



universität
wien

Diplomarbeit

Titel der Arbeit

Der Szenen-Objekt-Konsistenzeffekt und
Ratetendenzen bei Tiefpassfilterung von
Objektbildern

Verfasser

Florian Schrittwieser

Angestrebter akademischer Grad

Magister der Naturwissenschaften (Mag. rer. nat.)

Wien, im Juni 2012

Studienkennzahl: 298

Studienrichtung: Psychologie

Betreuer: Univ.-Prof. Dr. Ulrich Ansorge

Inhaltsverzeichnis

| | | |
|----------|---|-----------|
| 1 | <u>EINFÜHRUNG</u> | 5 |
| 2 | <u>STAND DES WISSENS</u> | 6 |
| 2.1 | ERKLÄRUNGSMODELLE KONSISTENZEFFEKT | 20 |
| 2.2 | TIEFPASSINFORMATIONEN UND TOP-DOWN-PROZESSE | 21 |
| 2.3 | FORSCHUNGSFRAGEN | 28 |
| 3 | <u>METHODE</u> | 29 |
| 3.1 | TEILNEHMER | 29 |
| 3.2 | MATERIALIEN UND GERÄTE | 29 |
| 3.3 | DESIGN UND ABLAUF | 30 |
| 3.3.1 | STATISTISCHE HYPOTHESEN | 34 |
| 3.4 | SCORING | 35 |
| 3.5 | ANMERKUNGEN | 37 |
| 4 | <u>RESULTATE</u> | 38 |
| 4.1 | DESKRIPTIVE AUSWERTUNG | 39 |
| 4.1.1 | FEHLERTYPENANALYSE | 41 |
| 4.2 | STATISTISCHE PRÜFUNG DER HYPOTHESEN | 48 |
| 5 | <u>DISKUSSION</u> | 49 |
| 6 | <u>ZUSAMMENFASSUNG</u> | 56 |
| 6.1 | DEUTSCH | 56 |
| 6.2 | ENGLISCH | 57 |
| 7 | <u>LITERATUR</u> | 58 |

1 Einführung

Unser Alltag ist von einer Vielzahl von auf uns einströmenden Reizen geprägt. Um Ordnung in dieses Chaos zu bringen, ist der Mensch bestrebt, die Informationsflut, mit der wir täglich konfrontiert sind, durch Kategorisierung überschaubarer zu gestalten. Diese Einteilung hilft uns, für den derzeitigen Moment unwichtige Bestandteile von Reizen herauszufiltern, was wiederum mehr Ressourcen für die weitere Verarbeitung des Wesentlichen freilässt. Durch vorangegangene Erfahrungen bilden wir Kombinationen von typischen Hintergrund-Objekt-Konstellationen und speichern diese ab, um sie bei späterer Aktivierung parat zu haben. Sehen wir also beispielsweise eine Bowlinghalle, stufen wir das Vorhandensein von Bowlingkugeln oder Pins als zu der Szene passend ein, wohingegen wir einen Motorradhelm oder einen Reifen als inkonsistent mit dem Kontext beurteilen würden. Dieses Zusammenspiel einer Szene (Hintergrund) mit den darin enthaltenen Objekten hilft uns also, schnell Wichtiges von Unwichtigem zu trennen und somit die Entscheidungszeit kurz zu halten.

Untersuchungen zur Bedeutung des Kontext bei der Objekt- und Szenenerkennung beschäftigen sich mit der Frage, inwieweit sich unser Wissen über plausible Kombinationen von Objekt und Szene auf unsere Wahrnehmung auswirkt. Wird eine Szene erkannt, führt dies zu einer Eingrenzung von Objekten, deren Auftreten wir in diesem Zusammenhang für wahrscheinlich halten. Des Weiteren ist von Interesse, ob sich durch Manipulation der Objekte hinsichtlich ihrer Eigenschaften deren Erkennung beeinflussen lässt und welche Rolle dabei die Szene spielt, in welche diese Objekte eingebettet sind. Objekte kommen im Alltag selten ohne Umgebung vor, sie stehen also meist in Verbindung zur ihrer Umwelt. In der vorliegenden Untersuchung wurde demzufolge versucht, Objekte hinsichtlich ihrer Erscheinung zu manipulieren, um anschließend festzustellen, ob und wie sich diese Veränderung auf das Zusammenspiel von Objekt und Szene auswirkt. In diesem Zusammenhang ist von besonderem Interesse, inwieweit sich Versuchspersonen durch die

Bearbeitung der Objekte (erschwerter Erkennbarkeit) zu Rateurteilen auf Basis vorangegangener Erfahrungen hinreißen lassen.

2 Stand des Wissens

Die Fragen, ob und wie die Erkennung einer gezeigten Szene die Objektwahrnehmung beeinflusst, wurden schon früh untersucht. Wird uns beispielsweise das Bild einer Badezimmerzene oder eines Bauernhofes gezeigt, führt dies zu entsprechenden Erwartungen hinsichtlich wahrscheinlicher Objekte, die wir in diesen Szenen antreffen könnten (wie etwa eine Zahnbürste in der ersten Szene oder einen Heuballen in der zweiten). Werden die Objekte jedoch vertauscht präsentiert, widerspricht dies unseren Erwartungen und die Identifikation fällt uns schwerer. Im Folgenden wird versucht, einen kleinen Überblick über die verschiedenen Ansätze zu bieten, mit denen die Zusammenhänge zwischen Objekten und Szenen erforscht wurden. Dabei werden sowohl unterschiedliche Verfahren als auch Forschungsinteressen näher beschrieben.

Biederman (1972) zeigte in seiner Studie zur Objekterkennung schon früh, dass seine Versuchspersonen einerseits zu einer raschen Szenenidentifikation fähig waren, andererseits Schwierigkeiten hatten, Objekte zu erkennen, sobald der Sinn der Szene durch Bildmanipulation gestört wurde. Er zeigte somit, dass Objekterkennung bei kohärenten Szenen mit höherer Wahrscheinlichkeit und präziser auftrat, als bei Szenen, deren Aussagekraft eingeschränkt wurde. Palmer (1975) hingegen wählte einen etwas anderen Zugang und präsentierte seinen Versuchspersonen Zeichnungen von Szenen, bevor er ihnen Objekte zur Identifikation vorlegte. Jene Objekte, die in Zusammenhang mit der gezeigten Szene standen, wurden daraufhin genauer beschrieben als jene, die in keiner Beziehung zur Zeichnung standen. Allerdings wurden in dieser Studie Objekte und Szene nicht gleichzeitig gezeigt, weshalb die Erkenntnisse nicht auf Szenen umgelegt werden können, welche das Objekt direkt enthalten. Die Erkenntnis, dass Objekte, die der Bedeutung einer Szene widersprechen,

mit größerer Schwierigkeit erkannt werden als jene, die dieser Bedeutung zuträglich sind, konnte jedoch nicht in allen Studien zu diesem Thema bestätigt werden (siehe Hollingworth und Henderson, 1998).

Auf ähnliche Weise versuchten Davenport und Potter (2004) die Beziehung zwischen Szene und Objekt zu erläutern. Sie präsentierten ihren Versuchspersonen Farbfotografien von Szenen mit eingefügten Objektbildern. Dabei zeigte sich, dass Objekte in konsistenten Szenen genauer identifiziert wurden als in inkonsistenten Szenen und dass umgekehrt auch Szenen präziser wahrgenommen wurden, wenn sie ein konsistentes Objekt enthielten als wenn sie ein inkonsistentes enthielten. Das daraus abgeleitete interaktive Modell von Davenport und Potter (2004) beschreibt somit eine ganzheitliche Erfassung der Szene, bei der semantische Information bezüglich der Konsistenz der Reize bei der schnellen Erfassung der Szene verfügbar wird und daraufhin sowohl die Hintergrund- als auch die Objekterwahrnehmung beeinflusst. Objekte und Szenen werden in diesem Modell somit als in Interaktion stehend und nicht voneinander isoliert betrachtet und dementsprechend verarbeitet (Davenport und Potter, 2004). Die Szene wird ihnen zufolge dabei jedoch nicht unbedingt zuerst verarbeitet: Objekte können genauso als kontextschaffend für die Szene angesehen werden wie umgekehrt der Hintergrund für die Vordergrundobjekte. Somit beeinflussen beide unsere Wahrnehmung (Davenport und Potter, 2004).

In weiteren Studien zur Beziehung zwischen Objekten und Hintergründen wollte Davenport (2007) herausfinden, inwieweit sich die Anwesenheit von mehreren verwandten Objekten auf den Objekt-Szenen-Konsistenzeffekt (auch Konsistenzeffekt oder Szenenkonsistenzeffekt) auswirkt. Dieser besagt laut Davenport (2007), dass Objekte genauer beschrieben werden können, wenn sie in dazu passenden (konsistenten) Szenen präsentiert werden und dass auch umgekehrt die Szenenidentifikation von darin enthaltenen konsistenten Objekten profitiert. Sie nahm an, dass der Effekt bei Präsentation mehrerer Objekte eventuell abnehmen könnte, da die Auswirkungen des Hintergrundes bei nur einem anwesenden Objekt

wirkungsvoller wären, da die Aufmerksamkeitsanforderung weniger anspruchsvoll sei, als bei mehreren präsentierten Objekten. Dieses Hinzufügen eines Objektes würde demnach zu einem Anstieg der Verarbeitungsprozesse führen und den Einfluss des Hintergrundes verringern, da weniger Ressourcen zur Verfügung stünden. Wäre das interaktive Modell jedoch korrekt, sollte die Wahrnehmung eines Objektes auch bei Anwesenheit eines weiteren Objektes durch die Umgebung beeinflusst werden (Davenport, 2007).

Weiters überprüfte Davenport (2007), ob sich in Beziehung stehende Objekte auf die Szenenwahrnehmung auswirken. Die Ergebnisse zeigten, dass Objekte genauer beschrieben wurden, wenn sie in konsistenter Umgebung präsentiert wurden, als wenn sie sich in inkonsistenter befanden. Die Anwesenheit eines weiteren verwandten Objektes wirkte sich dabei nicht negativ auf den Szenenkonsistenzeffekt aus und die Genauigkeit der Beschreibung von zwei Objekten unterschied sich nicht von der bei Präsentation einzelner (Davenport, 2007). In einem zweiten Experiment fand Davenport (2007) einen deutlichen Einfluss des Vordergrundobjektes/ der Vordergrundobjekte auf die Hintergrundwahrnehmung, allerdings zeigte sich bezüglich der Anzahl der Objekte kein Unterschied bei der Beschreibung des Settings. Bei näherer Betrachtung des Versuchsdesigns und der Wertung der Antworten fielen vor allem die sehr kurze Darbietungszeit der Objekte (80ms) sowie der Umgang mit eventuellen Rateeffekten auf. Erriet eine Versuchsperson das konsistente Objekt, obwohl gerade ein inkonsistentes präsentiert worden war (beispielsweise wurde gerade ein Schwein in arktischer Umgebung präsentiert und der Teilnehmer antwortete mit "Iglu" (Davenport, 2007)), so wurde dies als umgebungsbedingtes Raten gewertet und es wurde eine korrekte konsistente Antwort bei dieser Person abgezogen. Um diese Notlösung zu umgehen, wurden im vorliegenden Experiment Fehlerkategorien eingeführt, welche teilweise dieselbe Problematik des Ratens behandelten. Auf die Unterschiede in Aufbau und Auswertung dieser und Davenports (2007) Arbeiten wird im Verlauf der Diskussion näher eingegangen, da sich hier

einige Parallelen aber auch große Unterschiede in der Vorgabe und im Scoring finden ließen, die es zu klären gilt.

Chaigenau, Barsalou und Zamani (2009) beschäftigten sich mit der Rolle, die Situationsfaktoren bei der Objektkategorisierung spielen. Dazu präsentierten sie ihren Versuchspersonen funktionale Objekte, die Hebeln, Gewichten oder Griffen ähnelten und ermittelten anschließend die Antworten der Teilnehmer. Wurde keinerlei situationsbedingte Informationen geboten, beschränkten sich die Beschreibungen der Teilnehmer vorwiegend auf physikalische Eigenschaften der Objekte. Erhielten sie jedoch Informationen bezüglich der Situation (Kontext), erweiterten die Versuchspersonen ihre Objekteinschätzung auf ihnen bekannte Objekte (sie stellten also Vergleiche an) und aktivierten abgespeicherte Objektkategorien häufiger und schneller (Chaigenau, Barsalou und Zamani (2009)). Dies ist vor allem in Hinblick auf mögliche Rateeffekte (und Ergänzungen der Teilnehmer über das Gesehene hinaus) besonders interessant, da man erkennt, welchen starken Einfluss der Kontext auf das abgegebene Urteil hat.

Mudrik, Deouell und Lamy (2011) verwendeten ein Verfahren namens Binocular Rivalry (BR), bei dem zwei verschiedene Stimuli gleichzeitig präsentiert werden, allerdings getrennt für jedes Auge. Dabei wird jeweils ein Bild zur Gänze aufgenommen, während das andere von der bewussten Verarbeitung ausgeschlossen wird. Hohe Dominanz (ein Bild wird zugunsten des anderen ausgeblendet) steht dabei für aufmerksame Zuwendung und bewusste Verarbeitung. Mithilfe dieser Technik wollten sie feststellen, inwieweit kongruente und inkongruente Objekte Aufmerksamkeit beanspruchen. Dazu wurden Szenenbilder präsentiert, wobei einem Auge ein kongruentes Objekt, dem anderen ein inkongruentes gezeigt wurde. Die Szene blieb dabei für beide Augen dieselbe. Anschließend mussten die Teilnehmer mittels Knopfdruck angeben, welches Objekt sie wahrgenommen hatten. Die Ergebnisse führten laut Mudrik, Deouell und Lamy (2001) zu dem Schluss, dass inkongruente Objekte zu einer längeren Verweildauer der Aufmerksamkeit führen, diese jedoch vermutlich nicht von den Objekten angezogen wird. Warum dies so ist, könnten frühere

Studien zum Blickverhalten von Loftus und Mackworth (1978) beantworten. Erste Untersuchungen des Blickverhaltens von Mackworth und Morandi (1967), welche vermehrt Fixationen auf Regionen zeigten, die den Teilnehmern als subjektiv informativ erschienen, wurden in weiterer Folge von Friedman (1979) sowie Loftus und Mackworth (1978) ergänzt, welche hierzu das Blickverhalten der Versuchspersonen aufzeichneten, um aus den Ergebnissen Rückschlüsse auf die Wahrnehmung und Verarbeitung der gezeigten Objekte und Szenen zu ziehen. Dabei zeigten sich frühere, häufigere und längere Fixationen bei jenen Objekten, welche eine geringere Wahrscheinlichkeit hatten, in einer Szene zu erscheinen (die also den Erwartungen widersprachen). Loftus und Mackworth (1978) nahmen an, dass die längere Betrachtungszeit ein Hinweis auf einen höheren Verarbeitungsaufwand sei und dass die Reihenfolge der Betrachtung vom Informationsgehalt der Objekte abhängig wäre. Sie schlossen daraus, dass Objekte, die Aufklärung über die Bedeutung eines Bildes bieten, früher fixiert werden als uninformative. Als informativ wurde ein Objekt dann bezeichnet, wenn es in Anbetracht des übrigen Bildes eine geringe Wahrscheinlichkeit hatte, darin vorzukommen (Loftus, 1976; Loftus und Bell, 1975). Als Beispiel nannten Loftus und Mackworth (1978) eine Farmszene, welche einmal mit einem Traktor und ein anderes Mal mit einem Oktopus kombiniert wurde. Da die Wahrscheinlichkeit für das Vorhandensein eines Oktopus hier als äußerst niedrig angenommen werden kann, gilt er im Sinne der oben genannten Definition als informatives Objekt. Die Ergebnisse zeigten, dass informative Objekte tatsächlich früher gefunden und öfter und länger angesehen wurden. Die dahinterliegende Annahme war, dass zuerst die Szene in ihrer Bedeutung erfasst wird und dann die Objekte in der Peripherie auf Verträglichkeit mit der Szene geprüft werden. Diese Schritte lenken die Augenbewegungen in Richtung der weniger wahrscheinlichen Objekte (Loftus und Mackworth (1978)).

Underwood und Foulsham (2006) konnten diese Annahmen bestätigen, auch sie berichten von früheren Fixationen bei inkonsistenten Objekten. Sie verwendeten hierzu Fotografien, welche deplatzierte Objekte enthielten.

Dennoch ist es laut Davenport (2005) unklar, was längere Fixationszeiten bei inkonsistenten Objekten über die dahinterliegenden Verarbeitungsprozesse aussagen, da typische Fixationen rund 300ms andauern, der Bedeutungsgehalt einer Szene jedoch viel schneller erfasst werden kann, nämlich in rund 100ms (Intraub, 1981). So könnte eine längere Betrachtungszeit laut Davenport (2005) auf mehrere Arten von Arbeitsaufwand hindeuten: Einerseits wäre es denkbar, dass die Identifizierung des Objektes Zeit in Anspruch nimmt, andererseits könnte die Integration des Objektes in die Szene Ressourcen aufbrauchen. Schließlich wäre auch denkbar, dass eine längere Verweildauer bei einem bestimmten Reiz auch einfach erhöhtes Interesse an einer spezifischen Region abbilden könnte (Davenport, 2005).

Aufgrund zahlreicher Studien liegt die Schlussfolgerung nahe, dass Objekte in typischen Settings genauer beschrieben und eher erkannt werden, als Objekte in unwahrscheinlichen Settings (Boyce und Pollatsek, 1992; Boyce, Pollatsek und Rayner, 1989). Hollingworth und Henderson (1998) jedoch zeigten ihren Versuchspersonen Linienzeichnungen von Szenen, worauf diese entscheiden mussten, welches von zwei Objekten in diesen zu sehen war. Dabei stellten sie keine Unterschiede in der Genauigkeit der Beschreibung der Objekte fest, egal ob diese mit konsistenten oder inkonsistenten Szenen kombiniert präsentiert wurden. Dies führte Hollingworth und Henderson (1998) zu der Annahme, dass Objekte und Szenen unabhängig voneinander verarbeitet werden.

De Graef, Christiaens und d'Ydewalle (1990) wiesen darauf hin, dass die Effekte der Erwartungsverletzungen bei präsentierten Objekten erst im Zuge der genaueren Szenenanalyse auftraten und nicht schon bei ersten Fixationen. Henderson, Weeks, und Hollingworth (1999) zeichneten daher das Blickverhalten bei der Inspektion von Linienzeichnungen auf, welche teilweise ungewöhnliche Objekte enthielten. Sie konnten jedoch keine früheren Fixationen feststellen. Eine Möglichkeit für diese widersprüchlichen Ergebnisse könnte die Darbietung der Stimuli darstellen, welche bei Loftus und Mackworth (1978) kontrastreicher und auffälliger

ausfielen als bei De Graef, Christiaens und d'Ydewalle (1990) und Henderson, Weeks und Hollingworth (1999) und somit die Aufmerksamkeit der Versuchspersonen eventuell eher anzogen. Um zu klären inwiefern die Auffälligkeit des Reizes mit dessen früherer Betrachtung zusammenhängt, präsentierten Underwood et al. (2008) ihren Versuchsteilnehmern Bilderpaare, von denen sie entscheiden mussten, ob diese gleich oder verschieden waren. Die Bilder enthielten Hintergrundobjekte, von denen in einer Bedingung eines durch ein entweder konsistentes oder inkonsistentes Objekt ersetzt wurde. Die Salienzwerte (welche der Bestimmung der Auffälligkeit eines Reizes dienen) wurden in Anlehnung an Itti und Koch (2000) festgelegt. Sie folgerten aus ihren Ergebnissen, dass inkongruente Objekte, die anhand der Szenenwahrnehmung nicht zu erwarten sind, eher Aufmerksamkeit anziehen als kongruente und dass diese Bevorzugung nicht von deren visueller Auffälligkeit abhängt (Underwood et al., 2008).

Zu ähnlichen Ergebnissen kamen auch Bonitz und Gordon (2008), die natürliche Szenen präsentierten und von früheren Fixationen bei semantisch inkonsistenten Objekten berichteten. Allerdings kam es bei diesen Studien laut Vo und Henderson (2011) bereits vor diesen ersten Fixationen von inkonsistenten Objekten zu Augenbewegungen. Daraus schlossen sie, dass unsere Blickrichtung nicht vorrangig von diesen Objekten angezogen wird, sondern dass inkonsistente Objekte bloß etwas früher Beachtung finden als konsistente. Dieses Phänomen könnte laut Vo und Henderson (2011) durch Fixationen in unmittelbarer Umgebung der inkonsistenten Objekte entstehen. Vo und Henderson (2011) versuchten in ihren Experimenten eine isolierte Betrachtung der ersten Szenenverarbeitung und der darauffolgenden weiteren visuellen Verarbeitung zu erzielen, um zu klären, ob auftretende Effekte auf diese erste Verarbeitung oder spätere Zufallsfixationen (die ohne Intention im Laufe der Szenenexploration in die Nähe eines Zielobjekts fallen) innerhalb der Szene zurückzuführen sind. Damit einhergehend stellte sich für sie die Frage, ob inkonsistente Objekte am Rande des Sichtfeldes verarbeitet werden können und somit unser Blickverhalten beeinflussen. Ihre Erkenntnisse führten zu

der Annahme, dass es uns nicht möglich ist, Objekt-Szene-Inkonsistenzen in der visuellen Peripherie wahrzunehmen und entsprechend darauf zu reagieren (Vo und Henderson, 2011). Entgegengesetzte Erkenntnisse aus Studien von Underwood und Foulsham (2006) und Bonitz und Gordon (2008) mit früheren Fixationen von peripheren inkonsistenten Objekten führten Vo und Henderson (2011) auf die unterschiedliche Szenenverarbeitung zurück. Inkongruente Objekte wurden demnach erst nach etlichen Fixationen oder nach Verstreichen einiger Sekunden präferiert, es kam also zu einer späten lokalen Verarbeitung der Szene. Somit kann gesagt werden, dass Inkonsistenzen zwischen Objekt und Szene unsere Augenbewegungen nicht vor ersten Fixationen beeinflussen können (Vo und Henderson, 2011).

Bornstein und Mash (2011) beschäftigten sich mit dem Blickverhalten von Kindern und Erwachsenen und richteten dabei besonderes Augenmerk auf die Einflüsse von Objekt-Kontext-Beziehungen. Sie vertraten in ihren Untersuchungen die Meinung, dass dieses natürliche Auftreten von Objekten gemeinsam mit deren Hintergründen den Entwicklungsrahmen für die Evolution und individuelle Entwicklung unseres visuellen Systems und unserer visuellen Wahrnehmung bot und immer noch bietet. Unser Wissen (und dessen Erwerb) über Objekte basiere ihnen zufolge somit auf Lernprozessen über deren typischen, wahrscheinlichen oder erwarteten Auftretenskontext (Bornstein und Mash, 2011). Sie nahmen weiters an, dass sich Kinder und Erwachsene gleichermaßen durch Objekt-Szene-Konsistenzerwartungen in ihrem Urteil beeinflussen lassen würden und dass Kinder bei Erwartungsverletzungen den Kontext verstärkt in den Verarbeitungsprozess einbeziehen würden. Sie präsentierten ihren Teilnehmern Tier- und Fahrzeugbilder, welche in der kongruenten Bedingung mit Naturszenen beziehungsweise städtischen Umgebungen kombiniert wurden und umgekehrt in der inkongruenten. Ähnlich wie Vo und Henderson (2011) gingen Bornstein und Mash, (2011) von ersten Orientierungsfixationen aus, die nicht notwendigerweise die Szenenkonfiguration betrafen. Anschließend konzentrierten sich die Augenbewegungen laut

Bornstein und Mash (2011) auf informative Regionen, wobei hier vor allem die Fovea aufgrund ihrer hohen Auflösung und somit hinsichtlich der Informationsbeschaffung als besonders bedeutend angenommen wurde (Hochberg, 1978). In ihren Ergebnissen zeigte sich, dass Erwachsene bei der Betrachtung Objekte gegenüber Szenen bevorzugten, diese also öfter fixierten und dass sie kongruente Szenen öfter betrachteten als inkongruente. Bornstein und Mash (2011) führten dies auf die weiter fortgeschrittene Entwicklung von Erwachsenen zurück, welche eher in der Lage wären, Objekte von ihrem Hintergrund zu extrahieren, als Kinder. Erwachsene zeigten jedoch unflexiblere Fixationswechsel innerhalb inkongruenter Szenen, was die Autoren auf eine Verlangsamung durch stärkere Objektbindung (aufgrund der höher ausgeprägten Extraktionsfähigkeit) sowie auf verzögernde top-down-Prozesse (Analyse-richtung vom Allgemeinen zum Speziellen, dabei werden die inkongruenten Inhalte zu verifizieren versucht) zurückführten (Bornstein und Mash, 2011).

Castelhana und Heaven (2010) wollten feststellen, inwieweit der Szenenkontext (welchen sie in generelle semantische und spezifische visuelle Details unterteilten) unsere Suche nach den Zielobjekten beeinflusst und welche Rolle deren Eigenschaften dabei spielen. Den semantischen Teil bezeichneten die Autoren auch als den Hauptinhalt der Szene, welcher Aufschluss über mögliche Varianten des Zielobjektes sowie dessen wahrscheinlichen Aufenthaltsort gibt. Möglicherweise noch wichtiger für die Aufmerksamkeitsverteilung als der semantische Anteil könnten laut Castelhana und Heaven (2010) jedoch die visuellen Details der Szene für die Targetlokalisierung sein. Darunter verstanden sie erlernte Anhaltspunkte wie den Standort des Betrachters oder den Maßstab, welche Aufschluss darüber geben, wo die Zielreize aller Wahrscheinlichkeit nach auftreten werden. Das Ziel von Castelhanos und Heavens (2010) Experimenten war es also, die Auswirkungen verschiedener top-down-Einflüsse auf das Suchverhalten zu klären: einerseits unser Wissen über Szenenkontexte und andererseits über die Eigenschaften des Zielreizes. Es zeigte sich, dass visuelle Hinweisreize in Form von Bildern die Suchzeit stärker verkürzten,

als Hinweise in Form von Namen. Demzufolge waren in Castelhanos und Heavens (2010) Untersuchungen die visuellen Details der Szene ausschlaggebender für eine erfolgreiche Suche als semantische Information. Weiters zeigte sich keine Art von Interaktion zwischen den zwei Formen der top-down-Einflüsse, demnach waren diese eher als einander ergänzend anzusehen (Castelhana und Heaven, 2010). Interessanterweise konnten die Autoren keinen positiven Einfluss des Hauptinhaltes der Szene auf die Suchzeit feststellen, diese war nur von den visuellen Details der Szene abhängig. Als Erklärung dafür sahen die Autoren die Notwendigkeit, sich in einer Szene zuerst grob orientieren zu müssen, bevor semantische Information überhaupt genutzt werden konnte. Genau dafür waren die visuellen Details von entscheidender Bedeutung. Castelhana und Heaven (2010) betonten weiters, dass keine Interaktion zwischen dem Szenenkontext und den Objekteigenschaften stattfand, der Kontext also nicht als limitierend für die Eigenschaftssuche angesehen werden konnte und umgekehrt die Objekteigenschaften den Einfluss der Szene nicht beschränkten. Es zeigte sich jedoch, dass beide Informationstypen des Szenenkontexts die Suche stärker verbesserten, als jeweils eine alleine und auch, dass die Objekteigenschaften das Suchverhalten beeinflussten (Castelhana und Heaven, 2010).

Auf neurowissenschaftlicher Ebene untersuchten Ganis und Kutas (2003), zu welchem Zeitpunkt Szeneneffekte Wirkung auf die Objektwahrnehmung zeigten. Mittels ereigniskorrelierter Potentiale und anhand der Messung von Reaktionszeiten sollte das Verhalten der Versuchspersonen bei der Darbietung von Szenen in Kombination mit kongruenten und inkongruenten Objekten näher erläutert werden. Dabei zeigten sich kürzere Reaktionszeiten bei kongruenten Objekten, sowie erste Auswirkungen durch die Darbietung der Szene nach etwa 300ms (Ganis und Kutas, 2003). Vor diesen 300ms zeigen sich keine Einflüsse durch den semantischen Gehalt der gezeigten Szene. Erst danach erfolgte den Autoren zufolge die Aktivierung von semantischen Repräsentationen der Objekte, ausgelöst durch die Darbietung der Szene. Um die frühesten Interaktionen zwischen

Kontext- und Objektverarbeitung näher zu untersuchen, präsentierten Joubert et al. (2008) ihren Versuchspersonen Tiermotive, welche mit kongruenten und inkongruenten Kontexten kombiniert wurden. Sie kamen zu der Erkenntnis, dass Objekte und ihr Kontext parallel verarbeitet wurden und während dieser Verarbeitung in ständigem Austausch standen. Die Untersuchungen von Joubert et al. (2008) unterstützten die Annahmen des *perceptual schema models* (Henderson und Hollingworth, 1999), welches die Interaktion von Kontext und Objekt während deren Verarbeitung beschreibt.

Martens, Trujillo-Barreto und Gruber (2011) beschäftigten sich mit der Problematik, Hirnaktivitäten anhand ihres Auslösers einzuteilen. Ihnen zufolge kann bei Verwendung von Blutsauerstoffwerten und ereignis-korrelierten Potentialen nicht zwischen Objekt- und Szeneneinfluss unterschieden werden, da sich diese überlappen und somit eine Antwort aufgrund eines Gesamteindrucks abgeben würden. Als Lösungsansatz für diesen Kritikpunkt versuchten Martens, Trujillo-Barreto und Gruber, (2011), die neuronale Aktivität ihrer Versuchspersonen getrennt nach Objekt und Hintergrund aufzuzeichnen. Sie präsentierten dazu semantisch konsistente oder inkonsistente Objekte und ermittelten die Hirnaktivität mittels EEG. Die Ergebnisse ließen die Autoren darauf schließen, dass die visuelle Aufmerksamkeit der Versuchsteilnehmer auf den Hintergrund gerichtet wurde, sobald Objekt und Hintergrund in semantischer Beziehung standen. Demzufolge wurde das Objekt offenbar als Teil der Szene wahrgenommen und somit nicht näher betrachtet. Im Gegensatz dazu wurde bei semantisch abweichenden Objekten die Aufmerksamkeit zur Gänze von diesen in Anspruch genommen (Martens, Trujillo-Barreto und Gruber, 2011). Diese Erkenntnisse decken sich mit Studien zum Blickverhalten, bei denen inkonsistente Objekte vermehrt Aufmerksamkeit zugeteilt bekamen (Underwood et al., 2008).

Martens, Trujillo-Barreto und Gruber (2011) zeigten, dass die Aufmerksamkeit bei inkonsistenten Objekt-Szene-Kombinationen vom Hintergrund zum unpassenden Objekt umgelenkt wurde. Aufgrund der Einschränk-

ungen des Verfahrens konnten jedoch erst Aufmerksamkeitsprozesse ab 800ms nach Präsentation des Reizes ermittelt werden, was frühere Reaktionen leider unbeachtet ließ (Martens, Trujillo-Barreto und Gruber (2011)).

Mudrik, Lamy und Deouell (2010) untersuchten mithilfe von ereignis-korrelierten Potentialen den Einfluss von Kontextverletzungen auf die Szenenverarbeitung. Dazu wurden die Aufzeichnungen, die durch die Präsentation von kongruenten und inkongruenten Szenen hervorgerufen wurden, miteinander verglichen. Um zu gewährleisten, dass die Versuchspersonen nicht schon im Vorfeld eine Erwartungstendenz in Richtung eines bestimmten Objektes aufbauen konnten (was dann nicht nur zu einer Kontextverletzung sondern auch zu einer Erwartungsverletzung geführt hätte), wurden die Szenen gleichzeitig mit den Objekten präsentiert. Die Teilnehmer sahen Personen die eine Handlung durchführten, wobei in einer Bedingung ein inkongruentes und in der anderen Bedingung ein kongruentes Objekt verwendet wurde. Die Versuchspersonen mussten nun feststellen, wie viele Hände die gezeigte Person für ihre Handlung verwendete. Die gewonnenen Ergebnisse führten Mudrik, Lamy und Deouell (2010) zu dem Schluss, dass Kontextverletzungen in einer gezeigten Szene zu Verarbeitungsprozessen führten, ähnlich denen bei semantischer Aktivierung durch erkannte Objekte. Diese Prozesse wären laut den Autoren von den bisherigen gefundenen Abgleichen zwischen Erwartungen und Realität zu trennen, da die Szenen im vorliegenden Experiment gleichzeitig mit den Objekten präsentiert wurden. Mudrik, Lamy und Deouell (2010) nahmen daher eine Interaktion zwischen Szenen- und Objektverarbeitung und einen Einfluss der Kontextinformation auf die Objektselektion und auf weitere Stufen der semantischen Wissensaktivierung an.

Sampanes, Tseng und Bridgeman (2008) versuchten in ihrer Untersuchung festzustellen, welche Bedeutung der Inhalt einer Szene für die weitere Verarbeitung hatte. Dabei fanden sie heraus, dass der Inhalt einer Szene getrennt von den ihn konstituierenden Spezifika kodiert wurde. Die Haupt-

aussage einer Szene wurde somit zwar aus einzelnen Details derselben gebildet, sobald diese (Hauptaussage) jedoch einmal erkannt wurde, war sie offenbar unabhängig von diesen Einzelheiten. Sampanes, Tseng und Bridgeman (2008) erkannten, dass kleinere Veränderungen eher bemerkt wurden, wenn sie den Hauptinhalt der Szene betrafen, als Änderungen des selben Ausmaßes, die jedoch unabhängig von der Hauptaussage der Szene waren. Um zu klären, ob der Szenenkontext stets das Abrufen visueller Objektrepräsentationen erleichterte, wählten Nakashima und Yokosawa (2011) einen Versuchsaufbau, der die Gedächtnisleistungen ihrer Versuchspersonen bei der Präsentation von lokal begrenzten Szenenveränderungen, welche keine serielle ganzheitliche Erfassung der Szene voraussetzten, untersuchbar machte. Es zeigte sich, dass Objektrepräsentationen nicht immer an den Szenenkontext gebunden waren, sondern nur dann, wenn die Aufgabenstellung explizit lautete, viel Information beizubehalten (Nakashima und Yokosawa, 2011).

Cohen, Alvarez und Nakayama (2011) versuchten zu klären, inwieweit Aufmerksamkeit eine Voraussetzung für die bewusste Szenenwahrnehmung darstellte. Dazu präsentierten sie ihren Versuchsteilnehmern identische Scheiben, welche sich bewegten, während sich Bilder im Hintergrund rasch veränderten. In einer zweiten Aufgabe zeigten sie ihnen Buchstaben und Zahlen, wieder vor einem sich ändernden Hintergrund. Diese Hintergrundbilder bestanden aus natürlichen Szenen und Masken. Die Ergebnisse legten laut Cohen, Alvarez und Nakayama (2011) den Schluss nahe, dass die Verarbeitung von Szenen Aufmerksamkeit benötigte. Die Teilnehmer zeigten Schwierigkeiten bei der Szenenerkennung bei doppelter Aufgabenbelastung, auch wenn die Erkennungsrate deutlich über dem Zufall lag. Dafür könnten laut Cohen, Alvarez und Nakayama (2011) sowohl das zu niedrig angesetzte Schwierigkeitsniveau als auch eventuell nicht aufmerksamkeitspflichtige Anteile der Szenenerkennung verantwortlich gewesen sein. Weiters mutmaßten die Autoren, dass die unbewusste Verarbeitung der Szenen möglicherweise ohne Aufmerksamkeit funktionieren könnte, Aufmerksamkeit jedoch notwendig wäre, um

natürliche Szenen bewusst wahrnehmen zu können (Cohen, Alvarez und Nakayama, 2011).

Um einen aktuellen Einblick in den Stand des Wissens bezüglich Szenenwahrnehmung zu gewähren, sei hier noch Intraubs (2012) Beschäftigung mit einigen Modellen der Verarbeitung von Kontexten erwähnt. Intraub (2012) diskutierte deren Vor- und Nachteile und beschrieb die Vorgänge, die stattfinden müssten, um aus der Vielzahl von auf uns einströmenden Reizen das Bild einer ganzheitlichen Szene zu erschaffen. Dabei kristallisierten sich vor allem zwei konkurrierende Modelle heraus: das *traditional visual-cognitive model* und das *multisource model*. Ersteres vertritt laut Intraub (2012) die Meinung, dass die Repräsentation einer Szene einzig und allein auf den visuellen Input zurückzuführen ist. Das *multisource model* hingegen nutzt weniger visuelle als räumliche Informationen, um Szenenwahrnehmung zu gewährleisten. Es geht von einer egozentrischen Betrachtungsweise des Raums von Seiten des Beobachters aus (der selbst Teil der Szene ist und Objekte und deren Lage hinsichtlich seiner eigenen Position lokalisiert), der eine Simulation der Welt betrachtet, welche nur einen Ausschnitt der Realität abbildet. Dieser Ausschnitt wird in weiterer Folge durch amodale Wahrnehmung (Ergänzungen, die über das Gezeigte hinausgehen), konzeptuelles Wissen (die Versuchsperson erkennt die Szene und aktiviert damit assoziierte, abgespeicherte Inhalte - dieses Wissen steuert das weitere Blickverhalten, es bestimmt die Orte für weitere Fixationen) und Objekt-Kontext-Erwartungen (die Teilnehmer schließen von vorhandenen Objekten auf mögliche Hintergründe) ergänzt und es kommt somit zu einer Grenzerweiterung (Intraub, 2012). Ab der ersten Fixation wird eine Repräsentation aufgebaut, welche in weiterer Folge und im Zuge weiterer Fixationen graduell verfeinert und aktualisiert oder korrigiert werden kann. Das *traditional visual-cognitive model* bietet der Autorin zufolge demnach einen Einblick in die Aufnahme der visuellen Information, wird jedoch der Fülle unserer Verarbeitungsprozesse nicht ganz gerecht. Hier zeigt sich laut Intraub (2012) das *multisource model* als interessante Alternative, da es

multimodale und höhere kognitive Prozesse der Szenenwahrnehmung einschließt und so einen umfassenderen Abdruck unserer täglichen Eindrücke abbildet. Beide Modelle vereint nach Ansicht der Autorin die Erkenntnis, dass Szenen in äußerst kurzer Zeit identifiziert werden können. Intraub (2012) kritisierte jedoch Ergebnisse aus Verhaltensstudien, da diese keinen Einblick in die Prozesse zwischen Reizpräsentation und Antwort liefern würden und es somit unklar wäre, wann eine Klassifizierung tatsächlich stattfindet. Mittels ereigniskorrelierter Potentiale konnte jedoch festgestellt werden, dass die erste Fixation einer Szene ausreichend für deren Identifizierung ist (Intraub, 2012).

2.1 Erklärungsmodelle Konsistenzeffekt

Um Verletzungen der erwarteten Kontext-Objekt-Beziehung zu identifizieren, müssen nicht nur Hintergrund und Objekt verarbeitet werden, sondern es muss auch eine semantische Komponente (wie wahrscheinlich ist es, dass dieses Objekt in diesem Kontext erscheint) berücksichtigt werden.

Generell kann laut Henderson und Hollingworth (1999) bei der Objekterkennung zwischen verschiedenen Stufen unterschieden werden: Zuerst wird der eingehende Reiz in grobe visuelle Kategorien übersetzt (beispielsweise Lokalisierung von Kanten und Flächen), anschließend werden diese Informationen zur genaueren Einteilung hinsichtlich der strukturellen Beschaffenheit des Reizes genutzt. Als letzter Schritt folgt ein Abgleich mit gespeicherten Gedächtnisinhalten, welcher, wenn positiv (match), zur Objektidentifikation führt. Anhand dieser Stufen können verschiedene Modelle der Objektidentifikation eingeteilt werden: Das *perceptual schema model* schlägt vor, dass sich abgespeicherte Gedächtnisinhalte hinsichtlich des Aufbaus einer Szene (Schemata) auf die nachfolgende Objekterkennung auswirken (Henderson und Hollingworth, 1999). Dadurch soll die Analyse schemakonsistenter Objekte erleichtert und die Aufarbeitung schemainkonsistenter Objekte erschwert werden, was sich auch positiv auf die Genauigkeit der Aufarbeitung auswirkt. Demnach

kommt es in einer frühen Phase der Reizverarbeitung zu einer Erschwerung der Analyse von schema-inkongruenten Objekten bei Kontextverletzungen und zu einer Verbesserung der Analyse schema-kongruenter Objekte (Henderson und Hollingworth, 1999).

Das *priming model* setzt an der Stelle des Abgleichs der visuellen Eindrücke mit den abgespeicherten Bildern an. Bei Aktivierung eines Szeneschemas kommt es zum Priming der gespeicherten schemakonsistenten Objektrepräsentationen (Henderson und Hollingworth, 1999). Das bedeutet, dass diese im Gedächtnis verankerten Abbilder in Zukunft weniger Information benötigen, um bei einem Abgleich als match ausgewählt zu werden. Anders als das *perceptual schema model* nimmt das *priming model* laut Henderson und Hollingworth (1999) an, dass sich die Erkennung der Szene insofern auswirkt, als dass bestimmte Objekte wahrscheinlicher anzutreffen sind als andere. Bei ersterem Modell hingegen wird ein direkter Einfluss der Szenenerkennung auf die Objektanalyse propagiert. Das dritte Modell, das *functional isolation model* (Henderson und Hollingworth, 1998), nimmt an, dass die Objekterkennung isoliert von den Einflüssen der Szene stattfindet, es beschreibt demnach eine Trennung von Wahrnehmung und Kognition. Dem Modell zufolge wird Kontextinformation erst nach der Objektidentifikation wirksam und beeinflusst dann die Aktivierung semantischen Wissens und die Entscheidungsfindung (Henderson und Hollingworth, 1999). Weiters wird im *functional isolation model* die Meinung vertreten, dass die Aufmerksamkeit nicht gleich anfangs von inkongruenten Objekten angezogen wird aber dass der Blick, sobald eine Fixation stattgefunden hat, dort verweilt (Henderson und Hollingworth, 1999).

2.2 Tiefpassinformationen und Top-Down-Prozesse

Studien zur Objekterkennung auf neuronaler Ebene beschäftigen sich hauptsächlich mit bottom-up-Prozessen, bei denen der visuelle Reiz in einer Abfolge von kortikalen Regionen analysiert wird (Bar, 2003). Dabei wird Information von ansteigender Komplexität verarbeitet. Die visuellen Eigenschaften eines Bildes werden zuerst in niedrigeren kortikalen Arealen

extrahiert um dann in höhere Areale projiziert zu werden. Dort entsteht dann eine visuelle Repräsentation des zuvor gesehenen Bildes (Tanaka, 1996). Anschließend erfolgt der Vergleich mit abgespeicherten Bildern, welche im Gedächtnis vorhanden sind. Bei einem erfolgreichen Abgleich kommt es zum Erkennen der Repräsentation (Bar, 2003). Somit sind die Informationen aus niedrigeren kortikalen Arealen Grundlage für die weitere Verarbeitung, das Modell ist also hierarchisch. Laut Ullmann (1995) tragen sowohl bottom-up als auch top-down-Prozesse einen bedeutenden Beitrag zur Objekterkennung bei. Laut seinem Modell werden durch das präsentierte Objekt bottom-up- und top-down-Abläufe aktiviert, welche parallel nach Abgleichen suchen. Zu einem Erkennen kommt es daher, wenn diese Prozesse aufeinandertreffen und einen positiven Abgleich erreichen (Bar, 2003). Bei top-down-Prozessen stellt sich klarerweise die Frage, wie diese ausgelöst werden können, bevor der Reiz analysiert wurde. Laut Bar (2004) sind dafür folgende Faktoren erforderlich, welche in seinem vorgestellten *model for contextual facilitation* näher erläutert werden:

Dem Modell liegt die Annahme zugrunde, dass Abbildungen, welche nur globale Tiefpassinformation aufweisen, ausreichend für eine schnelle Objekt- und Szenenerkennung sind. Bar (2004) ist der Meinung, dass diese Informationen die schnelle Erkennung ermöglichen, da sie Szenenschemata und Bezugsrahmen aktivieren, an welchen sich die weitere Verarbeitung orientiert. Als Bezugsrahmen definiert Bar (2004) einen Verband prototypischer Information über einen bestimmten Kontext, der Daten über die räumliche Anordnung sowie die Identität von Objekten, welche üblicherweise gemeinsam in diesem Kontext auftreten, beinhaltet. Diese Information führt zu bestimmten Erwartungen hinsichtlich der Umwelt, welche im Zuge des Abgleichs mit eintreffenden Reizen überprüft werden. Die Ortsfrequenz dient in diesem Modell der Analyse eines Stimulus, da sie wichtige Informationen über dessen Aufbau enthält. Sie setzt sich aus Hoch- und Tiefpassanteilen zusammen: Hochpassinformation (high spatial frequency) umfasst plötzliche Änderungen im Verlauf (beispielsweise Kanten) eines Bildes und feine Details. Tiefpassinformation (low spatial

frequency) hingegen beinhaltet umfassendere Aspekte wie Proportionen und räumliche Orientierung (Bar, 2003). Diese Tiefpassinformationen führen laut Bar (2003) zu sogenannten *initial guesses*, welche den Kreis potentieller Kandidaten (im Sinne eines positiven Abgleichs zwischen Input-Reiz und gespeichertem Bild) verkleinern. Dadurch können der kognitive Arbeitsaufwand und damit einhergehend auch die Bearbeitungszeit erheblich verringert werden. Kommt es zu einem Abgleich, werden die restlichen potentiellen Kandidaten unterdrückt (deutet der Kontextrahmen etwa in Richtung einer Bowlingbahn und sind die potentiellen Kandidaten eine Bowlingkugel und ein Motorradhelm, so wird der wahrscheinlichere der beiden Anwärter gewählt und der andere unterdrückt). Kommen beide Objekte für einen spezifischen Kontext in Frage, müssen weitere Informationen einbezogen werden, um einen sicheren Entschluss zu gewährleisten. Dies geschieht durch die langsamere aber genauere Hochpassinformation (Bar 2003).

Bar (2003) nimmt also an, dass die Vorlage eines Objektbildes, welches lediglich Tiefpassinformationen aufweist, genug Information bietet, um eine Auswahl an potentiellen Interpretationen des Gezeigten zu treffen und unwahrscheinliche Deutungen auszuklammern. Der Ablauf bei Präsentation der Reize gestaltet sich dann laut Bar (2004) wie folgt: Das präsentierte, tiefpassfrequente Bild wird sehr früh und auf schnellem Weg von den visuellen Arealen in den präfrontalen und parahippocampalen Kortex projiziert. Im parahippocampalen Kortex wird durch dieses Bild ein Bezugsrahmen erstellt, welcher auf bisherigen Erfahrungen beruht und für den Reiz als passend erscheint. Die erstellte Kontextinformation wird dann in den inferioren temporalen Kortex geleitet, in welchem es zu Assoziationen kommt, die thematisch diesen Rahmen betreffen. Parallel zu diesem Prozess wird das gefilterte Bild in den präfrontalen Kortex projiziert, allerdings liegt diesmal die Fixation auf dem hervorgehobenen Objekt, welches sich in der Szene befindet. Anschließend findet wieder eine Auswahl möglicher Kandidaten für das präsentierte Objekt statt. Im inferioren temporalen Kortex treffen dann diese zwei Prozesse (Analyse der

Szene mit Bildung möglicher Bezugsrahmen sowie Analyse des Objektes mit Auswahl potentieller Anwärter) aufeinander und bilden gemeinsam eine Schnittmenge, aus welcher ein einziger, wahrscheinlicher Kandidat hervorgeht, welcher Informationen aus beiden Strömungen bei der Entscheidungsfindung berücksichtigt. Im Anschluss an diese Auswahl werden die Eigenschaften des Objektes genauer definiert (das Objekt ist beispielsweise ein Ball, bei genauerer Betrachtung kann er als Fußball identifiziert werden). Diese Verfeinerung der Objektwahrnehmung geschieht durch die Ankunft langsamerer aber detaillierterer Hochpassinformation, welche dann eine exakte Repräsentation festlegt (Bar, 2004). Während der Reizverarbeitung kommt es im Verlauf der Objekterkennung zu wiederholten Veränderungen der Fixationen und der Aufmerksamkeit, was zu einer größeren und ständig wechselnden Auswahl an potentiellen Kandidaten führen sollte, die in den inferioren temporalen Kortex projiziert wird. Da der Kontextrahmen im parahippocampalen Kortex allerdings konstant bleibt (weil es nur eine Szene gibt) und sich weiters zeigte, dass die Identifizierung des Kontextes und die erste Objektfixation für das Erkennen der gesamten Szene entscheidend sind, können diese weiteren Fixationen als für die Erkennung unbedeutend angesehen werden. Die eben beschriebenen Abläufe des Modells von Bar (2004) sind in Abbildung 1 nochmals zusammengefasst:

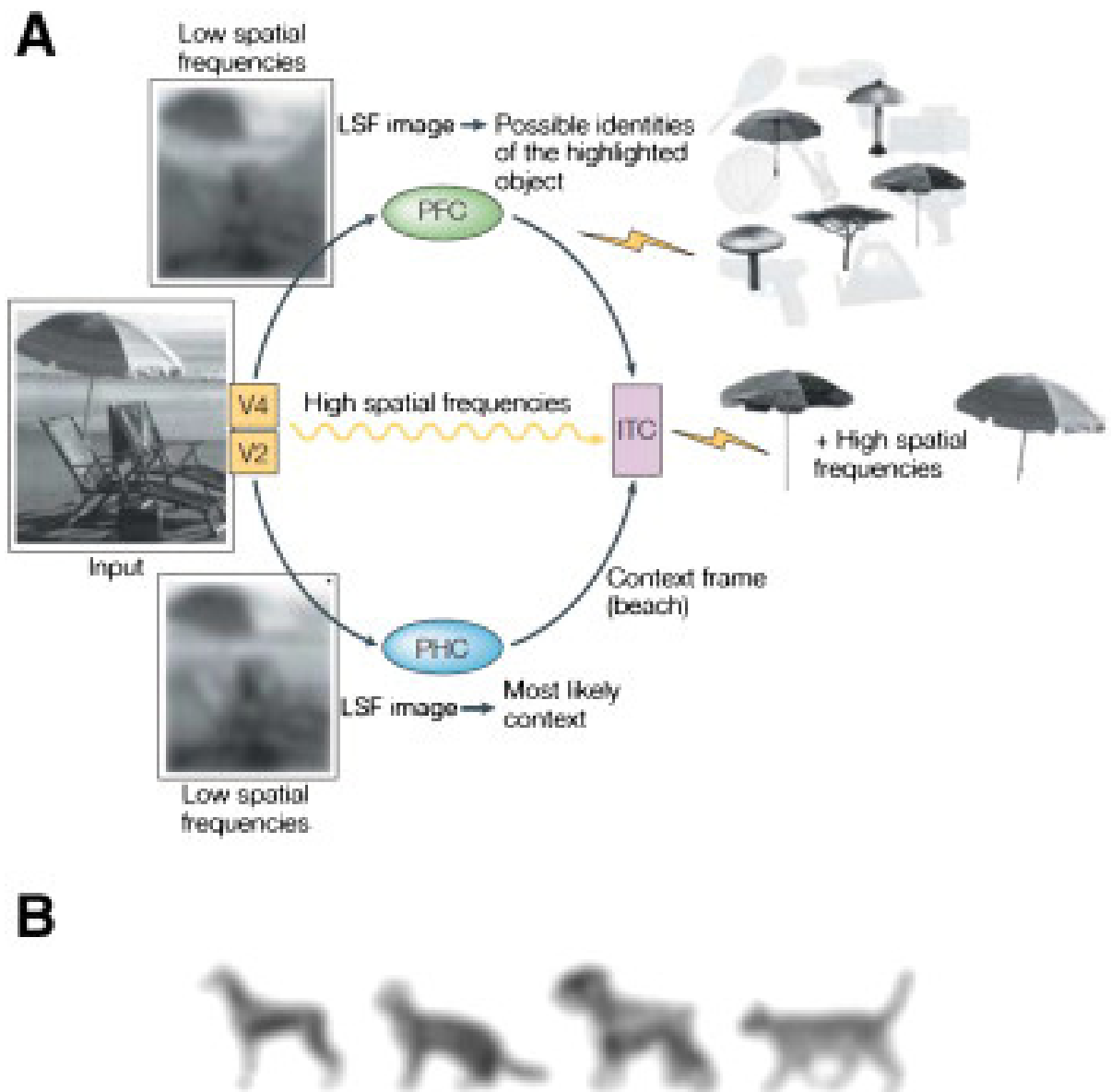


Abbildung 1. Modell der Szenen- und Objektverarbeitung oder auch *model for contextual facilitation* (Bar, 2004)

Vereinfacht dargestellt kommt es also zu folgenden Prozessen:

In seinem vorgestellten Modell fügt Bar (2003) also noch eine zweite Quelle ein, die top-down den bottom-up-Prozess der Objekterkennung beeinflusst, die schnelle Objekterkennung. Diese Pfade der schnellen Szenenanalyse, welche auf Tiefpassinformationen basiert, sowie die schnelle Objekterkennung (welche ebenfalls auf Tiefpassinformationen zurückgreift) und die genaue aber langsame Objekterkennung (welche auf Hochpassinformationen beruht) führen alle zu einer eigenen Hypothese, welche Objekte

an der in Frage kommenden Stelle präsent sein könnten. Diese werden dann beim Abgleich kombiniert und es wird die plausibelste Variante ausgewählt, welche eine Schnittmenge dieser drei Informationsquellen darstellt. Sind nun die Hypothesen der Szenenerkennung und der langsamen Objekterkennung inkongruent, führt dies zu einer Verlangsamung der Erkennung. Somit kann Bars (2003) Modell als Erweiterung des *priming models* (siehe Henderson und Hollingworth, 1999) angesehen werden, welches er durch die schnelle Objekterkennung ergänzt.

Diese top-down-Prozesse können laut Bar (2004) sowohl durch die Art der Aufgabenstellung als auch durch Erwartungen beeinflusst werden. Das bedeutet, dass bei höheren Anforderungen, wenn also beispielsweise die Identifizierbarkeit des Reizes durch verkürzte Präsentation oder Maskierung erschwert wird, die Interpretationen hinsichtlich der Identität des Reizes breiter streuen, da weniger Information zur Verfügung steht. Diese Unsicherheit führt zu einer größeren Anzahl an *initial guesses* (Bar, 2004). Umgekehrt führt die Vereinfachung des Reizes zu schnelleren und genaueren Vermutungen. Jene *initial guesses*, die vom präfrontalen Kortex an den inferioren temporalen Kortex geschickt werden, sind demnach präziser. Für Bar (2004) spielt nicht nur der Kontext eine wichtige Rolle bei der Objekterkennung, auch der umgekehrte Weg ist für ihn denkbar: wird ein präsentiertes Objekt erkannt, erleichtert dies die Identifikation der Hintergrundszene. Dieser Weg nutzt die Information, die der präfrontale Kortex an den inferioren temporalen sendet und schickt sie von dort an den parahippocampalen weiter, was vor allem bei atypischen Szenen von Bedeutung ist. Schließlich ist es auch möglich, dass ein nicht eindeutiges Objekt durch ein zweites (mit diesem in Verbindung stehendes) identifizierbar wird, weil das zweite Objekt auf dem bereits beschriebenen Weg einen Kontext schafft, der die Erkennbarkeit des uneindeutigen Objektes fördert (Bar, 2004).

Analog zu Bars (2004) Modell nehmen wir auch in dieser Untersuchung an, dass das tiefpassfrequente Bild in den präfrontalen und parahippocampalen Kortex projiziert wird. Im parahippocampalen Kortex wird dann ein Kontextrahmen erstellt (in diesem Fall also die Szene "Bowlingbahn"), welcher als Bezugspunkt für die weiteren Reize dient. Die Information wird an den inferioren temporalen Kortex weitergeschickt, welcher auf Basis dieses Rahmens Assoziationen erstellt (was passt zu der Szene? Bälle, Pins usw.) Im präfrontalen Kortex folgt eine Fixation auf das hervorgehobene Objekt (angedeutet durch den hellen Kreis) und eine Auswahl an potentiellen Kandidaten (etwa Fußball, Basketball, Wasserball, Bowlingkugel...). Anschließend erfolgt der bereits besprochene Abgleich dieser zwei Prozesse im inferioren temporalen Kortex und es wird eine Schnittmenge gebildet, die einen wahrscheinlichen Kandidaten hervorbringt (siehe Abbildung 2).

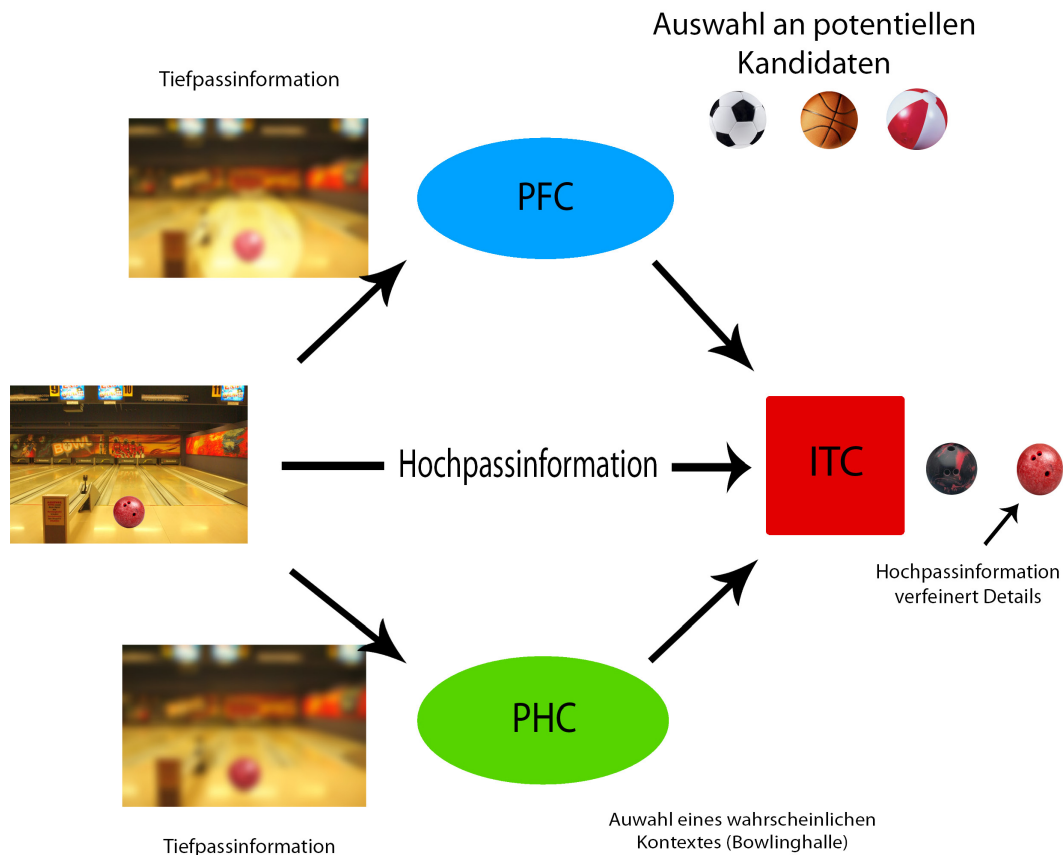


Abbildung 2. Angenommene Verarbeitungsmechanismen in Anlehnung an Bar (2004)

2.3 Forschungsfragen

Ziel der vorliegenden Untersuchung war es nun, einen eventuell auftretenden Szenenkonsistenzeffekt festzustellen, um zu klären, ob sich konsistente Objekt-Szene-Beziehungen positiv auf die Erkennungsleistung auswirken. Weiters wurde versucht, Art und Ausmaß des Einflusses von Objektfilterung auf diesen Konsistenzeffekt festzuhalten und Fehlantworten der Versuchsteilnehmer anhand von Fehlerkategorien in Hinblick auf mögliche Rateeffekte zu analysieren. Es wurde angenommen, dass sich die Objektfilterung verstärkend auf den Konsistenzeffekt auswirkt, da

mehr Kontextinformation für ein Urteil einbezogen werden muss und die Teilnehmer somit zu einer genaueren Analyse der Reize gezwungen werden.

3 Methode

Im folgenden Kapitel werden die Stichprobe, die Materialien und Geräte sowie das Versuchsdesign näher beschrieben, sowie einige Beispiele zur Verdeutlichung des Ablaufs gegeben. Anschließend werden die Fehlerkategorien näher vorgestellt und ein Einblick in die Antwortanalyse gegeben.

3.1 Teilnehmer

Als Teilnehmer wurden 41 PsychologiestudentInnen zwischen 18 und 28 Jahren gewählt, von denen 26 weiblich und 15 männlich waren. Vor dem Experiment wurden die Sehstärke und eventuelle Rot-Grün-Schwächen der Versuchspersonen überprüft, wobei hier von Versuchsperson Nummer 1 eine leichte Farbdifferenzierungsschwäche angegeben wurde, welche jedoch anhand des Testmaterials nicht bestätigt werden konnte. Die Teilnehmer bekamen als Aufwandsentschädigung Versuchspersonenstunden für ihr Studium bestätigt.

3.2 Materialien und Geräte

Als Stimuli dienten 72 Farbfotografien von typischen Szenen, welche mit Bildern von Objekten kombiniert wurden. In der konsistenten Bedingung wurde ein zu der Szene passendes Objekt verwendet, in der inkonsistenten eines, das nicht zum Gesamteindruck passte. Die so entstandenen Stimuli zeigten charakteristische Alltagsszenen wie beispielsweise einen Baumarkt mit Werkzeugkiste in der konsistenten Bedingung und in der inkonsistenten Bedingung den eben genannten Baumarkt kombiniert mit einem Paar Ski. Die Szenen enthielten teilweise auch andere Hintergrundobjekte, die nachträglich eingefügt hoben sich jedoch deutlich von diesen ab, da sie ohne die Szenenanordnung zu beachten in das Bild

eingefügt wurden. Somit waren sie als Zielreiz klar zu erkennen. Die Bilder der Objekte stammten aus dem Internet. Die Objektbilder wurden mithilfe von MATLAB hinsichtlich ihrer räumlichen Eigenschaften bearbeitet, sodass Tiefpasseigenschaften erhalten blieben und Hochpasseigenschaften weggefiltert wurden. Die so erhaltenen Stimuli wirkten verschwommen und wiesen keine feinen Kanten und abrupten Verlaufsänderungen auf. Die Szenenbilder wurden bezüglich Helligkeit und Schärfe aneinander angepasst und anschließend wurden die Objektbilder in den Kontext eingefügt. Die Darbietungsorte waren äquivalent zu Positionen in real aufgenommenen Fotos, die Objekte wurden gleichhäufig links, mittig und rechts eingefügt. Das Experiment wurde mithilfe von MATLAB programmiert und fand im Testraum der Fakultät für Psychologie in Wien statt. Der Raum war abgedunkelt und die Computermonitore wurden durch Hintergrundbeleuchtung erhellt. Die Szenen wurden auf handelsüblichen Computermonitoren mit Standardauflösung präsentiert. Um ablenkende Nebengeräusche zu vermeiden und um eine Störung durch andere Versuchsteilnehmer zu verhindern, wurde den Versuchspersonen die Möglichkeit gegeben, bereitgestellte Ohrstöpsel zu verwenden. Zur Sprachaufzeichnung wurden Kehlkopfmikrofone verwendet. Um einen gleichbleibenden Abstand zum Monitor zu gewährleisten kamen Kopfstützen zur Anwendung, welche im Vorfeld der Testung bereits fix eingestellt wurden. Die Testpersonen wurden instruiert, ihre Köpfe so zu positionieren, dass die Stirn am vorderen Rand der Vorrichtung anlag und das Kinn auf der dafür vorgesehenen Stütze ruhte. Weiters wurde darauf hingewiesen, dass eine Veränderung der Position nur durch Einstellung des Stuhles erlaubt sei. Aus Hygienegründen wurden die Kopfstützen mit Küchenrolle abgeklebt und nach jeder Versuchsperson gewechselt.

3.3 Design und Ablauf

Insgesamt wurden pro Versuchsperson 576 Szenen präsentiert, welche sich wie folgt aufteilten: 144 Szenen gehörten zur Bedingung konsistent/normal, 144 zu konsistent/LSF, weitere 144 zu inkonsistent/normal und die letzten

144 zu inkonsistent/LSF. Bei der Erstellung der Stimuli wurden jeweils zwei Szenen gepaart (etwa das Bild mit der Skiliftszene und das mit dem Baumarkt) und jedes konsistente Objektbild stellte ein inkonsistentes in der gepaarten Szene dar (das Bild der Werkzeugkiste wurde sowohl in der Baumarktszene als auch in der Skiliftszene verwendet). Weiters wurden die konsistenten und inkonsistenten Objektbilder in Größe und Form aneinander angepasst (so wurden beispielsweise in der Szene der Bowlinghalle eine Bowlingkugel und ein Motorradhelm verwendet). Jede Szene wurde achtmal präsentiert, da jeweils zwei konsistente und zwei inkonsistente Objektbilder eingefügt wurden und jedes dieser Bilder sowohl ungefiltert als auch gefiltert mit der Szene kombiniert wurde. Die verschiedenen Kombinationen einer Szene mit den einzelnen Objekten und deren Filterung werden in Abbildung 3 noch einmal verdeutlicht.

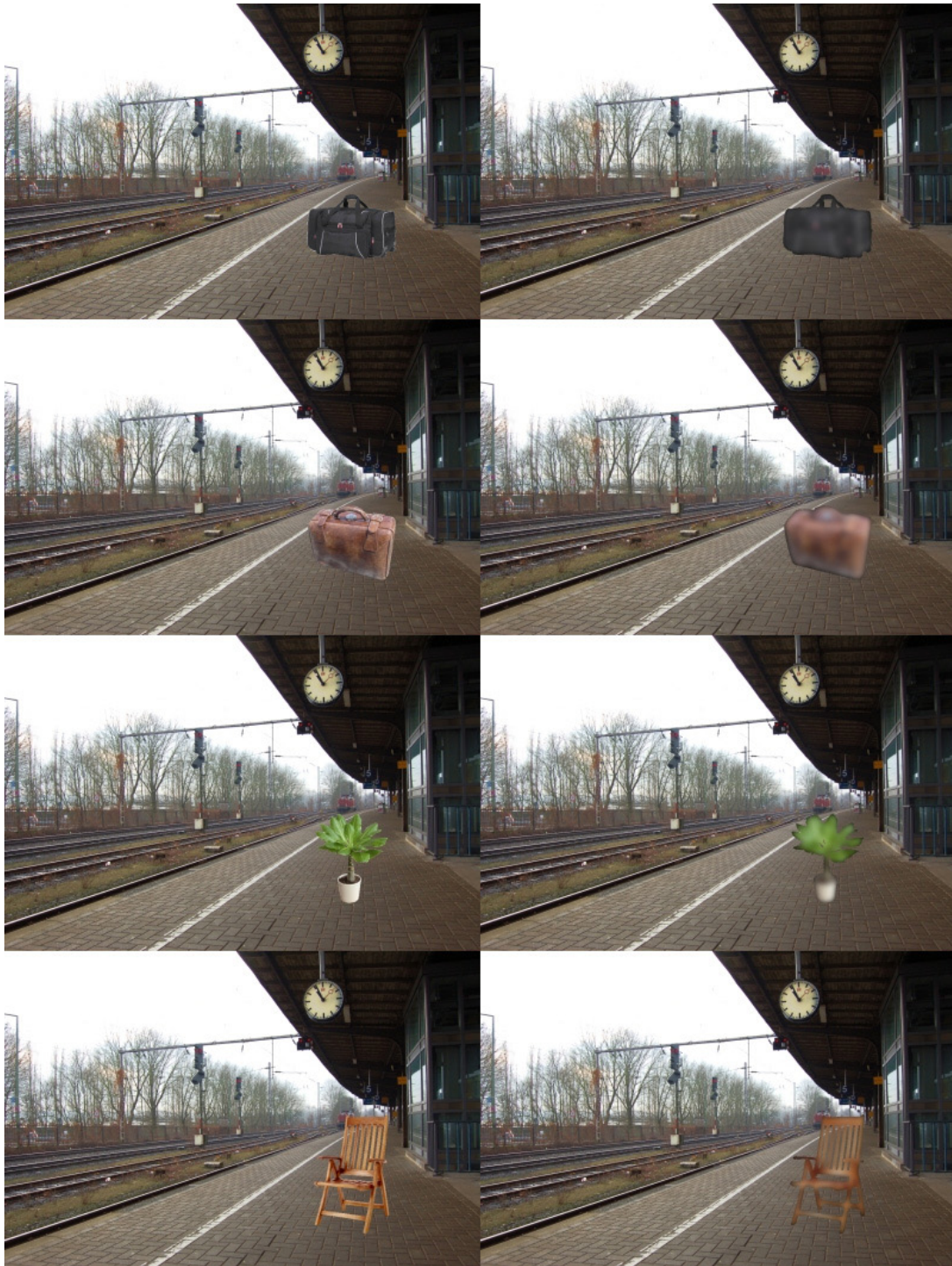


Abbildung 3. Kombination der Szene Bahnhof mit den zugewiesenen Objektbildern: Reisetasche - Reisetasche/tiefpassgefiltert, Koffer - Koffer/tiefpassgefiltert, Pflanze - Pflanze/tiefpassgefiltert, Klappstuhl - Klappstuhl/tiefpassgefiltert (v. l. n. r).

Jedes Trial bestand aus einem Fixationskreuz, welches 500ms eingeblendet wurde, dann folgte die Darstellung der Szene für 350ms, darauf erschien für 32ms eine Maske und anschließend wurde die bereits gezeigte Szene in verkleinerter Form für 5 Sekunden präsentiert, mit einem grauen Kreis als

Hinweis auf den Ort, an dem zuvor das Objekt zu sehen gewesen war. In diesem Zeitraum sollte eine verbale Rückmeldung abgegeben werden, danach blieb der Bildschirm für 1000ms weiß und es folgte das nächste Trial (siehe Abbildung 4).

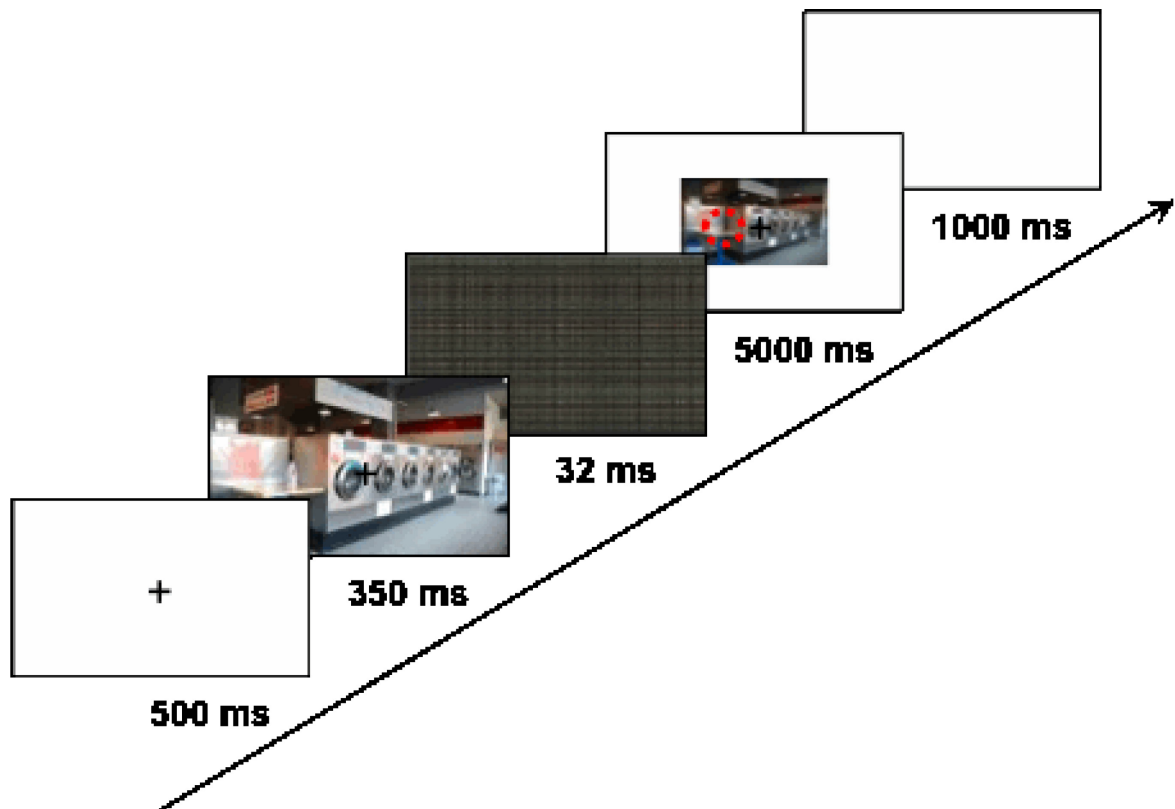


Abbildung 4. Trialabfolge mit Angabe der Darbietungszeiten

Die Teilnehmer erhielten zuerst die Information, dass die verwendeten Mikrofone auf ihre jeweilige Stimme und Antwortlautstärke kalibriert werden müssten, was in einem 5-10minütigen Prozedere unmittelbar vor der Testung geschah. Dazu mussten die Versuchspersonen die am Bildschirm dargestellten Wörter laut vorlesen und auf entsprechende Rückmeldungen des Programms reagieren. Anschließend folgte die Instruktion für die eigentliche Testung. Die Teilnehmer wurden informiert, dass sie nun gleich Darstellungen von Szenen präsentiert bekämen, welche ein künstlich eingefügtes Objekt enthielten. Diese Szenen wären nur für kurze Zeit zu sehen, danach folgte dieselbe Szene noch einmal, allerdings in

verkleinerter Form. Ein grauer Kreis wäre der Hinweis für den Ort des gesuchten Objekts, welches in der ersten Darbietung zu sehen gewesen wäre. Die Teilnehmer wurden instruiert, innerhalb dieser 5000ms andauernden Darbietung eine verbale Rückmeldung über das gesehene Objekt abzugeben, wobei sie auch die Anweisung erhielten, ihre Antworten in Form eines einzigen Wortes abzugeben und wenn möglich ohne jegliche Einleitung. Weiters sei es zulässig, auch nur die Kategorie des gesuchten Objekts zu nennen, als Beispiel wurde hier bei Erscheinen einer Palme die Antwort "Pflanze" genannt. Bei Nicht-Erkennen des gesuchten Objektes solle mit "Irgendwas" geantwortet werden. Im Vorfeld des Experimentes fand eine 30minütige Testung einer Studienkollegin statt, danach gab es die Möglichkeit zu einer 10minütigen Erholungsphase während der kleine Snacks und Getränke angeboten wurden. Die darauffolgende Testung erstreckte sich inklusive Instruktion über etwa 80 Minuten.

3.3.1 Statistische Hypothesen

Abgeleitet aus dem Versuchsaufbau und den Forschungsfragen ergeben sich für die vorliegende Untersuchung folgende Hypothesen:

1. Über einen Mittelwertvergleich soll geprüft werden, ob es einen Unterschied in der Genauigkeit der Antworten der Teilnehmer zwischen den Bedingungen konsistent/inkonsistent gibt, ob in diesem Fall also ein Konsistenzeffekt zu beobachten ist.
 - Als **Alternativhypothese** wird angenommen, dass sich die Genauigkeit der Antworten in den Bedingungen konsistent/inkonsistent signifikant voneinander unterscheiden und es in der konsistenten Bedingung zu signifikant mehr richtigen Antworten kommt als in der inkonsistenten.
 - Sollten keine signifikanten Unterschiede zwischen den Gruppen konsistent/inkonsistent hinsichtlich der Genauigkeit der gegebenen Antworten zu finden sein, wird die entsprechende **Nullhypothese** beibehalten, welche besagt, dass keine Unterschiede bezüglich der

Genauigkeit der Antworten zwischen den Gruppen konsistent/inkonsistent zu finden sind.

2. Über einen weiteren Mittelwertvergleich soll geprüft werden, ob es einen Unterschied in der Genauigkeit der Antworten zwischen den Bedingungen gefiltert/ungefiltert gibt.
 - Die entsprechende **Alternativhypothese** besagt also, dass sich die Genauigkeit der Antworten in den Bedingungen gefiltert/ungefiltert signifikant voneinander unterscheidet.
 - Sind keine signifikanten Unterschiede festzustellen, wird die entsprechende **Nullhypothese** beibehalten, nämlich dass keine Unterschiede zwischen den Gruppen gefiltert/ungefiltert hinsichtlich der Genauigkeit der Antworten zu finden sind.
3. Mittels Univariater Varianzanalyse wird weiters geprüft, ob es zu einer Interaktion zwischen den Faktoren Konsistenz und Filterung kommt. Im Speziellen soll festgestellt werden, ob die Anzahl der richtigen Antworten in der Bedingung konsistent/inkonsistent unter Berücksichtigung der Variable Objektfilterung variiert.
 - Als **Alternativhypothese** wird daher angenommen, dass der Einfluss eines Faktors von der Ausprägung des anderen Faktors abhängig ist.
 - Als **Nullhypothese** wird angenommen, dass sich die Ausprägungen des einen Faktors nicht für die einzelnen Ausprägungen des anderen Faktors unterscheiden.

3.4 Scoring

Die Bewertung der Antworten erfolgte anhand im Vorhinein festgelegter Kriterien. Allgemein wurden Antworten als richtig gewertet, sobald die Erfassung des Oberbegriffes erkennbar war. Eine sinngemäße Beschreibung des Präsentierten (beispielsweise die Antwort "Buch" bei Vorlage einer Speisekarte) wurde somit als richtig gewertet. Weiters wurden auch Nennungen des Oberbegriffes (zum Beispiel die Antwort "Werkzeug" im

Falle eines Schraubenziehers) als korrekt notiert. Die Fehleranalyse geschah dann anhand folgender Kriterien:

Fehlertyp 1

diesem Fehlertyp wurde eine Antwort dann zugerechnet, wenn ein Objekt genannt wurde, das zwar zur Szene passte, jedoch nicht gezeigt wurde. Weiters galt eine Antwort als Fehlertyp 1, wenn das genannte Objekt später in der Untersuchung noch als konsistentes Objekt gezeigt wurde, also bis dato noch nicht präsentiert wurde.

Fehlertyp 2

dieser Fehlertyp beinhaltete Objekte, die an einer anderen Stelle innerhalb der Szene zu sehen waren, jedoch nicht das präsentierte (und bearbeitete) Objekt darstellten.

Fehlertyp 3

unter diese Kategorie fielen Objekte, die weder gezeigt wurden noch einen Bezug zur präsentierten Szene aufwiesen.

Fehlertyp 4

Fehlertyp 4 umfasste Objekte, die bereits in einem vorangegangenen Trial bei dieser Versuchsperson gezeigt wurden und die in diesem Fall ein inkonsistentes Objekt darstellten.

Fehlertyp 5

ähnlich wie Fehlertyp 4 beinhaltete Typ 5 Objekte, die bereits bei diesem Teilnehmer präsentiert wurden, allerdings diesmal als konsistentes Objekt.

Fehlertyp 6

bei Nichterkennen des gezeigten Objektes wurden die Teilnehmer instruiert, mit "Irgendwas" zu antworten, was als Fehlertyp 6 gewertet wurde.

Die Fehlertypen 1, 4 und 5 sind hinsichtlich möglicher Rateeffekte von besonderem Interesse, da sie Antworttendenzen in Richtung Konsistenz (welches Objekt würde hier passen?) sowie Wiederholung (die Versuchsperson erinnert sich an ein bereits gezeigtes Objektbild und wiederholt die Antwort bei einer anderen Szene) abbilden könnten.

3.5 Anmerkungen

Generell kann die Qualität der Stimmaufzeichnungen als eher inhomogen bezeichnet werden, die Bandbreite reicht von "sehr gut verständlich" bis "undefinierbares Rauschen". Weiters neigten die Kehlkopfmikrofone bei hohen Stimmlagen vermehrt zu Fehlermeldungen, was bei weiblichen Teilnehmerinnen zu gehäuften Systemmeldungen führte ("lauter sprechen"). Auch die Audiokalibrierung erwies sich als fehleranfällig, schließlich mussten bei 11 Teilnehmern die Kalibrierungen nach mehreren Versuchen abgebrochen werden, was sich, wie sich bei der Auswertung jedoch herausstellte, nicht qualitätsmindernd auf die Sprachaufzeichnung auswirkte. Versuchsperson Nummer 7 zeigte während der Testung ausgeprägte Belastungserscheinungen (sie versuchte wiederholt aufzustehen und äußerte ihre Absicht die Testung zu verlassen, weiters fragte sie mehrmals nach der verbleibenden Testdauer) und brach diese nach 65 Minuten schließlich ab, weshalb die Daten unvollständig waren und nicht in die Auswertung einfließen konnten. Bei den Versuchspersonen 28, 31 sowie 41 fand aufgrund technischer Probleme keine Sprachaufzeichnung statt, weshalb deren Daten unberücksichtigt bleiben mussten. Weiters konnten einzelne Antworten aufgrund diverser Faktoren nicht in die Auswertung einbezogen werden: Einerseits kam es vereinzelt zu Sprachaufzeichnungs-

ausfällen, andererseits wurden teilweise während des vorgegebenen Antwortzeitraumes keine Rückmeldungen abgegeben. Für ein Ausbleiben der Antwort sind mehrere Erklärungen denkbar:

- die Antwort der Versuchsperson erfolgte nicht innerhalb des verfügbaren Zeitrahmens
- die Versuchsperson war während der Testung unaufmerksam oder abgelenkt und reagierte daher nicht entsprechend auf den präsentierten Reiz
- der Teilnehmer war bezüglich einer Interpretation des Gesehenen unsicher und zögerte mit seiner Antwort zu lange

Somit wurden die Daten der Versuchspersonen Nummer 7 (Abbruch der Testung seitens der Teilnehmerin), Nummer 28 (Fehler bei der Aufzeichnung der Antworten), 31 (ebenfalls technische Probleme bei der Aufzeichnung) sowie Nummer 41 (keine Aufzeichnung der Daten) aus der weiteren Bearbeitung ausgeschlossen. Aus ähnlichen Gründen konnten vereinzelt Antworten nicht berücksichtigt werden, weil sie entweder akustisch nicht verständlich waren oder die Antwort zu spät (nach Ablauf des Antwortfensters) oder gar nicht erfolgte. Insgesamt konnten somit 2428 Antworten nicht in die Auswertung einbezogen werden und wurden in der Tabelle mit "99" markiert.

4 Resultate

In diesem Kapitel wird einerseits eine deskriptive Auswertung vorgenommen, andererseits werden die Ergebnisse der statistischen Analysen präsentiert und dadurch die zuvor aufgestellten Hypothesen geprüft.

4.1 Deskriptive Auswertung

Bei Betrachtung der Daten hinsichtlich der Richtigkeit der Antworten unter Berücksichtigung der Kategorien konsistent/inkonsistent zeigt sich, dass bei den korrekten Antworten konsistente Objekte häufiger erkannt wurden als inkonsistente. Der umgekehrte Fall ist bei den falschen Antworten zu sehen, hier liegen die inkonsistenten Objektbilder vor den konsistenten (siehe Abbildung 5). Betrachtet man die Prozentwerte der richtigen Antworten, so sieht man, dass in der konsistenten Bedingung richtige Antworten zu 75% abgegeben werden, in der inkonsistenten liegt der Wert bei rund 70%.

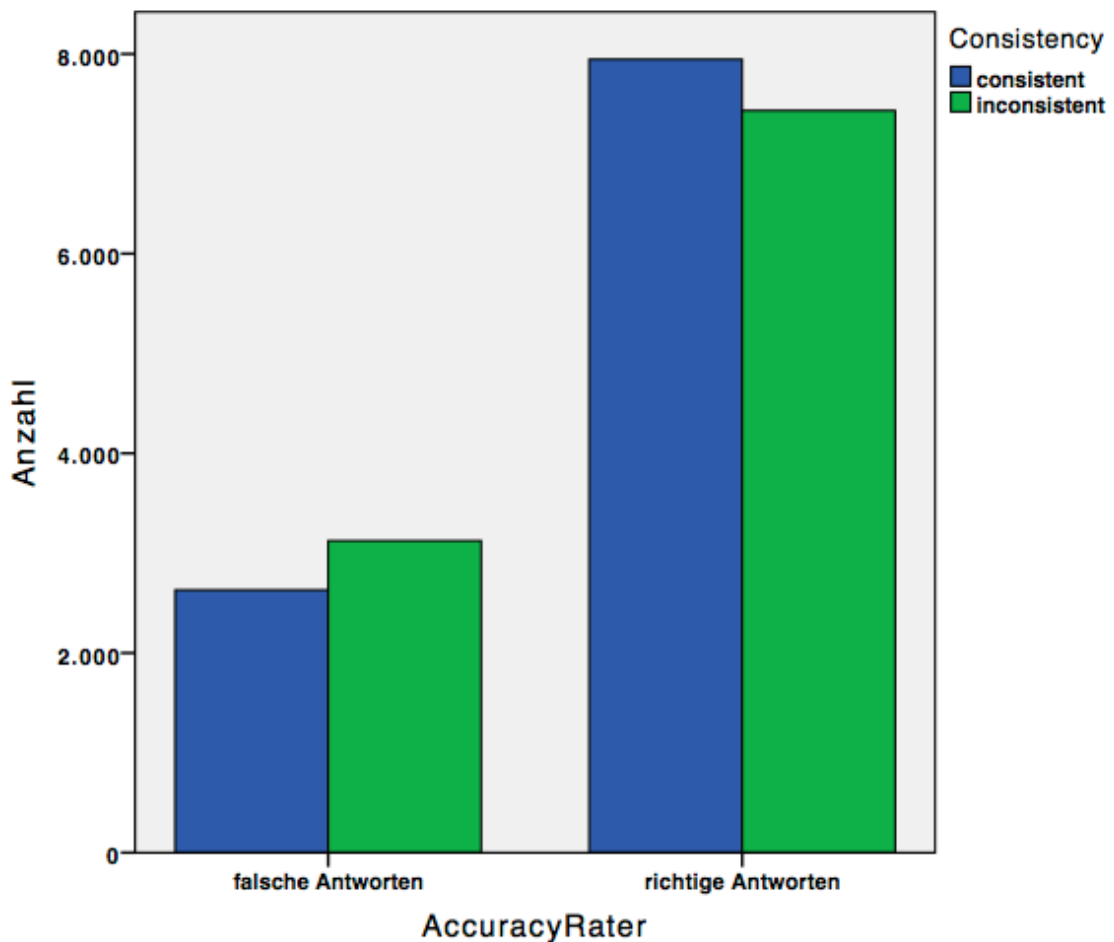


Abbildung 5. Falsche und richtige Antworten unter Berücksichtigung der Kategorien konsistent/inkonsistent

Wird nun zusätzlich die Objektfilterung berücksichtigt, ergeben sich Werte von 83% für ungefilterte Objekte in der konsistenten Bedingung sowie 79% für ungefilterte in der inkonsistenten (siehe Abbildung 6).

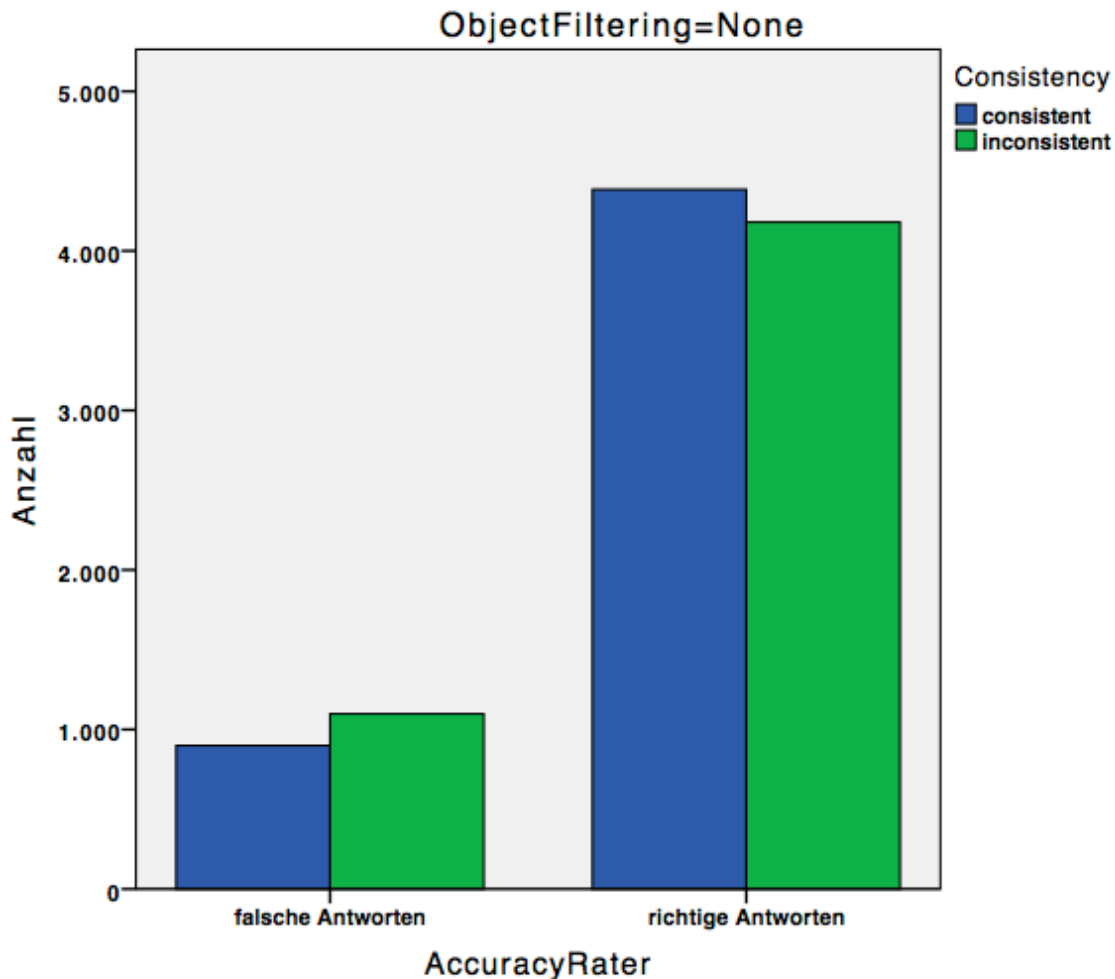


Abbildung 6. Falsche und richtige Antworten unter Berücksichtigung der Kategorien konsistent/inkonsistent bei ungefilterten Objektbildern

In der Filterbedingung wird zu 67% eine korrekte Antwort gegeben, wenn das präsentierte Objekt konsistent mit der Szene ist. Ist es inkonsistent, sinkt dieser Wert auf rund 62%, es lässt sich grafisch somit ein leichter Anstieg der Differenz zwischen konsistenter und inkonsistenter Bedingung im Vergleich zur ungefilterten Bedingung erkennen (siehe Abbildung 7).

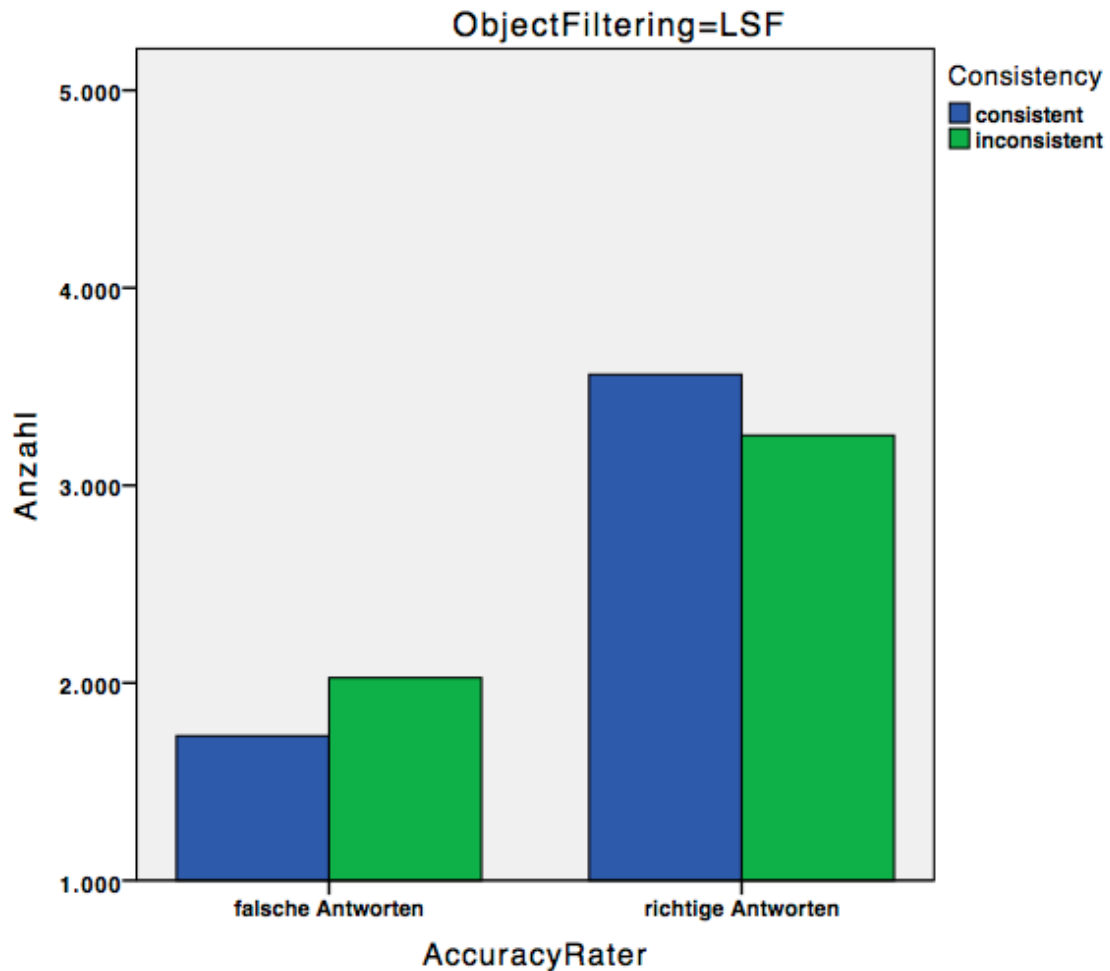


Abbildung 7. Einbezug der Objektfilterung und Auswirkung auf den Konsistenzeffekt

4.1.1 Fehlertypenanalyse

Weiters wurde in Hinblick auf mögliche Rateeffekte ein näherer Blick auf die einzelnen Fehlertypen geworfen. Insgesamt zeigten sich bei 21134 Datensätzen 15379 richtige Antworten, jedoch auch eine hohe Anzahl an Antworten des Fehlertyps 6 (Nichterkennen des Objektes, Meldung "Irgendwas"). In 4769 Fällen kam es somit zu einer ausweichenden Antwort der Versuchsteilnehmer, ein Umstand, dessen mögliche Ursachen im Verlauf der Diskussion näher erörtert werden. Fehlertyp 1 (es wurde ein Objekt genannt, welches zwar zur Szene passt, jedoch (noch) nicht gezeigt wurde) trat in 258 Fällen auf und ist somit deutlich häufiger vertreten als

die Fehlertypen 2 (das genannte Objekt war in der Szene vorhanden, jedoch nicht das gesuchte Objekt) mit 82 Fällen und 4 (das genannte Objekt wurde bereits als inkonsistentes Objekt gezeigt) mit gar nur 9 Fällen. Auch Fehlertyp 5 (das genannte Objekt wurde bereits als konsistentes Objekt gezeigt) zeigte mit 49 Fällen nicht die erwartete Häufigkeit (siehe Tabelle 1).

| Fehlertyp | Häufigkeit | Prozent |
|------------------|-------------------|----------------|
| 1 | 258 | 1,2 |
| 2 | 82 | 0,4 |
| 3 | 588 | 2,8 |
| 4 | 9 | 0,0 |
| 5 | 49 | 0,2 |
| 6 | 4769 | 22,6 |
| x | 15379 | 72,8 |
| Gesamt | 21134 | 100,00 |

Tabelle 1. Häufigkeitsverteilung der einzelnen Fehlertypen (x = richtige Antwort)

Betrachtet man die Verteilung der Fehlertypen unter Berücksichtigung der Objektfilterung, zeigt sich, dass es bei Darbietung der tiefpassgefilterten Objektbilder vermehrt zu Fehlantworten kam (siehe Tabelle 2).

Weiters zeigte sich ein verstärktes Auftreten des Fehlertyps Nummer 3 (es wurde ein Objekt genannt, das nicht zu sehen war und auch keinen Bezug zur gezeigten Szene aufwies. Um diesen Umstand näher zu analysieren muss die Verteilung dieses Fehlertyps genauer betrachtet werden. Einerseits könnte dies (wie es auch bei Fehlertyp Nummer 6 denkbar wäre) auf fehlende Motivation der Versuchsteilnehmer zurückzuführen sein, andererseits ist es bei Typ 3 und Typ 6 möglich, dass es zu einer Häufung

bei bestimmten Objekten und/oder gekommen ist, welche offenbar generell schwer zu identifizieren waren.

| Fehlertyp | LSF | None | Gesamt |
|---------------|-------|-------|--------|
| 1 | 178 | 80 | 258 |
| 2 | 49 | 33 | 82 |
| 3 | 362 | 226 | 588 |
| 4 | 7 | 2 | 9 |
| 5 | 27 | 22 | 49 |
| 6 | 3135 | 1634 | 4769 |
| x | 6814 | 8565 | 15379 |
| Gesamt | 10572 | 10562 | 21134 |

Tabelle 2. Häufigkeitsverteilung der einzelnen Fehlertypen unter Berücksichtigung der Objektfilterung (x = richtige Antwort)

Bezüglich des Fehlertyps 3 kann nach genauerer Analyse der Ergebnisse gesagt werden, dass vor allem die Szenen "Bankvorraum", "Fahrradladen", "Metzgerei" und "Restaurant" mit mindestens 17 Fällen dieses Fehlertyps besonders hervortraten (siehe Tabelle 3). Mit großem Abstand trat Fehlertyp 3 am Häufigsten bei der Szene "Fahrradladen" auf, hier konnten sogar 26 Fälle erfasst werden. Warum dies so ist, wird in der Diskussion versucht zu klären.

Bei näherer Betrachtung der einzelnen Objekte in Hinblick auf Fehlertyp 3 wird deutlich, dass sich einige als besonders anfällig für Falschantworten dieses Typs zeigten. Dies waren die Objekte "Bierdeckel", "Schwimmflügel" und "Zahnspiegel" mit jeweils mindestens 20 Fällen. Auch die Objekte "Beatmungsbeutel", "Cocktailshaker", "Fahrradschloss", "Salz_Pfeffer",

"Schuheinlagen" und "Wasserflasche" waren vermehrt von Antworten des Fehlertyps 3 betroffen (siehe Abbildung 8).

| Szene | Fehlertyp | | | | | | |
|--------------|-----------|---|----|---|---|-----|-----|
| | 1 | 2 | 3 | 4 | 5 | 6 | x |
| Bankvorraum | 4 | 1 | 18 | 0 | 0 | 127 | 145 |
| Fahrradladen | 8 | 3 | 26 | 0 | 2 | 57 | 199 |
| Metzgerei | 0 | 2 | 17 | 0 | 1 | 214 | 60 |
| Restaurant | 3 | 8 | 18 | 0 | 0 | 171 | 96 |

Tabelle 3. Auftretenshäufigkeit des Fehlertyps 3 bei ausgewählten Szenen

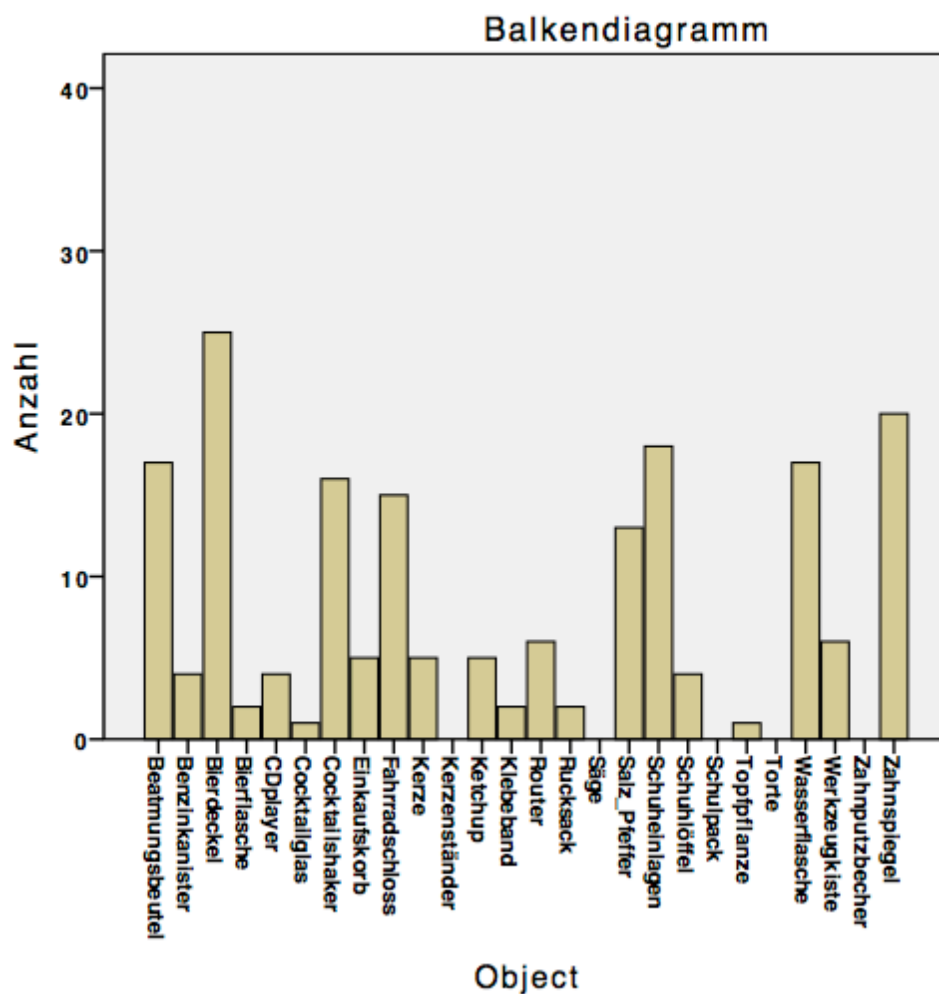


Abbildung 8. Auftretenshäufigkeit des Fehlertyps 3 bei ausgewählten Objekten

Betrachtet man den Verlauf des Fehlertyps 3 über die Testung hinweg, so können keine Tendenzen in Richtung Ermüdungseffekt (absichtlich unpassende oder unüberlegte Antworten aufgrund fehlender Motivation oder auch generell Unaufmerksamkeit) erkannt werden (siehe Abbildung 9).

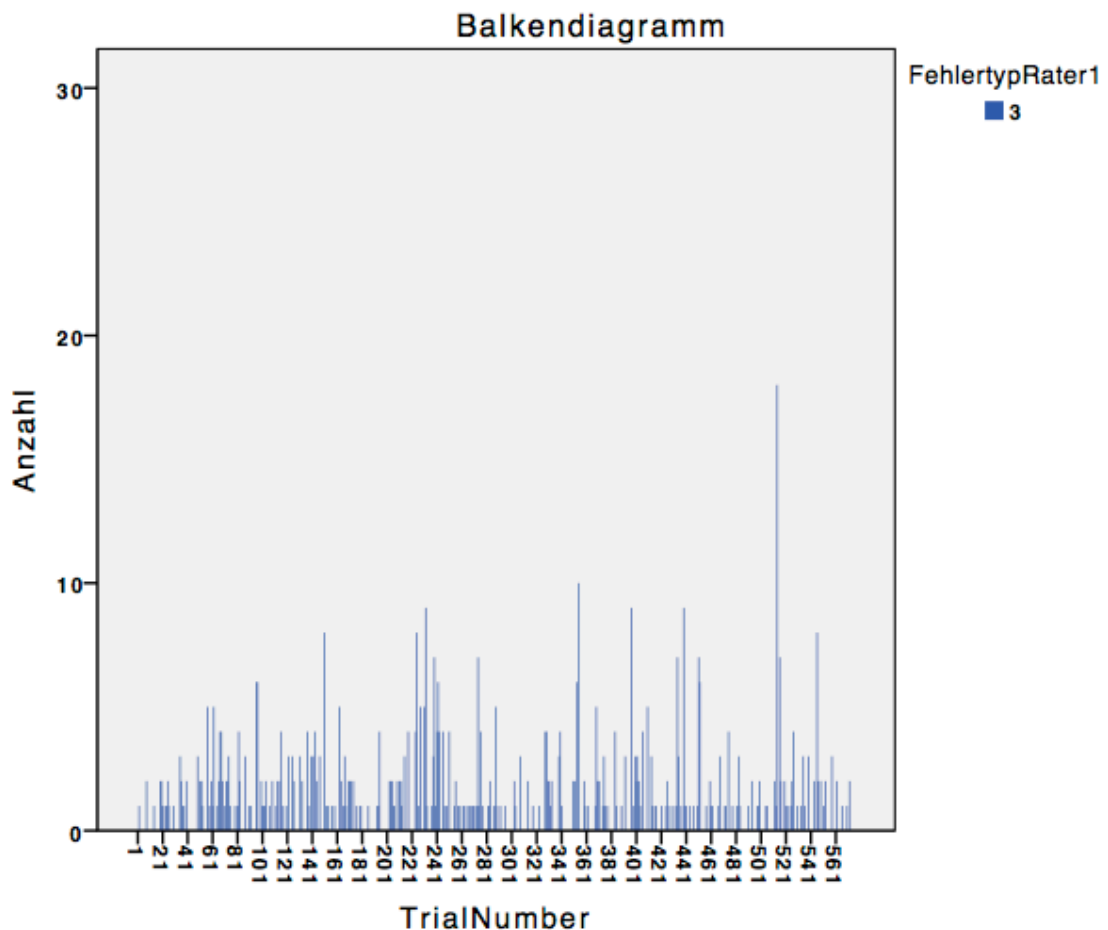


Abbildung 9. Verteilung des Auftretens des Fehlertyps 3 über den Verlauf der Testung

Bei Fehlertyp 6 stachen vor allem die Szenen "Büro", "Metzgerei", "Restaurant", "Serverraum" und "Vorlesungssaal" mit mindestens 170 bis zu 232 Fällen heraus (siehe Abbildung 10). Diese Szenen zeichneten sich durch große Komplexität aus, was eventuell zu Verwirrung und somit fehlender Entscheidungssicherheit bei den Experimententeilnehmern geführt haben könnte.

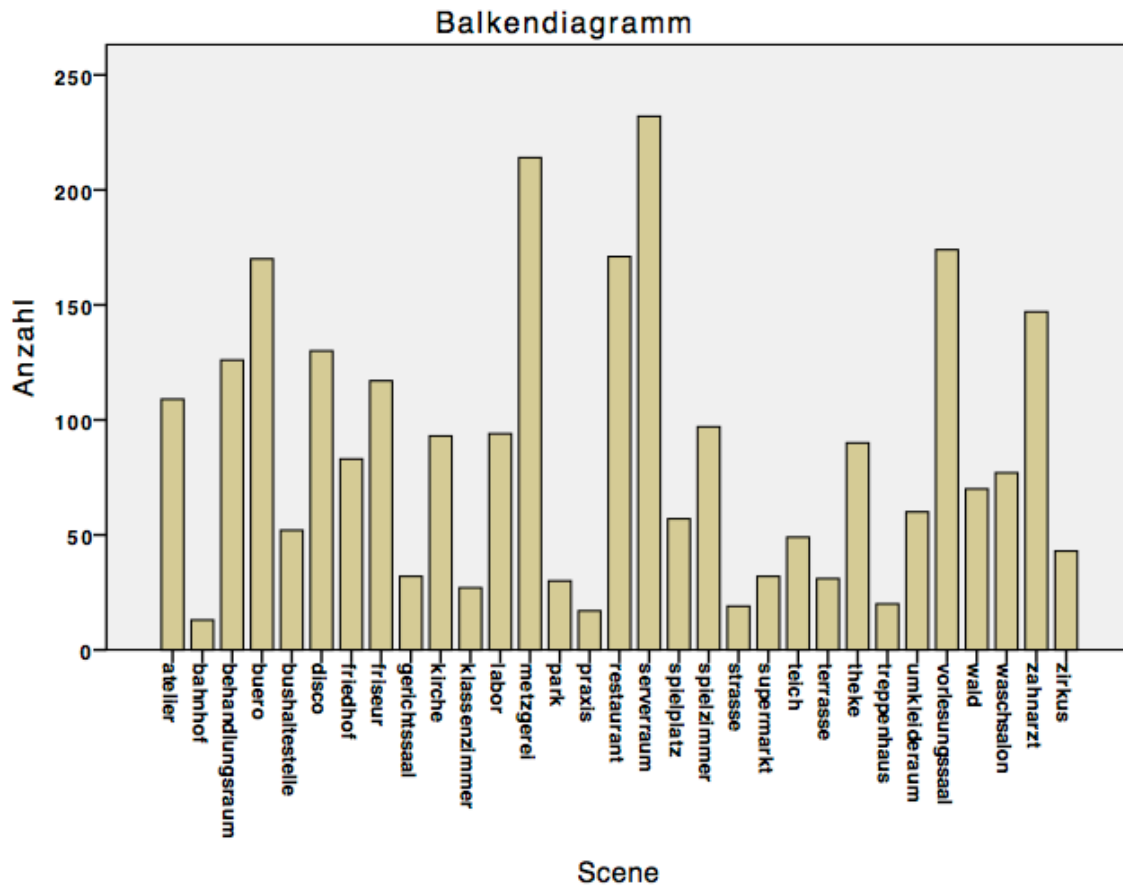


Abbildung 10. Auftretenshäufigkeit des Fehlertyps 6 bei ausgewählten Szenen

Bei genauerer Betrachtung der Fehlertypen in Hinblick auf die präsentierten Objekte zeigt sich auch hier, dass einige Stimuli vermehrt zu Unsicherheitsantworten bei den Teilnehmern geführt haben. So wurden beispielsweise die Objektbilder "Beatmungsbeutel", "Farbtube", "Fleischermesser", "Geldbeutel", "Lockenstab", "Router", "Schwamm" und "Zahnspiegel" jeweils mindestens in 100 Fällen nicht erkannt und mit der Antwort "Irgendwas" versehen (siehe Abbildung 11).

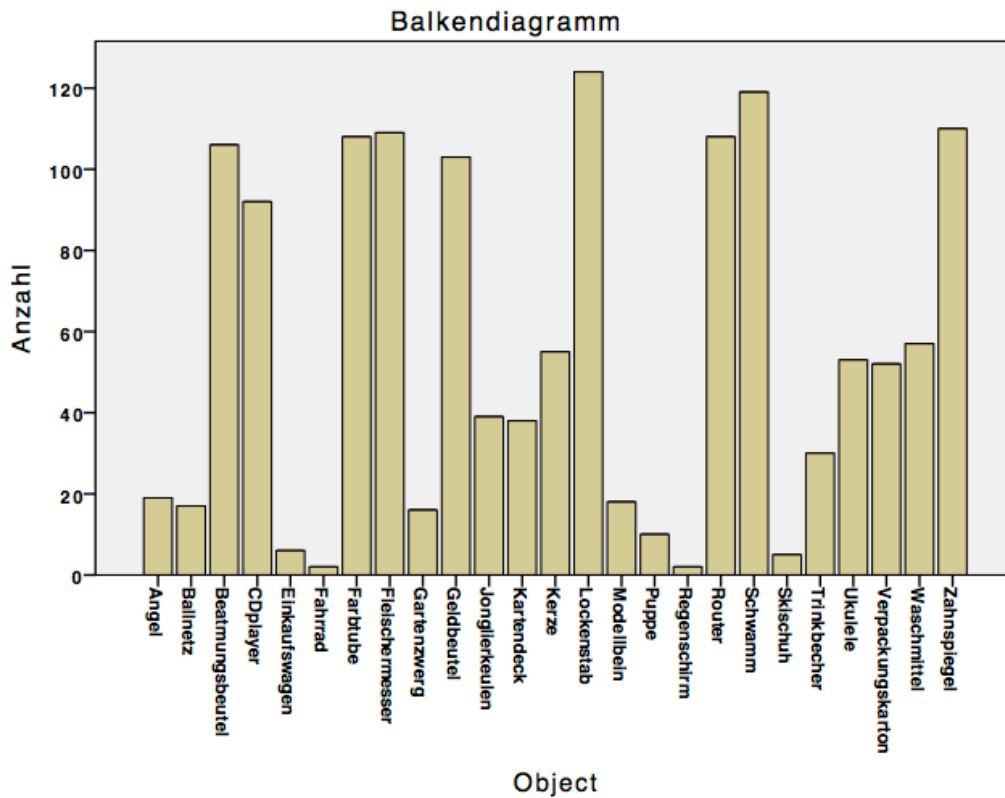


Abbildung 11. Auftretenshäufigkeit des Fehlertyps 6 bei ausgewählten Objekten

Bezüglich der Verteilung des Fehlertyps 6 über den Verlauf der Testung konnten keine Auffälligkeiten erkannt werden, somit sind geringe Motivationstendenzen seitens der Teilnehmer eher auszuschließen (siehe Abbildung 12).

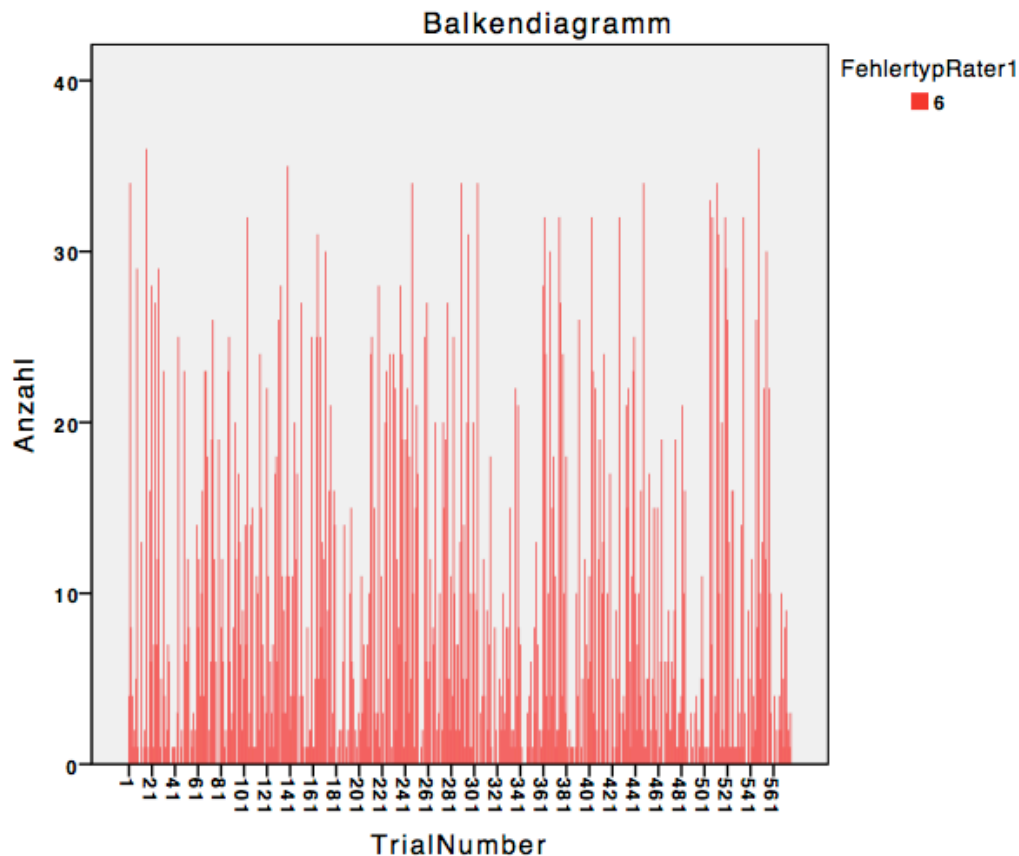


Abbildung 12. Verteilung des Auftretens des Fehlertyps 6 über den Verlauf der Testung

4.2 Statistische Prüfung der Hypothesen

Mithilfe einer ANOVA wurde untersucht, ob die Konsistenz der Objektbilder (konsistent mit der Szene/inkonsistent mit der Szene) Auswirkungen auf die Genauigkeit der Urteile zeigte. Es zeigte sich ein signifikanter Haupteffekt der Objekte-Szene-Konsistenz, Objekte in konsistenten Szenen ($M = 0.75$) wurden genauer beschrieben als in inkonsistenten ($M = 0.70$) [$F = 61.80, p < 0.000$].

Betrachtet man die Objektfiltrierung, zeigt sich auch hier ein signifikanter Haupteffekt [$F = 767.36, p < 0.000$], was verständlich ist, da die Erkennungsleistung bei schwerer identifizierbaren Objektbildern klarerweise sinkt. Es konnte weiters keine signifikante Interaktion zwischen Konsistenz und Objektfiltrierung festgestellt werden [$F = 2.41, p = 0.120$]. Ein

gemeinsamer Effekt würde bedeuten, dass Konsistenz beziehungsweise Filterung nur dann einen Einfluss hätten, wenn der jeweils andere Faktor eine bestimmte Ausprägung annehmen würde. Da dies jedoch nicht der Fall ist, muss davon ausgegangen werden, dass sich die Werte zwischen konsistenter und inkonsistenter Bedingung bei gefilterten Objekten nicht von jenen bei ungefilterten Objekten unterscheiden, der Konsistenzeffekt somit ähnlich stark ausgeprägt ist.

Somit wird bei Hypothese 1 die Alternativhypothese angenommen, die Unterschiede hinsichtlich der Genauigkeit der Antworten zwischen den Gruppen konsistent/inkonsistent unterscheiden sich signifikant, es wird somit ein Konsistenzeffekt beobachtet.

Auch bei Hypothese 2 wird die Alternativhypothese angenommen, es zeigte sich ein signifikanter Unterschied in der Genauigkeit der Antworten zwischen den Gruppen gefiltert/ungefiltert.

Hypothese 3 behandelt eine mögliche Interaktion zwischen den Faktoren Konsistenz und Filterung, hier konnten keine Zwischensubjekteffekte festgestellt werden, somit wird die Nullhypothese beibehalten (die Ausprägungen des einen Faktors unterscheiden sich statistisch gesehen nicht für die einzelnen Ausprägungen des anderen Faktors). Bei Objektfilterung ist der Unterschied zwischen konsistenter und inkonsistenter Bedingung somit nicht anders als in der ungefilterten Bedingung.

5 Diskussion

Das vorliegende Experiment versuchte die Zusammenhänge zwischen Objekten und Szenen und deren Einfluss auf die Wahrnehmung näher zu durchleuchten. Weiters wurden die Auswirkungen von Objektfilterung auf den Szenenkonsistenzeffekt untersucht. Die Ergebnisse zeigen einen stark ausgeprägten Szenenkonsistenzeffekt, Objekte in konsistenten Szenen wurden genauer beschrieben als Objekte in inkonsistenten Szenen.

Der hohe Ausprägungsgrad des Konsistenzeffektes deckt sich mit den

Erkenntnissen von Davenport und Potter (2004), welche eine ganzheitliche Erfassung der Szene und anschließende semantische Informationsbereitstellung propagieren, welche die darauffolgende Objektwahrnehmung beeinflusst. Dennoch müssen hier einige deutliche Unterschiede im Versuchsaufbau zwischen Davenports (2004) und der vorliegenden Untersuchung erwähnt werden. Davenport (2004) verwendete in ihren Experimenten keine Materialwiederholung, ihre Szenen- und Objektbilder wurden somit jeweils nur einmal gezeigt und nicht in verschiedenen Kombinationen wiederholt. Davenports (2004) Absicht dahinter war es, Rateeffekte durch vorausgegangene Auseinandersetzungen mit bereits gezeigten Stimuli zu vermeiden. Sie ist der Meinung, dass die wiederholte Darbietung einen möglichen Konsistenzeffekt vermindern würde. Um Rateeffekte aufgrund von Vorerfahrungen mit einzelnen Stimulusbildern kontrollieren zu können, wurden die Fehlertypen 4 und 5 verwendet, welche sich auf Antworten in Richtung bereits gezeigte Objektbilder konzentrierten. Da beide Kategorien unauffällig ausfielen, kann davon ausgegangen werden, dass sich die wiederholte Darbietung der Objektbilder und Hintergründe nicht zugunsten möglicher Rateeffekte auswirkte. Die Vorlage der Bilder wurde bei Davenport (2004) durch ein Fixationskreuz eingeleitet, das eigentliche Bild wurde dann für 80ms präsentiert und anschließend von einer Maske überblendet. Im Gegensatz dazu wurde im vorliegenden Experiment ein längeres Zeitfenster für die Präsentation der Szenenbilder gewählt, nämlich 350ms.

Ein weiterer wichtiger Unterscheidungspunkt ist der Umgang mit eventuellen Rateeffekten. In Davenports (2004) Fall wurden Datensätze, in denen die Versuchspersonen in inkonsistenten Bedingungen ein Objekt nannten, welches in der konsistenten vorgekommen wäre (beispielsweise gab der Teilnehmer bei Präsentation eines Wohnzimmers die Antwort "Couch", obwohl ein Zebra zu sehen gewesen war), wurden als "intrusion" (Davenport, 2004) gewertet und insofern korrigiert, als dass bei dieser Versuchsperson eine richtige Antwort bei konsistenten Trials subtrahiert wurde. Da dies hinsichtlich einer genauen Fehleranalyse zu unkritisch

scheint, wurden im vorliegenden Experiment Kategorien eingeführt, welche die Art der Fehler näher beleuchtet und somit genauere Schlüsse bezüglich möglicher Rateeffekte zulässt. Dazu wird im weiteren Verlauf der Diskussion nochmals genauer Stellung genommen.

Die gefundenen Ergebnisse deuten in Richtung des interaktiven Modells der Objekt-Szene-Beziehung, welches eine isolierte Betrachtung und Verarbeitung widerlegt und von einer ganzheitlichen Erfassung und Wechselwirkungen von Objekt und Szene ausgeht. Entgegen der Erkenntnisse von Henderson und Hollingworth (1998), welche zur Annahme eines *functional isolation models* geführt haben und somit eine isolierte Verarbeitung der Objekte unabhängig von der Szenenwahrnehmung favorisieren, stützen die vorliegenden Ergebnisse dieses Modell nicht. Dies ist vermutlich auf das unterschiedliche Stimulusmaterial und den abweichenden Antwortmodus der Teilnehmer zurückzuführen. Im Vergleich zu Linienzeichnungen beinhalten natürliche Fotografien mit eindeutigen Stimuli vermutlich mehr Konsistenzinformationen. In Hendersons und Hollingworths (1998) Experimenten musste außerdem teilweise zwischen zwei Möglichkeiten gewählt werden, was zusätzliche Informationen abseits des gezeigten Objektes bereitstellt und somit nachträgliche Schlussfolgerungen begünstigt.

Um zu klären, welchen Einfluss die Filterung der Objekte auf den Konsistenzeffekt hat, wurden die Objekte in einer Bedingung tiefpassgefiltert präsentiert. Es zeigte sich zwar ein signifikanter Effekt der Filterung alleine (die Vorgabe gefilterter Objekte führte zu verringerter Genauigkeit der Antworten), jedoch konnte keine Interaktion zwischen den Faktoren Konsistenz und Objektfilterung beobachtet werden. Es kann also nicht davon ausgegangen werden, dass die Wirkung eines der Faktoren von einer bestimmten Ausprägung des anderen abhängig wäre. Die Unterschiede zwischen konsistenten und inkonsistenten Objekten in der Filterbedingung entsprechen somit jenen in der ungefilterten Bedingung. Gründe für das Ausbleiben eines Effektes der Objektfilterung auf den Konsistenzeffekt ist eventuell das Antwortverhalten der Teilnehmer. Es

wurde angenommen, dass es bei schwer identifizierbaren Reizen (Filterbedingung) vermehrt zu Antworten der Fehlertypen 1, 4 und 5 kommen würde. Die Anzahl dieser Fehlertypen nahm zwar in der Filterbedingung zu, jedoch war die Gesamtanzahl dieser Fehler weiterhin auffallend gering. Somit kann hinsichtlich möglicher Rateeffekte gesagt werden, dass auffallend wenige Antworten auf Basis von Vorerfahrungen (die Versuchsperson hat das Objektbild nicht genau erkannt, nennt jedoch ein zur Szene passendes Objekt) abgegeben wurden. Die Teilnehmer antworteten eher ausweichend (Fehlertyp 6 "Irgendwas") oder generell ohne Bezug zu der Szene zu nehmen) und bezogen wenig Information aus dem Kontext mit ein. Dadurch konnte keine genaue Analyse der Reize gewährleistet werden, weshalb vermutlich auch eine verstärkende Wirkung auf den Konsistenzeffekt ausblieb.

Bezüglich der Fehlertypenanalyse müssen also einige Punkte näher betrachtet werden. Generell konnte die erwartete Häufung des Fehlertyps 1 (es wurde ein Objekt genannt, das zwar zur Szene passte, jedoch (noch) nicht gezeigt wurde), welche in Richtung Rateeffekte deuten würde (die Versuchsperson nennt irgendein Objekt, das für sie passend zur Szene erscheint), nicht beobachtet werden. Auch die Fehlertypen 4 und 5, welche eine Tendenz in Richtung Antwortwiederholung (die Versuchsperson lässt sich in ihrem Antwortverhalten von vorherigen Stimulipräsentationen beeinflussen), traten unerwartet selten auf. Dennoch zeigte sich, dass die Tiefpassfilterung der Objektbilder zu einem Anstieg der Fehlantworten führte, welche im Fall des Fehlertyps 1 in der Filterbedingung sogar mehr als doppelt so oft vorkamen wie in der ungefilterten Darbietungsbedingung. Der verminderte Informationsgehalt der dargebotenen Stimuli (Filterung) führt in diesen Fällen offenbar zur Inanspruchnahme von Kontextinformation, um genauere Urteile abgeben zu können.

Bei Fehlertyp 3 waren es vor allem die Szenen "Bankvorraum", "Fahrradladen", "Metzgerei" und "Restaurant", welche zu Falschantworten führten. Vor allem der Hintergrund "Fahrradladen" (siehe Abbildung 13) stach bei der Analyse besonders heraus, was vermutlich auf den hohen Grad

an Komplexität in dieser Szene zurückzuführen ist.



Abbildung 13. Szene Fahrradladen

Man sieht in dieser Szene deutlich eine große Anzahl an ablenkenden Objekten, welche eventuell zu unzureichender Verarbeitung des Zielreizes geführt haben könnten.

Sieht man sich jene Objekte, welche häufig Fehlertyp 3 verursachen, in Zusammenhang mit der kombinierten Szene an, zeigt sich, dass einige Objekte auch mit problematischen (in Hinblick auf eine Anfälligkeit für Fehlertyp 3) Hintergründen gepaart waren. Betrachtet man die Verteilung der Objekte auf die einzelnen Hintergrundbilder scheint es wohl einerseits eine Kombinationswirkung in Richtung Fehlantwort zu geben, andererseits aber auch einen isolierten Einfluss der Szenen und Objekte (s. Tabelle 4).

| Bankvorraum | Fahrradladen | Metzgerei | Restaurant |
|--------------------|---------------------|------------------|-------------------|
| Zahnspiegel | Fahrradschloss | | Schuheinlagen |
| | Bierdeckel | | |

Tabelle 4. Kombiniertes Auftreten von Objekt- und Szenebildern, welche vermehrt zu Antworten des Fehlertyps 3 führten

Einige Objektbilder führten wiederholt zu konsistenten Fehlantworten, wovon einige hier aufgelistet werden (siehe Tabelle 5), um sie bei weiteren Studien eindeutiger zu gestalten oder auszuschließen:

| Objektbild | alternative Antworten |
|-------------------|------------------------------|
| Bierdeckel | Ball |
| Schwimmflügel | Kissen |
| Zahnspiegel | Golfschläger |
| Beatmungsbeutel | Flasche |
| Cocktailshaker | Urne, Vase |
| Fahrradschloss | Seil, Schlauch |
| Salz_Pfeffer | Tasche |
| Wasserflasche | Gasflasche |

Tabelle 5. Wiederholt aufgetretene Falschantworten bei Präsentation bestimmter Objektbilder

Bei genauerer Analyse der Daten zeigte sich eine hohe Anzahl des Fehlertyps 6 (Nichterkennen des Reizes, Antwort "Irgendwas"), was einerseits auf einen hohen Schwierigkeitsgrad hinsichtlich der Identifikation der Stimulusmaterialien und andererseits auf eine geringe Motivation seitens der Versuchsteilnehmer aufgrund der Länge der Testprozedur zurückzuführen sein könnte. Geringe Motivationstendenzen wurden jedoch bereits im Abschnitt Results ausgeschlossen, daher konzentriert sich die weitere Diskussion vorwiegend auf die Stimulusmaterialien.

Bei näherer Betrachtung der Daten zeigt sich, dass einige Szenen eine auffällig hohe Anzahl des Fehlertyps 6 aufwiesen, was einerseits an einer möglicherweise ablenkenden oder uneindeutigen Umgebung und andererseits an den darin gezeigten Objekten liegen könnte. So wurden, wie schon bei den Resultaten berichtet, einige Objekte auffallend selten erkannt und daher mit der Antwort "Irgendwas" versehen. Betrachtet man jene Objekte, welche oft nicht identifiziert werden konnten, in Kombination mit den Szenenhintergründen, in denen sie präsentiert wurden, zeigt sich, dass einige Szenen wie "Büro" oder "Restaurant" unabhängig von den gezeigten Objekten zu vermehrten Antwortentendenzen in Richtung Fehlertyp 6 neigten, die Szenen "Metzgerei", "Serverraum" und "Vorlesungssaal" jedoch jeweils mindestens ein Objekt enthielten, welches mit einer starken Ausprägung des Fehlertyps 6 behaftet ist. Auch hier kommt es offenbar zu Einzel- und Kombinationswirkungen der Szenen und Objekte (siehe Tabelle 6).

| Büro | Metzgerei | Restaurant | Serverraum | Vorlesungssaal |
|------|-----------------|------------|-----------------|----------------|
| | Fleischermesser | | Router | Schwamm |
| | Farbtube | | Beatmungsbeutel | |

Tabelle 6. Kombiniertes Auftreten von Objekt- und Szenebildern, welche vermehrt zu Antworten des Fehlertyps 6 führten

Weiters ließen sich keine Unterschiede bezüglich der Verteilung des Fehlertyps 6 bei einzelnen Objekten hinsichtlich konsistenter versus inkonsistenter Hintergrund feststellen, offenbar sind auch hier einige Objekte generell schwerer zu identifizieren als andere.

Als einschränkender Faktor hinsichtlich der Aussagekraft der vorliegenden Untersuchung muss einerseits die lange Testdauer (rund 80 Minuten, zusätzlich eine 30minütige andere Testung im Vorfeld) erwähnt werden, welche zwar nicht unbedingt zu einem absteigenden aber eventuell generell niedrigen Motivationslevel geführt haben könnte. Die Versuchspersonen

wussten im Vorfeld der Testung, wie lange diese in etwa in Anspruch nehmen würde, was wiederholt zu Unmutsäußerungen geführt hatte. Insofern wäre eine gekürzte Fassung mit weniger Trials wünschenswert, um das Durchhaltevermögen und vielmehr die Ausgangsmotivation der Teilnehmer zu gewährleisten. Somit könnte die auffällig hohe Zahl an Antworten des Fehlertyps 6 (die Person hat das Objektbild nicht erkannt und antwortet "irgendwas") einerseits auf diese fehlende Startmotivation, andererseits auf Schwierigkeiten bei der Identifikation der Objekte zurückzuführen sein. Um weitere Möglichkeit auszuschließen, wäre eine Voruntersuchung mit Präsentation aller fraglichen Stimuli, bei der diese benannt werden müssen, von Vorteil. Hier würde sich abzeichnen, mit welchen Objektbildern die Teilnehmer Schwierigkeiten haben (eventuell aufgrund zu hoher Ähnlichkeit mit anderen Objekten oder weil die Bilder zu wenig oder zu stark detailliert abgebildet wurden). Weiters erwies sich die Fehlerkodierung aufgrund der vielen Fehlerkategorien und der hohen Anzahl an Trials ironischerweise als äußerst fehleranfällig, weswegen hier ein einfacheres Prozedere oder weniger Trials wünschenswert gewesen wären.

6 Zusammenfassung

6.1 Deutsch

Die vorliegende Untersuchung beschäftigt sich mit der Filterung von präsentierten Objektbildern und deren Auswirkung auf den Szenenkonsistenzeffekt. Weiters war von Interesse, ob es aufgrund der Objektbildfilterung zu Rateeffekten kommt. Im folgenden Experiment wurden hierzu Fotografien von Szenen präsentiert, die eingefügte Objektbilder enthielten, welche in der konsistenten Bedingung zu der Szene passten und in der inkonsistenten dem Hauptinhalt widersprachen. In einer weiteren Bedingung wurden die Bilder tiefpassgefiltert präsentiert, um zu sehen, welche Auswirkungen dies auf den angenommenen Szenenkonsistenzeffekt hat.

Die Ergebnisse deuten auf einen signifikanten Szenenkonsistenzeffekt hin, der jedoch nicht mit der Objektfilterung variiert. Es konnten somit keine Interaktionseffekte zwischen Konsistenz und Objektfilterung festgestellt werden. Schließlich wurde noch anhand einer Fehlertypenanalyse zur Bestimmung des Ausmaßes an möglichen Rateeffekten untersucht, inwieweit sich die Versuchsteilnehmer durch bereits präsentierte Objekte sowie antizipierte Konsistenzen zwischen Hintergrund und Objekt in ihrem Antwortverhalten beeinflussen ließen. Dabei zeigte sich kaum eine Tendenz in Richtung Rateeffekte, die Teilnehmer tendierten bei Falschantworten eher zu Begriffen, die in keinem Verhältnis zur gezeigten Szene standen oder gaben ausweichende Antworten ("Irgendwas").

6.2 Englisch

This specific study deals with the filtering of presented pictures of objects and the impact on consistency effects between objects and scenes. In addition, the effect of object filtering on the probability of guessing biases was tested. In the following experiment, pictures of scenes which included typical objects were presented. In one condition these objects were consistent with the background, in the other they contradicted the gist of the scene. Additionally, the pictures of objects were altered (low spatial frequency filtered) to analyse the impact on the consistency effect.

Results show a significant consistency effect, however, consistency does not interact with the filtering of objects. Finally, to control for guessing tendencies, categories were constructed to classify wrong answers. There were no noticeable trends towards guessing consistent object names, participants rather tended to give answers that were unrelated to the presented scene.

7 Literatur

Bar, M. (2003). A Cortical Mechanism for Triggering Top-Down Facilitation in Visual Object Recognition. *Journal of Cognitive Neuroscience*, *15*, 600-609. doi:10.1162/089892903321662976

Bar, M. (2004). Visual Objects in Context. *Nature*, *5*, 617-629. doi:10.1038/nrn1476

Biederman, I. (1972). Perceiving real-world scenes. *Science*, *177*, 77-80. doi:10.1126/science.177.4043.77

Bonitz, V. S., & Gordon, R. D. (2008). Attention to smoking-related and incongruous objects during scene viewing. *Acta Psychologica*, *129*, 255-263. doi:10.1016/j.actpsy.2008.08.006

Bornstein, M., & Mash, C. (2011). Perception of Object-Context Relations: Eye-Movement Analyses in Infants and Adults. *Developmental Psychology*, *47*, 364-375. doi:10.1037/a0021059

Boyce, S. J., & Pollatsek, A. (1992). Identification of objects in scenes: The role of scene background in object naming. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *18*, 531-543. doi:10.1037//0278-7393.18.3.531

Boyce, S. J., Pollatsek, A., & Rayner, K. (1989). Effect of background information on object identification. *Journal of Experimental Psychology: Human Perception & Performance*, *15*, 556-566. doi:10.1037//0096-1523.15.3.556

Castelhano, M., & Heaven, C. (2010). The relative contribution of scene context and target features to visual search in scenes. *Attention, Perception, & Psychophysics*, *72*, 1283-1297. doi:10.3758/APP.72.5.1283

- Chaigneau, S., Barsalou, L., & Zamani, M. (2009). Situational information contributes to object categorization and inference. *Acta Psychologica, 130*, 81-94. doi:10.1016/j.actpsy.2008.10.004
- Cohen, M., Alvarez, G., & Nakayama, K. (2011). Natural-Scene Perception Requires Attention. *Psychological Science, 22*, 1165-1172. doi:10.1177/0956797611419168
- Davenport, J. L., & Potter, M. C. (2004). Scene consistency in object and background perception. *Psychological Science, 15*, 559-564. doi:10.1111/j.0956-7976.2004.00719.x
- Davenport, J. (2005). Context in Object and Scene Perception. Dissertation, Massachusetts Institute of Technology.
- Davenport, J. (2007). Consistency effects between objects in scenes. *Memory & Cognition, 35*, 393-401. doi:10.3758/BF03193280
- De Graef, P., Christiaens, D., & d'Ydewalle, G. (1990). Perceptual effects of scene context on object identification. *Psychological Research, 52*, 317-329. doi:10.1007/BF00868064
- Friedman, A. (1979). Framing pictures: Role of knowledge in automatized encoding and memory for gist. *Journal of Experimental Psychology: General, 108*, 316-355. doi:10.1037//0096-3445.108.3.316
- Ganis, G., & Kutas, M. (2003). An electrophysiological study of scene effects on object identification. *Brain Research: Cognitive Brain Research, 16*, 123-144. doi:10.1016/S0926-6410(02)00244-6
- Henderson, J. M., & Hollingworth, A. (1999). High-level scene perception. *Annual Review of Psychology, 50*, 243-271. doi:10.1146/annurev.psych.50.1.243
- Henderson, J.M., Weeks, P. A., & Hollingworth, A. (1999). The effects of semantic consistency on eye movements during complex scene viewing. *Journal of Experimental Psychology. Human Perception and Performance, 25*, 210-228. doi:10.1037//0096-1523.25.1.210

- Hochberg, J. (1978). *Perception* (2nd ed.). Englewood Cliffs, NJ: Prentice Hall.
- Hollingworth, A., & Henderson, J. M. (1998). Does consistent scene context facilitate object perception? *Journal of Experimental Psychology: General*, *127*, 398-415. doi:10.1037/0096-3445.127.4.398
- Intraub, H. (1981). Rapid conceptual identification of sequentially presented pictures. *Journal of Experimental Psychology: Human Perception & Performance*, *7*, 604-610. doi:10.1037/0096-1523.7.3.604
- Intraub, H. (2012). Rethinking visual scene perception. *Cognitive Science*, *3*, 117-127. doi:10.1002/wcs.149
- Itti, L., & Koch, C. (2000). A saliency-based search mechanism for overt and covert shifts of visual attention. *Visual Research*, *40*, 1489-1506. doi:10.1016/S0042-6989(99)00163-7
- Joubert, O., Fize, D., Rousselet, G., Fabre-Thorpe, M. (2008). Early interference of context congruence on object processing in rapid visual categorization of natural scenes. *Journal of Vision*, *8*, 1-18. doi:10.1167/8.13.11
- Loftus, G. R. (1976). A framework for a theory of picture memory. In J. Senders & R. Monty (Eds.), *Eye Movements and Psychological Processes*. Hillsdale, N.J.: Lawrence Erlbaum Associates.
- Loftus, G. R., & Bell, S. M. (1975). Two types of information in picture memory. *Journal of Experimental Psychology: Human Learning and Memory*, *104*, 103-113.
- Loftus, G. R., & Mackworth, N.H. (1978). Cognitive determinants of fixation location during picture viewing. *Journal of Experimental Psychology: Human Perception and Performance*, *4*, 565-572. doi:10.1037//0096-1523.4.4.565
- Mackworth, N. H., & Morandi, A. J. (1967). Gaze selects informative details within pictures. *Perception & Psychophysics*, *2*, 547-552. doi:10.3758/BF03210264

- Martens, U., Trujillo-Barreto, N., & Gruber, T. (2011). Perceiving the Tree in the Woods: Segregating Brain Responses to Stimuli Constituting Natural Scenes. *The Journal of Neuroscience*, *31*, 17713-17718.
doi:10.1523/JNEUROSCI.4743-11.2011
- Mudrik, L., Deouell, L., & Lamy, D. (2011). Scene congruency biases Binocular Rivalry. *Consciousness and Cognition*, *20*, 756-767.
doi:10.1016/j.concog.2011.01.001
- Mudrik, L., Lamy, D., & Deouell, L. (2010). ERP evidence for context congruity effects during simultaneous object-scene processing. *Neuropsychologia*, *48*, 507-517.
doi:10.1016/j.neuropsychologia.2009.10.011
- Nakashima, R., & Yokosawa, K. (2011). Does scene context always facilitate retrieval of visual object representations? *Psychonomic Bulletin & Review*, *18*, 309-315. doi:10.3758/s13423-010-0045-x
- Palmer, S. E. (1975). Effects of contextual scenes on identification of objects. *Memory & Cognition*, *3*, 519-526.
- Sampanes, A., Tseng, P., & Bridgeman, B. (2008). The role of gist in scene recognition. *Vision Research*, *48*, 2275-2283.
doi:10.1016/j.visres.2008.07.01
- Tanaka, K. (1996). Inferotemporal cortex and object vision. *Annual Review of Neuroscience*, *19*, 109-139. doi:10.1146/annurev.neuro.19.1.109
- Ullman, S. (1995). Sequence seeking and counter streams: A computational model for bidirectional information flow in the visual cortex. *Cerebral Cortex*, *5*, 1-11. doi:10.1093/cercor/5.1.1
- Underwood, G., & Foulsham, T. (2006). Visual saliency and semantic incongruency influence eye movements when inspecting pictures. *Quarterly Journal of Experimental Psychology*, *59*, 1931-1949.
doi:10.1080/17470210500416342
- Underwood, G., Templeman, E., Lamming, L., & Foulsham, T. (2008). Is attention necessary for object identification? Evidence from eye

movements during the inspection of real-world scenes. *Consciousness and Cognition*, 17, 159-170. doi:10.1016/j.concog.2006.11.008

Vo, M., & Henderson, J. (2011). Object-scene inconsistencies do not capture gaze: evidence from the flash-preview moving-window paradigm. *Attention Perception & Psychophysics* 73, 1742-1753. doi:10.3758/s13414-011-0150-6

CURRICULUM VITAE

Florian Schrittwieser

| | |
|--------------------|--------------------------------------|
| Geboren am | 11. September 1985 |
| Familienstand | ledig |
| Staatsbürgerschaft | Österreich |
| Adresse | Plankengasse 9, 2700 Wiener Neustadt |
| Fremdsprachen | Englisch |

Ausbildung

| | |
|------------------|---|
| 1992 - 1996 | Volksschule |
| 1996 - 2004 | BG Babenbergerring Wiener Neustadt (Informatikschwerpunkt) |
| 10/2005 | Beginn des Psychologiestudiums an der Universität Wien mit Schwerpunkt Sozialpsychologie und Wirtschaftspsychologie |
| vorauss. 10/2012 | Magister der Naturwissenschaften (Mag. rer. nat.) |

Berufserfahrung

| | |
|-------------|--|
| 2004 - 2005 | Zivildienst beim Roten Kreuz Wiener Neustadt |
| seit 2005 | freiwilliger Mitarbeiter beim Roten Kreuz Wiener Neustadt |
| 2007 - 2008 | Mitarbeiter im Cineplexx Wiener Neustadt |
| seit 2008 | Mitarbeiter beim Internationalen Musik- und Medienzentrum (IMZ) in Wien |
| Sommer 2008 | Praktikant beim klinischen Psychologen Dr. Karl Schrittwieser in Wiener Neustadt |