natural Morphology (NM)? *Prague Linguistic Circle Papers,* 3, 135-144.

Dressler, W. U. 2002. Naturalness and functionalism. In: Dziubalska- Kołaczyk, K. & Weckwerth, J. (eds.). *Future Challenges for Natural Linguistics*. München: Lincom Europa 2002, 83-101.

Dressler, W. U. & Dziubalska-Kołaczyk, K.1994. Functional analysis in the study of second language acquisition. *Functions of Language*, 1 (2), 201-228.

Dressler, W. U. & Siptár, P. 1989. Towards a natural phonology of Hungarian. *Acta Linguistica Hungarica*, 39, 29-51.

Duanmu, S. 2000. *The Phonology of Standard Chinese*. Oxford: Oxford University Press.

Duanmu, S. 2006. Chinese (Mandarin): phonology. In: Brown, K. (ed.). *Encyclopedia of Language and Linguistics*, 2nd edition. Oxford, UK: Elsevier Publishing House, 351-355. Ms. [http://www-personal.umich.edu/~duanmu/ELL05.pdf]

Dulay, H., Burt, M., Krashen, S. 1982. *Language Two*. New York: Oxford University Press.

Dziubalska- Kołaczyk, K. (ed.) 2001. *Constraints and Preferences, Trends in Linguistics*. Berlin: Mouton de Gruyter, 43-68.

Dziubalska- Kołaczyk, K., Wrembel, M., Kul, M. (eds.) 2010. *New Sounds 2010. Proceedings of the 6th International Symposium on the Acquisition of Second Language Speech.* Poznan: Adam Mickiewicz University 2010.

Eckman, F. 1977 [1987]. Markedness and the contrastive analysis hypothesis. *Language Learning*, 27, 315-330. Republished in: Ioup, G. & Weinberger, S. (eds.). 1987. *Interlanguage Phonology. The Acquisition of a Second Language Sound System*. Cambridge: Newbury House, 53-69.

Eckman, F. 1991. The Structural Conformity Hypothesis and the acquisition of consonant clusters in the interlanguage of ESL learners. *Studies in Second Language Acquisition,* 13, 23-41.

Eckman, F. 2004. From phonemic differences to constraint rankings. Research on Second Language Phonology. *Studies in Second Language Acquisition*, 26 (4), 513-549.

Eckman, F. 2008. Typological markedness and second language phonology. In: Hansen Edwards, J. & Zampini, M. (eds.). *Phonology and Second Language Acquisition*. Philadelphia: John Benjamins, 95-115.

Eckman, F. & Iverson, G. 2000. Contrast and hypercontrast in interlanguage. *Studies in Phonetics, Phonology and Morphology*, 6 (2), 213-247.

Eckman, F., Iverson, G., Yung Song, J. 2013. The role of hypercorrection in the acquisition of L2 phonemic contrasts. *Second Language Research*, 29 (3), 257-283.

Eimas, P. 1975. Auditory and phonetic coding of the cues for speech discrimination of the /r-l/ distinction by young infants. *Perception & Psychophysics*, 18, 341-347.

Eimas, P. & Miller, J. (eds.) 1981. *Perspectives on the Study of Speech*, Hillsdale: Lawrence Erlbaum.

Eimas, P., Miller, J., Jusczyk, P. 1987. On infant speech perception and the acquisition of language. In: Harnad, S. (ed.). *Categorical Perception*. New York: Cambridge University Press.

Eimas, P., Siqueland, E., Jusczyk, P., Vigorito, J. 1971. Speech perception in infants. *Science*, 171, 303-306.

Elliot, A.R. 1995a. Field independence/dependence, hemispheric specialization, and attitude in relation to pronunciation accuracy in Spanish as a foreign language. *Modern Language Journal*, 79 (3), 356-371.

Elliot, A.R. 1995b. Foreign language phonology: field independence, attitude, and the success of formal instruction in Spanish pronunciation. *Modern Language Journal*, 79 (3), 530–542.

Ellis, N. 2002. Frequency effects in language processing: A review with implications for theories of implicit and explicit language acquisition. *Studies in Second Language Acquisition,* 24, 143-188.

Ellis, R. 1986. *Understanding second language acquisition*. Oxford: Oxford University Press.

Ellis, R. 1994. *The Study of Second Language Acquisition*. Oxford: Oxford University Press.

Endres, W. & Großmann, E. 1974. Manipulation of the time functions of vowels for reducing the number of elements needed for speech synthesis. In: Fant, G. (ed.). *Speech Communication. Proceedings of the Speech Communication Seminar*. Stockholm: Almqvist & Wiksell, 267-275.

Engel, U., Isbăşescu M., Stănescu, S. Nicolae O. 1993. *Kontrastive Grammatik Deutsch-Rumänisch*. Heidelberg: Julius Groos.

Engel, U. & Mrazovic, P. (eds.) 1986. *Kontrastive Grammatik Deutsch-Serbokroatisch*, 2 Bde. München: iudicium.

Epstein, S., Flynn, S., Martohardjono, G. 1996. Second language acquisition: Theoretical and experimental issues in contemporary research. *Behavioural and Brain Sciences*, 19, 677-714.

Escudero, P. 2002. *Learning the Sounds of a New Language. Adults can do it too*! Paper read at the First Pan-American/Iberian Meeting on Acoustics, Cancun, Mexico. [http://www.fon.hum.uva.nl/paola/#Papers]

Escudero, P. 2005. *Linguistic Perception and Second Language Acquisition. Explaining the Attainment of Optimal Phonological Categorization*, Utrecht: LOT. [http://www.fon.hum.uva.nl/paola/escudero_PhDmanuscript_final.pdf]

Escudero, P. & Boersma, P. 2001. Modelling the perceptual development of phonological contrasts with Optimality Theory and the Gradual Learning Algorithm. Proceedings of the 25th Penn Linguistics Colloquium (Penn Working Papers in Linguistics), *Rutgers Optimality Archive,* 439. [http://www.fon.hum.uva.nl/paul/papers/ot_cue_integration.pdf]

Escudero, P. & Boersma, P. 2004. Bridging the gap between L2 speech perception research and phonological theory. *Studies in Second Language Acquisition*, 26, 551-585.

Escudero, P. & Sharwood Smith, M. 2001. Reinventing the native speaker or What you never wanted to know about the native speaker so never dared to ask. In: Nizegorodcew, A. & Foster-Cohen, S. (eds.). *EuroSLA Yearbook 2001*. Amsterdam: John Benjamins, 275-286.

Escudero, P. & Williams, D. 2012. Native dialect influences second-language vowel perception: Peruvian versus Iberian Spanish learners of Dutch. *JASA*, 131, EL406-EL412.

Ettlinger, M. & Johnson, K. 2009. Vowel discrimination by English, French and Turkish speakers: Evidence for an exemplar-based approach to speech perception. *Phonetica*, 66, 222-242.

Evans, B. & Iverson, P. 2004. Vowel normalization for accent: An investigation of best exemplar locations in northern and southern British English sentences. *JASA* 115 (1), 352-361.

Faber, A. 1997. Genetic sub-grouping of the Semitic languages. In: Hetzron, R. (ed.). *Semitic Languages*. London: Routledge, 3-15.

Fant, G. 1960 [1970]. *Acoustic theory of speech production*. (2nd edition). 's-Gravenhage: Mouton.

Fant, G. 1989. Quantal theory and features. *JPhon*, 17, 79-86.

Fant, G. 2001. Swedish vowels and a new three-parameter model. *TMH-QPSR* 1/2001, 43-49. [http://www.speech.kth.se/prod/publications/files/qpsr/2001/2001_42_1_043-049.pdf]

Fant, G. 2004 [1980]. The Relations between Area Functions and the Acoustical Signal. In: Fant, G. (ed.). *Speech Acoustics and Phonetics. Selected Writings*. Dordrecht Boston London: Kluwer Academic Publishers, 29-57. Originally published in: *Phonetica,* 37 (1980), 55-86.

Farnetani, E. 1997. Coarticulation and Connected Speech Processes. In: Hardcastle, W. & Laver, J. (eds.). *The Handbook of Phonetic Sciences*, Malden, MA: Blackwell, 371-404.

Ferguson, Ch. 1959. Diglossia. *Word*, 15, 325-337.

Ferguson, Ch. 1991. The Arabic Koiné. *Language*, 35, 616-630.

Fiedler, W. 2003. Albanisch. In: Roelcke, Th. (ed.). *Variation Typology. A Typological Handbook of European Languages Past and Present*. Berlin New York: de Gruyter, 749-797.

Finegan, E. 2011. English. In: Comrie, B. (ed.). *The World's Major Languages*. New York: Routledge, 59-85.

Fischer-Jørgensen, E. 1990. Intrinsic F0 in tense and lax vowels with special reference to German. *Phonetica*, 47 (3-4), 99-140.

Flege, J. 1987. The production of "new" and "similar" phones in a foreign language: Evidence for the effect of equivalence classification. *JPhon*, 15, 47-65.

Flege, J. 1988a. Factors affecting degree of perceived foreign accent in English sentences. *JASA, 84*, 70-79.

Flege, J. 1988b. The production and perception of foreign language speech sounds. In: Winitz, H. (ed.). *Human Communication and its Disorders. A Review*. Norwood, NJ: Ablex, 224-401.

Flege, J. 1988c. Effects of equivalence classification on the production of foreign language speech sounds. In: James, A. & Leather, B. (eds.). *Sound Patterns in Second Language Acquisition*. Dordrecht: Foris, 9-39.

Flege, J. 1991. Orthographic evidence for the perceptual identification of vowels in Spanish and English. *Quarterly Journal of Experimental Psychology*, 43, 701-731.

Flege, J. 1992. Speech learning in a second language. In: Ferguson, C., Menn, L., Stoel-Gammon, C. (eds.). *Phonological Development: Models, Research, and Application*. Timonium, MD: York Press, 565-604.

Flege, J. 1993. Production and perception of a novel second-language phonetic contrast. *JASA*, 93, 1589-1608.

Flege, J. 1995. Second-language speech learning: Theory, findings, and problems. In: Strange, W. (ed.). *Speech Perception and Linguistic Experience: Issues in Cross-Language Research*. Timonium, MD: York Press, 233-277.

Flege, J. 1999. Age of learning and second-language speech. In: Birdsong , D. (ed.). *Second Language Acquisition and the Critical Period Hypothesis*. Hillsdale, NJ: Lawrence Erlbaum, 101-132.

Flege, J. 2002. Interactions between the native and second-language phonetic systems. In: Burmeister, P., Piske, T., Rohde, A. (eds.). *An Integrated View of Language Development: Papers in Honor of Henning Wode*. Trier: Wissenschaftlicher Verlag, 217-244.

Flege, J. 2003a. A method for assessing the perception of vowels in a second language. In: Fava, E. & Mioni, A. (eds.). *Issues in Clinical Linguistics*, Padova: Uni Press, 19-44.

Flege, J. 2003b. Assessing constraints on second-language segmental production and perception. In: Meyer, A. & Schiller, N. (eds.). *Phonetics and Phonology in Language Comprehension and Production, Differences and Similarities*. Berlin: Mouton de Gruyter, 319-355.

Flege, J. 2007. Language contact in bilingualism: Phonetic system interactions, in: Cole, J. & Hualde, J.I. (eds.). *Laboratory Phonology 9*. Berlin: Mouton de Gruyter, 353-382.

Flege, J. 2009. Give input a chance! In: Piske, T. & Young-Scholten, M. (eds.). *Input Matters in SLA*. Briston, England: Multilingual Matters, 175-190.

Flege, J., Bohn, O.-S., Jang, S. 1997. Effects of experience on non-native speakers' production and perception of English vowels. *JPhon*, 25, 437-470.

Flege, J. & Eefting, W. 1986. Linguistic and developmental effects on the production and perception of stop consonants. *Phonetica*, 43, 155-171.

Flege, J., Frieda, E., Nozawa, T. 1997. Amount of native-language (L1) use affects the pronunciation of an L2. *JPhon*, 25, 169-186.

Flege, J. & Liu, S. 2001. The effect of experience on adults' acquisition of a second language. *Studies in Second Language Acquisition*, 23, 527-552.

Flege, J. & MacKay, I. 2004. Perceiving vowels in a second language. *Studies in Second Language Acquisition*, 26, 1-34.

Flege, J. & MacKay, I. 2010. Age effects on second language acquisition. In: Dziubalska-Kołaczyk, K., Wrembel, M., Kul, M. (eds.). *New Sounds 2010. Proceedings of the 6th International Symposium on the Acquisition of Second Language Speech*, 113-118. [http://ifa.amu.edu.pl/newsounds/files/proceedings/proceedings_quotable_version.pdf]

Flege, J. & MacKay, I. 2011. What accounts for "age" effects on overall degree of foreign accent? In: Wrembel, M., Kul, M., Dziubalska-Kołaczyk, K. (eds.). *Achievements and Perspectives in the Acquisition of Second Language Speech: New Sounds 2010*, vol. 2. Bern: Peter Lang, 65-82.

Flege, J., Munro, M., Fox, R. 1994. Auditory and categorical effects on cross-language vowel perception. *JASA*, 95, 3623-3641.

Flege, J., Munro, M., MacKay, I. 1995. Factors affecting degree of perceived foreign accent in a second language. *JASA*, 97, 3125-3134.

Flege, J., Yeni-Komshian, G., Liu, S. 1999. Age constraints on second language learning. *Journal of Memory and Language*, 41, 78-104.

Fleischer, J. & Schmid, S. 2006. Zurich German. *Journal of the International Phonetic Association*, 36, 243-253.

Flemming, E. 1996. Evidence for constraints on contrast: the dispersion theory of contrast. *UCLA Working Papers in Phonology*, 1, 86-106.

Flemming, E. 2002. *Auditory Representations in Phonology.* (Outstanding Dissertations in Linguistics). New York London: Routledge.

Flemming, E. 2004. Contrast and perceptual distinctiveness. In: Hayes, B., Kirchner, R., Steriade, D. (eds.). *Phonetically based Phonology*. Cambridge: Cambridge University Press, 232-276.

Flemming, E. 2008. Speech Perception and Phonological Contrast. In: Pisoni, D. & Remez, R. (eds.). *The Handbook of Speech Perception*, Oxford: Blackwell, 156-181.

Fletcher, H.. 1950 [1972]. *Speech and Hearing in Communication*. New York, NY: Van Nostrand.

Fluck, H.-R., Zaiza, L., Qichang, Z. 1984. *Kontrastive Linguistik Deutsch-Chinesisch*. (Sprachvergleichende Arbeiten in den Bereichen Phonetik/Phonologie – Lexik/Morphologie/Syntax – Übersetzung). Heidelberg: Groos.

Fowler, C. 1986. An event approach to the study of speech perception from a direct-realist perspective. *JPhon*, 14, 3-28.

Fowler, C. & Brown, J.M. 1997. Intrinsic f0 differences in spoken and sung vowels and their perception by listeners. *Perception and Psychophysics*, 59 (5), 729-738.

Fox, R., Flege, J., Munro, M. 1995. The perception of English and Spanish vowels by native English and Spanish listeners: A multidimensional scaling analysis. *JASA*, 97, 2540-2551.

Francescato, G. 2009. Speech perception and the "basis of articulation". *Folia Linguistica*, 2 (3-4), 176-182.

Francis, A. & Nusbaum, H. 2002. Selective attention and the acquisition of new phonetic categories. *Journal of Experimental Psychology: Human Perception and Performance*, 28 (2), 349-366. [http://www.linguistics.berkeley.edu/~kjohnson/ling290e/francis_nusbaum_2002.pdf]

Freeman, A. 1996. *Andrew Freemans's Perspectives on Arabic Diglossia*. Ms. [http://innerbrat.org/Andyf/Articles/Diglossia/digl_96.htm]

Fridland, V. 2008. Patterns of /uw/, /ʊ/, and /ow/ fronting in Reno, Nevada. *American Speech*, 83, 432-454.

Frieda, E. & Nozawa, T. 2007. You are what you eat phonetically. The effect of linguistic experience on the perception of foreign vowels. In: Bohn, O.-S. & Munro, M. (eds.). *Language Experience in Second Language Speech Learning. In Honor of James Emil Flege*. Amsterdam Philadelphia: John Benjamins, 79-96.

Fries, C. 1945. *Teaching and Learning English as a Foreign Language*. Ann Arbor: University of Michigan Press.

Fry, D., Abramson, A., Eimas, P., Liberman, A. 1962. The identification and discrimination of synthetic vowels. *Language & Speech*, 5, 171-189.

Fujisaki, H. & Kawashima, T. 1969. On the modes and mechanisms of speech perception. *Annual Report of the Engineering Research Institute, Faculty of Engineering, University of Tokyo*, 28, 67-73.

Fujisaki, H. & Kawashima, T. 1970. Some experiments on speech perception and a model for the perceptual mechanism. *Annual Report of the Engineering Research Institute, Faculty of Engineering, University of Tokyo*, 29, 207-214.

Fujisaki, H. & Kawashima, T. 1971. A model of the mechanisms for speech perception: Quantitative analysis of categorical effects in discrimination. *Annual Report of the Engineering Research Institute, Faculty of Engineering, University of Tokyo*, 30, 59-68.

García Lecumberri, M. L., Cooke, M., Cutler, A. 2010. Non-native speech perception in adverse conditions: A review. *Speech Communication*, 52, 864-886. [http://laslab.org/upload/non-native_speech_perception_in_adverse_conditions:_a_review.pdf]

Gardner, R. 1991. Attitudes and motivation in second language learning. In: Reynolds, A. (ed.). *Bilingualism, Multiculturalism, and Second Language Learning*. Mahwah, NJ: Erlbaum, 43-63.

Garner, W. 1978. Selective attention to attributes and to stimuli. *Journal of Experimental Psychology: General*, 107, 287-308.

Gehrmann, S. 1995. *Deutsche Phonetik in Theorie und Praxis. Ein Lehr- und Übungsbuch auf lernpsychologisch-motorischer Grundlage. Anhang 1: Kontrastive Fehleranalyse kroatisch-deutsch,* 2nd edition. Zagreb: Školska knjiga.

Gerrits, E. & Schouten, M. 2004. Categorical perception depends on the discrimination task. *Perception & Psychophysics*, 66 (3), 363-376.

Ghazeli, S. 1979. Du statu des voyelles en arabe. Analyses-Théories. *Etudes Arabes*, 2-3, 199-219.

Gillon, G. 2004. *Phonological Awareness: from Research to Practice*. New York: Guilford Press.

Göksel, A. & Kerslake, C. 2005. *Turkish: A Comprehensive Grammar*. London: Routledge.

Goldinger, S. 1996. Words and voices: Episodic traces in spoken word identification and recognition memory. *Journal of Language and Social Psychology: Learning Memory, and Cognition,* 22 (5), 1166-1183.

Goldinger, S. 1998. Echoes of echoes? An episodic theory of lexical access. *Psychological Review*, 105, 251-279.

Goldinger, S. 2007. A complementary-systems approach to abstract and episodic speech perception. *ICPhS* XVI, Saarbrücken, 49-54.

Goldstein, L. 1977. Three studies in perception: features, relative salience, and bias. *Working Papers in Phonetics, Dept. of Linguistics, UCLA*, 39. [http://escholarship.org/uc/item/08c9j6cm]

Goldstone, R. 1994a. Similarity, interactive activation, and mapping. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 20, 3-28. [http://cognitrn.psych.indiana.edu/rgoldsto/pdfs/siam.pdf]

Goldstone, R. 1994b. The role of similarity in categorization: Providing a groundwork. *Cognition*, 52, 125-157. [http://cognitrn.psych.indiana.edu/rgoldsto/pdfs/sim&cat.pdf]

Goldstone, R., Braithwaite, D., & Byrge, L. 2012. Perceptual learning. In: Seel, N. (ed.) *Encyclopedia of the Sciences of Learning*. Heidelberg, German: Springer, 2616-2619. [Ms. http://cognitrn.psych.indiana.edu/rgoldsto/pdfs/perlearnESL.pdf]

Goldstone, R., & Day, S. 2013. Similarity. In: Pashler, H. (ed.) *The Encyclopedia of Mind*. Thousand Oaks, CA: Sage, 696-699. [Ms. http://www.indiana.edu/~pcl/papers/similaritysage.pdf]

Goldstone, R. & Yun Son, J. 2005. Similarity. In: Holyoak, K. & Morrison, R. (eds.). *The Cambridge Handbook of Thinking and Reasoning.* New York et al.: Cambridge University Press, 13-36.

Goldstone, R. & Yun Son, J. 2012. Similarity. In: Holyoak, K. & Morrison, R. (eds.). *The Oxford Handbook of Thinking and Reasoning*. New York et al.: Cambridge University Press, 155-176.

Gordon, M. & Applebaum, A. 2006. Phonetic Structures of Turkish Kabardian. *Journal of the International Phonetic Association*, 36 (2), 159-186. [http://www.linguistics.ucsb.edu/faculty/gordon/kabardianphonetics.pdf]

Gordon, M., Ghushchyan, E., McDonnell, B., Rosenblum, D., Shaw, P. 2012. Sonority and central vowels: A cross-linguistic phonetic study. In: Parker, St. (ed.). *The Sonority Controversy*. Berlin Boston: de Gruyter, 219-256. [http://www.linguistics.ucsb.edu/faculty/gordon/Gordonetal_Sonority.pdf]

Górka, J. 1973. *Der Einfluß der muttersprachlichen Artikulationsbasis auf die deutsche Aussprache polnischer Sprecher*. Dissertation. Kraków: Universytet Jagielloński Kraków.

Gósy, M. 1992. *Speech Perception*. (Forum Phoneticum 50). Frankfurt am Main: Hector.

Gósy, M. 2004. *Fonetika, a beszéd tudománya*. Budapest: Osiris.

Gottfried, T. 1984. Effects of consonant context on the perception of French vowels. *JPhon,* 12, 91-114. [http://www2.lawrence.edu/fast/gottfrit/Gottfried1984.pdf]

Gottfried, T. & Beddor, P. 1988. Perception of temporal and spectral information in French vowels. *Language and Speech*, 31, 57-75 (also in: Status Report on Speech Research, 93/94, 51-66. [http://www.haskins.yale.edu/sr/SR093/SR093_04.pdf]).

Granser, Th. (n.y.) *Sprachensteckbrief Albanisch*. [http://www.schule-mehrsprachig.at/fileadmin/schule_mehrsprachig/redaktion/sprachensteckbriefe/pdf/Sprachensteckbrief_Albanisch.pdf]

Greenberg, J. 1966. Language universals. *Current Trends in Linguistics*, 3, 61-112.

Greenberg, J. (ed.) 1978. *Universals of Human Language, Vol. 2: Phonology*. Stanford: Stanford University Press.

Gregor-Chiriţă, G. 1991. *Das Lautsystem des Deutschen und des Rumänischen*, (Deutsch im Kontrast, Bd. 11). Heidelberg: Julius Groos.

Grieser, D. & Kuhl, P. 1989. Categorization of speech by infants: Support for speech-sound prototypes. *Developmental Psychology,* 25, 577-588.

Griffiths, C. 2004. *Language Learning Strategies. Theory and Research*. Occasional Paper, Ms. [http://www.crie.org.nz/research-papers/c_griffiths_op1.pdf]

Griffiths, C. & Parr, J. 2001. Language-learning strategies: theory and perception. *ELT Journal* 55 (3), 247-254.

Grotjahn, R. 1998. Lernstile und Lernstrategien: Definition, Identifikation, unterrichtliche Relevanz. *Der Fremdsprachliche Unterricht – Französisch*, 34, 11-15.

Guenther, F., Husain, F., Cohen, M., Shinn-Cunningham, B. 1999. Effects of·categorization and discrimination training on auditory perceptual space. *JASA*, 106, 2900-2912.

Guion, S., Flege, J., Ahahane-Yamada, R., Pruitt, J. 2000. An investigation of current models of second language speech perception: The case of Japanese adults' perception of English consonants. *JASA,* 107, 2711-2725.

Guiora, A. et al. 1972. Construct Validity and Transpositional Research: Toward an Empirical Study of Psychoanalytic Concepts. *Comprehensive Psychiatry*, 13, 139-150.

Gullberg, M. & Indefrey, P. (eds.) 2010. *The Earliest Stages of Language Learning*. New York: Wiley-Blackwell.

Gussmann, E. 2007. *The Phonology of Polish*. Oxford: Oxford University Press.

Gut, U. 2010. Cross-linguistic influence in L3 phonological acquisition. *International Journal of Multilingualism*, 7 (1), 19-38.

Gvozdanović, J. 1980. *Tone and Accent in Standard Serbo-Croatian: with a Synopsis of Serbo-Croatian Phonology*. Wien: Verlag der Österreichischen Akademie der Wissenschaften.

Hachenberg, P. 2003. Probleme chinesischer Lerner mit dem deutschen Vokalsystem: Ein Überblick. In: Wannagat, U., Gerbig, J., Bucher. S. (eds.). *Deutsch als zweite Fremdsprache in Ostasien – neue Perspektiven*. Tübingen: Peter Lang, 97-128.

Hackert, S. 2012. *The Emergence of the English Native Speaker. A Chapter in Nineteenth-Century Linguistic Thought*. Boston, MA: de Gruyter Mouton.

Hagiwara, R. 1997. Dialect variation and formant frequency: The American English vowels revisited. *JASA*, 102, 655-658.

Hagiwara, R. 2006. Vowel production in Winnipeg. *The Canadian Journal of Linguistics*, 51 (2/3), 127-141. [http://home.cc.umanitoba.ca/~robh/files/cjlfinaldistr.pdf]

Hajek, J. & Maeda, S. 2000. Investigating universals of sound change: the effect of vowel height and duration on the development of distinctive nasalization. In: Broe, M. & Pierrehumbert, J. (eds.). *Papers in Laboratory Phonology V*. Cambridge: Cambridge University Press, 52-69. [http://repository.unimelb.edu.au/10187/99]

Hall, K. 2007. Pairwise perceptual magnet effects. *ICPhS,* XVI, Saarbrücken, 669-672. [http://www.icphs2007.de/conference/Papers/1602/1602.pdf]

Hancin-Bhatt, B. 1994. Segment transfer: a consequence of a dynamic system. *Second Language Research*, 10, 241-269.

Hancin-Bhatt, B. 2008. Second language phonology in optimality theory. In: Hansen Edwards, J. & Zampini, M. (eds.). *Phonology and Second Language Acquisition*. Amsterdam: John Benjamins, 117-148.

Hansen Edwards, J. & Zampini, M. (eds.) 2008. *Phonology and Second Language Acquisition*. Amsterdam: John Benjamins.

Hansen Edwards, J. 2008. Social factors and variation in production in L2 phonology. In: Hansen Edwards, J. & Zampini, M. (eds.). *Phonology and Second Language Acquisition*. Amsterdam: John Benjamins, 251-280.

Hardcastle, W. & Marchal, A. (eds.) 1990. *Speech Production and Speech Modelling*. Kluwer Academic Publishers. [http://getebook.org/?p=235153]

Harnad, S. (ed.) 1987. *Categorical Perception: The Groundwork of Cognition*. Cambridge: Cambridge University Press.

Harrington, J., Cox, F., Evans, Z. 1996. An acoustic study of broad, general and cultivated Australian English vowels. *Australian Journal of Linguistics*, 17 (2), 155-184. [http://www.tandfonline.com/doi/pdf/10.1080/07268609708599550]

Harrington, J., Hoole, P., Kleber, F., Reubold, U. 2011. The physiological, acoustic, and perceptual basis of high back vowel fronting: Evidence from German tense and lax vowels. *JPhon*, 39, 121-131.

Harrington, J., Kleber, F., Reubold, U. 2008. Compensation for coarticulation. /u/-fronting, and sound change in Standard Southern British: An acoustic and perceptual study. *JASA*, 123, 2825-2835.

Harris, J. & Lindsey, G. 1995. The elements of phonological representation. In: Durand, J. & Katamba, F. (eds.). *Frontiers of Phonology: Atoms, Structures, Derivations*. Harlow, Essex: Longman, 34-79. [http://www.phon.ucl.ac.uk/home/johnh/papers/Durakata01.pdf]

Harris, J. & Lindsey, G. 2000. Vowel patterns in mind and sound. In: Burton-Roberts, N., Carr, Ph., Docherty, G. (eds.). *Phonological Knowledge: Conceptual and Empirical Issues*. Oxford: Oxford University Press, 185-205. [http://www.phon.ucl.ac.uk/home/johnh/papers/Newky48.pdf]

Hary, J. & Massaro, D. 1982. Categorical results do not imply categorical perception. *Perception & Psychophysics,* 32 (5), 409-418. [http://www.phonetik.uni-muenchen.de/~reichelu/kurse/perz_fort/literatur/HaryPP1982.pdf]

Haspelmath, M. 2003. *Against Iconicity and Markedness*. Paper read at Stanford University. [http://www.eva.mpg.de/lingua/staff/haspelmath/pdf/IconicityMarkedness.pdf]

Haspelmath, M. 2006. Against markedness (and what to replace it with). *Journal of Linguistics*, 42, 25-70. [http://email.eva.mpg.de/~haspelmt/Againstmarkedness.pdf]

Haspelmath, M., Dryer, M., Gil, D., Comrie, B. (eds.) 2008. *The World Atlas of Language Structures*, vol. 1. Oxford: Oxford University Press.

Hawkins, S. & Midgley, J. 2005. Formant frequency of RP monophthongs in four age groups of speakers. *Journal of the International Phonetic Association*, 35, 183-199.

Hayes, B. 1979. The rhythmic structure of Persian verse. *Edebiyat*, 4, 193-242.

Hayes, B. 1999. Phonetically-driven phonology: the role of Optimality Theory and inductive grounding. In: Darnell, M., Moravscik, E., Noonan, M., Newmeyer, F., Wheatly, K. (eds.). *Functionalism and Formalism in Linguistics, Vol. 1: General Papers*. Amsterdam: John Benjamins, 243-285.

Hayes, B., Kirchner, R., Steriade, D. (eds.) 2004. *Phonetically based Phonology*. Cambridge: Cambridge University Press.

Hayes, B. & Puppel, S. 1985. On the Rhythm Rule in Polish. In: van der Hulst, H. & Smith, N. (eds.). *Advances in Nonlinear Phonology.* Dordrecht: Foris*,* 59-81.

Hayes, B. & Steriade, D. 2004. Introduction: The phonetic bases of phonological Markedness. In: Hayes, B., Kirchner, R., Steriade, D. (eds.). *Phonetically based Phonology*. Cambridge: Cambridge University Press, 1-33.

Hazan, V. & Barrett, S. 2000. The development of phonemic categorization in children aged 6-12. *JPhon*, 28, 377-396.

Hazan, V. 2007. Second language acquisition and exemplar theory. *ICPhS,* XVI, Saarbrücken, 39-42.

Hentschel, G. 1981. Die phonologische Interferenz des Polnischen im Bereich der Aussprache deutscher monophthongischer Vokale. *Kwartalnik Neofilologiczny*, 28 (1), 27-36.

Hentschel, G. 1982. *Die Perzeption der vorderen, labialen Vokale des Deutschen durch Native-Speaker des Polnischen*. (Slavistische Linguistik 6). Frankfurt am Main.

Hentschel, G. 1986. *Vokalperzeption und natürliche Phonologie. Eine kontrastive Untersuchung zum Deutschen und Polnischen*. München: Otto Sagner.

Herok, T. 2001. Zum Verhältnis von vokalischer Quantität und Qualität im Deutschen und Ungarischen. In: Eichner, H., Mumm, P.-A., Panagl, O., Winkler, E. (eds.). *Fremd und Eigen. Untersuchungen zu Grammatik und Wortschatz des Uralischen und Indogermanischen. In memoriam Hartmut Katz.* Wien: Edition Praesens, 77-153.

Hertrich, I. & Ackermann, H. 1997. Articulatory control of phonological vowel length contrasts: Kinematic analysis of labial gestures. *JASA*, 102 (1), 523-536.

Hillenbrand, J. 2003. American English: Southern Michigan. *Journal of the International Phonetic Association*, 33 (1), 121-126.

Hillenbrand, J. 2013. Static and dynamic approaches to vowel perception. In: Morrison, G. & Assmann, P. (eds.). *Vowel inherent Spectral Change. Modern Acoustics and Signal Processing*. Berlin, Heidelberg: Springer, 9-30. [http://homepages.wmich.edu/~hillenbr/Papers/Hillenbrand%202013%20Static%20and%20dynamic%20approaches%20to%20vowel%20perception.pdf]

Hillenbrand, J., Clark, M., Houde, R. 2000. Some effects of duration on vowel recognition. *JASA*, 108 (6), 3013-3022. [http://www.ncbi.nlm.nih.gov/pubmed/11144593]

Hillenbrand, J., Clark, M., Nearey, T. 2001. Effects of consonant environment on vowel formant patterns. *JASA*, 109 (2), 748-763.

Hillenbrand, J., Clark, M.J., Baer, C.A. 2011. Perception of sinewave vowels. *JASA*,129, 3991-4000. [http://homepages.wmich.edu/~hillenbr/Papers/Hillenbrand%20Clark%20Baer%202011%20Perception%20of%20sinewave%20vowels%20Final.pdf]

Hillenbrand, J., Getty, L., Clark, M., Wheeler, K. 1995. Acoustic characteristics of American English vowels. *JASA*, 97 (5), 3099-3111.

Hillenbrand, J. & Nearey, T. 1999. Identification of resynthesized /hVd/ utterances: Effects of formant contour. *JASA*, 79, 124-130.

Hirschfeld, U. 1994. *Untersuchungen zur phonetischen Verständlichkeit Deutschlernender*. (Forum Phoneticum 57). Hamburg: Buske.

Hohlfeld, A., Mierke, K., Sommer, W. 2004. Is word perception in a second language more vulnerable than in one's native language? Evidence from brain potentials in a dual task setting. *Brain and Language*, 89 (3), 569-579.

Hoole, P. 1999a. Articulatory-acoustic relations in German vowels. *ICPhS,* XIV, San Francisco, 2153-2156.

Hoole, P. 1999b. On the lingual organization of the German vowel system. *JASA,* 106 (2), 1020-1032.

Hoole, P. & Kroos, C. 1998. Control of larynx height in vowel production. *The 5th International Conference on Spoken Language Processing (ICSLP),* Sydney, Australia. [http://www.phonetik.uni-muenchen.de/~hoole/pdf/icslp98_larheight_proc.pdf]

Hoole, P. & Mooshammer, C. 2002. Articulatory analysis of the German vowel system. In: P. Auer, P. Gilles, Spiekermann, H. (eds.). *Silbenschnitt und Tonakzente*. Tübingen: Niemeyer, 129-152. [http://www.haskins.yale.edu/staff/mooshammer/ph_freiburg_art.pdf]

Houston, D. 2008. Speech Perception in Infants. In: Pisoni, D. & Remez, R. (eds.). *The Handbook of Speech Perception*. Malden, MA: Blackwell, 417-448.

Howie Marshall, J. 1976. *Acoustical Studies of Mandarin Vowels and Tones*. Cambridge: Cambridge University Press.

Hume, E. 2008. Markedness and the Language User. *Phonological Studies*, 11, 83-98. [http://www.ling.ohio-state.edu//~ehume/papers/Hume_Markedness&LgUser.pdf]

Hume, E. 2010. Deconstructing markedness. A predictability-based approach. *Proceedings of the Berkeley Linguistic Society 2004*. [http://www.ling.ohio-state.edu/~ehume/papers/Hume_markedness_BLS30.pdf]

Hume, E. 2011. Markedness. In: Van Oostendorp, M., Ewen, C., Hume, E., Rice, K. *The Blackwell Companion to Phonology*, vol. 1. Cambridge, Boston: Wiley-Blackwell, 79-106.

Hume, E. & Johnson, K. (eds.) 2001a. *The Role of Speech Perception in Phonology*. New York: Academic Press.

Hume, E. & Johnson, K. 2001b. A Model of the Interplay of Speech Perception and Phonology. In: Hume, E. & Johnson, K. (eds.). *Studies on the Interplay of Speech Perception and Phonology*. *Working Papers in Linguistics,* 55, 1-22.

Hunold, C. 2009. *Untersuchungen zu segmentalen und suprasegmentalenAusspracheabweichungen chinesischer Deutschlernernder*. Frankfurt am Main Wien: Lang.

Hura, S.L., Lindblom B., Diehl, R. 1992. On the role of perception in shaping phonological assimilation rules. *Language and Speech,* 35, 59-72.

Hyltenstam, K. 2011. *The native speaker as a methodological instrument – are there alternatives?* Paper read at EUROSLA 21, Annual Conference of the European Second Language Association. [http://eurosla.org/eurosla21_LLroundtable.html]

Iivonen, A. 1987a. Zur regionalen Variation der betonten Vokale im gehobenen Deutsch. In: Kahlas-Tarkka, L. (ed.). *Neophilologica Fennica. Neuphilologischer Verein 100 Jahre. Mémoires de la Société Néophilologique de Helsinki* XLV, 87-119.

Iivonen, A. 1987b. Regional Differences in the Realization of Standard German Vowels. *ICPhS,* XI, Tallinn: Academy of Sciences of the Estonian S.S.R., vol. 4, 161-164.

Iivonen, A. 1987c. Monophthonge des gehobenen Wiener Deutsch. *Folia Linguistica*, XXI/2-4, 293-336.

Iliescu, M. 2002. Rumänisch. In: Okuka, M. et al. (eds.). *Wieser Enzyklopädie des europäischen Ostens. Lexikon der Sprachen des europäischen Ostens, Bd. 10.* Klagenfurt et al.: Wieser, 145-166.

Iliescu, M. 2003. Rumänisch. In: Roelcke, Th. (ed.). *Variationstypologie. Ein sprachtypologisches Handbuch der europäischen Sprachen*. Berlin: de Gruyter, 530-572.

Ingram, J. & Park, S.-G. 1997. Cross-language vowel perception and production by Japanese and Korean learners of English. *JPhon*, 25, 343-370.

Inkelas, S. & Orgun, C. 1995. Level Ordering and Economy in the Lexical Phonology of Turkish. *Language*, 71 (4), 763-793.

Inkelas, S. & Orgun, C. 2003. Turkish stress: a review. *Phonology*, 2, 139-161.

Ioup, G. 2008. Exploring the role of age in L2 phonology. In: Hansen Edwards, J. & Zampini, M. (eds.). *Phonology in Second Language Acquisition*. Amsterdam: John Benjamins, 41-62.

IPA International Phonetic Association 1999. *Handbook of the International Phonetic Association. A Guide to the Use of the International Phonetic Alphabet*. Cambridge: Cambridge University Press.

Ito, M., Tsuchida, J., Yano, M. 2001. On the effectiveness of whole spectral shape for vowel perception. *JASA*, 110 (2), 1141-1149.

Iverson, P. & Evans, B. 2007. Auditory training of English vowels for first-language speakers of Spanish and German. *ICPhS,* XVI, Saarbrücken, 1625-1628.

Iverson, P. & Kuhl, P. 1995. Mapping the perceptual magnet effect for speech using signal detection theory and multidimensional scaling. *JASA*, 97, 553-562.

Iverson, P. & Kuhl, P. 2000. Perceptual magnet and phoneme boundary effects in speech perception: Do they arise from a common mechanism? *Perception & Psychophysics*, 62 (4), 874-886.

Iverson, P., Kuhl, P., Akahane-Yamada, R., Diesch, E., Tokhura, Y., Kettermann, A., Siebert, C. 2003. A perceptual interference account of acquisition difficulties for non-native phonemes. *Cognition*, 87, B47-B57.

Jacewicz, E., Fox, R. Salmons, J. 2011. Vowel change across three age groups of speakers in three regional varieties of American English. *JPhon*, 39, 683-693. [http://web.shs.ohio-state.edu/Faculty/Jacewicz/Jacewicz_Publications/2011_Jacewicz%20et%20al_JPhonetics.pdf]

Jakobson, R. 1941 [1968]. *Child Language, Aphasia and Phonological Universals*. The Hague: Mouton.

Jakobson, R. 1962 [1969]. *Kindersprache, Aphasie und allgemeine Lautgesetze*. Frankfurt: Suhrkamp.

Jakobson, R. 1971. *Studies on Child Language and Aphasia*. The Hague: Mouton.

Jakobson, R., Fant, C.G. & Halle, M. 1952. *Preliminaries to Speech Analysis. The distinctive features and their correlates*. Cambridge, MA: MIT Press.

Jakobson, R. & Halle, M. 1956. *Fundamentals of Language*. s'Gravenhage: Mouton.

Janda, L. 2010. Cognitive Linguistics in the year 2010. *International Journal of Cognitive Linguistics*, 1 (1), 1-30. [http://www.personal.psu.edu/bur12/index_files/IJCL.pdf]

Janda, R. & Auger, J. 1992. Quantitative evidence, qualitative hypercorrection, sociolinguistic variables – and French speakers' 'eadhaches with English h/Ø. *Language and Communication*, 12, 195-236.

Janota, F. 1965. *Auditory Evaluation of Synthetic Vowel Sounds*. Habilitation, University of Prague.

Jarvis, S. & Pavlenko, A. 2008. *Cross-Linguistic Influence in Language and Cognition*. New York: Routledge.

Jarvis, S. 2012. Crosslinguistic influence and multilingualism. In: Chapelle, C. (ed.). *The Encyclopedia of Applied Linguistics*. Boston: Wiley Blackwell DOI: 10.1002/9781405198431.wbeal0291 [http://onlinelibrary.wiley.com/doi/10.1002/9781405198431.wbeal0291/abstract].

Jassem, W. 1968. Vowel formant frequencies as cues to speaker discrimination. In: Jassem, W. (ed.). *Speech Ananlysis and Synthesis 1*, 9-41.

Jassem, W. 2003. Polish. *Journal of the International Phonetic Association*, 33 (1), 103-107. [http://ifa.amu.edu.pl/~krynicki/teaching/cg/files/06_phonetics_JASSEM.pdf]

Jenkins, J., Strange, W., Miranda, S. 1994. Vowel identification in mixed-speaker silent-center syllables. *JASA*, 95, 1030-1043.

Jessen, M., Marasek, K., Schneider, K., Clahßen, K. 1995. Acoustic Correlates of Word Stress and the tense/lax Opposition in the Vowel System of German. *ICPhS,* XIII, Stockholm, vol. 4, 428-431.

Jilka, M. 2009. Talent and proficiency in language. In: Dogil, G. & Reiterer, S. (eds.). *Language Talent and Brain Activity*, Berlin: de Gruyter, 1-16.

Jilka, M., Anufryk, V., Baumotte, H., Lewandowska, N., Rota, G., Reiterer, S. 2007a. Assessing individual talent in second language production and perception. In: Rauber, A., Watkins, M., Baptista, B. (eds.). *New Sounds 2007. Proceedings of the Fifth International Symposium on the Acquisition of Second Language Speech*. Florianópolis, Brazil: Federal University of Santa Catarina, 243-258.

Jilka, M., Baumotte, H., Lewandowski, N., Reiterer, S., Rota, G. 2007b. Introducing a Comprehensive Approach to Assessing Pronunciation Talent. *ICPhS,* XVI, Saarbrücken, 1737-1740.

Johnson, J. & Newport, E. 1989. Critical period effects in second language learning: The influence of maturational state on the acquisition of English as a second language. *Cognitive Psychology*, 21, 60-99.

Johnson, K. 1997. Speech perception without speaker normalization: An Exemplar Model. In: Johnson, K. & Mullenix, J.W. (eds.). *Talker Variability in Speech Processing*. San Diego: Academic Press, 145-165. [http://linguistics.berkeley.edu/~kjohnson/papers/SpeechPerceptionWithoutSpeakerNormalization.pdf]

Johnson, K. 2000. Adaptive dispersion in vowel perception. *Phonetica*, 57, 181-188. [http://linguistics.berkeley.edu/~kjohnson/papers/Adaptive.pdf]

Johnson, K. 2003. *Acoustic and Auditory Phonetics*, 2nd edition. Oxford: Blackwell.

Johnson, K. 2004. Cross-linguistic perceptual differences emerge from the lexicon. In: Agwuele, A., Warren, W., Park, S.-H. (eds.). *Proceedings of the 2003 Texas Linguistics Society Conference*. Somerville: Cascadilla Proceedings Project, 26-41. [http://linguistics.berkeley.edu/~kjohnson/papers/TLS_xling.pdf]

Johnson, K. 2006. Resonance in an exemplar-based lexicon: The emergence of social identity and phonology. *JPhon*, 34 (4), 485-499.

Johnson, K. 2008a. Speaker Normalization in Speech Perception. In: Pisoni, D. & Remez, R. (eds.). *The Handbook of Speech Perception*. Oxford: Blackwell, 363-389.

Johnson, K. 2008b. *Quantitative Methods in Linguistics*. Malden: Blackwell.

Johnson, S., Johnson, W. & Remers., P. 2010. *Patterns in Child Phonology*. Edinburgh University Press.

Jones, D. 1932. *An outline of English phonetics*, 3rd edition. Leipzig: Teubner.

Joos, Martin. 1948. Acoustic Phonetics. *Language*, 24, Suppl. 2 (=Language Monograph 23).

Jørgensen, H.P. 1969. Die gespannten und ungespannten Vokale in der norddeutschen Hochsprache mit einer spezifischen Untersuchung der Struktur ihrer Formantfrequenzen. *Phonetica*, 19, 217-245.

Josipović Smojver, V. 2010. Foreign accent and levels of analysis. Interference between English and Croatian. In: Waniek-Klimczak, E. (ed.). *Issues in Accents of English 2: Variability and Norm.* Newcastle upon Tyne: Cambridge Scholars Publishing, 23-35. [http://bib.irb.hr/datoteka/449577.Vinja-CSP.doc]

Jusczyk, P. 1980. Auditory vs. phonetic coding of speech signals during infancy. In: Mehler, J. Walker, E., Garrett, M. (eds.). *Perspectives on Mental Representation*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Jusczyk, P. 1997. *The Discovery of Spoken Language*. Cambridge, MA: MIT Press.

Jusczyk, P., Cutler, A., Redanz, N. 1993. Infants' preference for the predominant stress patterns of English words. *Child Development*, 23, 648-654.

Kabak, B. & Vogel, I. 2001. The Phonological Word and Stress Assignment in Turkish. *Phonology*, 18, 315-36.

Kabak, B. 2007. Hiatus resolution in Turkish: An underspecification account. *Lingua*, 117 (8), 1378-1411.

Kaden, K. 1976. *Das Laut- und Tonsystem der modernen chinesischen Hochsprache. Untersuchungen zur Fundierung des Chinesischunterrichts an Ausbildungsstätten der DDR*. Dissertation. Berlin.

Kager, R. 1999. *Optimality Theory*. Cambridge: Cambridge University Press.

Karypidis, C. 2007. Order effects and vowel decay in short-term memory: The neutralization hypothesis. *ICPhS,* XVI, Saarbrücken, 657-660.

Karypidis, C. 2009. *What d' scores can hide and what decision space representations should reveal: An account on asymmetries in vowel perception*. Ms. [http://hal.archives-ouvertes.fr/docs/00/50/27/91/PDF/karypidis2009_Fechner.pdf]

Kaun, A. 1995. *The typology of rounding harmony. An optimality theoretic approach*. UCLA Doctoral Dissertation in Linguistics, No. 8. [http://roa.rutgers.edu/files/227-1097/227-1097-KAUN-0-0.PDF]

Kaun, A. 2004. The typology of rounding harmony. In: Hayes, B., Kirchner, R., Steriade, D. (eds.). *Phonetically based Phonology*. Cambridge: Cambridge University Press, 87-116.

Kaye, A. 2011. Arabic. In: Comrie, B. (ed.). *The World's Major Languages*. New York: Routledge, 560-577.

Kaye, J., Lowenstam, J., Vergnaud, J.-R. 1985. The internal structure of phonological elements: a theory of charm and government. *Phonology Yearbook*, 2, 305-328.

Kazanina, N., Phillips, C., Idsardi, W. 2006. The influence of meaning on the perception of speech sounds. *Proceedings of the National Academy of Sciences*, 103 (30), 11381-11386.

Kellerman, E. & Sharwood Smith, M. 1986. *Crosslinguistic Influence in Second Language Acquisition*. New York: Oxford University Press

Kellerman, E. 1978. Giving learners a break: Native language intuitions as a source of predictions of transferability. *Working Papers on Bilingualism*, 15, 59-92.

Kellerman, E. 1979. Transfer and non-transfer: where we are now. *Studies in Second Language Acquisition*, 2, 37-57.

Kellerman, E. 1983. Now you see it, now you don't. In: Gass, S. & Selinker, L. (eds.). *Language Transfer in Second Language Learning*. Rowley, MA: Newbury House, 112-134.

Kellerman, E. 1995. Crosslinguistic influence: Transfer to nowhere. *Annual Review of Applied Linguistics*, 15, 125-150.

Kerschhofer, N. 1995. *Vokalsubstitutionen in der Fremdsprache*. Master Thesis. University of Vienna.

Kerschhofer-Puhalo, N. 1998a. The acquisition of German vowels by Polish learners. In: James, A. & Leather, J. (eds.). *New Sounds 97. Proceedings of the 3rd International Symposium on the Acquisition of Second Language Speech 1997,* 167-175.

Kerschhofer-Puhalo, N. 1998b. Sprachspezifische phonetische Eigenheiten und der Transfer von Phonemen aus der Muttersprache. In: *Lernsprache Deutsch,* 1998 (2).

Kerschhofer-Puhalo, N. 2010a. Preferences in L2 Vowel Perception. In: Dziubalska-Kolaczyk, K., Wrembel, M., Kul, M. (eds.), *New Sounds 2010, Proceedings of the 6thInternational Symposium on the Acquisition of Second Language Speech* 2010, 245-250.

Kerschhofer-Puhalo, N. 2010b. Sprache gehört gehört – Anregungen für den Schriftspracherwerb in der Zweitsprache. *Tribüne* 2010 (1) (Themenheft Sprache und Migration), LIT-Verlag, 19-23.

Kerschhofer-Puhalo, N. 2012. Phonetik und Phonologie im Schriftspracherwerb. *ÖDaF-Mitteilungen*, 2012 (1), 75-85.

Kerschhofer-Puhalo, N. (in press a). Fertigkeiten und ihre Wechselwirkungen im Schriftspracherwerb Deutsch als Zweitsprache. In: Pietzuch, A. (ed.), *Schriftspracherwerb im Kontext von Migration und Zweitsprache*, XV. Internationale Deutschlehrertagung IDT 2013 Proceedings. Bozen: Bozen-Bolzano University Press.

Kerschhofer-Puhalo, N. (in press b). ‚Ich habe solche Probleme mit den Umläuten‘ –Ähnlichkeitsbeziehungen zwischen deutschen Vokalen. In: Koroschetz Maragno, R. & Missaglia, F. (eds.). *Phonetik und Prosodie*, XV. Internationale Deutschlehrertagung IDT 2013 Proceedings. Bozen: Bozen-Bolzano University Press.

Kewley-Port, D. & Atal, B. 1989. Perceptual distances between vowels located in a limited phonetic space. *JASA,* 85, 1726-1740.

Kiefte, M., Enright, T., Marshall, L. 2010. The role of formant amplitude in the perception of /i/ and /u/. *JASA*, 127 (4), 2611-21.

Kiefte, M., Nearey, T., Assmann, P. 2013. Vowel perception in normal speakers. In: Ball, M. & Gibbon, F. (eds.). *Handbook of Vowels and Vowel Disorders*. Psychology Press, 160.

Kiliç, M.A. 2003. Türkiye Türkçesi'ndeki ünlülerin sesbilgisel özellikleri. In: Özsoy, A.S., Taylan, E.E., Aksu-Koç, A.,Akar, D., Nakipoğlu, M. (eds.). *Studies in Turkish Linguistics*. Istanbul: Boğaziçi University Press, 3-18.

Kiliç, M.A.& Öğüt, F. 2004. A high unrounded vowel in Turkish: is it a central or back vowel? *Speech Communication*, 43 (1-2), 143-154.

Kingston, J. & Diehl, R. 1994. Phonetic knowledge. *Language*, 70, 419-454

Kingston, J. 1992. The phonetics and phonology of perceptually motivated articulatory covariation. *Language and Speech*, 35, 99-113.

Kingston, J. 2003. Learning foreign vowels. *Language and Speech,* 46, 295-349.

Kiss, J. 2003. Ungarisch. In: Roelcke, Th. (ed.). *Variation Typology. A Typlogical Handbook of European Languages Past and Present*. Berlin New York: de Gruyter, 905-918.

Klatt, D. 1976. Linguistic uses of segmental duration in English. Acoustical and perceptual evidence. *JASA*, 59 (5), 1208-1221.

Klatt, D. 1982. Speech processing strategies based on auditory models. In: Carlson, R. & Granstrom, B. (eds.). *The Representation of Speech in the Auditory System*. Amsterdam: Elsevier, 181-196.

Kleiner, S. & Knöbl, R. 2011. Hochdeutsch und Hochdeutsch: Regionale Gebrauchsstandards im gesprochenen Deutsch. *Sprachreport*, 2/2011, Mannheim: Institut für Deutsche Sprache, 2-10.

Kochetov, A. 2008. Self-organization through misperception: Secondary articulation and vowel contrasts in language inventories. In: Avery, P., Elan Dresher, B., Rice, K. (eds.). *Contrast in Phonology. Theory, Perception, Acquisition*, Berlin: Mouton de Gruyter, 193-218.

Kohler, K. 1990. Segmental reduction in connected speech: Phonological facts and phonetic explanations. In: Hardcastle, W. & Marchal, A. (eds.). *Speech Production and Speech Modelling*. Dordrecht: Kluwer Academic Publishers, 69-92.

Kohler, K. 1999. German. In: IPA (ed.). *Handbook of the International Phonetic Association*. Cambridge: Cambridge University Press, 86-89.

Konopka, M. 2003. Polnisch. In: Roelcke, Th. (ed.). *Variationatypologie. Ein sprachtypologisches Handbuch der europäischen Sprachen in Geschichte und Gegenwart.* Berlin New York: de Gruyter, 657-683.

Kontra, M. 1995. On current research into spoken Hungarian. *International Journal of the Sociology of Language*, 111, 9-20.

Kontra, M. 2002. Where is the 'most beautiful' and the 'ugliest' Hungarian spoken? In: Long, D. & Preston, D. (eds.). *Handbook of Perceptual Dialectology*, vol. 2, 205-218.

Kornfilt, J., 1997. *Turkish*. London: Routledge.

Kostić, D. & Das, R.S. 1969. An analytical description of interrelationships among Serbo-Croatian vowels. *Phonetica*, 20, 193-204.

Kozhevnikov, V. & Chistovich, L. 1965. *Speech: Articulation and Perception*. Washington, D. C.: Joint Publications Research Service.

Kozma, G. 1987. An interlingual typological examination of vowels. *Magyar Füzetek*, 17, 132-162.

Krámský, J. 1990. Phonetics and Phonology. In: Hazai, G. (ed.), *Handbuch der türkischen Sprachwissenschaft 1*, Harrasowitz: Wiesbaden, 302-334.

Krull, D., Traunmüller, H., van Dommelen, W. 2003. The effect of local speaking rate on perceived quantity: a comparison between three languages. *ICPhS,* XV, Barcelona, 1739-1742.

Krumhansl, C. 1978. Concerning the applicability of geometric models to similarity data: The interrelationship between similarity and spatial density. *Psychological Review*, 85 (5), 445-463.

Kruschke, J. 1992. ALCOVE: An exemplar-based connectionist model of category learning. *Psychological Review*, 99 (1), 22-44. [http://cognitrn.psych.indiana.edu/busey/q270/categorization/Kruschke1992.pdf]

Kruschke, J. 2011. Models of attentional learning. In: Pothos, E.M. & Wills, A. (eds.) *Formal Approaches in Categorization*. Cambridge: Cambridge University Press, 120-152. [http://www.indiana.edu/~kruschke/articles/Kruschke2011PWed.pdf]

Kuhl, P. 1991. Human adults and human infants exhibit a 'perceptual magnet effect' for the prototypes of speech sounds, monkeys do not. *Perception and Psychophysics,* 50 (2), 93-107.

Kuhl. P. 1992. Speech prototypes: studies in the nature, function, ontogeny and phylogeny of the 'centers' of speech categories. In: Tohkura, Y., Vatikiotis-Bateson, W., Sagisaka, Y. (eds.). *Speech Perception, Production and Linguistic Structure*. Tokyo: Ohmsha, 197-219.

Kuhl, P. 1993. An examination of the 'perceptual magnet' effect. *JASA*, 93, 2423.

Kuhl, P. 1998a. Effects of language experience on speech perception. *JASA,* 103, 2931.

Kuhl, P. 1998b. The development of speech and language. In: Carew, T., Menzel, R., Shatz, C. (eds.). *Mechanistic Relations between Development and Learning*. Chichester: John Wiley & Sons, 53-74.

Kuhl, P. 2000a. Language, mind, and brain: Experience alters perception. In: Gazzaniga, M. (ed.). *The New Cognitive Neurosciences*, 2nd edition. Cambridge, MA: MIT Press, 99-115.

Kuhl, P. 2000b. A new view of language acquisition. *Proceedings of the National Academy of Sciences of the United States of America*, 97 (22), 11850-11857. [http://www.pnas.org/content/97/22/11850.full]

Kuhl, P., Conboy, B., Coffey-Corina, S., Padden, D., Rivera-Gaxiola, M., Nelson, T. 2008. Phonetic learning as a pathway to language: new data and native language magnet theory expanded (NLM-e). *Philosophical Transactions of the Royal Society Biological Sciences*, vol. 363, no. 1493, 979-1000. [http://rstb.royalsocietypublishing.org/content/363/1493/979.full.pdf]

Kuhl, P. & Iverson, P. 1995. Linguistic experience and the 'perceptual magnet effect'. In: Strange, W. (ed.). *Speech Perception and Linguistic Experience: Issues in Cross-Linguistic Research*. Timonium, MD: York Press, 121-154.

Kuhl, P., Stevens, E., Hayashi, A., Deguchi, T., Kiritani, S., Iverson, P. 2006. Infants show a facilitation effect for native language phonetic perception between 6 and 12 months. *Developmental Science*, 9, F13-F21. [http://ilabs.washington.edu/kuhl/pdf/Kuhl_etal_2006.pdf]

Kuhl, P., Tsao, F.M., Liu, H.M. 2003. Foreign-language experience in infancy: Effects of short-term exposure and social interaction on phonetic learning. *Proceedings of the National Academy of Sciences*, *of the United States of America*, 100, 9096-9106. [http://www.pnas.org/content/100/15/9096.full.pdf+html]

Kuhl, P., Williams, K., Lacerda, F., Stevens, K., Lindblom, B. 1992. Linguistic experiences alter phonetic perception in infants by 6 months of age. *Science*, 255, 606-608.

Kunzmann-Müller, B. 2003. Kroatisch und Serbisch. In: Roelcke, Th. (ed.). *Variationstypologie. Ein sprachtypologisches Handbuch der europäischen Sprachen*. Berlin: de Gruyter, 704-730.

Labov, W. 1966. *The Social Stratification of English in New York City*. Wahington, DC: The Center of Applied Linguistics.

Labov, W. 1972. *Sociolinguistic patterns*. Philadelphia, PA: The University of Pennsylvanian Press.

Labov, W. 1994. *Principles of Linguistic Change, Vol. 1: Internal Factors*. Oxford Cambridge: Blackwell.

Labov, W. 2006. *The Social Stratification of English in New York City*, 2nd edition. Cambridge: Cambridge University Press.

Labov, W. 2010. *Principles of Linguistic Change. Cognitive and Cultural Factors*. vol. 3. Oxford et al.: Wiley-Blackwell.

Labov, W., Ash, S., Boberg, C. 1997. *A National Map of the Regional Dialects of American English*. New York: Mouton de Gruyter. [http://www.atlas.mouton-content.com/]

Labov, W., Ash, S., Boberg, C. 2006. *The Atlas of North American English. Phonetics, Phonology, and Sound Change*. Berlin: Mouton de Gruyter.

Lacerda, F. 1995. The perceptual magnet-effect: An emergent consequence of exemplar-based phonetic memory. In: Ellenius, K. & Branderud, P. (eds.). *ICPhS,* XIII, Stockholm: KTH and Stockholm University, 140-147.

Ladefoged, P. 1982. *A Course in Phonetics*, 2nd edition. New York: Harcourt Brace Jovanovich.

Ladefoged, P. 1993. *A Course in Phonetics*, 3rd edition. Fort Worth: Harcourt Brace Jovanovich.

Ladefoged, P. 1999. American English. In: *Handbook of the International Phonetic Association*, Cambridge: Cambridge University Press, 41-44.

Ladefoged, P. 2001. *Vowels and Consonants. An Introduction to the Sounds of Languages*. Oxford: Blackwell.

Ladefoged, P. 2003. *Phonetic Data Analysis: an Introduction to Fieldwork and Instrumental Techniques*. Oxford: Blackwell.

Ladefoged, P. & Maddieson, I. 1990. Vowels of the world's languages. *JPhon,* 18, 93-122.

Ladefoged, P. & Maddieson, I. 1996. *The Sounds of the World's Languages*. Oxford: Blackwell Publishers.

Lado, R. 1957. *Linguistics Across Cultures*. Ann Arbor: University of Michigan Press.

Lahiri, Aditi. 2007. Non-equivalence between phonology and phonetics. *ICPhS,* XVI, Saarbrücken, 31-34.

Landau, E., Lončarića, M., Horga, D., Škarić, I. 1999. Croatian. In: IPA (ed.) *Handbook of the International Phonetic Association: A guide to the Use of the International Phonetic Alphabet*. Cambridge: Cambridge University Press, 66-69.

Lang, C.E. & Ohala, J.J. 1996. Temporal cues for vowels and universals of vowel inventories. In: *Proceedings of the 4th International Conference on Spoken Language Processing ICSLP*, Wilmington: Philadelphia, 434-437.

Larsen-Freeman, D. & Long, M. 1991. *An Introduction to Second Language Acquisition Research.* London: Longman.

Laver, J. 1994. *Principles of Phonetics*. Cambridge: University Press.

Leather, J. & James, A. (eds.) 1990. *New Sounds 90. Proceedings of the Amsterdam Symposium on Second-Language Speech*. Amsterdam: University of Amsterdam Press.

Leather, J. & James, A. 1991. The acquisition of second language speech. *Studies in Second Language Acquisition*, 13, 305-341.

Leather, J. & James, A. (eds.) 1992. *New Sounds 92. Proceedings of the Second International Symposium on the Acquisition of Second-Language Speech*. Amsterdam: University of Amsterdam Press.

Leather, J. & James, A. (eds.) 1997. *New Sounds 97. Proceedings of the Third International Symposium on the Acquisition of Second-Language Speech*. Klagenfurt: University of Klagenfurt.

Leather, J. (ed.) 1999. *Phonological Issues in Language Learning*. Oxford: Blackwell.

Leather, J. & James, A. (eds.) 2000. *New Sounds 2000. Proceedings of the Fourth International Symposium on the Acquisition of Second-Language Speech*. Klagenfurt: University of Klagenfurt.

Lee, W.-S. & Zee, E. 2003. Illustrations of the IPA: Standard Chinese (Beijing). *Journal of the International Phonetic Association*, 33, 109-112.

Lehiste, I. & Ivić, P. 1963. *Accent in Serbocroatian. An Experimental Study*. Ann Arbor: Department of Languages and Literatures.

Lehiste, I. & Ivic, P. 1986. *Word and Sentence Prosody in Serbocroatian*. Cambridge, MA: MIT Press.

Lehiste, I. 1970. *Suprasegementals*. Masachusetts: MIT Press.

Lehiste, I. 1976. Influence of fundamental frequency pattern on the perception of duration. *JPhon*, 4, 113-117.

Lehnert-LeHouillier, H. 2007a. The influence of dynamic F0 on the perception of vowel duration: Cross-linguistic evidence. *ICPhS,* XVI, Saarbrücken, Germany.

Lehnert-LeHouillier, H. 2007b. *The Perception of Vowel Quantity: a Cross-Linguistic Investigation*. Dissertation. Graduate School of the State University of New York.

Lehnert-LeHouillier, H. 2010. A cross-linguistic investigation of cues to vowel length perception. *JPhon*, 38 (3), 472-482.

Lenneberg, E. 1967. *Biological foundations of language*. New York: John Wiley & Sons.

Levi, S.V. 2005. Acoustic correlates of lexical accent in Turkish. *Journal of the International Phonetic Association*, 35, 73-97.

Levy, E. & Strange, W. 2008. Perception of French vowels by American English adults with and without French Language Experience. *JPhon*, 36, 141-157.

Levy, E. 2009. Language experience and consonantal context effects on perceptual assimilation of French vowels by American English learners of French. *JASA*, 125 (2), 1138-1152. [http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2677367/]

Li, Ch.N. & Thompson, S. 1981. *Mandarin Chinese. A Functional Reference Grammar*. Berkeley Los Angeles: University of California Press.

Liberman, A.M. 1996. *Speech: A Special Code*. Cambridge, MA: MIT Press.

Liberman A.M., Cooper F.S., Shankweiler, D.S., Studdert-Kennedy, M. 1967a. Perception of the speech code. *Psychological Review*, 74, 431-461. [http://www.haskins.yale.edu/Reprints/HL0069.pdf]

Liberman, A.M., Cooper, F.S., Shankweiler, D.P., Studdert-Kennedy, M. 1967b. Language beyond our grasp. What mirror neurons can and cannot do for the evolution of language. *Psychological Review*, 74, 431-461.

Liberman, A.M., Harris, K., Hoffman, H., Griffith, B. 1957. The discrimination of speech sounds within and across phoneme boundaries. *Journal of Experimental Psychology*, 54, 358-368.

Liberman, A.M., Harris, K., Kinney, J., Lane, H. 1961. The discrimination of relative onset time of the components of certain speech and nonspeech patterns. *Journal of Experimental Psychology,* 61, 379-388.

Liberman, A.M. & Mattingly, I. 1985. The motor theory of speech perception revised. *Cognition*, 21 (1), 1-36.

Liberman, A.M. & Mattingly, I. 1989. A specialization for speech perception. *Science*, 243, 489-494.

Lightbown, P. & Spada, N. 1999. *How Languages are learned*. Oxford: Oxford University Press.

Liljencrants, J. & Lindblom, B. 1972. Numerical simulation of vowel quality systems: the role of perceptual contrast. *Language*, 48, 839-862.

Lin, Y.W. 2007. *The Sounds of Chinese*. Cambridge: Cambridge University Press.

Lindau, M. 1978. Vowel features. *Language,* 54, 541-560.

Lindblom, B. 1983. Economy of speech gestures. In: MacNeilage, P.F. (ed.). *The Production of Speech*. New York: Springer, 217-245.

Lindblom, B. 1986. Phonetic universals in vowel systems. In: Ohala, J. & Jaeger, J. (eds). *Experimental Phonology*. Orlando: Academic Press, 13-44.

Lindblom, B. 1990. Explaining phonetic variation: a sketch of the H&H theory. In: *Speech Production and Speech Modelling NATO ASI Series*, 55, 403-439. [http://link.springer.com/chapter/10.1007%2F978-94-009-2037-8_16#page-1]

Lindblom, B., Diehl, R., Creeger, C. 2009. Do ‚Dominant Frequencies' explain the listeners's response to formant and spectrum shape variations? *Speech Communication*, 51 (7), 622-629. [http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2760739/pdf/nihms129947.pdf]

Lindblom, B. & Sundberg, J. 1971. Acoustical consequences of lip, tongue, jaw, and larynx movement. *JASA*, 50, 1166-1179.

Lindner, G. 1966. Beurteilung synthetisch erzeugter vokalartiger Klänge durch deutschsprachige Hörer. *Zeitschrift für Phonetik, Sprachwissenschaft und Kommunikationswissenschaft*, 19, 45-90.

Linell, P. 1982. The concept of phonological form and the activities of speech production and speech perception. *JPhon,* 10, 37-72.

Linker, W. 1982. Articulatory and Acoustic Correlates of Labial Activity in Vowels: A Cross-Linguistic Study. *Working Papers in Phonetics*, 56, University of California, Los Angeles, CA. [http://escholarship.org/uc/item/0wq546xq#page-1]

Llisteri, J. 1995. Relationships between speech production and speech perception in a second language. *ICPhS,* XIII, Stockholm, vol. 4, 92-99. [http://liceu.uab.es/~joaquim/publicacions/LListerri_95_L2_Production_Perception.pdf]

Lloshi, Xh. 1999. Albanian. In: Hinrichs, U. (ed.). *Handbuch der Südosteuropa-Linguistik*. Wiesbaden: Harrasowitz, 277-300.

Löfqvist, A., Baer, T., McGarr, N., Story, R.S. 1989. The cricothyroid muscle in voicing control. *JASA*, 85, 1314-1321.

Long, M. 1990. Maturational Constraints on Language Development. *Studies in Second Language Acquisition*, 12, 251-285.

Luce, R.D. 1963. Detection and recognition. In: Luce, R.D., Bush, R.R., Galanter, E. (eds.) *Handbook of Mathematical Psychology, I*. New York: Wiley,103-189. [http://www.socsci.uci.edu/~rdluce/bio/pre1990/1963/Luce_Book%20Chapter_1963.pdf]

Luce, R.D. & Green, D. 1974. Detection, discrimination, and recognitions. Reprinted from *Handbook of Perception, Vol. 2: Psychophysical Judgment and Measurement*. New York et al.: Academic Press, 300-342. [http://www.socsci.uci.edu/~rdluce/bio/pre1990/1974/LuceGreen_Book%20Chapter_1974b.pdf]

Maas, U. 1999. Phonologie. *Einführung in die funktionale Phonetik des Deutschen*. Göttingen: Vandenhoeck & Ruprecht.

MacKay, I. & Flege, J. 2004. Effects of the age of second-language (L2) learning on the duration of L1 and L2 sentences: The role of suppression. *Applied Psycholinguistics*, 25, 373-396.

Macmillan, N., Goldberg, R., Braida, L. 1988. Resolution for speech sounds: Basic sensitivity and context memory on vowel and consonant continua. *JASA*, 84, 1262-1280.

Maddieson, I. 1984. *Patterns of Sounds*. Cambridge: Cambridge University Press.

Maddieson, I. 1997. Phonetic universals. In: Hardcastle, W. & Laver, J. (eds.). *The Handbook of Phonetic Sciences*, Oxford: Blackwell, 619-639.

Maddieson, I. 1999. In search of universals. In: Ohala, J., Hasegawa, Y., Ohala, M., Granville, D., Bailey, A. (eds.). *ICPhS,* XIV, San Francisco, vol. 3, 2521-2528.

Maddieson, I. 2008a. Vowel quality inventories. In: Haspelmath, M., Dryer, M., Gil, D., Comrie, B. (eds.). *The World Atlas of Language Structures*, *Vol. 1*. Oxford: Oxford University Press. Or in: Dryer, M. & Haspelmath, M. (eds.). *The World Atlas of Language Structures Online*. Munich: Max Planck Digital Library. [available online at http://wals.info/chapter/2]

Maddieson, I. 2008b. Front rounded vowels. In: Haspelmath, M., Dryer, M., Gil, D., Comrie, B. (eds.). *The World Atlas of Language Structures*, *Vol. 1*. Oxford: Oxford University Press, 50-54.

Maddieson, I. 2013. Front rounded vowels. In: Dryer, Matthew S. & Haspelmath, Martin (eds.) *The World Atlas of Language Structures Online*. Leipzig: Max Planck Institute for Evolutionary Anthropology. [http://wals.info/chapter/11]

Maddieson, I. & Precoda, K. 1990. Updating UPSID. *UCLA Working Papers in Phonetics*, 74, 104-111. [http://escholarship.org/uc/item/71s1701m#page-3]

Maddox, W. & Ashby, G. 1993. Comparing decision bound and exemplar models of categorization. *Perception and Psychophysics*, 53 (1), 49-70. [http://homepage.psy.utexas.edu/HomePage/Group/MaddoxLAB/Publications/1990-1994/maddox%20and%20ashby%201993.pdf]

Mády, K. 2001. Kontrastive Phonetik deutsch-ungarisch im Hinblick auf zu erwartenden Interferenzphänomene. In: Szigeti, I. (ed.). *Germanistische Linguistik aus dem Ambrosianum. Festschrift for György Hell*. Pilicsaba, 30-51. [http://www.phonetik.uni-muenchen.de/~mady/pub/mady_ppke_2001.pdf]

Mády, K. 2010. Shortening of long high vowels in Hungarian: a perceptual loss? *Proceedings of the Workshop "Sociophonetics at the Crossroads of Speech Perception, Processing and Communication"*. Pisa, December 2010. [http://edizioni.sns.it/it/downloadable/download/sample/sample_id/12/]

Magner, T. 1998. *Introduction to the Croatian and Serbian Language*. University Park: Pennsylvania State University Press.

Magner, T. & Matejka, L. 1971. *Word Accent in Modern Serbo-Croatian*. University Park: Pennsylvania State University Press.

Maher, M. 1983. Schwerpunkte einer deutsch-arabischen kontrastiven Grammatik. *Jahrbuch Deutsch als Fremdsprache*, 9, 164-180.

Major, R. 1987a. A model for interlanguage phonology. In: Ioup, G. & Weinberger, S. (eds.). *Interlanguage Phonology: The Acquisition of a Second Language Sound System*. New York: Newbury House/Harper & Row, 101-124.

Major, R. 1987b. English voiceless stops production by Brazilian speakers of English. *JPhon*, 15, 197-202.

Major, R. 1987c. Phonological Similarity, Markedness, and Rate of L2 Acquisition. *Studies in Second Language Acquisition*, 9 (1), 63-82.

Major, R. 1994. Current Trends in Interlanguage Phonology. In: Yavas, M. (ed.). *First and Second Language Phonology*. San Diego: Singular Publishing Group, 181-204.

Major, R. 2001. *Foreign Accent. The Ontogeny and Phylogeny of Second Language Phonology*. Mahwah, NJ: Lawrence Erlbaum.

Major, R. 2008. Transfer in second language phonology. A review. In: Hansen Edwards, J. & Zampini, M. (eds.). *Phonology and Second Language Acquisition*. Philadelphia: John Benjamins, 63-94.

Major, R. & Kim, E. 1999. The Similarity Differential Rate Hypothesis. *Language Learning*, 49, 151-183.

Mallinson, G. 1986. *Rumanian*. London et al.: Croom Helm.

Mallinson, G. 1990. Rumanian. In: Harris, M. & Vincent, N. (eds.). *The Romance Languages.* Oxford University Press, 391-419.

Mallinson, G. 2011. Rumanian. In: Comrie, B. (ed.). *The World's Major Languages*, 2nd edition. London New York: Routledge, 253-268.

Mannel, R. 2008. *Resolving vowel formants in Australian English*. Ms. [http://clas.mq.edu.au/perception/speechperception/resolving_vowel_formants.html]

Markel J. & Gray, A. 1976. *Linear Prediction of Speech*; Berlin Heidelberg New York: Springer.

Markham, D. 1997. *Phonetic Imitation, Accent, and the Learner*. PhD Dissertation. Travaux de l'Institut de Linguistique de Lund, 33. Lund University Press.

Markman, A. & Gentner, D. 1993. Structural alignment during similarity comparisons. *Cognitive Psychology*, 25, 431-467.

Massaro, D. 1987a. Categorical partition: A fuzzy logical model of categorical behavior. In: Harnad, S. (ed.). *Categorical Perception: The Groundwork of Cognition*. Cambridge: Cambridge University Press, 254-283.

Massaro, D. 1987b. *Speech Perception by Ear and Eye: A Paradigm for Psychological Inquiry*. Hillsdale, NJ: Erlbaum.

Mattingly, I. & Studdert-Kennedy, M. 1991. *Modularity and the Motor Theory of Speech Perception*. Hillsdale: Lawrence Erlbaum.

Maye, J. & Gerken, L.2000. Learning phonemes without minimal pairs. *Proceedings of the 24th Annual Boston University Conference on Language Development*, 522-533.

Mayo, L., Florentine, M., Buus, S. 1997. Age of second-language acquisition and perception of speech in noise. *Journal of Speech, Language, and Hearing Research*, 40, 686-693.

Mayr, R. 2010. What exactly is a front rounded vowel? An acoustic and articulatory investigation of the NURSE vowel in South Wales English. *Journal of the International Phonetic Association,* 40 (1), 93-112.

McAllister, R. 1991. Perceptual foreign accent: L2 user's comprehension ability. *PERILUS,* 12, 55-67.

McAllister, R. 1996. Aspects of the relation between the production and perception of a second language. *TMH-QPSR*, 37 (2), 69-72. [http://www.speech.kth.se/prod/publications/files/qpsr/1996/1996_37_2_069-072.pdf]

McAllister, R. 1998. Second language perception and the concept of foreign accent. In: *STiLL Speech Technology in Language Learning*, 1998, 155-158. [http://www.isca-speech.org/archive_open/archive_papers/still98/stl8_155.pdf]

McAllister, R., Flege, J., Piske, T. 2002. The influence of L1 on the acquisition of Swedish quantity by native speakers of English, Spanish and Estonian. *JPhon,* 30, 229-258.

McArthur, T. 2002. *The Oxford Guide to World English*. Oxford: Oxford University Press.

McBride-Chang, C. 1995. What is phonological awareness? *Journal of Educational Psychology*, 87, 179-192.

McCarthy, J. 1994. The phonetics and phonology of Semitic pharyngeals. *Linguistic Department Faculty Publication Series*, Paper 86. [http://scholarworks.umass.edu/linguist_faculty_pubs/86]

McCarthy, J. 2002. *A Thematic Guide to Optimality Theory*. Cambridge: Cambridge University Press.

McClelland, J. & Elman, J. 1986. The trace model of speech perception. *Cognitive Psychology*, 18, 1-86.

Medin, D. & Aguilar, C. 1999. Categorization. In: Wilson, R.A. & Keil, F. (eds.) *The MIT Encyclopedia of the Cognitive Sciences (MITECS)*, Cambridge: MIT Press, 104-106. [http://ai.ato.ms/MITECS/Entry/medin.html]

Medin, D. & Schaffer, M. 1978. Context theory of classification learning. *Psychological Review*, 85, 207-238.

Meier, Helmut 1967. *Dt. Sprachstatistik*. Hildesheim: Olms.

Mermelstein, P. 1976. *Distance measures for speech recognition – perceptional and instrumental. Status Report on Speech Research*, 47, Yale: Haskins Laboratories [http://www.haskins.yale.edu/sr/SR047/SR047_07.pdf]

Meunier, F. & Segui, J. 2002. Cross-modal morphological priming in French. *Brain and Language*, 81, 89-102.

Mielke, J. 2002. Turkish /h/ deletion: evidence for the interplay of speech perception and phonology. *ZAS Papers in Linguistics,* 28, 55-72.

Mielke, J. 2003. The Interplay of Speech Perception and Phonology: Experimental Evidence from Turkish. *Phonetica*, 60 (3), 208-229

Mielke, J. 2008. *The Emergence of Distinctive Features*. Oxford: Oxford University Press.

Miller, J. 1981. Effects of speaking rate on segmental distinctions. In: Eimas, P. & Miller, J. (eds.). *Perspectives on the Study of Speech*. Hillsdale: Lawrence Erlbaum, 39-74.

Miller, J. 1989. Auditory-perceptual interpretation of the vowel. *JASA*, 85, 2114-2134.

Miller, J. 1994. On the internal structure of phonetic categories: a process report. *Cognition*, 50, 271-285.

Miller, J. & Eimas, P. 1983. Studies on the categorization of speech by infants. *Cognition*, 13 (2), 135-165.

Mitchell, A. & Delbridge, A. 1965. *The Pronunciation of English in Australia*. Sydney: Angus & Robertson.

Mohanan, K. 1993. Fields of attraction in phonology. In: Goldsmith, J. (ed.). *The Last Phonological Rule – Reflections on Constraints and Derivation*. Chicago: University of Chicago Press, 61-116.

Moore, B. 1997. Aspects of auditory processing related to speech perception. In: Hardcastle, W. & Laver, J. (eds.). *The Handbook of Phonetic Sciences*. Oxford: Blackwell, 539-565.

Moore, B. 2008. Basic auditory processes involved in the analysis of speech sounds. *Philosophical Transactions of the Royal Society*, B, 363, 947-963. [http://rstb.royalsocietypublishing.org/content/363/1493/947.full.pdf]

Mooshammer, C. 1998. *Experimentalphonetische Untersuchungen zur artikulatorischen Modellierung der Gespanntheitsopposition im Deutschen* [Diss. HU Berlin]. Forschungsberichte des Instituts für Phonetik und Sprachliche Kommunikation der Universität München (FIPKM) 36, 3-192.

Mooshammer, C. 2000. *Is Tenseness in German a segmental feature or a property of the syllable? An experimental study*. Ms. [http://www.haskins.yale.edu/staff/mooshammer/edinburgh_syll.pdf, accessed 2012-03-03]

Mooshammer, C., Hoole, P., Alfonso, P., Fuchs, S. 2001. Intrinsic pitch in German: A puzzle? *JASA*, 110, 2761 (A).

Mooshammer, C., Perrier, P., Fuchs, S., Geng, C., Payan Y. 2001. The control of token-to-token variability: An experimental und modelling study. *Proceedings of the 4th International Speech Motor Conference*, 78-81. Nijmegen. [http://www.haskins.yale.edu/staff/mooshammer/Mooshammer_et_alNij.pdf,]

Moosmüller, S. 2007. *Vowels in Standard Austrian German. An Acoustic Phonetic and Phonological Analysis*. Habilitationsschrift. Universität Wien.

Moosmüller, S. & Granser, T. 2001. The schwa in Albanian. *Proceedings of the 7th European Conference on Speech Communication and Technology, Eurospeech 2001,* 317-320. [http://perso.telecom-paristech.fr/~chollet/Biblio/Congres/Audio/Eurospeech01/CDROM/papers/page317.pdf]

Moosmüller, S. & Granser, T. 2003. The Vowels of Standard Albanian. *ICPhS* XV, Barcelona, 659-662. [http://www.kfs.oeaw.ac.at/publications/2003_moosmueller_granser_the_vowels_of_standard_albanian.pdf]

Moosmüller, S. & Granser, T. 2006. The spread of Standard Albanian: An illustration based on an analysis of vovels. *Language, Variation and Change*, 18 (2), 121-140.

Moravcsik, E. & Wirth, J. 1986. Markedness. An overview. In: Eckman, F., Moravcsik, E., Wirth, J. (eds.). *Markedness*. New York: Plenum Press, 1-11.

Morciniec, N. 1979. Review of Prędota 1978. In: *Kwartalnik Neofilologiczny,* 25 (3), 407-412.

Morciniec, N. & Prędota, S. 1973. *Fonetyka kontrastywna języka niemieckiego*. Wrocław/Warszawa: PWN.

Morciniec, N. & Prędota, S. 1982. *Podręcznik wymowy niemieckiej*. Warszawa: PWN.

Morrison, G. & Assmann, P. (eds.) 2013. *Vowel Inherent Spectral Change. Modern Acoustics and Signal Processing*. Berlin Heidelberg: Springer.

Moulton, W. 1962. *The Sounds of English and German. A Systematic Analysis of the Contrasts between the Sound Systems*. Chicago London: University of Chicago Press.

Moulton, W. 1974. *The Sounds of English and German. A Systematic Analysis of the Contrasts between the Sound Systems*, 7th impr. Chicago: University of Chicago Press.

Moyer, A. 1999. Ultimate Attainment in L2 Phonology: The Critical Factors of Age, Motivation, and Instruction. *Studies in Second Language Acquisition,* 21, 81-108.

Moyer, A. 2004. *Age, Accent and Experience in Second Language Acquisition*. Clevedon, UK: Multilingual Matters.

Müller, U. & Chudoba, G. 1999. Serbokroatisch. Ms. publ. as „Serbisch (mit Bosnisch und Kroatisch)“. In: Hirschfeld, U., Kelz, H., Müller, U. (eds.) (2003ff). *Phonetik international: Von Afrikaans bis Zulu. Kontrastive Studien für Deutsch als Fremdsprache.* Waldsteinberg: Heidrun Popp. [http://www.phonetik-international.de/]

Mumm, P.A. 2009. *Strukturkurs Neupersisch*. Ms. [http://www.indogermanistik.uni-muenchen.de/downloads/publikationen/publ_mumm/struktur_persisch.pdf]

Muñoz, C. 2011. Input and long-term effects of starting age in foreign language learning. *International Review of Applied Linguistics in Language Teaching*, 49 (2), 113-133.

Muñoz, C. & Singleton, D. 2007. Foreign accent in advanced learners. *EUROSLA Yearbook*, 7, 171-190.

Munro, M. 1992. *Production and Perception of English Vowels by Native Speakers of Arabic*. Unpubl. PhD Dissertation. University of Alberta, Edmonton, Alberta.

Munro, M. 1993. Production of English vowels by native speakers of Arabic: Acoustic measurements and accentedness ratings. *Language and Speech,* 36, 39-66.

Munro, M. 2008. Foreign accent and speech intelligibility. In: Hansen Edwards, J. & Zampini, M.(eds.). *Phonology and Second Language Acquisition*. Amsterdam: John Benjamins, 193-218.

Munro, M. & Derwing, T. 1995. Processing time, accent, and comprehensibility in the perception of native and foreign-accented speech. *Language and Speech*, 38, 289-306.

Munro, M. & Derwing, T. 1999. Foreign Accent, Comprehensibility, and Intelligibility in the Speech of Second Language Learners. *Language Learning*, 49 (1), 285-310.

Näätänen R. 2001. The perception of speech sounds by the human brain as reflected by the mismatch negativity (MMN) and its magnetic equivalent (MMNm). *Psychophysiology*, 38, 1-21.

Näätänen, R., Lehtokoski, A., Lennes, M., Cheour, M., Huotilainen, M., Iivonen, A., Vainio, M., Alku, P., Ilmoniemi, R.J., Luuk, A., Allik, J., Sinkkonen, J., Alho, K. 1997. Language-specific phoneme representations revealed by electric and magnetic brain responses. *Nature*, 385, 432-434.

Näätänen, R., Paavilainen, P., Rinne, T., Alho, K. 2007. The mismatch negativity (MMN) in basic research of central auditory processing: A review. *Clinical Neurophysiology*, 118, 2544-2590. [http://web.mit.edu/carrien/Public/speechlab/papers/naatanen-paavilainen-rinne-alho-2007.pdf]

Nathan, G. & Donegan, P. (forthcoming). Natural phonology and sound change. In: Honeybone, P. & Salmons, J.C. (eds.). *The Handbook of Historical Phonology*. Oxford: Oxford University Press. [http://www.ling.hawaii.edu/faculty/donegan/Papers/201Xhistphon.pdf]

Naylor, K.E. 1980. Serbo-Croatian. In: Schenker, A.M. & Stankiewicz, E. (eds.). *The Slavic Literary Languages: Formation and Development*. New Haven: Yale Concilium on International and Area Studies, 65-83.

Nearey, T. & Assmann, P. 1986. Modeling the role of inherent spectral change in vowel identification. *JASA*, 80, 1297-1308.

Nearey, T. 1989. Static, dynamic, and relational properties in vowel perception. *Journal of the Acoustical Society of America*, 85, 2088-2113. [http://www.utdallas.edu/~assmann/aud6306/nearey89.pdf]

Neel, A. 2008. Vowel space characteristics and vowel identification accuracy. *Journal of Speech, Language, and Hearing Research*, 51, 574-585. [http://jslhr.asha.org/cgi/content/full/51/3/574]

Newman, D. & Verhoeven, J. 2002. Frequency analysis of Arabic vowels in connected speech. *Antwerp Papers in Linguistics,* 100, 77-86. [http://www.dur.ac.uk/daniel.newman/Apil1.pdf]

Newman, D. 2002. The phonetic status of Arabic within the world's languages: the uniqueness of the lughat al-daad. *Antwerp Papers in Linguistics*, 100, 65-75. [http://dro.dur.ac.uk/4420/1/35759.pdf]

Newmark, L., Hubbard, P., Prifti, P. 1982. *Standard Albanian*. Stanford: Stanford University Press.

Niedersächsisches Kultusministerium, 1991ff. *Hilfen für den Deutschunterricht mit Kindern und Jugendlichen, deren Erstsprache nicht Deutsch ist*. Auszug aus der Ausgabe von 1991. [http://www.nibis.de/nli1/fid/pdf/Hilfen%20fuer%20den%20Deutschunterricht1996Sprachvergleiche.pdf]

Nittrouer, S. 2001. Challenging the notion of innate phonetic boundaries. *JASA*, 110 (3), 1598-1605.

Nittrouer, S. 2002. A reply to "Innate phonetic boundaries revisited". *JASA*, 112 (4), 1261-1264.

Nittrouer, S. & Miller, M. 1997. Predicting developmental shifts in perceptual weighting schemes. *JASA*, 101, 2253-2266.

Noll, A. 2001. *STX - A Programming and Execution Environment for Signal-Processing Applications and Acoustical Experiments. Reference Manual*. Wien: Austrian Academy of Science.

Noll, A., Deutsch, W., Balazs, P., White, J. 2001. *Intelligent Sound Processing S_TOOLS - STX User's Guide*. Wien: Austrian Academy of Science. [http://www.kfs.oeaw.ac.at/publications/2001_noll_deutsch_balazs_white_stx_manual.pdf]

Norušis, M. 2007. *SPSS 15.0. Advanced Statistical Procedures Companion*. Upper Saddle River: Prentice Hall.

Nosofsky, R. 1986. Attention, similarity, and the identification-categorization relationship. *Journal of Experimental Psychology: General*, 115 (1), 39-57. [http://www.cogs.indiana.edu/nosofsky/pubs/1986_rmn_jep-g_attention.pdf]

Nosofksy, R. 1987. Attention and learning processes in the identification and categorization of integral stimuli. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 13 (1), 87-108. [http://psych.stanford.edu/~jlm/pdfs/Nosofsky87GCM.pdf]

Nosofksy, R. 1988. Similarity, frequency, and category representations. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 14 (1), 54-65.

Nosofsky, R. 1991. Stimulus bias, asymmetric similarity, and classification. *Cognitive Psychology*, 23, 94-140. [http://www.cogs.indiana.edu/nosofsky/pubs/1991_rmn_cs_Stimulus.pdf]

Nosofksi, R. & Zaki, S. 2002. Exemplar and prototype models revisited: Response strategies, selective attention, and stimulus generalization. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 28 (5), 924-940.

Nowak, P. 2006. *Vowel Reduction in Polish*. Dissertation. University of California. [http://www.linguistics.berkeley.edu/phonlab/annual_report/documents/2006/Nowak_dissertation_2006.pdf]

Odlin, T. 1989. *Language Transfer: Cross-Linguistic Influence in Language Learning*. Cambridge: Cambridge University Press.

Odlin, T. 2005. Cross-linguistic influence. In: Doughty, C. & Long, M. (eds.). *The Handbook of Second Language Acquisition*. Malden: Blackwell, 436-486.

Oflazer, K. & Inkelas, S. 2003. *A Finite State Pronunciation Lexicon for Turkish*. Ms. [http://people.sabanciuniv.edu/oflazer/archives/papers/pronlex.pdf]

Ohala, J. 1981. The listener as a source of sound change. In: C. S. Masek, R. A. Hendrick, & M. F. Miller (eds.), *Papers from the Parasession on Language and Behavior*. Chicago: Chicago Ling. Soc. 178-203. [http://linguistics.berkeley.edu/~ohala/papers/listener_as_source.pdf]

Ohala, J. 1983. The origin of sound patterns in vocal tract constraints. In: MacNeilage, P.F. (ed.). *The Production of Speech*. New York: Springer, 189-216.

Ohala, J. 1989. Discussion of Lindblom's 'Phonetic invariance and the adaptive nature of speech'. In: Elsendoorn, B.A.G. & Bouma, H. (eds.). *Working Models of Human Perception*. London: Academic Press, 175-183. [http://linguistics.berkeley.edu/~ohala/papers/comments_on_lindblom.pdf]

Ohala, J. 1993a. Coarticulation and Phonology. *Language and Speech*, 36, 155-170.

Ohala, J. 1993b. The phonetics of sound change. In: Jones, Ch. (ed.). *Historical Linguistics: Problems and Perspectives*. London: Longman, 237-278.

Ohala, J. 1996. Speech perception is hearing sounds, not tongues. *JASA*, 99 (3), 1718-1725. [http://linguistics.berkeley.edu/~ohala/papers/hearing_sounds.pdf]

Olsson, M. 1992. Hungarian Phonology and Morphology. *Travaux de l'Institut de Linguistique de Lund*. Lund: University Press.

O'Malley, J. & Chamot, A. 1990. *Learning Strategies in Second Language Acquisition.* Cambridge: Cambridge University Press.

Ortmann, W.D. (ed.) 1976. *Lernschwierigkeiten der deutschen Aussprache. Ergebnisse einer Befragung von Lehrern für Deutsch als Fremdsprache*. Goethe-Institut München, Arbeitsstelle für wissenschaftliche Didaktik, Projekt Phonothek, 3 Bde., München: Goethe Institut.

Oturan, K. 2001. Formantfrequenzwerte der türkischen Vokale: Eine statistische Untersuchung der ersten zwei Formantfrequenzwerte der Vokale von drei Sprechern des Türkischen. *Anadolu University Journal of Social Sciences*, vol. 1 (1), 147-168.

Oturan, K. 2002. *Vokalperzeption und -produktion in der Zweitspache: eine Untersuchung mit türkischen Muttersprachlern*. Aachen: Shaker.

Oxford, R. 1990. *Language Learning Strategies: What every teacher should know*. Boston: Heinle & Heinle.

Özen, E. 1985. *Untersuchungen zu einer kontrastiven Phonetik Türkisch-Deutsch*. Hamburg: Buske.

Paikeday, T. 2003. *The Native Speaker Is Dead!*. Toronto & New York: Lexicography, Inc. [http://www.paikeday.net/speaker.pdf]

Pallier, C., Colomé, A., Sebastián-Gallés, N. 2001. The influence of native-language phonology on lexical access: Concrete exemplar-based vs. abstract lexical entries. *Psychological Science*, 12, 445-449.

Pape, D. & Mooshammer, C. 2004. Intrinsic pitch in German – Examining the whole fundamental frequency contour of the vowel. *Proceedings of the CFA/DAGA*, Strasbourg, France, 897-898. [http://www.haskins.yale.edu/staff/mooshammer/pape_daga_2004.pdf]

Pape, D. & Mooshammer, C. 2006. Is intrinsic pitch language-dependent? Evidence from a cross-linguistic vowel pitch perception experiment. *Proceedings of the ISCA Tutorial and Research Workshop (ITRW) on Multilingual Speech and Language Processing,* Stellenbosch, South Africa. [http://www.haskins.yale.edu/staff/mooshammer/pape_mul.pdf]

Pape, D. 2009. *Microprosodic Differences in a Cross-Linguistic Vowel Comparison of Speech Production and Speech Perception*. Berlin: Weißensee.

Pape, D., Mooshammer C., Fuchs, S., Hoole, P. (2005). Intrinsic pitch differences between German vowels /i:/, /I/ and /y:/ in a cross-linguistic perception experiment. *Proceedings of the ISCA Workshop on Plasticity in Speech Perception* (PSP 2005), London, UK. [http://www.haskins.yale.edu/staff/mooshammer/PapePSP05_london.pdf]

Parker, St. 2002. *Quantifying the Sonority Hierarchy*. Doctoral Dissertation. University of Masachusetts Amherst. [http://www.ai.mit.edu/projects/dm/theses/more/parker02.pdf]

Pater, J. 1998. *From Phonological Typology to the Development of Receptive and Productive Phonological Competence: Applications of minimal Violation*. Rutgers Optimality Archive. [http://roa.rutgers.edu/files/296-0299/roa-296-pater-1.pdf]

Patkowski, M. 1980. The sensitive period for the acquisition of syntax in a second language. *Language Learning*, 30, 449-472.

Patkowski, M. 1990. Age and accent in a second language: A reply to James Emil Flege. *Applied Linguistics*, 11, 73-89.

Penfield, W. & Roberts, L. 1959. *Speech and Brain Mechanisms*. Princeton, NJ: Princeton University Press.

Perry, J.R. 2005. *A Tajik Persian Reference Grammar*. Leiden: Brill.

Peterson, G. & Barney, H. 1952. Control methods used in a study of the vowels. *JASA*, 24 (2), 175-184.

Petrovic, E. 1956. La fonction phonologique des „diphthongues“ roumaines ^{e}a, ^{i}a et leur rapport avec les phonèmes consonantiques palatalisés. *Révue de Linguistique,* 1, 25-28.

Phanintra, T. 2007. The intrinsic pitch and intrinsic length of high and low vowels in Mon-Khmer language. *Mon-Khmer Studies*, 37, 41-52. [http://sealang.net/sala/archives/pdf8/phanintra2007intrinsic.pdf]

Pickett, J.M. 1999. *The Acoustics of Speech Comunication. Fundamentals, Speech Perception Theory, and Technology*. Boston et al.: Allyn & Bacon.

Pierrehumbert, J. 2000. The phonetic grounding of phonology, *Bulletin de la Communication Parlee,* 5, 7-23. [http://faculty.wcas.northwestern.edu/~jbp/publications/phonetic_grounding.pdf]

Pierrehumbert, J. 2001. Exemplar-dynamics: Word-frequency, lenition and contrast. In: Bybee, J. & Hopper, P. (eds.). *Frequency and the Emergence of Linguistic Structure*. Philadelphia: John Benjamins, 137-157.

Pierrehumbert, J. 2002. Word-specific phonetics. *Laboratory Phonology,* 7. Berlin New York: de Gruyter, 101-139.

Pierrehumbert, J. 2003. Phonetic diversity, statistical learning, and acquisition of phonology. *Language and Speech*, 46 (2-3), 115-154. [http://faculty.wcas.northwestern.edu/~jbp/publications/phonetic_diversity.pdf]

Pike, K.L. 1943. *Phonetics: A Critical Analysis of Phonetic Theory and a Technique for the Practical Description of Sounds*. Ann Arbor: University of Michigan Press.

Piske, T. 2008. Phonetic awareness, phonetic sensitivity and the second language learner. In: Cenoz, J. & Hornberger, N. (eds.). *Encyclopedia of Language and Education, Vol. 6: Knowledge about Language*, 2nd edition. Berlin: Springer, 155-166.

Piske, T., Flege, J, MacKay, I., Meador, D. 2002. The production of English vowels by fluent early and late bilinguals. *Phonetica*, 59, 49-71.

Piske, T., MacKay, I., Flege, J. 2001. Factors affecting degree of foreign accent in an L2: A review. *JPhon*, 29, 191-215.

Pisoni, D. 1971. On the nature of categorical perception of speech sounds. *Status Report on Speech Research*, 27, (Supplement). New Haven: Haskins Laboratories.

Pisoni, D. 1973. Auditory and phonetic codes in the discrimination of consonants and vowels. *Perception & Psychophysics*, 13, 253-260.

Pisoni, D. 1976. Fundamental frequency and perceived vowel duration. *JASA*, 59, S39.

Pöchtrager, M. 2010. Does Turkish Diss Harmony? *Acta Linguistica Hungarica*, 57 (4), 458-473.

Polivanov, E.D. 1931. La perception des sons d’une langue étrangère. *Travaux du Cercle Linguistique de Prague* 4, 79-96.

Polka, L. 1995. Linguistic influences in adult perception of non-native vowel contrasts. *JASA*, 97, 1286-1296.

Polka, L. & Bohn, O.-S. 1996. A cross-language comparison of vowel perception in English-learning and German-learning infants. *JASA*, 100, 577-592.

Polka, L. & Bohn, O.-S. 2003. Asymmetries in vowel perception. *Speech Communication*, 41, 221-231.

Polka, L. & Bohn, O.-S. 2011. Natural Referent Vowel (NRV) framework: An emerging view of early phonetic development. *JPhon*, 39 (4), 467-478.

Polka, L. & Werker, J. 1994. Developmental changes in perception of nonnative vowel contrasts. *Journal of Experimental Psychology: Human Perception and Performance*, 20, 421-435.

Polka, L., Juszyk, P., Rvachew, S. 1995. Methods for studying speech perception in infants and children. In: Strange, W. (ed.). *Speech Perception and Linguistic Experience.* Timonium, MD: York Press, 49-89.

Pouplier, M., Buchwald, A., Stone, M. 2004. A tagged cine-MRI and ultrasound study of the German vowel system. *Working Papers and Technical Reports, Vocal Tract Visualization Laboratory*, 6 (University of Maryland Dental School, Baltimore), 18-55. [http://www.phonetik.uni-muenchen.de/~pouplier/Vowel_paper.pdf]

Pouplier, M. & Stone, M. 2005. A tagged Cine-MRI investigation of German vowels. *Proceedings of the 9th European Conference on Speech Communication and Technology (INTERSPEECH 2005, Lisbon),* 2157-2160. [http://www.phonetik.uni-muenchen.de/~pouplier/1670anav.pdf]

Prędota, S. 1979. Die polnisch-deutsche Interferenz im Bereich der Aussprache. (Prace Wrocławskiego Towarzystwa Naukowego, Seria A nr 194), Wrocław: Ossolineum 1979.

Prince, A. & Smolensky, P. 1997. Optimality theory: from neural networks to universal grammar. *Science*, 275 (5306), 1604-1610.

Prince, A. & Smolensky, P. 2004. *Optimality Theory*. Oxford: Blackwell.

Proctor, M. 2007. The Organisation of Phonological Inventories. In: Trouvain, J. & Barry, W.J. (eds.). *ICPhS, XVI*, Saarbrücken, 1377-1380.

Ramers, K.H. 1988. *Vokalquantität und -qualität im Deutschen*. Tübingen: Niemeyer.

Ramus, F., Nespor, M., Mehler, J. 1999. Correlates of linguistic rhythm in the speech signal. *Cognition*, 72, 1-28. [http://www.lscp.net/persons/ramus/docs/Cognition99.pdf]

Rastegar-El-Zarka, D. 1997. *Prosodische Phonologie des Arabischen*. Dissertation, Universität Graz.

Rauber, A., Watkins, M., Baptista, B. (eds.) 2007. *New Sounds 2007. Proceedings of the Fifth International Symposium on the Acquisition of Second Language Speech*. Florianópolis, Brazil: Federal University of Santa Catarina.

Rebuschat, P. & Williams, J. (eds.) 2012. *Statistical Learning and Language Acquisition*. New York: de Gruyter.

Reinholt-Peterson, N. 1986. Perceptual compensation for segmentally-conditioned fundamental-frequency perturbations. *Phonetica*, 43, 31-42.

Rennison, J. 1990. On the elements of phonological representations: the evidence from vowel systems and vowel processes. *Folia Linguistica*, 24, 175-244.

Renwick, M. 2011. Phoneme type frequency in Romanian. *University of Pennsylvania Working Papers in Linguistics*, 17 (1), 195-204. [http://repository.upenn.edu/pwpl/vol17/iss1/22/]

Renwick, M. 2012. *Vowels of Romanian. Historical, Phonological and Phonetic Studies*. Dissertation. Cornell University. [http://conf.ling.cornell.edu/peggy/Renwick_2012_Vowels-of-Romanian.pdf]

Repp, B. 1984. Categorial perception: Issues, methods, findings. In: Lass, N.J. (ed.). *Speech and Language*. Orlando et al.: Academic Press, 243-335.

Repp, B., Healy, A., Crowder, R. 1978. Categories and context in the perception of isolated steady-state vowels. *Haskins Laboratories Status Report on Speech Research,* 54, 1-30. [http://www.haskins.yale.edu/sr/SR054/SR054_01.pdf]

Restle, D. & Mooshammer, C. 1999. The influence of the tense-lax contrast in vowels on the production of post-vocalic consonants in Standard German. *ICPhS,* XIV, San Francisco, vol. 1, 531-534.

Ridjanović, M. 2012. *Bosnian for Foreigners. With a Comprehensive Grammar*. Sarajevo: Rabić Publ.

Ringbom, H. 2007. *Cross-Linguistic Similarity in Foreign Language Learning*. Clevedon: Multilingual Matters.

Rips, L., Smith, E., Medin, D. 2012. Concepts and Categories: Memory Meaning, and Metaphysics. In: Holyoak, K. & Morrison, R. (eds.). *The Oxford Handbook of Thinking and Reasoning*. New York: Oxford University Press, 177-209.

Roach, P. 2000. *English Phonetics and Phonology: A Practical Course,* 3rd edition. Cambridge: Cambridge University Press.

Rochet, B. 1995. Perception and production of second-language speech sounds by adults. In: Strange, W. (ed.). *Speech Perception and Linguistic Experience.* Timonium, MD: York Press, 379-410.

Roelcke, Th. 1997. *Sprachtypologie des Deutschen: historische, regionale und funktionale Variation*. Berlin: de Gruyter.

Roelcke, Th. (ed.) 2003. *Variation Typology. A Typological Handbook of European Languages Past and Present*. Berlin New York: de Gruyter.

Rohany Rahbar, E. 2008. On distinguishing Persian vowels. *Paper read at the Canadian Linguistics Association Congress 2008*. [http://lib-ocs.lib.sfu.ca:8087/fedcan/index.php/cla/acl-cla2008/paper/viewFile/139/84]

Rohany Rahbar, E. 2009. On contrast in the Persian vowel system. *Toronto Working Papers in Linguistics*, 31. [http://twpl.library.utoronto.ca/index.php/twpl/article/viewFile/6092/3093]

Ronneberger-Sibold, E. 1999. Ambisyllabic consonants in German. Evidence from dialectal pronunciation of lexical creations. In: Rennison, J. & Kühnhammer, K. (eds.). *Phonologica 1996: Syllables?!*. The Hague: Thesus, 247-271. [http://www.univie.ac.at/linguistics/tagungen/phon96/phonologica1996.pdf]

Rosch, E. 1973. Natural Categories. *Cognitive Psychology*, 4 (3), 328-350.

Rosch, E. 1975. The Nature of Mental Codes for Color Categories. *Journal of Experimental Psychology: Human Perception & Performance,* 1 (4), 303-322.

Rothstein, R. 1993. Polish. In: Comrie, B. & Corbett, G. (eds.). *The Slavonic Languages*. London: Routledge. 686-758.

Rothstein, R. 2006. Polish. In: Brown, K. (ed.). *Encyclopedia of Language & Linguistics*, 2nd edition. Amsterdam et al.: Elsevier, 672-676.

Rubach, J. 1984. *Cyclic and Lexical Phonology. The Structure of Polish*. Dordrecht: Foris Publications.

Rubach, J. 2006. Polish: Phonology. In: Brown, K. (ed.). *Encyclopedia of Language & Linguistics*, 2nd edition. Amsterdam et al.: Elsevier, 676-679.

Rubach, J. & Booij, G. 1985. A grid theory of stress in Polish. *Lingua*, 66 (4), 281-320.

Rubin, J. 1975. What the 'good language learner' can teach us. *TESOL Quarterly*, 9, 41-51.

Rubin, J. 1987. Learner strategies: theoretical assumptions, research history and typology. In: Wenden, A. & Rubin, J. (eds.). *Learner Strategies in Language Learning.* UK: Prentice Hall, 15-19.

Saadah, E. 2011. *The Production of Arabic Vowels by English L2 Learners and Heritage Speakers of Arabic*. Dissertation. University of Illinois at Urbana-Champaign. [https://www.ideals.illinois.edu/bitstream/handle/2142/24104/Saadah_Eman.pdf?sequence=1]

Salehi, F., Bahrami, A., Nouri, F., Ghanbari, H., Baranzehi, T. 2008. Formant frequency of Persian vowels in 18-24 years old university students. *Journal of Research in Rehabilitation Sciences*, 4 (1), 35-44. [http://www.jrrs.mui.ac.ir/index.php/jrrs/article/view/27/26]

Samuel, A. 1982. Phonetic prototypes. *Perception and Psychophysics*, 31, 307-314.

Sapir, S. 1989. The intrinsic pitch of vowels: Theoretical, physiological and clinical considerations. *Journal of Voice*, 3, 44-51.

Sato, C. 1984. Phonological Processes in second language acquisition: another look at interlanguage syllable structure. *Language Learning*, 34 (4), 43-58.

Saville-Troike, M. 2006. *Introducing Second Language Acquisition*. Cambridge: Cambridge University Press.

Sawicka, I. 1995. Fonologia. In: Dukiewicz, L. & Sawicka, I. (eds.). *Gramatyka współczesnego języka polskiego*. Kraków: PWN, 105-195.

Sawicka, I. 1995. The Balkan Sprachbund in the light of phonetic features. *ICPhS,* XIII, Stockholm, vol. 3, 386-388.

Şayli, Ö. 2002. *Duration Analysis and Modelling for Turkish Text-to-Speech Synthesis*. Boğaziçi University. M.S. Thesis.

Schader, Basil 2008. Albanisch. In: Colombo-Scheffold, S. et al. (eds.): *Ausländisch für Deutsche. Sprachen der Kinder – Sprachen im Klassenzimmer*. Freiburg i. Br.: Fillibach, 43-57.

Schader, B. 2009. Albadeutsch und Shqipanisht: Untersuchungen zum albanisch-deutschen Sprachkontakt und Codeswitching. In: Ahrenholz, B. (ed.): *Empirische Befunde zu DaZ-Erwerb und Sprachförderung. Beiträge aus dem 3. Workshop Kinder mit Migrationshintergrund*. Freiburg i. Br.: Fillibach, 273-290.

Schader, B. 2010. *Albanisch und Deutsch: Stichworte zu einem Sprachvergleich Exkurs: Potenzielle Deutschprobleme von Albanischsprachigen*. Ms.
Schmid, P. & Yeni-Komshian, G. 1999. The effects of speaker accent and target predictability on perception of mispronunciations. *Journal of Speech, Language, and Hearing Research*, 42, 56-64.
Scholes R. 1967. Phoneme categorization of synthetic vocalic stimuli by speakers of Japanese, Spanish, Persian, and American English. *Language & Speech*, 10, 46-68.
Schouten, M. & van Hessen, A. 1992. Modelling phoneme perception. I. Categorical perception. *JASA 92* (4), 1841-1855.
Schröder, A. & Stadie, N. 2014. Phonemfrequenzen [auf Basis von Aichert, I., Marquardt, C., Ziegler, W. 2005)]. Universität Potsdam, Treatment Lab. [ttp://www.uni-potsdam.de/fileadmin/projects/treatmentlab/assets/Phonemfrequenzen_TreatmentLab.pdf]
Schröder, C. & Şimşek, Y. 2010. Kontrastive Analyse Türkisch-Deutsch. In: Krumm, H.-J., Fandrych, C., Hufeisen, B., Riemer, C. (eds.). *Deutsch als Fremd- und Zweitsprache. Ein internationales Handbuch*, vol. 1, 719-726.
Schwartz, B. & Sprouse, R. 1996. L2 cognitive states and the Full Transfer/Full Access model. *Second Language Research*, 12, 40-72.
Schwartz, B. & Sprouse, R. 2000. When syntactic theories evolve. Consequences for L1 acquisition research. In: Archibald, J. (ed.). *Second Language Acquisition and Linguistic Theory*. Malden: Blackwell, 156-186.
Schwartz, J. Abry, C., Boë, L.-J., Ménard, L., Vallée, N. 2005. Asymmetries in vowel perception in the context of the Dispersion-Focalisation Theory. *Speech Communication*, 45, 425-434.
Schwartz, J.-L., Boe, L.-J., Vallée, N., Abry, C. 1997a. Major trends in vowel system inventories. *JPhon* 25, 233-253.
Schwartz, J.-L., Boe, L.-J., Vallée, N., Abry, C. 1997b. The Dispersion-Focalization Theory of vowel systems. *JPhon*, 25, 255-286.
Scovel, T. 1988. *A Time to speak: A Psycholinguistic Inquiry into the Critical Period for Human Speech*. Rowley, MA: Newbury House.
Scovel, T. 2000. A critical review of critical period research. *Annual Review of Applied Linguistics,* 20, 213-223.
Sek, A. & Moore, B. 1995. Frequency discrimination as a function of frequency, measured in several ways. *JASA*, 97 (4), 2479-2486.
Selen, N. 1979. *Söyleyiş sesbilimi, akustik sesbilim ve Türkiye Türkçesi*. Ankara: Türk Dil Kurumu.
Seliger, H. 1984. Processing universals in second language acquisition. In: Eckman, F., Bell, L., Nelson, D. (eds.). *Universals of Second Language Acquisition*. Rowley, MA: Newbury House, 36-47.
Selinker, L. 1969. Language transfer. *General Linguistics*, 9, 67-92.
Selinker, L. 1972. Interlanguage. *International Review of Applied Linguistics*, 10 (3), 209-241.
Selinker, L. 1992. *Rediscovering Interlanguage*. London: Longman.
Sendlmeier, W. 1981. Der Einfluß von Qualität und Quantität auf die Perzeption betonter Vokale des Deutschen. *Phonetica*, 38, 291-308.
Sendlmeier, W. 1985. *Psychophonetische Aspekte der Wortwahrnehmung*. Hamburg: Buske.
Sendlmeier, W. & Seebode, J. (n.y.). Formantkarten des deutschen Vokalsystems. Ms. TU Berlin, [https://www.kw.tu-berlin.de/fileadmin/a01311100/Formantkarten_des_deutschen_Vokalsystems_01.pdf]
Sereno, J. & Wang, Y. 2007. Behavioral and cortical effects of learning a second language. The acquisition of tone. In: Bohn, O.-S. & Munro, M. (eds.). *Language Experience in Second Language Speech Learning. In Honor of James Emil Flege*. Amsterdam Philadelphia: John Benjamins, 239-258.
Serniclaes, W. 2011. Features are phonological transforms of natural boundaries. In: Clements, N. & Ridouane, R. (eds.). *Where do Phonological Features come from?* Amsterdam Philadelphia: John Benjamins, 237-257.
Sezer, E. 1983. On non-final stress in Turkish. *Journal of Turkish Studies*, 5, 61-69.
Sezer, E. 1986. An autosegmental analysis of compensatory lengthening in Turkish. In: Wetzels, W.L. & Sezer, E. (eds.). *Studies in compensatory lengthening*. Dordrecht: Foris Publications, 227-250. [http://www.ai.mit.edu/projects/dm/featgeom/sezer86.pdf]
Shadish, W., Cook, T., Campbell, D. 2002. *Experimental and Quasi-Experimental Designs for Generalized Causal Inference*. Boston: Houghton Mifflin.

Shafer, V., Yu, Y., Datta, H. 2010. Maturation of Speech Discrimination in 4- to 7-year-old childrenas indexed by event-related potential mismatch responses. *Ear and Hearing*, 31 (6), 735-745.

Sharwood Smith, M. 2011. Is the native speaker a psycholinguistic impossibility? Paper held at *EUROSLA 21, Annual Conference of the European Second Language Association*. [http://eurosla.org/eurosla21_LLroundtable.html]

Shaw, J., Gafos, A.I., Hoole, P., Zeroual, C. (2009). Temporal evidence for syllabic structure in Moroccan Arabic: data and model. *Phonology*, 26 (1), 187-215.

Shepard, R. 1957. Stimulus and response generalization: A stochastic model relating generalization to distance in psychological space. *Psychometrika*, 22, 325-345.

Shepard, R. 1972. Psychological representation of speech sounds. In: David, E. & Denes, P. (ed.). *Human Communication: A Unified View*. New York: McGraw-Hill, 67-113. [http://courses.cit.cornell.edu/ling4476/shepard-on-speech-sounds.pdf]

Shepard, R. 1980. Multidimensional Scaling, Tree-Fitting, and Clustering. *Science*, 210, 390-398. [http://serendip.brynmawr.edu/local/scisoc/emergence/shepard80.pdf]

Shriberg, E. 1992. Perceptual restoration of filtered vowels with added noise. *Language and Speech*, 35 (1), 127-136.

Sidwell, A & Summerfield, Q. 1986. The auditory representation of symmetrical CVC syllables. *Speech Communication*, 5 (3-4), 283-297

Siemund, P. 2003. English. In: Roelcke, Th. (ed.). *Variation Typology. A Typological Handbook of European Languages Past and Present*. Berlin New York: de Gruyter, 1-29.

Silverman, D. 2011. Usage-based Phonology. In: Kula, N., Botma, E. Botma, B., Nasukawa. K. (eds.). *The Continuum Companion to Phonology*. London New York: A&C Black. 369-394. [http://seedyroad.com/academics/usage-based-phonology.pdf]

Silverman, K. 1987. *The structure and processing of fundamental frequency contours*. PhD Dissertation. Cambridge University.

Singleton, D. & Lengyel, Z. (eds.) 1995. *The Age Factor in Second Language Acquisition*. Clevedon: Multilingual Matters.

Siptár, P. & Törkenczy, M. 2000. *The Phonology of Hungarian*. Oxford: Oxford University Press.

Skehan, P. 1989. *Individual differences in second-language learning*. London: Edward Arnold.

Skinner, B. 1957. *Verbal Behavior*. New York: Appleton-Century-Crofts.

Smiljanić, R. 2004. *Lexical, Pragmatic, and Positional Effects on Prosody in Two Dialects of Croatian and Serbian. An Acoustic Study*. New York London: Routledge.

Smith, J. 2004. *The American-British British-American Dictionary: For English speaking People. American and British Pronunciation Differences*. Corvallis: CodeSmith Publ. [http://members.peak.org/~jeremy/dictionaryclassic/chapters/pronunciation.php]

Smith, J. 2008. Phonological constraints are not directly phonetic. In: *Proceedings of CLS 41*, vol. 1, 457-471. [http://www.unc.edu/~jlsmith/home/pdf/cls41jls.pdf]

Smith, Neilson V. 1973. *The Acquisition of Phonology*. Cambridge: Cambridge University Press.

Smolensky, P. 1996. On the comprehension/production dilemma in child language. *Linguistic Inquiry,* 27, 720-731.

Soos, M. 2013. The reversal of the BÄREN-BEEREN merger in Austrian Standard German. *The Mental Lexicon*, 8 (3), 353-371.

Sovilj-Nikic, S. 2007. *Trajanje vokala kao jedan od prozodijskih elemenata u sintezi govora na srpskom jeziku*. M.S. thesis, Fakultet tehnickih nauka, Novi Sad, 2007.

Sovilj-Nikic, S. 2010. Trajanje glasova i faktori koji ga određuju u srpskom i drugim jezicima. *18th Telekomunikacioni forum TELFOR 2010 Srbija*, Beograd, 602-605. [http://2010.telfor.rs/files/radovi/TELFOR2010_05_09.pdf]

Sovilj-Nikic, V. Delic, I. Sovilj-Nikic, M. Markovic 2014. Tree-based Phone Duration Modelling of the Serbian Language. *Elektonika Elektrotechnika*, 20 (3), 2014. [http://matsc.ktu.lt/index.php/elt/article/download/4090/3461]

Sovilj-Nikić, S., Secujski, M., Delić, V., Sovilj-Nikić, I. 2006. Analysis of different factors influencing vowel duration in the Serbian language. *Proceedings of the 10th Symposium on Social Communication*, Santiago de Cuba, vol. 1, 385-389. [http://www.santiago.cu/hosting/linguistica/simposios.php]

Stadtschulrat Wien, 2007. *Wiener Schulpflichtmatrik 2006*. Wien: Stadtschulrat für Wien.
Stampe, D. 1969. The acquisition of phonetic representation. In: Binnick, R., Davison, A., Green, G., Morgan, J. (eds.). *Proceedings of the Fifth Regional Meeting of the Linguistic Society*. Chicago: Chicago Linguistic Society, 443-454.
Stampe, D. 1973a. *A Dissertation on Natural Phonology*. PhD Dissertation. University of Chicago. (distributed 1979 by Indiana University Linguistics Club, Bloomington).
Stampe, D. 1973b. On chapter nine. In: Kenstowicz, M. & Kisseberth, C. (eds.). *Issues in Phonological Theory: Proceedings of the Urbana Conference on Phonology*, April 23-24, 1971. The Hague: Mouton, 44-52.
Stampe, D. 1987. On phonological representation. In: Dressler, W., Luschützky, H., Pfeiffer O., Rennison, J. (eds). *Phonologica 1984*. London: Cambridge University Press, 287-300.
Steriade, D. 1995. Underspecification and markedness. In: Goldsmith, J. (ed.). *The Handbook of Phonological Theory.* Cambridge: Blackwell, 114-174.
Steriade, D. 1997. *Phonetics in phonology: the case of laryngeal neutralization*. Ms. [http://www.linguistics.ucla.edu/people/steriade/papers/PhoneticsInPhonology.pdf]
Steriade, D. 2001a. Directional asymmetries in place assimilation: a perceptual account. In: Hume, E. & Johnson, K. (eds.). *The Role of Speech Perception in Phonology*. New York: Academic Press, 219-247.
Steriade, D. 2001b. The Phonology of Perceptibility Effects: the P-map and its consequences for constraint organization, Ms. [http://www.linguistics.ucla.edu/people/steriade/papers/P-map_for_phonology.doc]
Steriade, D. 2003. Knowledge of perceptual similarity and its phonological uses: evidence from half-rhymes. In: Solé, M.J, Recasens, D., Romero, J. (eds). *ICPhS, XV*, Barcelona: Universitat Autònoma de Barcelona, 363-366.
Steriade, D.. 2008. A pseudo-cyclic effect in Romanian morphophonology. In: Bachrach, A. & Nevins, A. (eds.). *Inflectional Identity*. Oxford: Oxford University Press, 313-355.
Stern, H.H. 1975. What can we learn from the good language learner. *Canadian Modern Language Review*, 34, 304-318.
Stevens, K. & House, A. 1955. Development of a quantitative description of vowel articulation. *JASA*, 27, 484-495.
Stevens, K. & House, A. 1963. Perturbation of vowel articulations by consonantal context: An acoustical study. *Journal of Speech and Hearing Research*, 6, 111-128.
Stevens, K. 1972. The quantal nature of speech: evidence from articulatory-acoustic data. In: Davis, E. & Denes, P. (eds.). *Human Communication: A Unified View*. New York: McGraw-Hill, 51-66.
Stevens, K. 1989. On the quantal nature of speech. *JPhon,* 17, 3-45.
Stevens, K. 1997. Articulatory-acoustic-auditory relationships. In: Hardcastle, W. & Laver, J. (eds). *The Handbook of Phonetic Sciences*, 462-506.
Stevens, K. 1998. *Acoustic phonetics*. Cambridge, MA: MIT Press.
Stevens, K. 2002. Toward a model for lexical access based on acoustic landmarks and distincitve features. *JASA*, 111, 1872-1891.
Stevens, K. & Keyser S. 1989. Primary features and their enhancement in consonants. *Language*, 65, 81-106.
Stevens, K. & Keyser, S. 2010. Quantal theory, enhancement and overlap. *JPhon*, 38 (1), 10-19.]
Stevens, K., Liberman, A., Studdert-Kennedy, M., Öhman, S. 1969. Cross language study of vowel perception. *Language Speech*, 12, 1-23.
Stevens, K. & Halle, M. 1967. Remarks on analysis by synthesis and distinctive features. In: Wathen-Dunn, W. & Woods, L. (eds.). *Models for the Perception of Speech and Visual Form*. Cambridge: MIT Press, 88-102.
Stockwell, R.P., Donald Bowen, J., Martin, J.W. 1965. *The Grammatical Structures of English and Spanish*. Chicago: University of Chicago Press.
Stoll, G. 1984. Pitch of vowels: Experimental and theoretical investigation of its dependence on vowel quality. *Speech Communication*, vol. 3, 137-150.
Stone, Gerald. 2011. Polish. In: Comrie, B. (ed.). *The World's Major Languages*. London New York: Routledge, 289-304.
Strange, W. 1989a. Evolving theories of vowel perception. *JASA*, 85 (5), 2081-2087.
Strange, W. 1989b. Dynamic Specification Theory. *JASA,* 85, 2135-2153.

Strange, W. 1992. Learning non-native phoneme contrasts: Interactions among subject, stimulus und task variables, In: Tohkura, Y., Vatikiotis-Bateson, W., Sagisaka, Y. (eds.). *Speech Perception, Production and Linguistic Structure*. Tokyo: Ohmsha, 197-219.

Strange, W. (ed.) 1995a. *Speech Perception and Linguistic Experience: Theoretical and Methodological Issues*. Timonium, MD: York Press.

Strange, W. 1995b. Cross-language studies of speech perception: A historical review. In: Strange, W. (ed.). *Speech Perception and Linguistic Experience*. Timonium, MD: York Press, 3-45.

Strange, W. 1996. Phonetics of Second-Language Acquisition: Past, Present, Future. *ICPhS,* XIII, Stockholm, vol. 4, 76-83.

Strange, W. 2007a. Cross-language phonetic similarity of vowels: Theoretical and methodological issues. In: Bohn, O.-S. & Munro, M. (eds.). *Language Experience in Second Language Speech Learning. In Honor of James Emil Flege*. Amsterdam Philadelphia: John Benjamins, 35-55.

Strange W. 2007b. Selective perception, perceptual modes, and automaticity in first- and second-language processing. *JASA*, 122, 2970.

Strange, W. 2011. Automatic selective perception (ASP) of first and second language speech: A working model. *JPhon*, 39 (4), 453-710.

Strange, W. & Bohn, O.-S. 1998. Dynamic specification of coarticulated German vowels: Perceptual and acoustical studies. *JASA*, 104 (1), 488-504.

Strange, W., Bohn, O.-S., Nishi, K., Trent, S. 2005. Contextual variation in the acoustic and perceptual similarity of North German and American English vowels. *JASA*, 118 (3), 1751-1762.

Strange, W., Bohn, O.-S., Trent, S.A, Nishi, K. 2004. Acoustic and perceptual similarity of North German and American English vowels. *JASA*, 115 (4), 1791-1807.

Strange, W., Jenkins, J., Johnson, T. 1983. Dynamic specification of coarticulated vowels. *JASA,* 74 (3), 695-704.

Strange, W., Levy, E., Law II, F. 2009. Cross-language categorization of French and German vowels by naïve American listeners. *JASA*, 126 (3), 1461-1476.

Strange, W. & Shafer, V. 2008. Speech perception in second language learners. The re-education of selective perception. In: Hansen Edwards, J. & Zampini, M. (eds.). *Phonology and Second Language Acquisition*. Amsterdam: John Benjamins, 153-192.

Strange, W., Weber, A., Levy, E., Shafiro, V., Hisagi, M., Nishi, K. 2007. Acoustic variability within and across German, French, and American English vowels: phonetic context effects. *JASA*, 122 (2), 1111-1129.

Strutyński, Janusz, 1999. *Gramatyka polska*. Kraków: Wydawnictwo Tomasz Strutyn.

Studdert-Kennedy, K., Shankweiler, D., Pisoni, D. 1972. Auditory and phonetic processes in speech perception. Evidence from a dichotic study. *Cognitive Psychology*, 3, 455-466.

Syrdal, A. & Gopal, H. 1986. A perceptual model of vowel recognition based on the auditory representation of American English vowels. *JASA*, 79, 1086-1100.

Szende, T. 1994. Illustrations of the IPA: Hungarian. *Journal of the International Phonetic Association*, 24 (2), 91-94.

Szende, T. 1999. Hungarian. In: *Handbook of the International Phonetic Association*. Cambridge: Cambridge University Press, 104-107.

Takahashi, Y. 1987. *Language Universals and Transfer Effects: Acquisition of English Vowel Durations by Japanese Learners*. PhD Dissertation. Stanford University.

Tauberer, J. & Evanini, K. 2009. *Intrinsic vowel duration and the post-vocalic voicing effect: Some evidence from dialects of North American English*. Interspeech 2009 Proceedings, 2211-2214. [http://razor.occams.info/pubdocs/interspeech2009.pdf]

Teodorescu, H.-N., Pistol, L., Feraru, M. 2010. *Voiced Sounds of the Romanian Language*. Online [http://www.etc.tuiasi.ro/sibm/romanian_spoken_language/en/vocale_en.htm]

Terbeek, D. 1977. *A Cross-Language Multi-Dimensional Scaling Study of Vowel Perception*. PhD Dissertation. University of California, Los Angeles. [http://escholarship.org/uc/item/3nx4g138?display=all#page-1]

Tesar, B. & Smolensky, P. 2000. *Learnability in Optimality Theory*. Cambridge, MA: MIT Press.

Thelwall, R. 1990. Illustrations of the IPA: Arabic. *Journal of the International Phonetic Association,* 20 (2), 37-41.

Thomas, E. 2001. *An Acoustic Analysis of Vowel Variation in New World English*. Durham, NC: Duke University Press.

Toma, I. 2004. *Limba română contemporană: fonetică-fonologie, lexicologie*. Bukarest: Editura Fundaţiei România de mâine.
Tomaschek, F., Truckenbrodt, H., Hertrich, I. 2011. Processing German vowel quantity: Categorical perception or perceptual magnet effect. *ICPhS,* XVII, Hong Kong, 2002-2005.
Toosarvandani, M. 2004. Vowel Length in Modern Farsi. *Journal of the Royal Asiatic Society*, 14 (3), 241-251.
Torgersen, W. 1965. Multidimensional scaling of similarity. *Psychometrika*, 30, 379-393.
Traunmüller, H. 1981. Perceptual dimension of openness in vowels. *JASA*, 69, 1465-1475.
Traunmüller, H. & Krull, D. 2003. The effect of local speaking rate on the perception of quantity in Estonian. *Phonetica*, 60, 187-207.
Traxler, M.J. (ed.) 2006. *Handbook of Psycholinguistics*, 2nd edition. Amsterdam et al.: Elsevier.
Triskova, H. 2008. The sounds of Chinese and how to teach them. [Review article on Lin, Y.-H. 2007. The Sounds of Chinese. Cambridge University Press]. *Oriental Archive*, 76.
Trommer, J & Grimm, A. 2004. *Albanian Word Stress*. Paper read at Manchester Phonology Meeting. [http://www.uni-leipzig.de/~jtrommer/papers/alb.pdf]
Trubetzkoy, N. 1929. Zur allgemeinen Theorie der phonologischen Vokalsysteme. *Travaux du Cercle Linguistique de Prague*, 1, 39-67.
Trubetzkoy, N. 1931. Die phonologischen Systeme. *Travaux du Cercle Linguistique de Prague,* 4, 96-116.
Trubetzkoy, N. 1939. *Grundzüge der Phonologie*. Prague: Travaux du Cercle Linguistique de Prague, 7.
Trubetzkoy, N. 1958. *Grundzüge der Phonologie*. Göttingen: Vandenhoeck & Ruprecht.
Trubetzkoy, N. 1969. *Principles of Phonology*. Berkeley, CA: University of California Press.
Trubetzkoy, N. 2001. A theory of phonological oppositions. In: Trubetzkoy, N. 2001. *Studies in General Linguistics and Language Structure*. (with an introduction by A. Liberman), Durham London: Duke University Press 14-21.
Trudgill, P. & Hannah, J. 2002. *International English. A Guide to the Varieties of Standard English*, 4th edition. London: Arnold.
Tsuji, S. & Cristia, A. 2014. Perceptual attunement in vowels: A meta-analysis. *Developmental Psychobiology*, 56, 179-191.
Türk, O., Sayli, O., Ozsoy, A.S., Arslan, L.M., 2004. Türkçe'de Ünlülerin Formant Frekans Ýncelemesi. *18. Ulusal Dilbilim Kurultayý 2004*, Ankara, Turkey. [http://www.busim.ee.boun.edu.tr/~speech/publications/Turkish_phonetics/oytun_c004.pdf]
Turner, P. & Verhoeven, J. 2011. Intrinsic vowel pitch: A gradient feature of vowel systems. *ICPhS,* XVII, Hong Kong.
Tversky, A. 1977. Features of similarity. *Psychological Review*, 84, 327-352.
Underhill, R. 1986. Turkish. In: Slobin, D.I. & Zimmer, K. (eds.). *Studies in Turkish Linguistics*. Amsterdam Philadelphia: John Benjamins, 7-21. [http://de.scribd.com/doc/32322192/Turkish-by-Robert-Underhill]
Vago, R. 1980. *The Sound Pattern of Hungarian*. Washington, D.C.: Georgetown University Press.
Vaissière, J. 2011. On the acoustic and perceptual characterization of reference vowels in a cross-language perspective. *ICPhS,* XVII, Hongkong, 52-59. [http://www.icphs2011.hk/resources/OnlineProceedings/PlenaryLecture/Vaissiere/Vaissiere.pdf]
Valaczkai, L. 1998. *Atlas deutscher Sprachlaute. Instrumentalphonetische Untersuchung der Realisierung deutscher Phoneme als Sprechlaute*. Wien: Edition Praesens.
van der Hulst, H. & van de Weijer, J. 1991. Topics in Turkish Phonology. In: Boeschoten, H. & Verhoeven, L. (eds.). *Structure and Use of Turkish.* Leiden: Brill, 11-59. [http://www.webklik.nl/user_files/2009_10/72522/Papers/Topics%20Turkish%20Phonology.pdf]
van Dommelen, W.A.1993. Does dynamic F0 increase perceived duration? New light on an old issue. *JPhon*, 21, 367-386.
van Heuven, V. 1985. Some acoustic characteristics and perceptual consequences of foreign accent in Dutch spoken by Turkish immigrant workers. In: van Oosten, J. & Snapper, J.F. (eds.). *Dutch Linguistics at Berkeley*. Papers presented at the Dutch Linguistics Colloquium held at the University of California, Berkeley on November 9th 1985, 67-84. [https://openaccess.leidenuniv.nl/bitstream/handle/1887/2831/167_028.pdf?sequence=1]
van Hoof, S. & Verhoeven, J. 2011. Intrinsic vowel F0, the size of vowel inventories and second language acquisition. *JPhon*, 39, 168-177.

van Santen, J.P.H. 1992. Contextual effects on vowel duration. *Speech Communication*, 11 (6), 513-546.

van Son, R. 1993. Vowel Perception: A closer look at the literature. *IFA Proceedings*, 17, Institute of Phonetic Sciences, University of Amsterdam, 33-64. [http://www.fon.hum.uva.nl/rob/Publications/Proc_Van_Son_1993.pdf]

Vasiliu, E. 1965. *Fonologia limbii române*. Bucureşti: Editura Ştiinţifică.

Vaux, B. & Samuels, B. 2006. *Explaining vowel systems: Dispersion Theory vs. Natural Selection*. Ms. [http://www.academia.edu/207443/Explaining_vowel_systems_Dispersion_Theory_vs._Natural_Selection]

Vennemann, T. 1983. Causality in language change. Theories of linguistic preferences as a basis for linguistic explanations. *Folia Linguistica Historica,* 6 (1), 5-26.

Vennemann, T. 1991. Syllable structure and syllable cut prosodies in Modern Standard German. In: Bertinetto, P., Kenstowicz, M., Loporcaro, M. (eds.). *Certamen Phonologicum II: Papers from the 1990 Cortona Phonology Meeting*. Torino: Rosenberg & Sellier, 211-243.

Verhoeven, J. & van Hoof, S. 2008. Intrinsic vowel pitch in an interlanguage perspective. *Studies van de Belgische Kring voor Lingüistiek* (Belgische Kring voor Linguïstiek), 3 (2008), 1-9. [http://webh01.ua.ac.be/linguist/SBKL/sbkl2008/ver2008.pdf]

Viëtor, W. 1890. *Phonetische Studien*, vol. 3. Marburg: N. G. Elwert. [https://archive.org/details/phonetischestud00victgoog]

Walley, A. 1993. More developmental research is needed. *JPhon*, 21, 171-176.

Walley, A. 2008. Speech Perception in Childhood. In: Pisoni, D. & Remez, R. (eds.). *The Handbook of Speech Perception*. Malden, MA: Blackwell, 449-468.

Walley, A. & Flege, J. 1999. Effect of lexical status on children's and adults' perception of native and non-native vowels. *JPhon,* 27, 307-332.

Walley, A. & Metsala, J. 1992. Young children's perception of spoken words varying in age-of-acquisition. *Memory & Cognition*, 20, 171-182.

Wang, M. 1993. *Kontrastierung der Phonemsysteme des Chinesischen und des Deutschen*. Frankfurt am Main: Lang.

Wang, W.S.-Y., Lehiste, I., Chuang, C.-K., Darnovsky, N. 1976. Perception of vowel duration. *JASA*, 60 (S1), S92.

Wängler, H. & Weiss, R. 1975. Experimental approach to the study of vowel perception in German. *Phonetica*, 32, 180-199.

Wardhaugh, R. 1970. The contrastive analysis hypothesis. *TESOL Quarterly,* 4, 123-130.

Watkins, M., Rauber, A., Baptista, B. (eds.) 2010. *Recent Research in Second Language Phonetics/Phonology. Perception and Production*. Cambridge New Castle: Cambridge Scholars Publishing.

Watson, J. 2007. *The Phonology and Morphology of Arabic*. Oxford: Oxford University Press.

Wedel, A. 2006. Exemplar models, evolution and language change. *The Linguistic Review*, 23, 247-274.

Weinreich, U. 1953. *Languages in Contact*. New York: Linguistic Circle of New York.

Weiss, R. 1972. Perceptual parameters of vowel duration and quality in German. In: Rigault, A. & Charbonneau, R. (eds.). *Proceedings of the 7th International Congress of Phonetic Sciences, 1971*. (*ICPhS VII*, Montreal). The Hague: Mouton, 633-636.

Weiss, R. 1974. Perception of German vowels by native Germans and American students of German: A preliminary study. *Hamburger Phonetische Beiträge*, 13, 283-294.

Weiss, R. 1976. *The Perception of Vowel Length and Quality in German. An Experimental Phonetic Investigation*. Hamburg: Buske.

Weiss, R. 1978. The German vowel: Phonetic and phonemic considerations. *Hamburger Phonetische Beiträge*, 25, 461-475.

Werker, J. & Lalonde, C. 1988. Cross-language speech perception: initial capabilities and developmental change. *Developmental Psychology*, 24 (5), 672-683.

Werker, J. & Logan, J. 1985. Cross-language evidence for three factors in speech perception. *Perception and Psychophysics*, 37, 35-44.

Werker, J. & Tees, R. 1983. Developmental changes across childhood in the perception of nonnative speech sounds. *Canadian Journal of Psychology*, 57, 278-286.

Werker, J. & Tees, R. 1984a. Phonemic and phonetic factors in adult cross-language speech perception. *JASA*, 75, 1866-1878.

Werker, J. & Tees, R. 1984b. Cross-language speech perception: evidence for perceptual reorganization during the first year of life. *Infant Behaviour and Development*, 7, 49-63.

Werker, J. & Tees, R. 1992. The organization and reorganization of human speech perception. *Annual Review of Neuroscience*, 15, 377-402.

Werker, J. & Tees, R. 1999. Influences on infant speech processing: Towards a new synthesis. *Annual Review of Psychology*, 50, 509-535.

Werker, J. & Tees, R. 2005. Speech perception as a window for understanding plasticity and commitment in language systems of the brain. *Developmental Psychobiology (Special Issue: Critical Periods Re-examined: Evidence from Human Sensory Development)*, 46 (3), 233-251. [http://129.199.80.1/~ruthdb/articles/werker05.pdf]

Whalen, D. & Levitt, A. 1995. The universality of intrinsic F0 of vowels. *JPhon*, 23, 349-366.

Whaley, L. 1997. *Introduction to Typology. The Unity and Diversity of Language*. Thousand Oaks: Sage.

White, L. 1987. Markedness and second language acquisition: The question of transfer. *Studies in Second Language Acquisition,* 9, 261-286.

White, L. 1990. Second language acquisition and universal grammar. *Studies in Second Language Acquisition,* 12, 121-133.

White, L. 2000. Second language acquisition: From initial to final state. In: Archibald, J. (ed.). *Second Language Acquisition and Linguistic Theory*. Malden: Blackwell, 130-155.

White, L. 2003. *Second Language Acquisition and Universal Grammar*. Cambridge: Cambridge University Press.

Wiese, R. 1996. *The Phonology of German*. Oxford: Clarendon Press.

Williams, D.P. 2013. *Cross-Language Acoustic and Perceptual Similarity of Vowels. The Role of Listeners' Native Accents*. Doctoral Dissertation. University of Sheffield. [http://etheses.whiterose.ac.uk/3716/]

Wilson, C. 2006. Learning phonology with substantive bias: An experimental and computational study of velar palatalization. *Cognitive Science*, 30, 945-982.

Windfuhr, G. 1979. *Persian Grammar. History and State of its Study*. The Hague: Mouton.

Windfuhr, G. 2011. Persian. In: Comrie, B. (ed.). *The World's Major Languages*, 2nd edition. London et al.: Routledge, 445-459.

Winkler, I., Kujala, T., Tiitinen, H., Sivonen, P., Alku, P., Lehtokoski, A., Czigler, I., Csépe, V., Ilmoniemi, R., Naatänen, R. 1999. Brain responses reveal the learning of foreign language phonemes. *Psychophysiology*, 36, Cambridge: Cambridge University Press, 638-642.

Winters, S., Lichtman, K., Weber, S. 2013. The role of linguistic knowledge in the encoding of words and voices in memory. In: Voss, E., Tai, S.-Y.D., Li, Z. (eds.). *Selected Proceedings of the 2011 Second Language Research Forum*. Cascadilla Proceedings Project. Sommerville, MA, 129-138. [http://www.lingref.com/cpp/slrf/2011/paper2911.pdf]

Wode, H. 1981. *Learning a Second Language*. Tübingen: Narr.

Wode, H. 1990. Die Entwicklung des sprachlichen Hörens und seine Bedeutung für einen zeitgemäßen Deutschunterricht. *Der Deutschunterricht*, 42 (5), 19-34.

Wode, H. 1993. The development of phonological abilities. In: Hyltenstamm, K. & Viberg, A. (eds.). *Progression and Regression in Language: Sociocultural, Neuropsychological and Linguistic Perspectives*. New York: Cambridge University Press, 415-438.

Wode, H. 1994a. Nature, nurture, and age in language acquisition: The case of speech perception. *Studies in Second Language Acquisition*, 16, 325-345.

Wode, H. 1994b. Perzeption, Produktion und die Lernbarkeit von Sprachen. In: Ramers, K., Vater, H., Wode, H. (eds.). *Universale phonologische Strukturen und Prozesse*. Tübingen: May Niemeyer Verlag, 169-187.

Wode, H. 1994c. Speech perception and the learnability of languages. *International Journal of Applied Linguistics*, 4, 143-168.

Wode, H. 1994d. Introduction. L1 and L2 phonology: Looking ahead. In: Yavaş, M. (ed.), *First and second language phonology*. San Diego, CA: Singular, 175-179.

Wood, S. 1975. The weakness of the tongue-arching model of vowel articulation. *Working Papers Dept. of Linguistics, University of Lund,* 11, 55-108.

Wood, S. 1979. A radiographic analysis of constriction location for vowels. *JPhon,* 7, 25-43. [http://www.ai.mit.edu/projects/dm/featgeom/wood1979-vowelconstric.pdf]

Wood, S. 1982a. X-ray and model studies of vowel articulation. *Working Papers Dept. of Linguistics, University of Lund,* 23, 1-49.

Wood, S. 1982b. The acoustical consequences of tongue, lip and larynx articulation in rounded palatal vowels. *Working Papers Dept. of Linguistics, University of Lund,* 23, 77-117.

Wood, S. 1986. The acoustical significance of tongue, lip, and larynx manoeuvers in rounded palatal vowels. *JASA*, 80 (2), 391-401.

Wood, S. 1990. Vertical, monovocalic and other "impossible" vowel systems: a review of the articulation of the Kabardian vowels. *Working Papers Dept. of Linguistics, University of Lund*, 36, 191-212.

Wood, S. 1991. X-ray data on the temporal coordination of speech gestures. *JPhon*, 19, 281-292.

Wood, S. 1992. A radiographic and model study of the tense-lax contrast in vowels. In: Dressler, W.U., Luschützky, H.C., Pfeiffer, O.E., Rennison, J. R. (eds.). *Phonologica 1988*. Cambridge: Cambridge University Press. 283-291.

Wood, S. 1994. A spectrographic analysis of vowel allophones in Kabardian. *Working Papers Dept. of Linguistics, University of Lund*, 42, 241-250.

Wrembel, M. 2014. VOT patterns in the acquisition of third language phonology. *Proceedings of the International Symposium on the Acquisition of Second Language Speech. Concordia Working Papers on Applied Linguistics*, 5, 750-770. [http://doe.concordia.ca/copal/documents/48_Wrembel_Vol5.pdf]

Wright, R. 2004. A review of perceptual cues and cue robustness. In: Hayes, B., Kirchner, R., Steriade, D. (eds.). *Phonetically based Phonology*. Cambridge: Cambridge University Press, 34-57.

Yavaş, M. 1982. Borrowing and its Implications for Turkish Phonology. *Kansas Working Papers in Linguistics*, 3, 34-44. [http://kuscholarworks.ku.edu/dspace/bitstream/1808/722/1/ling.wp.v3.n1.paper03.pdf]

Yen, Ch.-J. 1992. *Kontrastive Untersuchungen zur segmentalen Phonetik und Phonologie des Chinesischen und Deutschen*. Erlangen: Palm & Enke.

Yu, A.C.L. (2010). Tonal effects on perceived vowel duration. In: Fougeron, C., Kuhnert, B., D'Imperio, M., Vallée, N. (eds.). *Papers in Laboratory Phonology 10*. Berlin: Mouton de Gruyter.

Zee, E. & Lee, W.-S. 2001. An acoustical analysis of the vowels in Beijing Mandarin. *Proceedings of the 7th European Conference on Speech Communication and Technology, Eurospeech 2001*, 643-646.

Zee, E. & Lee, W.-S. 2007. Vowel Typology in Chinese*.* *ICPhS,* XVI, Saarbrücken, 1429-1432.

Zeftawi, S. 2005. MMN to natural Arabic CD syllables: 2 – cross-language study. *Hearing Research*, 210 (1-2), 24-29.

Zhang, Y., Kuhl, P., Imada, T., Kotani, M., Tohkura, Y. 2005. Effects of language experience: Neural commitment to language-specific auditory patterns. *NeuroImage*, 26 (3), 703-720.

Zimmer, K. & Orgun, O. 1992. Turkish. *Journal of the International Phonetic Association*, 22, 43-45.

Zimmer, K. & Orgun, O. 1999. Turkish. In: *Handbook of the International Phonetic Association: A Guide to the Use of the International Phonetic Alphabet*. Cambridge University Press: Cambridge, 154-156.

Zwicker, E. & Fastl, H. 1990. *Psychoacoustics: Facts and Models*. Berlin: Springer.

Zwicker, E. & Terhardt, E. 1980. Analytical expression for critical-band rate and critical bandwidth as a function of frequency. *JASA*, 68 (5), 1523-1525.

Curriculum Vitae and Publications

Studien

1987 – 1995	Studium der Allgemeinen und Angewandten Sprachwissenschaft und Orientalistik an der Universität Wien
1991 – 1992	Studium an der Humboldt-Universität Berlin (Deutsch als Fremdsprache) und an der Freien Universität Berlin (Orientalistik)
2001 – 2003	Universitätslehrgangs für Markt- und Meinungsforschung, Universität Wien

Studienabschlüsse

10/1992	Zertifikat Deutsch als Fremdsprache, Humboldt-Universität Berlin
06/1995	Studienabschluss in Sprachwissenschaft und Orientalistik an der Universität Wien (mit Auszeichnung)
07/2003	Abschluss des 2-jährigen Universitätslehrgangs für Markt- und Meinungsforschung, Universität Wien (mit Auszeichnung)

Wissenschaftliche Tätigkeit und akademische Lehre

09/1993 – 08/1995	Studienassistentin am Lehrstuhl Deutsch als Fremdsprache, Institut für Germanistik, Universität Wien
09/1995 – 04/1999	Vertragsassistentin am Lehrstuhl Deutsch als Fremdsprache, Institut für Germanistik, Universität Wien
1996 – 1999	Lehrbeauftragte anm Institut für Germanistik, Universität Wien
seit 2010	Lehrbeauftragte an der Pädagogischen Hochschule Wien
02/2011 – 04/2012	Researcher am Acoustics Research Institute/Institut für Schallforschung der Österreichischen Akademie der Wissenschaften
09/2013 – 11/2014	Projektleitung „Individuelle Erwerbsverläufe im sinnerfassenden Lesen mehrsprachiger SchülerInnen“ im Auftrag des Bundesministeriums für Bildung und Frauen und Arbeiterkammer Wien, Institut für Sprachwissenschaft, Universität Wien
2014	Lehrbeauftragte an der Universität Innsbruck, ULG Deutsch als Fremdsprache
11/2014 – 10/2016	Projektleitung „My Literacies. Zugänge zu Schriftlichkeit im Kontext von Multimedialität und Mehrsprachigkeit aus Sicht von Kindern“, Sparkling Science, Bundesministerium für Wissenschaft, Forschung und Wirtschaft, Institut für Sprachwissenschaft, Universität Wien

Wissenschaftliche Publikationen

Monographien

Plutzar, Verena/Kerschhofer-Puhalo, Nadja, Hrsg. (2009), *Nachhaltige Sprachförderung. Zur veränderten Aufgabe des Bildungswesens in einer Zuwanderergesellschaft. Bestandsaufnahmen und Perspektiven.*, Innsbruck: Studien-Verlag. 2009, 236 pp.

Kerschhofer-Puhalo, Nadja/ Krumm, Hans-Jürgen, Hrsg. (1997), *Curricula für den Deutschunterricht und die Deutschlehrerausbildung in den nichtdeutschsprachigen Ländern.* hrsg. von Nadja Kerschhofer-Puhalo und Hans-Jürgen Krumm in Zusammenarbeit mit dem Internationalen Deutschlehrerverband, Institut für Germanistik/Deutsch als Fremdsprache, Wien 1997, 260 pp.

Beiträge in Zeitschriften, Sammelbänden, Handbüchern und Proceedings

Kerschhofer-Puhalo, Nadja (in Druck a), „Fertigkeiten und ihre Wechselwirkungen im Schriftspracherwerb Deutsch als Zweitsprache", in: Pietzuch, Anja (Hrsg.), *Schriftspracherwerb im Kontext von Migration und Zweitsprache*, Proceedings der XV. Internationalen Deutschlehrertagung IDT 2013. Bozen: Bozen-Bolzano University Press.

Kerschhofer-Puhalo, Nadja (in Druck b), „'Ich habe solche Probleme mit den Umläuten'. – Ähnlichkeitsbeziehungen zwischen deutschen Vokalen", in: Koroschetz Maragno, Renata/ Missaglia, Federica (Hrsg.), *Phonetik und Prosodie*, Proceedings der XV. Internationalen Deutschlehrertagung IDT 2013. Bozen: Bozen-Bolzano University Press.

Kerschhofer-Puhalo, Nadja & Werner Mayer (2014), „Individuelle Lesestrategien ein- und mehrsprachiger SchülerInnen – ein videographischer Zugang", in: Lütke, Beate/Petersen, Inger (Hrsg.), *Deutsch als Zweitsprache – erwerben, lernen und lehren.* (Reihe Beiträge aus dem 9. Workshop Kinder mit Migrationshintergrund). Stuttgart: Fillibach bei Klett, 115-141.

Kerschhofer-Puhalo, Nadja (2012), „Phonetik und Phonologie im Schriftspracherwerb", *ÖDaF-Mitteilungen* 1/2012, 75-85.

Kasess, Christian, Wolfgang Kreuzer, Ewald Enzinger, Nadja Kerschhofer-Puhalo (2012). *Estimation of vocal tract shape of nasals using a variational Bayesian scheme.* INTERSPEECH 2012. 13th Annual Conference of the International Speech Communication Association, 9.-13. September 2012, Portland, Oregon.

Pucher, Michael, Schabus, Dietmar, Hofer, Gregor, Kerschhofer-Puhalo, Nadja, Moosmüller, Sylvia (2012), Regionalizing Virtual Avatars - Towards Adaptive Audio-Visual Dialect Speech Synthesis**.** Proceedings of the 5th International Conference on Cognitive Systems (Vienna, Austria, February 2012).

Kerschhofer-Puhalo, Nadja/Plutzar, Verena (2009), „Nachhaltige Sprachförderung", in: Plutzar, Verena/Kerschhofer-Puhalo, Nadja (eds.), *Nachhaltige Sprachförderung. Zur veränderten Aufgabe des Bildungswesens in einer Zuwanderergesellschaft. Bestandsaufnahmen und Perspektiven.* Innsbruck: Studien-Verlag. 2009, 9-13.

Pucher, Michael, Nadja Kerschhofer-Puhalo, Dietmar Schabus (2011), "Phone set selection for HMM-based dialect speech synthesis", *1st Workshop on Algorithms and Resources for Modelling of Dialects and Language Varieties (DIALECTS 2011). EMNLP 2011: Conference on Empirical Methods in Natural Language Processing*, Edinburgh, UK.

Kerschhofer-Puhalo, Nadja (2011), „Literalität und Illiteralität in der Migration", in: Schlögl, Peter, Wieser, Regine, Dér, Krisztina (Hrsg.), *Kalypso und der Schlosser. Basisbildung als Abenteuer im Land des Wissens und Könnens*. LIT-Verlag, 85-109.

Kerschhofer-Puhalo, Nadja (2010), "Preferences in L2 Vowel Perception", in: Katarzyna Dziubalska-Kolaczyk, Magdalena Wrembel, Malgorzata Kul (eds.), *New Sounds 2010, Proceedings of the 6th International Symposium on the Acquisition of Second Language Speech*, 2010, 245-250.

Kerschhofer-Puhalo, Nadja (2010), „Sprache gehört gehört – Anregungen für den Schriftspracherwerb in der Zweitsprache". *Tribüne* 1/2010 (Themenheft Sprache und Migration), LIT-Verlag, 19-23.

Kerschhofer-Puhalo, Nadja (2010), „Gruppenarbeit" (p. 111), „Kleingruppenarbeit" (p. 152-153), „Minimalpaar" (p. 215), „Polysynthetische Sprachen" (p. 255), Wörterbuchartikel in: Barkowski, Hans/Hans-Jürgen Krumm (Hrsg.), Fachlexikon Deutsch als Fremd- und Zweitsprache, Tübingen: Narr.

Kerschhofer-Puhalo, Nadja (2009), „Mehrsprachigkeit und interkultureller Dialog an Schulen", in: BMUKK (Hrsg.), *Probebühne Zukunft. Schule und interkultureller Dialog*, Tagungsband, Wien 2009, 129-134.

Kerschhofer-Puhalo, Nadja (2009), „Erwachsenenbildung. Qualität, Kontinuität und Nachhaltigkeit von Sprachlernprozessen bei Erwachsenen", in: Plutzar, Verena/Kerschhofer-Puhalo, Nadja (eds.), *Nachhaltige Sprachförderung. Zur veränderten Aufgabe des Bildungswesens in einer Zuwanderergesellschaft. Bestandsaufnahmen und Perspektiven*. Innsbruck: Studien-Verlag. 2009, 174-188.

Kerschhofer-Puhalo, Nadja (2009), „Erstsprachliche Lautwahrnehmung und ihr Einfluss auf den Schriftspracherwerb in der Zweitsprache", in: Werner Viater; Gerda Videsott (eds.), *Migration und Mehrsprachigkeit. Von der individuellen zur gesellschaftlichen Mehrsprachigkeit*, Wien/München/Bern: Verlag Peter Lang, 2009, 217-238.

Kerschhofer-Puhalo, Nadja (2008), „Wie viel Theorie brauchen wir in der Ausbildung von SprachlehrerInnen", in: Fritz, Thomas u.a. (Hrsg.), *Dokumentation zur Auftakttagung des Projekts SAPA 21: Sprachenlernen im 21. Jahrhundert: Szenarios – Anforderungen – Profile – Ausbildung*, 31.5.-1.6. 2007 Wien, Verband Wiener Volksbildung, 102-105.

Kerschhofer-Puhalo, Nadja (2001), „Organisationsformen von Lernen", in: Anz, Heinrich; Lutz Götze; Gerhard Helbig; Gert Henrici; Hans-Jürgen Krumm; Walter Veit (eds.), *Deutsch als Fremdsprache. Ein internationales Handbuch*. (HSK Handbooks of Linguistics and Computational Society 35, 1. ed.), Berlin/New York: Walter de Gruyter & Co., 761-776.

Kerschhofer-Puhalo, Nadja (1998), „Vowel substitutions in German as a foreign language", in: James, Allan R./ Jonathan Leather (eds.), *New Sounds 97. Proceedings of the 3rd*

International Symposium on the Acquisition of Second Language Speech, Klagenfurt 1998, 167-175.

Kerschhofer-Puhalo, Nadja (1998), „Transfer, Interferenz und die Rolle der Muttersprache", in*: Theorie und Praxis. Österreichische Beiträge zu Deutsch als Fremdsprache 2, 1998,* Innsbruck: Studien-Verlag. 103-129.

Kerschhofer-Puhalo, Nadja (1998), „Sprachspezifische phonetische Eigenheiten und der Transfer von Phonemen aus der Muttersprache", in: *Lernsprache Deutsch* 2/1998.

Kerschhofer-Puhalo, Nadja (1997), „Perzeption und Produktion von Vokalen in der Fremdsprache", in: de Cillia, Rudolf; Martin Stegu (eds.), *Fremdsprachendidaktik*. Wien/München/Bern: Peter Lang, 1997, 263-277.

Kerschhofer-Puhalo, Nadja (1996), „Hörübungen zur Verbesserung der Aussprache" in: *ÖDaF-Mitteilungen* 2/1996, 4-14.

Kerschhofer-Puhalo, Nadja (1995), „Sozialformen", in: Bausch, Karl-Richard; Herbert Christ; Hans-Jürgen Krumm (eds.), *Handbuch Fremdsprachenunterricht*. 3. überarb. und erw. Auflage, Tübingen/Basel: Francke UTB, 1995, 489-495.

Wissenschaftliche Abschlussarbeiten

Similarity, Cross-linguistic Influence and Markedness in Non-Native Vowel Perception, Universität Wien, Dissertation (in Abschlussphase).

Vokalsubstitutionen in der Fremdsprache. Eine kontrastive Untersuchung im Rahmen der Natürlichen Phonologie zum Erwerb des deutschen Vokalsystems durch Lernende mit polnischer Muttersprache, Diplomarbeit, Universität Wien, 1995, 196 pp.

Semiotik im Spannungsfeld zwischen Theorie und Praxis. Der Beitrag einer angewandten Medien-Semiotik für Werbung und Marketing. Abschlussarbeit Universitätslehrgang für Markt- und Meinungsforschung, Institut für Publizistik, Universität Wien, 2003.

Didaktische und bibliographische Materialien, Berichte

Kerschhofer-Puhalo, Nadja/Mayer, Werner (2014). Individuelle Erwerbsverläufe im sinnerfassenden Lesen mehrsprachiger SchülerInnen. Projektendbericht im Auftrag des BMBF und der AK Wien, Wien.

Kerschhofer-Puhalo, Nadja (2009), *Spielend Deutsch lernen. Spiele zur Sprachförderung in der Zweitsprache Deutsch*. Wien: wienXtra-spielebox.

Kerschhofer-Puhalo, Nadja (1998), *CD-ROM Deutsche Intonation*, Skriptum und Tonaufnahme erstellt im Rahmen eines Projekts der Europäischen Union, Aktion Offener Unterricht und Fernlehre, Kooperation zwischen Österreich, Deutschland, Großbritannien, Frankreich und Italien, Manuskript und Audio-Aufnahmen.

Kerschhofer-Puhalo, Nadja/Brenner, Andrea/ Katsikas, Sergios (1997), „Österreichische Publikationen zu DaF/DaZ, interkulturellem Lernen und österreichischem Deutsch", in: *Zugänge. Sonderheft 1997 der ÖDaF-Mitteilungen*, Wien 1997, 118-139.