universität wien

MASTERARBEIT

Titel der Masterarbeit

"Approaching User Experience: Moving between Modelling and In-Depth"

Verfasser

Michael Glanznig, BSc

angestrebter akademischer Grad

Master of Science (MSc)

Wien, 2013

Studienkennzahl lt. Studienblatt:	A 066 013
Studienrichtung lt. Studienblatt:	Joint Degree Programme MEi:CogSci Cognitive Science
Betreuer:	assoc prof. dr. Urban Kordeš

Abstract

User experience (UX) research in the field of Human-Computer Interaction tries to understand how humans experience the interaction with technological artefacts (e.g. computers, mobile phones, cameras, etc.). This field of research is quite young. It emerged in the nineties of the last century and is moving and evolving rapidly. Therefore, most concepts including UX are not clearly defined nor agreed upon. There is significant debate about what the concept is or represents, how to research it and some bickering between theory and practise. There are two movements that are more or less competing against each other: Hassenzahl's model of user experience and McCarthy & Wright's framework of user centred design. One focuses on uncovering the objective in the subjective, on the precise and the formal, while the other one stresses the ambiguous, the human and suggests to live with the subjectivity that is inherent in the concept of (user) experience.

Most researchers are in favour of one or the other and only few use both approaches, which has been criticised. There have been warnings not to emphasize the methodological stance to an extend where it damages research quality and others proposed to work towards a unified view. The situation in UX research might seem unique but this is not the case. It is rather a discussion between first person and third person approaches that is taking place also in other disciplines (e.g. cognitive science) and with this discussion there are also efforts to integrate both perspectives with each other. In UX research comparisons of methods are only rarely attempted and often not critically reflected. Here I am attempting a comparison, where there are two groups and group A evaluates a real estate website by using one methodology, while group B uses the other methodology. Additionally, I am reflecting on both methods from the stakeholder's (i.e. the client or customer) and researcher's viewpoint. I look at how time-consuming each approach is and how helpful the results of each approach are.

Results suggest that the method based on Hassenzahl's model is easy to employ but misses to provide explanations for its results. On the other hand, the method based on McCarthy & Wright's framework provides detailed and informative insights, but is very costly to employ and requires a skilled researcher. There is the need to add detailed context information to Hassenzahl's questionnaire while an in-depth approach like the here used Descriptive Experience Sampling should be easier to employ. This is possible through expert researchers and streamlined tools for analysis. Generally, a methodological stance should not be emphasised to the extend of damaging research quality and an integrated view from different perspectives seems to be more valuable than a unified view. In user experience research *talking* to the users in one or the other way is not a matter of choice.

Zusammenfassung

Forschung zu User Experience (UX; Nutzererfahrung) im Gebiet der Mensch-Maschine Interaktion versucht zu verstehen wie Menschen die Interaktion mit technologischen Artefakten (z.B. Computer, Mobiltelefone, Kameras, etc.) erleben. Dieser Forschungsbereich entstand in den Neunzigern des vergangenen Jahrhunderts und verändert sich schnell. Aus diesem Grund sind die meisten Konzepte wie UX nicht genau definiert und die Definitionen sind nicht allgemein akzeptiert. Es gibt eine ausführliche Debatte darüber, was UX ist oder repräsentiert, wie das Phänomen erforscht werden soll und ein paar Streitigkeiten zwischen Theorie und Praxis. Es existieren zwei Strömungen, die in Konkurrenz zueinander stehen: Hassenzahls Modell der User Experience und McCarthy & Wrights Denkphilosophie über benutzerzentriertes Design. Ein Ansatz strebt mehr nach dem Objektiven im Subjektiven und platziert den Fokus auf dem Präzisen und Formalen, während der andere Ansatz den Schwerpunkt auf das Mehrdeutige und das Menschliche legt und vorschlägt mit dem inhärenten Subjektiven in UX zu leben.

Die meisten Forscher bevorzugen den einen oder anderen Ansatz und nur wenige nutzen beide, was kritisiert wurde. Es gab Warnungen die eigenen methodologischen Vorstellungen nicht soweit hervorzuheben, bis sie die Forschungsqualität beschädigen. Andere schlugen vor auf eine vereinheitlichte Sichtweise hinzuarbeiten. Die Situation der Forschung zu User Experience scheint einzigartig zu sein, aber bei genauerer Betrachtung stellt sich heraus, dass dies nicht der Fall ist. Es handelt sich dabei mehr um eine Diskussion zwischen Ansätzen, die als first person und third person research bezeichnet werden. Diese findet auch in anderen Disziplinen (z.B. der Kognitionswissenschaft) statt. Mit dieser Diskussion einher gehen Bemühungen beide Perspektiven zu verbinden. Bei UX-Forschung sind Vergleiche zwischen Methodiken nur sehr spärlich zu finden und wenn vorhanden, dann wurden diese Vergleiche oft nicht kritisch hinterfragt. Ich versuche hier einen Vergleich mit zwei Gruppen, wo Gruppe A eine Immobilien-Webseite mithilfe einer Methodik evaluiert und Gruppe B die andere Methodik verwendet. Zusätzlich reflektiere ich über beide Ansätze aus der Sichtweise der Projektbeteiligten und des Forschers. Ich sehe mir an, wie aufwändig beide Methoden durchzuführen sind und wie nützlich die Ergebnisse der Methoden sind.

Die Ergebnisse legen nahe, dass die Methode basierend auf Hassenzahls Modell leicht einzusetzen ist, aber es verabsäumt Begründungen für ihre Ergebnisse zu liefern. Andererseits liefert die Methode basierend auf McCarthy & Wrights Denkphilosophie detaillierte Einsichten, ist aber sehr aufwändig einzusetzen. Es besteht die Notwendigkeit detaillierte Kontextinformationen zu Hassenzahls Fragebogen hinzuzufügen, während ein in die Tiefe gehender Ansatz wie das hier verwendete Descriptive Experience Sampling leichter einzusetzen sein sollte. Generell kann ich festhalten, dass eine ganzheitliche Sichtweise aus verschiedenen Blickwinkeln wertvoller als eine vereinheitlichte Sichtweise ist. Die UX-Forschung hat keine andere Wahl als mit den Benutzern in der einen oder anderen Form zu *sprechen*.

Acknowledgements

A project like this thesis cannot be completed without the help of several people and over the course of one year I needed quite some help. While I'm not able to mention all of you I still can say: Thanks to you all, without you I'd have been lost!

Many thanks to my supervisor *Urban Kordeš* who kept me on track, but let me work independently. His supervision style couldn't be any closer to the true meaning of the word *mentor*.

Thank you very much to *TAO Software* in Graz, part of *TAO Beratungs- und Management GmbH*, and especially to *Ulrich Andres* and *Roman Weinberger* for letting me evaluate your product, for discussing with me and supporting me beyond business ties.

Thank you so much to my friends who helped me getting this finished. Especially to *Harry* who always listened to my complaints and uncertainties, to *Johannes* who discussed with me and readily helped me in the pilot-study, to *Geli* who never said that she was busy (although she was) if I needed her help and to *Anna & Chris* for brightening up my stays in Graz.

Thank you to the team of MEi:CogSci and especially to *Markus Peschl* and *Elisabeth Zimmermann* for creating and running a curriculum that - unlike the current trends in university education - counts on freedom of choice, many possibilities and active support in getting into research.

Many thanks to my sister *Anita* for proof-reading this work despite her other duties and for pointing out parts that were not quite understandable.

Vielen Dank auch an meine Eltern, die mich nach so vielen Jahren noch immer mit aller Kraft unterstützen. Ihr seid die Besten!

Contents

List of Figures

List of Tables

1 Introduction

User experience[1] (UX) research in the field of Human-Computer Interaction (HCI) tries to understand how humans experience the interaction with technological artefacts (e.g. computers, mobile phones, cameras, etc.). This field of research is quite young. It emerged in the nineties of the last century and is moving and evolving rapidly. Therefore, most concepts including UX are not clearly defined nor agreed upon. HCI is an interdisciplinary field and benefits heavily from different views on the problem. On the downside different backgrounds and vocabularies do not make progress easier as we will see later where for example *dialogue* is used in different ways.

What follows is the description of a journey I have undertaken in the last one or so years to understand what UX is, how it can be researched and how to make sense of different approaches. But before I get ahead of myself I start at the very beginning: the definition of UX. While the lack of a unique definition for UX has sometimes been seen as deficiency (Law, Roto, et al. 2009) it enables us now to take a closer look at different views and their differences and commonalities.

1.1 Usability: Where it Begins

The roots of UX can be found in *usability*. It is a connected concept and some see it as enclosing UX while others say it is being enclosed by UX (Law, Vermeeren, et al. 2007). What looks like an unimportant subtlety reveals quite large differences in theoretical grounding when watching closely. ISO 9241-210 defines usability as the

> extent to which a system, product or service can be used by specified users to *achieve specified goals* with *effectiveness*, *efficiency* and *satisfaction* in a specified context of use. (ISO 9241-210 2010, p. 7, my emphasis)

I call this *Engineer's definition*. It emphasizes goal achievement and contains quantitatively measurable behavioural variables with one exception - satisfaction. While effectiveness and efficiency are measured by error rates and task completion times, satisfaction is approached with thinking aloud techniques and questionnaires. In usability engineering satisfaction was

[1] This chapter is an edited and extended version of Glanznig, M. (2012b). User experience research: modelling and describing the subjective. *Interdisciplinary Description of Complex Systems*, *10*(3), 235–247. doi:10.7906/indecs.10.3.3.

with some exceptions traditionally neglected. Possibly, because it was more difficult to measure.

This neglection became more and more unsatisfactory over time, because usage of technology changed. Computers moved out of the workplace and entered the homes. Technology-mediated leisure (e.g. multimedia, games) became more important. Recently, ubiquitous computing (e.g. smartphones) added to this progress. All of this contributed to a shift of focus from efficiency to satisfaction, which in turn caused the emergence of user experience as distinct concept (Cockton 2006; Kort et al. 2007; Bargas-Avila and Hornbæk 2011). Some have seen the emergence of UX as "old wine in new bottles" (as mentioned in Hassenzahl 2008b, p. 11; Wright and Blythe 2007, p. 66), which in my opinion overemphasizes the utility of the satisfaction part of usability and underestimates the shift of focus that has occurred.

If we look at the definition of user experience in ISO 9241-210 we find the following:

> A person's perceptions and responses that result from the use or anticipated use of a product, system or service. (ISO 9241-210 2010, p. 7)

This gives a general idea, but contains a vagueness which continues throughout ISO 9241-210. Nowadays in UX research the Engineer's definition is seen as superseded. However, nothing of the above renders usability irrelevant. It is an important, established and quite easily testable concept, which just does not tell us much about satisfaction or even experience of technology interaction. On the other hand, UX studies cannot investigate usability *per se* in detail. As we will see, UX is a more abstract and evaluative construct and (good) usability is a precondition for (good) UX. It is important that UX and usability are not being confused.

1.2 Different Views on User Experience

If we now turn to user experience we are confronted with different views. There is a vivid discussion in progress how the phenomenon should be researched. At least two movements are competing and are viewed by their proponents to be more or less opposing (Law, Vermeeren, et al. 2007). To illustrate the two competing approaches let us now move to two other definitions of UX by researchers that assume a key position in the discussion about the direction of UX research.

> An experience is an episode, a chunk of time that one went through-with sights and sounds, feelings and thoughts, motives and actions; they are closely knitted together, **stored** in memory, **labelled**, **relived** and **communicated** to others. An experience is a **story**, emerging from the **dialogue** of a person with her or his world through **action**. *User Experience* is not much different from experience *per se.* (Hassenzahl 2010, p.8, my emphasis in bold)

In what I call *Psychologist's Definition* Marc Hassenzahl (ibid.) emphasizes that an experience is a complex construct, which emerges through interaction with the world. User experience is very similar to experience in general.

> For [John] Dewey, experience is constituted by the **relationship between self and object**, where the self is always already engaged and comes to every situation with personal interests and ideologies. … **action is situated** and creative. … For [Mikhail] Bakhtin, the unity of felt experience and the meaning made of it are never available *a priori* but must always be accomplished **dialogically**. (Wright and McCarthy 2010, pp. 17-18, my emphasis in bold)

In what I call *Humanist's Definition* John McCarthy and Peter Wright (ibid.) place the focus on the holistic nature of an experience and how meaning is made of it. Both definitions use overlapping vocabulary (e.g. dialogue), but they attach different meaning to it. While Hassenzahl stays heavily grounded in psychological research and its methods, McCarthy & Wright take a more interpretive and qualitative approach. We will come back to the two accounts and their differences and similarities later.

1.3 UX Research & Engineering: Theory vs. Practice?

Let us compare the two latter definitions of user experience with the definitions of usability and UX in ISO 9241-210. We might notice that there are quite some differences between the engineer's point of view and that of UX researchers. These differences result in difficulties when both talk to each other and also when results of UX research try to influence software or systems engineering. UX researchers criticise engineering for still not looking beyond functionalism: "When the focus of a community is so tightly trained on the functionality of systems and how they can be made more accessible and usable, experience is an outsider concept" (McCarthy and Wright 2004, p. 3), "a product should not longer be seen as simply delivering a bundle of functional features and benefits" (Hassenzahl 2005, p. 31). The response then may sound polemic: "Don't have to know what it is like to be a bat to build a radar reflector" (Kerkow 2007). Both positions seem somewhat reasonable. Researchers worked hard to justify the claim "functionality and usability are just not enough" (Hassenzahl 2005, p. 31) and create what is known as user-centred design (McCarthy and Wright 2004; Stegemann and Fiore 2006; Hassenzahl 2010; Wright and McCarthy 2010). In contrast engineering often calls for a pragmatic concept (Kerkow 2007) that can be embraced in a cost-effective and easy way.

So far the focus has been more on the sometimes problematic relationship between user experience and usability and the debate between UX research and engineering. This enabled us to see the area of tension in which UX research as (still) emerging field finds itself. For an excellent critical analysis of empirical studies in UX that addresses these issues see Bargas-Avila

and Hornbæk (2011). For an overview of the history of HCI see Cockton (2006). Let us now explore the two different strands of UX research we looked at earlier.

1.4 Modelling User Experience

The psychologist Marc Hassenzahl (2010) uses James Russell's account of emotional experience, hierarchical goals and related action theories to develop his own model of user experience. As stated in his definition he views UX as not being very different from experience as such, the difference being the focus on a specific mediator of experiences – e.g. interactive products. He stays heavily grounded in psychological research and its methods. He explicitly distinguishes himself from authors such as John McCarthy and Peter Wright (McCarthy and Wright 2004; Wright and McCarthy 2010), who are proponents of a holistic and dialogical approach. Hassenzahl calls this "phenomenological" (Hassenzahl 2010, p. 73) approach and also hints at a possible extension of his model with such approaches (ibid., p. 74).

A main point of critique towards Hassenzahl's research is its reductionist nature (Stegemann and Fiore 2006) that sees the user as action/reaction system (Wright and McCarthy 2010, p. 6) while UX being a complex and possibly irreducible construct. In this vein the validity of the model is questioned. The author's argument against this kind of critique is that his research is not so much a reduction than a necessary categorisation and usage of well-researched (psychological) models and theories. Additionally, he suspects experiences with technology to be far less unique and variable as the critics might imply (Hassenzahl 2008b). Following psychologist James Russell (2003) Hassenzahl views emotional experience as consequence of self-perception and categorisation and as construction of a coherent and emergent, albeit complex, narrative in dialogue with the world. The great amount of single aspects that are integrated into an experience make it seem to be so unique and irreducible. Emotions and experiences may not be fully explainable and predictable from single underlying elements but they are not detached from them. (Hassenzahl 2010, p. 4)

1.4.1 Essential Properties of Experience

For Hassenzahl experience has the following attributes: subjective, holistic, situated, dynamic and positive (in the sense of worthwhile) (ibid., pp. 9-31). *Subjective* (ibid., pp. 9-11) means that experience is created and remains in the experientor's head. Objective values (e.g. task completion time) may be experienced differently (subjectively). However, this gap or mismatch can be described by rules. Therefore it is possible to shape experiences by knowing and using these rules.

Given a hierarchy of goals such as motor-goals, do-goals and be-goals (listed bottom to top), which may be "dialling in numbers", "making a telephone call" and "feeling related to others",

interaction design traditionally focused only on do-goals and below. The author refers to the necessary extension of HCI with the meaning providing be-goals as *holistic* (ibid., pp. 11-16).

He also acknowledges the *situatedness* (ibid., pp. 16-19) of single experiences - two of them are never alike. Descriptive approaches are therefore at a lost position (ibid., p. 17). Instead, categorisation of experiences enables us to compare reality to prototypes of experiences. This is possible because accounts of particular experiences might differ, but the essence of the experience itself does not. Hassenzahl develops a form of categorisation based on needs which he calls experience patterns (ibid., pp. 17, 76). It has been shown that needs are relatively independent from each other and (positive) experiences are often marked by a particular need (ibid., p. 47). Experience patterns can be seen as a blueprint of various experiences, a condensed, idealised and optimised version.

Experiences change over time. They are *dynamic* (ibid., pp. 19-27). Hassenzahl sees an experience as story. It is packaged, interpreted and labelled, and constructed, but not an objective account of the experience. However, he views the actual construction as only happening once and then being remembered unaltered.

In contrast to usability engineering, which focuses on problems and their removal (the difference between a bad and acceptable experience), an experiential approach strives to make an experience *positive* (pleasurable, good) (ibid., pp. 27-31). "Positive experiences we went through hold more power to increase well-being than any material possession." (ibid., p. 40) Need satisfaction (as motivation for an experience) is rarely an explicit goal, but an emergent property.

1.4.2 The Model

Hassenzahl calls his model the hedonic/pragmatic model of user experience (Hassenzahl 2005; Hassenzahl 2007). It has two different quality dimensions: *pragmatic* and *hedonic quality*. We have already learned about the hierarchy of goals he builds upon: motor-goals, do-goals and be-goals (bottom to top). Pragmatic quality now refers to the product's perceived ability to support the achievement of do-goals (e.g. making a telephone call). Hedonic quality means the product's perceived ability to support the achievement of be-goals (e.g. being related to others). (Hassenzahl 2010, p. 49) These dimensions open up a two dimensional space in which a product can be placed with high values on both dimensions being desirable (Hassenzahl et al. 2003). Pragmatic quality is more focussed on the product, while hedonic quality focuses on the Self (Hassenzahl 2007). The main assumption of the model is that these dimensions are viewed as unrelated. Hassenzahl: "In fact, all studies published so far support this notion." (Hassenzahl 2010, p. 50)

How does usability relate to user experience in this model? Hassenzahl argues that the fulfilment of be-goals is the driver of experience (Hassenzahl 2008b). Usability is rather associated to the product and to do-goals. User experience is associated to the Self and be-goals. Lack of usability can be a barrier to the fulfilment of be-goals, but usability is in itself not desired (ibid.). In other words, (good) usability is only a precondition of (good) UX.

All right, but how can the product's perceived ability to support the achievement of do- and be-goals be assessed? Here Hassenzahl believes that it is possible to describe and characterise people's experiences with the help of a questionnaire, which he sees as promising strategy for HCI (Hassenzahl 2010, p. 56). For this purpose the AttrakDiff questionnaire (Hassenzahl et al. 2003) has been developed and validated. It comes in the form of a so-called semantic differential with twenty-one seven-point Likert scaled bipolar items with verbal anchors (e.g. confusing – clear, good – bad, ugly – beautiful etc.) (ibid.). The questionnaire has three subscales: perceived pragmatic quality (PQ), perceived hedonic quality-stimulation (HQ-S) and perceived hedonic quality-identification (HQ-I).

1.4.3 Using The Model

We now leave the theoretical realm of Hassenzahl's model but kind of stay in the lab to look at some work that has been done with the AttrakDiff questionnaire. We start with two studies by Marc Hassenzahl (Hassenzahl 2008a) where the interplay between perceived pragmatic attributes (PQ), hedonic attributes (HQ) and beauty of MP3-player skins has been investigated. Related work on beauty and usability has been done by Tractinsky et al. (2000) on ATM layouts. While pragmatic and hedonic attributes are perceived qualities, beauty is an evaluative construct. Hassenzahl emphasizes the fact that "perceptions of hedonic or pragmatic attributes can *potentially* lead to a positive evaluation but they must not necessarily do so." (Hassenzahl 2008a, pp. 322-323, original emphasis)

The results of the first study did not support the clear relation between usability (PQ) and beauty that has been reported by Tractinsky et al. (2000). Comparing ugly and beautiful skins (rated by participants) revealed greatest differences for hedonic quality-identification (HQ-I), followed by hedonic quality-stimulation (HQ-S) and pragmatic quality (PQ). As a major limitation participants in the first study only saw the interfaces, but never interacted with them (Hassenzahl 2008a, p. 333). Therefore, in the second study participants also interacted with the product after rating the interface and were allowed to revise their rating after interaction. (ibid., p. 335) Interestingly, pragmatic attributes were affected by experience, but hedonic attributes remained stable in both ratings (ibid., p. 340). A related study has investigated the constructs beauty and goodness further using websites (van Schaik and Ling 2008).

Another study has investigated the influence of usage mode (explorative vs. task-oriented) on perceived quality (Wechsung et al. 2010). The research question was motivated out of the impression that "it is likely that success rates in traditional usability tests are higher than in natural settings." (ibid., p. 189) The participants interacted with an "ultra mobile personal computer" and had either to perform a task-oriented block and then an explorative block or vice versa. Additionally, they could choose between the input modalities touch input or voice control. The results showed that task-oriented settings reduce the experienced identification with the system and the overall attractiveness (ibid.). Pragmatic quality was strongly correlated to

overall attractiveness in both usage modes, which is contradictory to what has been found by Hassenzahl (Hassenzahl 2008a, p. 323).

1.5 Describing User Experience

The computer scientist Peter Wright and the psychologist John McCarthy (McCarthy and Wright 2004; Wright and McCarthy 2010) use John Dewey's pragmatist philosophy of experience and aesthetics (Dewey 1934) and Mikhail Bakhtin's account of dialogue as grounding to develop their approach towards experience-centred design. They see the term *user* in user experience as problematic as it suggests a limited view on a person, like that of a tool user. In their view one has to think of persons holistically: What they do, how they feel about it and how they give meaning to it. People have a past, a present and a future. Their history is part of what defines them as a person, embedded in complex and changing social networks. (Wright and McCarthy 2010, p. 63)

The authors therefore suggest taking a more interpretive and qualitative approach towards user experience. They see experience-centred design as designing for the richness of human experience (ibid., p. 2). For them experience-centred design is not simply about technology, it is about people's lived and felt experience (their felt life), which is sometimes mediated by technology (ibid., p. 3). The authors despise any attempts to exploit their concept for business use only: "Experience-centred design must not become exclusively a business strategy" (ibid., p. 9). McCarthy & Wright also reject the usage of methods as recipes (ibid., p. 90), because they think that research on experience is "not suited to fixed research designs and procedures" (ibid., p. 83).

In the authors' view individuals as embodied in their lifeworld (ibid., p. 14) that have to make sense of it. This sense making is a highly subjective and introspective process, which is also irreducibly social and is connected to voice and narrative (ibid., p. 19). In this sense sharing an experience involves a common history, a common ground, something of which stories can be made. Stories can be seen as edited versions of our lived experience (ibid., p. 20). Meaning is not inherent in them (and in experience) and cannot be a logical inference of it (ibid., p. 21). Therefore, separations and reductions (e.g. as in usability engineering and affective computing) oversimplify the lived experience and miss the crucial point (ibid., p. 14). Also, the user is traditionally seen as subject and the designer as objective gatherer of data, which is problematic. Understanding experience requires involvement and not just observation (ibid., p. 23). It requires dialogue and not just surveying (ibid., p. 70).

Doing research in experience-centred design can be viewed as the construction and reconstruction of stories of people's experiences with technology (ibid., p. 37). However, stories of experiences come not ready-formed. Instead they are brought into being in dialogue and emerge between speaker and listener(s) (ibid., p. 39). Dialogue or dialogism puts the emphasis on the process between communicating people instead of what happens within each of them

(Wright and McCarthy 2010, p. 51). There is also a similar notion in art theory that is named dialogical or relational aesthetics (cf. Dewey's 1934 notion of interaction between subject and object in art). The authors believe that new meaning arises through engagement with the other person (Wright and McCarthy 2010, p. 54). The dialogical approach treats relationships and communication as privileged to understanding experience (ibid., p. 86). But simply sitting down with people saying "tell us your story" will not work. That is because people are used to construct scripted and stereotypical accounts of themselves (cf. Jerome Bruner's 2004 research on life narratives). The result may be accounts that are carefully tailored to what the persons think is needed by the researcher. In addition the whole picture also entails much that is not even obvious to the persons themselves. (Wright and McCarthy 2010, p. 64) Other researchers use similar notions. For example Russell Hurlburt et. al. use something they call *expositional interview* for their descriptive experience sampling technique (Hurlburt and Heavey 2006): "We call it the expositional interview to indicate that our intent is to expose (to make known, bring to light) what is hidden from us but present to the subject (though not necessarily, at first, clearly known to the subject either)." (ibid., p. 86)

McCarthy & Wright's (2004) work has received some criticism questioning the reliability of their approach. They used Wright and McCarthy (2010) to clarify their position, but did not explicitly respond to their critics. Hassenzahl (Hassenzahl 2008b; Hassenzahl 2010) has tried to distinguish his own research from McCarthy & Wright's position and doubts that the immense richness and diversity in experience as suggested by McCarthy & Wright exists. Accounts of experiences might differ: "A poet may find beautiful words" (Hassenzahl 2008b, p. 14), but experience or at least the essence of it does not. At the same time Hassenzahl also acknowledges that a "phenomenological-oriented" approach is better suited to provide a detailed understanding of the people and the context (ibid.).

1.5.1 Threads of Experience

The authors provide us with four guiding threads to describe experience of technology. (McCarthy and Wright 2004, pp. 79-104) These threads should not be understood as fundamental elements or categories. They are: the sensual, the emotional, the compositional and the spatio-temporal thread. (ibid., p. 80)

Through our sense organs we participate directly in the world around us. The *sensual thread* of experience is about our sensory engagement with our environment, which orients us to the visceral character of experience. Part of this sensory engagement and therefore the interaction is also the body and the physicality of the technology. (ibid., pp. 80-83)

The *emotional thread* refers to value judgements that, according to our needs and desires, make other people and things important to us. Perceiving, thinking and deciding are not the computational processes we might think, instead they are influenced by values, needs, desires and goals. Thus, we do not perceive an objective representation of the world but a unique version that is coloured by our values. (ibid., pp. 83-85)

The *compositional thread* refers to relationships between the parts and the whole of an experience (like the relation between elements of a painting and between painting, viewer and setting). (ibid., pp. 87-91)

A *spatio-temporal* component is inherent in all experiences. For example our sense of time might change when we are bored or within an intense experience. Frustrating experiences can transform a space into something confining. (ibid., pp. 91-94) We might first enjoy the vastness of the landscape on a mountaintop and later be frightened by the steepness of a cliff edge on the same mountain.

1.5.2 Making Sense of Experience

McCarthy & Wright emphasize the sense making process of experience that occurs dialogically: "Understanding or making sense of an experience occurs in the tension between self and other." (ibid., p. 73) In this dialogue the experience is relived and also altered. The produced narratives of experience are selective interpretations that are tailored to a specific audience. (ibid., pp. 118-119) The authors present six processes of sense making with no implication of linear and causal relations between these processes. They are: anticipating, connecting, interpreting, reflecting, appropriating and recounting. (ibid., pp. 124-127)

We do not arrive at an experience without expectations. We *anticipate* something. This not only happens prior the experience but also continues later on. (ibid., p. 124)

The term *connecting* refers to the immediate, pre-conceptual and pre-linguistic sense of an encountered situation. This may be an apprehension of speed or movement or stillness. It may also mean an immediate sense of tension or a thrill of novelty, a sense of relief or the anticipation of something happening. (ibid., p. 125)

When *interpreting* an occurring experience we have to discern the narrative structure, the involved agents and action possibilities. We look at what has happened and think about what is likely to happen. This can result in anxiety of not knowing. We may feel disappointment at unmet expectations. (ibid., p. 125)

At the same time of interpreting an experience we may also *reflect* on it and make judgements about it. We may want to see how we feel about things and if we have reached our goals (if there were any). This is like an inner dialogue that helps us to meaningfully recount the experience to others. (ibid., p. 126)

Appropriating means making the experience our own by relating it to our Self, our personal history and our anticipated future. By putting the experience in the context of a past and a future we create a meaning that is more personal to us. (ibid., p. 126)

Recounting involves telling the experience to others or ourselves. It gives us the opportunity to savour it again, place it in the context of other experiences and find new meanings in it. (ibid., p. 127)

1.5.3 A Toolbox for Practice

When it comes to methods McCarthy & Wright think that research on experience is "particularly difficult to express in a procedure" (Wright and McCarthy 2010, p. 83). And indeed, they do not offer one. What they are offering is a framework of thought where certain methods fit into. Namely, methods that "open up dialogue between designers, researchers and participants" (ibid., p. 83). These methods mostly originate in art practice, in the humanities and in the social sciences. McCarthy & Wright note that researching experience "requires an individual to develop the sensibilities of a good ethnographic researcher" (ibid., p. 83). Apart from some "homegrown" methods the authors list some methods from the social sciences: ethnography, interviewing, diary studies, focus groups, repertory grids and card sorting.

There are a variety of other methods for design or evaluation (see Wright and Blythe (2007) for a more detailed overview) that can be used. Quite well known are *cultural probes* by Gaver et al. (1999), where participants are given probe packages to provoke inspirational responses. Another popular method is *experience prototyping* by Buchenau and Suri (2000) that builds upon the "experience it yourself" stance. McCarthy & Wright do not mention it, but it certainly fits here: *Descriptive experience sampling* (DES) by Hurlburt and Heavey (2006), where participant's experiences are randomly sampled and later on it is tried to uncover the essence of the sampled experiences through interviews. Not that different to DES is the *day reconstruction method* by Kahneman et al. (2004). Here participants systematically reconstruct their activities and experiences of the preceding day while trying to minimize recall biases. The *fictional inquiry technique* by Dindler and Iversen (2007) tries to create partially fictional settings and artefacts through a shared narrative. This should provide a space for collaborative design activities and help participants imagine desirable futures. Blythe & Wright use fiction as a resource in their *pastiche scenarios* method (Blythe and Wright 2006) to write character-based scenarios. They re-use existing (well-known) characters from fiction to recruit "a pre-existing rich understanding of the character-users and the use context" (ibid., p. 1142). Bertelsen & Pold draw upon aesthetics and literary or art criticism to advance their *interface criticism* technique (Bertelsen and Pold 2004). Swallow et al. (2005) developed techniques such as *persona matching*, where participants are recruited according to predefined personas and *"Do something"-challenges*. Here participants were able to select some emotional adjectives from a list and then carry out activities with the artefact they found to be representative with these descriptions (e.g. Do something funny / sexy / surprising … with your mobile phone.).

Let us now look at how to analyse the data. The above-mentioned methods mostly produce qualitative data so the researcher will end up with field notes or some transcript. This data could then be analysed with e.g. Grounded Theory, Content Analysis, Narrative Analysis etc. Whatever method is used, it should be able to capture the holistic and dialogical qualities of experience (Wright and McCarthy 2010, p. 85). McCarthy & Wright stress the point that, when analysing the data, one has to bear in mind that design implications cannot be inferred without any creative or imaginative intervention of the person doing it. It is not possible to do it in a logical deductive manner. It is more like seeing a situation from different perspectives.

Theories can serve as a guiding filter and a resource for dialogue but one should avoid the finalizing tendencies of approaches that assume that there is one correct theory or one possible best solution. (ibid., p. 67)

1.6 A Step Back: First Person and Third Person Approaches

At first sight the situation in UX research with two competing strands of research by Marc Hassenzahl (Hassenzahl 2010) and John McCarthy & Peter Wright (McCarthy and Wright 2004; Wright and McCarthy 2010) might seem unique. When looking closer it turns out that this is not the case. In consciousness research in general we find similar arguments as we have seen in UX research, like the one that conscious experience is not approachable by reductive theories (see Markič 2012). Kordeš provides an illustrative example of competing methods in researching cognition. He writes that currently cognitive neuroscience is seen as promising path to understanding the functioning of the brain. Looking at history, however, reveals that there have been already a couple of disciplines that have raised similar hopes: artificial intelligence, psychology, cybernetics and philosophy – and have failed to fulfil them (Kordeš 2012, p. 224). In the end it is the view from different perspectives that might succeed and it is the task of cognitive science to provide a framework for integration (Markič 2012, p. 214). Back at the example this integration in the case of cognitive neuroscience could be trying first to understand human experience before its neural correlates are being studied (Kordeš 2012, p. 224). Actually this is already being done by first attempts in neurophenomenology with e.g. the combination of experiential reports and MRI imaging.

What does this mean for UX research? The different approaches try to study the same phenomenon - user experience. Apart from that the two presented strands of research are quite different. One focuses on uncovering the objective in the subjective, on the precise and the formal, while the other one stresses the ambiguous, the human and suggests to live with the subjectivity that is inherent in the concept of (user) experience. One focuses on evaluation rather than design and the other one rather on design than evaluation. One is a model and the other one rather a framework of thought. Both can be criticised. The model can be questioned in terms of validity and the results of the other approach do not easily generalise across contexts – their reliability can be questioned. A unified view in UX research is sometimes emphasised as desirable (Law, Hvannberg, et al. 2006; Law, Vermeeren, et al. 2007; Mahlke 2007). This is difficult as the approaches are based on different traditions, which are subject to intense debate and no resolution of this debate in near future can be expected. For example, Cockton expresses his displeasure on determinism in computer science: “Objectivity is preferred over subjectivity, precision over looseness, automation over human agency, and formality over ambiguity” (Cockton 2006, p. 102) This also applies to other disciplines and while I agree with his concerns others might as well criticise this subjectivity, looseness and ambiguity as unscientifically.

Still, combinations of methods are possible and could be beneficial. Unfortunately, this is not done very often as Bargas-Avila and Hornbæk criticise in an analysis of empirical studies of UX: "Some studies overemphasize their methodological stance to the extent of damaging research quality" (Bargas-Avila and Hornbæk 2011, p. 2696). Only few studies try to combine what they call "uniqueness studies" (like McCarthy & Wright; ibid., p. 2696) and "dimension studies" (like Hassenzahl; ibid., p. 2696). One of these few is the study by Karapanos et al. (2009) that investigates the temporality of user experience (i.e. its development over time) with iPhone users. They used the day reconstruction method (Kahneman et al. 2004) to capture "rich qualitative accounts" (Karapanos et al. 2009, p. 731) of experience. Participants were asked to pick the three most impactful experiences of one day and write a small story about it, which the authors call experience narration. For each narration participants rated the product using a shortened version of the AttrakDiff questionnaire. The collected experience narratives were analysed using a conventional qualitative content analysis and different phases of product adoption were identified. These phases were then related to the overall perceived quality of the product using the results of the questionnaire. The integration of qualitative and quantitative methods strengthened their arguments, better enabled them to relate their findings to other studies and increased the possibility of informing other research.

What is lacking so far is reflecting on how well each approach works in practice and what kind of data are more helpful for which questions and for which stakeholders. Methodological comparisons are rare. As Bargas-Avila and Hornbæk note: "New methods are merely used without comparison to other methods, or the comparisons are weak. We see much opportunity here to improve our understanding of the relative merits of methods aimed at assessing or evaluating UX" (Bargas-Avila and Hornbæk 2011, p. 2696).

1.7 Sampling Inner Experience

So far we have learned about one method for researching UX: Hassenzahl's model and the AttrakDiff questionnaire. We also learned about McCarthy & Wright's framework of thought where several methods of section 1.5.3 (or similar ones) would fit into. We now have to *pick* one of those methodologies depending on whether the focus lies on design or evaluation. Here it will be evaluation and I have chosen an Experience Sampling flavoured methodology. Of this kind at least two related methods and several modifications exist. The two are Experience Sampling Method (ESM, Csikszentmihalyi and Larson 1992) and Descriptive Experience Sampling (DES, Hurlburt and Heavey 2006). Both build upon *in situ* measurements and deliver random beeps (alarms) to participants. For each beep they are required to answer questions or take notes about their activities or experiences. In my opinion any of the modifications would fit into McCarthy & Wright's framework of thought (McCarthy and Wright 2004) if it included a crucial element - dialogue.

Dialogue is a central element of DES and is called *expositional interview*. There the beeped experiences are relived and it is tried to uncover their content and structure. For that reason DES

requires some training for participant and researcher. Performing the expositional interview is an art for itself and the participant might not be used to pay attention to what is going on at the beep (Hurlburt and Heavey 2006, p. 2). Therefore, data that have been acquired at the first sampling day should be discarded (ibid., p. 15).

Hurlburt and Heavey adopt a similar stance as McCarthy & Wright. They see their participants as co-researchers: "You have something I need in order for this project to advance – your inner experience. You are the expert about that, and I am totally lost without your participation. On the other hand, I have a method and expertise in exploring that inner experience, so I can help you understand the details of your experience. Together, we have a good shot at learning something interesting, but neither of us can do it alone" (ibid., p. 81). For DES this does not only mean viewing the participant as partner as McCarthy & Wright do it, but also that participants have to be prepared to search within themselves. Nevertheless, this co-researcher stance and the expositional interview are the reasons why I think that DES fits perfectly into McCarthy & Wright's framework. We will come back to the expositional interview later, let us start with the important aspects of first-person data acquisition.

1.7.1 Introspection and Memory

Inner experience is not directly accessible, often not even to the experientors themselves. It needs *looking into* oneself (introspecting), either alone or with the help of another person in dialogue. Unfortunately, introspection was not very successful as a method at the beginning of the twentieth century and was subsequently discarded by psychology. Hurlburt and Heavey argue that not introspection as such was flawed but the way how it was conducted. In their view it is "possible to make accurate introspective observations if, but only if, an adequate method is used" (ibid., p. vii) and they believe that DES is such a method.

A central aspect for gaining access to inner experience is memory. What we learn from psychology, however, is that memory is prone to a variety of errors. Therefore, it is important to keep in mind that retrospective self-accounts are often incorrect even when the person is very confident about them. (ibid., p. 47) It is more likely to get accurate descriptions of inner experience when we refer to recent experiences (ibid., p. 48). Some features of inner experience that might be important to introspection (e.g. what was said in an inner voice) do not usually belong to the meaningful parts of an experience that are encoded. These features are not likely to be reported accurately unless the expositional interview takes place very soon (within 24 hours) after the experience (ibid., p. 37). Characterisations of experience over time are likely to be distorted by self-theories of the person. Therefore, introspection should target clearly identified moments and keep the amount of information to be remembered as small as possible (ibid., pp. 49-50).

At the beginning of sampling people often find it difficult to pay attention to their inner experience. This might be due to the fact that our society does not train people in doing that com-

pared to e.g. Eastern societies (Hurlburt and Heavey 2006, p. 32). However, as sampling progresses it typically gets easier and easier for them to observe what is ongoing within them.

1.7.2 Characteristics of the Method

The method is designed in a way that helps making a clear distinction between descriptions of inner experience that have been collected under specified conditions and descriptions of inner experience that are retrospective or general (ibid., p. 10). It also helps in making a clear distinction between ongoing thoughts before the beep that would have happened anyway and thoughts that were triggered by the beep (ibid., p. 15).

When starting with sampling people often refer to their inner experience as *thinking*. Despite this similarity there is a lot of variability between persons in what they intend with the word *thinking*. Some might say something to themselves, others might see a visual image and others might have no symbolized thoughts at all at this moment (ibid., p. 36).

The DES method does not explore the unconscious. It can only describe aspects that have been somehow known to the person even if they were hidden before (ibid., p. 7). Co-researchers should not be asked to provide more than just descriptions of phenomena. DES is not designed to explore causation and people cannot be expected to answer *why-questions* accurately. (ibid., pp. 54-55)

DES places special emphasis on the randomness of the beeps and the frequency of the phenomena. Randomly chosen moments may discover important characteristics of a person and each phenomenon gets more important the more frequent it occurs in different samples (ibid., pp. 7, 69). Relying on sampling frequency gives more accurate insight into inner experience than questionnaires can typically offer, but this comes at the cost of higher effort for sampling. Hurlburt and Heavey offer an example. They showed that people, who endorsed such a sentence as *I am sad all the time* from a depression inventory, were actually sad in fewer than half of their sampled experiences. Therefore, they argue, there is no substitute for collecting a large number of samples from randomly chosen moments (ibid., p. 37).

1.7.3 The Procedure

The DES method requires co-researchers to carry a beeper when they go about their daily business. A beeper is simply any device that emits a signal at random times - the *beep*. In my case it will be an app for Android smart phones, which is described later. When the beep comes persons should take notes describing their inner experience ongoing the moment just before the beep. These notes are primarily for themselves so that they can remember and relive the experience at the interview.

Inner experience is private per definition. Sometimes one does not want to talk about an experience or about portions of it. The authors clearly state that in this case people should just

simply say that a beep is none of the interviewer's business instead of just omitting sensitive parts. Typically, in the explicitation interview interviewers will keep questioning until they feel that they have a clear understanding of the experience. Saying that something is none of the interviewer's business avoids lengthy questioning (ibid., p. 82). On the other hand, if the co-researchers tell something about their inner experience it should be the truth, and nothing but the truth. Similarly, sometimes questions in the interview can seem to be hard or impossible to answer or they are simply not understandable. In this cases the authors encourage the co-researchers to simply say that they do not know or something like that (ibid., p. 84).

Above we have talked about the fact that the restriction to a certain moment helps in introspection because it reduces the amount of information that have to be remembered and reduces the probability of distortions by self-theory. Hurlburt and Heavey describe this moment as "that microsecond just before your awareness was disturbed by the beep" (ibid., p. 84), which does not include any reactions to it. "It's rather like a flash photograph. The very beginning of the beep is the flash. I'm interested in what was ongoing right at the moment of the flash. The flash very often makes you blink, but the photo records your face immediately prior to the blink." (ibid., p. 84)

Hurlburt and Heavey also emphasize that the method is primarily interested in "*what* the experience was like, *not* in *why* it had the content or features it had. Furthermore, I'm not interested in whether a particular sample is typical or unusual." (ibid., p. 84, original emphasis) Typical experiences will usually occur often during sampling and these are the experiences that are important.

1.7.4 Sampling a HCI Context

DES was designed for sampling inner experience as it happens during a normal day. The authors did not impose any restrictions on sampling context. When restricting the context to a specific one, in my case Human-Computer Interaction, some things are different and have to be re-considered. First, in the original DES procedure the random beeps come on average between half an hour and one hour. For HCI this interval is way too long. A typical interaction with a technological artefact (e.g. your phone) is quite brief, set aside the fact that some spend hours with their computers or devices. Even then, if we are interested in the interaction with a specific software and not any one the usage period might be considerably shorter. Therefore, the average interval between two subsequent beeps has to be rather short. Making it too short will annoy the user and might not capture much differences in inner experience. Making it too long will only yield one or even no sample per interaction with the device or software. I have used an average interval of five minutes with a maximum of ten minutes.

Another aspect is the moment of the beep. It is an important time anchor point that minimizes reflection, memory load and distortion by self-theories. Hurlburt and Heavey stress the fact that DES practitioners should stay at the moment of the beep under all circumstances. While

I agree with them I think it is desirable to weaken this rule just a little bit. When sampling an interaction we might lose essential parts of this interaction by being too strict about one specific moment in time. An interaction is extended in time so we should allow the coverage of a small time frame. Under no circumstances we should extend this time frame too much, because that will again introduce a lot of reflection and self-theories.

In the original DES method the authors state that they "have found that the *how* of experience is more important than the *about what*" (Hurlburt and Heavey 2006, p. 29, original emphasis). While this is certainly true, for HCI purposes it sometimes will be sufficient to know that something was present (e.g. feeling angry or frustrated) but not necessarily in detail how this (e.g. anger or frustration) was experienced. Here we have to be careful not to skip something important, so it is safest to explore the experience quite detailed at the interview. Similarly, the context of the beep or what the person was doing at the moment of the beep becomes more important in the HCI condition, because we want to know what might have *caused* this experience. Here we also need to be careful as the inference of causation usually needs a substantial amount of data and the simple concurrence of experience and action might also be coincidental. Nevertheless, McCarthy & Wright help us out when they write that it is mostly unjustified to think "that the design implications can be inferred without any creative or imaginative intervention of the person doing the analysis" (Wright and McCarthy 2010, p. 67).

Because of this differences the approach I have used for HCI should be called a *modified form* of Descriptive Experience Sampling. I think it is close enough to the original method that calling it like that is justified.

2 Study Design

So far we have heard a lot about different concepts and methodologies that can be used in user experience (UX) research to evaluate a product or to guide design. My focus will be on evaluation. Before I go on with describing all the details of how I have attempted this evaluation and what is the product that is being evaluated, let me shortly summarize what I am aiming at.

2.1 The Aim: Comparing and Reflecting on Methodologies

We have seen that the debate between first person and third person approaches is not unique to UX research. In other disciplines (e.g. neuroscience and philosophy) there are similar concerns about the *right* way and we see efforts to integrate both perspectives with each other (see Kordeš 2012). The problem of integration can be seen in various ways (see Markič 2012): We can see it simply as different levels of analysis, as different approaches or as based on different epistemological foundations or traditions. For UX research Bargas-Avila and Hornbæk (2011) emphasize the need for combinations and comparisons of methods. Unfortunately, such combinations or comparisons are only rarely attempted (see Karapanos et al. (2009) for an exception) and often not critically reflected.

2.1.1 Research Questions

This lack of comparisons and reflection motivates my research questions:

- What is missed out and discovered by each approach in comparison to the other?
- Why where facts missed out or discovered?
- How time-consuming or costly is each approach?
- How helpful is the outcome of each approach as perceived by stakeholders (i.e. client or customer)?
- How helpful is the outcome of each approach as perceived by author?

2.1.2 Methodology

I am attempting a comparison with two groups (between-group design), where group A evaluates the product by using one methodology, while group B uses the other one. A combination would have required participants to evaluate the product using *both* methodologies (within-group design). For a successful comparison both groups need to carry out the same task while staying within the constraints of each method. Each method should follow a typical design adapted to this specific evaluation. After analysis of the data follows a reflection phase with stakeholders. This stakeholder evaluation has been done qualitatively in a focus group-like setting combined with presenting the results of both approaches.

2.2 The Product: A Real Estate Website

A real estate website made by TAO Software has been evaluated. This website uses a map-centric interaction mode, which is quite different from what is available today. People can search for real estate in a location-based way. A map is the main element on which flats and houses are located (offered). These flats and houses can be searched, filtered and starred (marked as favourite). For each object there are descriptions (e.g. price, area, number of rooms), pictures and videos. If the object is of interest the owner or a real estate agent can be contacted directly using the website. Various additional data like public transport, location of stores, restaurants and cafés shall help to make a decision. Figure 2.1 shows screenshots of the user interface.

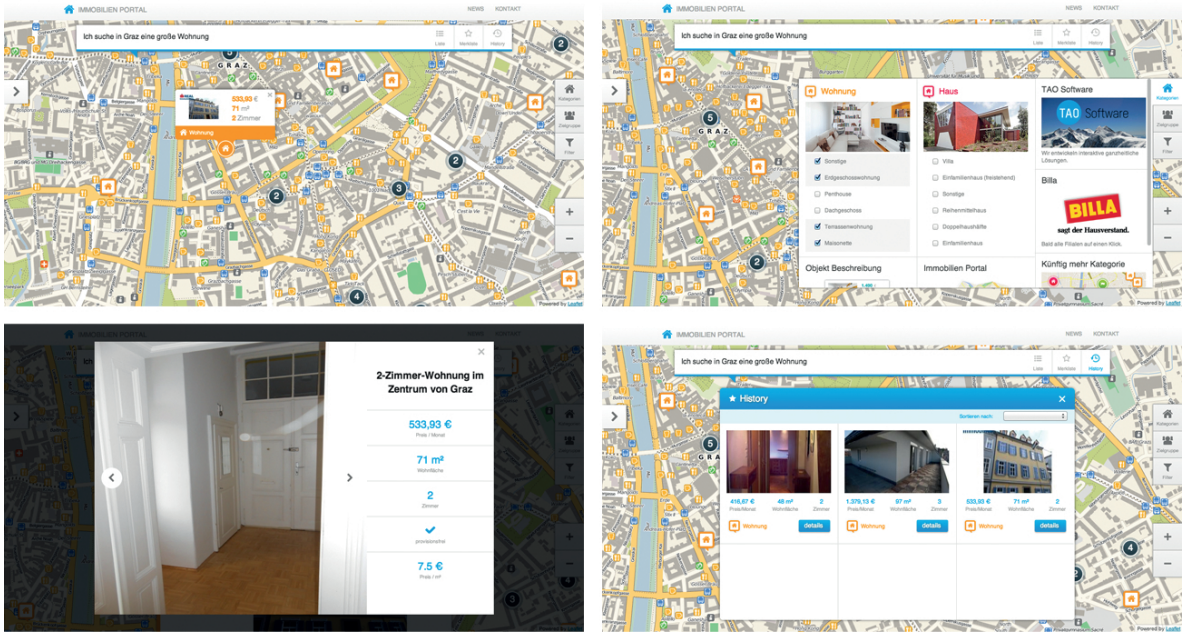

Figure 2.1: Screenshots of the website's user interface (© TAO Software): real estate object (top left), filter (top right), detail view (bottom left), history (bottom right)

Compared to other real estate websites that offer mostly searchable lists of objects this site focuses on a location-based view and map interaction. Maps are not uncommon these days and offer a lot of possibilities. Due to the map-centred design the website is quite interactive and unconventional but may be also challenging for some users.

2.3 Design of the Modelling Approach

This approach is based on Hassenzahl's model of user experience (see section 1.4) and it is characterised by requiring quite a number of participants (the bigger the sample size the better) while not engaging very deeply with each participant. The sample should resemble the target groups for which the product was designed. Participants can be invited in various ways. Here I have used some sort of snowball sampling since there was the objective not to publicly expose the prototype. The sample should be reasonably large to ensure that the power of statistical tests is not affected too much by the (possible) non-normality of the data. Sometimes 30 people are mentioned in this regard.

The (official) German version of *AttrakDiff* (Hassenzahl et al. 2003) has been used for the evaluation. The process was very similar to other studies that used the questionnaire like Hassenzahl (2008a). AttrakDiff is a questionnaire with twenty-one seven-point Likert scaled bipolar items with verbal anchors (e.g. confusing – clear, good – bad, ugly – beautiful etc.) (Hassenzahl et al. 2003) and has been made freely available by its authors. It consists of three subscales: perceived pragmatic quality (PQ), perceived hedonic quality-stimulation (HQ-S), perceived hedonic quality-identification (HQ-I) and contains also items about attractiveness (ATT). For statistical purposes the following additional data were collected: age, sex, occupation and highest completed education level. The questionnaire has been implemented in a proprietary survey system developed by TAO Software (Figure 2.2), which enabled participants to fill out the questionnaire by using their web browser. It was secured with an individual access token that became invalid after the questionnaire had been completed.

Participants were invited to use the website once and evaluate its handling and appearance. Access to the website was again secured with an individual access token. The participants evaluated the website unsupervised and context has not been controlled. They viewed the website on their own different devices wherever and whenever they wanted. There were no tasks to complete. Participants have been instructed to discover the site freely and to their liking for 15 minutes or so. They were advised to use all functionality of the site they needed and to stop if they did not know what to do anymore. The goal was to create a scenario as diverse and realistic as possible. Problems that could arise from different configurations of the devices were deliberately accepted, because these cannot be detected and controlled in a realistic usage scenario either. Nevertheless they are part of the evaluation. After their use of the website participants were asked to fill out the questionnaire. They should just select the first thing that came to mind and should not contemplate a lot about their choices. After

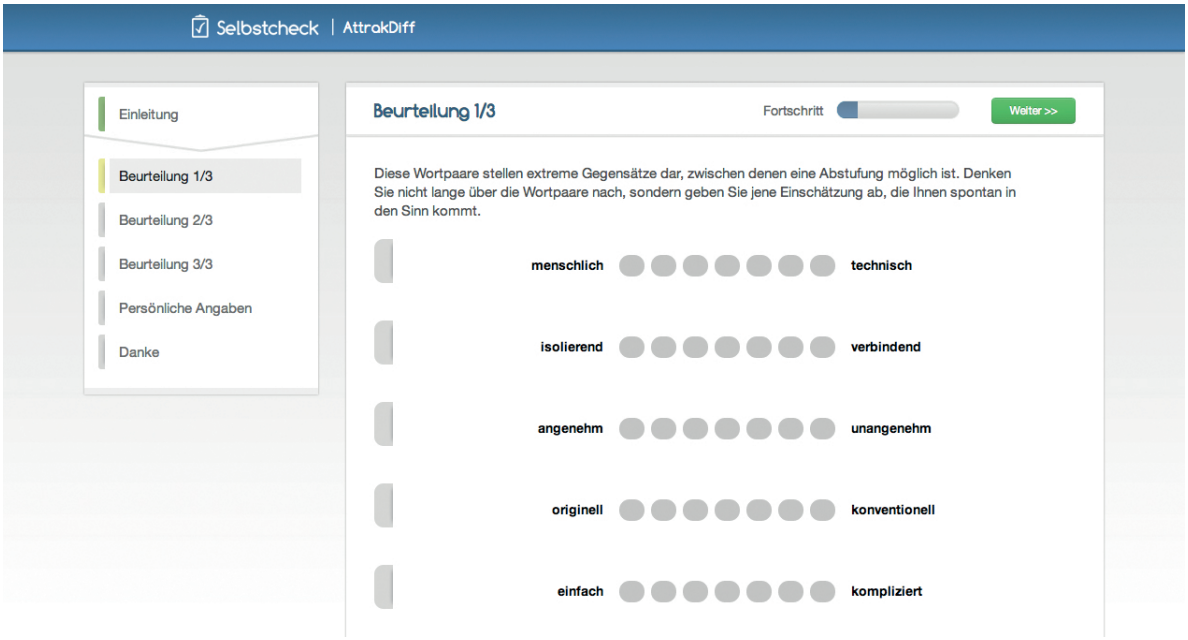

Figure 2.2: Proprietary on-line survey system where AttrakDiff was implemented. (© TAO Software)

successful completion of the questionnaire each participant was eligible to win one of three vouchers with total value of 50 euro.

2.4 Design of the In-Depth Approach

McCarthy & Wright provide the underlying philosophy of thought of this approach (see section 1.5). The central element is dialogue with users. With only a few participants there is in-depth interaction. As specific methodology for McCarthy & Wright's framework I have chosen a modified form of Descriptive Experience Sampling (DES; see section 1.7). Unlike in the modelling approach the participants should not resemble target groups but represent *maximal variation* to get a diverse range of data (Flick 2008, pp. 27 sq.). For the real estate website this could include the following personas:

- frequently moved or searched for a flat
- hardly moved or searched for a flat
- recently moved or searched for a flat
- web or computer expert
- web or computer novice
- works or has worked in the real estate industry

There are only few participants because DES is quite labour-intensive. Also qualitative studies are case studies and do not necessarily require lots of participants if there is the desired amount of variation. Participants were invited to an individual introductory discussion where the method (see section 1.7.3) and the software (see section 2.5) were explained in detail. I have

adopted the same co-researcher stance as Hurlburt and Heavey do. Participants were informed that results will be published, but their anonymity and privacy will be respected (ibid., p. 75) by changing names or other sensitive information. They were free to quit anytime, but if they completed the whole procedure they would be eligible to win one of three vouchers with total value of 50 euro. All data that they have collected with the app should enable them to relive the experience at the interview. However, I told them that I would like to do some statistics with average beep times and that I would produce a write-up or summary for each of their experiences and would discuss these with them if they would like to do so.

DES needs some initial training for both interviewer and interviewee. Both have to know what each person is talking about, what is asked and so on. Hurlburt and Heavey recommend to view the first sampling day as training and to discard all data of this day (Hurlburt and Heavey 2006, p. 15). I used the first two sampling days as training and let the participants sample any activities during their day they wanted. In this setting the beeps came on average each thirty minutes with maximum intervals of one hour. Participants should do 4-6 beeps, which would take them 2-3 hours. They could take a photo for each sample, enter a title, a description and answer open questions regarding their mood and presence. Within 24 hours the interview was scheduled which lasted between 30 minutes to an hour and took place in the conference room of TAO Software.

In the next phase participants should use the website two times, each usage being on different days. This usage has been unsupervised and the participants' own different devices have been used to their liking. There were no tasks to complete. Access to the website was secured with an individual access token. In this setting the beeps came on average each five minutes with maximum intervals of ten minutes. Participants should collect three beeps or more but they could stop anytime if they were bored. Within 24 hours again an interview was scheduled which lasted between 30 minutes to an hour. Participants have been asked some reflective questions after discussing the samples. These are based on the sense making threads proposed by McCarthy & Wright (see section 1.5.2) and were originally asked in German. The (translated) questions are:

1. What were your expectations?
2. Where they met? Why? Why not?
3. How will it - in your opinion - continue?
4. Are you keen on continuing? Why? Why not?
5. Name the most prominent quality the product currently has for you.

Question 1 refers to the *anticipating* thread and to our expectations and possibilities. Question 2, 3 and 4 refer to the threads *interpreting* and *reflecting* and to what has happened, what is likely to happen and if it is fulfilling (anxiety or boredom). Question 5 refers to the threads *connecting* and *appropriating*, tries to uncover the pre-conceptual sense of a situation and relates it especially to our Self. There is another thread that is covered by all questions and all interviews: *recounting* or telling others.

2.5 BeepMe: A Smart Phone App

Descriptive Experience Sampling or DES (see section 1.7) that is used in the in-depth approach (see section 2.4) needs a device that delivers random beeps and ideally also lets people record their experiences or take notes. Hurlburt and Heavey (2006) use a simple *beeper* that just can be turned on and off, emits random beeps and has an earpiece. As small computing devices become more and more popular and smart phones are carried along during the whole day a beeper software could conveniently run on these devices. Since the device is normally used a lot by the owners it is also familiar to them. The acceptance should be higher as with an additional device they would have to carry with them.

Others have also taken this route, but software I have found before was mostly outdated and seemingly abandoned. These were running on PDA or Handheld PC devices that were popular before the smart phone boom. Other ones did appear after I had started my own development or were in a closed testing phase. Here is an incomplete list of projects I have found:

- Emotion Sense `http://emotionsense.org`
- DEScribe `http://code.google.com/p/describe`
- MyExperience `http://myexperience.sourceforge.net`
- MyServiceFellow `http://www.myservicefellow.com`

Included in this list are also applications that would maybe be categorised as mobile ethnography apps rather than experience sampling apps. I think there is a gradual transition. Because no usable software was available I had to develop my own app (Glanznig 2012a). But which platform to choose? Particularly Android (by Google) and iOS (by Apple) devices (e.g. iPhone or Nexus handsets) are extremely popular today. Since Android has a higher market share than iOS according to IDC (2012) and others, I own a Android handset for testing, it is cheaper to publish apps in the Google Play store and easier to distribute apps without a store I have chosen Android as platform for the beeper software. Unfortunately, time and budget restrictions did not allow to develop a multi-platform software.

2.5.1 Characteristics of the Beep

Hurlburt and Heavey provide desirable characteristics for the signal, which they call the beep. It should be *unambiguous* and basically mean "Pay attention to your awareness now!" (Hurlburt and Heavey 2006, p. 73), so it is not advisable to use some sort of ring tone. It should be *easily detectable*, but not being too soft or too loud. "If it's too loud, it will startle you, and you'll likely forget what was in your inner experience. If it's too soft, you'll find yourself asking yourself, *Is that the beep?*" (ibid., p. 83) and that would be the content of the person's inner experience. The signal should have a *rapid onset* because the moment under investigation should be clearly defined.

The beep should be *private* as the persons should be able to pay attention to their inner experience as undisturbed as possible (ibid., p. 73). Hurlburt and Heavey use an earphone for that. I think the earphone does not change much regarding privacy. Persons have to wear it visibly and might have to give explanations about why they are doing so. Additionally, they cannot hide their note-taking after a particular beep. Therefore, I have chosen not to use an earphone as it just means additional hassle for the persons. There are two more desirable characteristics. The signal should be able to follow the persons wherever their experience takes place, it should be *portable*. It should *require a response*, because this forces the persons to take some action in response to the beep and makes it harder to ignore it (ibid., p. 73).

2.5.2 App Features

Mainly the app has to deliver random beeps to the user. This involves tasks such as starting and pausing the timer, scheduling beeps in the required intervals, adapt to different phone states like silent mode and deliver the signal itself, maybe pausing already playing other audio first and not interfering with phone calls. A simple task like *deliver random beeps* can get quite complex to implement if you think about it for some time. Additionally, the users should be able to take notes about their experience, take photos, list and look at their samples and edit them later on until twenty-four hours after the beep. Figure 2.3 shows the user interface of different app screens.

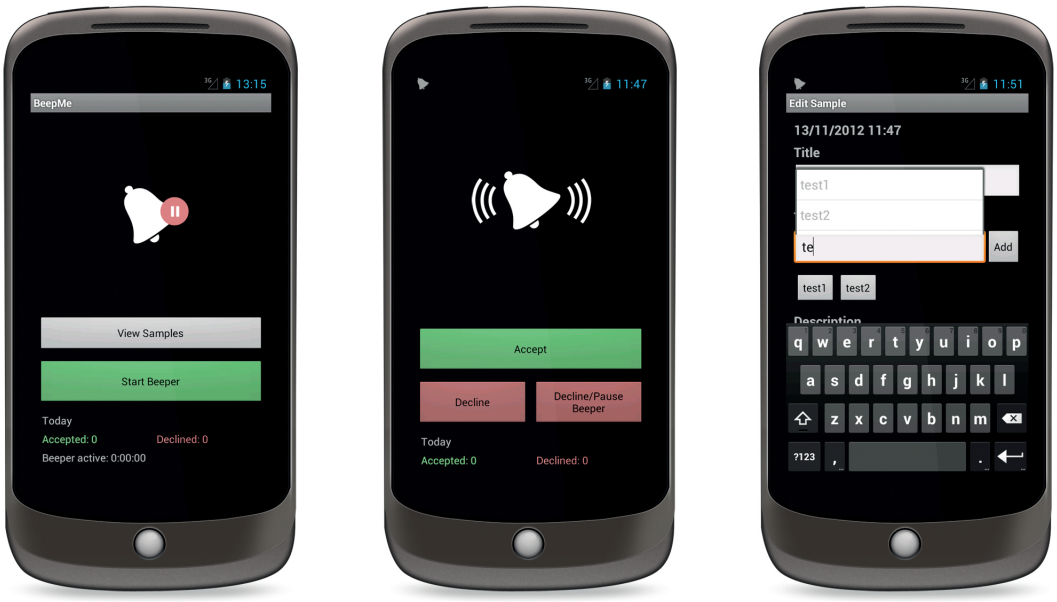

Figure 2.3: BeepMe user interface. Home screen (left), the beep (middle), edit a sample (right).

A beep can be accepted or declined, this is similar to just turning Hurlburt & Heavey's beeper off without taking any notes. However, after too many subsequent declined beeps the intervals become shorter if the timer stays active until a beep was accepted. The average length of a

interval is specified by a *beeper profile* of which several can exist. As noted in section 1.7.4 the average interval has to be shorter when sampling a HCI context. The software can now adapt to different contexts by changing the beeper profile.

Taking a photo after a beep can help to remember inner experience at that moment, because the photo captures the surroundings as they where at that time. Of course, this does not tell us much about inner experience but it removes ambiguity from outer circumstances and can be a helpful aide. On the other hand, it may overlay the mental image (if there is any). Therefore, there is the risk that any differences between the mental image and the real picture are lost when looking at it again.

Because typing on a touch screen can be tedious users may not want to enter much text for each beep. This problem can be circumvented by letting the user enter keywords and let the system autocomplete them or make suggestions for already entered ones. If quite similar keywords for all users are expected (e.g. when asking for people's mood) a predefined but extendable vocabulary could be used. However, this should be used with care since it may lead to reduced diversity in alternatives or missed details.

2.6 Testing Designs With a Small Pilot-Study

As already noted I have used both methods for the first time. Generally, it is a good idea to run a small pilot-study to be prepared for most eventualities and in addition doing DES explicitation interviews requires a good amount of training. Hurlburt and Heavey make it quite clear that you should not assume that "you're good at DES" (Hurlburt and Heavey 2006, p. 129) and you should not assume that "you can learn DES faster than you can learn to play the violin" (ibid., p. 129). So I started to practise doing interviews and went through the whole sampling procedure with two friends by using a conceptual similar website (`http://bikemap.net`). Additionally, I distributed an online version of the AttrakDiff questionnaire to 15 friends and asked them to use and rate the same website. Interestingly, some people were not satisfied with the questionnaire's possibilities to state their opinion. One friend even called me to tell me some things she discovered but could not voice in the questionnaire. The pilot-study did not reveal any show-stoppers in the study design.

3 Results

Since we now know *how* the evaluation has been done it is time to look at *what* the outcome of each approach was. The results are listed and are discussed for each method separately.

3.1 Modelling Approach

Analysing the data of the modelling approach is done through statistical analysis that is based mostly on comparing subscale and item means and their confidence intervals. The effect of age on subscale means is investigated with MANOVA.

3.1.1 Structure of the Sample

The sample should resemble the target groups of the evaluated product. Requirements in this regard stayed vague, young people and students were named as possible target groups. The selection of participants was complicated by the objective not to publicly expose the prototype. Finally 43 people (15 females) were invited to the evaluation in a snowball sampling-like approach with the help of TAO Software's employees. 36 people filled in the questionnaire, which corresponds to a response rate of 84%. One result has been discarded for analysis because the participant stated in his response that he did not search for flats in the target city and therefore did not see any real estate objects.

The remaining 35 participants (13 or 37% females) are available for analysis. Average age in the sample is 33.9 years. The youngest participant is 18 years old and the oldest participant being 62 years old. Some participants are students (11 people), other occupations are for example chef, teacher, information technology engineer and salesperson. Most participants specified university (14 people) as highest completed education level. It was followed by high school [Matura] (11 people) and college [Akademie] (5 people).

If we look at the age structure of the sample (Figure 3.1) we see two age groups and a gap between 40 and 50 years. This can be explained by the snowball sampling. One group represents the employees of TAO Software and their friends, the other group represents the senior generation (e.g. parents).

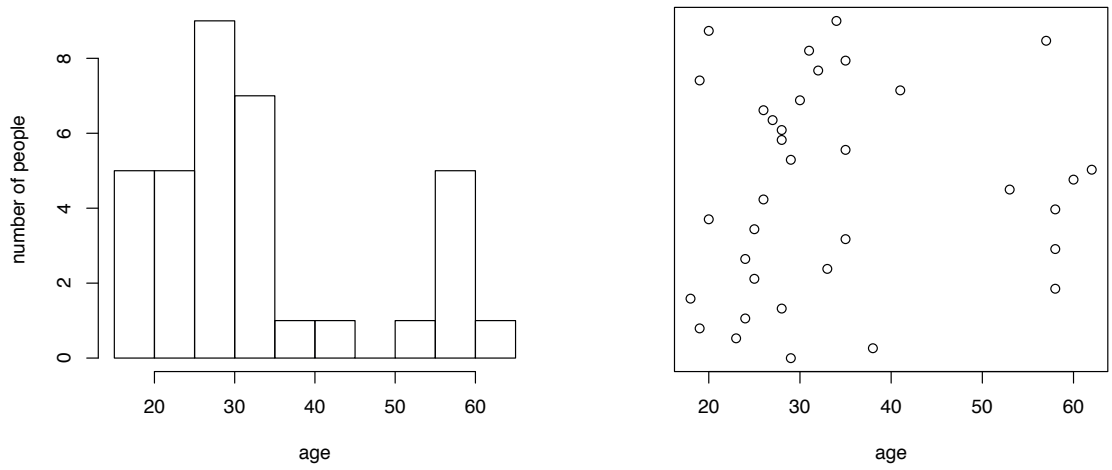

Figure 3.1: Age structure of the sample as histogram (left) and scatter plot (right)

3.1.2 Results

Hassenzahl's model of user experience has two quality dimensions - pragmatic quality (PQ) and hedonic quality (HQ) - that are connected to the product (PQ) and to the Self (HQ). The left part of Figure 3.2 shows the mean values of these dimensions, which have been computed from grouped questionnaire items. In the right part the subscales PQ, hedonic quality-identification (HQ-I), hedonic quality-stimulation (HQ-S) and attractivity (ATT) are shown. Figure 3.3 displays mean values of the questionnaire's items with positive alternatives being located on the right side. The labels on the far left side tell to which subscale an item belongs (see section 1.4.2). The lowest ratings received *undemanding - challenging* (HQ-S), *technical - human* (PQ) and *isolating - connective* (HQ-I). The highest ratings were obtained for *unpresentable - presentable* (HQ-I), *unimaginative - creative* (HQ-S) and *conventional - inventive* (HQ-S).

The error bars show the 95% confidence intervals of the mean values. The real population's mean falls within this interval with a probability of (approximately[1]) 95%. The underlying data do not follow a normal distribution and also the distributions of the mean values have negative skew. See Table 3.1 for results of the Shapiro-Wilk normality test. Nevertheless, the distribution of sample means (the sampling error) approaches normality at a certain sample size (central limit theorem). A sample size of 35 should be sufficient, therefore the confidence intervals are depicted. Also ANOVA (analysis of variance) is quite robust against non-normality, but small deviations are possible (Box 1953).

The smallest confidence intervals (highest confidence) have *dull - captivating* (HQ-S), *unimaginative - creative* (HQ-S) and *repelling - appealing* (ATT). Largest confidence intervals (lowest

[1]Deviations with magnitude below five percent are possible due to the rather small sample size and the non-normality of the data.

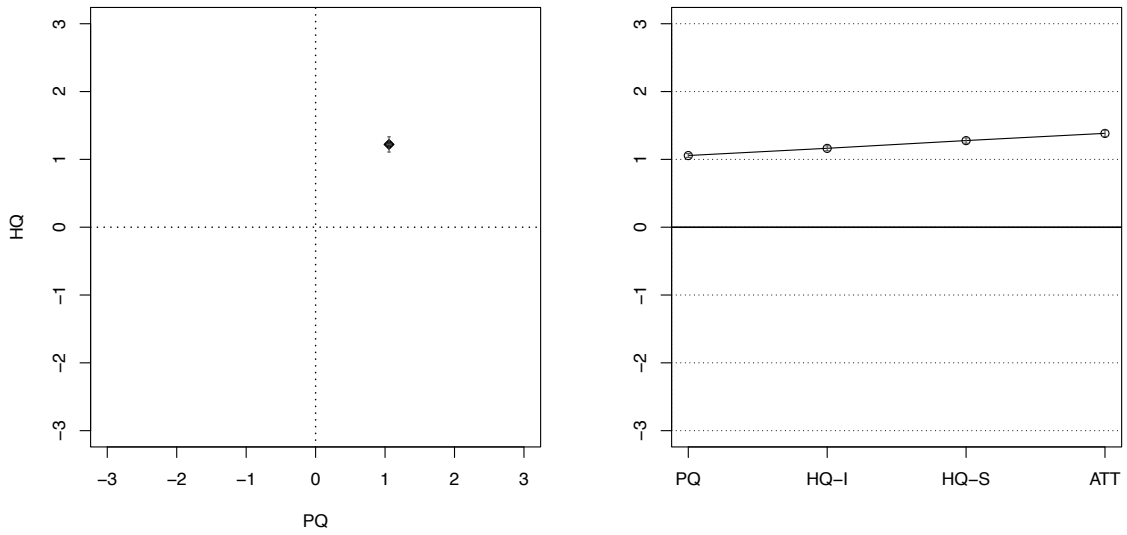

Figure 3.2: Mean values of pragmatic quality (PQ) and hedonic quality (HQ) on the left side, mean values of subscales PQ, hedonic quality-identification (HQ-I), hedonic quality-stimulation (HQ-S) and attractivity (ATT) on the right side, each with 95% confidence intervals (N=35)

	W	p
PQ	0.9227	0.01701
HQ-I	0.9146	0.01001
HQ-S *	0.9624	0.27
ATT	0.9143	0.009771

Table 3.1: Shapiro-Wilk normality test for subscale means with W_{sig}=0.1883 and p_{sig}=0.05 at N=35, * normal distributed according to test

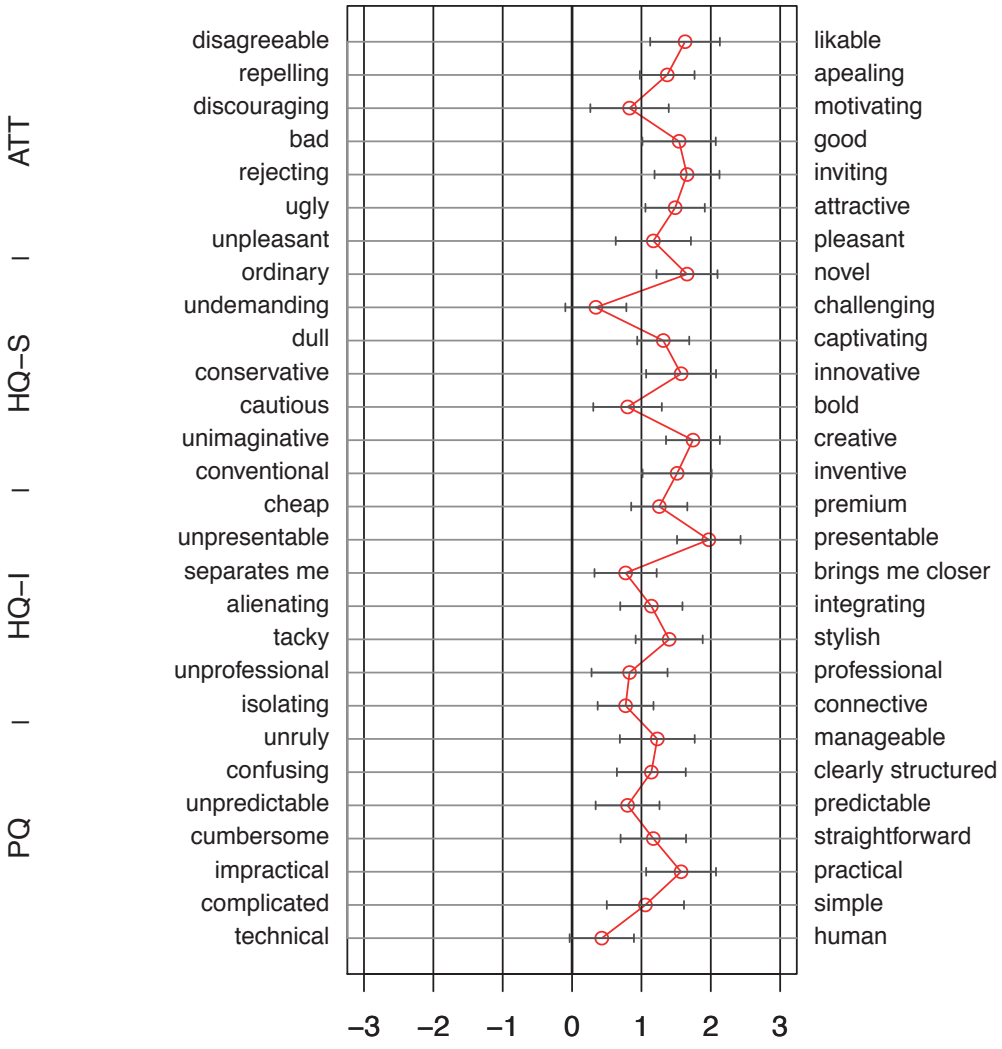

Figure 3.3: Mean values of AttrakDiff items with 95% confidence intervals (N=35)

confidence) can be found for *discouraging - motivating* (ATT), *complicated - simple* (PQ) and *unprofessional - professional* (HQ-I).

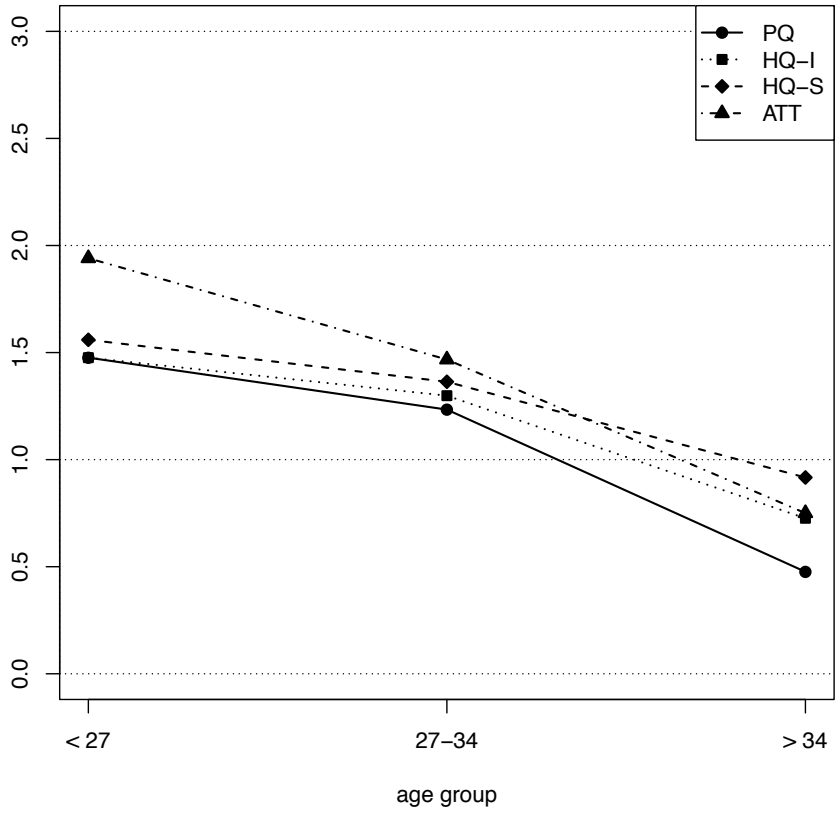

Figure 3.4: Subscale means of pragmatic quality (PQ), hedonic quality-identification (HQ-I), hedonic quality-stimulation (HQ-S) and attractiveness (ATT) for three age groups of the sample

	Df	Pillai	approx F	num Df	den Df	Pr(>F)
age	2	0.20977	0.87883	8	60	0.5395

Table 3.2: Results of MANOVA with dependent variables PQ, HQ-I, HQ-S, ATT and independent variable age

Subscale means of three different age groups are depicted in Figure 3.4 (ordinate similar to Figure 3.2 right with the lower negative part removed). The group borders have been selected in a way that group sizes are approximately equal. Despite the age differences in the sample MANOVA (multivariate analysis of variance) with dependent variables PQ, HQ-I, HQ-S, ATT and the independent variable age with depicted age groups did not yield a statistically significant result (see Table 3.2).

3.1.3 Discussion

In the left side of Figure 3.2 the website is positioned in the upper right quadrant which means that it is characterized by participants as pragmatic and hedonic. Despite this good result the values of both dimensions are located near the lower third of the quadrant which indicates the possibility of improvement. When looking at subscale means (Figure 3.2 right) it is striking that pragmatic quality (PQ) or usability with negative adjectives like unruly, confusing, unpredictable, cumbersome, impractical, complicated and technical received the lowest score in comparison to other subscales. Attractivity (ATT) with positive adjectives like likable, appealing, motivating, good, inviting, attractive and pleasant received the highest score compared to the other subscales. The website's score for stimulation (HQ-S) is higher than for identification (HQ-I). Nevertheless, all means are located in the positive (upper) area, but their position also indicates room for improvement. Why usability (PQ) did receive the lowest score compared to other subscales is not clear. Maybe that is a result of bugs or interaction flaws in the prototype or of initial unfamiliarity with its usage.

All item means in Figure 3.3 are located in the positive (right) half and even the confidence intervals indicate that population means are also located in the positive area. There are only two items with lowest scores where the confidence intervals extend slightly into the negative area. Participants see the website rather in direction of *undemanding*, *technical* and *isolating* with the possibility that the means of undemanding - challenging and technical - human fall within the negative area. On the other hand, the participants rated the website also as being *presentable*, *creative* and *novel* where the highest scores have been reached. Also *likable* and *practical* received high scores. Participants agreed that the website is *somewhat captivating*, *creative* and *rather appealing*. They rather disagreed on rating the website as *not so much motivating*, *not that simple* and *somewhat professional*.

If we look at the subscale means for different age groups of the sample (Figure 3.4) we see that scores get lower with increasing age. Younger people in the sample give better ratings than elderly people. Interestingly, the site is particularly attractive for younger people, while this is not the case for elderly people. Usability (PQ) has received the lowest score compared to other subscales in all age groups and is rated particularly low compared to other subscales in the oldest age group. The hypothesis that age influences the rating in the overall population yielded no statistically significant result and could not be verified. It could be possible that the age differences manifest only in this particular sample. Unfortunately, the age structure of the sample could be better (Figure 3.1). There is a gap between 40 and 50 years and there are only few participants older than fifty. This could be the reason for the insignificant result.

3.1.4 Action Possibilities

The website received quite good ratings on both dimensions pragmatic quality (usability) and hedonic quality (stimulation, identification). The rating for attractivity is particularly high. However, usability is rated lowest compared to the other subscales. The reason for this rating

could not be investigated. Maybe the cause are bugs or interaction flaws in the prototype or initial unfamiliarity with its usage. A small usability evaluation with few participants or a longitudinal UX evaluation may be able to provide further insights.

It was not possible to statistically verify the existence of degradation of ratings with increasing age in the overall population. This would require an improved age structure of the sample and an increased sample size. The possibility of degradation of ratings with increasing age and hence also the possibility of increased problems with the website with increasing age should be considered when choosing target groups and enhancing the website.

3.2 In-Depth Approach

For the in-depth approach the interviews have been recorded and transcribed and summaries of samples have been produced. These summaries have been coded and the codes have been grouped into themes and have been placed into a hierarchy according to frequency of samples and number of participants where the theme manifested.

3.2.1 Participants

The group of possible participants was rather small as the method is quite time consuming (see section 2.4) and due to budget restrictions only a voucher as prospective prize could be offered to participants. A prerequisite was also the ownership of an Android smart phone because the beeper software runs on that system. It was not possible to provide the participants with a device but this would not have been a good idea anyway since they would have had to learn how to operate it first. Because of the smart phone requirement elderly people unfortunately were nearly excluded from participation. Participants were recruited largely from my social environment with active support of the manager of TAO Software. Nevertheless, it was taken care to fulfill the criteria for maximum variation that have been specified in section 2.4 (opportunity sampling).

In the end 6 people (2 females) in the age group of late twenties to early thirties participated in the evaluation. 3 people were students or had an university degree, others had completed high school [Matura]. The participants[2] can be characterised as follows:

- **Andreas**: Has moved now and then between different cities and knows what to look for. He is a normal user, but fancies technology.
- **Barbara**: Has barely moved but worked for a real estate agency for a short time. She does not have problems with technology as long as it works.
- **Markus**: Has just completed the search for a new flat and has intimate knowledge of all the real estate search engines around. He is a normal user, but fancies technology.

[2]To protect the participants' privacy all names have been changed.

- **Susi**: Has just renovated her house but has barely moved and does not really fancy technology.
- **Thomas**: As an engineer he engages in software development. He has moved a few times.
- **Werner**: He really fancies technology and dedicates himself to the newest trends and developments. He has moved a few times.

3.2.2 Analysis

All interviews have been recorded. The beeper software recorded timestamps of beeps and intervals between beeps. Analysis was done through qualitative text analysis to produce a Grounded Theory (cf. Glaser and Strauss 1967). This process requires coding (assigning labels to parts of the interview), grouping of similar codes or recurrent occurrences of codes to concepts, summarising groups of similar concepts into categories and generate a collection of explanations for the data (a theory).

The recorded interviews have been transcribed (see Appendix for an example). For each sample a write-up (summary) was produced, which could have been shown to the participant for verification. However, no participant wanted to do this.

Each write-up was divided into connected parts or themes, which were associated to different data categories. These are:

- **ACT**: Activity, describes actions by the participants
- **EXP**: Experience, describes inner experience
- **FEEL**: Mood (feeling), describes emotions or moods
- **PRES**: Presence, describes degrees of focus
- **REFL**: Reflection, describes (judgemental) thoughts which are based on self-theories or theories about the external world
- **SUGG**: Suggestion, describes suggestions for improvement or wishes for the website

The above categories are not to be seen as strict divisions. Especially EXP and REFL could show significant overlap and strictly speaking SUGG is a subset of REFL. The categories are meant to be a rough classification and differentiation between several types of themes. This classification emerged after conducting several interviews as it describes recurring patterns.

The themes were consolidated, grouped and were subsequently moved into a hierarchy. Ideally, the assigning of data categories, grouping and consolidation should have been done at least partly by at least two different persons. This concept is known as inter-rater reliability (cf. Stemler 2001). Unfortunately, time and budget restrictions prevented having more than one person doing the analysis.

3.2.3 Results

All data result from the second phase of sampling where the participants used the website. In total 12 interviews have been conducted which lasted on average for 32 minutes. We talked about 22 beeps which came in average intervals of 6 minutes and 45 seconds. The participants used the website on average for 23 minutes. For each sample they answered an open question regarding their presence (such as present, concentrated, day dreaming etc.). The consolidated answers are listed in Table 3.3.

Presence	No. of Samples
average	14
very present	11
concentrated	9
less present	5
not present	2

Table 3.3: Presence of participants. The answer *present* was rated as average and concentrated is regarded as more present than very present.

Similarly, participants answered an open question regarding their mood. Multiple answers were possible. Moods are depicted as word cloud in Figure 3.5 where word frequency corresponds to font size. Word clouds have been generated with Wordle (`wordle.net`). The mood *good*, which originally occurred 17 times, was regarded as default answer and therefore discarded.

Figure 3.5: Moods of participants. Word frequency is proportional to font size and similar moods are coloured similarly.

Similar moods were grouped into categories. These are not to be considered definite and are listed in Table 3.4. Moods in Figure 3.5 have been coloured according to these categories.

Category	Moods	No.
positive	interested, curious, amused, keen, relaxed, enjoyable, pleasant	24
slightly negative	confused, astonished, irritated, impatient	20
neutral to negative	uninterested, bored, tired, surprised	12
very negative	angry, disappointed, frustrated, demotivated	10
very positive	cheerful, happy, pleased, super	7

Table 3.4: Moods grouped into (non-definite) categories

The frequency of words in summaries of all samples might reveal different themes (Figure 3.6). Similar words have been grouped and trivial words, participant names and data categories have been excluded. Prevalent words that were to be expected such as *looking*, *reading*, *flat*, *website*, *clicking* and *searching* have also been excluded (in total nine words).

Figure 3.6: Content of all samples as word cloud. Word frequency is proportional to font size. Similar words have been grouped and trivial words, participant names, data categories and expected prevalent words have been excluded.

Themes

Table 3.5 gives an overview over all themes that have been found during the analysis. The table also includes the number of samples and persons where the theme manifested, which might help to classify the importance of the theme. For a detailed list of all sub-themes refer to the Appendix. Below I give a short description of each theme sorted by frequency and beginning with the most frequent one.

Name	No. Samples	No. Persons
steep learning curve	19	6
get an overview, look around, "rambling"	14	5
site and content are a unity	11	6
photos are important	11	4
issues with cut-off detail view	9	3
website offers freedom of search	8	4
sensibility of loading time, reaction time, visual feedback	7	6
description text's presentation is subpar	7	5
photos and groundplan enable envisioning of flat	6	4
advertisement stands out and polarises	5	5
high expectations towards the site	5	3
groundplan is important	5	3
description text is unreliable	3	3
starring feature is hidden	3	3
reset feature is hidden	3	3

Table 3.5: Overview over all themes and number of samples and persons where the theme manifested

steep learning curve The impression of a steep learning curve emerged in 19 samples of all 6 people and can be summarized with Susi's statement that one has to try to understand the site. The theme consists of two sub-themes: *initial confusion, being overwhelmed.* In 16 samples of all 6 people - coming almost exclusively from the first day of usage - appeared statements like: You have to orient yourself, navigation is complicated, filters are hidden, a menu was overlooked and the site itself being somewhat confusing. The role of markers and cluster markers were not clear and the site behaved oddly after careless clicks. Three people were searching for a key to the map or they wanted to have one. Statements like you're fine if you know where everything is (Susi) and that at the beginning everything is displayed instantly and then you have to select (Markus) complete the picture. Nevertheless, most participants were able to overcome those initial difficulties as good ratings at the end suggest. *trying-out necessary.* In 8 samples 5 people talk about trying out a feature or just clicking ahead (Susi). Also Werner's statement of having to make a move fits here.

get an overview, look around, "rambling" This theme appeared in 14 samples of 5 people. The map enables you to get an overview (Barbara, Werner, Markus), you can look around (Werner, Susi, Barbara) and find out what's available in the vicinity (Werner). Cluster markers reduce complexity (Andreas) and display available flats in a region (Barbara, Markus). Using the map feels like browsing around or flying over the city (Barbara). The possibility that this theme only appeared because none of the participants really searched for a flat cannot be ruled

out. This fact could have led to a more relaxed atmosphere.

site and content are a unity In 7 samples of 5 people it became clear that participants did not differentiate between the website as infrastructure and data or content which are provided by estate agents or landlords. Both components are seen as unity and data quality is part of the website's rating. Bad data quality gets in the way of successfully searching for real estate and therefore the rating suffers from it. In this context participants wanted to see the exact address of a flat (Andreas, Werner). Ongoing updates raise the interest for the site and good content is important for the website (Markus). Susi reported that a flat did not have a balcony despite it was selected as criterion by her. After her filtering hardly any choices were left (Susi). There should be more objects in the database (Thomas). Particularly critically evaluated were photos: *Photos are distorted or poorly made.* Werner, Susi and Barbara criticised that photos were distorted or pixelated. Susi said that envisioning a room is hampered by poorly made photos. Barbara wanted to have better photos and Markus and Susi noticed odd or useless photos where for example the house entrance was depicted.

photos are important There was a lot of interaction with photos by 4 people in 11 samples (see also *photos and groundplan enable envisioning of flat*). Werner noticed that photos stand out. Markus thinks that photos are important. Barbara recognized a photo of a previously seen flat while browsing through the list and Susi was able to look at the uncut photos by clicking on them and using the lightbox.

issues with cut-off detail view The fact that the overlay of the detail view was not completely visible on the screen after opening it (see Figure 3.7) caused problems for 3 people (Werner, Susi, Barbara) in 9 samples. The other three participants did not report the problem. Either because they did not see it or they solved it automatically by moving the map. Barbara, Susi and Werner tried to move the overlay by using their mouse wheel as they would normally scroll a website. On the map the mouse wheel is associated with the zoom function. This had the unpleasant effect of zooming the map, closing the overlay and resulting in participants being completely lost which caused quite some amount of frustration. Barbara and Werner did later on find out how it worked but Werner reported that it still got into his way sometimes. Susi did not succeed in moving the overlay, she then discovered the lightbox which at least enabled her to look at the uncut photos. She also reported that she wanted to leave the site after the problem had appeared for the first time. She did not do it because she wanted to complete the evaluation.

website offers freedom of search Four people in 8 samples were quite fond of the different possibilities the site offers them for searching. Different parameters can be specified (Andreas) and there are different possibilities for filtering and sorting (Andreas, Werner, Barbara). You need only one tool to reach your goal (Andreas), you are able to approach the site in your own

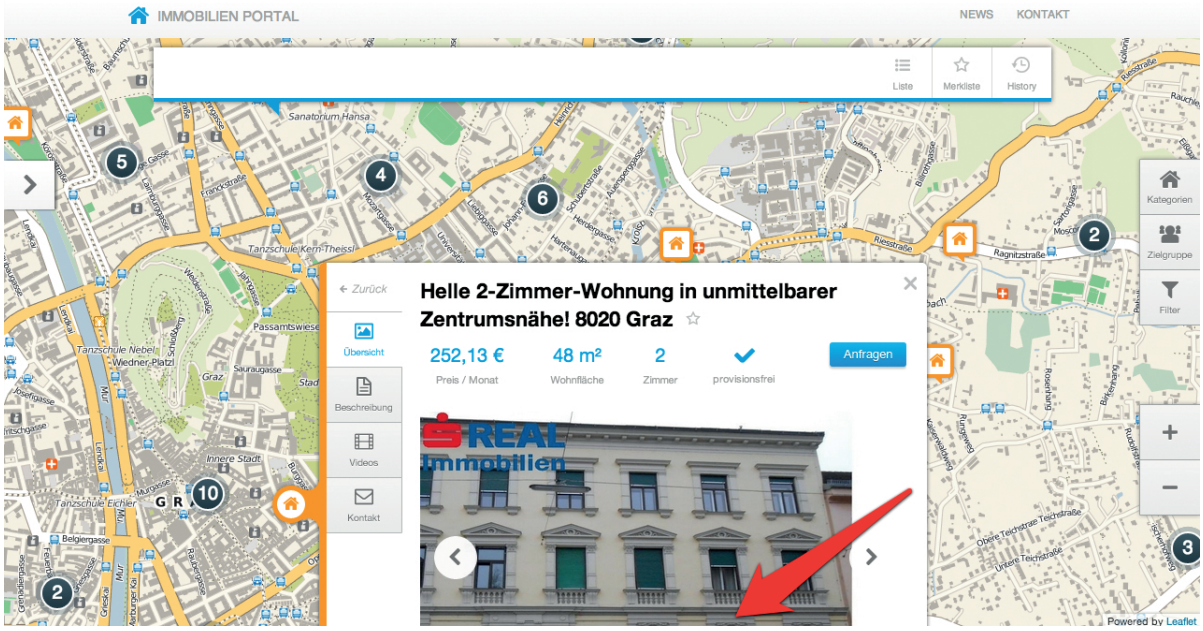

Figure 3.7: the overlay of the detail view is cut-off at the bottom after opening it (© TAO Software)

way (Andreas, Werner), everything is displayed instantly and you can select afterwards and the site has everything you need for real estate search (Markus). Barbara remarked that you gain a good overview with the map but the list is better suited for comparing.

sensibility of loading time, reaction time, visual feedback All participants showed a certain sensibility concerning delays in 7 samples. The reason for that might be found in the otherwise high interactivity and dynamics of the site (see ratings below). Little problems with loading time were prevalent (Andreas, Werner, Thomas, Markus) which could have appeared due to problems with Internet connections. Thomas, Markus and Barbara reported delays when clicking on checkboxes at the right menu which caused them to click on the checkboxes several times in a row. Markus and Susi showed some uncertainty about the behaviour of the map after they had used a filter.

description text's presentation is subpar 5 participants (Werner, Thomas, Susi, Barbara, Markus) were not really satisfied with the description text's appearance. In three samples they said that they prefer looking at the photos to reading the description. They would read it only when highly interested (see also *description text is unreliable*). Werner, Thomas and Susi noticed that the content of the description repeated itself. Werner suggested a better structure of the text and Barbara said that the text was tedious to read as it had small font size with narrow spaces.

photos and groundplan enable envisioning of flat In 6 samples of 4 people photos or the groundplan enabled participants to envision the flat in their head. Often this was quite vividly

experienced. Andreas was exploring different situations in a room. Markus saw the anteroom and was looking into the other rooms. When looking at another flat he had kind of a film strip of photos of a flat he inspected nearby in his head, one photo per room. Barbara sunk from the bird's eye view of the groundplan towards the flat and entered the anteroom through the door (see Appendix for a detailed description).

advertisement stands out and polarises The logo of Billa [Austrian grocery chain] was standing out for 5 people in 5 samples. There were different reactions but the logo was largely perceived as advertisement and that was seen in a negative way. Werner quickly identified it as advertisement and ignored it. Thomas was wondering what relationship Billa and the target groups could have. Andreas deduced from it that grocery chains and others were already marked on the map. Barbara asked herself: "What about the other chains?" Markus had the feeling that yet another site is plastered with advertisements (see also *high expectations towards the site*) and envisioned the map containing a lot of giant yellow Billa bags.

high expectations towards the site Three people showed high or increased expectations towards the site. Andreas said in 3 samples that the site has increased his expectations. Werner concluded that there are quite some problems to solve on the site and Markus had the feeling that yet another site is plastered with advertisements when he saw the Billa logo.

groundplan is important The groundplan plays an important role for 3 people in 5 samples (see *photos and groundplan enable envisioning of flat*). Andreas wanted to have a groundplan for every flat if possible in a distinct area and zoomable. Andreas, Markus, Susi and Barbara said that they have looked at a plan. Barbara and Andreas remarked that the groundplan supports their freedom to inform them thoroughly.

description text is unreliable Three people remarked in 3 samples that they do not really trust the description text as this text tends to be biased (Andreas), is written by estate agents who try to cover up flaws (Werner), is untrustworthy (Barbara). The participants relied more on other sources like map, photos and groundplan.

starring feature is hidden Flats can be starred to add them to the list of favourites. To do this one has to click on a small star below the title in the detail view (see Figure 3.8). Some participants did intentionally want to try out this feature but they could not figure out how to add the flat or house to the list. This problem appeared in 3 samples and with 3 people (Susi, Andreas, Thomas). Susi discovered the star but remarked that it was quite hidden. Andreas did observe it somehow but thought it to be a rating star. He tried to drag the flat towards the list or thought there might be a context menu. Thomas searched actively for a star ("just like in Firefox"), but simply did not find it.

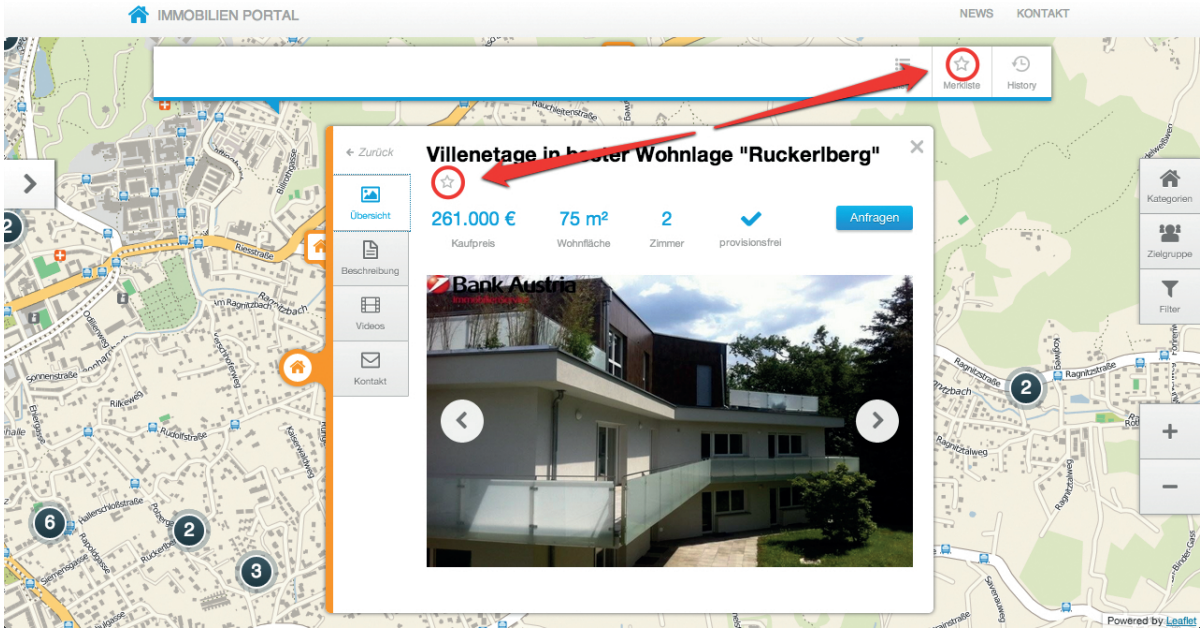

Figure 3.8: The star for adding a flat or house to the list of favourites is quite hidden (© TAO Software)

reset feature is hidden Two participants (Susi, Markus) in 3 samples wanted to make a fresh start and reset their input and filters. Both clicked on the logo of the website (top left) to achieve this, but in the beginning they were not completely sure how to do it.

Ratings of Participants

During the interviews and especially in the questions at the end the website was rated by the participants. Below there are ratings which have not been added to any themes above. Andreas compared the website to other real estate websites and said that the others are 2D and this one is 3D. Susi said the site is quite different. You can save time (Andreas), be swift (Andreas, Markus, Thomas) and efficiently look at a lot of flats (Markus). They remarked that the site looks bright and friendly (Werner), fresh (Susi) and like a comic (Barbara). There were moments where the site worked trouble-free (Andreas, Markus). It is interactive (Werner) and dynamic (Thomas), the interaction is pleasant (Thomas) and effortless (Andreas). Changes (like filtering) seem to happen in real-time (Andreas, Werner, Thomas). The map was seen as neat (Barbara) and better than a list (Markus, Barbara). On it parameters like distances are displayed objectively (Andreas). When Barbara discovered the target groups menu she thought that this was clever. Thomas said that it is a pleasant way of searching for a flat.

3.2.4 Discussion

While using the website participants were rather present (see Table 3.3). This suggests that the participants were not bored with the website but were rather challenged. Occasionally there

was boredom or day dreaming after longer usage or at the second time of use.

If we look at the participants' moods in Figure 3.5 we see that moods like *angry*, *confused*, *astonished*, *irritated*, but also *interested*, *keen*, *amused*, *curious* and *relaxed* have occurred rather often. This points towards an ambivalent relationship of the participants with the site. Obviously it elicits stronger moods. The colouring in Figure 3.5 and the rough categorisation in Table 3.4 shows that positive moods occurred most often but they are already followed by slightly negative moods.

The ratings of participants (see section 3.2.3) show that the website left a positive impression. It was characterized as significant improvement over other real estate sites, as interactive, dynamic and with pleasant and effortless interaction. So where does this ambivalent relationship come from? A look at the themes (Table 3.5 and section 3.2.3, detailed list in the Appendix) that have been found during analysis tells us that participants particularly at the beginning had problems with the site. It has a rather *steep learning curve*. This steep learning curve manifests through *initial confusion, being overwhelmed* and through the site requiring or promoting *trying-out* and *discovering* of its features. In the samples both show through searching for or suggesting a key to the map. Interaction with the site is quite effortless once it has been learned. Still, it is quite complex and the user needs knowledge from similar interactions (usage of an interactive map). If the user does not have this knowledge the learning process can get tedious. Clicking ahead and trying-out is not everybody's taste. Some people have reservations against careless or unknown clicks and the consequences they might elicit. This phenomenon is known as *self-efficacy* (cf. Compeau and Higgins 1995).

The learning phase at the beginning is embedded into a challenging situation. There, alleged minor interaction problems can lead quickly to some amount of frustration as the theme *issues with cut-off detail view* shows. There the overlay of the detail view is not centred on the screen after opening it but cut-off at the bottom (see Figure 3.7). It can be quickly moved with the map to a better position. However, for Werner and Barbara it led to problems at the beginning and annoyed them also later on. Susi was not able to solve the problem and therefore was unable to look at uncut photos. For Andreas, Markus and Thomas the problem either did not occur or they solved it immediately (with Thomas being an software engineer).

If we are interested in how the site is used and how participants searched for flats Figure 3.6 and Table 3.5 give us some clues. Figure 3.6 shows that the map is the prevalent element. Also the list appears often, photos are used or looked at more often than the description, also cost seems to be an important criterion and participants had expectations. They used the groundplan, the filter (and the map) to get an overview of the vicinity. They also looked at the categories, the target groups and the favourites (that have a mysterious connection to a star, see *starring feature is hidden*). Additionally there are the words *fast*, *interesting*, *beautiful*, *simple* and *pleasant*.

Apart from the experience of initial confusion or being overwhelmed and having to explore the site mostly real estate objects have been envisioned by using the information (pictures, groundplan, location) that was available. Sometimes this information has been connected

with knowledge about parts of the city and sometimes it has been related to past experiences of inspecting flats. This experiences have been quite vivid and included standing in the flat and looking into rooms or acting out of specific situations. Especially the groundplan and pictures sparked these experiences and they also created moods and opinions about the real estate objects. For example, the picture of a sunset scenery elicited a good mood and well done pictures or the presence of a groundplan elicited a feeling of professionalism and may have set an initial level of trust. Other experiences included remembering previously seen pictures as slide show while reading the description text and comparing flats in the head by relating the hard facts to each other.

Searchers want to get comprehensive and objective information about objects independently without much interaction with estate agents and landlords. They want to sort out for them inappropriate objects quickly and early without contacting anybody or even inspecting the object. Therefore, they rather use and trust alleged objective parameters like the map (location), cost, number of square metres, photos and groundplan. This manifests in themes like *photos are important*, *groundplan is important*, *photos and groundplan enable envisioning of flat* and *get an overview, look around, "ramble"*. They rather mistrust the description text, because language is fuzzy and could be biased (*description text is unreliable*). While I have not engaged with estate agents or landlords it is not difficult to state that these people have other interests. Landlords do not want to illustrate shortcomings that nearly every object in one or the other way has too obviously since it might lower their chances to sell or rent the object.

Both groups have opposing interests. The operator of a real estate platform has to aggregate those interests without adversely affecting one or the other group, because this would result in either not enough offers or not enough searchers on the platform. The evaluated site supports searchers in their demand for objective and independent informing. It gives them the freedom to conduct the search to their liking (*website offers freedom of search*, *get an overview, look around, "ramble"*), but this freedom requires some amount of self-responsibility which can be challenging at the beginning (*everything is displayed instantly, then having to select*). In addition to the possibilities that are provided by the operator (map, filters, interaction, favourites, search) also data (photos, groundplan, description, additional information) are part of this freedom. The operator of a real estate platform cannot easily influence the quality of the data that are provided by estate agents or landlords. However, the user is not explicitly aware of this fact when searching. The site is viewed as a whole (*site and content are a unity*).

3.2.5 Action Possibilities

Above mentioned themes result in the following action possibilities that could improve the already good user experience even further.

solve usability problems

Some minor usability problems are easily solved. The overlay of the detail view should open centred on the screen (see *issues with cut-off detail view*). The description text should be made more readable (see *description text's presentation is subpar*). The star for adding an object to the favourite list (starring) should be better visible or the interaction should be changed (e.g. dragging object onto list, see *starring feature is hidden*). The participants noticed delays when clicking on items in the right menu, which caused them to click several times. This delay should be removed.

flatten learning curve

The steep learning curve could be flattened by extending the start screen with further filtering possibilities (wizard), an initial information overlay that displays where each function is located and by making the menus on the left and right more visible (e.g. by colouring). Instead of the arrow for the left menu a more meaningful symbol should be used or it should be fused with the menu *Filter*. The possibility to reset all input by clicking on the logo (top left) was used but not immediately recognized (see *reset feature is hidden*). A better visible feature *start over* could help here.

the need for editing content should be considered

The problem of data quality for objects in the database could be reduced through the definition of quality criteria and by reviewing the first offers of new suppliers. Of course this is costly. It is also important to have a policy for objects that have already been sold or rented but are still in the database.

4 Discussion & Reflection

It is now time to bring all pieces together. As I have written in section 2.1 I am aiming at comparing results and reflecting on both methods from the stakeholder's and researcher's viewpoint. For the stakeholder evaluation a group meeting in a focus group-like setting (Flick 2008, pp. 31 sq.) including presentation of results has been conducted.

4.1 Reflection on Results

Reflection on results includes: what are the differences and commonalities, how the analysis is done and what were the pros and cons for stakeholders and author. To get an overview Table 4.2 lists all the facts, which are further explained in the text.

These facts have been obtained by having a group meeting at the office of TAO Software with participants being developers of the real estate website including management which was heavily involved in shaping the vision and in designing the product. Each approach was presented with ongoing discussion and afterwards there was a reflection about what was learned from the data and about the methods as such. The participants did express preference for a special order of presentation so I presented the methods in the same order as in this thesis: modelling approach first, followed by in-depth approach. The focus group was recorded and key arguments have been noted.

The results and kind of data are quite different for both approaches. The modelling approach yields tabular data which is rather broad and gives an overview. On the other hand, the in-depth approach yields a more or less large body of text if the interviews are transcribed or a collection of audio or video files if the recordings are coded directly. This data are rather detailed and on different levels. For example the method can capture actions, reflections about actions, moods and experiences. Because of this differences also the methods of analysis are quite different. While for the modelling approach statistical analysis is the method of choice for the in-depth approach qualitative analysis (e.g. Grounded Theory) is used.

4.1.1 Methods of Analysis

Which options are available for statistical analysis depends greatly on the type of data. For AttrakDiff all questions use Likert scales, for which there is debate if they can be seen as or-

	Modelling Approach	In-Depth Approach
type of results	broad, overview	detailed, deep
kind of data	tabular data	body of text or set of audio or video files
way of analysis	statistical analysis that may include mean, standard deviation, confidence intervals, ANOVA, MANOVA and certain types of graphs	qualitative analysis like Grounded Theory, involves coding, grouping and frequencies of occurrence, frequency charts and word clouds for visualisation
captured facts	overall good rating of product with (relatively) lowest rating for usability and (relatively) highest rating for attractiveness, (possible) degradation of ratings at increasing age, separation between usability and UX	characterization of product as dynamic, interactive with pleasant and effortless interaction, significant improvement compared to other products, steep learning curve and other themes
missed facts	*why* the rating was that way, e.g. steep learning curve	(possible) degradation of ratings at increasing age
Stakeholder's View		
pros of results	data confirm or refute opinions about the product	detailed feedback is interesting, helps with prioritising further development
cons of results	in-depth view is missing, values need to be classified into practical relevant categories	
Author's View		
pros of results	broad overview, visualization of results, condensed conclusion, terminology & concepts the same across all studies	detailed view, uncovers subtle topics, explanations extendable to similar products, concepts adapt to special circumstances, results spark discussion
cons of results	possibilities of generation of explanations for results are limited, pre-defined concepts like HQ-I and HQ-S that cannot easily adapt	difficult to get overview over results, concepts and terminology might change and decrease generalisation abilities, skill is needed

Table 4.2: Facts about results and analysis of modelling and in-depth approach.

dinal or interval scale. In the AttrakDiff questionnaire the scale items are not labelled so they should be perceived as rather equidistant. It may be plausible to see them as being on an interval scale, which allows us to compute means and standard deviations. Another prerequisite of analysis is often normality of the data, which is clearly not the case for Likert scales. However, the measurement error of the mean approaches normal distribution at a certain sample size (central limit theorem), which allows us to compute confidence intervals, ANOVA or MANOVA. The means and confidence intervals are often visualised. In the case of AttrakDiff this is done for the two dimensions of the model and for the individual word pairs, which allows easy comparison between different (versions of) products and word pairs with each other. Tendencies (improvements or degradations) are also easily visualised. In other words: the generalisability and therefore reliability is high. What remains hidden in the data are the assumptions that went into the model. There are certain methods to establish validity, which are more or less trusted depending on whom you ask.

For qualitative analysis there are different methods available, but all of them more or less depend on some form of coding, grouping and counting of frequencies. In the case of Grounded Theory there are labels assigned to parts of the data, similar or recurring codes are grouped and this groups are further grouped into categories out of which a set of explanations for the data is generated. Frequencies of words, codes and groups help to judge the importance of topics. Problems similar to statistical analysis' issues with scale types and normality are here inter-rater reliability, where different people may produce different coding or grouping. For visualisation tables sorted by frequencies or word clouds which display font size proportional to frequency are suited. Here it is more difficult to get an overview because a certain degree of immersion into the data is necessary. Improvements or degradations can only be detected at a deep level of immersion. Also comparison between products is more difficult, the reliability is lower.

4.1.2 Captured & Missed

When we now turn to actual results of the evaluation we see that each method captured the overall good rating of the product while the in-depth method being more detailed. Additionally, the in-depth method captured a set of explanations for the ratings (e.g. steep learning curve) that are based on my view from within the data after much engagement with participants. Explanations for the results are missing for the modelling approach. On the other hand, the modelling approach captured a possible degradation of ratings with increasing age that the in-depth approach missed. To a certain degree this is the case because both uses of the methods could be improved. For the in-depth approach all participants were in their late twenties to early thirties and older participants would certainly have added valuable insights to the data. On the other hand, the questionnaire of the modelling approach could have been employed twice or more often. An increase in ratings might have generated an explanation for a steep learning curve. Nevertheless it has to be noted that the possibilities for the modelling approach to generate explanations for results are rather limited.

4.1.3 Stakeholder's View

The stakeholders commented on the results after their presentation. For the modelling approach they explicitly asked if there had been a field for additional comments in the questionnaire. There was none because AttrakDiff originally also does not contain such a field (see attrakdiff.de). They also were especially interested in the relationship of rating and age. There was debate about the perceived ambiguity of questionnaire items and what is to be considered a positive or negative adjective. For them data of the questionnaire are good suited to confirm or refute opinions about the product. But they did miss "context", an "in-depth view", "what did the persons think when they filled out the questionnaire", "why did they fill it out like they did". During the presentation of the modelling approach's results there was not much discussion.

When the stakeholders were presented with the results of the in-depth approach some themes sparked an intense discussion about the implications for the product, what was already thought of and what has to be done or changed - including a discussion of the future direction of the project. Once the discussion interrupted my presentation for around five to ten minutes. When it ended they apologised and said that they had to use this chance for discussing. Reports of users inspired the developers to try to put themselves into the situation of the user and think about how it could have been for them. They found the detailed feedback to be interesting and it helped them to prioritise their further development. There were no remarks about what was missing in the results for the in-depth approach.

4.1.4 Author's View

In my view the strengths of the modelling approach lie in the broad overview the data give, the easy visualization of statistical results and the condensed conclusion of positioning the product in a two-dimensional space. This position clearly tells: Your product is that good or bad. Therefore it is particularly useful for evaluations. In addition the terminology and concepts do not change across studies which allows for easy comparison of results. For example several products or different versions of the same product can be positioned in the same space to make them comparable. On the downside the possibilities to deliver explanations of *why* results are this way are rather limited. The model uses pre-defined concepts that may not be able to adapt to different circumstances. Validity and reliability have been established when the questionnaire has been created. Reliability is mostly not questioned, although the validity of the model in general is sometimes criticised, e.g. by McCarthy & Wright.

As the name of the in-depth approach suggests it delivers a detailed and deep but specialised view on the data. It is able to uncover subtle topics and generalised explanations can extend onto similar products. Therefore it is best suited for explorations and product design. Concepts can vary and can therefore adapt to special circumstances. This is also a major weakness because this decreases the ability of generalisation and therefore reduces reliability. It is difficult to get an overview over the mostly vast body of data, which requires a certain amount

of immersion by the researcher. This immersion is strength and weakness altogether because it opens up the possibility of an expert view from within the data but also introduces inter-rater bias which means that two persons might come to different conclusions. There have been estimates of validity and reliability of DES by Hurlburt and Heavey, but in general these are harder to establish as for Hassenzahl's model. The reliability of descriptive approaches in general is sometimes criticised, e.g. by Hassenzahl.

4.2 Reflection on Methods

When looking at the methods it is interesting how costly they are, what are the pitfalls when using them, and what are opinions of stakeholders and author. Table 4.4 lists all the facts that are further explained in the text.

The differences in analysis for both approaches have already been discussed in section 4.1. Regarding preparations for conducting each approach these consist of acquiring knowledge about the approach, its parts and instruments and how to use them. There might be some training required before the method can be successfully applied and certain artefacts or documents have to be prepared.

4.2.1 Conducting the Modelling Approach

For the modelling approach knowledge about Marc Hassenzahl's model of user experience (see section 1.4), about the structure and possibilities of questionnaires and about statistical analysis is required. Knowledge about statistical analysis includes to know what instruments are available and when it is possible to use them. Using one instrument despite unmet prerequisites might lead to false conclusions. Therefore, doing statistical analysis might require some training. In my case I had knowledge about statistical analysis, but it required renewal and consolidation. That is why the amount of training was slightly higher than after repeated use of the instruments.

Before the study the questionnaire has to be created. It can be either on paper or be some sort of on-line questionnaire. In my case I had to use a questionnaire system developed by the company where it was quite cumbersome to create the questionnaire.

Recruiting participants is not that difficult as the amount of time required by them to participate is not that high (at most 30 minutes). Normally it would be possible to put up notices on bill-boards or to distribute invitations by e-mail. Here the company required some degree of confidentiality and therefore recruiting participants was rather difficult.

Cost of running the study is also rather low. One has to invite and remind participants, keep track of completed questionnaires and based on that make decisions about further invitations.

	Modelling Approach	In-Depth Approach
required knowledge	statistical analysis, Hassenzahl's model, questionnaires	qualitative analysis, dialogue, DES
amount of training	questionnaire creation, statistical analysis, not that high	explicitation interview, qualitative analysis, rather high
preparations	create questionnaire (on-line)	develop beeper app, prepare transcription sheet
recruiting participants	rather easy as participant effort is not that high, here due to restrictions rather difficult	rather difficult as participant effort is high, personal interest of participants required
cost of conduction	rather low, inviting & reminding of participants, checking progress	rather high, inviting participants, introductory meetings, conducting interviews
cost of analysis	high, depending on knowledge about statistical analysis, transformation of data and visualisation	very high, depending on skill, transcribing interviews, immersing into data, writing and visualising explanations
Stakeholder's View		
pros of method	questionnaire is easy to answer, everybody can do it	
cons of method		DES is costly, they doubted that they could have afforded DES for the project if they would have had to pay me
Author's View		
pros of method	easy to conduct, easy to distribute	in-situ measurements of DES
cons of method		DES is very costly to conduct and analyse

Table 4.4: Facts about conducting modelling and in-depth approach. For the in-depth approach facts depend on the actually used method that fits into McCarthy & Wright's framework, here this is Descriptive Experience Sampling (DES).

When it comes to analysis the cost here is also rather low depending on the amount of knowledge and training. Basically the data have to be transformed and visualisations have to be generated. When doing this for the first time or again after some time efforts are slightly higher as was the case here. Analysis can be largely automated, which is helpful for repeated usage.

4.2.2 Conducting the In-Depth Approach

For the in-depth approach knowledge about McCarthy & Wright's framework of thought (dialogue; see section 1.5) is required. Then one has to pick a suiting method and acquire knowledge about it (Descriptive Experience Sampling or DES; see section 1.7). Also knowledge about qualitative analysis in general is required. DES or more exactly the explicitation interview requires quite some amount of training, because it is not easy to ask the right questions. Hurlburt and Heavey relate this to learning to play the violin (Hurlburt and Heavey 2006, p. 129).

As preparation for the study I had to develop the beeper app (see section 2.5) which took me several weeks as I also had to get to know the Android platform. Of course the app can be reused and there might be already a beeper device or app available. Also sheets for transcription with time codes have to be prepared once.

Recruiting participants is quite difficult as their effort is very high (several hours on different days). When budget allows sufficient compensation, which was not the case here, it might help. Still, you have to find participants that are interested in exploring their inner experience. I had to rely on the participants' personal interest in helping me.

The work load of conducting the study higher compared to the modelling approach as you have to invite participants, have to schedule introductory meetings and have to conduct the interviews were each of them at least lasts half an hour. The cost of analysis depends on the skill of the researcher, but is higher than analysing the data of the modelling approach unless interviews are not transcribed. Analysis may require transcription of interviews, writing summaries of beeps, immersing yourself into the data (listening again to all recordings and looking at summaries), code and group concepts and produce visualisations. As I did the analysis for the first time I might have used nearly twice as much time than after repeated use. However, here lies also the possibility of cost reduction. A skilled researcher can do analysis considerably faster and also interviews may not have to being transcribed but recordings could be coded directly. Discovering patterns in the data requires skill, so a skilled researcher might discover them immediately. It also has to be noted that overall cost of the in-depth approach depends on the selected method. DES may require more time than some other method that fits into McCarthy & Wright's framework.

4.2.3 Stakeholder's & Author's View

In the opinions of stakeholders and author also the relatively easy employment of AttrakDiff and the costly conduction of DES were prevailing. Stakeholders noted that the questionnaire is easy to answer and everybody could do it. They saw DES as costly and doubted that they could have afforded to pay someone to do it for their project. In my view the questionnaire is easy to conduct and distribute. The strengths of DES are its in-situ measurements but its costly conduction and analysis lie clearly on the downside.

4.3 Reflection on Using DES for HCI

In section 1.7.4 I have listed several modifications that are necessary to the original Descriptive Experience Sampling (DES) method to make it work for Human-Computer Interaction (HCI). One modification concerns the average interval between subsequent beeps. For HCI this average interval has to be significantly shorter as in the original method. I chose 5 minutes as average distance between two beeps and 10 minutes as maximum. In practice this settings worked quite well, maybe the intervals could be a bit shorter. Perceptions of appropriate interval lengths differed between participants. The same interval lengths may be perceived as too short when engaged in challenging activities and may be perceived as too long when bored. It may be appropriate to choose slightly asymmetric intervals with a shorter distance between minimum and average and a greater distance between average and maximum.

Another difference concerns training. Since a specific software or device should be investigated it is not possible to do the training on that software or device because it may reduce the amount of data available for analysis. In my case it was not possible to let the participants use the website for training since they might have been already bored of it afterwards. Therefore I instructed participants to sample without restricted context (e.g. their daily activities) for training. This included a change of context from training to the actual usage of the website. I have observed that some participants did not cope very well with this change of context. While they had already accustomed themselves to the method in the unrestricted condition everything seemed rather new in the HCI condition despite nothing changed regarding the method. Solutions to this problem might be to explicitly tell participants they should do everything in the same way as before the context change or to let the training also involve a technological artefact.

HCI requires to extend the time interval of investigation a bit beyond the beep in order to capture the interaction. However, the researcher has to resist the urge to extend this time interval too much. If the moment of the beep is left there is substantial risk that self- and other theories are introduced that distort the data. Instead, it is better to reduce the intervals between subsequent beeps to capture more moments.

5 Conclusion

I started out by being interested in the concept of *user experience* in the field of human-computer interaction. Soon I realised that there is a significant debate about what the concept is or represents, how to research it (see sections 1.1, 1.2) and not to forget the old bickering between theory and practise (see section 1.3). There seemed to be two movements that were competing against each other. There was Hassenzahl's model of user experience (see section 1.4) and McCarthy & Wright's framework of user centred design (see section 1.5). Most researchers were in favour of one or the other and only few used both approaches e.g. Karapanos et al. (2009). This splitting of the field has been criticised. There have been warnings not to emphasize the methodological stance to an extend where it damages research quality (Bargas-Avila and Hornbæk 2011) and others proposed to work towards a unified view (Law, Hvannberg, et al. 2006; Law, Vermeeren, et al. 2007; Mahlke 2007).

My first questions here were: Is a unified view possible? Is it beneficial? To answer this questions it first has to be clear what is meant by unified view. For the authors above it takes the form of "integrating different theoretical perspectives" (Law, Vermeeren, et al. 2007, p. 206) and "relating the approach described here to the framework by McCarthy & Wright. Nonetheless, I think that a solution is necessary to avoid a situation in that energy is wasted in discussions between these two perspectives …" (Mahlke 2007, p. 29). What is meant by *integrating* and *relating to* remains vague. At first sight the call for a unified view of those two competing strands in UX research seem unique. When looking closer it turns out that this is not the case. It is rather a debate between what is called first person approaches and third person approaches (see section 1.6) and similar debates exist in other fields (e.g. cognitive science). There are epistemological implications that are rather unlikely to be solved by UX research and I will not attempt to solve them here either. So let us set aside these implications.

My goal was now to compare both approaches. I did not really want to see which approach is generally better as I thought that both are useful and their goodness cannot be easily decided. This opinion has not changed during this work. I also have to note that I am closer to McCarthy & Wright's worldview which also did not change during this work. Nevertheless, I think that it is necessary to bridge the gap between worldviews which this comparison tried to do by exposing which approach is useful in which circumstances, what kind of results each of them yields and what are the implications when using them in practise (see section 2.1 for details how I approached this). The expected result was that it is necessary to use them both together to be able to get a complete picture, which was confirmed. Hassenzahl's model is more useful for evaluation but has limited capabilities of generating explanations for its results

while McCarthy & Wright's framework (and here Descriptive Experience Sampling) is most useful for guiding design or exploration, is able to uncover most facts without other methods, but is very costly in conduction. But let us look at the details.

We can compare if one approach discovered something the other one has missed out. This requires that we assume that the assumptions of Hassenzahl's model are correct and we trust its validity. There are procedures to establish validity of a questionnaire but these rely mostly on comparing expert judgements with questionnaire results. We also have to trust the researcher who immersed himself into the data to draw the right conclusions. This is much harder to verify, but there are measures like inter-rater agreement. Here I have to trust myself. We then see (chapter 4) that the questionnaire discovered a (possible) degradation of website ratings with increasing age that DES did not capture. We also see that the questionnaire did not provide much explanations for its results (e.g. a steep learning curve might have caused lower ratings for usability) while DES was quite detailed and helpful here. Partly this can be attributed to limitations of the study. If elderly people would have participated in DES it may have been possible to see a decline in rating. If AttrakDiff would have been employed repeatedly there an increase in rating might have become visible. Clearly, AttrakDiff's perceived inability to provide explanations for *why* the ratings have been that way cannot be attributed to limitations of the study. When we now look at how each approach is conducted we see that AttrakDiff is quite easy to employ while DES is quite costly to employ and depends on researcher skill. Also directly coding recordings (needs skill) and using the right tools for analysis is helpful. Still I am sure that it remains more costly than AttrakDiff.

To sum it up we could say that AttrakDiff is easy to employ but misses to provide explanations for its results. On the other hand, DES provides detailed and informative insights, but is quite costly to employ and needs a skilled researcher. There is the need to add detailed context information to AttrakDiff while an in-depth approach like DES should be easier to employ. This is possible through expert researchers and streamlined tools for analysis. Marc Hassenzahl acknowledges the need for exploring the context (Hassenzahl 2010, p. 75), but he does not state that leaving it out is simply insufficient. Nevertheless, AttrakDiff in its official version does exactly that. On the other hand, McCarthy & Wright might have a too mechanistic view of modelling and statistics. They state that it is impossible to infer design implications "without any creative or imaginative intervention of the person doing the analysis" (Wright and McCarthy 2010, p. 67). While being certainly true and acknowledging that statistics has a big set of rules when to do what if some preconditions are met it is not true that this type of analysis is simply a deduction of results. This is illustrated by the fact that I have discovered the possible degradation of ratings with increasing age rather by chance when looking at the data and went on to investigate it further.

If this conclusion now should result in an advice it includes the rephrased criticism of Bargas-Avila and Hornbæk: Don't emphasise your methodological stance to the extend of damaging research quality. An integrated view from different perspectives is more valuable than a unified view. *Talking* to the users in one or the other way is not a matter of choice.

Disclosure

I am employed part-time at TAO Software. This is the company that developed the evaluated real estate website. Nevertheless, I have at no time been involved in the design and implementation of the website. On the contrary, it has strictly been taken care of that I know as little as possible about the product before the start of the evaluation. The study has been conducted unpaid in my free time as part of my study. Therefore I believe that considerable care has been taken that there is no conflict of interest.

Appendix

Example of a Transcribed Interview with Summaries

After the transcription of the interview follow the summaries of the beeps with inserted labels of data categories and answers to additional questions at the end. Only the summaries have been translated. The titles for the beeps have been entered into the app by the participants. For descriptions of the data categories see section 3.2.2.

Interview Barbara-03 02.03.2013 20:02
Name of audio file: barbara_03_20130302.wav

00:00	
00:15	
00:30	
00:45	
01:00	1. Beep, Zimmerfokus
00:15	
01:30	Es war nach 10 min, wo ich auf der Homepage herumgeschaut habe
01:45	Aufmerksam war ich schon, eigentlich durchgehend, Stimmung, ich war gerade positiv überrascht, weil ich gerade in dem Moment etwas gecheckt habe
02:00	Da hab ich mir gedacht, ah, so ist das
02:15	Positiv überrascht, dass ich das kapiert habe und das dass auf der HP so war, weil ich das gut gefunden habe
02:30	
02:45	Hab mir gedacht, wenn das so ist, dann such ich jetzt nach dem und dem, es war so ein entschlossenes Gefühl
03:00	
03:15	Ich bin draufgekommen, dass es Zimmer und Whg in getrennten Farben gibt, und ich habe beschlossen ich konzentriere mich jetzt auf die Zimmersuche
03:30	Nur weil so und so viele Zimmer steht, weil bei den Häusern steht 0 Zimmer
03:45	Bei den Whg stehen halt immer die Zimmer dabei, aber manchmal ist es von 400 – 1000 Euro und manchmal ist es im zweistelligen Tausender-Bereich, da hab ich dann kapiert, das eine ist Whg mieten und das andere ist Whg kaufen
04:00	Das hat aber die gleiche Farbe, weil alle Whg die gleiche Farbe haben, nur Whg und Häuser haben unterschiedliche, dann hab ich mir gedacht, mal schauen wie viel das kostet eine Whg zu mieten

04:15
04:30 Am Anfang hab ich irgendwo komisch geklickt und das war gleich weg, das Fenster, bevor ich mich damit befassen konnte
04:45 Da war ich ein bisschen überrascht und leicht verärgert, dann hab ich irgendwo hingeklickt und dann kam gleich irgendwo hergezoomt und aufgegangen, irgendwas und wieder weg und ich hab mich überhaupt nicht ausgekannt
05:00 Und ich hab nicht gecheckt was diese Ringerl in den Kreisen sollen, denn ich dachte das ist nr 1, nr 2, nr 3 und nicht da sind 3 Objekte, da bin ich dann draufgekommen
05:15
05:30 In meiner Erinnerung haben kaufen und mieten gleich ausgeschaut
05:45 Alarm war gerade als ich draufgekommen bin mit mieten und kaufen
06:00 Ganz am Anfang hab ich geschaut, das ist die Stadt, dann hab ich hinausgezoomt und geschaut bis wie weit es in der Stadt solche Punkte gibt
06:15 Dann hab ich mir überlegt was ein guter Bezirk zum Wohnen ist, aber das sieht man natürlich auf der Karte nicht, dann hab ich geschaut wo es die meisten Whg gibt
06:30 Lustigerweise mitten drinnen, und dann hab ich mir gedacht ich vergleiche mal wie das ist eine Wohnung ganz außen zu haben und ganz innen und dann hab ich auf das alleräußerste geklickt was ich gefunden hab
06:45 Das erste war ein Haus, das zweite war dann eben die Whg wo ich dann das gecheckt hab
07:00
07:15
07:30 Es war so ein bisschen ein neuer Schwung wieder drinnen, weil ich wieder ein neues Ziel hatte, weil ich beschlossen hab ich mache jetzt das, ich suche Whg die man mieten kann
07:45 Hab gedacht, ok, geht schon, mach ma, such ma
08:00
08:15
08:30 Ich hab ziemlich random auf den nächsten Punkt geklickt, den ich gefunden hab
08:45 Dann hab ich da wieder geschaut, das waren voll viele, nein nicht random, das war das wo es die meisten gab, da gab es die meisten zu kaufen
09:00 Die hab ich alle nicht angeklickt, sondern nur die, die vom Preis so ausgeschaut haben, als ob das Mietwhg wären
09:15 2. Beep: Whg betreten
09:30 Da hab ich gerade, das war die zweite Whg, die ich von diesen vielen hatte, da hab ich mir vorher die Fotos angeschaut und bei den Fotos ist das zweite/dritte der Plan gewesen
09:45 Der Grundriss von der Whg, dann hab ich mir den angeschaut

10:00 Immer wenn ich einen Whg-Plan hab, dann stelle ich mir vor wie ich da reingehe oder durchgehe, ich versetze mich praktisch in den Plan hinein, gerade wie ich ins Vorzimmer hineingegangen bin durch die Eingangstür hat das Ding geläutet

10:15 Da war ich fast noch mehr versunken in dem Ganzen, weil ich mich darauf konzentrierte das vorzustellen

10:30 Die Whg und wie sie ausschaut, es war gespannt, interessiert

10:45

11:00 Da ich vorher die Fotos gesehen hab, wusste ich ein bisschen was ich mir vorstellen muss, aber es war kein Foto vom Vorzimmer, das konnte ich mir nicht vorstellen

11:15 Mein Hirn hat so ein bisschen, das wo es gewusst hat dass es allgemein gültig ist, hat es übernommen, z. B. das Parkett überall ist, da hab ich gewusst wie es ausschaut, auch die Wände hab ich gewusst wie die aussehen auf den Fotos

11:30 Sonst war noch alles leer und weiß

11:45 Es war keine Tür da, nur der Türrahmen

12:00 Es war genau so unter der Schwelle

12:15 Am Anfang sehe ich den Plan und dann sehe ich das noch von oben, als ob ich mich selber von der Vogelperspektive sehen würde, wenn ich reingehe

12:30 Ich sehe mich nicht selber, aber ich sehe die Whg von oben, aber ich sehe nicht den Plan sondern die Whg

12:45 Aber noch von oben

13:00 Wenn ich weitergegangen wäre, dann wäre ich drinnen gewesen, dann hätte ich die Wände auf Augenhöhe gesehen oder auf der Höhe wie sie sind

13:15 aber so hab ich es noch von oben gesehen, weil ich gerade erst noch beim Reingehen war

13:30 Es war eher so ein Hineinzoomen

13:45 Es war so ein bisschen Harry Potter Style

14:00 Wie wenn sie am Besen fliegen und dann so runterstechen, deshalb hab ich auch gesagt unter der Schwelle

14:15 Weil ich gerade so in der Bewegung war, ich war gerade in diesem Schwiuuuu

14:30 Es war genau so wie es ist, praktisch, die Whg

14:45 Außerhalb von der Whg, die waren so Beton, so grauer, nasser, unverputzter Beton, als ob das wie ein Modell irgendwo stehen würde

15:00

15:15

15:30

15:45

16:00 Es war komisch, weil die Bilder haben ein bisschen verzerrt ausgeschaut

16:15 Sie haben ausgeschaut, als ob sie in die Länge gezogen wären, ich dachte uä, entweder das ist ur schiarch und schmal oder sie sind in die Länge gezogen, das nächste war auch so, dann hab ich mir gedacht, ok, es sind die Bilder in die länge gezogen

16:30 Das hätte man schon besser fotografieren können, wenn man es verkaufen will, nach dem Alarm hab ich sie nochmal durchgeschaut, dann waren sie nicht mehr so verzerrt

16:45 Entweder ich hab das nur falsch wahrgenommen oder ich hab mich dann dran gewöhnt gehabt oder es hat sich was geändert

17:00 3. Beep Carls Whg

17:15 Da war ich auch wieder bei einer Whg

17:30 Da hab ich wieder die Fotos angeschaut, das erste Foto hat ausgeschaut, ein bisschen, wie die Whg von einem Freund von mir in Graz

17:45 Da dachte ich mir, ah, lustig, das schaut aus wie die Whg vom Carl, ich glaube es war das Wohnzimmer drauf

18:00 Man sieht so ein Fenster oder eineinhalb Fenster, einen Schrank und eine Couchhälfte, ich dachte mir das wäre ja lustig wenn das seine Wohnung wäre und wenn man die da drin finden würde

18:15 Wenn ich das über die Seite erfahren würde oder finden würde

18:30

18:45 Dann hab ich mir gedacht, nein, eigentlich schaut sie nicht ganz so aus

19:00 Dann wüsste ich das sich was getan hat, weil die Whg war eine WG und dann würde ich wissen, dass sich die WG aufgelöst hat

19:15 Die Aufmerksamkeit war vielleicht schon ein bisschen vermindert, weil das war sicher schon die fünfte sechste Wohnung, die ich angeschaut hab

19:30 Da tut man dann nur mehr so durchblättern, ein bissl oberflächlicher

19:45 Aber die Stimmung war trotzdem noch heiter, würde der Wetterbericht sagen

20:00 Ich war gut gelaunt, so gestimmt, ja, schaun wir uns noch was an, es war nachwievor gut

20:15

20:30

20:45 Den Beschreibungstext hab ich mir nur bei der ersten Whg durchgelesen, dann hab ich mir gedacht das ist mir zu mühsam

21:00 Weil Beschreibungstexte sind meistens nur so, ach es ist so schön und so toll und blablabla, dabei kannst du denen überhaupt nicht glauben, ich finde Whg muss man beurteilen nach Fotos und nach den Fakten einfach

21:15 Diese Texte sind sowieso immer nur ein Geschwafel, Fakten sind wieviel es kostet, wie groß es ist, wo es liegt, wie die Zimmereinteilung ist

21:30 Der Plan ist voll wichtig finde ich und die Fotos halt, die Fotos sind auch Fakten

21:45 Nicht 100%-Fakten, denn wenn ich lese 120 m^2 dann weiß ich wieviel das ist, wenn ich das Foto sehe dann weiß ich nicht immer wieviel das ist

22:00 Fotos sind schon Fakten, aber sind eine andere Kategorie wie Zahlen, Preis, aber sind trotzdem wichtig zum beurteilen

22:15

22:30

22:45

23:00 Ich mit der Wohnung angefangen, weil sie ganz unten war, ich wollte die Preise zwischen Innenstadt und Stadtgrenze vergleichen, dann musste ich wieder weiter ins Zentrum

23:15 Dann dachte ich mir ich gehe ganz in die Mitte, ins Ballungszentrum wo die meisten Whg sind, dann hab ich den Kreis angeklickt mit den meisten Whg

23:30 Ich bin mir nicht sicher ob ich danach die ganze Zeit in einem Kreis war oder ob die letzte Whg außerhalb von dem in einem dritten Bereich war

23:45

24:00

24:15 Es waren vier Seiten glaub ich und man konnte einfach so weiterblättern, da habe ich immer auf die Whg geklickt die nach Preis ausgesehen haben als ob es Mietwhg wären

24:30

24:45 Ich hab mich zuerst gefragt, weil da steht überhaupt keine Beschreibung dabei, links ist so ein Pfeil in einem extra weißen Kastl, was klappe ich damit auf, wenn ich da draufklicke

25:00 Dann bin ich draufgekommen das ist so ein Filter, den ich auch nicht ganz durchschaut habe, zuerst ist Geld, dann Quadratmeter und dann Zimmer

25:15 Du musst einen Preisrahmen eingeben, von – bis, da kannst du zweimal den Preis eingeben, das erste Ding hab ich nicht ganz kapiert

25:30

25:45

26:00

26:15

26:30 Ich habs nicht ausprobiert, weil ich mir gedacht habe, ich suche nicht ernsthaft eine Whg und deshalb habe ich es nicht gemacht

26:45 Aber sonst hätte ich bestimmt diese Filteroption genommen, denn an sich ist das sehr geschickt

27:00

27:15 Das Menü hab ich kurz vorm ersten Alarm bemerkt

27:30

27:45

28:00

28:15

28:30

28:45 Nachdem ich gesehen hab auf der einen Seite gibt's was, dann dachte ich mir ich schau auch auf die andere Seite

29:00 Da war auch Filter, an das erste kann ich mich nicht mehr erinnern, da war auch Zielgruppe hat das geheißen, das habe ich auch sehr clever gefunden

29:15 Da hat man gemerkt, dass sie sich was überlegt haben, mit Studenten, dass man da schaut das es provisionsfrei ist und das es öffentliche Verkehrsanbindungen hat, dass es WG-tauglich ist, bei Senioren, dass es barrierefrei ist

29:30 Ich hab mir die Kategorien angeschaut und dann gab es ein Ding, dieses Billa, das hab ich lustig gefunden, gesponsert von Billa
29:45
30:00 Das war schräg, so nach dem Motto, und was ist mit den anderen Supermärkten
30:15 Draufgeklickt hab ich dann auf wir haben laufend neue … für sie
30:30 Das Zeichen war die Karte von dem Ganzen
30:45 Aber da kam nix
31:00
31:15
31:30 Da hab ich den Filter ausprobiert von dem Rechten
31:45 Ich hab diesen Filter eingestellt auf Provisionsfrei und mit Balkon, glaub ich
32:00 Dann haben sich neben mir auf dem Hauptstadtplan haben sich die Zahlen deutlich verringert in diesen Kreisen und da hab ich dann auf den geklickt, der am meisten hatte
32:15
32:30 Dann haben sich neben mir auf dem Hauptstadtplan haben sich die Zahlen deutlich verringert in diesen Kreisen und da hab ich dann auf den geklickt, der am meisten hatte
32:45
33:00
33:15
33:30 Ich weiß nicht ob ich Erwartungen hatte, eigentlich war ich hauptsächlich gespannt wie es aussehen wird, wie es ist, es war auf jeden Fall besser als ich es mir erwartet hab
33:45 Ich hab an so ein Ding mit so einer Liste mit Whg drauf, so eine langweilige, gedacht
34:00 Sie wurden übertroffen, weil es viel besser aufbereitet war als ich mir gedacht habe
34:15 Es war nicht so eine Liste unter der man sich nichts vorstellen kann, es war echt übersichtlich mit dem Plan, ein bisschen verwirrend war es, es wäre gut gewesen wenn irgendwo gestanden wäre
34:30 Wie das ganze funktioniert, durch dahin und dorthin klicken zoomen sie wo hin
34:45 Und vielleicht auch eine Art Legende, um zu erklären was diese Kreise da sind, man kommt aber eh drauf
35:00 Am Anfang hab ich mich schon ein bisschen geärgert, weil ich irgendwohin geklickt und es hat dann immer was anderes gemacht als es hätte sollen
35:15 Ich hab versucht in das Feld der Whg zu klicken, dann war ich offensichtlich nicht genau drauf und dahinter hat sich der Stadtplan wieder verschoben
35:30 Dann kommt so ein größeres Ding, das kommt aber nicht zentriert, das ist unten abgeschnitten

35:45	Das wollte ich raufschieben und das konnte ich nicht raufschieben mit dem scrollen, obwohl ich geglaubt hatte dass ich die Maus drin hatte, weil ich hab dann den Stadtplan herumgeschoben, dann war die Whg wieder weg
36:00	
36:15	
36:30	
36:45	
37:00	Ich würde noch weiter schauen, um noch zu vergleichen, oder ich schaue mir noch Häuser an, aber alles gesehen habe ich noch nicht
37:15	
37:30	Wenn ich Wohnungssuchende wäre, wäre das eine Seite wo ich sicher gerne suchen würde
37:45	Das wäre sicher ein nettes Herumschauen, weil man gerade das Gefühl hat, dass man ein bisschen herumsurft
38:00	
38:15	
38:30	
38:45	
39:00	Bunt, nett gezeichnet
39:15	Ich hab ein bisschen das Gefühl, dass es wie ein Comic ist
39:30	anschaulich

1. beep: focus on rooms

Barbara is not able to enter a zip code because she clicks somewhere and the window disappears suddenly. ACT She is surprised and also slightly upset, because something zoomed and opened and she was completely lost. FEEL She also doesn't know what the numbers within the circles should mean. First she thought it would be no. 1, no. 2 and so on instead of the number of objects. She finds out afterwards. REFL First she zooms out and gets an overview over the city. Then she thinks about where a good district to live would be, but she does not see it on the map. ACT She discovers an arrow without description ACT and she asks herself what will expand if she clicks on it. Then she sees that it is a filter, but she cannot completely figure out how it works - at least regarding to the price. She can enter a price two times, which she finds odd. REFL She doesn't check out the filter because she is not really looking for a flat. After she has discovered the left side she also looks to the right. There is also a filter and a button for target groups ACT, which she finds clever. Obviously the creators of the site did really think about it. REFL

She is looking at the categories and recognizes the Billa [Austrian grocery chain] sponsoring ACT, which she finds funny and weird. She asks herself what has happened to the other grocery stores. REFL She checks out the filter with free-of-commission and balcony and has to click several times because the setting isn't saved immediately. On the map the numbers

within the circles get smaller and she clicks onto the circle with the largest number. ACT At the beep she understands that flats to rent and to buy have the same color. She can distinguish them by price. Flats to rent cost 400 to 1000 Euro and flats to buy cost mostly more than 50.000 Euro. Now a new drive for searching emerges, because she has a new goal: searching for flats to rent. EXP In order to do this she clicks through the list and only looks at those entries that look like flats to rent according to their price. ACT

presence: attentive
moods: surprised, decisive, slightly upset, funny, weird, confused, curious

2. beep: entering the flat

Barbara looks at the photos of a flat which are oddly distorted. ACT She thinks that the photos could be better since the flat is tried to be sold. REFL One photo contains the groundplan which she views in detail. ACT She imagines how she enters the flat and moves through the door. Form the photos she roughly knows how the flat looks like, although the photo of the anteroom is missing. So she creates her own in her head, it is white, without furniture but with parquet flooring. There is no door only the the door frame. At the beginning she sees the flat from above and she is moving downwards until she sees the walls at eye level. It is a movement like in the Harry Potter movies when they sit on their brooms and are nosediving down. At the beep she is exactly beneath the door but is not yet at the ground. The walls outside the flat are grey, wet, unplastered concrete walls that look like a model or mockup. EXP

presence: very attentive
moods: keen, interested, confused

3. beep: my friend's flat

Barbara looks again at the photos of a flat. She is just browsing through them. The first photo looks like the flat of a friend. It was a living room with one window, a cupboard and the half of a couch on it. ACT That would be funny if that would be his flat and if should would find it here. FEEL Then she would know that they had moved out. But then she thinks that the flat actually doesn't really look like the flat of her friend. REFL She is not reading the description ACT because it is too tedious and most of the time you cannot trust these texts anyway. REFL

presence: not so attentive
moods: cheerful, good mood, indifferent

Additional Questions

She is not sure if she had expectations. Actually she was mostly curious how it would look like. It was in any case better than expected. She thought about a site with a boring list of flats. So her expectations were fulfilled or even exceeded since it was much better visualized than she would have thought. It was not just only a list where you cannot imagine anything. It was really neat with the map. REFL It was a bit confusing and it would have been good if somewhere had been written how everything works and maybe also some sort of key SUGG to explain what the circles are. But then she got it anyway. REFL At the beginning she was slightly upset because she has clicked somewhere and the site behaved completely different than it should have behaved. REFL

For example, she wanted to click onto the marker of the flat. Obviously she was not exactly above it and the map behind it moved. REFL The detailed view was not centred but cut off at the bottom. She wanted to move it, but she couldn't do it by scrolling. Then she changed the map and the flat was gone. REFL

If she would search for a flat she would use the website gladly. It would be a nice digging around because you have the feeling that you surf around or fly over the city. REFL It is colourful and nicely drawn. It is somehow like a comic. REFL

All Themes including Sub-Themes

Below a detailed list of all sub-themes to the themes of Table 3.5 including data categories (see section 3.2.2) and number of samples and persons where the theme manifested itself is provided.

steep learning curve

Name	Data Category	No. Samples	No. Persons
You have to try to understand the site.	REFL	1	1

initial confusion, being overwhelmed

Name	Data Category	No. Samples	No. Persons
finding cluster marker (function?)	ACT	5	5
Is it me or is it the website	REFL	4	4
uncertainty about marker's role	REFL	2	2
key to the map	SUGG	2	2
filters are hidden	REFL	2	2

orienting	ACT	2	1
scrolling in list zooms map in background	ACT	1	1
searching for key	ACT	1	1
uncertainty about role of upper box ("Searching for ...")	EXP	1	1
uncertainty about meaning of splitting into target groups and category	REFL	1	1
uncertainty about price selection in left menu	REFL	1	1
navigation is complicated	REFL	1	1
overlooked left menu	REFL	1	1
uncertainty if settings are saved	REFL	1	1
slightly confusing	REFL	1	1
everything is displayed instantly, then having to select	REFL	1	1
you're fine if you know where everything is	REFL	1	1
site behaves oddly after careless clicks	ACT	1	1
would be nice to have a manual	SUGG	1	1
careless clicks are moving the map	REFL	1	1

trying-out necessary

Name	Data Category	No. Samples	No. Persons
try out zooming	ACT	3	2
try out functioning of cluster marker	ACT	2	2
try out right menu	ACT	2	2
try out functioning of marker	ACT	1	1
(having to) make a move	REFL	1	1
You have to try out things, have to click ahead	REFL	1	1

get an overview, look around, "ramble about"

Name	Data Category	No. Samples	No. Persons
get an overview	REFL	6	2
look around on the map	ACT	5	3
cluster marker shows available flats in a region	REFL	3	2
find out what is available	REFL	2	1

cluster marker reduces complexity	REFL	1	1
find out what is available in the vicinity	REFL	1	1
surf around, fly over the city	REFL	1	1

site and content are a unity

Name	Data Category	No. Samples	No. Persons
exact address of flat	REFL / SUGG	2	2
ongoing updates raise interest	REFL	1	1
good content is important for the website	REFL	1	1
flat has no balcony despite selection	REFL	1	1
hardly any choices left after filtering	ACT	1	1
would be good to have more objects in database	REFL	1	1

photos are distorted or poorly made

Name	Data Category	No. Samples	No. Persons
photos are distorted or pixelated	REFL	4	3
odd, useless photos of flat	REFL	2	2
envisioning of room hampered by poorly made photos	REFL	1	1
photos could be of better quality	REFL	1	1

photos are important

Name	Data Category	No. Samples	No. Persons
looking at photos	ACT	10	4
photos stand out	REFL	1	1
photos are important	REFL	1	1
magnify photos by clicking on them (lightbox)	ACT	1	1
recognise photo	EXP	1	1

issues with cut-off detail view

Name	Data Category	No. Samples	No. Persons
detail view cut-off, at wrong position	ACT	7	3
scrolling to move detail view	ACT	3	3
dragging photo to re-position detail view	ACT	1	1
magnify photos by clicking on them (lightbox)	ACT	1	1

website offers freedom of search

Name	Data Category	No. Samples	No. Persons
different possibilities for filtering and sorting	REFL	3	3
it is possible to approach the site differently	REFL	3	2
searching with different parameters	REFL	1	1
you need only one tool to reach your goal	REFL	1	1
everything is displayed instantly, then having to select	REFL	1	1
the site has everything needed for real estate search	REFL	1	1
it is better to use the list for comparisons	REFL	1	1

sensibility of loading time, reaction time, visual feedback

Name	Data Category	No. Samples	No. Persons
being sensible of loading time, delays	FEEL	4	4
delayed reaction of checkboxes in right menu	ACT	3	3
uncertainty about how map behaves	REFL	3	2

description text's presentation is subpar

Name	Data Category	No. Samples	No. Persons
description is repeating itself	REFL	3	3
not reading through description	REFL	3	2
description text is tedious to read	REFL	3	1
better structure for description text	SUGG	1	1

photos and groundplan enable envisioning of flat

Name	Data Category	No. Samples	No. Persons
envisioning the flat in your head	EXP	4	3
groundplan enables envisioning of the flat	REFL	3	3
zoom into the photo thought-wise	EXP	1	1
alter the groundplan in your head	EXP	1	1

advertisement stands out and polarises

Name	Data Category	No. Samples	No. Persons
looking at Billa [Austrian grocery chain] advertisement	ACT	4	4
ignoring Billa advertisement	REFL	1	1
wondering that Billa is part of the target groups menu	FEEL	1	1
Billa advertisement stands out in a negative way	REFL	1	1
envisioning a map with Billa advertisement	EXP	1	1

high expectations towards the site

Name	Data Category	No. Samples	No. Persons
raised expectations	REFL	3	1
there are some problems to solve	REFL	1	1
another site filled with advertisements	REFL	1	1

groundplan is important

Name	Data Category	No. Samples	No. Persons
looking at groundplan	ACT	4	4
groundplan for each flat desired	SUGG	1	1
groundplan should have its own section	SUGG	1	1
zooming the groundplan	SUGG	1	1
groundplan supports freedom of thorough informing	REFL	1	1

description text is unreliable

Name	Data Category	No. Samples	No. Persons
description is fuzzy	REFL	1	1
descriptions of estate agents tend to be biased	REFL	1	1
you cannot trust descriptions	REFL	1	1

starring feature is hidden

Name	Data Category	No. Samples	No. Persons
did not find out how to add flat to favourites	ACT	2	2
overlooked star for starring	REFL	2	2
starring feature is hidden	REFL	1	1
dragging flat onto favourites icon	ACT	1	1
seen star for starring as star for rating	REFL	1	1

reset feature is hidden

Name	Data Category	No. Samples	No. Persons
start from the beginning	ACT	3	2
did not know how to undo wrong selection at start screen	ACT	1	1

Bibliography

Bargas-Avila, J. A. & Hornbæk, K. (2011). Old wine in new bottles or novel challenges: a critical analysis of empirical studies of user experience. In *Proceedings of the 2011 Annual Conference on Human Factors in Computing Systems* (pp. 2689–2698). CHI '11. Vancouver, BC, Canada: ACM. doi:10.1145/1978942.1979336. at pp. 2, 3, 12, 17, 51, 52

Bertelsen, O. W. & Pold, S. (2004). Criticism as an approach to interface aesthetics. In *Proceedings of the 3rd Nordic Conference on Human-Computer Interaction* (pp. 23–32). ACM. doi:10.1145/1028014.1028018. at p. 10

Blythe, M. & Wright, P. (2006). Pastiche scenarios: fiction as a resource for user centred design. *Interacting with Computers*, *18*(5), 1139–1164. doi:10.1016/j.intcom.2006.02.001. at p. 10

Box, G. E. P. (1953, Dec.). Non-normality and tests on variances. *Biometrika*, *40*(3/4), 318–335. at p. 26

Bruner, J. S. (2004). Life as narrative. *Social Research*, *71*(3), 691–710. at p. 8

Buchenau, M. & Suri, J. F. (2000). Experience prototyping. In *Proceedings of the 3rd conference on Designing interactive systems: processes, practices, methods, and techniques* (pp. 424–433). ACM. doi:10.1145/347642.347802. at p. 10

Cockton, G. (2006). Valuing user experience. In E. L.-C. Law, E. T. Hvannberg, & M. Hassenzahl (Eds.), *User Experience – Towards a Unified View: Second International COST294-MAUSE Open Workshop* (pp. 100–105). Oslo, Norway. at pp. 2, 4, 11

Compeau, D. R. & Higgins, C. A. (1995). Computer self-efficacy: development of a measure and initial test. *Management Information Systems Quarterly*, *19*(2), 189–211. at p. 40

Csikszentmihalyi, M. & Larson, R. (1992). Validity and reliability of the experience sampling method. *The experience of psychopathology: Investigating mental disorders in their natural settings*, 43–57. at p. 12

Dewey, J. (1934). *Art as experience*. Perigee Books. at pp. 7, 8

Dindler, C. & Iversen, O. S. (2007). Fictional inquiry–design collaboration in a shared narrative space. *CoDesign*, *3*(4), 213–234. doi:10.1080/15710880701500187. at p. 10

Flick, U. (2008). *Designing qualitative research*. Qualitative Research Kit. SAGE Publications. at pp. 20, 21, 43

Gaver, B., Dunne, T., & Pacenti, E. (1999). Design: cultural probes. *interactions*, *6*(1), 21–29. doi:10.1145/291224.291235. at p. 10

Glanznig, M. (2012a). Beepme: android experience sampling app. Retrieved from `http://beepme.glanznig.com`. at p. 22

Glanznig, M. (2012b). User experience research: modelling and describing the subjective. *Interdisciplinary Description of Complex Systems*, *10*(3), 235–247. doi:10.7906/indecs.10.3.3. at p. 1

Glaser, B. G. & Strauss, A. L. (1967). *The discovery of grounded theory: strategies for grounded research*. Aldine de Gruyter. at p. 32

Hassenzahl, M. (2005). The thing and i: understanding the relationship between user and product. In M. Blythe, K. Overbeeke, A. Monk, & P. Wright (Eds.), *Funology* (Vol. 3, pp. 31–42). Human–Computer Interaction Series. Springer. doi:10.1007/1-4020-2967-5_4. at pp. 3, 5

Hassenzahl, M. (2007). The hedonic/pragmatic model of user experience. In E. L.-C. Law, A. P. O. S. Vermeeren, M. Hassenzahl, & M. Blythe (Eds.), *Towards a UX Manifesto. COST294-MAUSE affiliated workshop.* (pp. 10–14). Lancaster, UK. at p. 5

Hassenzahl, M. (2008a). The interplay of beauty, goodness, and usability in interactive products. *Human-Computer Interaction*, *19*(4), 319–349. doi:10.1207/s15327051hci1904_2. at pp. 6, 7, 19

Hassenzahl, M. (2008b). User experience (UX): towards an experiential perspective on product quality. In *Proceedings of the 20th International Conference of the Association Francophone d'Interaction Homme-Machine* (pp. 11–15). IHM '08. Metz, France: ACM. doi:10.1145/1512714.1512717. at pp. 2, 4, 5, 8

Hassenzahl, M. (2010). Experience design: technology for all the right reasons. *Synthesis Lectures on Human-Centered Informatics*, *3*(1), 1–95. doi:10.2200/S00261ED1V01Y201003HCI008. at pp. 2–6, 8, 11, 52

Hassenzahl, M., Burmester, M., & Koller, F. (2003). AttrakDiff: Ein Fragebogen zur Messung wahrgenommener hedonischer und pragmatischer Qualität [Attrakdiff: a questionnaire for measuring perceived hedonic and pragmatic quality]. In G. Szwillus & J. Ziegler (Eds.), *Mensch & Computer 2003: Interaktion in Bewegung* (pp. 187–196). Stuttgart, Germany: Teubner. at pp. 5, 6, 19

Hurlburt, R. T. & Heavey, C. L. (2006). *Exploring inner experience: the descriptive experience sampling method*. Advances in Consciousness Research. John Benjamins Publishing Company. at pp. 8, 10, 12–16, 21–24, 47, 49

IDC. (2012). Android marks fourth anniversary since launch with 75 percent market share in third quarter, according to idc [Press release]. International Data Corporation, Framingham, MA, USA. Retrieved from `http://www.idc.com/getdoc.jsp?containerId=prUS23771812`. at p. 22

ISO 9241-210. (2010). Ergonomics of human-system interaction – part 210: human-centred design for interactive systems. International Organization for Standardization, Geneva, Switzerland. Retrieved from http://www.iso.org/iso/iso_catalogue/catalogue_ics/catalogue_detail_ics.htm?csnumber=52075. at pp. 1–3

Kahneman, D., Krueger, A. B., Schkade, D. A., Schwarz, N., & Stone, A. A. (2004). A survey method for characterizing daily life experience: the day reconstruction method. *Science*, *306*(5702), 1776–1780. doi:10.1126/science.1103572. at pp. 10, 12

Karapanos, E., Zimmerman, J., Forlizzi, J., & Martens, J.-B. (2009). User experience over time: an initial framework. In *Proceedings of the 27th International Conference on Human Factors in Computing Systems* (pp. 729–738). CHI '09. Boston, MA, USA: ACM. doi:10.1145/1518701.1518814. at pp. 12, 17, 51

Kerkow, D. (2007). Don't have to know what it is like to be a bat to build a radar reflector - functionalism in UX. In E. L.-C. Law, A. P. O. S. Vermeeren, M. Hassenzahl, & M. Blythe (Eds.), *Towards a UX Manifesto. COST294-MAUSE affiliated workshop.* (pp. 19–25). Lancaster, UK. at p. 3

Kordeš, U. (2012). Thinking of experience, experiencing thinking. *Interdisciplinary Description of Complex Systems*, *10*(3), 223–234. doi:10.7906/indecs.10.3.2. at pp. 11, 17

Kort, J., Vermeeren, A. P. O. S., & Fokker, J. E. (2007). Conceptualizing and measuring user experience. In E. L.-C. Law, A. P. O. S. Vermeeren, M. Hassenzahl, & M. Blythe (Eds.), *Towards a UX Manifesto. COST294-MAUSE affiliated workshop.* (pp. 57–64). Lancaster, UK. at p. 2

Law, E. L.-C., Hvannberg, E. T., & Hassenzahl, M. (2006). User experience – towards a unified view. In E. L.-C. Law, E. T. Hvannberg, & M. Hassenzahl (Eds.), *User Experience – Towards a Unified View: Second International COST294-MAUSE Open Workshop* (pp. 1–3). Oslo, Norway. at pp. 11, 51

Law, E. L.-C., Roto, V., Hassenzahl, M., Vermeeren, A. P., & Kort, J. (2009). Understanding, scoping and defining user experience: a survey approach. In *Proceedings of the 27th International Conference on Human Factors in Computing Systems* (pp. 719–728). CHI '09. Boston, MA, USA: ACM. doi:10.1145/1518701.1518813. at p. 1

Law, E. L.-C., Vermeeren, A. P. O. S., Hassenzahl, M., & Blythe, M. (2007). Towards a UX manifesto. In *Proceedings of the 21st British HCI Group Annual Conference on People and Computers: HCI...but not as we know it - Volume 2* (pp. 205–206). BCS-HCI '07. University of Lancaster, United Kingdom: British Computer Society. at pp. 1, 2, 11, 51

Mahlke, S. (2007). User experience: usability, aesthetics and emotions in human-technology interaction. In E. L.-C. Law, A. P. O. S. Vermeeren, M. Hassenzahl, & M. Blythe (Eds.), *Towards a UX Manifesto. COST294-MAUSE affiliated workshop.* (pp. 26–30). Lancaster, UK. at pp. 11, 51

Markič, O. (2012). First- and third-person approaches: the problem of integration. *Interdisciplinary Description of Complex Systems*, *10*(3), 213–222. doi:10.7906/indecs.10.3.1. at pp. 11, 17

McCarthy, J. & Wright, P. (2004). *Technology as experience*. MIT Press. at pp. 3, 4, 7–9, 11, 12

Russell, J. A. (2003). Core affect and the psychological construction of emotion. *Psychological Review*, *110*(1), 145. doi:10.1037/0033-295X.110.1.145. at p. 4

Stegemann, S. K. & Fiore, S. G. (2006). Designing unscientifically for experience. In E. L.-C. Law, E. T. Hvannberg, & M. Hassenzahl (Eds.), *User Experience – Towards a Unified View: Second International COST294-MAUSE Open Workshop* (pp. 90–93). Oslo, Norway. at pp. 3, 4

Stemler, S. (2001). An overview of content analysis. *Practical assessment, research & evaluation*, *7*(17), 137–146. at p. 32

Swallow, D., Blythe, M., & Wright, P. (2005). Grounding experience: relating theory and method to evaluate the user experience of smartphones. In *Proceedings of the 2005 Annual Conference of the European Association of Cognitive Ergonomics* (pp. 91–98). EACE '05. Chania, Greece: University of Athens. at p. 10

Tractinsky, N., Shoval-Katz, A., & Ikar, D. (2000). What is beautiful is usable. *Interacting with Computers*, *13*(2), 127–145. doi:10.1016/S0953-5438(00)00031-X. at p. 6

van Schaik, P. & Ling, J. (2008). Modelling user experience with web sites: usability, hedonic value, beauty and goodness. *Interacting with Computers*, *20*(3), 419–432. doi:10.1016/j.intcom.2008.03.001. at p. 6

Wechsung, I., Naumann, A., & Möller, S. (2010). The influence of the usage mode on subjectively perceived quality. In G. Lee, J. Mariani, W. Minker, & S. Nakamura (Eds.), *Spoken Dialogue Systems for Ambient Environments* (Vol. 6392, pp. 188–193). Lecture Notes in Computer Science. Springer. doi:10.1007/978-3-642-16202-2_20. at p. 6

Wright, P. & Blythe, M. (2007). User experience research as an inter-discipline: towards a UX manifesto. In E. L.-C. Law, A. P. O. S. Vermeeren, M. Hassenzahl, & M. Blythe (Eds.), *Towards a UX Manifesto. COST294-MAUSE affiliated workshop.* (pp. 65–70). Lancaster, UK. at pp. 2, 10

Wright, P. & McCarthy, J. (2010). Experience-centered design: designers, users, and communities in dialogue. *Synthesis Lectures on Human-Centered Informatics*, *3*(1), 1–123. doi:10.2200/S00229ED1V01Y201003HCI009. at pp. 3, 4, 7, 8, 10, 11, 16, 52

Michael Glanznig
Curriculum Vitae

Education

Oct. 2010 – June 2013 **Joint Master Study Middle European Interdisciplinary Master Programme in Cognitive Science (MEi:CogSci)**, *University of Vienna*.
with emphasis on first person research and Human-Computer Interaction, Master of Science

Sep. 2011 – Feb. 2012 **Semester abroad for MEi:CogSci**, *University of Ljubljana*, Slovenia.

Oct. 2007 – Nov. 2009 **Diploma Study Meteorology**, *University of Vienna*.

Oct. 2004 – June 2008 **Bachelor Study Media and Computer Science**, *Vienna University of Technology*.
with emphasis on design, Bachelor of Science with distinction

Work Experience in Research

May 2009 – July 2011 **Research Assistant (Wissenschaftlicher Mitarbeiter)**, *University of Vienna, Department of Meteorology and Geophysics*, Vienna.
Participating in the research project MANNO (Meteorological Analysis and Nowcasting)

Publications

Glanznig, M. (2012). User Experience Research: Modelling and Describing the Subjective. In *Interdisciplinary Description of Complex Systems*, 10(3), 235–247, doi:10.7906/indecs.10.3.3

Glanznig, M., Malik, M.M., and Gröller, M.E. (2009). Locally Adaptive Marching Cubes through Iso-Value Variation. In *Proceedings of the 17th International Conference in Central Europe on Computer Graphics, Visualization and Computer Vision*, 33–40

Languages

Deutsch **native language**

English **fluent**

Slovenščina **basic knowledge**

We apologise for the inconvenience.

God's Final Message to His Creation

So Long, and Thanks for All the Fish
The Hitchhiker's Guide to the Galaxy
Douglas Adams