

Universität Konstanz
FB Informatik und Informationswissenschaft
Master-Studiengang Information Engineering

Masterarbeit

Theorie und Praxis von Benutzerbefragungen, Benutzungsstudien und
Benutzbarkeitstests in der Automobilindustrie.

*zur Erlangung des akademischen Grades eines
Master of Science (M.Sc.)*

Studienfach: Information Engineering
Schwerpunkt: Computer Science
Themengebiet: Angewandte Informatik

von

Andrea Söter

andrea.soeter@gmail.com

Matrikel-Nr.: 01/556143
Erstgutachter: Prof. Dr. Harald Reiterer
Zweitgutachter: Prof. Dr. Marc. H. Scholl
Einreichung: 2009-12-03

Zusammenfassung

Es finden immer mehr neue Technologien Einzug in das Fahrzeug. Unterhaltungselektronik wie Radio, CD-Player oder auch Mobiltelefone sind im automotiven Kontext schon seit vielen Jahren nicht mehr ungewöhnlich. Das Fahrzeug erlebt einen Wandel vom reinen Transportmittel zu einem Informationssystem, was am 15. November 2008 besonders deutlich wurde, als BMW das erste Fahrzeug mit Internetzugang auf den deutschen Markt brachte [BMW 08].

Bei der Entwicklung automotiver Systeme sind jedoch aufgrund des besonderen Anwendungskontextes diverse Rahmenbedingungen zu beachten, die sich von denen am Desktop-PC unterscheiden und aus diesem Grunde besondere Herausforderungen bei der Konzeption, Entwicklung und Evaluation darstellen. Zur Hinführung an die Thematik der Arbeit werden diese Rahmenbedingungen sowie die damit verbundenen Forschungsfragen erläutert. Anschließend werden konventionelle Methoden von Benutzerbefragungen, Benutzungsstudien und Benutzbarkeitstests der Mensch-Computer Interaktion vorgestellt und es wird diskutiert, inwiefern diese auch in der Automobilindustrie Anwendung finden könnten und es auch tun. Ein beispielhafter Lifecycle soll dabei aufzeigen, wie die theoretischen Methoden praktisch umgesetzt werden können. Der Schwerpunkt liegt hierbei auf den Datenerhebungs- und Evaluationstechniken, die im Rahmen der Entwicklung eines automotiven Webbrowsers angewandt wurden, um zwei Evaluationsstudien zu planen, durchzuführen und auszuwerten.

Die erste Studie war von qualitativer Natur und hatte das Ziel die vielversprechendsten Interaktionskonzepte dreier Browservarianten zu identifizieren, damit diese in einem nächsten Iterationsschritt zu einem finalen Design führen konnten. Diese Untersuchung fand sowohl im Stand, als auch in der Fahrt statt. Da der reale Nutzungskontext der Fahrt die stationär erlebten Bedienungsprobleme verstärkte, wurde dieser Aspekt des Testdesigns bei der zweiten, quantitativ orientierten Evaluationsstudie beibehalten. Diese hatte es zum Ziel, den optimierten Webbrowser mit einer funktional reduzierten Basisvariante experimentell zu vergleichen. Aus diesem Grunde wurde bei der Auswahl der Probanden das Hauptaugenmerk auf eine homogene Nutzergruppe gelegt. Die Ergebnisse dieser zweiten Studie zeigen, dass der hier gewählte iterative Prozess sowie die Untersuchung des Systems in seinem realen Nutzungskontext wichtig waren, um die Bedienungsprobleme zu identifizieren und die Prioritäten der Benutzer beim automotiven Browsen zu erkennen. Nur so konnte ein gebrauchstaugliches und unterhaltsames System zu entwickeln werden.

Die Arbeit schließt mit einer Zusammenfassung der drei wichtigsten Herausforderungen bei der Planung und Durchführung von automotiven Evaluationsstudien und gibt Handlungsempfehlungen, wie diesen sowohl in der Automobilindustrie als auch in der universitären Forschung begegnet werden kann.

Abstract

New technologies are continually introduced to the concept of the vehicle. Entertainment electronics such as radio and CD players, as well as mobile phones, have ceased to be uncommon to the automotive field many years ago. The vehicle is experiencing a change from a dedicated means of transportation towards an information system. This was made clear on 15 November, 2008 when BMW introduced into the German market the first vehicle with an onboard internet connection [BMW 08].

In order to develop automotive systems, however, it is necessary to consider certain general constraints which arise from a special context of application different from desktop PCs and therefore pose special challenges to concept design, development, and evaluation. As an introduction to the subject of the thesis, these general constraints as well as the appropriate research questions will be explained. Subsequently, conventional methods of consumer surveys, usage studies and usability tests pertaining to human-computer interaction will be introduced, and the applicability, both theoretical and actual, of said methods to the automotive industry will be discussed. An exemplary lifecycle will show the practical application of theoretical methods. The focus will be on the data collection and evaluation techniques used to plan, conduct, and analyse two evaluation studies in the context of the development process of an automotive internet browser.

The first study was qualitative and aimed to identify the most promising interaction concepts of three browser variants, in order to transform them into a final design in the next iterative step. This analysis was conducted both stationary and driving. Since the real operational context of driving increased the usability problems experienced in stationary operation the second, quantitatively oriented evaluation study retained this aspect of the test design. The objective of the second study was to experimentally compare the optimised internet browser with a functionally scaled down basic variant. To this end the process to acquire test subjects focused on creating a homogenous user group. The results of this second study show that the iterative process used here, as well as the analysis of the system in its real context of usage were crucial in identifying usability problems and in recognising the priorities of users operating an automotive browser. This was necessary in order to develop a usable and entertaining system.

This thesis will conclude with a summary of the three most important challenges experienced in planning and conducting automotive evaluation studies and will recommend strategies to meet these challenges both in the automotive industry and in university research.

Inhaltsverzeichnis

1. Einleitung und Motivation.....	7
2. Rahmenbedingungen und Forschungsfragen bei der benutzerzentrierten Entwicklung im automotiven Kontext	11
2.1. Rahmenbedingungen für die Entwicklung im automotiven Kontext	11
2.2. Forschungsfragen in der Automobilindustrie.....	14
3. Benutzerbefragungen, Benutzungsstudien, Benutzbarkeitstests in der Automobilindustrie.....	17
3.1. Allgemeine Datenerfassungstechniken	19
3.2. Benutzungsorientierte Methoden.....	24
3.3. Benutzerorientierte Methoden der Anforderungsanalyse.....	30
3.3.1. Die Befragten	32
3.3.2. Interviews	34
3.3.3. Fokusgruppen	35
3.3.4. Fragebögen	38
3.3.5. Beobachtungen	39
3.3.6. Tagebuch-Methode.....	41
3.3.7. Gas Station Flash Survey.....	42
3.4. Benutzerorientierte Methoden in der Phase der Evaluation	44
3.4.1. Eine Evaluation planen.....	44
3.4.2. Evaluationstechniken zum richtigen Zeitpunkt einsetzen	46
Skizze	47
Lo-Fi-Prototyp.....	47
Hi-Fi-Prototyp	48
Fertiges Produkt	49
3.4.3. Probanden aus dem Probandenpool und externe Probanden.....	51
3.4.4. Forschungsfragen und erfasste Daten.....	53
3.4.5. Dual Task	56
3.4.5.1. Theoretische Modelle	57
Multi-Tasking GOMS	57
ACT-R.....	58
CarE.....	60
Zusammenfassung: theoretische Modelle	61

3.4.5.2. Studien im Fahrsimulator	61
3.4.5.3. Kontrolliertes Fahren.....	67
3.4.5.4. Fahren im Straßenverkehr	70
3.5. Rücklauf in der Automobilindustrie.....	72
4. Fallstudien	75
4.1. Fallstudie A: Benutzerbefragungen und Benutzertest im automotiven Kontext anhand von Methoden aus der Mensch-Computer Interaktion.....	76
4.1.1. Anforderungsanalyse.....	77
4.1.2. Evaluation.....	80
4.1.3. Herausforderungen und Einschränkungen	90
4.2. Fallstudie B: Benutzertest im automotiven Kontext.....	93
4.2.1. Das zu evaluierende System.....	93
4.2.2. Ziele, Hypothesen und Aufgaben.....	96
4.2.3. Testdesign und Auswahl der Probanden	101
4.2.4. Aufbereitung der Daten, Auswertung und Ergebnisse	104
Auswertung der qualitativen Daten	113
4.2.5. Herausforderungen und Einschränkungen	117
5. Zusammenfassung und Fazit	120
Fazit.....	123
6. Literaturverzeichnis.....	125
7. Abbildungsverzeichnis.....	134
8. Tabellenverzeichnis.....	136
9. Anhang	137

1. Einleitung und Motivation

In den letzten Jahren fanden immer mehr neue Technologien ihren Einzug in Automobilbranche. Dabei treffen verschiedenste Funktionen in einem Gerät zusammen. Heutzutage kann im Fahrzeug nicht mehr nur das Radio oder der CD-Player bedient werden. Vielmehr sind Navigationsgerät, das Einstellen von Komfort-Funktionen oder auch eine Fernseh-Applikation in moderne Infotainmentsysteme integriert. Zusätzlich bringt der Fahrer weitere, eigene Geräte wie iPod oder Mobiltelefon ins Fahrzeug, welche ebenfalls bedient werden müssen. Die Tendenz ist steigend.

Durch diese Vielzahl verschiedenartiger Funktionen und teilweise auch Interaktionskonzepte erhöht sich die Komplexität der Bedienung. Problematisch ist hierbei, dass es einen negativen Zusammenhang zwischen der Komplexität eines Systems und dessen Gebrauchstauglichkeit gibt. Bei steigender Komplexität sinkt zumeist die Gebrauchstauglichkeit des gesamten Systems, da der Benutzer den Überblick über sehr viele Funktionen behalten muss und sich gegebenenfalls auf verschiedene Interaktionskonzepte einlassen muss.

Viele Produkte liegen auf der Trendlinie zwischen niedriger Komplexität bei hoher Gebrauchstauglichkeit und hoher Komplexität bei niedriger Gebrauchstauglichkeit und nur einige wenige Produkte liegen unter- oder im Idealfall oberhalb dieser Linie (vgl. Abbildung 1.01) [Wiklund 94]. Das Ziel sollte es jedoch sein, trotz hoher Komplexität, die auf dem Vorhandensein großer Funktionalität begründet ist, eine hohe Gebrauchstauglichkeit zu erzielen. Dies ist jedoch nur möglich, wenn im gesamten Designprozess Benutzer involviert sind und Evaluationen auch schon frühzeitig durchgeführt werden [Wiklund 94].

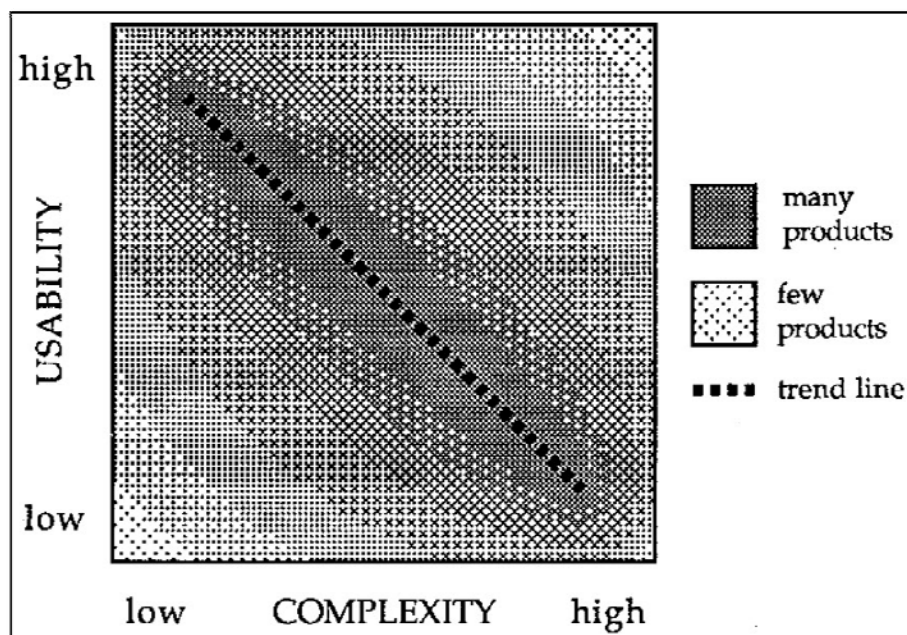


Abbildung 1.01: Das allgemeine Verhältnis zwischen Gebrauchstauglichkeit und Komplexität (Punkte entsprechen Produkten). (Quelle: [Wiklund 94])

In den vergangenen Jahren ging der Fokus zudem weg von der reinen Funktionalität und hin zu erhöhten Anforderungen an eine einfache und intuitive Bedienbarkeit [Hamacher 02]. In der Automobilindustrie spielt Kundenzufriedenheit eine bedeutende Rolle: Da das Kaufintervall in dieser Branche verhältnismäßig lang ist, gilt es primär, Kundenbeziehungen zu stabilisieren [Diez 04]. Dies gelingt jedoch nur bei zufriedenen Kunden, deren wachsende Loyalität sich in einer gesteigerten Wiederkauftrate ausdrückt. Somit ist es umso wichtiger, gebrauchstaugliche Systeme zu entwickeln.

Bei der Entwicklung interaktiver Systeme im automotiven Umfeld gilt es aufgrund des speziellen Anwendungsgebietes „Fahrzeug“ besondere Anforderungen zu betrachten. Neben einer teilweise eingegrenzten Zielgruppe wie beispielsweise dem Autofahrer, dem Beifahrer oder dem Fondpassagier sind auch die besonderen äußeren Bedingungen bei der Benutzung dieser Systeme zu beachten. Eine erhöhte Geräuschkulisse oder Erschütterungen während der Fahrt, sind nur zwei von vielen Umwelteinflüssen. Auch das Fehlen von präzisen Eingabegeräten wie der Maus, die aufgrund der Erschütterungen ohnehin ungeeignet wäre, stellt eine besondere Anforderung an die Entwicklung und Darstellungs- und Interaktionskonzepten für automotive Systeme dar.

Wird ein System für den Fahrzeugführer spielt zusätzlich der Aspekt der Sicherheit eine übergeordnete Rolle. Es gilt nicht nur den Fahrzeugführer selbst, sondern auch seine Mitfahrer sowie andere Verkehrsteilnehmer durch die Interaktion mit einem System nicht zu gefährden.

§23 der Straßenverkehrsordnung sieht aus diesem Grunde vor, dass weder Sicht, noch Gehör durch Geräte im Fahrzeug beeinträchtigt werden dürfen [StVO 09]. Daher ist es beispielsweise seit 2001 auf deutschen Straßen Pflicht, eine Freisprecheinrichtung für Mobiltelefone zu benutzen, damit die Verkehrssicherheit nicht aufgrund des Abnehmens und Haltens des Gerätes beeinträchtigt wird. Doch nicht nur die zusätzliche motorische Ablenkung stellt eine Gefahr dar. Auch die zusätzliche visuelle und kognitive Belastung muss minimal bleiben.

Neben seiner Funktion als Transportgerät übernimmt das Fahrzeug heutzutage mehr und mehr die Aufgaben eines multimodalen Informationssystems. Mit diesem Fortschritt in Bezug auf elektronische Geräte wächst jedoch auch die Informationsmenge im Fahrzeug und macht es nötig, neue Strategien für den Fahrer zu entwickeln, damit er mit dieser Informationsmenge umgehen kann [Ablassmeier 06]. Die kognitive Belastung sowie die Ablenkung des Fahrers müssen so niedrig wie möglich bleiben. Aus diesem Grund ist bei interaktiven Systemen für den Fahrzeugführer Effektivität und Effizienz besonders wichtig, da die Bedienung dieser Systeme neben der Primäraufgabe, dem Fahren, erfolgen muss. Die subjektive Zufriedenheit darf bei der Interaktion nicht vernachlässigt werden, doch müssen in erster Linie Mindestanforderungen in Bezug auf die Fahrzeugsicherheit erfüllt werden [Fragebogen 09] [Gespräche].

Die Entwicklung im automotiven Bereich wirft durch die oben genannten, besonderen Rahmenbedingungen auch sehr spezielle Forschungsfragen auf, die teilweise spezielle Befragungs-

und Evaluationsmethoden erfordern. Dies gilt bei der Entwicklung neuer Fahrzeuge genauso wie bei der Entwicklung neuer Technologien für das Fahrzeug. Benutzertests finden aus diesem Grunde nicht selten in einem Fahr Simulator oder realem Fahrzeug und unter dem Einbezug potentieller Benutzer in den Designprozess statt. Auf diese Weise können die Befragten den Kontext der Benutzung besser nachvollziehen und die produzierten Ergebnisse sind sehr nahe an der Realität.

Diese Arbeit entstand im Rahmen einer seit Oktober 2007 bestehenden Kooperation zwischen dem Lehrstuhl für Mensch-Computer Interaktion an der Universität Konstanz mit der Volkswagen AG. Ziel der Kooperation war die Entwicklung eines benutzerfreundlichen, innovativen Webbrowsers für den automotiven Gebrauch. Im Oktober 2007 wurde im Rahmen der Vorlesung „Visuelles Requirement Engineering für das Interaction Design“ [HCI KN 07] ein Pilotprojekt durchgeführt, welches das Potential von Webbrowsern im automotiven Kontext erfassen sollte. In diesem ersten Schritt wurden von drei Entwicklerteams relativ frei und unter Einfluss weniger Einschränkungen drei verschiedene Interaktions- und Darstellungskonzepte entwickelt. Ziel des Pilotprojektes war es, ein Browserkonzept für den Audi A8 zu entwickeln, welches vorrangig von Geschäftsleuten zum Zwecke ihrer Arbeit genutzt werden sollte. Der Fokus lag bei allen Teams neben dem Darstellungskonzept auf der Einbindung alternativer Eingabegeräte, die im Gegensatz zur bewegungsempfindlichen Maus, auch für den automotiven Bereich geeignet waren.

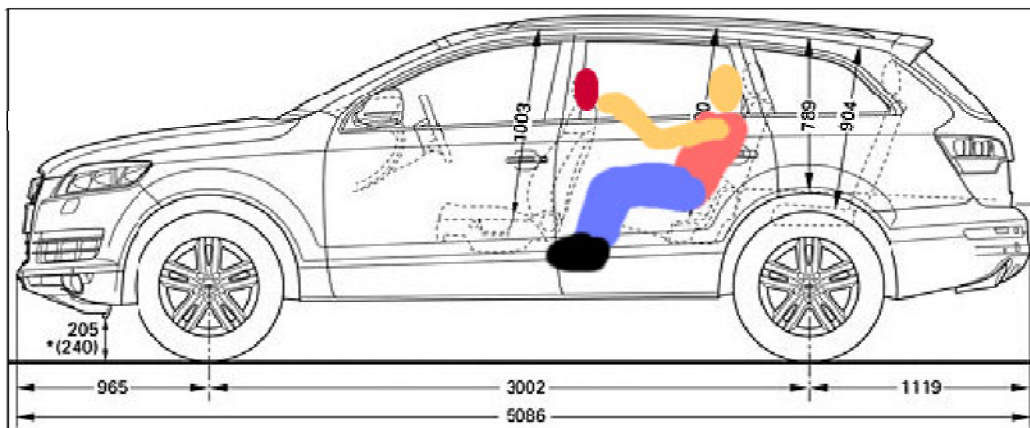


Abbildung 1.02: Audi Q7 (Quelle: [Audi 09a]) mit Screen für das RSE (rot) und potentiellm Benutzer des RSE.

Das Projekt wurde im Frühjahr 2008 zu Bachelor- und Masterprojekten erweitert und neue Rahmenbedingungen wurden definiert. Statt dem Audi A8 war nun der Audi Q7 Einsatzort des automotiven Browsers (vgl. Abbildung 1.02). Es mussten dabei erneut Anforderungen erhoben werden, um der neuen Zielgruppe gerecht zu werden und darüber hinaus wurden Evaluationen Bestandteil des Projekts. Das finale Browserkonzept entstand in einem iterativen und benutzer- und benutzungsorientierten Prozess und wurde mehrfach evaluiert.

Gegenstand dieser Arbeit ist es die Methoden der Benutzerbefragungen, Benutzungsstudien und Benutzbarkeitstests in der Automobilindustrie aufzuzeigen und an einem Fallbeispiel zu dokumentieren, wie diese in der Praxis eingesetzt werden können. Dies erfolgt anhand eines exemplarischen Lifecycle für die Entwicklung eines Darstellungs- und Interaktionskonzepts im automotiven Bereich. Der Fokus liegt hierbei auf den Methoden der Evaluation.

Im nachfolgenden Kapitel werden die besonderen Rahmenbedingungen automotiver Systeme aufgezeigt und mit den Rahmenbedingungen von Desktop-PCs verglichen. Die durch den besonderen Nutzungskontext vorgegebenen Forschungsfragen werden detailliert beschrieben. Zudem wird erklärt, weshalb besondere Methoden von Benutzerbefragungen, Benutzungsstudien und Benutzbarkeitstests in der Automobilindustrie benötigt werden.

In Kapitel 3 werden anschließend kurz Methoden und Werkzeuge vorgestellt, wie sie bei traditionellen Anforderungsermittlungsprozessen und Untersuchungen in der Mensch-Computer Interaktion Verwendung finden. Darüber hinaus wird hier aufgezeigt, inwieweit diese Methoden in der Automobilindustrie eingesetzt werden können und tatsächlich eingesetzt werden. Die Ergebnisse hierzu entstammen neben der Literaturrecherche und der Befragung von Mitarbeitern der Automobilindustrie auch den persönlichen Erfahrungen der Autorin, die im Laufe der Kooperation gesammelt wurden oder sich durch Gespräche ergeben haben.

Im vierten Kapitel wird schließlich anhand des Fallbeispiels „automotiver Webbrowser“ ein exemplarischer Lifecycle aufgezeigt, wie er im automotiven Kontext durchgeführt werden könnte um ein gut funktionierendes, gebrauchstaugliches System zu erhalten. Das Ziel dieses Kapitels ist es, anhand eines praktischen Beispiels aufzuzeigen, wie unter Einbezug der Methoden aus Kapitel 3 Evaluationen im automotiven Bereich durchgeführt werden können. Der Lifecycle beginnt in Kapitel 4.1. mit einer benutzer- und benutzungsorientierten Anforderungsanalyse, die an anderer Stelle detaillierter beschrieben ist und zur Entwicklung dreier verschiedener Browserkonzepte führte [Söter 08]. Eine Evaluation unter Einbezug potentieller Nutzer sollte hier helfen, die vielversprechendsten Konzepte zu extrahieren um sie im nächsten Iterationsschritt weiter zu verfolgen. Diese Konzepte sowie das aus den Ergebnissen dieser ersten Evaluation resultierende Darstellungs- und Interaktionskonzept werden in [Specht 10] erörtert.

Der in diesem Kapitel vorgestellte Lifecycle mündet schließlich in der Evaluation des finalen Prototyps, wie sie im Rahmen eines universitären Umfeldes unter Einbindung der Werkzeuge der Automobilindustrie stattfinden kann. Ihre Charakteristiken und Ergebnisse werden Kapitel 4.2. geschildert.

In Kapitel 5 werden schließlich die Haupt-Unterschiede zwischen Untersuchungen in der Automobilindustrie und in der Mensch-Computer Interaktion zusammengefasst und es wird gezeigt, wie den alternativen Herausforderungen in der universitären Forschung begegnet werden kann.

2. Rahmenbedingungen und Forschungsfragen bei der benutzerzentrierten Entwicklung im automotiven Kontext

Wie im vorigen Kapitel erwähnt, stellt der Nutzungskontext ‚Fahrzeug‘ besondere Anforderungen an die Entwicklung interaktiver Systeme. Im Gegensatz zur Entwicklung von Systemen für die Bedienung am Desktop-PC müssen einige besondere Rahmenbedingungen beachtet werden, die zu sehr speziellen Forschungsfragen führen. Diese wiederum erfordern zum Teil das Verwenden besonderer Methoden während der Anforderungsanalyse und der Evaluation. Die nachfolgenden Kapitel behandeln die Besonderheiten des automotiven Nutzungskontexts und die damit verbundenen Forschungsfragen. Die hierfür geeigneten Methoden werden in Kapitel 3 diskutiert.

2.1. Rahmenbedingungen für die Entwicklung im automotiven Kontext

Soll in einem Fahrzeug mit einem Informations- oder Unterhaltungssystem interagiert werden, führen neben dem System selbst noch weitere Faktoren zu Erfolg oder Misserfolg der Interaktion. Da das Automobil sich von einem normalen Raum stark unterscheidet, spielen gewisse externe Faktoren bei der Interaktion eine große Rolle. Tabelle 2.01 fasst diese Unterschiede zusammen.



Abbildung 2.01: Nutzungskontext Schreibtisch mit Desktop-PC.

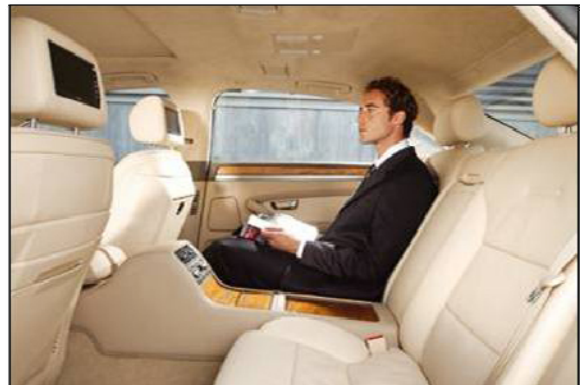


Abbildung 2.02: Nutzungskontext Fahrzeug mit Infotainment-/Rear-Seat-Entertainmentsystem.

Der Benutzer an einem Desktop-PC kann – je nach Art des zu entwickelndem Systems – einer sehr breiten Zielgruppe angehören. Ein handelsübliches Betriebssystem beispielsweise muss sowohl für Kinder, als auch für ältere Personen verständlich sein. Geht es jedoch um die Entwicklung eines Betriebssystems für ein automotives Infotainmentsystem, müssen sehr spezielle Zielgruppen betrachtet werden. Je nach Fahrzeugmarke oder –klasse müssen Personen eines bestimmten Alters, einer

bestimmten Gehaltsklasse oder mit bestimmten familiären Verhältnissen betrachtet werden. Insbesondere rücken bei der automotiven Entwicklung jedoch Fahrer in den Mittelpunkt.

	Nutzungskontext Schreibtisch	Nutzungskontext Fahrzeug
Zielgruppe	Je nach System: nahezu beliebig.	Je nach System: Fahrer, Beifahrer, Passagier im Fondbereich.
Benutzer	Flexibel.	In der Bewegung eingeschränkt.
Umgebung	Zimmer / Büro. Stationär. Viel Raum.	Fahrzeug. Mobil. Eingeschränkter Raum.
Externe Einflüsse (Umwelt)	Etwas variable Lichtverhältnisse.	Witterung, Beschaffenheit der Straße, Ampeln, Erschütterungen, stark variable Lichtverhältnisse, Natur & Sehenswürdigkeiten
Externe Einflüsse (Personen)	Telefonate, Besuche.	Mitfahrer, Telefonate, Straßenverkehr, Passanten.
Aufgaben	Interaktion mit dem System.	Interaktion mit dem System. Im Falle des Fahrers: zusätzliche Fahraufgabe.
Eingabegeräte	Nahezu beliebig, mobil.	Eingeschränkt, fest verbaut.
Technologisches Umfeld	Weitere Monitore, Lautsprecher, Telefon, alternative Eingabemöglichkeiten in beliebiger Anordnung.	Fahrerassistenzsystem, Navigationssystem, Radio, CD-/mp3-player, Telefon. Größtenteils fest verbaut.
Geräuschkulisse	CD-/mp3-Player/Radio, Telefon, Besucher.	CD-/mp3-Player/Radio, Telefon, Navigationssystem, Motorengeräusche, Mitfahrer, Straßenverkehr.
Displaygröße	Beliebig groß. Typischerweise 17 bis 24 Zoll	Eingeschränkt, eher klein. Typischerweise 10 bis 11 Zoll.
Kontext der Nutzung	Arbeit oder Freizeit.	Freizeit.

Tabelle 2.01: Unterschiede der Nutzungskontexte Desktop-PC und Automobil

Bei Entertainmentsystemen, welche nicht für den Fahrer, sondern für den Beifahrer oder die Passagiere im Fond entwickelt werden, gilt es, eben diese ins Zentrum der Beobachtung zu rücken. Diese Zielgruppen hängen jedoch auch vom Fahrer selbst ab: soll beispielsweise ein interaktives System für den Fondbereich eines Kombi entwickelt werden, würde dieses höchstwahrscheinlich von

den Kindern des Fahrzeugbesitzers bedient werden. Je nach Eigenschaften des Besitzers haben diese beispielsweise ein bestimmtes Alter. Diese detaillierten Informationen führen zu einer sehr genauen Kenntnis der jeweiligen Zielgruppen. Somit können Personen, die für Benutzerbefragungen, Benutzungsstudien und Benutzbarkeitstests in Frage kommen, gezielt ausgewählt werden.

Ein weiterer Unterschied der beiden Nutzungskontexte ist die Mobilität des Benutzers im Vergleich zur Mobilität der Umgebung. Der Benutzer an einem Desktop-PC kann sich bei Bedarf frei bewegen, doch der Ort der Interaktion – zumeist ein Zimmer oder Büro – bleibt unveränderlich. Im Gegensatz dazu kann sich der Benutzer eines automotiven Infotainmentsystems während der Fahrt nicht von seinem Platz fortbewegen, wohingegen der Ort an welchem er interagiert – das Fahrzeug – sich stets in Bewegung befindet. Dieser Umstand spielt bei der Entwicklung eine gewichtige Rolle. Das interaktive System am stationären Desktop-PC kann mit nahezu beliebigen Eingabegeräten bedient werden, die durch die Unbeweglichkeit des Umfeldes eine sehr hohe Präzision ermöglichen. Eine Interaktion mit Tastatur, Maus, Trackball oder Trackpad sind sehr gut geeignet. Doch auch eine blickgesteuerte oder sprachbasierte Steuerung ist denkbar. Die Präzision einer Maus am Desktop-PC lässt sich im fahrenden Auto jedoch nicht reproduzieren, da hier viele externe Faktoren zu einem unruhigen Interaktionskontext führen. Je nach Beschaffenheit der Straße kann die Interaktion von mehr oder weniger Erschütterungen beeinträchtigt werden. Auch Witterungsverhältnisse oder der Straßenverkehr haben einen Einfluss auf das Maß an Bewegung innerhalb eines Fahrzeugs. So wird die Interaktion auch durch sich stauenden Verkehr und deshalb ständig wechselndem Beschleunigen und Abbremsen beeinträchtigt. Durch diese Erschütterungen während der Fahrt sind die meisten der bekannten Eingabemodalitäten aus dem Desktop-Kontext nicht geeignet. Doch nicht nur aufgrund der mangelnden Präzision sind einige dieser Geräte im Fahrzeug unbrauchbar. Auch die Tatsache, dass diese Geräte bei einer starken Bremsung unter den Sitz fallen könnten und darüber hinaus die Sicherheit der Fahrzeuginsassen gefährden könnten, verbietet ihre Nutzung. Deshalb sollten sowohl Ein- als auch Ausgabegeräte im Fahrzeug fest montiert sein. Dieser Aspekt führt in Fahrzeug auch zu einer nahezu statischen Anordnung der zu bedienenden Geräte. Im Gegensatz zum Desktop-PC kann der Monitor nicht verrückt oder das Telefon an eine andere Stelle gelegt werden, da diese Geräte bzw. deren Halterungen einen festen Platz im Fahrzeuginnenleben haben. Trotz dieser durch die Umgebung vorgegebenen Einschränkungen an die Anordnung der Ein- und Ausgabegeräte, muss die Interaktion komfortabel bleiben und dem Benutzer ein gewisses Maß an Unterhaltung bieten.

Eine weitere Herausforderung stellt die Größe und die Platzierung der Ausgabegeräte dar. Während Monitore am Schreibtisch nahezu beliebig groß sein können und zudem nach Belieben platziert werden können, müssen im automobilen Bereich gewissen Restriktionen beachtet werden, die beispielsweise durch den mangelnden Platz auferlegt werden. Doch auch auf kleinen Displays müssen die benötigten Informationen – sowohl bei Systemen für den Fahrer, als auch für den Mitfahrer – ausreichend lesbar dargestellt werden.

Ein weiterer unterscheidender Faktor sind die im Gegensatz zum Zimmer veränderlichen Lichtverhältnisse während der Fahrt. Dies spielt nicht nur beim natürlichen abendlichen Dunkelwerden eine Rolle, sondern auch bei der Fahrt durch Tunnels oder beim Auftreten von plötzlichen Nebelschwaden oder Regenwolken. Das Display automotiver Systeme muss demnach auch unter veränderlichen Umgebungsbedingungen gut sichtbar sein und dem Benutzer die benötigten Informationen vermitteln können. Darüber hinaus sollte die Interaktion mit dem System trotz störender Umgebungsgeräusche wie Motoren- oder Straßenlärm, dem Prasseln von Regen auf dem Autodach oder Gesprächen der weiteren Mitfahrer nicht beeinträchtigt werden.

Oben genannte Rahmenbedingungen treffen sowohl für den Fahrer, als auch für die Mitfahrer zu. Soll das automotive System nur für den Fahrzeugführer entwickelt werden, spielen neben diesen Faktoren auch diverse weitere Aspekte eine gewichtige Rolle. Die Interaktion muss nach wie vor in erster Linie subjektiv zufriedenstellend gestaltet sein, doch in diesem Fall müssen auch einige Kriterien der Sicherheit erfüllt werden. Die Interaktion muss sekundär bleiben und darf die Aufmerksamkeit des Fahrers nicht zu stark von seiner Primäraufgabe, dem Fahren eines Fahrzeugs, ablenken. Je nach Art und Menge des Verkehrs ist die kognitive Belastung des Fahrers ohnehin hoch und darf durch die zusätzliche Interaktion nicht erhöht werden. Sowohl Augen, als auch Ohren müssen in diesem Fall auf den Straßenverkehr gerichtet sein und die Hände sollten nicht unnötig lange vom Lenkrad entfernt werden. Die Straßenverkehrsordnung [StVO 09] und beispielsweise auch die Alliance of Automobile Manufacturers (kurz: AAM) [AAM 02] legen einige Prinzipien fest, die in Bezug auf die Sicherheit Mindestanforderungen an ein automotives System stellen.

2.2. Forschungsfragen in der Automobilindustrie

Die oben genannten Rahmenbedingungen werfen spezielle Forschungsfragen auf, die für die Automobilindustrie zum Teil eine alternative Herangehensweise bei der Durchführung von Benutzerbefragungen, Benutzungsstudien und Benutzbarkeitstests erfordern, als es in der traditionellen Mensch-Computer Interaktion der Fall ist.

Neben effektiver, effizienter und subjektiv zufriedenstellender Bedienung, sind auch Aspekte der Sicherheit, des Rechts, der Handhabung und Platzierung der Ein- und Ausgabegeräte und auch der Markenidentität von großer Bedeutung.

In den vergangenen Jahren nahm die Komplexität der fahrzeugeigenen und zusätzlichen Geräten, Infotainmentsystemen, Fahrerassistenzsysteme sehr stark zu [Müller & Fiedland 09], was zu einer höheren Belastung des Fahrzeugführers führte. Das zusätzliche Ausführen sekundärer Aufgaben beim Führen eines Fahrzeuges beeinflusst immer die Fahrleistung und kann durchaus auch Unfälle nach sich ziehen [Bach et al. 08]. Aus diesem Grund sind in diesem Anwendungskontext besondere Konzepte gefragt, die die Sicherheit des Autofahrers sowie der anderen Verkehrsteilnehmer gewährleisten. Evaluationen automotiver Systeme sind unumgänglich.

Da das Steuern des Fahrzeugs die Primäraufgabe bleiben muss, muss bei der Entwicklung automotiver Systeme die doppelte Belastung durch geteilte Aufmerksamkeit sowie die Ablenkung des Fahrers berücksichtigt und gemessen werden. Die Interaktion muss schnell und unter geringer kognitiver Belastung erfolgen. Bei Untersuchungen solcher Systeme ist aus diesem Grunde darauf zu achten, dass das Maß der Ablenkung durch die Sekundäraufgabe minimal ist. Die Sekundäraufgabe sollte einen möglichst geringen Anteil der Aufmerksamkeit des Fahrers beanspruchen, sodass sich dieser weiterhin den Gegebenheiten des Straßenverkehrs anpassen kann.

Darüber hinaus sollte die Fehlerrate bei der Bedienung eines interaktiven Systems so gering wie möglich sein, damit der Frustrations- und Erregungszustand des Interagierenden möglichst gering bleibt (vgl. [Gruenstein 09]) und das System dessen Fokus nicht aus Gründen der Fehlerbehebung von der Fahraufgabe ablenkt. Bei Untersuchungen im automotiven Sektor gilt es daher, Faktoren wie die Blickabwendung, Spurabweichung, Fahrfehler, Bedienzeiten, Reaktionszeiten und die Unterbrechbarkeit und Wiederaufnahme der Interaktion zu ermitteln, da diese einen großen Einfluss auf den oben genannten Aspekt der Sicherheit haben [Ablassmeier et al. 06] [Fragebogen 09].

Oftmals werden in Untersuchungen auch verschiedene Varianten eines Systems gegeneinander getestet, um zu erforschen, welches die beste Performance bei gleichzeitig höchsten Ansprüchen in Bezug auf die Sicherheit liefert (z.B. [Ablassmeier et al. 06] [Fragebogen 09] [Kern et al. 09] [Minin 08] [Wäller 08]).

Doch neben der Sicherheit spielt auch in der Automobilindustrie die subjektive Zufriedenheit mit dem System eine sehr große Rolle. Selbst wenn ein System alle Anforderungen an die Sicherheit erfüllt und in Bezug auf die Performance hervorragende Werte liefert, wird es nur dann Einzug in das Fahrzeug finden, wenn es von seinen Benutzern als gebrauchstauglich und subjektiv zufriedenstellend empfunden wird [Fragebogen 09]. Aus diesem Grunde müssen auch bei Untersuchungen in der Automobilindustrie qualitative Daten wie Bedienungsprobleme, Ausmaß der Belastung oder der Zufriedenheit, oder auch Verbesserungsvorschläge erfasst werden.

Auch bei interaktiven Systemen für die Mitfahrer müssen einige dieser Aspekte, wie beispielsweise Bedienzeiten, Fehlerzahl oder die Unterbrechbarkeit der Benutzung untersucht werden – der Aspekt der Sicherheit in Form von Spurabweichung oder Fahrfehlern spielt in diesem Kontext jedoch keine Rolle. Für diese Zielgruppe sind Performancedaten in einem kleineren Ausmaß wichtig. In diesem Fall ist es von höherer Bedeutung, ein hohes Maß an Unterhaltung zu gewährleisten.

Eine weitere Besonderheit von Technologien, die für die Automobilindustrie entwickelt werden, sind spezielle Ein- und Ausgabegeräte. Während bei der Interaktion mit Systemen am Schreibtisch sehr präzise Eingabegeräte, wie beispielsweise die Maus verwendet werden können, ist dies im fahrenden Fahrzeug nicht ohne Weiteres möglich. Nicht nur Erschütterungen während der Fahrt, sondern auch die Tatsache, dass die Hände des Fahrzeugführers beim Steuern des Fahrzeugs möglichst lange am

Lenkrad bleiben sollten, erfordern alternative Eingabemedien. Manche Automobilhersteller haben diese in Form von Dreh-Drückstellern, Spracheingabe oder gar multimodalen Eingabemöglichkeiten gefunden. Diese Modalitäten ermöglichen es, dass die Augen und Hände des Interagierenden so wenig wie möglich beansprucht werden. Für den Fahrer sind solche Eingabeinstrumente ideal, da seine Augen auf dem Verkehr und seine Hände am Lenkrad bleiben können.

Auch Mitfahrer benötigen Eingabegeräte, die ohne Blickzuwendung bedient werden müssen, da im Kontext eines Fahrzeugs Ein- und Ausgabegerät zumeist weiter voneinander entfernt liegen, als bei einem Notebook oder handelsüblichen Desktop-PC. Das passende Eingabemedium sowie die passende Interaktionsart für den jeweiligen Anwendungsbereich zu finden, stellt im automotiven Umfeld also eine große Herausforderung dar. Oftmals müssen hier mehrere Möglichkeiten in Bezug auf ihre Gebrauchstauglichkeit und Performance gegeneinander getestet werden.

3. Benutzerbefragungen, Benutzungsstudien, Benutzbarkeitstests in der Automobilindustrie

In der traditionellen Mensch-Computer Interaktion liegt der Fokus auf dem Benutzer und seiner Interaktion mit einem System auf einem PC, Mobiltelefon oder Ähnlichem. Damit ein solches System erfolgreich ist, sollte es einen iterativen Entwicklungsprozess durchlaufen. Die Analyse von Anforderungen vor der Entwicklungsphase sowie Evaluationen während und nach der Entwicklung gehören idealerweise zu diesem Prozess und helfen dabei, das System gebrauchstauglich und benutzerfreundlich zu gestalten. Es gibt zu diesem Zwecke in der traditionellen Mensch-Computer Interaktion verschiedene Datenerfassungstechniken, die sich bei der einen oder der anderen Zielsetzung besser eignen.

In der Automobilindustrie herrschen jedoch, wie im vorigen Kapitel beschrieben, stark unterschiedliche Rahmenbedingungen, die sehr spezielle und vielfältige Anforderungen an die Entwicklung automotiver Interaktionskonzepte stellen. Die eingeschränkte Zielgruppe und die mobile Interaktionsumgebung sind nur zwei von vielen einflussnehmenden Faktoren. Legale und sicherheitstechnische Aspekte sind jedoch die Faktoren mit der größten Bedeutung und erfordern aus diesem Grunde eine besondere Betrachtung. Benutzertests für interaktive Systeme müssen deshalb neben den klassischen Anforderungen an Usability- und User Experience auch diese Aspekte ermitteln. Dies ist nur mittels besonderer Testdesigns möglich, die ab Kapitel 3.3. beleuchtet werden.

Doch schon während der Anforderungsanalyse müssen gewisse Faktoren beachtet werden, die in der traditionellen Mensch-Computer Interaktion eine geringere Bedeutung spielen. Einerseits ist die Zielgruppe, für die ein interaktives, automotives System entwickelt wird, sehr klar definiert. Wird ein solches System für den Fahrzeugführer entwickelt, so sind nur seine Bedürfnisse von Belang und der Aspekt der Ablenkung und somit der Sicherheit erfährt eine hohe Bedeutung. Wird das System hingegen für Beifahrer oder Passagiere der Rückbank entwickelt, spielt die Ablenkung eine Nebenrolle und andere Faktoren gelangen in den Mittelpunkt.

Neben diesen Faktoren spielt jedoch auch der Automobilhersteller eine besondere Rolle. Oftmals sind beispielsweise nicht nur Fahrer im Allgemeinen die Zielgruppe einer Entwicklung, sondern Fahrer einer bestimmten Fahrzeugmarke oder einer bestimmten Gehaltsklasse. Der Hersteller bringt jedoch auch noch weitere wichtige Faktoren in den Entwicklungsprozess mit ein: beispielsweise muss das System trotz aller sicherheitstechnischen und Usability-Aspekte dem Persönlichkeitsbild der Marke entsprechen und oftmals müssen Untersuchungen unter bestimmten Richtlinien der Geheimhaltung durchgeführt werden, die ein freies Entwickeln und Testen einschränken.

Die aus den vielen Besonderheiten resultierenden Forschungsfragen erfordern in der Automobilindustrie oftmals eine alternative Herangehensweise und machen die Verwendung besonderer Methoden notwendig. Diese unterscheiden sich teilweise nur in geringem Maße von den

traditionellen Methoden der Mensch-Computer Interaktion, doch teilweise sind auch gänzlich neue Methoden erforderlich.

In diesem Kapitel werden die Methoden der klassischen Mensch-Computer Interaktion im Allgemeinen beschrieben und es wird herausgearbeitet, wie sie sich in den Phasen der Anforderungsanalyse und der Evaluation einbinden lassen. Zusätzlich wird aufgezeigt, welche Methoden auch im automotiven Umfeld Anwendung finden können und welche hierfür weniger geeignet sind. Die Gemeinsamkeiten und Unterschiede bei der Anwendung im klassischen und automotiven Kontext werden erläutert.

In dieses Kapitel fließen neben Informationen aus einer umfassenden Literaturrecherche zusätzlich auch persönliche Erfahrungen aus dem Laufe des Kooperationsprojekts, sowie Gespräche mit Personen aus der Automobilindustrie ein. In der Automobilindustrie werden aufgrund von Geheimhaltungsrichtlinien nur selten konkrete Beispiele von Anforderungsanalysen oder gar Evaluationen publiziert. Die Anwendungsbeispiele der nachfolgenden Kapitel entstammen aus diesem Grunde hauptsächlich aus Forschungsk Kooperationen der Automobilindustrie mit beispielsweise einer Universität oder allgemeinen universitären Forschungsprojekten mit automotivem Hintergrund. Handelt es sich bei den Projekten um Kooperationen, wird zumeist nicht erwähnt, mit welchem Automobilhersteller hier zusammengearbeitet wurde. Um dennoch einen Überblick über die Methoden in der Automobilindustrie in ihrer praktischen Anwendung zu erhalten, wurde von der Autorin dieser Arbeit ein Fragebogen entwickelt, welcher an Personen geschickt wurde, die in der Automobilindustrie tätig sind und hier Benutzerbefragungen, Benutzungsstudien und/oder Benutzbarkeitstests durchführen.

Der Fragebogen umfasst 29 Fragen die in die drei Teile Anforderungsanalyse, Evaluationen und Probanden eingeteilt sind. Im ersten Teil sollten die Befragten angeben, welches die wichtigsten Quellen für neue Produkte sind. Auf diese Weise sollte zunächst ermittelt werden, inwieweit potentielle Benutzer in den Anforderungsermittlungsprozess eingebunden sind. Die nachfolgenden Fragen bezogen sich auf die Methoden, mit Hilfe derer Anforderungen erhoben und Daten erfasst werden. Das Ziel war es herauszufinden, ob bereits in dieser Phase der Produktentwicklung Methoden angewandt werden, die sich von den klassischen Methoden der Mensch-Computer Interaktion unterscheiden. Eine Frage zu den Charakteristiken der Befragten sollte überprüfen wie die Probanden jeweils ausgewählt werden. Die Autorin dieser Arbeit hat im Laufe der Kooperation erfahren, dass hauptsächlich konzerninterne Personen in den Entwicklungsprozess mit einbezogen werden. Diese Frage sollte überprüfen, ob dies schon in der Phase der Anforderungsermittlung der Fall ist.

Der Evaluations-Teil des Fragebogens sollte herausstellen, in welchen Stadien der Produktentwicklung auf welche Weise evaluiert wird. Hierzu sollten beispielhaft einige Forschungsfragen genannt werden, um den jeweiligen Sachverhalt verdeutlichen. Weitere Fragen

wurden zu der Art der erhobenen Daten gestellt und zu den besonderen Methoden, die in dieser Phase zur Erhebung angewandt werden. Das Ziel war es auch hier zu überprüfen, ob andere als die aus der klassischen Mensch-Computer Interaktion bekannten Methoden Anwendung finden. Anschließend sollten die Befragten beantworten, wie der Rückfluss der Ergebnisse einer Evaluation in der weiteren Produktentwicklung von statten geht. Diese Frage war insofern wichtig, weil die Autorin überprüfen wollte ob der im Rahmen der Kooperation durchgeführte Lifecycle (vgl. Kapitel 4) auch für die Automobilindustrie typisch ist.

Der letzte Block des Fragebogens beschäftigte sich schließlich mit den Probanden, die zu Evaluationen in der Automobilindustrie eingeladen werden. Mit diesen Fragen sollte ermittelt werden, in welchem Maße konzerninterne und externe Personen eingebunden werden und welche Eigenschaften diese Personen teilen.

Dieser Fragebogen wurde mit Hilfe der Onlinesoftware QuestionPro (<http://www.questionpro.com/>) zusammengestellt und der Link wurde per E-Mail an sechs Personen geschickt, die in der Automobilindustrie tätig sind oder es waren. Diese wurden darum gebeten, den Fragebogen auszufüllen und ihn an Kollegen weiterzuleiten, die sich ebenfalls mit der Thematik beschäftigten. Von 22 Personen, die den Link zum Fragebogen öffneten, beantworteten ihn acht vollständig während weitere zwei Personen zumindest die Fragen zum Thema Anforderungsanalyse bearbeiteten. Diese beiden Fragebögen wurden jedoch trotzdem in die Auswertung mit einbezogen. Die Ergebnisse der Befragung sind in den nachfolgenden Kapiteln zu finden. Der Fragebogen selbst ist auf der DVD im Anhang zu finden.

3.1. Allgemeine Datenerfassungstechniken

Es gibt verschiedene Möglichkeiten, Daten im Rahmen einer Anforderungsanalyse oder einer Evaluation zu erfassen. Einige davon benötigen die Unterstützung von potentiellen Benutzern, die den Untersuchungsleiter ihre Ansichten und Meinungen zur Verfügung stellen. Andere Methoden sind nicht benutzerzentriert und bedürfen der Bearbeitung durch Experten. Diese Methoden werden in Kapitel 3.2. beschrieben. Der Fokus von Kapitel 3 liegt jedoch auf Techniken, die den potentiellen Benutzer in einem hohen Maße involvieren.

Sowohl in der Phase der Anforderungsanalyse, als auch für eine Evaluation gilt: bevor man mit der Untersuchung beginnt, muss ein Plan aufgestellt werden. Dieser Plan beinhaltet zunächst das Ziel einer Untersuchung [Sharp et al. 07]. Ein Ziel zu Beginn des Entwicklungsprozesses könnte beispielsweise sein, die Anforderungen an ein bestimmtes System zu ermitteln. In späteren Phasen könnte das Ziel einer Untersuchung sein, zu ermitteln, ob man mit den aktuellen Ideen und einem mittlerweile erstellten Prototypen auf dem richtigen Weg ist. Sobald die Ziele gesteckt sind, muss festgestellt werden, welche Art von Daten benötigt werden, um die jeweilige Fragestellung zu

bearbeiten und welche Methoden angewandt werden müssen, um diese Daten zu beschaffen [Sharp et al. 07].

Der nächste Schritt einer Untersuchung ist es, ein Dokument zu verfassen, welches den Befragten vor der Durchführung der Untersuchung vorgelegt wird. In diesem Dokument wird den Befragten erklärt, auf welche Art Daten erfasst werden und wie diese Daten verwendet werden. Zusätzlich sollte dem Befragten zugesichert werden, die Untersuchung beenden zu können, falls er sich nicht mehr wohl dabei fühlt. Ein solches Dokument ist sowohl für die Untersucher-, als auch für die Befragtenseite wichtig. Dem Untersucher sichert es zu, die aufgezeichneten Daten für seine Studie verwenden zu dürfen und dem Befragten sichert es zu, dass seine Daten nicht für andere Zwecke verwendet werden [Sharp et al. 07].

Als ein letzter Schritt vor der tatsächlichen Untersuchung, ist es sinnvoll, eine Pilotstudie durchzuführen. Diese kann helfen, noch vorhanden Problemstellen zu identifizieren. Diese können sowohl die Funktionalität der Hardware und Software, wie beispielsweise der Aufzeichnungsgeräte oder des zu testenden Systems aufdecken [Sharp et al. 07]. Ebenso kann mit Hilfe einer solchen Pilotstudie überprüft werden, inwieweit der Testaufbau stimmig ist, ob die Testaufgaben oder Interviewfragen verständlich sind und ob der vorgegebene zeitliche Rahmen angemessen ist und nicht überschritten wird.

Klassische Untersuchungen in der Mensch-Computer Interaktion finden in einer kontrollierten Umgebung statt, beispielsweise einem Usability Labor. Ein solches Labor kann das natürliche Umfeld der Benutzung mit wenig Aufwand nachbilden und die teilnehmende Person kann ohne externe, störende Einflüsse befragt werden oder mit einem zu testenden System interagieren. Für gewisse Bereiche aus der automotiven Benutzbarkeits-Forschung eignen sich solche Labore auch. Insbesondere in den frühen Stadien der Produktentwicklung finden Usability Labore Einsatz. Die hier vorherrschenden Bedingungen sind jedoch mit einer Studie im realen, automotiven Nutzungsumfeld nicht zu vergleichen, da in der Laborumgebung natürliche Distraktoren wie der Straßenverkehr oder Erschütterungen nicht vorhanden sind.

Es gibt sowohl in der traditionellen Mensch-Computer Interaktion, als auch in der Automobilindustrie verschiedene Methoden für die Erfassung der Benutzerdaten. Jede birgt im jeweiligen Kontext ihre eigenen Vor- und Nachteile. Kombiniert man bei einer Untersuchung mehrere Methoden, ermöglicht dies das Erfassen von mehreren Perspektiven, was die gefundenen Ergebnisse festigen kann [Sharp et al. 07].

Eine der einfacheren Methoden der Datenerfassung ist das Notieren von Aussagen mittels Papier und Bleistift. Diese Methode ist für den Befragten weniger aufdringlich als beispielsweise eine Videokamera und auch weniger störend, als würde mittels Laptop mit klickender Tastatur protokolliert werden [Sharp et al. 07]. Allerdings kann es bei dieser Methode dazu kommen, dass verschiedene Protokollanten durch ihre verschiedenen Fokusse bei der selben Befragung zu verschiedenen Notizen



kommen. Unerfahrene Protokollanten könnten zudem relevante Informationen übersehen, da sie nicht wissen, worauf sie beim Notieren achten müssen. Darüber hinaus fehlt es bei dieser Art der Datenerfassung an visuellen Daten, wie beispielsweise dem Gesichtsausdruck des Befragten, während er eine Frage beantwortet. Auch solche Daten können Aufschluss über die Ansichten des Befragten geben, können jedoch mittels handschriftlicher Notizen oftmals nicht festgehalten werden. Eine weitere Einschränkung dieser Methode ist das Problem von überhörten Aussagen, da es recht schwierig ist, gleichzeitig Fragen zu stellen, zuzuhören und die Antworten zu notieren. Zudem könnte es für den Befragten irritierend sein, wenn manche seiner Aussagen, die er selbst für wichtig erachtet, vom Protokollanten nicht aufgeschrieben werden [Sharp et al.07a]. Vorteilig bleibt jedoch die Einfachheit, da mit einem Papier und Notizblock bereits alle benötigten Materialien vorhanden sind. Außerdem können für eine erste Analyse der Ergebnisse, die notierten Daten schnell und einfach ausgewertet werden.

Für Studien im Labor ist diese Methode aufgrund der genannten Vorteile sehr gut geeignet, sollte aber im Idealfall mit anderen, verlässlicheren Techniken kombiniert werden. Auch in der Automobilindustrie wird diese Art der Datenerfassung geschätzt und wird daher sehr oft gewählt. Acht der zehn Befragten nutzen diese Methode alleine oder in Kombination mit anderen Methoden. Grund für die häufige Nutzung ist das Verhältnis von Aufwand zu Nutzen und die hohe Effizienz, die diese Methode mit sich bringt [Fragebogen 09]. Die so erfassten Daten können zudem sehr schnell und einfach ausgewertet werden und eignen sich hervorragend zur Gewinnung eines ersten Überblicks.

Bei Untersuchungen am Tisch oder Usability Labor ist diese Art der Datenerfassung unbedingt zu empfehlen. Etwas problematischer wird sie jedoch im Kontext eines Fahrsimulators, in welchem der Protokollant eine stabile Unterlage benötigt, um saubere Notizen zu machen. Doch selbst dann wird die Schrift weniger leserlich als beim Protokollieren an einem Tisch im Labor. Bei mobilen Studien ist dies noch schwieriger, da durch die Erschütterungen während der Fahrt eine saubere Handschrift kaum möglich ist. Wird das Protokoll zudem von unterschiedlichen Personen geführt und ausgewertet, erhöht dies die genannte Problematik sodass diese Art der Datenerfassung hier noch weniger geeignet ist.



Eine Alternative zur handschriftlichen Datenerfassung ist das Aufzeichnen von Audiodaten mittels Diktiergerät oder Ähnlichem. Diese Methode ist im Gegensatz zu Videoaufzeichnungen ebenfalls wenig aufdringlich und ermöglicht durch seine Portierbarkeit eine höhere Mobilität. Zudem kann – im Gegensatz zu den handschriftlichen Notizen – dem Befragten mehr Aufmerksamkeit geschenkt

werden, da während dem Zuhören nicht gleichzeitig mitgeschrieben werden muss. Nachteilig bleibt weiterhin das Fehlen visueller Daten. Audiodaten haben den Vorteil, dass sie vollständiger und somit zuverlässiger sind als handschriftliche Notizen. Des Weiteren sind sie eine gute Erinnerungshilfe, falls im späteren Verlauf des Entwicklungsprozesses etwas unklar sein sollte. Für eine Auswertung können die gesamte Aufnahme oder aber nur relevanten Passagen transkribiert werden.

Diese Methode ist ebenfalls für Laborstudien sehr gut geeignet, kann jedoch im automotiven Kontext Schwierigkeiten bergen. Auch im Umfeld eines Labors können durch beispielsweise ein offenes Fenster oder einen Ventilator Hintergrundgeräusche mit aufgezeichnet werden, die teilweise das Gesprochene übertönen. Das Fenster kann jedoch geschlossen, die Drehzahl des Ventilators kann vermindert werden. Bei Studien im Fahrzeug sind die Geräuschquellen noch vielfältiger und können im Gegensatz zum Laborumfeld nicht so einfach kontrolliert werden. Im stehenden Fahrzeug kann zwar die eventuell angeschaltete Klimaanlage oder Heizung ausgeschaltet werden, was jedoch die Testumgebung gegebenenfalls unangenehmer machen würde. Die Geräusche, die im fahrenden Fahrzeug beispielsweise durch den Motor entstehen, können jedoch nicht verhindert werden.

Wohl aus diesem Grunde scheint diese Methode während der Anforderungsanalyse in der Automobilindustrie wenig Anklang zu finden. Sie wird zwar von sechs der zehn befragten Personen verwendet, jedoch nur als zusätzliches Mittel in Kombination mit anderen Methoden der Datenerfassung [Fragebogen 09].

Eine ganzheitliche, aber etwas aufdringlichere Methode der Datenerfassung stellt die Videoaufzeichnung dar. Die Herausforderung hierbei besteht darin, die Kamera so unauffällig zu



platzieren, dass sie dem Befragten nicht während der gesamten Untersuchung auffällt. Dadurch soll die Aufdringlichkeit, die diese Methode per se mit sich bringt, gemindert werden. Eine Videokamera muss vor Beginn der Untersuchung justiert werden, sodass der gewünschte Bereich komplett aufgezeichnet wird. Umgebungsgeräusche können hier, genauso wie bei einer Audioaufzeichnung, störend sein. Insbesondere gilt dies auch

für die Verwendung in automotiven Studien. Durch das Vorhandensein visueller Daten können jedoch Silben, die eventuell von Nebengeräuschen übertönt wurden, anhand der Lippenbewegung erkannt werden und die Stimmung des Befragten kann – sofern dies nicht über die Tonlage geschehen kann – anhand seines Gesichtsausdrucks festgestellt werden. Videoaufzeichnungen erfreuen sich deshalb in der Automobilindustrie großer Beliebtheit. Insbesondere während der Anforderungsanalyse werden sie häufig eingesetzt. Acht der zehn Befragten nutzen diese Methode, da sie den Vorteil haben, die Situation ganzheitlich zu erfassen und dem Untersuchungsleiter die Möglichkeit bieten, im Nachhinein nach weiteren Informationen zu suchen [Fragebogen 09].

Eine Videokamera hat jedoch den Nachteil, dass sie komplexer zu bedienen ist, als Papier und Bleistift oder ein Diktiergerät. Insbesondere in der Hektik einer Untersuchung besteht die Gefahr, viele Fehler zu machen. Doch im Gegenzug dazu bietet eine Videoaufzeichnung vollständige und verlässliche Daten.

Neben diesen drei grundlegenden Methoden wird bei Untersuchungen in der Mensch-Computer



Interaktion nicht selten auch Software zur Erfassung des Bildschirminhaltes eingesetzt. Eine der bekannteren Anbieter für solche Software ist TechSmith, deren Produkt Morae® nicht nur die Bildschirminhalte erfassen kann, sondern diese zusätzlich mit den Daten einer Videokamera synchronisiert. Zusätzlich können mit Hilfe dieser

Software die Interaktionen des Probanden geloggt werden, sodass bei der Auswertung Mausclicks eindeutig identifiziert werden können. Die Verwendung solcher Software eignet sich nicht nur zur Evaluation eigener Prototypen, sondern kann auch im Rahmen des Anforderungsermittlungsprozesses eingesetzt werden, um eventuelle Bedienungsprobleme der aktuell benutzten Software zu finden und einen Einblick in die Bedienabläufe des Benutzers zu erhalten.

Dass diese Art der Datenerfassung in der automotiven Usability Forschung sehr gut eingesetzt werden kann, zeigt Kapitel 4 und auch einer der Befragten des Fragebogens gab an, dass diese Methode in der Automobilindustrie verwendet wird.

Etwas häufiger wurde die Verwendung von Logdaten genannt, die eine sehr genaue Methode der Datenerfassung darstellen. Wie Software zur Erfassung der Bildschirminhalte, kann auch diese Methode im Rahmen der Anforderungsanalyse dabei helfen, die aktuellen Bedienabläufe der Benutzer zu verstehen. Nicht nur Bedienzeiten können mit Hilfe von Logdaten genauestens erfasst werden, sondern auch die Fehler, die der Benutzer macht oder die Positionen seiner Klicks.

Wie in der traditionellen Mensch-Computer Interaktion werden auch in der Automobilindustrie Kombinationen mehrerer Datenerfassungstechniken verwendet, um Befragungen oder Beobachtungen aufzuzeichnen. Videoaufzeichnungen nutzen dabei acht der zehn Befragten als einziges Mittel oder in Kombination mit anderen Werkzeugen. Genauso viele schreiben in der Phase der Anforderungsermittlung handschriftliche Protokolle über ihre Gespräche mit potentiellen Benutzern. Zwei der Befragten gaben an, im Rahmen der Anforderungsanalyse Logdaten zu erstellen. Diese Art der Daten wird jedoch nur in Kombination mit handschriftlichen Protokollen und Videoaufzeichnungen erhoben [Fragebogen 09]. Die Kombination aus Videoaufzeichnungen und handschriftlichen Protokollen ist sehr beliebt, da durch die handschriftlichen Protokolle eine erste,

schnelle Auswertung der Daten erfolgen kann und auf diese Weise erste Ergebnisse schnell vorgezeigt werden können. Die zusätzlichen Videoaufzeichnungen ermöglichen ein späteres, detaillierteres Auswerten der Daten. Diese Kombination wird von sieben der zehn Befragten bevorzugt.

Die verwendeten Werkzeuge zur Datenerfassung sind in der Automobilindustrie jedoch immer vom zeitlichen und finanziellen Budget abhängig, das von Auftraggeber zur Verfügung gestellt wird.

Benutzerorientierte Methoden, die in der Anforderungsanalyse oder der Evaluation Anwendung finden, bedienen sich der hier genannten Datenerfassungsmethoden. Diese werden nachfolgend im Detail vorgestellt. Zunächst erfolgt jedoch eine Abgrenzung zu benutzungsorientierten Methoden der Mensch-Computer Interaktion.

3.2. Benutzungsorientierte Methoden

Die oben genannten Methoden der Datenerfassung beinhalten jeweils das Vorhandensein potentieller Benutzer zur Befragung beziehungsweise Beobachtung. Sind jedoch potentielle Benutzer nicht in ausreichendem Maße verfügbar oder ist man Verfechter des benutzungsorientierten Ansatzes, kann zur Erfassung der Daten der Fokus weg vom Benutzer und hin auf die Benutzung gelegt werden. Manche Anhänger der benutzungsorientierten Herangehensweise betrachten benutzerzentrierte Methoden zudem als schädlich für ein erfolgreiches Projekt [Norman 05]. Sie rücken Aktivitäten in den Vordergrund und argumentieren damit, dass viele Systeme, die nicht dem natürlichen Verhalten des Benutzers entsprechen, sehr gut und global funktionieren. Als Beispiele werden Musikinstrumente, Uhren oder das Schreiben genannt. Diese funktionieren deshalb so gut, weil sich der Benutzer des jeweiligen Systems der unterliegenden Aktivität bewusst ist. Wenn der Benutzer die Aktivität versteht und das System so gestaltet wurde, dass es die Aktivität unterstützt, kann der Benutzer auch das System selbst verstehen. Darüber hinaus führt zu großer Fokus auf eine bestimmte Benutzergruppe dazu, dass das entwickelte System für diese Gruppe zu diesem Zeitpunkt maßgeschneidert ist. Für Andere oder sobald sich die Individuen dieser Gruppe weiterentwickelt haben, ist es unangemessen. Weiter haben Anhänger des benutzungsorientierten Ansatzes festgestellt, dass allein entsprechend der benutzerzentrierten Methoden gestaltete Systeme mit jeder Überarbeitung komplexer und zugleich schlechter verständlich werden (vgl. Abbildung 1.01) [Norman 05]. Zusätzlich birgt das zu starke Fokussieren auf den Benutzer das Problem, dass das Entwickeln innovativer Systeme erschwert wird, da Benutzer oftmals Altbekanntes bevorzugen und vor Neuerungen zunächst zurückschrecken (vgl. [Söter 08] Seiten 16, 22/23). Tatsächliche Benutzer werden aus diesem Grund bei diesem Ansatz nur am Rande befragt.

Im Fokus der benutzungsorientierten Anforderungsanalyse, Entwicklung und Evaluation steht daher der Usability-Experte. Dieser bringt in jedem der genannten Phasen sein Experten-Wissen über Gebrauchstauglichkeit und Benutzerfreundlichkeit ein. In einem ersten Schritt muss er die jeweilige Zielgruppe befragen, oder idealerweise beobachten. Beobachtungen eignen sich laut Verfechter dieses

Ansatzes besser, da sie die Handlungsweisen und Charakteristiken ungefiltert wiedergeben, was bei einer Befragung nicht der Fall wäre. Da Personen zudem dazu neigen, Details auszulassen, weil sie ihnen als unwichtig erscheinen, sind Beobachtungen hier besonders hilfreich. Auf diese Weise lernt der Experte den Nutzungskontext und die Aktivitäten der späteren Benutzer in einem hohen Detailgrad kennen. Diese Informationen dienen als Basis für die Entwicklung von Personas und Szenarien, die die Ergebnisse der Beobachtungen und Befragungen bündeln und für die Anforderungsanalyse ein wichtiger Schritt sein können.

Personas sind Beschreibungen fiktiver Personen, die sich aus den Charakteristiken einer bestimmten Zielgruppe zusammensetzen. Um Anforderungen an ein System extrahieren zu können, werden für jede Persona auch ihr technisches Fachwissen, ihre Ziele und Motive sowie Aufgaben und Benutzungsmuster notiert. Die Notation kann in Freitextform (vgl. [Sharp et al. 07], S. 484) oder tabellarisch (vgl. [Söter 08], S. 53) erfolgen und beinhaltet stets einen fiktiven Namen sowie ein Bild.

Persona Peter Müller	
Persönliche Informationen.	<ul style="list-style-type: none"> - 31 Jahre alt. - Glücklicher verheirateter Familienvater. - 3 Kinder im Alter von 28, 22 und 12 Jahren. - Diplom Volkswirt. - Sehr ehrgeizig in allen Lebensbereichen. - Interessiert sich für das aktuelle Weltgeschehen. - Spielt mit seinem jüngsten Sohn am Wochenende auf der Wii und Playstation.
Arbeitspezifische Informationen.	<ul style="list-style-type: none"> - Von Beruf Key Account Manager, der oft im Auto zu seinen Kunden gefahren wird. - Darf auch während den langen Fahrten keine Arbeitszeit verlieren. - Muss sich vor jedem Termin genauestens über seinen jeweiligen Kunden informieren. - Hat eine 60-Stunden Woche. - Muss oft mit seinen Mitarbeitern und Kunden in Kontakt bleiben. - Benötigt seine Daten (Dokumente, Bookmarks, Adressbuch, ...) auch unterwegs. - Muss das Öfteren Tabellen und Texte einsehen, scannen, verwalten.
Technischer Kenntnisstand.	<ul style="list-style-type: none"> - Benutzt seit vielen Jahren mehrfach täglich Laptop und PC. - Benutzt bei seiner Arbeit sehr häufig Mobiltelefon, PDA und Co. - Surft beruflich und privat viel im Internet – insgesamt mehrere Stunden täglich. - Interessiert sich für neue Erfindungen im Bereich der Technologie. - Kennt sich dank seiner Konsolen-Erfahrung mit alternativen Eingabegeräten aus.
EDV Ausstattung	<ul style="list-style-type: none"> - PC daheim, Laptop bei der Arbeit, Blackberry, Mobiltelefon. - Nintendo Wii, Playstation 2.
Ziele, Motive, Bedenken.	<ul style="list-style-type: none"> - Die wichtigsten Aufgaben von Peter sind die telefonische und schriftliche Kommunikation mit seinen eigenen Mitarbeitern und seinen Kunden sowie das Einholen von Informationen über den Kunden und über das aktuelle Weltgeschehen – letzteres aus eigenem Interesse und um nicht-arbeitspezifische Gesprächsthemen für das Treffen mit dem Kunden zu haben. - Für Peter ist es essentiell seine Hauptaufgaben effizient erledigen zu können, da er in seinem Job sehr viel arbeiten muss und er danach zu seiner Familie gehen will, um auch mit ihr Zeit verbringen zu können. - Peter macht es nichts aus eine gewisse Zeit zu investieren, um eine Software zu erlernen, die er für sinnvoll hält und die ihn dabei unterstützt, seine Arbeit so schnell wie möglich zu erledigen. Zudem macht es ihm auch oft Spaß, neue Technologien zu ertratseln, da das Wissen darüber auch zu seinen Interessen gehört.

Abbildung 3.01: Beispielhafte, tabellarische Darstellung einer Persona (Quelle: [Söter 08]).

Durch ihre glaubwürdige und realistische Beschreibung können Personas als tatsächliche, mögliche Benutzer eines Systems angesehen werden, was das Gestalten für diese Benutzergruppe vereinfachen kann [Sharp et al. 07]. Auch beim benutzerzentrierten Ansatz kann es hilfreich sein, zunächst diesen

Weg einzuschlagen, da auf diese Weise noch vor der Beteiligung realer Nutzer ein Überblick über die Eigenschaften der Zielgruppe geschaffen werden kann. Nach Einbezug tatsächlicher Nutzer im weiteren Verlauf der Anforderungsanalyse kann die Gruppe der Personas sowie die Beschreibung einzelner Personas beliebig erweitert werden.

Da der Fokus benutzungsorientierter Methoden auf den Aufgaben und Zielen der Benutzer liegt, muss das zu entwickelnde System so konzipiert sein, dass es diese erfolgreich unterstützt. Hierfür können in einem weiteren Schritt die Aufgaben der Personas anhand von Essential Use Cases (vgl. [Constantine & Lockwood 99]) modelliert werden. Das Hauptaugenmerk bei dieser Technik ist die Interaktion des Benutzers mit einem System. Daher werden hier die Intention des Benutzers bei seiner Interaktion mit dem System, sowie die Antwort des Systems auf die Eingaben des Benutzers beschrieben. Klassische Use Cases beziehungsweise Essential Use Cases können in Use Case Diagrammen zusammengefasst werden [Constantine & Lockwood 99] [Sharp et al. 07]. Diese bieten einen Überblick über die Aufgaben, die das System unterstützen soll sowie über die Beziehungen zwischen den Aufgaben und können während der weiteren Entwicklung des Systems eine hilfreiche Gedankenstütze sein (vgl. [Söter 08], S. 8). Insbesondere bei größeren Projekten, wie sie in der Automobilindustrie zu finden sind, unterstützen Aufgabenmodelle den Designer dabei, elementare Aufgaben nicht aus den Augen zu verlieren. Auf diese Weise kann das System so gestaltet werden, dass es den Benutzer effektiv und effizient unterstützen kann.

Auch Szenarios sind eine bewährte und weniger formale Methode, Anforderungen an ein System zu ermitteln, ohne reale Benutzer zu stark in den Prozess zu integrieren. Sie können von Usability-Experten in Freitextform verfasst werden und beinhalten eine Beschreibung des Kontextes der Nutzung sowie die damit verbundenen Ziele und Anforderungen denen eine Person bei der Interaktion mit einem System begegnet. Auch sie stellen idealerweise eine gebündelte und strukturierte Zusammenfassung der Ergebnisse einer vorangehenden Beobachtung oder Befragung dar. Im Gegensatz zu Essential Use Cases liegt der Fokus hier nicht auf der Interaktion mit dem System, sondern auf den Eigenheiten der Aktionen, die der Benutzer vollführen will sowie auf dessen Zielen. Sie können beispielsweise auch den Pfad durch einen Use Case in einer detaillierten Art und Weise beschreiben. Szenarios helfen dabei, ein Verständnis für den Nutzungskontext zu generieren um hieraus Erfordernisse und Nutzungsanforderungen abzuleiten (vgl. [DATech 09] sowie [Söter 08], S. 61). Ein Nachteil, den die Erstellung von Szenarios mit sich bringt, ist der hohe Aufwand bei der Erstellung und Auswertung. Die ganze Bandbreite an Anforderungen innerhalb von Szenarios zu beschreiben würde sehr viel Zeit kosten. Deshalb wird empfohlen, sich hier auf die Wichtigsten zu beschränken. Szenarios erhöhen zwar das Verständnis des Designers für verschiedene Aspekte der Benutzung, doch sollten diese möglichst bald in eine formale Ordnung transformiert werden (vgl. [DATech 09]) um sie in den folgenden Phasen der Entwicklung einfacher interpretieren zu können.

Sowohl bei der Verwendung von Personas, als auch Szenarien empfiehlt es sich, die gefundenen Ergebnisse anhand von weiteren Befragungen oder Fokusgruppen zu evaluieren und sie gegebenenfalls zu ergänzen.

Die Benutzung von Szenarios ist auch in der Automobilindustrie denkbar, doch der damit verbundene hohe Aufwand verlangt ab einem gewissen Punkt das Ausweichen auf andere Methoden, wie beispielsweise Personas und Essential Use Cases.

Auch bei Anforderungsanalysen im automotiven Umfeld werden teilweise auf benutzungszentrierte Methoden zurückgegriffen. Als Grund hierfür wird genannt, dass Interviews, Fokusgruppen oder auch Fragebögen nur eine kleine Teilmenge der Zielgruppe erreichen können. Insbesondere wenn es um die Entwicklung von Ein- und Ausgabemethoden über verschiedene Märkte hinweg geht, haben sich benutzungsorientierte Methoden als erfolgreichere Alternative erwiesen. In diesen Fällen werden der Markt oder beliebige andere gut funktionierende und beliebte Produkte während der Anforderungsanalyse exploriert. Die Ergebnisse und daraus extrahierte Anforderungen werden anschließend im Team diskutiert. Eine Analyse durch Experten aus Vertrieb, Marketing oder Entwicklung entscheidet schließlich, welche Ideen weiterverfolgt werden [Fragebogen 09]. Diejenigen Personen, die während der Anforderungsanalyse diesen Ansatz wählen, geben jedoch an, während der anschließenden Produktentwicklung bereits in frühen Entwicklungsphasen Nutzerstudien durchzuführen.

Evaluationen können im benutzungsorientierten Ansatz ohne den Einbezug von Benutzern stattfinden und bilden so eine schnelle und kostengünstige Möglichkeit, die Gebrauchstauglichkeit eines Systems zu bestimmen. Da hier Usability-Experten oder Vorhersagemodelle die Rolle des Benutzers übernehmen, können diese analytischen Methoden nur Mutmaßungen über mögliche Bedienungsprobleme oder Bedienungsdauern hervorbringen. Der Vorteil von Inspektionen durch Experten ist, dass diese auch schon in frühen Stadien der Produktentwicklung angewandt werden können, noch bevor erste vollfunktionale Prototypen vorhanden sind.

Der Experte, der eine Evaluation eines Systems oder Prototypen durchführt, sollte ein gewisses Fachwissen in Usability Engineering, Design von Benutzerschnittstellen, Grafikdesign, Ergonomie, Mensch-Computer Interaktion, industrieller oder kognitiver Psychologie oder Softwareengineering und -entwicklung besitzen [Constantine & Lockwood 99]. Je erfahrener der Experte ist, desto besser sind die Ergebnisse seiner Inspektion. Die Herangehensweise beim Test durch einen Experten kann unterschiedlich sein, doch meist nimmt der Experte die Rolle des unerfahrenen Nutzers an und blendet dabei sein Wissen in Bezug auf Software, Computer und Benutzerschnittstellen aus [Constantine & Lockwood 99]. Hierbei nimmt er ihren ersten Eindruck über das System zur Kenntnis, bevor er bei der weiteren Inspektion ins Detail geht. Zudem versucht er, jeden Bereich des Systems zu besuchen und zu beurteilen. Stress-Testen ist ebenfalls eine häufig genutzte Methode, um das

Verhalten der Software bei Fehlern und Überforderung zu überprüfen [Constantine & Lockwood 99]. Das Ergebnis einer Evaluation durch einen Experten ist die Beurteilung des Systems durch diesen, welche häufig in Form einer Liste der Probleme sowie Vorschlägen zur Verbesserung erfolgt. Ist der Evaluierende erfahren, kann er mit dieser Methode 60 bis 70% der Bedienungsprobleme aufdecken [Constantine & Lockwood 99]. Als Nachteil einer solchen Inspektion muss der Fakt genannt werden, dass der Entwickler des Systems anschließend zwar weiß, was er falsch gemacht hat. Die Gründe hierfür oder wie er diesen Fehler in der Zukunft vermeiden kann, lernt er jedoch nicht.

Auch in der Automobilindustrie ist das Testen durch Experten eine geläufige Methode. Durch das direkte Testen der Prototypen in möglichst realen Situationen und unter Einbezug normaler Störungen werden Problemstellen schon im Vorfeld identifiziert. Oftmals werden hierbei vergleichende Untersuchungen durchgeführt. Dieser Vorgang umfasst in der Automobilindustrie etwa 20 Experten, jedoch werden auch hier in einem weiteren Schritt tatsächliche oder potentielle Nutzer in den Evaluationsprozess eingebunden, um weitere Informationen über die Güte des getesteten Systems zu erhalten [Fragebogen 09].

Eine besondere Form von Inspektionen ist die heuristische Evaluation. Im ersten Schritt wird hier das gesamte System durch einen Usability-Experten gesichtet, um seine Gesamtstruktur zu verstehen. Im zweiten Schritt untersucht der Experte anhand einer Liste von Heuristiken einzelne Elemente der Benutzeroberfläche und überprüft, ob diese den Heuristiken entsprechen [Constantine & Lockwood 99]. Nielsen schlägt für diese Art der Evaluation seine mehrfach überarbeitete Liste von zehn generellen Heuristiken vor [Nielsen 94], empfiehlt aber die zusätzliche Entwicklung eigener, individueller Heuristiken, um die Bedienungsprobleme in spezifischen Anwendungsbereichen aufzudecken [Sharp et al. 07]. Um den Großteil der Bedienungsprobleme aufzudecken, empfiehlt Nielsen die Beteiligung von drei bis fünf Evaluierenden [Nielsen 90]. Grundsätzlich gilt hier jedoch die Regel „mehr ist mehr“, da die Beteiligung von mehr als fünf Usability-Experten dazu führt, dass die Anzahl gefundener Usability-Probleme steigt. Der Nachteil dieser Methode ist, dass viele Probleme ohne Vorschläge zur Verbesserung aufgedeckt werden und sie gewöhnlich nicht zu bahnbrechenden Neuerungen beim evaluierten Design führt. Daher empfiehlt Nielsen zusätzliche Ressourcen in alternative Evaluationsmethoden zu investieren [Nielsen 90]. Große Vorteile bringt die Methode heuristischer Evaluation insbesondere bei mobilen Systemen mit sich, da sie in diesem Anwendungsgebiet einfacher durchzuführen ist als Beobachtungen und Feldstudien [Sharp et al. 07].

Ein alternativer Ansatz um Bedienungsprobleme ohne Benutzertests zu bestimmen, sind Walkthroughs. Hierbei bearbeitet der Usability-Experte oder auch ein Team aus selbigen, Entwicklern sowie Benutzern Aufgaben mit einem System und notiert währenddessen problematische Merkmale des Systems. Der Fokus bei kognitiven Walkthroughs liegt, im Gegensatz zu heuristischer Evaluation, nicht auf dem gesamten System oder auf größeren Teilen davon, sondern auf der detaillierten Untersuchung kleinerer Teile des Systems. Der Grund liegt unter Anderem darin, dass es für den

untersuchenden Designer oder Usability-Experten viel zu aufwändig wäre, jede mögliche Aktion in einem System daraufhin zu untersuchen

- ob für den Benutzer verständlich ist, was er beim aktuellen Schritt tun muss,
- wie er dies tun muss und
- ob ihm das Feedback, welches er erhält, ausreicht um zu wissen, ob seine Aktion korrekt war oder nicht.

Diese zu betrachtenden Aspekte betreffen jeweils kognitive Vorgänge bei der Interaktion, weshalb diese Art von Walkthrough auch den Namen „kognitiver Walkthrough“ erhalten hat [Sharp et al. 07]. Diese Methode kann bei der benutzerzentrierten Vorgehensweise auch sehr gut mit zusätzlichen Benutzertests kombiniert werden, was dazu führt, dass eine Vielzahl von Bedienungsproblemen aufgedeckt werden kann.

Eine andere Art von Walkthroughs sind pluralistische Walkthroughs, an denen ebenfalls neben Usability-Experten und den Entwicklern selbst auch Benutzer teilnehmen. Hierbei werden Szenarien in Form von Bildern einer Benutzerschnittstelle präsentiert und die Teilnehmer schreiben zunächst einzeln auf, wie sie sich von einem Bildschirm zum nächsten bewegen würden. Entwickler und Experten sollen dabei versuchen, sich in die Rolle typischer Benutzer hineinzusetzen. Anschließend werden die Ergebnisse präsentiert, wobei typischerweise die tatsächlichen Benutzer beginnen, damit sie von den Aussagen der Experten nicht beeinflusst werden. Die jeweils vorgeschlagenen Aktionen werden in der Gruppe diskutiert, bevor zum nächsten Szenario gewechselt wird. Da bei dieser Art der Evaluation ein starker Fokus auf den Aufgaben des Benutzers liegt und diese sehr detailliert betrachtet werden, eignet sich diese Art der Evaluation insbesondere bei sicherheitskritischen Anwendungen [Sharp et al. 07]. Durch das multidisziplinäre Evaluationsteam kommen verschiedene Meinungen zum Verlauf der Interaktion zur Sprache, was für die Weiterentwicklung hilfreich sein kann. Wie kognitive Walkthroughs, sind jedoch auch pluralistische Walkthroughs sehr zeitaufwändig, sodass auch hier nur eine beschränkte Anzahl von Szenarien besprochen werden kann [Sharp et al. 07].

Im automotiven Kontext können Walkthroughs zwar ebenfalls zur Evaluation eingesetzt werden, doch aufgrund des hohen Zeitaufwandes und des hohen Detailgrades in einzelnen kleineren Bereichen ist diese Methode wenig praktikabel. Im Gegensatz dazu können heuristische Evaluationen aufgrund ihrer flexiblen Einsetzbarkeit sehr gut eingesetzt werden, da sie den Anforderungen der mobilen Automobilindustrie genügen. Durch die finanzielle Flexibilität, wie sie im hier im Gegensatz zur universitären Forschung gegeben ist, ist die Beteiligung mehrerer Usability-Experten einfacher zu bewerkstelligen, was zusätzlich dazu führt, dass die Mehrzahl der Bedienungsprobleme aufgedeckt werden kann.

Im Allgemeinen haben benutzungsorientierte Methoden in der Anforderungsanalyse oder bei der Evaluation den Vorteil, dass sie mit wenig Aufwand durchgeführt werden müssen. Da nicht einzelne potentielle Nutzer zu Befragungen eingeladen oder in Fokusgruppen zusammengebracht werden

müssen, ist man mit diesem Ansatz zeitlich und auch örtlich flexibel. Zudem erhält man mit diesem Ansatz in jeder Entwicklungsphase schnell diejenigen Informationen, die für die weitere Entwicklung nötig sind. Je nach Qualifikation des befragten Usability-Experten können diese Informationen von sehr hoher Qualität sein und daher auch sehr hilfreich für den weiteren Prozess. Des Weiteren sind Evaluationen mit diesem Ansatz bereits in frühen Stadien der Produktentwicklung möglich, sodass bereits hier eine Vielzahl an möglichen Bedienungsproblemen aufgedeckt und bereinigt werden kann.

3.3. Benutzerorientierte Methoden der Anforderungsanalyse

Um Systeme zu schaffen, die den Benutzer bei seiner Tätigkeit unterstützen, gilt es möglichst viele Informationen über seine Charakteristiken, Tätigkeiten und den Kontext dieser Tätigkeiten zu erfahren. Diese Phase im Prozess der Produktentwicklung nennt sich Anforderungsanalyse. Sie ist ein iterativer Prozess, da durch das Ermitteln einiger Anforderungen deutlich werden kann, welche anderen Daten noch erhoben werden müssen, um die bisherigen Anforderungen zu bestätigen oder zu erweitern. Auch nach einer ersten Evaluation des Systems durch potentielle Benutzer (vgl. Kapitel 3.4. bzw. Kapitel 4.1.) oder Usability-Experten (vgl. Kapitel 3.2.) muss gegebenenfalls zur Anforderungsanalyse zurückgekehrt werden, um die neuen Erkenntnisse zu bündeln und um sie mit den bestehenden Anforderungen zu vereinen.

Anforderungen lassen sich in funktionale und nicht-funktionale Anforderungen unterteilen. Während funktionale Anforderungen bestimmen, welche Funktionen das System erfüllen soll, legen nicht-funktionale Anforderungen die Eigenschaften des Systems fest. Funktionale Anforderungen können von System zu System verschieden sein. Für nicht-funktionale Anforderungen gibt es verschiedene Listen von Qualitätsmerkmalen, wie beispielsweise nach ISO/IEC 9126 (vgl. Tabelle 3.01) oder Volere.

Hauptmerkmal	Kurzbeschreibung
Funktionalität	Vorhandensein eines Satzes von Funktionen mit spezifizierten Eigenschaften.
Zuverlässigkeit	Fähigkeit der Software, ihr Leistungsniveau unter festgelegten Bedingungen über einen festgelegten Zeitraum aufrechtzuerhalten.
Benutzbarkeit	Aufwand, der zur Benutzung erforderlich ist, und individuelle Beurteilung einer Benutzung durch die vorausgesetzte Gruppe von Benutzern.
Effizienz	Verhältnis zwischen Leistungsniveau der Software und dem Umfang der eingesetzten Betriebsmittel unter festgelegten Bedingungen.
Änderbarkeit	Aufwand, der zur Durchführung vorgegebener Änderungen notwendig ist.
Übertragbarkeit	Eignung einer Software, von einer Umgebung in eine andere übertragen zu werden.

Tabelle 3.01: Qualitätsmerkmale für Software nach ISO/IEC 9126-1 (Quelle: [Hohler 07]).

Zu den nicht-funktionalen Anforderungen gehören auch Faktoren, wie das Aussehen und die Handhabung des Systems („Look and Feel“) oder seine Benutzbarkeit, die sich unter anderem aus seiner Verständlichkeit und Erlernbarkeit zusammensetzt. Auch Faktoren wie Kosten oder Termin der Fertigstellung gehören zur Gruppe der nicht-funktionalen Anforderungen. Während diese beiden Faktoren häufig vom Auftraggeber festgelegte Einschränkungen sind, betreffen einige der anderen nicht-funktionalen und funktionalen Anforderungen den Benutzer und seine Interaktion mit dem System. Daher ist es bei der benutzerorientierten Systementwicklung unerlässlich, den Benutzer in den Prozess der Anforderungsanalyse zu integrieren. Hierbei werden häufig Methoden angewandt, die in den nachfolgenden Kapiteln erläutert werden. Für den Anforderungsermittlungsprozess ist es wichtig, sämtliche Zielgruppen in mit einzubeziehen, um ein umfassendes Bild ihrer Aufgaben, Ziele und Bedürfnisse zu erhalten [Sharp et al. 07]. Da jeder Benutzer sehr individuelle Ansichten zu seiner Aufgabe, seinem Umfeld und zu seiner präferierten Art der Interaktion hat, empfiehlt es sich zudem, mehr als nur eine Person pro Zielgruppe zu involvieren.

Soll in der Automobilindustrie ein neues Eingabegerät, Fahrzeugdesign oder System entwickelt werden, so entstammen 46,21%* der Ideen hierfür dem Brainstorming und der Diskussion im Entwicklerteam. Die Betrachtung von Konkurrenzprodukten spielt mit 23,91% ebenfalls eine gewichtige Rolle. Zu 13,69% werden jedoch auch potentielle Nutzer nach Ihren Vorstellungen und Ideen befragt, wenn ein neues System entwickelt werden soll. Aus dieser Quelle stammen immerhin mehr Ideen als vom Vorgesetzten (10,74%) oder von Marktforschungsinstituten (5,46%). Neben diesen Quellen werden auch Produkte aus dem nicht-automotiven Bereich betrachtet, um Ideen für ein neues Konzept zu generieren oder es werden Kreativmethoden angewandt. Doch auch Messebesuche, Fachkonferenzen und Kongresse können Quelle neuer Ideen sein [Fragebogen 09]. In Nachfolgenden werden jedoch nur diejenigen Methoden näher betrachtet, die den Benutzer mit einbeziehen, da diese den Fokus dieser Arbeit darstellen.

Zum Teil finden sich im Rahmen von Anforderungsanalysen in der Automobilindustrie Methoden der klassischen Mensch-Computer Interaktion wieder. Um schon in diesem Stadium der Produktentwicklung möglichst verlässliche und statistisch signifikante Resultate zu erzielen, wählen viele Untersuchungsleiter aus der Automobilindustrie ihre Methoden entsprechend der einschlägigen Literatur. Die Anzahl der Befragten dieser Phase setzt sich in solchen Fällen aus Literaturvorgaben und Erfahrungswerten zusammen, wobei organisatorische Grenzen ebenfalls eine gewichtige Rolle spielen [Fragebogen 09]. In manchen Fällen ist es jedoch aufgrund der in Kapitel 2 genannten besonderen Rahmenbedingungen und Forschungsfragen nötig, eine alternative Herangehensweise zu finden.

* Die Zahlen sind Mittelwerte der von den Befragten aus der Automobilindustrie genannten prozentualen Anteile.

Ziel der nachfolgenden Kapitel ist es, verschiedene Methoden der klassischen Mensch-Computer Interaktion vorzustellen und anschließend zu erläutern, ob und wie diese in der Automobilindustrie verwendet werden. Im Zentrum dieses Kapitels steht die Anwendung im Rahmen einer Anforderungsanalyse. Inwiefern diese Methoden im Rahmen einer Evaluation Verwendung finden können und welche besonderen Methoden dort nötig sind, wird in Kapitel 3.4. detailliert erläutert.

3.3.1. Die Befragten

Die Akquise von Befragten für Anforderungsanalysen in der traditionellen Mensch-Computer Interaktion hängt stark vom zu entwickelnden System und vom Anwendungskontext ab. Die Möglichkeiten sind hier nahezu unbegrenzt und oftmals kommt eine breite Zielgruppe in Betracht. Je nach zu bearbeitenden Forschungsfragen können die gewünschten Personen über E-Mail-Listen, Internetforen, Aushänge an Arbeits- oder Ausbildungseinrichtungen oder auch auf der Straße akquiriert werden.

Die Zielgruppe, die im Rahmen einer Anforderungsanalyse für Systeme im automotiven Bereich befragt werden kann, ist ebenfalls breit gefächert, doch gilt es zumeist in erster Linie Autofahrer zu befragen. Hier kann man weiter zwischen bestehenden und potentiellen neuen Kunden unterscheiden und in manchen Fällen kann es auch wichtig sein, nur Personen zu befragen, die ein Fahrzeug einer bestimmten Marke oder einer bestimmten Klasse fahren.

Aus organisatorischen Gründen werden in der Automobilindustrie diese Personen zumeist aus dem unternehmensinternen Probandenpool oder aus externen Partnerpools akquiriert. Der Probandenpool der Volkswagen AG beispielsweise ist eine seit 2004 existierende Einrichtung der Volkswagen Konzernforschung, dessen Auftraggeber aus verschiedenen Konzernbereichen stammen. Nahezu wöchentlich werden mit Hilfe seiner Mitglieder Projekte durchgeführt [VW 09].

Seit seiner Entstehung ist der Pool auf aktuell über 2000 Mitglieder gewachsen. Diese sind ausschließlich Mitarbeiter des Unternehmens, die sich selbstständig und freiwillig anmelden können [VW 09]. Die Altersspanne der Mitglieder beginnt bei etwa 20 Jahren und endet mit dem Vorruhestand. Ist es für die Produktentwicklung notwendig, auch ältere Personen zu befragen, so werden diese in der Automobilindustrie aus externen Partnerpools akquiriert.

Neben der Mitarbeit im Unternehmen ist der Besitz eines Führerscheins ein Aufnahmekriterium. Doch auch eine Verträglichkeit des Fahrsimulators ist insbesondere für den Einsatz der Personen im Rahmen einer Evaluation vorteilhaft [Fragebogen 09] [Gespräche]. Obwohl die Mitglieder des Pools im eigenen Unternehmen arbeiten, kennen Sie nicht alle Konzepte und Systeme, die in Fahrzeugen der Marke VW Anwendung finden. Der Anteil der Personen, die keine Erfahrungen mit Fahrerassistenzsystemen, wie beispielsweise einem Tempomat haben, ist hoch [Gespräche].

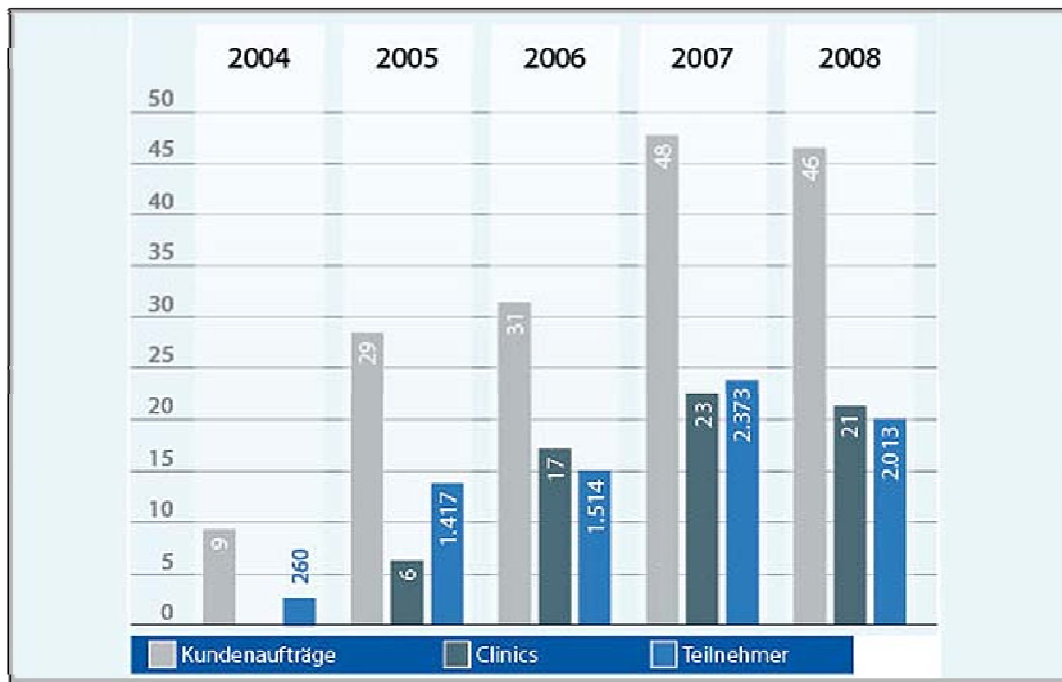


Abbildung 3.02: Entwicklung des Probandenpools seit seiner Gründung 2004 (Quelle: [VW 09]).

Der Vorteil von unternehmenseigenen Probandenpools ist unter anderem die Schnelligkeit, Kosteneffizienz oder die Kundennähe, die dadurch gewährleistet wird, dass viele Mitarbeiter des Unternehmens zumeist auch selbst Kunden sind. Ein ebenfalls wichtiger Aspekt ist das hohe Maß an Geheimhaltung, das mit Hilfe eines Probandenpools gewährleistet werden kann [VW 09]. Dieser Faktor ist für die Automobilbranche ein sehr wichtiger Aspekt, da die Konzepte und Systeme, die entwickelt oder getestet werden zumeist erst viele Jahre später auf den Markt kommen und die Hersteller verhindern müssen, dass die Konkurrenz von dem Vorhaben erfährt.

All diese Vorteile haben jedoch auch eine Kehrseite [Dumas & Redish 99]. Der Aspekt der Geheimhaltung kann beispielsweise auch durch die sorgfältige Auswahl der Probanden sowie durch das Unterschreiben einer Geheimhaltungsverpflichtung gewährleistet werden. Der Faktor Kosteneffizienz relativiert sich gegebenenfalls durch die Tatsache, dass die Mitglieder des Probandenpools während der Befragung oder Studie nicht ihren regulären Arbeiten nachkommen können. Zudem sollten sie in gewissem Maße für ihre Mitwirkung im Projekt entlohnt werden, was weitere Kosten nach sich zieht [Dumas & Redish 99].

Ein wichtiger Aspekt, der gegen die Einbindung eigener Mitarbeiter spricht, ist jedoch die Tatsache, dass sie nicht den tatsächlichen Benutzern entsprechen. So haben sie bestimmte Vorstellungen davon, wie ein System aussehen soll. Dieses entspricht jedoch nicht notwendigerweise der gebrauchstauglichsten Variante. Auch während der Evaluation können durch den fehlenden Einbezug externer Probanden schwerwiegende Probleme auftreten. Diese werden in Kapitel 3.4.3. erläutert.

Abgesehen vom Probandenpool kann auch über Fahrzeughändler oder Marktforschungsinstitute der Zugang zur gewünschten Zielgruppe, insbesondere zur Gruppen der Kunden, hergestellt werden.

Tankstellen können auch ein geeigneter Ort sein, um verschiedenste Zielgruppen für automotiv Anforderungsanalysen zu akquirieren. Die meisten Personen, die sich an diesem Ort aufhalten sind Autofahrer oder deren Mitfahrer. Je nach Fragestellung können hier auch sehr spezielle Zielgruppen identifiziert und um eine Teilnahme beim Anforderungsermittlungsprozess gebeten werden.

3.3.2. Interviews

Ein Interview ist in der klassischen Mensch-Computer Interaktion eine sehr gut geeignete Methode, um die Meinung eines potentiellen Benutzers bezüglich eines Konzepts oder Systems zu erhalten. Hierbei gibt es verschiedene Ansätze: Je nach Kontrolle durch den Untersuchungsleiter und nach Stadium im Entwicklungszyklus können Interviews mehr oder weniger strukturiert sein.

Bei unstrukturierten Interviews werden offene Fragen gestellt und der Befragte kann hierzu so ausführlich antworten, wie er möchte. Ein Beispiel für eine Frage in einem unstrukturierten Interview könnte sein: „Welche Aufgaben würden sie gerne auch während der Fahrt erledigen können“. Diese Art der Interviews erlaubt mehr Flexibilität als die strukturierte Form, da der Interviewer bei Bedarf nachhaken kann und auf diese Weise mehr in die Tiefe gegangen werden kann. Trotz dieser Flexibilität sollte jedoch immer das Ziel der Befragung im Auge behalten werden und die hierfür wichtigen Themen sollten dennoch abgearbeitet werden. Vorteil von unstrukturierten Interviews ist, dass verschiedene Sachverhalte zur Sprache kommen können, die der Interviewer zuvor möglicherweise nicht bedacht hat. Nachteilig ist der hohe Zeitaufwand bei der Analyse der Aufzeichnungsdaten. Zudem ist durch die hohe Flexibilität jedes Interview individuell und kann daher nicht auf die selbe Weise wiederholt werden [Sharp et al. 07].

Im Gegensatz zu einem unstrukturierten Interview geht der Interviewer in einem strukturierten Interview anhand eines zuvor festgelegten Fragenkatalogs vor. Diese Fragen sind zumeist geschlossen und erfordern eine Antwort innerhalb eines festgelegten Bereichs [Sharp et al. 07]. Beispielsweise könnte hier gefragt werden, ob eine bestimmte Webseite bekannt ist oder wie häufig diese auf einer vorgegebenen Skala besucht wird.

Semi-strukturierte Interviews sind ein Hybrid aus strukturierten und unstrukturierten Interviews. Hierbei werden ebenfalls bestimmte Fragestellungen behandelt, doch der Interviewer geht dabei so lange in die Tiefe, bis keine weiteren Informationen vom Befragten mehr zu erwarten sind [Sharp et al. 07]. Beispielsweise könnte hierbei nach dem Favoriten zwischen mehreren Varianten gefragt werden. Nach der Antwort des Befragten, könnte der Interviewer nach den Beweggründen für diese Entscheidung fragen. Wenn der Befragte nichts mehr äußert, könnte der Interviewer sicherheitshalber nochmals nachfragen, ob es noch weitere Gründe gibt, die bisher noch nicht genannt wurden.

Bei allen drei Arten von Interviews ist es wichtig, dass der Interviewer Suggestivfragen vermeidet und dem Befragten ausreichend Zeit zum Überlegen und zum Antworten gibt. Zur Protokollierung von Interviews eignet sich die Aufzeichnung mittels Diktiergerät oder Videokamera sehr gut, da auf diese Weise keine wichtigen Aussagen verloren gehen. Außerdem kann der Interviewer sich so voll auf seine Fragen und die Antworten des Befragten konzentrieren und gegebenenfalls nach weiteren Details zu den Antworten fragen. Gibt es mehr als einen Untersuchungsleiter, könnte eine zweite Person zusätzlich mittels Papier und Bleistift oder am PC ein Protokoll erstellen, welches für eine erste, schnelle Auswertung zu Rate gezogen werden kann.

Das Interview ist eine Methode, die im Rahmen von Anforderungsanalysen eher wenig Anklang in der Automobilindustrie findet. Für nur eine der zehn Befragten ist diese Methode die wichtigste zur Erhebung von Anforderungen [Fragebogen 09]. Die Begründung hierfür liegt darin, dass Anforderungsprofile sehr interindividuell sind. Mit weniger strukturierten Interviews kann der Untersuchungsleiter auf diese Profile und auf die Befragten selbst sehr individuell eingehen und erhält so sehr vielfältige und detaillierte Informationen. Interviews ermöglichen im automotiven Kontext das Erhalten von kreativem Output der Befragten. Vorgefertigte Frage-Antwort-Formate hingegen bieten diese Möglichkeit nicht.

Nichtsdestotrotz sind Interviews in ihrer Durchführung sehr zeit- und damit auch kostenintensiv: während ihrer gesamten Dauer können nur die Informationen von einer einzelnen Person abgerufen werden. Aufgrund der individuellen Art der Befragung fehlen hier zudem normierte Auswertungsmethoden, wie sie bei Fragebögen beispielsweise vorhanden sind. Der hohe Individualitätsfaktor von Interviews spiegelt sich daher negativ in der Auswertungszeit wieder, sodass die anderen Befragten alternative Methoden bei der Anforderungsanalyse präferieren [Fragebogen 09].

3.3.3. Fokusgruppen

Eine besondere Art von Interviews sind Fokusgruppen. Diese stammen aus der Marktforschung der 20er Jahre [Kitzinger 94] und unterscheiden sich von den oben beschriebenen Arten von Interviews insofern, dass mehrere Teilnehmer gleichzeitig befragt werden. Ziel einer Fokusgruppe ist es, innerhalb einer Gruppendiskussion die Ansichten, Bedürfnisse und Erfahrungen der Teilnehmer zu einem spezifischen Thema zu erkunden. Der Fokus wird durch eine kollektive Aktivität gegeben, wie das gemeinsame Betrachten eines Filmes, Vortrages, Konzeptes oder Systems oder aber durch das gemeinsame Erarbeiten einer Reihe von Fragen [Kitzinger 94].

Ein möglichst erfahrener Moderator leitet hierbei die Diskussion, ohne dabei den Fluss an Ideen und Meinungen zu behindern. Er treibt die Diskussion voran, indem er jeden Teilnehmer ermuntert, seine Meinung mitzuteilen. Während die Fokusgruppe für den Teilnehmer als freie und flexible Aktivität erscheint, ist sie für den Moderator fest vorgegeben, da er die für die Untersuchung benötigten Themen anhand eines Leitfadens abarbeitet [Morgan 97] [Nielsen 97].

Was die ideale Größe einer Fokusgruppe angeht, gibt es verschiedene Meinungen. Während Sharp und ihre Kolleginnen sich mit drei bis zehn Teilnehmern zufrieden geben [Sharp et al.07], erwartet Nielsen mindestens sechs bis neun Teilnehmer [Nielsen 97]. Morgan schlägt eine Teilnehmerzahl von sechs bis zehn vor [Morgan 97]. Bei weniger als sechs Teilnehmern sei es möglicherweise schwierig, die Diskussion aufrecht zu erhalten, wohingegen es bei mehr als zehn schwierig werden könne, diese zu kontrollieren [Morgan 97]. Problematisch bei einer größeren Gruppe könnte zudem sein, dass mehrere Leute gleichzeitig anfangen zu reden. Da es schwierig ist, solche Situationen auf einer Aufzeichnung auszuwerten, könnten so wertvolle Informationen verloren gehen. Darüber hinaus könnte es bei einer großen Gruppe passieren, dass sich Untergruppen bilden, die parallel miteinander diskutieren, was ebenfalls für den Auswertungsprozess abträglich ist [Morgan 97]. Fokusgruppen mit einer geringeren Teilnehmerzahl als sechs können jedoch gut funktionieren, wenn alle ein gewisses Interesse für das behandelte Thema mitbringen. Ebenso wichtig, wie eine geeignete Gruppengröße zu wählen ist es jedoch auch, die Teilnehmer so auszusuchen, dass sie die benötigte Zielgruppe repräsentieren.



Abbildung 3.03: Teilnehmer einer Fokusgruppe bei der Diskussion über automotiv Interaktionskonzepte.

Während die Meinungen bezüglich der Gruppengröße leicht differieren, ist man sich bezüglich der Anzahl der Gruppen einig: es sollte mehr als eine Fokusgruppe durchgeführt werden, um sicherzustellen, dass die Ergebnisse tatsächlich der Wahrheit entsprechen und nicht durch Gruppendynamik beeinflusst wurden [Morgan 97] [Nielsen 97]. Morgan schlägt das Durchführen von drei bis fünf Fokusgruppen vor, wobei die Anzahl sowohl von der Variabilität unter den Teilnehmern, als auch von der Strukturiertheit der Interviewfragen abhängt: je strukturierter die Fokusgruppe durchgeführt wird, umso weniger Gruppen bedarf es um die benötigten Informationen zu erhalten [Morgan 97].

Nachteilig bei einer Fokusgruppe ist, dass es je nach Charakteristiken der einzelnen Teilnehmer der Gruppe eine Tendenz in Richtung Konformität geben kann, wenn beispielsweise Teilnehmer ihre Meinungen vorenthalten, die sie in einem Einzelinterview geäußert hätten. Das andere Extrem ist ebenfalls möglich: eine Tendenz in Richtung Polarisierung kann erfolgen, wenn Teilnehmer ihre

Meinungen extremer darstellen, als sie es in klassischen Interviews getan hätten [Morgan 97]. Zudem erfährt man mit dieser Methode nur, was die Teilnehmer sagen und nicht das, was sie in Wirklichkeit tun [Nielsen 97]. Um das tatsächliche Tun eines Teilnehmers zu erforschen, wären Beobachtungen nötig. Im Gegensatz zu diesen, ist die Fokusgruppe durch das Vorgeben eines Diskussionsthemas durch den Moderator kontrollierter [Morgan 97]. Dadurch ist es für den Moderator jedoch auch möglich, die Informationen durch gemeinsame Meinungsfindung in der Gruppe zu erhalten, die er benötigt [Sharp et al. 07] [Morgan 97]. Um diesen Prozess zu fördern, ist es auch möglich in einer Fokusgruppe Prototypen zum vorgegebenen Thema zu zeigen. Das kann die Kreativität der Teilnehmer fördern und deren Meinungen zu spezifischen Eigenschaften der jeweiligen Prototypen ans Licht bringen, was für eine weitere Entwicklung hilfreich sein kann. In Fokusgruppen können zudem durch die Diskussion in der Gruppe verschiedene Sachverhalte und Probleme zur Sprache gebracht werden, die bei Einzelinterviews womöglich nicht bedacht worden wären [Sharp et al. 07].

Durch die Interaktion der einzelnen Teilnehmer ist eine Fokusgruppe weniger kontrolliert als ein Einzelinterview. Die Fokusgruppe bildet also einen Mittelweg zwischen Beobachtungen und klassischen Interviews, der die Stärken beider Methoden kombiniert. Allerdings erfordert die Auswertung von Fokusgruppen viel Zeit, insbesondere wenn für die bessere Auswertung eine Aufzeichnung mittels Videokamera erfolgte. Diese ist einer Aufzeichnung mittels Diktiergerät vorzuziehen, da es bei einer Audioaufzeichnung schwer fallen würde, die Aussagen den jeweiligen Teilnehmern korrekt zuzuordnen. Zudem kann der Moderator auf einem Video erkennen, ob jeder die Chance hatte, seine Meinung zu äußern.

Zusätzlich kann durch eine weitere Person, die nicht der Moderator ist, ein hand- oder maschinenschriftliches Protokoll erstellt werden, welches anschließend für die Zusammenfassung erster Ergebnisse hilfreich sein kann. Die Aufmerksamkeit des Moderators sollte nämlich auf seinen Leitfaden und vor Allem auf die Teilnehmer gerichtet sein.

Auch bei Untersuchungen im automotiven Bereich können Fokusgruppen problemlos durchgeführt werden. Aus diesem Grunde nennen drei der zehn Befragten des Fragebogens Fokusgruppen als wichtigste Methode, um potentielle Benutzer in den Anforderungsermittlungsprozess einzubinden. Die anderen Methoden wurden hingegen nur von einer oder maximal zwei Personen als wichtigste Methode der Anforderungsanalyse betrachtet. Der Vorteil, der in Fokusgruppen gesehen wird ist, dass ein lockerer Dialog mit den potentiellen Benutzern zustande kommt und dass durch deren Diskussion in der Gruppe die Stärken und Schwächen bestehender Systeme besser identifiziert werden können. Zudem wird diese Methode als wenig hemmend angesehen und ermöglicht die gegenseitige Stimulation zu neuen Denkansätzen [Fragebogen 09]. Die Vorteile dieser Methode für die traditionelle Mensch-Computer Interaktion werden also auch in der Automobilbranche als solche angesehen. Im Gegensatz dazu müssen jedoch in der Automobilindustrie gewissen Kriterien der Geheimhaltung berücksichtigt werden. Aus diesem Grunde können im Rahmen von Fokusgruppen je nach Hintergrund der Teilnehmer nur bedingt Prototypen oder erste Ideen präsentiert werden. Entstammen

die Teilnehmer dem unternehmensinternen Probandenpool vermindert sich die Gefahr der unbefugten Informationsweitergabe.

Wie in der Literatur empfohlen, wird auch in der Automobilindustrie im Rahmen einer Anforderungsanalyse mehr als eine Fokusgruppe durchgeführt. Die Teilnehmerzahl wird laut Antworten des Fragebogens anhand von Erfahrungswerten festgelegt, und entspricht mit etwa 8 Teilnehmern ebenfalls den Vorgaben aus der Literatur [Fragebogen 09].

3.3.4. Fragebögen

Fragebögen sind ein Werkzeug der Datenerfassung, welche sich von Interviews und Fokusgruppen insofern unterscheidet, dass die Frage nicht von einem Interviewer gestellt wird, sondern auf einem Blatt Papier, einer Webseite oder Ähnlichem formuliert ist.

Fragebögen können die demografischen Daten der Befragten betreffen oder deren Meinung. Insbesondere erstere sollten bei allen Untersuchungen Anwendung finden, da bei einer späteren Auswertung beispielsweise verglichen werden kann, ob Personen mit ähnlichen Antworten möglicherweise ähnliche Charakteristiken aufweisen oder nicht.

Fragebögen können sich aus offenen, geschlossenen oder halboffenen Fragen zusammensetzen, wobei es bei offenen Fragen keine Antwortvorgaben gibt. Beispielsweise kann hier nach der Art der Ausbildung einer Person gefragt werden, die diese in einem hierfür vorgesehenen Feld notieren kann. Bei geschlossenen Fragen sind die möglichen Antwortalternativen bereits vorgegeben und der Befragte muss zwischen diesen auswählen. Die Frage nach Geschlecht des Befragten oder nach der favorisierten Variante zwischen mehreren Designalternativen könnte dieser Kategorie angehören. Halboffene Fragen werden ebenfalls von einer vorgegebenen Anzahl an Fragen gefolgt, ermöglichen jedoch auch weitere Alternativen durch die zusätzliche Antwortoption „Sonstige“.

Alle drei Fragetypen haben jedoch gemein, dass sie klar und verständlich formuliert sein sollten und nur eine Aussage pro Frage überprüfen sollten. Suggestivfragen sind hier – ebenso wie bei Interviews – zu vermeiden. Ein weiterer Aspekt von Fragebögen, der strukturierten Interviews ähnelt, ist ihr Aufbau. Allerdings muss die Motivation, alle Fragen zu bearbeiten, hier höher sein, da kein Interviewer vorhanden ist [Sharp et al. 07].

Bei der Erstellung von Fragebögen kann zwischen verschiedenen Frage- und Antwortformaten gewählt werden. Fragen können die Auswahl zwischen mehreren Alternativen, wie beispielsweise Geschlecht oder Händigkeit betreffen oder aber die Auswahl zwischen mehreren Bereichen, wie das bei der Altersspanne möglich ist.

Fragebögen können aus Beurteilungsskalen wie der Likert-Skala oder dem semantischen Differential bestehen, mittels derer es möglich ist, die Beurteilung von Sachen oder Systemen zu erfahren. Aus diesem Grund wird dieses Format häufig für Evaluationsstudien verwendet [Sharp et al. 07]. Bei einer Likert-Skala wird eine Eigenschaft präsentiert und der Befragte entscheidet sich, wie sehr er dieser Eigenschaft auf einer Skala von beispielsweise eins bis fünf zustimmt. Eigenschaften können

hier die Einfachheit, die Benutzbarkeit, die Attraktivität von Systemen betreffen. Im Gegensatz hierzu werden dem Befragten beim semantischen Differential gegensätzliche Wortpaare vorgelegt und er muss seine Meinung durch ein Kreuz auf der Skala zwischen den beiden Begriffen äußern [Sharp et al.07a]. Die größte Herausforderung beim semantischen Differential liegt darin, Wortpaare zu finden, die einerseits zur Fragestellung der Untersuchung passen und andererseits unmissverständlich sind.

Der wohl größte Vorteil von Fragebögen ist, dass sie preisgünstig sind und viele Personen gleichzeitig erreichen können – unabhängig von deren Position auf dem Globus [Kirakowski 00] [Sharp et al. 07]. Zudem haben die Befragten eines Fragebogens während der Bearbeitung ausreichend Zeit, ihre Antwort zu überlegen, wohingegen sie beim Interview oftmals recht spontan antworten müssen.

Nachteilig ist, dass eventuell missverständliche Frage bei einer asynchronen Bearbeitung des Fragebogens nicht erläutert werden können, da kein Untersuchungsleiter in der Nähe ist [Sharp et al. 07]. Aus diesem Grund muss bei der Erstellung eines Fragebogens die Wortwahl besonders beachtet werden. Die asynchrone Bearbeitungsweise birgt zudem den Nachteil, dass im Gegensatz zu einem Interview nicht weiter in die Tiefe gegangen werden. Desweiteren ist die Rücklaufquote bei postalischer oder Onlinebefragung meist gering, weshalb es vorteilig sein kann die Motivation durch einen bestimmten Anreiz, wie der Chance auf einen Gewinn oder eine finanzielle Vergütung zu erhöhen.

Zwei der befragten Personen aus der Automobilindustrie sehen Fragebögen als wichtigstes Mittel der Anforderungsanalyse an. Die Begründung für diese Entscheidung ist, dass hierbei die Auswertung normiert erfolgen kann und der Aufwand dadurch geringer bleibt, als bei den anderen Methoden. Dieser Vorteil spielt in der Automobilindustrie eine sehr große Rolle, da hier je nach Ziel der Untersuchung sehr viele Menschen befragt werden können und die Auswertung von beispielsweise Einzelinterviews viele zeitliche und personelle Ressourcen in Anspruch nehmen würde. Allerdings wird darauf hingewiesen, dass durch diese Methode die kreativen Ideen der Befragten nicht erfasst werden können und zu diesem Zwecke andere Methoden, wie beispielsweise Interviews oder Fokusgruppen herangezogen werden müssen.

3.3.5. Beobachtungen

Beobachtungen in frühen Stadien eines Entwicklungsprozesses eignen sich sehr gut, um die Umgebung, Aufgaben und Ziele des Nutzers zu verstehen. In späteren Stadien kann durch Beobachtung herausgefunden werden, wie gut der neu entwickelte Prototyp den Benutzer bei der Erreichung dieser Aufgaben und Ziele unterstützt.

Beobachtungen können direkt erfolgen, indem der Beobachter Zeit mit dem Beobachteten verbringt und seine Aktivität zu dem Zeitpunkt beobachtet, zu dem sie passiert. Bei einer indirekten Beobachtung hingegen werden Aufnahmen der Aktivitäten im Nachhinein untersucht. Eine Kombination beider Varianten ist ebenfalls möglich.

Je nach Ziel der Beobachtung, wird das Beobachtungsumfeld gewählt. Dieses kann einerseits dem natürlichen Umfeld des Beobachteten entsprechen, in welchem der Beobachter diesen bei der Verrichtung seiner alltäglichen Tätigkeiten beobachtet. In diesem Fall spricht man von Feldstudien. Beobachtungen können jedoch auch in einer kontrollierten Umgebung, wie einem Usability Labor erfolgen, in welchem der Beobachtete bei der Verrichtung spezifischer, vorgegebener Aufgaben observiert wird.

Die Variante der Feldstudie wird oftmals gewählt, weil Menschen nur schwer beschreiben können, was sie genau tun [Sharp et al. 07]. Manchmal lassen sie wichtige Details weg, weil sie sie als irrelevant oder als selbstverständlich erachten. Solche Details können jedoch bei einer Beobachtung durch externe Personen aufgespürt werden. Bei Beobachtungen, die über längere Tage gehen, kann der Beobachter zur Protokollierung seiner Erkenntnisse beispielsweise auf Diaries zurückgreifen [Sharp et al. 07]. Auch handschriftliche Notizen oder ein Diktiergerät sind geeignete Werkzeuge zur Erfassung der Daten. Zusätzlich kann der Beobachter Fotos machen, um seine Erkenntnisse damit anzureichern. Da eine Videokamera zum Einen aufdringlich und zum Anderen wenig portabel ist, eignet sie sich für diese flexible Art der Datenerfassung weniger gut.

Bei einer schlechten Planung können Beobachtungen sehr viele Daten hervorbringen, die zum Teil irrelevant sind. Daher ist es wichtig, den Fokus der Beobachtung nicht aus den Augen zu verlieren, selbst wenn man als teilnehmender Beobachter in das Umfeld der Beobachtung integriert ist [Sharp et al. 07].

Eine besondere Form der Feldstudie ist die Ethnographie, bei der der Beobachter die Rolle eines Mitglieds der Gruppe einnimmt und versucht das Leben und die Strukturen dieser Gruppe aus deren Sichtweise zu verstehen.

Da es bei der Entwicklung in der Automobilindustrie darum geht, Systeme für Kunden und potentielle Neukunden zu produzieren, ist der Feldforschungs-Ansatz der Beobachtungen oder gar die teilnehmende Beobachtung im Rahmen einer Anforderungsanalyse weniger einfach zu vollziehen. Hier müsste den besagten Personen gefolgt werden und diese müssten in ihrem natürlichem Umfeld, dem Fahren, beobachtet werden. Eine solche Art der Untersuchung könnte für den Beobachteten unangenehm sein. Zudem wäre dies bei mehreren Beobachteten mit einem erheblichen Aufwand verbunden. Interviews oder Fragebögen sind hier eine zweckmäßigere Alternative.

Während bei Feldstudien der gesamte Kontext und die Interaktion der Beobachteten untereinander wichtig sind, liegt der Fokus bei Beobachtungen in einer kontrollierten Umgebung auf den Details der Tätigkeiten der Beobachteten. Der Aufbau der Ausrüstung ist bei dieser Art von Beobachtungen statisch: eine Videokamera sollte den Gesichtsausdruck des Beobachteten erfassen und gegebenenfalls sollten auch seine Körpersprache sowie seine Interaktionen aufgezeichnet werden [Sharp et al. 07]. Solche Aufbauten sind auch im automotiven Kontext denkbar, um die Interaktion mit bestehender Software im realen Kontext ihrer Benutzung zu testen und um die ökologische Validität der

Untersuchung zu erhöhen. Je nachdem, ob es sich bei der Beobachtung um eine stationäre oder eine mobile Untersuchung handelt, kann die Umgebung besser oder schlechter kontrolliert werden. Werden Untersuchungen beispielsweise in einem Modellfahrzeug in einem Labor vorgenommen, kommt dies der kontrollierten Umgebung eines klassischen Usability Labors sehr nahe. Werden die Untersuchungen in einem fahrenden Fahrzeug im Straßenverkehr getätigt, kann dieses Maß an Kontrolle nicht gewährleistet werden.

Der Nachteil einer Beobachtung ist, dass der Beobachter die Gedanken des Beobachteten nicht erfassen kann, wie dies in einer Interview-Situation möglich ist. Das Stellen von Fragen während einer Beobachtung hingegen ist nicht praktikabel, weil es den Beobachteten von seiner Aufgabe ablenkt. Dieses Problem wird durch das Anwenden der „Laut Denken“-Technik vermindert. Bei dieser Technik sagt der Beobachtete, was er denkt und was er versucht zu tun. So externalisiert er seinen Denkprozess und lässt den Beobachter daran teilhaben [Sharp et al. 07]. Der Beobachter wird auf diese Weise auf Problemstellen sowie dem Grund ihrer Entstehung aufmerksam gemacht: zum Beispiel darauf, dass eine unpassende Wortwahl zu Verwirrungen führt.

Zwei der befragten Personen aus der Automobilindustrie empfinden Beobachtungen als wichtigste Methode im Rahmen einer Anforderungsanalyse. Der Vorteil liegt auch für sie in der Natürlichkeit der Methode und ihrer Fähigkeit, Probleme mit bestehenden Systemen objektiv und ungefiltert wiederzugeben. Bei Befragungen hingegen ist dies nicht möglich, da die Aussagen vom subjektiven Empfinden beeinflusst werden. Bei Beobachtungen in der Automobilindustrie werden Probanden zumeist beim Bearbeiten von Bedienungsaufgaben observiert.

3.3.6. Tagebuch-Methode

Wenn der Beobachter bei einer Studie nicht vorhanden sein kann, weil diese sonst beispielsweise zu aufdringlich wäre, können beispielsweise Tagebücher als Werkzeug der indirekten Beobachtung Anwendung finden. Am Ende der Interaktion mit einem System oder am Ende eines Tages schreiben die Teilnehmer einer Studie mit dieser Methode auf, was sie wann getan haben und was ihnen dabei leicht oder schwer fiel. Der Vorteil von Tagebüchern ist, dass sie kostengünstig sind, keine Vorkenntnisse erfordern, auch bei einer örtlich verstreuten Zielgruppe angewandt werden können und sie für Langzeitstudien geeignet sind. Sind sie zusätzlich mittels einiger Stichworte oder Fragen vorformatiert, ermöglicht dies ein einfacheres Ausfüllen sowie eine einfachere Analyse. Problematisch ist hier, dass die Teilnehmer ihre Tagebücher regelmäßig ausfüllen müssen, um eine vollständige Analyse gewährleisten zu können. Darüber hinaus neigen Personen oftmals zur subjektiven Verzerrung ihrer Empfindungen, wenn sie diese im Nachhinein notieren müssen. So können Ereignisse als angenehmer beziehungsweise unangenehmer oder als langwieriger beziehungsweise kürzer erinnert werden, als sie tatsächlich waren. Der Einsatz weiterer Medien, wie das selbstständige Benutzen einer Fotokamera oder Tonaufzeichnungen können den Teilnehmern in solchen Situationen beim Rekapitulieren der Ereignisse helfen.

Tagebücher erscheinen im Rahmen einer Anforderungsanalyse in der Automobilindustrie als weniger praktikabel, da es den potentiellen Befragten als seltsam erscheinen könnte, regelmäßig über ihre Erlebnisse während dem Fahren Buch zu führen. Dies ist vermutlich der Grund dafür, dass keine der Befragten des Fragebogens diese Methode als wichtig ansah [Fragebogen 09]. Im Rahmen einer Evaluation könnte diese Methode jedoch auch in der Automobilindustrie Anwendung finden (vgl. Kapitel 3.4.2.).

3.3.7. Gas Station Flash Survey

Bei der Anforderungsanalyse in der automotiven Produktentwicklung kann auch eine besondere Art von Interviews angewendet werden. Diese Methode, die sich die Akquise von Autofahrern an Tankstellen zu Nutze macht, wurde 2007 von [Kern & Schmidt 07] am Lehrstuhl für Pervasive Computing und User Interface Engineering an der Universität Duisburg Essen entwickelt. Sie nennt sich „Gas Station Flash Survey“ und zeichnet sich durch zwei Kerncharakteristiken aus: den Ort, an dem die Befragung durchgeführt wird sowie die Dauer der Befragung. Die Grundidee ist, dass Autofahrer während dem Betanken ihres Fahrzeuges an einer Tankstelle befragt werden. Da eine Tankfüllung zwischen zwei bis fünf Minuten dauert, ist die Befragung dementsprechend kurz. Die Bereitschaft zur Teilnahme an der Befragung ist sehr hoch, wie die Durchführung eines solchen Tankstellenblitzinterviews gezeigt hat. Der Grund hierfür ist darin zu sehen, dass die Befragten während des Tankvorganges typischerweise keine anderen Aufgaben haben und daher eher bereit sind, ihre Zeit zu investieren [Kern & Schmidt 07].

Sofern einige Punkte vor der Durchführung beachtet werden, ermöglicht diese Methode vielfältigen Input in einem kurzen Zeitraum. Wichtig ist es, nicht nur eine große Tankstelle auszuwählen, sondern auch einen Tag, an welchem viel Verkehr an der Tankstelle herrscht. In Deutschland eignen sich hierfür insbesondere Montage, Freitage und Samstage [Kern & Schmidt 07].

Essentiell für das Gelingen einer solchen Befragung ist das sorgfältige und rationelle Auswählen der Interviewfragen. Diese sollten ein spezifisches Ziel verfolgen und leicht verständlich sein. Dabei können sie offen oder geschlossen sein, doch es ist es ist von größter Wichtigkeit eine Pilotstudie durchzuführen, um zu gewährleisten, dass die Fragen und Antworten in einer angemessenen Zeitspanne bleiben. Ansonsten müsste das Interview frühzeitig abgebrochen werden. In diesem Zusammenhang empfiehlt es sich, Antwortmöglichkeiten zu notieren, um den Befragten gegebenenfalls Denkanstöße liefern zu können. Auf diese Weise wird auch das Protokollieren des Interviews vereinfacht, da der Interviewer die Antworten lediglich auf seinem Formular ankreuzen muss. Eine zusätzliche Aufzeichnung von Audiodaten kann dennoch von Vorteil sein.

Die von [Kern & Schmidt 07] durchgeführte Studie fand an einem Freitag von 10:00-12:00 Uhr sowie von 14:00-16:00 Uhr statt. In diesem Zeitraum wurden 58 Tankstellenbesucher während dem Betanken ihres Fahrzeugs angesprochen und alle erklärten sich bereit, bei der Befragung zu partizipieren. Zwei offene Fragen zum Thema „Telefonieren im Fahrzeug“ wurden gestellt, wobei auf

die erste Frage mit 144 spontanen Antworten reagiert wurde. Weitere 55 Antworten lieferten die Befragten nach Nennung der Denkanstöße. Die zweite Frage lieferte weitere 127 Antworten. Da sie der ersten Frage ähnlich war, wurden hier keine Denkanstöße mehr gegeben. Die Befragung lieferte somit innerhalb von nur vier Stunden mehr als 300 qualitativ hochwertige Antworten [Kern & Schmidt 07].

Die Vorteile dieser Methode für die Anforderungsanalyse sind unter anderem in den geringen zeitlichen und finanziellen Kosten zu sehen. Darüber hinaus ist die Methode gut geeignet um die Zielgruppe Fahrer oder auch Beifahrer zu befragen. In diesem Zusammenhang können auch weitere Einschränkungen in Bezug auf die Zielgruppe leicht vorgenommen werden. So können beispielsweise nur Fahrer einer bestimmten Fahrzeugmarke oder Fahrzeugklasse herausgefiltert werden [Kern & Schmidt 07]. Im Gegensatz dazu kann der Untersuchungsleiter beispielsweise bei der Auswertung von Online-Fragebögen nicht wissen, wer die Fragen beantwortet hat. Ebenfalls im Gegensatz zu Online-Fragebögen oder auch im Gegensatz zu Befragungen an anderen Orten, wie einem Einkaufszentrum, fällt die hohe Bereitschaft zur Teilnahme ins Auge. Während es beispielsweise Personen in Einkaufszentren zumeist eilig haben, ist die Zeit der Tankfüllung typischerweise ungenutzt und eignet sich daher gut zur Beantwortung einiger Fragen [Kern & Schmidt 07].

Die Tankstelle als Befragungsort birgt den weiteren Vorteil, dass die Fragen direkt aus dem Fahrkontext heraus beantwortet werden können. Der Befragte kann sich gut in die Fahrsituation hineinversetzen, da diese nur wenige Sekunden zurückliegt. Bei Online-Fragebögen ist dies zumeist nicht der Fall.

Ebenfalls positiv zu bewerten ist, dass durch die Blitzinterviews ein initialer Kontakt zu der Zielgruppe hergestellt wird und die Befragten dabei um Erlaubnis gebeten werden können, sie anzurufen um sie für spätere Interviews oder Benutzerstudien zu gewinnen [Kern & Schmidt 07].

Nachteilig ist die geringe Zeitspanne, die für das Interview verfügbar ist. Nach der Tankbefüllung wollen die Befragten ihren Weg fortsetzen und das Interview muss notfalls abgebrochen werden. Für längere und detailliertere Interviews ist diese Methode nicht geeignet. Das Austeilen von Fragebögen ist aus Zeitgründen ebenfalls nicht handhabbar. Zudem waren die 56 der 58 befragten Personen männlich. Um eine geschlechterunabhängige Befragung zu gewährleisten müsste an einer größeren Tankstelle und vermutlich über einen längeren Zeitraum besser gefiltert werden.

Um dem Problem der geringen Zeitspanne entgegenzuwirken, ist es denkbar, die Befragung an einer Autowaschanlage durchzuführen [Kern & Schmidt 07]. Auch hier kommen die Befragten direkt aus dem Fahrkontext und haben während dem Waschvorgang typischerweise keine anderen Aufgaben, sodass auch hier mit einer hohen Teilnahmebereitschaft gerechnet werden kann. Da die Autofahrer in diesem Szenario länger warten müssen, können hier mehr Fragen gestellt werden, die in einem höheren Detailgrad beantwortet werden könnten.

3.4. Benutzerorientierte Methoden in der Phase der Evaluation

Eine Evaluation eines Systems kann und soll in verschiedenen Stadien der Entwicklung stattfinden. Frühe Evaluationen der Anforderungen beispielsweise können dabei helfen, diese zu verifizieren, damit eine bessere Anforderungsspezifikation erreicht wird. Etwas später im Entwicklungsprozess können durch den frühen Einsatz von Evaluationen Schwachstellen im grundlegenden Konzept rechtzeitig identifiziert werden [Zimmermann et al. 07]. Am Ende eines Projekts steht schließlich die summative Evaluation, die die Benutzbarkeit und Gebrauchstauglichkeit des finalen Systems erfasst. Wie ein einfacher Lifecycle im Interaktionsdesign aussehen könnte, zeigt Abbildung 3.04.

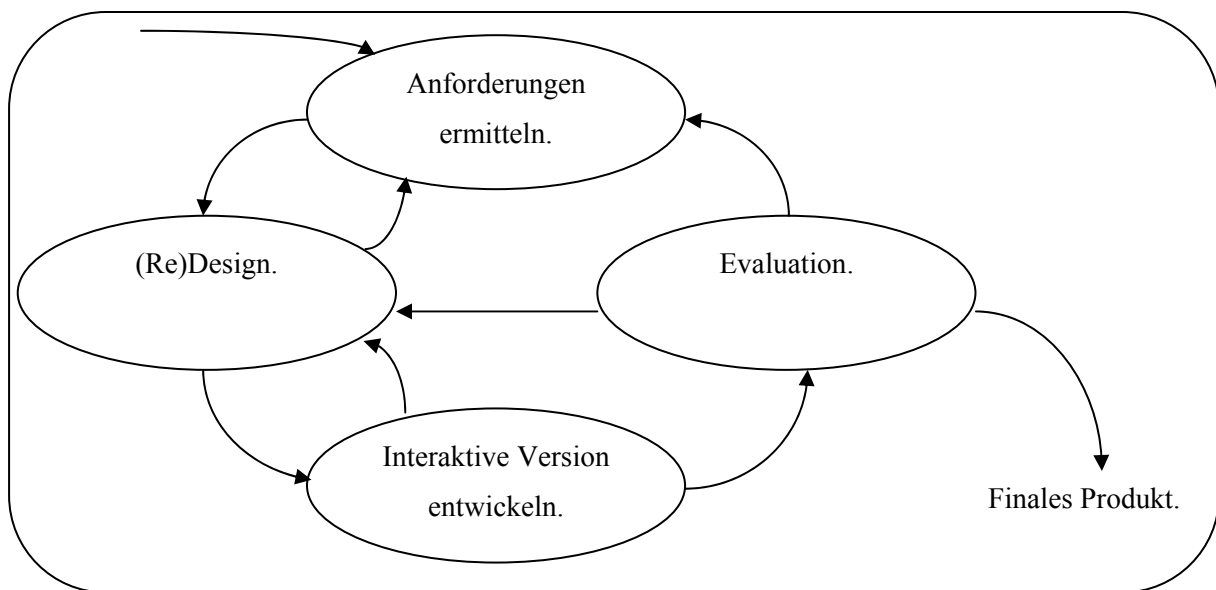


Abbildung 3.04: Einfacher Lifecycle des Interaktionsdesign nach [Sharp et al. 07]

Die Informationen, die in Kapitel 3.4. und sämtliche Unterkapitel eingeflossen sind, entstammen wie die bisherigen der umfassenden Literaturrecherche sowie den persönlich gemachten Erfahrungen der Autorin im Laufe der Kooperation und der in diesem Zusammenhang geführten Gespräche mit Mitarbeitern der Automobilindustrie. Im Gegensatz zum vorigen Kapiteln fließen in dieses jedoch nur die Antworten von acht Fragebögen mit ein, da zwei der Befragten die Beantwortung des selbigen nach den Fragen zur Anforderungsanalyse abgebrochen haben.

3.4.1. Eine Evaluation planen

Am Anfang einer Evaluation steht der Evaluationsplan beziehungsweise das Framework der Evaluation. Sharp und ihre Kolleginnen schlagen als Leitfaden für eine erfolgreiche Evaluation das DECIDE Framework vor [Sharp et al. 07]. Dieses setzt sich aus folgenden sechs Schritten zusammen:

1. **D**etermine the Goals – Ziele festlegen.
2. **E**xplore the Questions – Fragen untersuchen.
3. **C**hoose the Evaluation Approach and Methods – Geeignete Methoden und Evaluationsansatz auswählen.
4. **I**dentify the Practical Issues – Praktische Fragen klären.
5. **D**ecide how to deal with the ethical issues – Umgang mit ethischen Fragen klären.
6. **E**valuate, analyze, interpret, and present the data – Daten evaluieren, analysieren, interpretieren und präsentieren.

D: Das Aufstellen der übergeordneten Ziele ist der erste Schritt einer Evaluation. Je nach Stadium in der Produktentwicklung können Ziele unterschiedlich sein. Sie haben jedoch gemeinsam, dass sie die Rahmenbedingungen der Evaluation festlegen [Sharp et al. 07]. Bei der nachfolgenden Evaluation gilt es, die in diesem Schritt aufgestellten Hypothesen zu überprüfen, in deren Rahmen die abhängigen und unabhängigen Variablen der Untersuchung erläutert werden [Dix et al. 04]. Um dies zu ermöglichen, müssen die Ziele konkret, quantitativ bestimmbar, objektiv und messbar sein [Lindgaard 94]. Je nach Art der festgelegten Ziele ist das Design der Evaluation sehr unterschiedlich: Die Frage, ob das entwickelte System schnell und intuitiv verständlich ist, erfordert ein anderes Testdesign als die Frage, ob die Benutzer des Systems nach einer längeren Nutzung erweiterte Funktionen nutzen können [Dumas & Redish 99].

E: Um das Ziel der Evaluation untersuchen zu können, muss die Kernfrage in viele konkrete Unterfragen aufgeteilt werden [Sharp et al. 07]. Auf diese Weise kann bei der späteren Evaluation der Fokus korrekt gesetzt werden und der Evaluierende weiß, welche Auffälligkeiten er notieren muss, um die Forschungsfragen zu beantworten.

C: Anhand der Ziele und Forschungsfragen müssen im nächsten Schritt geeignete Methoden ausgewählt werden, um diese zu beantworten. Die Auswahl der Methoden wird auch von anderen Aspekten beeinflusst, wie den verfügbaren zeitlichen und finanziellen Ressourcen oder dem Zugang zu möglichen Probanden [Sharp et al. 07]. Diese Aspekte werden im nachfolgenden Schritt diskutiert. Wie in Kapitel 3.3. beschrieben, können verschiedene Methoden kombiniert werden um einen umfassenderes Bild zu den gewählten Zielen und Fragen zu erhalten.

I: Nachdem die Ziele, Fragen und Methoden geklärt sind, müssen die praktischen Fragen der Evaluation geklärt werden. Diese betreffen unter Anderem den Ort und die Dauer der Evaluation, das verfügbare Budget und wie damit umgegangen wird. Eine wichtige Entscheidung, die einerseits von der Fragestellung und andererseits vom verfügbaren Budget abhängt, betrifft das grundlegende Testdesign: so kann zwischen einem Between-Subjects und Within-Subjects Design gewählt werden [Dix et al. 04]. An dieser Stelle muss auch geklärt werden, wie die benötigten Probanden akquiriert werden, damit sie die Zielgruppe des Systems möglichst gut abdecken [Sharp et al. 07].

Auch Fragen bezüglich der Technik gehören zu diesem Schritt. So müssen vor der Evaluation sämtliche Aufzeichnungsgeräte wie Videokameras, Software zur Aufzeichnung des Bildschirms oder

Notizblock und Bleistift verfügbar und funktionstüchtig sein. Die Aufgaben für die Benutzertests, eventuelle Fragebögen und Ähnliches müssen ausgedruckt sein und bereit liegen. Es müssen gegebenenfalls ausreichend Lizenzen für die benötigte Aufzeichnungs- oder Auswertungssoftware verfügbar sein. Auch die Stromversorgung darf während eines Tests nicht unterbrochen werden. Es ist sinnvoll vor der eigentlichen Testreihe eine Pilotstudie durchzuführen, um eventuelle Probleme im Vorfeld aufzudecken und zu beheben.

D: Werden Evaluationen mit Menschen durchgeführt, gilt es einige ethische Grundsätze zu beachten. Datenschutz gehört zu den wichtigsten Aspekten. Dem Probanden muss zugesichert werden, dass seine Daten anonymisiert und nur zum Zwecke der Untersuchung genutzt werden. Dabei sollte er darauf hingewiesen werden, wer Zugang zu seinen Daten erhält [Sharp et al. 07]. Bei einem Bericht über die Studie dürfen Probanden nicht identifiziert werden können, es sei denn sie sind damit ausdrücklich einverstanden [Sharp et al. 07]. Werden Audio- oder Videoaufzeichnungen gemacht, muss der Proband hierfür sein Einverständnis geben. Ein weiterer ethischer Aspekt betrifft das Wohlbefinden des Probanden: sollte er sich während der Untersuchung nicht mehr wohlfühlen, muss er das Recht haben, diese zu jeder Zeit abzubrechen. Ein wichtiger Grundsatz besagt hier, dass eine Person, die bei einer Untersuchung teilnimmt, diese nicht in einem verschlechterten mentalen oder physischen Zustand verlassen darf [Noyes & Baber 99]. Grundsätzlich sollte den Probanden höflich gegenüber getreten werden.

E: Bevor nun mit der Evaluation begonnen werden kann, müssen noch einige Entscheidungen bezüglich der Daten getroffen werden. Insbesondere muss geklärt werden, welche Daten notwendig sind, um die Forschungsfragen zu beantworten und wie diese erfasst werden. Diese Frage wird zwar zum Großteil von der verwendeten Methode beantwortet (vgl. Kapitel 3.3.), doch sollen die Daten beispielsweise zusätzlich statistisch ausgewertet werden, kann dies weitere Techniken der Datenerfassung erfordern. Auch Fragen bezüglich der Reliabilität (sind die Daten reproduzierbar?) und der Validität (misst die Methode, was gemessen werden soll?) müssen in diesem letzten Schritt vor der tatsächlichen Evaluationsstudie beantwortet werden.

3.4.2. Evaluationstechniken zum richtigen Zeitpunkt einsetzen

Wenn potentielle Benutzer an einer Evaluation teilnehmen sollen, so sollte diese zu einem Zeitpunkt stattfinden, in dem zumindest eine Idee in Form einer Skizze oder einem Lo-Fi-Prototypen vorliegt. Idealerweise existiert jedoch ein funktionierender Prototyp des zu untersuchenden Systems, da ein solcher für externe Personen vorstellbarer und greifbarer ist, als eine Skizze. Ein solcher Prototyp kann eine Simulation des Systems ohne Vorhandensein jeglicher Funktionalität oder auch ein komplett ausimplementiertes System sein [Dix et al. 04].

Um die Benutzerfreundlichkeit eines Systems anhand von qualitativen Daten zu analysieren, eignet sich Usability Testing. Hierbei erhält der Benutzer meistens im Kontext einer kontrollierten Umgebung ein Aufgabenset, das er mit dem zu testenden System bearbeitet. Bei der

Zusammenstellung der Aufgaben gilt es nicht nur, die globalen Ziele und Forschungsfragen der Evaluation zu betrachten. Vielmehr sollten auch Dokumente aus der Phase der Anforderungsanalyse integriert werden, damit das Aufgabenset die damals ermittelten Aufgaben und Ziele des Benutzers möglichst vollständig abdeckt [Lindgaard 94].

In diesem Kapitel wird beschrieben, welche Methoden der Evaluation in der Automobilindustrie in den jeweiligen Stadien der Produktentwicklung zum Einsatz kommen. Die Datenbasis für dieses Kapitel ist der von den Mitarbeitern der Automobilindustrie ausgefüllte Fragebogen.

Skizze

Evaluationen erfolgen in der Automobilindustrie bereits in frühen Stadien. Die Befragten des Fragebogens geben an, dass erste Interviews oder in seltenen Fällen auch Fragebögen oder Fokusgruppen stattfinden, sobald erste Skizzen vorliegen. In diesen Fällen werden konzerninterne Experten oder Mitglieder des Probandenpools an einem PC oder Tisch gebeten, um die vorhandenen Ideen zu bewerten und zu klassifizieren. Dies hat zum Ziel, Ideen im Vorfeld größerer Projekte zu erweitern oder zu reduzieren. Dabei wird diskutiert, welche Konzepte weiterverfolgt werden sollen. Bei Vorhandensein mehrerer Konzeptvarianten, wird teilweise schon in diesem Stadium der Produktentwicklung diskutiert, welche davon Potential haben und welche wieder verworfen werden [Fragebogen 09].

In diesem Stadium eignet sich eine Diskussion mit konzerninternen Experten besonders gut, da in dieser Phase noch die Möglichkeiten abgeschätzt werden müssen. Wie bereits oben erwähnt sollten externe Probanden erst zu einem späteren Zeitpunkt in die Entwicklung eingebunden werden.

Lo-Fi-Prototyp

Erste richtige Tests finden dann in der Phase der Produktentwicklung statt, in der bereits ein Lo-Fi-Prototyp des Konzeptes existiert. Die meisten Untersuchungen werden hier an einem Tisch mit PC durchgeführt, doch auch Untersuchungen im Usability Labor oder Fahr Simulator sind zu diesem Zeitpunkt gebräuchlich. Seltener werden Untersuchungen mit Lo-Fi Prototypen laut Fragebogen im Kontext eines stehenden oder fahrenden Fahrzeugs durchgeführt. Der Grund hierfür könnte sein, dass in diesem Stadium noch viele Änderungen am Prototypen möglich sind und eine Integration ins reale Fahrzeug mit mehr Kosten als Nutzen verbunden wäre.

Hauptsächlich durch die genaue Beobachtung der Probanden beim Bearbeiten von Aufgaben oder beim Ausprobieren des Lo-Fi-Prototyps wird in dieser Phase der Gebrauchstauglichkeit und Bedienungsproblemen einzelner Funktionen objektiv nachgegangen [Fragebogen 09]. Um die Gründe für diese Problemstellen zu ermitteln, ist es sinnvoll wenn die Probanden die Technik des lauten Denkens anzuwenden. Um die Daten der Benutzer zu erfassen, eignen sich viele der in Kapitel 3.1.

aufgezählten Techniken. Während das handschriftliche Notieren eine kostengünstige und schnell zu analysierende Möglichkeit darstellt, können beispielsweise mittels Videoaufzeichnung sämtliche Aussagen und Aktionen des Probanden aufgezeichnet werden.

Im Fragebogen wurde gesagt, dass in dieser Phase der Produktentwicklung neben Beobachtungen zum Teil auch Interviews und Fragebögen und in seltenen Fällen auch Fokusgruppen eingesetzt werden, die das Ziel verfolgen, Spezifikationsfehler zu identifizieren und zu eliminieren, bevor zusätzliche Zeit und Kosten in die Weiterverfolgung der Konzepte investiert wird. Die in diesem Stadium recht häufig angewandten Interviews bieten im Gegensatz zu Beobachtungen einen strukturierteren Ansatz zur Erfassung der Evaluationsdaten. Eine gewisse Struktur liefert ein gewisses Maß an Konsistenz zwischen verschiedenen Interviews [Dix et al. 04] und ermöglicht so eine bessere Vergleichbarkeit der Ergebnisse. Darüber hinaus bleibt der Fokus der Evaluation stets präsent. Semi-strukturierte Interviews haben in der Phase der Evaluation den Vorteil, dass auf Bedienprobleme detaillierter eingegangen werden kann, um deren Ursprung zu ergründen. Darüber hinaus helfen sie dabei, die subjektiven Empfindungen des Probanden nach der Evaluation nochmals zusammenzufassen und können gegebenenfalls zu Vorschlägen zur Behebung von Problemstellen führen.

Hi-Fi-Prototyp

Existiert bereits ein Hi-Fi-Prototyp eines Systems, wird das Testsetting vom Tisch zum Fahr Simulator oder gar zum echten Fahrzeug verlagert und es können andere Aspekte untersucht werden. Die Bedienbarkeit und Akzeptanz des zu untersuchenden Systems spielen auch hier eine wichtige Rolle und werden mittels Beobachtungen und Fragebögen gemessen. Etwas seltener finden Interviews und auch Fokusgruppen im Kontext des Fahr Simulators Anwendung [Fragebogen 09]. Der Grund dafür, dass Fokusgruppen eher in den anderen Szenarien stattfinden könnte der sein, dass das Fahrzeug oder auch der Fahr Simulator ein Einzelsetting ist. Hier mehrere Personen im Rahmen einer Fokusgruppe einzuladen, würde wenig Sinn machen, da nicht ausreichend Platz zur Verfügung steht. Zudem empfiehlt sich für Fokusgruppen ein Setting, in welchem alle Beteiligten bequem an einem Tisch Platz nehmen können, da diese eine längere Aktivität sind, die zudem oftmals die Interaktion zwischen den Teilnehmern erfordert.

Sobald nun ein Hi-Fi Prototyp vorhanden sind, können zum ersten Mal essentielle, sicherheitskritische Aspekte des entwickelten Systems untersucht werden. Der Fahr Simulator wird in der Automobilindustrie häufig dazu verwendet, um Dual-Task-Untersuchungen durchzuführen. Hierbei fährt der Proband auf einer vordefinierten Strecke unter vordefinierten Bedingungen (vgl. Kapitel 3.4.5.2.) und muss dabei einer zweiten Aufgabe, wie beispielsweise dem Bedienen eines Fahrerassistenzsystems nachkommen. Ausgewählte Kriterien wie die Blickzu- und -abwendung, Spurabweichung, Fahrfehler, Bedien- und Reaktionen werden dabei gemessen, um ein Maß für die Fahrleistung sowie die Ablenkung durch die Sekundäraufgabe zu erhalten [Fragebogen 09]. Diese

Testumgebung eignet sich bei der Untersuchung von Systemen für den Fahrer sehr gut, da die oben genannten Faktoren in einer sicherheitsunkritischen Umgebung untersucht werden können. Im Straßenverkehr wäre dies nicht ohne Weiteres möglich. Eine genauere Betrachtung von Fahrsimulatorstudien sowie ihrer Vor- und Nachteile findet sich in Kapitel 3.4.5.2.

Wenn Systeme für Personen auf dem Beifahrersitz oder im Fondbereich gestaltet werden, macht eine Untersuchung im Fahrsimulator wenig Sinn. In solchen Fällen sind Untersuchungen im realen Fahrzeug besser geeignet, da sie ein realistischeres Bild der Umgebung bieten. Doch im realen Fahrzeug lassen sich auch Hi-Fi-Prototypen von Fahrersystemen sehr gut untersuchen, wenn die Forschungsfrage nicht den Aspekt der Ablenkung durch eine Sekundäraufgabe betrifft. Diese Umgebung eignet sich insbesondere, um die Bedienbarkeit des Systems als Primäraufgabe zu untersuchen und um zu überprüfen, welche grundlegenden Probleme das aktuelle System birgt. Neben Bedienproblemen und Fehlerraten können jedoch auch subjektive Aspekte wie beispielsweise die Akzeptanz in einer solchen Umgebung sehr gut überprüft werden [Fragebogen 09].

Da sich der Proband im realen Fahrzeug und somit in der tatsächlichen, späteren Umgebung des Systems befindet, empfindet er die Situation als realistisch und kann eher die Handhabbarkeit, Wertigkeit und Anmutung des zu testenden Systems beurteilen. Fragen in Bezug auf die Ergonomie des Systems werden aus diesem Grund in der Automobilindustrie oftmals in einem stehenden Fahrzeug untersucht.

Fertiges Produkt

Ist das System schließlich fertig, sind Fokusgruppen die favorisierte Methode der Evaluation [Fragebogen 09]. Fragebögen werden in diesem Stadium der Entwicklung auch mit einer gewissen Regelmäßigkeit eingesetzt, wohingegen Beobachtungen oder Interviews etwas seltener Gebrauch finden. Bei der Durchführung von Fokusgruppen beginnt typischerweise der Moderator mit der Vorstellung des Systems und fährt anschließend mit der Besprechung einiger im Voraus bestimmter Sachverhalte fort [Noyes & Baber 99]. Verschiedene Aspekte des Systems können auf diese Weise in der Gruppe diskutiert werden. Ähnlich wie beim Interview reagieren die Teilnehmer von Fokusgruppen spontan auf das vorgeführte System und machen dementsprechend intuitive Aussagen, wohingegen Fragebögen eine gewisse Zeit zum Überlegen einer Antwort einräumen. Subjektive Präferenzen einer Gruppe in Bezug auf die Gestaltung eines Systems können dank Fokusgruppen schnell ermittelt werden, doch bietet diese Evaluationstechnik nur eine geringe Reliabilität und Validität [Noyes & Baber 99].

Die bevorzugte Testumgebung in diesem Stadium der Produktentwicklung ist das reale Fahrzeug. In seltenen Fällen wird das finale System jedoch auch im Fahrsimulator getestet. Auch in diesem Stadium wird die Bedienbarkeit des Systems überprüft und Fragen bezüglich der Ergonomie und Akzeptanz werden geklärt. Ein für die Automobilindustrie sehr wichtiger Aspekt ist die

Preisbereitschaft des potentiellen Kunden. Der ideale Ort, um diese zu überprüfen ist das reale, stehende oder fahrende Fahrzeug, da es den Probanden mit realen und vollständigen Umgebungsinformationen versorgt und sich selbiger in dieser Umgebung das beste Bild vom tatsächlichen Nutzungskontext machen kann.

Manchmal wird das fertige System auch im fahrenden Fahrzeug getestet. Im Gegensatz zum stehenden Fahrzeug ist selbiges hier jedoch als Sekundäraufgabe angesiedelt. Hier wird zudem – im Gegensatz zu Untersuchungen im Fahrsimulator – weitestgehend auf Tests der Ablenkungswirkung verzichtet. Die Frage, die in dieser Testumgebung beantwortet werden soll, ist eher, ob während der Fahrt Probleme mit dem System auftauchen. Insbesondere eignet sich diese Testumgebung, wenn gewisse Umgebungsbedingungen für die Untersuchung der Forschungsfrage essentiell sind, in einem Fahrsimulator jedoch nicht simuliert werden können. Die Kräfte der Beschleunigung oder Erschütterungen sind Beispiele für solche Bedingungen [Fragebogen 09].

Das fahrende Fahrzeug stellt zwar die realistischste der automotiven Testumgebungen dar, jedoch ist es bei diesen Untersuchungen von größter Bedeutung, eine unnötige Gefährdung der Probanden zu verhindern. Aus diesem Grund wird das fahrende Fahrzeug eher im Kontext von Wettbewerbsvergleichen oder der vergleichenden Konzeptbewertung durchgeführt [Fragebogen 09]. Teilweise finden diese Untersuchungen auch über einen längeren Zeitraum statt, in welchem der Proband idealerweise mehrere Fahrzeuge mit verschiedenen Systemen fährt. Solche Untersuchungen im Feld erhöhen zum einen die ökologische Validität der Evaluation und können zum anderen auch Veränderungen des Benutzerverhaltens erfassen. Da bei manchen Fragestellungen kein Untersuchungsleiter dabei sein kann, erfolgt die Datenerfassung in diesen seltenen Fällen auch mittels Tagebuch-Methode [Fragebogen 09]. Dabei notiert der Proband Auffälligkeiten und seine Bedienungsprobleme im Laufe des Tests in einem Tagebuch. Durch das zusätzliche Loggen von Nutzerdaten können jedoch auch bei dieser Methode quantitative Daten erfasst werden. Dies hat den Vorteil, dass das Testen unter natürlicheren Bedingungen stattfinden kann und der Proband nicht das Gefühl hat, selbst getestet zu werden. Solche Langzeitstudien eignen sich insbesondere, wenn bereits bekannte Eingabegeräte oder Systeme mit neuen Darstellungs- und Interaktionskonzepten verglichen werden oder um zu überprüfen, ob Bedienungsprobleme auch noch nach längerer Nutzung vorhanden sind [Gerken & Reiterer 09].

Häufiger als die Tagebuch-Methode finden jedoch auch in diesem Stadium Fragebögen Anwendung. Mit deren Hilfe kann eine Einstufung der getesteten Systeme und Konzepte vorgenommen werden kann. Sie werden in allen Entwicklungsstadien mit einer gewissen Regelmäßigkeit und Häufigkeit durchgeführt und stellen im Gegensatz zu Interviews oder Fokusgruppen eine alternative, aber weniger flexible Art der Befragung dar. Durch ihre stärkere Strukturiertheit ist jedoch ihre Analyse einfacher und schneller zu bewerkstelligen [Dix et al. 04]. Im Rahmen einer Evaluation können sie helfen, bestimmte Aspekte eines Systems zu beleuchten.

Fragebögen eignen sich im Allgemeinen eher für das Erfassen subjektiver Maße wie der Zufriedenheit des Befragten [Kirakowski 00]. Objektive Maße wie die Effektivität oder Effizienz eines Systems können mit Fragebögen nicht beantwortet werden, da das Empfinden dieser Maße subjektiv beeinflusst ist. Für diese Werte bedarf es einer anderen Datenerfassungstechnik wie beispielsweise der Beobachtung oder dem automatischen Erstellen einer Logdatei.

Es existiert eine Vielzahl standardisierter und valider Fragebögen, mit deren Hilfe die subjektive Zufriedenheit oder die Gebrauchstauglichkeit eines Systems untersucht werden können. Mit Hilfe des Fragebogens AttrakDiff lässt sich beispielsweise aus Sicht des Benutzers die Attraktivität des untersuchten Systems in Hinblick auf Aussehen und Bedienbarkeit ermitteln [Hassenzahl et al. 09]. Anhand von Wortpaaren auf einer semantischen Differentialskala wird die hedonische und pragmatische Qualität eines Systems analysiert. Die pragmatische Qualität bezieht sich dabei auf die Bedienbarkeit des Systems, wohingegen sich die hedonische Qualität mit dem emotionalen Empfinden des Befragten in Bezug auf das System beschäftigt. Auch in der Automobilindustrie findet dieser Fragebogen zur Bestimmung der subjektiven Zufriedenheit Einsatz.

Ein anderer vorgefertigter Fragebogen, der ebenfalls nach dem Entwicklungsprozess angewandt werden kann, ist der ErgoNorm-Benutzerfragebogen, welcher die subjektiv empfundene Effektivität und Effizienz des Systems beschreibt und so Hinweise zu Bedienungsproblemen liefert [Dzida et al. 01]. Er ist Teil des ErgoNorm-Prüfverfahrens, welches bei der Bestimmung der Gebrauchstauglichkeit von Softwareprodukten auf der Grundlage von DIN EN ISO 9241-110 und -111 angewendet wird [ISO 9241-11] [ISO 9241-110]. Die Fragen des Fragebogens betreffen die sieben Grundsätze der Dialoggestaltung: Aufgabenangemessenheit, Selbstbeschreibungsfähigkeit, Erwartungskonformität, Lernförderlichkeit, Steuerbarkeit, Fehlertoleranz und Individualisierbarkeit [Geis 06] und der Befragte hat die Möglichkeit diese mit „ja“, „nein“ und „trifft nicht zu“ zu beantworten. Im Falle von Problemen hat er zusätzlich die Möglichkeit, diese genauer zu erläutern und sie gegebenenfalls mit „ist besonders störend“ zu kennzeichnen [Dzida et al. 01]. Am Ende des Fragebogens wird zudem Raum für weitere Anmerkungen geboten. Angewandt wird dieser Fragebogen bei einer summativen Evaluation typischerweise nach der Auslieferung des Systems.

Weitere Fragebögen die sich zur Evaluation einer Software eignen, sind beispielsweise der seit 1987 existierende QUIS zur Messung der Benutzerzufriedenheit [QUIS 09] oder der 1993 zuerst veröffentlichte SUMI zur Beurteilung der Nutzungsqualität aus Sicht der Benutzer [SUMI 09]. Ein Fragebogen, der auf DIN EN ISO 9241-110 beruht, ist IsoMetrics [Willumeit 97]. Diese wurden jedoch weder in den Gesprächen mit Mitarbeitern der Automobilindustrie, noch in den Fragebögen als gebräuchliche Werkzeuge im automotiven Kontext genannt.

3.4.3. Probanden aus dem Probandenpool und externe Probanden

Evaluationen werden in der Automobilindustrie hauptsächlich mit Hilfe der Mitglieder des konzerneigenen Probandenpools vorgenommen. Da sich beispielsweise für den Probandenpool der

Volkswagen AG jeder Mitarbeiter vom Manager zur Dame an der Kantine registrieren kann, ist hier die Altersverteilung sehr hoch und es werden viele unterschiedliche Zielgruppen abgedeckt [Fragebogen 09] [Gespräche] [VW 09]. Da außerdem die meisten Mitarbeiter eines Automobilherstellers zugleich Kunden sind, werden bei den so durchgeführten Evaluationen direkt die Kunden des Herstellers adressiert. Zum Teil werden auch die Kollegen aus der eigenen oder benachbarten Arbeitsgruppe als Probanden für die eigene Evaluation herangezogen [Gespräche]. In der Regel laufen Evaluationen in der Automobilindustrie als mehrstufiger Prozess ab. Zunächst testen etwa 20 konzerninterne Experten, wie beispielsweise oben genannte Kollegen, das System. Anschließend wird mit Hilfe des unternehmensinternen Probandenpools getestet, bevor schließlich bei Bedarf externe Kunden in den Evaluationsprozess eingebunden werden [Fragebogen 09].

Wie in Kapitel 3.3.1. im Rahmen der Anforderungsanalyse erwähnt, haben auch Evaluationsstudien mit Hilfe des Probandenpools organisatorische Vorteile wie Schnelligkeit, Kosteneffizienz und eine unbürokratische Projektabwicklung [Fragebogen 09] [VW 09]. Auch die Kundennähe ist als großer Vorteil anzusehen [VW 09]. Werden für die Evaluation Zielgruppen benötigt, die im Probandenpool nicht abgedeckt sind, so werden diese entweder aus externen Pools [Fragebogen 09] oder aus dem Umfeld der Personen im Probandenpool akquiriert [Gespräche]. Auf diese Weise können beispielweise auch Personen im Rentenalter oder Kinder und Jugendliche für Evaluationen gewonnen werden.

Wird der Probandenpool für Evaluationen herangezogen, so hat dies nicht immer nur Vorteile. Einige Unternehmen benutzen beispielsweise einen bestimmten Wortschatz, der intern zwar sehr gut verständlich ist, externen Personen jedoch Schwierigkeiten bereiten kann. Probleme eines Konzepts in Bezug auf seinen Sprachgebrauch können aus diesem Grunde mit der alleinigen Integration von Mitarbeitern nicht aufgedeckt werden [Dumas & Redish 99].

Darüber hinaus haben Mitarbeiter eines Automobilherstellers oftmals ein bestimmtes Bild von der Marke und ihren Eigenschaften, die nicht immer der gebrauchstauglichsten Variante entsprechen. So kam es in der Automobilindustrie durchaus vor, dass ein neues Design eines Armaturenbrettes mit Hilfe des Probandenpools getestet wurde und in der Evaluation sehr schlecht abschnitt. Der Grund hierfür war, dass für die Beleuchtung der Anzeigetafeln eine für die Marke bisher untypische Farbe eingesetzt wurde. Für die konzerninternen Probanden war dies ungewohnt und führte zur Ablehnung. Aus diesem Grunde wurde ein externes Unternehmen mit der Evaluation beauftragt. Dieses hatte keinen Zugriff auf den Probandenpool und wählte daher die Probanden entsprechend der Zielgruppe selbst aus. Diese zweite Evaluation zeigte, dass das neue Design des Cockpits viel gebrauchstauglicher war, als das alte und dass die Anzeige durch die neue Beleuchtung nun viel besser zu lesen war [Gespräche].

Aus diesem Grund ist es sinnvoll, externe Probanden zu Evaluationen im automotiven Bereich einzuladen und sie, wie bereits in Kapitel 3.3.1. erwähnt, zur Geheimhaltung zu verpflichten.

Alternativ kann jedoch wie im oben genannten Beispiel ein externes Unternehmen zur Evaluation herangezogen werden.

Im Rahmen der Evaluationen in diesem Forschungsprojekt hatten die Entwickler und Untersuchungsleiter keinen Zugriff auf den unternehmensinternen Probandenpool. Auch bei den Kooperationsprojekten mit Universitäten oder ähnlichen Einrichtungen, die im Rahmen der Literaturrecherche ermittelt wurden, schien dies der Fall zu sein (vgl. z.B. [Ablassmeier et al. 06], [Bach et al. 08], [Kern et al. 09]). Der organisatorische Aufwand ist dadurch zwar höher, doch haben diese Evaluationen den großen Vorteil, von der Marke unabhängige Personen zu befragen. Es wird dadurch ein breiteres Spektrum an Markenzugehörigkeiten abgedeckt und wenn einige der ausgewählten Probanden doch Kunden des jeweiligen Automobilherstellers sind, so ist dies zumeist zufällig.

Bei diesen automotiven Forschungsprojekten außerhalb der Automobilindustrie ist ebenfalls eine breit gefächerte Altersverteilung möglich und zumeist wurden die Probanden auch dementsprechend ausgewählt. Bei den meisten gefundenen Untersuchungen wurden Probanden zwischen ca. 20 und ca. 50 Jahren zu den Evaluationen eingeladen (vgl. z.B. [Ablassmeier et al. 06], [Gruenstein et al. 09], [Minin et al. 08], etc.). Allerdings entstammen die Probanden nichtsdestotrotz oftmals aus dem näheren Umfeld der Untersuchungsleiter, das heißt: Personen mit höherer Bildung oder aus der eigenen Arbeitsgruppe (vgl. [Ablassmeier et al. 06]). Die Akquise externer Probanden ist für Entwicklungsteams ohne den Zugriff auf eine Probandendatenbank mit einem sehr hohen organisatorischen Aufwand verbunden, der nur selten zum Erfolg führt. Dies wurde insbesondere bei der Planung der in Kapitel 4.1. durchgeführten Studie festgestellt. Der Grund hierfür ist, dass viele berufstätige Personen ihrer Arbeit nachkommen müssen und daher nur begrenzt Zeit finden können, um an einer Evaluationsstudie mitzuwirken. Berufstätige an einer Universität haben hingegen kürzere Wege und sind eher bereit zwischen zwei Terminen oder nach Feierabend Zeit für eine Evaluation im universitären Kontext zu investieren. Insbesondere gilt dies auch, weil sie eine engere Bindung zu den Untersuchungsleitern und Entwicklern empfinden und daher eher gewillt sind zu helfen, als externe Berufstätige.

Aus diesem Grunde wurden bei der in Kapitel 4.2. vorgestellten Studie die berufstätigen Probanden aus dem universitären Umfeld gewählt. Es wurde jedoch dennoch darauf geachtet, dass ein breites Spektrum an beruflichen Hintergründen und technischer Affinität abgedeckt wird.

3.4.4. Forschungsfragen und erfasste Daten

Wie auch bei Untersuchungen in der traditionellen Mensch-Computer Interaktion, werden auch in der Automobilindustrie zunächst die Forschungsfragen definiert, bevor eine Evaluation stattfindet. Ein Aspekt, welcher dabei auch in der Automobilindustrie eine große Rolle spielt, ist die Gebrauchstauglichkeit des neuen Systems sowie die Akzeptanz des Benutzers. Die Gebrauchstauglichkeit setzt sich auch hier aus den in ISO 9241-11 definierten Aspekten der

Effektivität, Effizienz und der subjektiven Zufriedenheit zusammen [ISO 9241-11]. Die Effektivität bestimmt hierbei das Maß, in welchen das System sein Ziel erfüllt. Die Effizienz besagt, welche zeitlichen oder auch mentalen Kosten mit dem Erfüllen des Ziels verbunden waren. Die subjektive Zufriedenheit schließlich macht eine Aussage über das Nutzererleben.

Oftmals wird in der automotiven Forschung untersucht, ob das neue System eine Verbesserung der Bedienung gegenüber dem bestehenden System liefert. Dabei werden mindestens zwei Systeme, das alte und das neue und gegebenenfalls weitere Varianten des neuen, getestet und deren Güte anhand von Logdaten und der subjektiven Bewertung der Probanden beurteilt. Wie auch bei klassischen Untersuchungen anhand der Methoden der Mensch-Computer Interaktion, werden zur Bestimmung der Performance Bedienzeiten und Fehlerraten gemessen. In der Automobilindustrie sind jedoch nicht nur Fehler in Bezug auf die falsche Bedienung des Systems von Bedeutung, sondern auch Fahrfehler. Zudem spielen bei der Ermittlung der Effizienz eines Systems nicht nur Bedienzeiten, sondern auch die empfundene mentale Belastung bei der Bedienung eine sehr wichtige Rolle. Die Belastung des Probanden kann in der Automobilindustrie auf verschiedene erhoben werden. Einerseits eignet sich hierfür die Skala zur Erfassung subjektiv erlebter Anstrengung (=SEA-Skala), anhand welcher die Gesamtbeanspruchung auf einer metrischen Skala bewertet wird [Eilers et al. 86] [Pataki et al. 05]. Zusätzlich oder alternativ kann der Proband den NASA-TLX („Task Load Index“) ausfüllen. Mit Hilfe dieses Fragebogens werden verschiedene Dimensionen der Beanspruchung, wie beispielsweise mentale oder körperliche Beanspruchung, Anstrengung und Frustration erhoben. Bei der Auswertung kann ermittelt werden, wie sich die Gesamtbelastung des Probanden zusammensetzt [Pataki et al. 05]. Die subjektive Zufriedenheit mit dem zu testenden System wird in der Automobilindustrie zumeist mit Hilfe standardisierter Skalen oder Fragebögen erhoben. Der Proband demonstriert dabei seine subjektive Zufriedenheit anhand einer Bewertung auf einer meist fünfstufigen Likert-Skala, deren Wertebereich zumeist von -- bis ++ reicht. Eine solche Bewertung findet oftmals während und nach dem Test statt [Fragebogen 09]. Auch bei Untersuchungen im Fahrzeug oder Fahrsimulator kann der Wertebereich einer Skala dem Probanden stets angezeigt werden, indem beispielsweise ein beschriftetes Brett mit den möglichen Dimensionen neben dem Armaturenbrett befestigt wird. Auf diese Weise muss der Proband nur auf den jeweiligen Wert deuten und der Untersuchungsleiter notiert sich diesen. Durch diese Art der Messung wird eine schnellere und einfachere Auswertung der Daten ermöglicht, was für die Mitarbeiter der Automobilindustrie einen großen Pluspunkt darstellt, da das Bedürfnis nach einer effizienten Auswertung mehrfach betont wurde. Außerdem eignet sich das Zeigen auch für Untersuchungen während der Fahrt, da es wenig invasiv und ablenkend ist. Die auf diese Weise ermittelten Daten ermöglichen zudem einen schnellen Vergleich zwischen mehreren Varianten eines Systems.

Ein weiteres Werkzeug, welches in der Automobilindustrie zur Ermittlung des subjektiven Empfindens gegenüber einem System verwendet wird, ist das semantische Differential. Insbesondere

der Fragebogen AttrakDiff wird oft genutzt [Fragebogen 09]. Dieser eignet sich ebenfalls gut zum Vergleich mehrerer Konzepte und ermöglicht, genauso wie Likert-Skalen, eine schnelle Auswertung. Zum Teil werden Fragebögen jedoch auch individuell erstellt und in Kombination mit Interviewfragen während und nach dem Test bearbeitet.

Ein sehr wichtiger subjektiver Aspekt für die Entwicklung in der Automobilindustrie ist die optische Gefälligkeit eines Systems. Diese wird nach dem Testen in Form von Interviewfragen ermittelt. In diesem Zusammenhang wird auch das Maß der vom Probanden empfundenen Hochwertigkeit des Systems ergründet, welches Hand in Hand mit der Preisbereitschaft geht. Im Allgemeinen sind die Akzeptanz und die Kundenzufriedenheit wichtige, da kaufentscheidende Kriterien in der Automobilindustrie. Fällt ein Produkt in diesem Zusammenhang durch, ist es sehr unwahrscheinlich, dass es auf den Markt kommt.

Auch Kommentare der Probanden während der Bedienung des Systems fließen in die Ergebnisse mit ein, doch ist diese Art der qualitativen Datenerhebung in der Automobilindustrie eher eine Ausnahme [Fragebogen 09]. Qualitative Inhaltsanalysen erfolgen in der Phase der Evaluation, sofern sie überhaupt durchgeführt werden, eher in frühen Phasen der Produktentwicklung, in denen die Aussagen und Verbesserungsvorschläge der Probanden noch in den weiteren Entwicklungsprozess integriert werden können. In den meisten Fällen erfolgt die subjektive Datenerfassung anhand von Mitschriften des Untersuchungsleiters. Die Analyse erfolgt oftmals basierend auf Software, die die systematische Analyse und Interpretation von Texten ermöglicht (z.B. **MAXQDA**) [Fragebogen 09]. Diese Art der Datenerfassung und -analyse ist wenig zeitaufwändig und auch ihre Auswertung geht schneller von statten. Das sind Aspekte, deren Wichtigkeit von Mitarbeitern der Automobilindustrie immer wieder betont wird.

Aus diesem Grunde erfolgt die subjektive Bewertung zumeist, wie bereits oben beschrieben, anhand quantitativer Daten, wie sie mittels standardisierter Fragebögen, Skalen oder einem subjektiven Ranking verschiedener Systeme erfasst werden können.

Mit fortschreitender Produktentwicklung nehmen in der Automobilindustrie die Anteile der quantitativen Untersuchung zu. Wenn ein grundlegendes, subjektiv zufrieden stellendes, einfach zu bedienendes Konzept vorliegt, muss es optimiert werden, sodass die neuen Funktionen die Mindestanforderungen an Systeme im Fahrzeug erfüllen. Diese gelten insbesondere für Systeme, die durch den Fahrer bedient werden sollen. Da dieser während der Bedienung eines solchen Systems nicht von seiner primären Fahraufgabe abgelenkt werden darf, haben verschiedene nationale und internationale Verbände Richtlinien für eine sichere Interaktion im Fahrzeug in diversen Regelwerken festgehalten. Ein solches grundlegendes Regelwerk ist in Deutschland die Straßenverkehrsordnung. Diese erfordert vom Fahrer, es sicherzustellen, dass weder seine Sicht noch sein Gehör durch die Besetzung, Ladung, Tiere oder auch Geräte beeinträchtigt werden [StVO 09]. Doch es ist auch

wichtig, dass der Automobilhersteller dies selbst bei seinen integrierten Systemen in einem gewissen Maße gewährleisten kann. Aus diesem Grunde wird bei Evaluationen von Fahrersystemen auch entsprechend bestimmter Richtlinien wie beispielsweise ESoP oder AAM-Guidelines getestet [Fragebogen 09].

Das European Statement of Principles (=ESoP) wurde 1999 von der Europäischen Kommission entwickelt, und fordert Automobilhersteller und Zulieferer auf, einige Sicherheitsaspekte bei der Gestaltung von automotiven Informations-, Kommunikations- und Unterhaltungssystemen zu berücksichtigen. Die Prinzipien betreffen unter anderem die grundlegende Gestaltung des Systems, die Darstellung der Information sowie die Interaktion mit dem Display und den Bedienelementen [ESOP 99]. Auch die Alliance of Automobile Manufacturers (=AAM) hat 2002 eine Sammlung von Prinzipien entwickelt, welche ebenfalls sowohl von Automobilherstellern als auch von Zulieferern bei der Entwicklung von Informations- und Kommunikationssystemen berücksichtigt werden sollten, um die Sicherheit bei der Interaktion zu gewährleisten [AAM 02]. Diese Prinzipien, die den Namen AAM Guidelines tragen, bauen auf dem ESoP auf und betreffen ebenfalls unter anderem die Darstellung von Informationen sowie die Interaktion mit Display und Bedienelementen. Im Gegensatz zum etwa vage formulierten ESoP sind diese Guidelines konkret überprüfbar [Kuhn 05]. Auch sie werden in der Automobilindustrie im Rahmen von Evaluationen herangezogen [Fragebogen 09].

3.4.5. Dual Task

Wie bereits mehrfach angedeutet spielt bei automotiven Systemen neben der subjektiven Zufriedenheit und der Akzeptanz der Benutzer ein weiterer Aspekt eine gewichtige Rolle. Bei Systemen für den Fahrer ist unbedingt zu beachten, dass die Nutzung eines interaktiven Systems lediglich eine Sekundäraufgabe ist und den Fahrer keinesfalls von seiner Primäraufgabe, dem Fahren eines Fahrzeugs ablenken darf. Dieser Aspekt wird Dual-Task genannt und wird als größter Unterschied zu den Untersuchungen in der klassischen Mensch-Computer Interaktion angesehen [Fragebogen 09]. Bei der Untersuchung von Fahrersystemen müssen daher neben den bekannten Logdaten wie Bedienzeiten und Fehlerraten, weitere Daten erfasst werden, die die Belastung des Autofahrers anzeigen.

Bevor es zu einer Evaluation mit richtigen Probanden kommt, können jedoch auch theoretische Modelle herangezogen werden, um zu bestimmen, inwieweit das vorliegende Konzept Potential hat und wie stark seine Ablenkungswirkung sein könnte. Sie haben den Vorteil, dass sie auch ohne das Vorhandensein eines Prototyps durchgeführt und daher zu einem sehr frühen Stadium der Produktentwicklung eingesetzt werden können. In diesen Phasen wäre eine empirische Bewertung mit Hilfe von Probanden beispielsweise noch nicht möglich. Diese werden zudem dahingehend kritisiert, dass sie erst spät stattfinden können, sodass für die laufende Entwicklung kein Nutzen mehr gewährleistet werden kann [Hamacher et al. 02].

Mit Hilfe theoretischer Modelle können hingegen beispielsweise die Bedienzeiten, die der Benutzer benötigen würde um einzelne Aufgaben mit dem interaktiven System zu bearbeiten frühzeitig

abgeschätzt werden. Darüber hinaus können theoretische Modelle Informationen über die potentielle Ablenkungswirkung eines Systems liefern oder auch modellieren, wie schnell der Fahrer in einer kritischen Situation reagieren kann. Da es sich in diesem Szenario um ein sicherheitskritisches Umfeld handelt, ist das frühzeitige Vorhandensein solcher Informationen besonders wichtig. Doch auch bei der Entwicklung für Beifahrer oder Mitreisende auf dem Rücksitz, ist es hilfreich, schon früh im Entwicklungsprozess Informationen über die potentielle Gebrauchstauglichkeit des Systems zu erhalten.

3.4.5.1. Theoretische Modelle

Theoretische Modelle simulieren das Benutzerverhalten und können so die Gebrauchstauglichkeit oder Ablenkungswirkung eines interaktiven Systems vorhersagen. Die subjektive Zufriedenheit mit einem System, wie es bei der Bestimmung der Gebrauchsfähigkeit nach ISO 9241 – Teil 11 der Fall ist, können sie zwar nicht ermitteln, doch sie können die Effektivität und die Effizienz eines Systems durch die Simulation der Interaktion zwischen Benutzer und dem System durchaus vorhersagen [Hamacher et al. 02].

Es gibt verschiedene Ansätze formaler Modelle, die sich zur Modellierung der Interaktion mit einem System eignen. Einige davon lassen sich sehr gut auf den automotiven Kontext übertragen und werden im Folgenden vorgestellt. Wie bereits erwähnt, eignen sich einige dieser theoretischen Modelle bereits in frühen Stadien der Produktentwicklung wie beispielsweise in der Definitions- oder Spezifikationsphase. Bei anderen Modellen bedarf es eines Prototyps, um Vorhersagen über die Gebrauchstauglichkeit treffen zu können. Auch wenn hier bereits die Möglichkeit für empirische Untersuchungen gegeben wäre, kann die Integration theoretischer Modelle sinnvoll sein, um das zeitliche und finanzielle Budget zu schonen.

Multi-Tasking GOMS

Eine Möglichkeit, die Beeinträchtigung der Primäraufgabe durch eine zusätzliche Aufgabe zu berechnen, bietet Multi-Tasking GOMS [Urbas & Leuchter 08].

Das zugrunde liegende Konzept wurde 1983 von Stuart Card und Kollegen entwickelt und berücksichtigt die kognitiven Aktivitäten des Benutzers zur Informationsverarbeitung [Card et al. 83]. Es beruht auf dem Prinzip, dass Benutzer rational handeln, um ihre Ziele zu erreichen und beschreibt die Aktionen, die Menschen mit interaktiven Geräten durchführen [Hamacher 02]. Das GOMS-Modell besagt, dass die kognitive Struktur des Benutzers sich aus den vier Faktoren *Goals* (= Ziele), *Operators* (= Operatoren), *Methods* (= Methoden) und *Selection Rules* (= Auswahlregeln) zusammensetzt [Card et al. 83]. *Ziele* definieren den Zustand, den der Benutzer erreichen möchte. Übergeordnete Ziele bestehen dabei oft aus mehreren Teilzielen. Um diese zu erreichen, wendet der

Benutzer *Operatoren* an, die aus kognitiven Prozessen und physischen Aktionen bestehen. Um Operatoren auszuführen bedarf es erlernter Prozeduren, den sogenannten *Methoden*, welche aus einzelnen, kleinen Schritten bestehen. Mit Hilfe der *Auswahlregeln* können schließlich geeignete Methoden ausgewählt werden, falls es für den jeweiligen Aufgabenteil mehrere Möglichkeiten zur Auswahl gibt. Die Wahl der Auswahlregel hängt hierbei von den Aufgabenparametern, Training oder Gewohnheiten des Benutzers ab [Card et al. 83] [Hamacher et al. 02] [Sharp et al. 07]. Dieses Modell hilft dabei, Fehlerraten und die Zeit vorherzusagen, die zur Erledigung bekannter Abläufe benötigt wird. Darüber hinaus kann es auch qualitative Aussagen hinsichtlich der Umsetzung von Funktionen liefern [Hamacher 02]. Besonders zuverlässig ist seine Anwendung bei Aufgaben, die automatisiert erfolgen und nur ein geringes Maß an höheren kognitiven Prozessen, wie beispielsweise Planen, erfordern. Dies ist im Anwendungsgebiet der „In Vehicle Information Systems“ gegeben [Urbas & Leuchter 08].

Die Anwendung von Methoden der GOMS-Familie ist jedoch auf die Bedienung von Systemen im stehenden Fahrzeug eingeschränkt, da das Standard Modell die Modellierung der Aufgaben nur in ungestörter Umgebung erlaubt [Urbas & Leuchter 08]. Für den Fahrkontext bedarf es einer Erweiterung, die [Urbas & Leuchter 08] in Form von Multi-Tasking GOMS entwickelt haben. Dieser Ansatz besteht aus drei Teilen. Das *Resource Profile* repräsentiert vereinfacht die Belastung des Fahrers während seiner primären Fahraufgabe. Es gibt mehrere Profile, die verschiedene Fahrsituationen wie Stadtverkehr, Autobahn, etc. repräsentierten. Das *Task Model* ist in einer GOMS-ähnlichen Notation mit Multi-Tasking Erweiterungen beschrieben und stellt die sekundäre Aufgabe, das Benutzen des „In Vehicle Information Systems“, sehr detailliert dar. Im Gegensatz zum traditionellen GOMS-Modell werden hier Unterbrechungen und Wiederaufnahme des Arbeitsflusses ebenfalls betrachtet. Der zentrale Teil ist die *Interference Engine*, welche die Beschreibungen der primären und sekundären Aufgabe kombiniert und das Multi-Tasking Modell erstellt, das anschließend in Simulationen verwendet werden kann [Urbas & Leuchter 08]. Das Konzept Multi-Tasking GOMS sowie seine Implementation mtGOMS hat sich als valides Hilfsmittel im Rahmen der Entwicklung von „In Vehicle Information Systems“ erwiesen [Urbas & Leuchter 08].

ACT-R

Einen alternativen Ansatz zur Modellierung der Kognition des Benutzers stellt die kognitive Architektur ACT-R (Atomic Components of Thought - Rational) dar. Mit Hilfe dieser Theorie können die Schritte der Wissensverarbeitung, insbesondere das Lernen und Behalten von Informationen, erklärt werden [Hamacher 02]. Hierbei werden die symbolischen Strukturen der Kognition durch ein Produktionensystem dargestellt. Ihre subsymbolischen Strukturen werden durch eine Vielzahl paralleler kognitiver Prozesse bestimmt, die durch mathematische Gleichungen beschrieben werden

können [Dzaack 08]. Der Wissenserwerb entspricht dabei dem Lernen von Produktionen beziehungsweise von Produktionsfolgen [Hamacher 02]. Bei diesem Konzept werden kognitive Strukturen und Prozesse in verschiedenen Modulen zusammengefasst, wie beispielsweise im visuellen Modul oder im motorischen Modul. Die Module sind dabei voneinander unabhängig. Das Verhalten der Module sowie das gesamte künstliche kognitive System werden durch das zentrale Produktionssystem koordiniert [Dzaack 08]. Die Verwendung dieser Architektur liefert unter anderem auch Aussagen über den Zeitbedarf einzelner Prozesse [Hamacher 02].

Um zu überprüfen, ob eine Simulation der kognitiven Gefüge mittels ACT-R auch im Vergleich zum Menschen gültig ist, können die Ergebnisse der Modellsimulationen mit einer empirischen Datenbasis verglichen werden. [Dzaack 08] integriert Blickbewegungsdaten in seinen Vergleich, um einen umfassenderen Vergleich zwischen modellierter und tatsächlicher Kognition zu erhalten. Das visuelle Modul von ACT-R hilft ihm dabei, indem es die Wahrnehmung von Umgebungsinformationen durch die aktive Aufmerksamkeitsverschiebung des visuellen Systems simuliert [Dzaack 08]. Eine Einschränkung der aktuellen Version von ACT-R ist die Tatsache, dass Aspekte wie Ablenkung und peripheres Sehen nicht umgesetzt sind [Dzaack 08].

Das Werkzeug SimTrA (Simulation Trace Analyzer), welches an der Technischen Universität in Berlin entwickelt wurde, unterstützt den Entwickler bei der Analyse und dem Vergleich von Leistungsdaten kognitiver Benutzermodelle und empirischer Daten [Dzaack 08]. Das Werkzeug kann – wie auch Multi-Tasking GOMS – bereits in frühen Phasen des Entwicklungsprozesses eingesetzt werden, um eine frühzeitige, formale Bewertung verschiedener Benutzerschnittstellen zu erhalten. Bis zu drei unterschiedliche Datensätze (z.B. Mensch, Modell 1, Modell 2) können mittels SimTrA verglichen werden. Für seinen Versuch hat [Dzaack 08] drei Alternativen für die Mensch-Maschine-Schnittstelle einer Prozesssteuerungsanlage entwickelt, die jeweils mit 34 Versuchspersonen, einem Top-Down-ACT-R-Modell und einem Bottom-Up-ACT-R-Modell bearbeitet wurden (eine detaillierte Beschreibung der Modelle ist in [Dzaack 08] zu finden).

Zu Beginn des Experimentes wurden drei Hypothesen aufgestellt: 1. Die drei Schnittstellen induzieren unterschiedliches Verhalten. 2. Die zwei kognitiven Modelle zeigen unterschiedliches Verhalten. 3. Das Bottom-Up-Modell prädiziert das menschliche Verhalten besser als das Top-Down-Modell. Die Ergebnisse der Performance-Daten konnten die erste Hypothese nicht bestätigen, da innerhalb der drei Gruppen keine Unterschiede in Bezug auf die Performance festgestellt werden konnten. Daher wurden die Daten der drei Szenarien für die drei Gruppen für eine weitere Betrachtung zusammengefasst [Dzaack 08]. Mittels der Performance-Daten konnte die zweite Hypothese bestätigt werden. Die dritte Hypothese konnte nur tendenziell bestätigt werden.

Die zusätzliche Analyse der Blickdaten hilft einen umfassenderen Eindruck der drei Benutzerschnittstellen und deren Problemstellen zu erhalten. Zusätzlich liefern sie weiteren Input für die Bestätigung oder Ablehnung der Hypothesen. Das Ergebnis der Blickdaten-Analyse bestätigt alle drei Hypothesen. Bei der ersten Hypothese hat sich gezeigt, dass diese durch die Performance-Daten

alleine noch nicht bestätigt werden kann und die Analyse von Blickbewegungsdaten für eine genauere Untersuchung notwendig ist [Dzaack 08]. Die dritte Hypothese, die durch die Performance-Daten nur tendenziell bestätigt wurde, konnte dank der Analyse der Blickbewegungsdaten gänzlich bestätigt werden: Es zeigen sich hier deutliche Ähnlichkeiten zwischen dem menschlichen Verhalten und dem mittels ACT-R simulierten Bottom-Up-Modell [Dzaack 08].

Die Ergebnisse der Studie zeigen, dass für eine Bewertung der Benutzerschnittstellen von Mensch-Maschine-Systemen durchaus auch kognitive Modellierungsmethoden geeignet sein können. Insbesondere das Bottom-Up-Modell (= der Prozess der Informationsaufnahme und -verarbeitung wird durch externe Reize der Arbeitsumgebung gesteuert [Dzaack 08]) ähnelt dem menschlichen Verhalten. Es hat sich auch gezeigt, dass das Analysieren einfacher Leistungsdaten (hier: die Performance der Prozesssteuerung) noch keine differenzierte Aussage bezüglich der Gestaltung der drei Benutzerschnittstellen sowie bezüglich der unterschiedlichen Theorien der kognitiven Modelle erlaubt. Daher wird die Integration von Blickbewegungsdaten bei der Simulation menschlichen Verhaltens empfohlen, um Problemstellen frühzeitig zu erkennen und um intuitiv bedienbare Benutzerschnittstellen zu gestalten.

CarE

Eine Alternative zur Bewertung von Benutzeroberflächen durch theoretische Modelle, bietet das Werkzeug CarE (Car-Evaluator) an. Es wurde für einen deutschen Fahrzeughersteller entwickelt, damit Ingenieure schon während der Entwicklung im automotiven Bereich Fahrzeugkomponenten wie „Bedieneinheiten und Bedienelemente in Bezug auf ihre kognitiv-ergonomische Gestaltung standardisiert bewerten können“ [Dubrowsky et al. 01]. Dieses Software-Tool ist – wie oben genannte theoretische Modelle – bereits in frühen Phasen der Entwicklung, wie beispielsweise in der Modellphase, einsetzbar. Es kann jedoch auch beim fertigen Fahrzeug zum Einsatz kommen.

Acht Bewertungsdimensionen und 31 dazugehörige Skalen, die in der Literatur verwendet werden, wurden in das System integriert. Die Anwendung erfolgt anhand eines Pen-PC, da dieser die benötigte Mobilität bei der Nutzung gewährleistet [Dubrowsky et al. 01].

Nach Auswahl der zu beurteilenden Elemente, entscheidet der untersuchende Ingenieur auf welcher Dimension (z.B. Steuerbarkeit) und mit welcher Skala (z.B. Dauer der Bedienung, Genauigkeit, etc.) er diese bewerten möchte. Das System sucht die zur Auswahl zugehörigen Kriterien aus der Datenbank und formuliert sie in einen Fragekatalog um, den der untersuchende Ingenieur mittels eines fünfstufigen Ratings beantwortet [Dubrowsky et al. 01].

CarE wurde so entwickelt, dass es mit dem schnellen Fortschreiten der Entwicklungen in der Fahrzeugindustrie mithalten kann. So können spezifische Profile erstellt werden, um die umfangreiche Kriterien-Liste einzuschränken oder neu zu kombinieren. Darüber hinaus ist es möglich, neue Bedieneinheiten und Bedienelemente in das System zu integrieren oder diese zu editieren

[Dubrowsky et al. 01]. Auf diese Weise bietet das Werkzeug dem entwickelnden Ingenieur die Möglichkeit, auch ohne umfassende Kenntnisse der Mensch-Computer Interaktion und der Physiologie des Menschen, gebrauchstaugliche Systeme zu erkennen und zu entwickeln. Der Vorteil dieser Methode ist – wie bei der Anwendung theoretischer Modelle – dass sie auch schon in frühen Entwicklungsstadien angewandt werden kann und die Ergebnisse aus diesem Grund noch in die weitere Produktentwicklung einfließen können. Darüber hinaus werden hier weder Usability-Experten noch potentielle Benutzer benötigt und auch das Bereitstellen eines Prototypen ist nicht unbedingt notwendig. Auf diese Weise können mit Hilfe dieses Systems sowohl Zeit als auch Kosten eingespart werden.

Zusammenfassung: theoretische Modelle

Der Vorteil theoretischer Modelle liegt insbesondere darin, dass sie schon frühzeitig im Entwicklungsprozess eingesetzt werden können. Auf diese Weise kann die Gebrauchstauglichkeit einer Benutzeroberfläche schon frühzeitig ermittelt werden und die Ergebnisse können gut in den weiteren Entwicklungsprozess integriert werden. Bei empirischen Untersuchungen, die erst in späteren Phasen der Entwicklung stattfinden, ist eine Einbindung der Ergebnisse nicht mehr ohne Weiteres möglich [Hamacher 02].

Darüber hinaus müssen bei theoretischen Modellen keine tatsächlichen Benutzer integriert werden, was einerseits Zeit und Geld kosten und andererseits einen Prototypen erfordern würde. Dadurch kann mittels dieser Methoden in einem kürzeren Zeitraum evaluiert und schon früh und schnell zwischen Designalternativen entschieden werden [Urbas & Leuchter 08]. Auch die Einbindung von Usability-Experten ist nicht notwendig, doch muss der Entwickler und Beurteiler sich mit der Erstellung und Anwendung theoretischer Modelle auskennen, da nur manche Systeme eine automatische Erstellung selbiger ermöglichen (z.B. TREVIS auf der Grundlage von ACT-R [Hamacher 02]) und die automatisch erstellten Modelle zumeist manuell überarbeitet werden müssen.

Ein weiterer Vorteil theoretischer Modelle ist, dass mit ihrer Hilfe Benutzerschnittstellen entwickelt werden können, die an die menschlichen kognitiven Prozesse angepasst sind und aus diesem Grund einfacher und intuitiver zu bedienen sind [Dzaack 08]. Dennoch werden sie eher in automotiven Forschungsprojekten als in der Automobilindustrie eingesetzt. Die Befragten des Fragebogens erwähnten keine theoretischen Modelle im Rahmen der von ihnen durchgeführten Evaluationen.

3.4.5.2. Studien im Fahrsimulator

Wurden mit Hilfe theoretischer Modelle diejenigen Varianten eines Systems identifiziert, welche das meiste Potential haben, können sie in Form von Prototypen umgesetzt werden. Diese können anschließend unter der Integration von potentiellen Benutzern empirisch untersucht werden.

In der Automobilindustrie werden mehrere Aspekte untersucht, um die Ablenkungswirkung und die Belastung des Fahrers während der Fahrt durch ein interaktives System zu ermitteln. Aus sicherheitstechnischen Gründen erfolgt eine solche Untersuchung jedoch zumeist nicht in einem richtigen Fahrzeug im wirklichen Straßenverkehr, sondern in einem Fahrsimulator oder einem ähnlichen Aufbau.

Es gibt verschiedene Arten von Fahrsimulatoren. Welche Art bei den Studien Verwendung findet, hängt größtenteils von den finanziellen Möglichkeiten des Forschungsteams zusammen. So finden manche Untersuchungen an einem Tisch statt, an welchem die Fahrt auf einem etwas größeren Monitor simuliert wird. Bei anderen Untersuchungen, insbesondere in der Automobilindustrie, gibt es Sitzkisten oder Fahrzeuge in welchen der Proband Platz nehmen kann und die Fahrt wird auf Leinwände projiziert.

Grundsätzlich gibt es zwei Faktoren, die den Realismus einer Testumgebung bestimmen: die physische und die funktionale Wiedergabetreue. Die physische Wiedergabetreue besagt wie sehr die Testumgebung der realen Umgebung entspricht. Je höher diese ist, umso eher nimmt der Proband seine Aufgaben während der Evaluation ernst. Im Fahrsimulator bestimmen Aspekte wie die Auflösung, in der die Straße und die Umgebung dargestellt sind oder das Vorhandensein von Bäumen, die Darstellung des Himmels oder Lichteffekte, die den Übergang zwischen Tag und Nacht simulieren, das Maß der physischen Wiedergabetreue. Auch auditives Feedback wie der Wind, Fahr-, Motoren- oder Reifengeräusche sind ein bestimmender Faktor und helfen zudem dem Probanden im Fahrsimulator die Geschwindigkeit zu ermitteln [Bach et al. 08]. Je nach Fahrsimulator kann das physische Feedback sehr hoch sein.



Abbildung 3.05: Lo-Fi Fahrsimulator mit geringer physischer und funktionaler Wiedergabetreue (Quelle: [Kern et al. 08]).

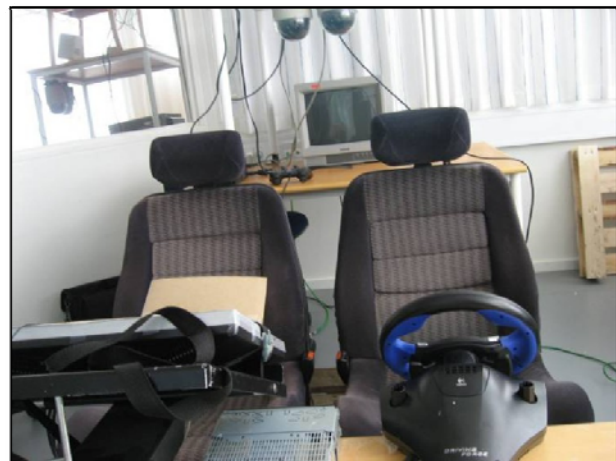


Abbildung 3.06: Medium-Fi Fahrsimulator mit mittlerer physischer Wiedergabetreue (Quelle: [Bach et al. 08])

Lo-Fi-Fahrsimulatoren, die auf einem PC in einem Usability-Labor installiert sind, vor denen der Proband auf einem Stuhl sitzt und bestenfalls ein Spielkonsolen-Lenkrad bedient (vgl. Abbildung 3.05), haben eine geringere physische Wiedergabetreue, als Simulatoren mit richtigen Autositzen und einer großen, auf eine Leinwand projizierte Sicht (vgl. Abbildung 3.06). Es gibt jedoch auch sehr fortschrittliche Fahrsimulatoren, die nicht nur wie ein Fahrzeug aufgebaut sind und daher den Nutzungskontext sehr real wiedergeben, sondern dem Probanden zudem 360°-Ansichten inklusive Rückspiegel-Ansichten anbieten (vgl. Abbildungen 3.07 und 3.08) oder teilweise auch Motorengeräusche und Erschütterungen simulieren können. Diese Simulatoren haben eine sehr hohe physische Wiedergabetreue und stellen so eine realere Testumgebung dar, die zu valideren Ergebnissen führt.



Abbildung 3.07: Hi-Fi Fahrsimulator mit 360°-Ansicht. Blick durch die Frontscheibe (Quelle: [MT 08]).



Abbildung 3.08: Hi-Fi Fahrsimulator mit 360°-Ansicht und Rückspiegel (Quelle: [MT 08]).

Die funktionale Wiedergabetreue besagt, in welchem Maße sich die Testumgebung wie die reale Umgebung verhält. Aspekte wie das örtlich-visuelle oder sensorische Feedback, das durch Bewegung beziehungsweise Beschleunigung entsteht, spielen hierbei eine Rolle. Es mag Fahrsimulatoren geben, die diese Aspekte in gewissem Maße erfüllen können, doch ist die Autorin dieser Arbeit weder in der realen Welt, noch im Rahmen ihrer Literaturrecherche auf einen solchen gestoßen.

Wie bereits erwähnt finden sehr viele Evaluationen im automotiven Bereich in einem Fahrsimulator statt. Dieser bietet dem Untersuchungsleiter einige große Vorteile, die im normalen Stadtverkehr nicht gegeben sind. Zum Einen können durch die vorbestimmte Strecke, die jeder Proband fahren muss die späteren Ergebnisse besser verglichen werden. Zumeist erfolgt die Untersuchung von beispielsweise Fahrerassistenzsystemen nach dem gleichen Muster: der Proband fährt im Fahrsimulator eine bestimmte Strecke ab. Der erste Durchgang erfolgt hierbei ohne Probleme, wohingegen beim zweiten Durchgang sehr starke Probleme simuliert werden. So wird beispielsweise in einer Kurve das

Assistenzsystem ausgeschaltet und dem Fahrer eine Übernahmeaufforderung gegeben. Nun muss dieser reagieren und übernehmen, sonst wird er aus der Kurve geworfen [Gespräche]. Der Fahrsimulator hat den Vorteil, dass solche Probleme stets zum selben Zeitpunkt simuliert werden können. Die gemessenen Reaktionszeiten die die verschiedenen Probanden mit gegebenenfalls unterschiedlichen Systemen haben, können so gemessen und gegenübergestellt werden. Auch wenn Probleme durch den weiteren Straßenverkehr oder durch sich ändernde Witterungs- oder Lichtverhältnisse auftreten, kann dies in einem Simulator besser simuliert werden, als es in der realen Welt möglich wäre. Zudem können die Probleme, die hier simuliert werden, auch wenn sie in der Realität schwerwiegende oder gar tödliche Konsequenzen hätten, nicht die Sicherheit des Probanden gefährden.

Ein weiterer Vorteil des Fahrsimulators ist, dass sich sämtliche Daten sehr genau erfassen lassen. Die Aufzeichnung der Bedienzeiten, Fehlerraten und Systemzeiten helfen dabei, die Effektivität und Effizienz des Systems zu ermitteln. Zudem kann auch seine Gebrauchstauglichkeit im Hinblick auf den Straßenverkehr untersucht werden, da lange Bedienzeiten und insbesondere hohe Fehlerraten nicht nur ein höheres Maß der Ablenkung bewirken, sondern auch zur Frustration des Fahrers führen können, die sich wiederum negativ auf sein Fahrverhalten auswirken. Das Maß der Frustration lässt sich subjektiv durch beispielsweise Interviews bestimmen. Da jedoch die Frustrationstoleranz jedes Probanden unterschiedlich hoch ist und sich der eine oder andere Proband als möglichst frustresistent darstellen möchte, können zusätzlich objektive Methoden angewandt werden, um den Erregungszustand zu bestimmen. In diesem Zusammenhang können Faktoren wie die Herzfrequenz, die Hautleitfähigkeit und die Atmung mittels physiologischer Sensoren an Hand beziehungsweise Bauch gemessen werden [Gruenstein et al. 09]. Eine solche Messung wäre im realen Straßenverkehr zu invasiv und deshalb nicht praktikabel. Während [Pataki et al. 05] im Jahre 2005 noch kritisierten, dass in der automotiven Usability-Forschung lediglich Fragebogendaten erhoben werden und auf die Erfassung von Leistungs- und Performancedaten zur Nutzenbestimmung verzichtet wird, scheint sich das in den letzten Jahren geändert zu haben. Schenkt man den Forschern aus dem automotiven Bereich Glauben, werden sehr wohl verschiedenste Daten aufgezeichnet oder geloggt. Insbesondere bei Usability Tests im Fahrsimulator kann dies durch die Installation verschiedener Aufzeichnungsmöglichkeiten sauber und verlustlos getan werden.

Während in einem Fahrzeug im Stadtverkehr nur eine bestimmte Anzahl an Kameras verwendet werden kann, ohne allzu invasiv zu wirken, sind bei Evaluationen im Simulator-Kontext nahezu beliebig viele Kameras denkbar. Oftmals werden hier deshalb nicht nur der Blick auf die Straße, sondern auch die Interaktionen des Probanden mit dem System und seine Blicke erfasst (vgl. Abbildung 3.09) [Bach et al. 08].

Die Interaktionen des Probanden spielen insofern eine Rolle, als dass die Bedienfolge ermittelt werden kann, mit Hilfe derer der Proband eine bestimmte Aufgabe erfüllen möchte [Fragebogen 09]. Diese sollte idealerweise nicht mehr Schritte enthalten, als unbedingt nötig, sodass die Hände des Probanden die meiste Zeit zum Lenken des Fahrzeugs verwendet werden können.



Abbildung 3.09: Vier Videokameras erfassen im Simulatorkontext verschiedene Aspekte der Interaktion (Quelle: [Bach et al. 08]).

Das Erfassen der Blickbewegung spielt eine große Rolle bei der Untersuchung der Ablenkungswirkung und des Einflusses der Interaktion auf die Fahrleistung [Fragebogen 09]. So werden im automotiven Kontext verschiedene Kategorien von Blicklängen unterschieden. [Bach et al. 08] schlagen die Einteilung in die drei Kategorien „kürzer als 0,5 Sekunden“, „0,5 bis 2 Sekunden“ und „mehr als 2 Sekunden“ vor. Der Grund hierfür ist, dass Fixationen ein Indikator dafür sind, auf welchem Objekt sich die Aufmerksamkeit des Fahrers momentan befindet. Typischerweise erfolgen Fixationen ab einer Dauer von 0,5 Sekunden.

Untersuchungen haben gezeigt, dass Fahrer nur ungern länger als 2 Sekunden auf die visuellen Informationen des Straßenverkehrs verzichten [Bach et al. 08] [Pataki et al. 05]. Aus diesem Grunde wurde die Grenze zwischen der zweiten und dritten Kategorie entsprechend dieser „2-Sekunden-Regel“ gewählt. Während die Analyse von Eyetracking-Experimenten in der universitären automotiven Forschung je nach Fragestellung öfter zum Einsatz kommen kann, geben Mitglieder der Automobilindustrie an, dass Videoanalysen zur Beschreibung der Blickdaten in der Praxis eher eine Seltenheit sind. Grund hierfür ist, dass eine solche Analyse sehr zeit- und kostenintensiv ist [Fragebogen 09].

Ein weiteres Indiz für die Ablenkung des Fahrers ist das Beibehalten der vorgegebenen Geschwindigkeit sowie der Verbleib auf der vorgegebenen Spur. Deshalb wird in Simulatorstudien der Proband meist gebeten, die Spur zu halten, zu wechseln oder einem voranfahrenden Fahrzeug mit konstantem Abstand zu folgen, während er eine zusätzliche Aufgabe ausführen muss. Dabei wird gemessen, wie sehr die zusätzliche Interaktion seine Fahrleistung beeinflusst.

Um die Ablenkungswirkung eines automotiven Systems zu bestimmen, können beispielsweise Fehler der lateralen oder longitudinalen Kontrolle gemessen werden. Laterale Fehler sind solche, bei denen die vorgegebene Spur verlassen wird. In manchen Untersuchungen werden diese Fehler forciert, indem der Simulator so eingestellt ist, dass ein ständiges Nachlenken notwendig ist, um auf der Spur zu bleiben [Wäller et al. 08]. Longitudinale Fehler werden gemacht, wenn die vorgegebene Geschwindigkeit nicht gehalten werden kann, weil der Fahrer zu schnell oder zu langsam fährt. Solche Fehler tauchen bei der Interaktion mit automotiven Systemen des Öfteren auf, weil Fahrer dazu tendieren die Geschwindigkeit zu verringern, wenn sie eine zusätzliche Aufgabe durchführen [Kern et al. 09].

Eine weitere Methode, die Ablenkung mit Hilfe der Spurhaltung zu messen, ist der Lane Change Task (LCT). Hierbei wird der Proband gebeten beim Auftreten eines bestimmten Verkehrsschildes im Simulator auf die jeweils angezeigte Spur zu wechseln. Dabei wird die mittlere Abweichung von einer definierten Idealstrecke gemessen. Die Fläche zwischen dieser Idealstrecke und der tatsächlichen Fahrstrecke zeigt die Fahrqualität an. Diese wird von der Wahrnehmung, Reaktion, dem Manövrierverhalten sowie der Spurhaltung des Probanden beeinflusst (vgl. Abbildungen 3.10 und 3.11) [Kuhn 05].

Der LCT eignet sich insbesondere bei der Ermittlung der Ablenkung von visuell oder kognitiv beanspruchenden Aufgaben. Das sind solche, bei denen der Proband den Blick abwenden muss, um etwas zu suchen oder bei denen er angestrengt nachdenken muss. Weniger gut lässt sich mit dem LCT die motorische Ablenkung bei der Interaktion mit einem zusätzlichen System berechnen [Kuhn 05].

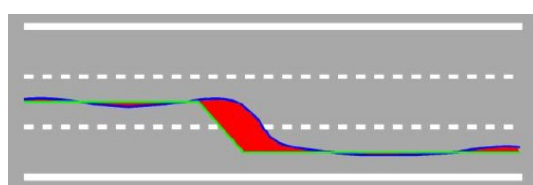


Abbildung 3.10: Gute Spurhaltung = Kleine Fläche
(Quelle: [Kuhn 05]).

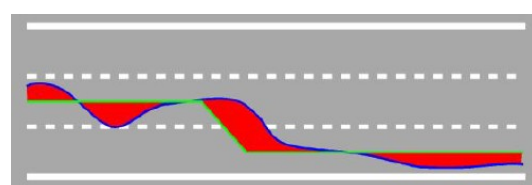


Abbildung 3.11: Schlechte Spurhaltung = große Fläche
(Quelle: [Kuhn 05]).

Die Werte des LCT oder der lateralen und longitudinalen Fehler sind Maße, die im Fahrsimulator im Gegensatz zu Untersuchungen im Straßenverkehr genauestens erfasst werden können. Während die Simulatorsoftware bereits kleinste Veränderungen der Geschwindigkeit registriert und die Position auf

der vorgegebenen Spur genau bestimmen kann, ist die Messung der Fahrfehler im Straßenverkehr von der Erkennung durch den Untersuchungsleiter abhängig. Beim Einbeziehen von Menschen kann jedoch niemals das Maß an Genauigkeit erreicht werden, welches eine Software ermöglicht.

Neben dem automatischen Loggen der Performancedaten wird auch in Fahrsimulatorstudien das subjektive Empfinden der Probanden untersucht. Dies geschieht in der Automobilindustrie zumeist mittels standardisierter Skalen und Fragebögen, die während beziehungsweise nach dem Test ausgefüllt werden. Ein Fragebogen, welcher in der Automobilindustrie zu diesem Zwecke oft eingesetzt wird, ist der NASA-TLX (vgl. Kapitel 3.4.4.) [Fragebogen 09] [Pataki et al. 05].

Der Fahrsimulator hat zwar sehr viele Vorteile für eine Untersuchung im automotiven Kontext, doch bergen sie grundsätzlich das Problem, dass die Ergebnisse ein Produkt der Fähigkeit des Simulators sind, die reale Welt wiederzugeben. Der Nachteil, dass funktionales Feedback nicht ausreichend gegeben ist, bleibt bestehen. Das Fehlen solchen Feedbacks bewirkt, dass das Fahrzeug weniger gut kontrollierbar wirkt und sich Probanden in den Studien aus diesem Grunde in einem höheren Maße unwohl fühlen, wenn sie ihren Blick und ihre Konzentration von der Straße abwenden müssen [Bach et al. 08]. Andere Simulatoren haben das Problem, dass dieses Feedback eine andere Stärke, Dauer oder Richtung hat, als in der realen Welt [Bach et al. 08] [Goodman et al. 97]. Dies kann insbesondere bei älteren und weiblichen Probanden zu Übelkeit führen [Gespräche].

Zudem agieren Probanden in Simulatorstudien anders, weil Fahrfehler keine gravierenden Konsequenzen für sie haben [Bach et al. 08] [Goodman et al. 97]. Daher kann es bei Untersuchungen, in welchem die Sekundäraufgabe nicht sicherheitsgefährdend wirkt, sinnvoll sein, die Evaluation in einem realen Fahrzeug durchzuführen.

3.4.5.3. Kontrolliertes Fahren

Die funktionale Wiedergabetreue kann im automotiven Kontext am besten unter Einbezug eines tatsächlichen Fahrzeugs während der Fahrt gewährleistet werden. Aus diesem Grunde werden manche Untersuchungen in diesem Bereich unter der Bedingung des kontrollierten Fahrens durchgeführt. Dabei fährt der Proband in einem richtigen Fahrzeug auf einer ihm vorgeschriebenen Strecke und führt dabei bestimmte Aufgaben in Bezug auf die Fahrzeugsteuerung und auf die Interaktion mit dem zu testenden automotiven System aus, während der Untersuchungsleiter als Beifahrer im Fahrzeug sitzt. Die Strecken, die im Rahmen des kontrollierten Fahrens genutzt werden, müssen aus sicherheitstechnischen Gründen wenig befahren sein. Gebräuchlich für solche Experimente sind daher Verkehrsübungsplätze (vgl. [Bach et al. 08]), wenig befahrene Strecken in abgelegenen Gebieten (vgl. [Lange et al. 06]) oder auch Waldstücke und manchmal Wohngebiete in den Abendstunden (vgl. [Pataki et al. 05]).

Beim kontrollierten Fahren kann der Fahrer beispielsweise gebeten werden, in einem bestimmten Abschnitt mit einer bestimmten Geschwindigkeit zu fahren. Ampeln und Verkehrsschilder können dabei helfen, den tatsächlichen Fahrkontext realitätsgetreuer nachzubilden. Slalomfahrten um Pylone können besonders kurvige Strecken oder ähnliche Probleme simulieren (vgl. Abbildung 3.12). Dennoch empfinden Fahrer das kontrollierte Fahren anders als das reale Fahren. Grund hierfür ist, dass die Umgebung und die Teststrecke vorhersagbar sind und dadurch als unnatürlich empfunden werden. Das Fehlen weiterer Agenten, wie zusätzlicher Verkehrsteilnehmer bestärkt dieses Empfinden. Zudem wird bei Evaluationen im automotiven Bereich zumeist ein präpariertes Testfahrzeug verwendet und nicht das Fahrzeug des Probanden, was den künstlichen Effekt verstärkt. Etwas mehr Realismus könnte durch Unvorhersehbarkeit wie das Hinzufügen autonomer Agenten oder veränderlicher Beschilderung erfolgen. Auch der Wegfall einer vorgeschriebenen Strecke könnte realistischer wirken. Nichtsdestotrotz ist das kontrollierte Fahren nicht mit dem natürlichen Fahren vergleichbar, da hierbei – wie auch bei Untersuchungen in der klassischen Mensch-Computer Interaktion – vom Untersuchungsleiter vorgeschriebene Aufgaben erledigt werden müssen und der Proband mit dem Fahrzeug nicht zu einem selbst bestimmten Ziel fährt, an welchem er zu einem bestimmten Zeitpunkt ankommen muss. Dennoch sind solche Untersuchungen näher an der Realität und produzieren daher validere Ergebnisse als Simulatorstudien.



Abbildung 3.12: Kontrolliertes Fahren auf einer Übungsstrecke (Quelle: [Bach et al. 08]).

[Bach et al. 08] haben eine Untersuchung durchgeführt, um die Aufmerksamkeit und das Fahrverhalten zwischen Fahrsimulator und kontrolliertem Fahren zu vergleichen. Sie haben herausgefunden, dass die Unterschiede nicht immer groß sind und teilweise eher im Detail liegen, dass es aber dennoch Aspekte gibt, die mit einer realistischeren Testumgebung deutlicher werden.

In ihrer Untersuchung haben sie drei verschiedene Interaktionskonzepte für Musikplayer sowohl in einer Simulatorumgebung, als auch bei der kontrollierten Fahrt auf einem Verkehrsübungsplatz miteinander verglichen. Dabei haben sie unter anderem festgestellt, dass die Bedienzeiten beim kontrollierten Fahren um 9% höher waren, auch wenn dieser Effekt sich als nicht signifikant herausstellte. Dieser Unterschied könnte darauf beruhen, dass die Probanden in der realeren Testumgebung des Öfteren beide Hände benötigten, um das Fahrzeug zu steuern [Bach et al. 08]. Hier waren auch die Unterschiede zwischen den Musikplayern größer. Zudem fiel auf, dass in der Simulatorumgebung der Blick des Probanden öfter auf den Musikplayer gerichtet war. Dieser Effekt wurde in einem T-Test zwar ebenfalls als nicht signifikant eingestuft, doch stach hier einer der Musikplayer besonders hervor: entgegen dem Trend bei den anderen beiden Geräten zog der dritte Musikplayer auf der Teststrecke mehr als doppelt so viele Blicke auf sich als im Simulator, obwohl die Anzahl der Interaktionsfehler in der Simulatorumgebung signifikant höher war.

Bei der Betrachtung der Blickverteilung fällt auf, dass sich die beiden Testumgebungen in der Anzahl der ersten und zweiten Blickkategorie kaum unterscheiden (vgl. Kapitel 3.4.5.2.). Lediglich bei den Blicken über der „2-Sekunden-Regel“ ist ein signifikanter Unterschied dahingehen zu erkennen, dass die Probanden deutlich mehr solcher Blick in der kontrollierten Umgebung tätigten. Der Grund dafür ist, dass sich die Probanden in dieser Umgebung sicherer fühlen als im Fahr Simulator [Bach et al. 08]. Durch das im Simulator fehlende örtlich-visuelle und sensorische Feedback wirkt die Situation künstlich und das Fahrzeug wirkt dadurch weniger kontrollierbar, sodass die Probanden sich hier stärker auf das Fahren selbst als auf die Sekundäraufgabe konzentrieren müssen. Dennoch werden im Kontext des Simulators mehr Fehler gemacht. Während die Unterschiede bezüglich der lateralen Fehler wie beispielsweise bei der Spurhaltung nicht signifikant sind, sind sie es sehr wohl in Bezug auf die longitudinalen Fehler wie dem Halten der vorgegebenen Geschwindigkeit. Dies liegt, wie bereits in Kapitel 3.4.5.2. erwähnt, daran dass die funktionale Wiedergabetreue in der Simulatorumgebung dürftig ist und der Proband auf diese Weise die Geschwindigkeit des Fahrzeugs nicht so gut abschätzen kann, wie bei einer richtigen Fahrt.

Bei ihren Untersuchungen ist [Bach et al. 08] aufgefallen, dass die kontrollierte Teststrecke den Probanden all jene Informationen bot, die in der Simulatorumgebung vermisst wurden. Die Teststrecke basierte jedoch nicht auf diesen Kommentaren, da die Tests im Simulator erst hinterher durchgeführt wurden. Dennoch empfanden die Probanden auch das kontrollierte Fahren auf einer Teststrecke nicht als ausreichend realistisch.

Bezüglich der Ergebnisse beim Vergleich der drei Musikplayer fällt auf, dass der Trend zumeist in die selbe Richtung geht und die teilweise vorhandenen, geringfügigen Unterschiede nicht signifikant sind [Bach et al. 08]. Dennoch wird das kontrollierte Fahren als realer empfunden und die Probanden empfinden ein größerer Maß an Kontrolle über das Fahrzeug, sodass die hier zustanden kommenden Ergebnisse vermutlich eher den Ergebnissen einer realen Fahrt entsprechen.

Es muss jedoch beim Planen einer Evaluation bedacht werden, dass das kontrollierte Fahren sowohl in der Durchführung, als auch in der Auswertung mehr Zeit und mehr Geld kostet, als eine Studie im Fahrsimulator [Bach et al. 08] [Lange et al. 06]. Während für einen Simulator lediglich ein Usability Labor sowie einige Autoteile und eine Simulatorsoftware benötigt werden, muss beim kontrollierten Fahren eine Teststrecke angemietet werden und ein richtiges Fahrzeug muss betankt und mit der benötigten Software zur Aufzeichnung diverser Daten ausgestattet werden. Die Vorbereitung und Durchführung der Tests auf der Teststrecke kosteten [Bach et al. 08] doppelt so viel Zeit, wie die Vorbereitung und Durchführung der Simulatortests. Bei der Auswertung der Daten lässt sich ein ähnlicher Trend erkennen: während in einem Simulator auch Daten wie die Geschwindigkeit und Spurhaltung genauestens erfasst werden können, bedarf es bei einem unpräparierten Fahrzeug des Untersuchungsleiters oder der Videoaufzeichnungen, um die lateralen und longitudinalen Fehler zu ermitteln. Zudem waren die Video- und Logdaten in der kontrollierten Bedingung minderwertiger als beim Fahrsimulator, sodass deren Auswertung anspruchsvoller war.

Welche Variante hier gewählt wird, muss demnach entsprechend des zeitlichen und finanziellen Budgets getroffen werden, welches dem Forschungsteam für die Evaluation zur Verfügung steht.

3.4.5.4. Fahren im Straßenverkehr

Eine weitere Testumgebung, in welcher Evaluationen im automotiven Kontext durchgeführt werden können, ist der reale Straßenverkehr. Dieser wirft jedoch einige gewichtige Fragen auf, die vor Durchführung eines Tests geklärt werden müssen und die zudem der Grund dafür sind, dass diese Art der Testumgebung nur in sehr seltenen Fällen gewählt wird. Die erste Frage betrifft ethische Aspekte, die insbesondere bei der Untersuchung der Ablenkungswirkung interaktiver Systeme an großer Bedeutung gewinnen. Der fünfte Punkt des DECIDE-Frameworks besagt, dass das Wohlbefinden der Probanden an oberster Stelle stehen muss [Sharp et al. 07] und dass eine Person, die bei einer Untersuchung teilnimmt, diese nicht in einem verschlechterten mentalen oder physischen Zustand verlassen darf [Noyes & Baber 99]. Ihr Zustand sollte demnach auch nicht durch bewusst zugeführte Ablenkung im realen Fahrkontext gefährdet werden.

Der zweite Aspekt betrifft einen der Vorteile eines Fahrsimulators, der in der realen Welt nicht gewährleistet werden kann: die Reproduzierbarkeit der Studie. Da der Straßenverkehr, Ampelschaltungen, Passanten und Witterungsverhältnisse nicht kontrolliert werden können, ist es unmöglich einen zweiten Test unter den identischen Bedingungen durchzuführen. Der Vergleich zweier Testdurchläufe wird also sehr schwierig.

Der dritte Aspekt betrifft schließlich die allgemeinen Einschränkungen von Experimenten, wie sie auch bei Untersuchungen in der Mensch-Computer Interaktion vorhanden und bekannt sind: Zum einen repräsentieren die ausgewählten Probanden nicht die Verschiedenheit der tatsächlichen Bevölkerung. Zum anderen sind Experimente stets auf bestimmte Arten kontrolliert und der Proband wird während seiner Interaktion überwacht, was ihm meistens bewusst ist und was oftmals dazu

führen kann, dass er sich anders verhält, als er es in einer alltäglichen Situation tun würde. Ein weiterer Faktor, welcher bei automotiven Experimenten auftritt, ist die Tatsache dass der Proband zumeist nicht sein eigenes Fahrzeug fährt, sondern ein unbekanntes Testfahrzeug, welches gegebenenfalls anders ausgestattet ist als sein eigenes und in welchem meistens zusätzlich ein Untersuchungsleiter drin sitzt. Diese Aspekte bewirken, dass die Evaluation im richtigen Straßenverkehr noch keine Garantie für hundertprozentig verlässliche Ergebnisse ist, sodass bei sicherheitsgefährdenden Forschungsfragen alternative Testumgebungen vorzuziehen sind.

Dennoch kann diese Testumgebung bei manchen Forschungsfragen problemlos eingesetzt werden. In der Automobilindustrie werden beispielsweise Langzeitstudien oftmals im realen Fahrzeug durchgeführt. Wie in Kapitel 3.4.2. beschrieben, wird diese Testumgebung oftmals im Rahmen von Kundenkliniken eingesetzt, in welchen der Proband über einen längeren Zeitraum nacheinander mehrere Fahrzeuge zur Verfügung gestellt bekommt. Diese sind zumeist mit verschiedenen Varianten eines interaktiven Systems ausgestattet und der Proband wird darum gebeten, das Fahrzeug für seine alltäglichen Tätigkeiten zu nutzen. In diesem Kontext kann selbstverständlich kein Untersuchungsleiter manuelle Mitschriften über die Auffälligkeiten machen und darum werden manchmal die Nutzungsdaten geloggt. Immer werden jedoch nach dem Testdurchlauf subjektive Daten in Form einer Befragung, eines Fragebogens oder mit Hilfe der Tagebuchmethode erhoben, um ein Ranking der getesteten Konzepte zu erhalten [Gespräche].

Auch bei Forschungsfragen, in welchen nicht die Ablenkungswirkung interaktiver Systeme während der Fahrt untersucht wird, können Evaluationen im Straßenverkehr durchgeführt werden. Ein Beispiel in welchem die Ablenkung im Fahrzeug nur minimal ist und daher eine Evaluation in Straßenverkehr bedenkenlos durchgeführt werden konnte, stellt das System InCA (=„In-Car Communication Agent“) dar [Kern et al. 07]. Dieses System zeigt, ähnlich wie es bei der Verfügbarkeitsanzeige von Instant Messaging Systeme der Fall ist, die Situation des Autofahrers an. Es beruht auf der Tatsache, dass nicht nur das Halten eines Mobiltelefons eine Gefahr für den Autofahrer und die anderen Verkehrsteilnehmer darstellt, sondern dass auch die mentale Belastung beim Führen eines Telefonats im Fahrzeug hoch ist und die Sicherheit gefährden kann. Wird das Mobiltelefon, wie es das deutsche Gesetz seit Anfang 2001 vorsieht [StVO 09] in der Freisprechanlage platziert, zeigt InCA zunächst generell die Situation als „im Auto“ an. Abhängig von Geschwindigkeit, Verkehrssituation oder auch manuell einstellbar wird die Situation nun in drei weitere Stufen aufgeteilt: grün, gelb und rot. Während in der grünen Stufe Anrufe problemlos entgegengenommen werden können, indiziert die gelbe Stufe, dass Anrufe ungünstig sind und daher nur getätigt werden sollen, wenn es nötig ist. Die rote Stufe hingegen zeigt an, dass Anrufe nicht erwünscht sind und nur getätigt werden sollen, wenn es zwingend erforderlich ist. Die Untersuchung der Fahrerseite wurde im realen Straßenverkehr durchgeführt, wobei die Probanden einmal mit und einmal ohne InCA 15km auf einer Teststrecke

bestehend aus Landstraßen, Autobahn und Stadtverkehr fahren mussten. Dabei saß der Untersuchungsleiter im Fahrzeug und initiierte die Anrufe. In der InCA-Bedingung wurden nur dann Anrufe getätigt, wenn der Status grün war. Ohne InCA wurden auch bei gelbem Status Anrufe getätigt. Es stellte sich heraus, dass die Fahrer den Statuswechsel nicht nachvollziehen könnten und daher eine manuelle Auswahl des aktuellen Status eventuell hilfreicher wäre. Doch es stellte sich auch heraus, dass die Anrufe die in der Ohne-InCA-Bedingung getätigt wurden, als signifikant störender empfunden wurden [Kern et al. 07].

Diese Untersuchung durfte bedenkenlos im realen Straßenverkehr durchgeführt werden, da durch Vorhandensein einer Freisprechanlage die Vorgaben der StVO befolgt wurden. Außerdem wurden Anrufe nur im Rahmen des grünen oder gelben Status initiiert, welche eine geringere Gefährdung der Probanden bewirken, als Anrufe bei roten Status es gewesen wären. Da es sich bei dem Projekt InCA um ein Kooperationsprojekt des Fraunhofer-Instituts und der BMW Group handelte, stammten sämtliche Probanden aus der technischen Forschungs- und Entwicklungsabteilung der BMW Group stammen. Es wird angenommen, dass die Teilnahme auf freiwilliger Basis stattfand, sodass sich die Probanden einer potentiellen Gefährdung durchaus bewusst und damit einverstanden waren. Zudem erinnert dieses Experiment an die Erkenntnisse aus Gesprächen mit Mitarbeitern der Automobilindustrie, wonach konzerninterne Probanden, wie unmittelbare Arbeitskollegen auch schon in frühen Phasen der Produktentwicklung in richtigen Fahrzeugen testen.

Insbesondere können jedoch solche Evaluationen im richtigen Straßenverkehr durchgeführt werden, in welchen Systeme getestet werden, die nicht vom Fahrer bedient werden. Bei solchen Untersuchungen macht es verständlicherweise wenig Sinn, den Fahrsimulator als Testsetting zu wählen und dafür auf funktionales Feedback zu verzichten. Mit dem Probanden als Mitfahrer kann das System getestet werden, während das Fahrzeug von einer dritten Person gesteuert wird, die keine Ablenkung erfährt. Die Gefährdung der Probanden bei solchen Untersuchungen ist zwar etwas höher als bei stationären Tests oder bei Tests im Fahrsimulator, doch führen solche Untersuchungen teilweise zu anderen und realistischeren Ergebnissen. Selbstverständlich muss der Proband – oder die Erziehungsberechtigten, falls es sich um minderjährige Probanden handelt (vgl. Kapitel 4.2.3.) – mit der Durchführung eines solchen Tests einverstanden sein.

3.5. Rücklauf in der Automobilindustrie

Bei universitären Forschungsprojekten werden interaktive Systeme, sowohl beim automotiven Einsatz, als auch beim Einsatz an einem Desktop-PC, zumeist im Rahmen eines iterativen Prozesses entwickelt. Dabei werden im Idealfall mehrere Evaluationen durchgeführt, um auch schon bei den ersten Zwischenergebnissen zu überprüfen, ob mit der Entwicklung der richtige Weg eingeschlagen wurde. Die bei solchen Evaluationen auftretenden Bedienungsprobleme werden, wie in

Abbildung 3.04 bereits angedeutet wurde, im nächsten Entwicklungsschritt beseitigt und es werden so viele Zyklen durchlaufen, bis das entwickelte System den erforderlichen Ansprüchen genügt.

In der Automobilindustrie ist Effizienz ein sehr wichtiges Kriterium, wie beispielsweise bereits in Kapitel 3.1. im Rahmen der verwendeten Methoden zur Datenerfassung festgestellt werden konnte. Daher stellt sich nun die Frage, inwiefern auch hier ein iterativer und dadurch möglicherweise längerer Prozess bei der Produktentwicklung in Kauf genommen wird.

Bezüglich des Rückflusses in der Automobilindustrie gibt es sehr unterschiedliche Ansichten. Die meisten Befragten sind der Ansicht, dass die qualitativen und quantitativen Ergebnisse gleichermaßen wichtig sind und deshalb in gleichem Maße in die weitere Produktentwicklung einfließen. Der Grund hierfür ist, dass quantitative Daten insbesondere bei der Entwicklung von Systemen für den Fahrer ein Maß für die Sicherheit des Fahrzeugs darstellen. Doch auch wenn diese sehr positiv sind, heißt das nicht, dass das System auch als verständlich und leicht zu bedienen empfunden wird und daher subjektiv gut bewertet wird. Im Gegensatz dazu kann sich der Proband bei der Bedienung eines Systems zwar sicher fühlen, doch kann die objektive, quantitative Datenbasis andere Ergebnisse aufweisen [Fragebogen 09].

Den Antworten auf den Fragebogen kann man jedoch auch entnehmen, dass in verschiedenen Stadien der Produktentwicklung unterschiedliche Daten einen höheren Einfluss auf die weitere Entwicklung haben. So sind qualitative Daten in den frühen Stadien sehr wichtig, da sie bestimmen, in welche Richtung die Entwicklung fortschreiten muss, damit ein ansprechendes und subjektiv zufrieden stellendes System resultiert. In späteren Phasen der Produktentwicklung, wenn das Konzept größtenteils feststeht und eher Feintuning betrieben werden muss, werden eher quantitative Daten erhoben. Für die Forscher sind beide Datenbasen gleichwertig.

Einige der Befragten geben an, dass die Ergebnisse der entwicklungsbegleitenden Untersuchungen zu 100% das weitere Vorgehen bei der Produktentwicklung bestimmen. Je nach Art führen sie hier dabei zu Optimierungen des bestehenden Systems oder zu einer gänzlichen Neuausrichtung. Dieser Aspekt ist insofern interessant, als dass die Erfahrungen, die die Autorin im Rahmen des Kooperationsprojektes gemacht hat, sich hiervon etwas unterscheiden. Bei dem Lifecycle, der vom Entwicklungsteam dieses Projektes vollzogen wurde und der in Kapitel 4 beschrieben ist, spielten die durchgeführten Evaluationen und die dort gefundenen Ergebnisse eine große Rolle für den nächsten Iterationsschritt. Gegenüber der Autorin dieser Arbeit wurde erwähnt, dass ein solcher Lifecycle, der die Ergebnisse in so starkem Maße bei der weiteren Entwicklung mit einbezieht in der Automobilindustrie eher eine Seltenheit ist.

Wohl aus diesem Grunde geben die meisten Befragten wohl an, dass die durchgeführten Studien eher Handlungsempfehlungen geben und die gewonnenen Ergebnisse lediglich als Argumentationsbasis für Diskussionen mit Entscheidungsträgern dienen. Der Entscheidungsprozess selbst mündet jedoch in den meisten Fällen in einer Vorstandsentscheidung [Fragebogen 09] [Gespräche].

Bei der Frage, welche Daten für die Entscheidungsträger wichtiger sind, scheiden sich erneut die Geister. Während die einen angeben, dass qualitative Daten wichtiger sind und eher zum Weiterverfolgen eines Konzepts führen, sind die anderen davon überzeugt, dass die Entscheidungen anhand von quantitativen Daten erfolgen [Fragebogen 09]. Während qualitative Daten der subjektiven Zufriedenheit entsprechen und daher eine wichtige Rolle bei der Kaufentscheidung spielen, bestimmen quantitative Daten die Sicherheit und Gebrauchstauglichkeit der Bedienung. Letztlich ist sicherlich die Kombination beider Datenbasen entscheidend.

4. Fallstudien

Die vorangegangenen Kapitel haben in der Theorie deutlich gezeigt, dass die automotive Produktentwicklung einige Herausforderungen aufwirft, die sich von den Herausforderungen bei der klassischen Softwareentwicklung unterscheiden. Zum einen ist der Benutzerkreis sehr speziell, zum anderen erfordert der Nutzungskontext besondere Maßnahmen bei der Entwicklung und Evaluation und auch die Aspekte der Geheimhaltung dürfen nicht vergessen werden. Diese theoretischen Herausforderungen finden sich jedoch auch in der Praxis wieder, wie die nachfolgenden Fallbeispiele zeigen sollen.

Die Fallstudien entstammen einer im Winter 2007/2008 beginnenden Kooperation zwischen der Volkswagen AG und dem Lehrstuhl für Mensch-Computer Interaktion an der Universität Konstanz. Zunächst wurde ein Pilotprojekt im Rahmen der Vorlesung „Visuelles Requirement Engineering für das Interaction Design“ durchgeführt [HCI KN 07]. Dieses hatte zum Ziel, das Potential von Webbrowsern im automotiven Bereich im Allgemeinen zu erfassen. Aus diesem Grunde waren die drei Projektteams in ihrer Konzeptentwicklung sehr frei: Abgesehen von der Anwendungsumgebung Audi A8 sowie der Displaygröße waren keine Einschränkungen vorgegeben, sodass die Projektteams ihre Anforderungen anhand interner Brainstorming-Methoden und Diskussionen entwickelten (vgl. [Söter 08]). Diese freie Entwicklung führte zu drei sehr unterschiedlichen Webbrowsern, deren unterliegende Konzepte Potential für eine weitere Entwicklung darstellten.

Daher wurde die Kooperation im Frühjahr 2008 zu Bachelor- und Masterprojekten erweitert und es wurden neue Rahmenbedingungen beschlossen. Diese führten zu einer erneuten Erhebung von Anforderungen – dieses Mal unter Einbezug potentieller Benutzer – und schließlich zu sehr andersartigen Darstellungs- und Interaktionskonzepten. Diese Konzepte wurden anschließend im Januar 2009 in einem Audi A8 evaluiert, der dem Projektteam von der Volkswagen AG zu diesem Zwecke zur Verfügung gestellt wurde. Die Ergebnisse dieser ersten Evaluationsstudie führten zu weiteren Anforderungen, die in einem weiteren Iterationsschritt zum finalen Darstellungs- und Interaktionskonzept führten. Dieses wurde im Juli / August 2009 in einer zweiten Evaluationsstudie überprüft (vgl. Kapitel 4.2.). Statt erneut anhand von Methoden der klassischen Mensch-Computer Interaktion zu evaluieren, wurde der Fokus bei dieser Evaluation auf die Methoden der Automobilindustrie gelegt. Ziel der Studie war es, selbige im Kontext einer universitären Evaluation anzuwenden.

Die nachfolgenden Kapitel stellen einen exemplarischen Lifecycle im automotiven Kontext dar, wobei der Schwerpunkt auf zwei Evaluationsstudien mit jeweils unterschiedlichem Fokussen liegt. Während Kapitel 4.1. beschreibt, wie verschiedene Interaktionskonzepte im Rahmen einer Evaluation verglichen wurden, zeigt Kapitel 4.2. wie eine summative Evaluation eines finalen Prototypen durchgeführt werden kann. Das Praxisbeispiel „automotiver Webbrowser“ soll dabei aufzeigen,

welche Methoden im Rahmen einer Evaluation im Nutzungskontext „Fahrzeug“ angewandt werden können und welche Herausforderungen sich dabei stellen. Zudem wird in den nachfolgenden Kapiteln dargelegt, wie die Zwischenergebnisse in weiteren Iterationsschritten eingearbeitet werden können, um ein gebrauchstaugliches Produkt zu entwickeln, welches nicht nur die gestellten Anforderungen erfüllt sondern dessen Benutzung auch Spaß macht.

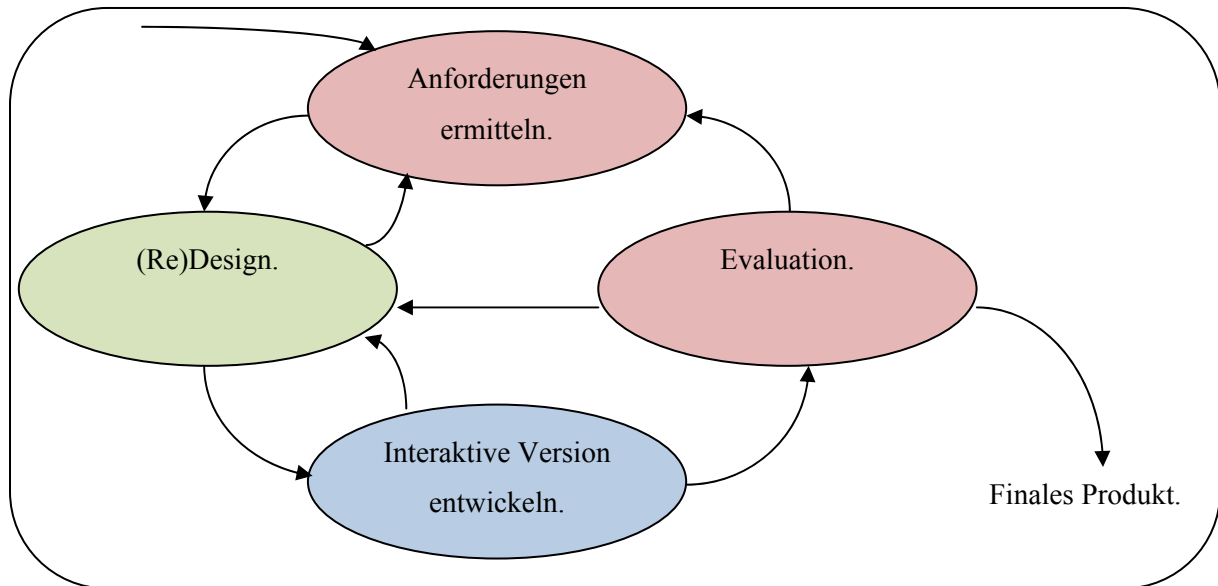


Abbildung 4.01: Einfacher Lifecycle des Interaktionsdesign nach [Sharp et al. 07]

Die Autorin dieser Arbeit war für die Anforderungsanalyse sowie für die Planung und Durchführung beider Evaluationsstudien verantwortlich (vgl. rote Bereiche in Abbildung 4.01). Zudem war sie an der Konzeption der Darstellungs- und Interaktionskonzepte beteiligt (vgl. grüner Bereich in Abbildung 4.01). Die technische Implementierung (vgl. blauer Bereich in Abbildung 4.01) wurde von Marcus Specht geleistet [Specht 10].

4.1. Fallstudie A: Benutzerbefragungen und Benutzertest im automotiven Kontext anhand von Methoden aus der Mensch-Computer Interaktion.

Im Frühjahr 2008 wurden für das Projekt „automotiver Webbrowser“ neue Rahmenbedingungen seitens der Volkswagen AG gestellt, die eine detailliertere Auseinandersetzung mit potentiellen Benutzern und ihren Aufgaben erforderte. Es wurden traditionelle Methoden der Mensch-Computer Interaktion eingesetzt, um die Anforderungen und Bedürfnisse der neuen Zielgruppe zu ermitteln. Kapitel 4.1.1. fasst diesen Prozess kurz zusammen. Eine detailliertere Beschreibung ist in [Söter 08] zu finden.

Nach Implementierung der Browserkonzepte wurde eine Evaluationsstudie geplant und durchgeführt, welche ebenfalls auf den in Kapitel 2 beschriebenen Methoden beruht. Die Herangehensweise bei dieser Studie sowie ihre Ergebnisse, die zum finalen Prototypen führten, sind in Kapitel 4.1.2.

dokumentiert. Die Studie barg gewisse Herausforderungen, die zum Teil allgemeiner Natur waren und zum anderen Teil auf dem Anwendungskontext beruhten. In Kapitel 4.1.3. werden diese Herausforderungen und wie ihnen begegnet wurde beschrieben.

4.1.1. Anforderungsanalyse

Statt weiterhin den Fokus auf Geschäftsleute zu legen, wie es bei der Entwicklung für den Audi A8 der Fall war, wurde die Zielgruppe im Frühjahr 2008 erweitert. Da der Webbrowser für die Benutzung in einem Audi Q7 entwickelt werden sollte, mussten die Eigenschaften und Bedürfnisse einer größeren Altersgruppe in Betracht gezogen werden. Um die neue Zielgruppe besser eingrenzen zu können, wurden anhand des benutzungsorientierten Ansatzes nach [Constantine & Lockwood 99] (vgl. Abbildung 4.02) Personas erstellt. Diese fiktiven Personen wurden mit verschiedenen Eigenschaften, Kenntnissen und Wünschen ausgestattet, welche in einem nächsten Schritt zu potentiellen Aufgaben führten. Die Aufgaben wurden in Form von Essential Use Cases gesammelt und zusammengeführt.

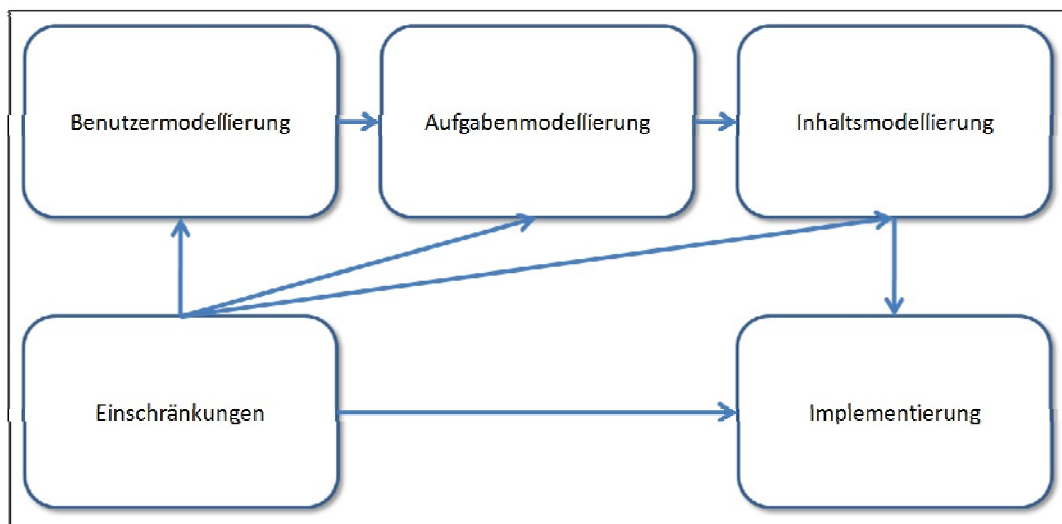


Abbildung 4.02: Modell des Usage-Centered Design nach Constantine und Lockwood [Constantine & Lockwood 99].

Anhand der erstellten Personas wurden drei Berufstätige und sechs Studenten, die diesen Personas in einigen Charakteristiken ähneln, eingeladen um an Fokusgruppen teilzunehmen. Ziel der Fokusgruppen war es, Anforderungen zu ermitteln, die bei der Erstellung der Personas möglicherweise noch nicht aufgefallen waren.

Sechs männliche und drei weibliche Personen im Alter von 22 bis 50 Jahren mit mittlerer bis sehr guter Erfahrung mit technischen Geräten und mittlerem bis sehr großem Interesse an neuen Technologien diskutierten in drei Gruppen zu zwei bis vier Leuten über die Internetnutzung im Fahrzeug. Bei den Befragten handelte es sich um Studenten und Mitarbeiter der Universität Konstanz

sowie um Absolventen mit einer Berufserfahrung von mindestens sechs Monaten. Ein Teil der ausgewählten Personen genoss eine sprachliche, der andere eine naturwissenschaftliche und der dritte Teil eine volkswirtschaftliche Ausbildung. Die erste Fokusgruppe mit zwei Personen war ein Testdurchlauf mit einer Expertengruppe, der jedoch ebenfalls ausgewertet wurde. Diese Expertenrunde sollte nicht nur ihre eigenen Bedürfnisse in die Anforderungsanalyse mit einbringen, sondern auf eventuelle weitere grundlegende Anforderungen hinweisen, die bis dahin womöglich noch nicht berücksichtigt worden sind. Durch die große Altersspanne und Breite an beruflichen und edukativen Hintergründen der Teilnehmer sollte sichergestellt werden, dass eine möglichst heterogene Menge an Bedürfnissen zur Sprache gebracht wird. Dadurch sollte es möglich sein, ein System zu entwickeln, welches Benutzer generationenübergreifend anspricht. Ein gewisses Interesse für neue Technologien sollte gewährleisten, dass sich alle Teilnehmer rege an den Diskussionen beteiligen und trotz der geringen Gruppengröße ausreichend wertvoller Input gewonnen wird.

Nach einer kurzen Einführung zu den Möglichkeiten des mobilen Internets wurden die Personen nach ihrem eigenen Nutzungsverhalten beim Surfen befragt und sollten erzählen, welches für sie persönlich die wichtigsten Funktionen dabei sind. Anschließend sollten sie diese in der Gruppe diskutieren und eine Priorisierung vornehmen. Es stellte sich hierbei beispielsweise heraus, dass für den Großteil der Befragten der „Zurück“-Knopf des Browsers wichtiger ist, als der „Vor“-Knopf (vgl. [Söter 08]). Diese Priorisierung war für das spätere Browserkonzept sehr wichtig, da die am häufigsten gewünschten Funktionen so platziert werden sollten, dass sie mit einem Handgriff zugänglich sind. Optionale Funktionen hingegen können auch an weniger prominenten Stellen platziert werden.

Nach der Diskussion über Browserfunktionen wurden den Teilnehmer der Fokusgruppen die zuvor entwickelten Personas vorgestellt und sie sollten versuchen, sich in diese fiktiven Personen hineinzusetzen. Da sich einige der Fokusgruppenteilnehmer mit einigen Personas recht gut identifizieren konnten oder Familienmitglieder oder Freunde in den Charakteristiken der Personas wiedererkannten, konnten sie weitere potentiell wichtige Aufgaben nennen, die mit einem automotiven Webbrowser erfüllbar sein müssten. Die Kenntnis der für die Teilnehmer wichtigen Aufgaben war für die spätere Entwicklung des Browsers essentiell: es konnte in Erfahrung gebracht werden, dass Informieren und Kommunizieren die Hauptaufgaben bei der Autofahrt darstellen würden und diese daher im Besonderen unterstützt werden müssten. Damit der spätere Benutzer sich informieren kann, muss beispielsweise das Lesen längerer Texte so unterstützt werden, dass es auch bei der Fahrt auf unebenen Straßen möglichst angenehm bleibt.

Mit Hilfe des Inputs aus den Fokusgruppen konnten die Personas und die Liste der Anforderungen, die zunächst auf Brainstorming des Projektteams basierten, um die Eigenschaften, Ideen und Wünsche der Fokusgruppenteilnehmer erweitert werden. Darüber hinaus konnte festgestellt werden, welche Funktionen eines handelsüblichen Browsers auch im automotiven Einsatz essentiell sind und welche Funktionen optional oder gar unerwünscht sind.

Durch Kenntnis der für die Teilnehmer relevanten Aufgaben und Funktionen konnten entsprechend dem Leitfaden Usability der Deutschen Akkreditierungsstelle Technik (=DATech) Kontextszenarien erstellt werden, welche die Aufgaben der jeweiligen Nutzer in ihrem Kontext beschreiben [DATech 09]. Diese wurden tabellarisch erfasst und über mehrere Schritte in Nutzungsanforderungen überführt, welche schließlich in finalen Systemspezifikationen resultierten (vgl. [Söter 08]).

Die Anforderungen seitens der Volkswagen AG wurden weiterhin beachtet. So stand unter Anderem fest, dass das Browserkonzept mit dem MMI-Terminal bedient werden sollte, welches auch das bisherige Eingabegerät für das Infotainmentsystem darstellte (vgl. Abbildung 4.03). Darüber hinaus sollten im Browser vier Softkeys in den Ecken die jeweils verfügbaren Funktionen anzeigen, so wie es auch in den anderen Anwendungen des MMI-Systems der Fall ist (vgl. Abbildung 4.04, 4.05 und 4.06).



Abbildung 4.03: Das MMI-Terminal als Eingabegerät für das Audi MMI-System (Quelle: [Technofile 08]).



Abbildung 4.04: Bildschirmaufbau des Audi MMI-Systems (Quelle: [Audi 08]).



Abbildung 4.05: Low-Fidelity Prototyp des grundlegenden Browserkonzeptes.

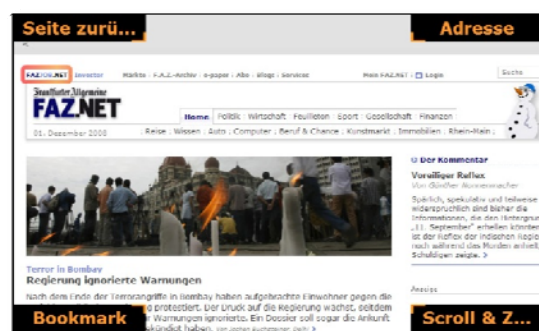


Abbildung 4.06: High-Fidelity Prototyp von Browservariante B.

Die Ergebnisse der Fokusgruppen ergaben im Zusammenspiel der Anforderungen seitens der Volkswagen AG eine lange Liste an Nutzungsanforderungen, Einschränkungen, Realisierungsmöglichkeiten sowie Ideen für die finale Systemspezifikation, die anschließend in ein

Interaktionskonzept überführt werden sollten. Aufgrund der sehr individuellen Charakteristiken, Vorlieben und Anforderungen der Teilnehmer wurden jedoch drei verschiedene Interaktionskonzepte entwickelt und prototypisch umgesetzt.

Das erste Interaktionskonzept, Variante A, ermöglichte dem Benutzer einen einfachen Zugriff auf alle verfügbaren Funktionen. So konnte der Benutzer beispielsweise mit einem einzelnen Handgriff Scrollen und Zoomen, musste jedoch präzise arbeiten, um Links zu treffen. Scrollen wurde über das Bewegen des Mauszeigers auf dem MMI-Terminal an die Bildschirmkanten realisiert. Zoomen konnte der Benutzer, indem er das Drehrad auf dem MMI-Terminal bewegte. Das zweite Konzept, Variante B, erlaubte dem Benutzer ein genaues Treffen von Links, trotz unpräziserer Arbeitsweise, da das Drehen des Drehrades hier zu einer zeilenweisen Traversierung aller Links einer Webseite führte. Scrollen war weiterhin über die Bildschirmkanten möglich. Um jedoch zoomen zu können, musste der Benutzer den Modus mittels Druck auf einen bestimmten Softkey wechseln. Das Zoomen war in diesem Modus erneut über das Drehrad zugänglich und es konnte weiterhin mit Hilfe des Mauszeigers an den Bildschirmkanten gescrollt werden. Das dritte Interaktionskonzept, Variante C, bestand ebenfalls aus zwei Modi. Im ersten Modus wurde das Linktraversierungskonzept zu Gunsten eines schnelleren Scrollens durch eine so genannte Segmenttraversierung ersetzt. Der zweite Modus in dieser Variante bot die selben Funktionen wie der zweite Modus in Variante B.

Eine detailliertere Beschreibung der Vorgehensweise bei der Anforderungsanalyse sowie die Ergebnisse der Fokusgruppen können in [Söter 08] nachgelesen werden. Die drei verschiedenen Interaktionskonzepte, sind in [Specht 10] detaillierter beschrieben.

4.1.2. Evaluation

Eine erste Evaluationsstudie im Januar 2009 sollte nun die drei verschiedenen Browservarianten auf ihre Gebrauchstauglichkeit hin überprüfen. Entsprechend dem DECIDE-Framework (vgl. Kapitel 3.4.1.) wurde zunächst das globale Ziel dieser Evaluationsstudie festgelegt: Das Ziel dieser Studie war es nicht, die beste Variante zu identifizieren. Vielmehr ging es darum, aus den grundlegenden Interaktionskonzepten der drei Varianten die vielversprechendsten für die Weiterentwicklung zu extrahieren. Die Stärken und Schwächen der jeweiligen Konzepte sollten aus Benutzersicht erfasst werden und die Probanden sollten ihr allgemeines Urteil zur Darstellung, Bedienung und Einfachheit des Systems äußern. Der Charakter dieser ersten Evaluationsreihe war explorativ und so wurden hier zunächst nur qualitative Daten erhoben.

Eine der generellen Forschungsfragen, die mit der Untersuchung beantwortet werden sollten, war, wie die drei Varianten im Allgemeinen von den Benutzern wahrgenommen werden. Darüber hinaus galt es drei speziellere Fragen zu beantworten:

F₁: Welche Bedienungsprobleme tauchen bei den jeweiligen Formen der Interaktion auf?

F₂: Welche Variante wird von den Probanden bevorzugt und warum?

F₃: Welche Funktionen sind essentiell, wünschenswert oder optional?

Dem Projektteam wurde von der Volkswagen AG ein Audi A8 zur Verfügung gestellt, welcher nicht nur für die stationären Tests eine realere Kulisse bot, als wenn man die Untersuchungen traditionell an einem Rechner mit installierter Software durchgeführt hätte. Zusätzlich konnten mit diesem Testfahrzeug auch mobile Tests durchgeführt werden, die eine Überprüfung der Bedienbarkeit in realer Umgebung ermöglichten. Auf diese Weise konnte ein Vergleich der Ergebnisse zwischen Stand und Fahrt gezogen werden.

Da es in der realen Umgebung eines automotiven Browsers zu Unterbrechungen durch Telefonate oder Ähnlichem kommen kann, wurde die Methode des Usability Testing etwas abgewandelt. Die Testumgebung sollte weniger kontrolliert sein und so wurde es den Probanden beispielsweise nicht verboten, ihre Mobilfunkgeräte beim Test eingeschaltet zu lassen. Dies gewährleistete ein noch realeres Szenario und es konnte eine höhere ökologische Validität der Tests erzielt werden.

Zur Untersuchung der Forschungsfragen wurde die Methode der Beobachtung eingesetzt, um die Reaktionen der Probanden in Bezug auf das System zu erkunden. Hierbei sollten die Probanden die Technik des lauten Denkens anwenden, damit der Untersuchungsleiter die Ursachen für eventuelle Problemstellen erfährt. Semi-strukturierte Interviews sollten nach jeder getesteten Variante und im Anschluss der Untersuchung die Vor- und Nachteile der jeweiligen Variante beleuchten und den jeweiligen Favoriten begründen.

Vor der Studie wurden vier Hypothesen aufgestellt, die nun überprüft werden sollten. Alle vier Hypothesen beruhen auf einigen Annahmen, die im Laufe der Konzeptionsphase vom Projektteam getätigt wurden. Diese werden im Folgenden erörtert.

H₁: Probanden nutzen in allen Varianten hauptsächlich die Mauszeigersteuerung für das Scrollen durch Webseiten.

H₂: Variante A (modusfreie Variante) ist am leichtesten nachvollziehbar und bietet daher das höchste Ease-of-Learning.

H₃: Probanden nach intensiver Nutzung Variante C (Segmenttraversierung) benutzen.

H₄: Die Auswahl aktiver Elemente (z.B. Links) erfolgt mittels Variante B (Linktraversierung) präziser.

Grund für H₁ ist, dass diese Art der Mauszeigersteuerung der Interaktion mit dem Desktop PC ähnelt und sie den Probanden daher vermutlich vertrauter vorkommt als alternative Methoden. Darüber hinaus wurden alternative Bedienkonzepte in den Fokusgruppen abgelehnt. H₂ hat das Verwenden verschiedener Interaktionsmodi zum Gegenstand und hat den Hintergrund, dass eine der Varianten modusfrei ist und somit alle möglichen Funktionen mit einem Handgriff zugänglich sind. So muss sich der Proband nicht auf einen Moduswechsel einlassen und sich ebenfalls nicht merken, in welchem Modus welche Funktionen verfügbar sind. Der Grund für H₃ könnte sein, dass das Drehen des Drehrades am MMI-Terminal um zu Scrollen der Bedienung einer Webseite mit dem Musrad nahe kommt und für die Probanden daher familiär wirken könnte. Darüber hinaus ist die Segmenttraversierung eine Interaktionsform, die schnelles Scrollen ermöglicht und der Proband auf diese Weise schnell einen Überblick über die zu Grunde liegende Webseite erhält. H₄ betrifft schließlich das präzise Arbeiten und hat den Hintergrund, dass die Mauszeigerpositionierung insbesondere während der Fahrt aufgrund von Erschütterungen beeinträchtigt wird und die Linktraversierung für dieses Problem eine adäquate Lösung bieten kann.

Im nächsten Schritt wurden die praktischen Fragen geklärt, die nötig waren, um die Hypothesen auf ihre Richtigkeit zu untersuchen. So wurde das Testfahrzeug mit einer Kamera ausgerüstet, die die Mimik, Gestik und Aussagen der Probanden festhalten sollte. Die Bildschirminhalte und Nutzeraktionen wurden mit der Aufzeichnungssoftware Morae® von TechSmith (vgl. Abbildung 4.07) erfasst. Bei den stationären Tests saß der Versuchsleiter neben dem Probanden und machte sich zusätzlich Notizen. Bei den mobilen Tests saßen zwei Versuchsleiter und ein Fahrer zusammen mit dem Probanden im Testfahrzeug. Für den Fall dass während der Tests Probleme auftreten sollten, waren technische Assistenten stets in der Nähe oder zumindest telefonisch erreichbar.

Die Probanden wurden entsprechend der überarbeiteten Personas ausgewählt. Dabei wurde darauf geachtet, sowohl die Generation der Audi Q7-Käufer, als auch deren Kinder abzudecken. Aus diesem Grunde wurden zwei Schüler im Alter von 15 Jahren, drei Studenten im Alter von 20 bis 24 Jahren sowie vier Berufstätige im Alter von 43 bis 52 Jahren akquiriert. Fünf Probanden waren männlich, vier waren weiblichen Geschlechts und sechs der neun Probanden hatten eine Sehschwäche. Zwei Probanden hatten beim stationären Test ihre Sehhilfe nicht dabei, einer beim mobilen Test. Da auch überprüft werden sollte, wie gut die Anforderungen aus den Fokusgruppen in den drei Prototypen

umgesetzt wurden, wurden zwei Personen als Probanden eingeladen, die bereits bei den Fokusgruppen teilgenommen hatten.

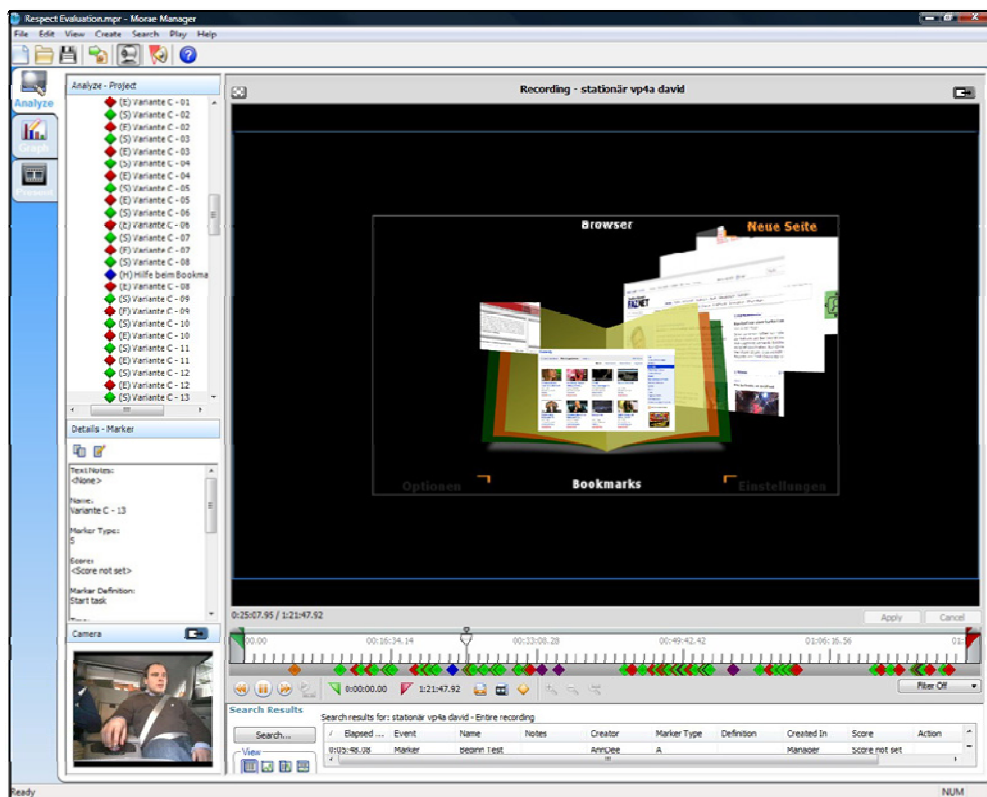


Abbildung 4.07: Der Morae® Manager bei der Auswertung einer Aufzeichnung.

Bevor die Probanden an den Tests teilnahmen, mussten sie einen Fragebogen ausfüllen. In diesem befanden sich Fragen zu den demographischen Daten der Probanden sowie zu deren Sehfähigkeit. Es wurde auch gefragt, ob der Proband Rechts- oder Linkshänder ist, um ihn beim Test auf der Seite sitzen zu lassen, die das bequeme Bedienen des MMI-Terminals mit seiner Führhand ermöglichte. Neben Fragen zur Computer- und Internetnutzung gab es zusätzlich die Frage nach dem Kenntnisgrad diverser Webseiten, die beim Test besucht wurden. Sollte der Proband gewisse Aufgaben nicht korrekt bearbeiten können, würde diese letzte Frage dabei helfen, die Ursache hierfür zu identifizieren.

Bevor mit dem Test begonnen wurde, wurden die ethischen Aspekte geklärt, indem den Probanden ein Willkommensschreiben vorgelegt wurde, in welchem der Zweck der Untersuchung geschildert wurde. Anschließend wurde um ihre Einverständnis zu den Videoaufnahmen gebeten und es wurde ihnen zugesichert, den Test zu jedem Zeitpunkt beenden zu können, falls sie sich dabei nicht mehr wohl fühlen sollten. Dieses Angebot wurde jedoch weder bei den stationären, noch bei den mobilen Tests wahrgenommen.

Während beim stationären Test Aufgaben auf realen Webseiten mit Zugriff auf das Internet bearbeitet wurden, wurden für den mobilen Test Webseiten aus dem Cachespeicher abgerufen. Ersteres

ermöglichte eine höhere Aktualität der Webseiten auf Kosten der Vergleichbarkeit. Letzteres erhöhte hingegen die Vergleichbarkeit der Ergebnisse und verminderte zusätzlich das Risiko von Verbindungsschwierigkeiten, die durch eventuelle Funklöcher während der Fahrt entstanden wären. Die stationären Tests dauerten im Schnitt 90 Minuten – die mobilen Tests waren mit einer durchschnittlichen Bearbeitungszeit von 40 Minuten etwas kürzer. Die kürzere Dauer sollte die Gefahr von Unwohlsein während der Fahrt minimieren. Tatsächlich war einem der fünf Probanden am Ende des mobilen Tests etwas übel. Ab diesem Zeitpunkt musste der Proband jedoch keine weiteren Aufgaben mehr bearbeiten. Das abschließende Interview wurde dennoch durchgeführt.

Bei dieser Evaluationsstudie handelte es sich um ein Within-Subject Design. Jeder Proband sollte hierbei alle drei Interaktionskonzepte testen, damit anschließend der persönliche Favorit gewählt werden und eine Rangfolge der Konzepte gebildet werden konnte.

Die Reihenfolge der drei verschiedenen Interaktionskonzepte wurde permutiert, sodass jede Variante mit gleicher Häufigkeit als erstes, zweites oder drittes getestet wurde. Hiermit sollte verhindert werden, dass Usability Probleme oder die Favorisierung von Lerneffekten beeinflusst wurden. Der schematische Ablauf der Tests lässt sich Abbildung 4.08 entnehmen.

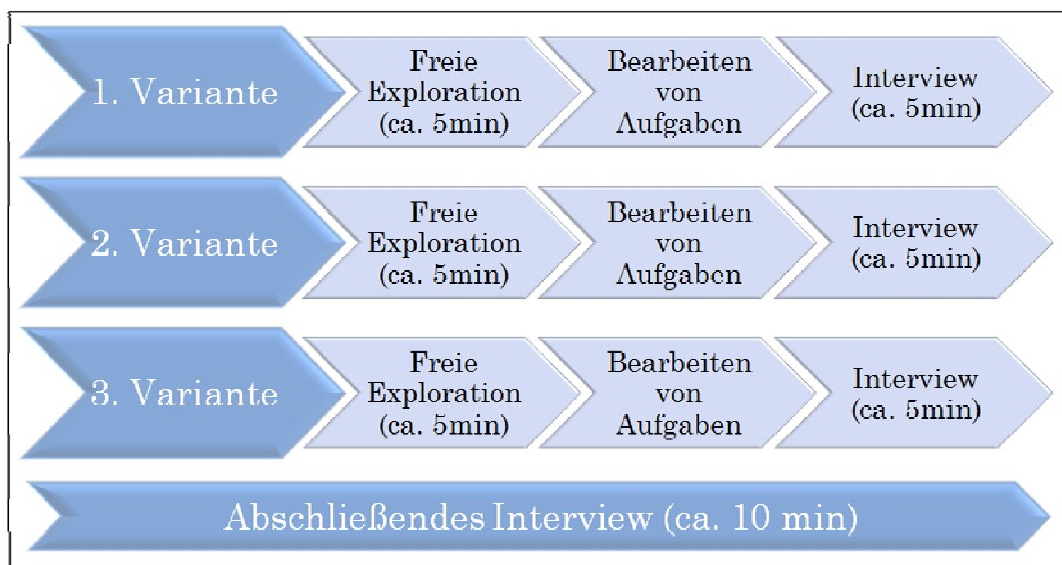


Abbildung 4.08: Ablauf der Benutzertests der explorativen Evaluationsstudie.

Vor jeder Variante wurde dem Probanden eine Explorationsphase eingeräumt, in der er die Funktionen selbstständig entdecken konnte. Auf diese Weise konnte sich jeder Proband zunächst an das neue Interaktionskonzept gewöhnen, bevor er damit Aufgaben bearbeiten musste. Diese Phase durfte selbstständig beendet werden und anschließend wurden jedem Probanden vom Untersuchungsleiter die Funktionen gezeigt, die er bis dahin nicht entdeckt hatte. Danach erhielt jeder Proband nacheinander Aufgaben auf Karteikarten, die er laut vorlesen und anschließend bearbeiten sollte. Während des

gesamten Tests sollte er die Methode des lauten Denkens verwenden um auf Ereignisse hinzuweisen, die er nicht erwartet hätte oder die ihm besonders gut gefallen haben. Mit Hilfe der so getätigten Aussagen jedes Probanden wurden nicht nur die Stärken und Schwächen der verschiedenen Konzepte identifiziert, sondern es wurde auch überprüft, wie gut und wie schnell der Proband das neue Interaktionskonzept verinnerlicht hat und welche Aspekte die jeweils größte mentale Beanspruchung erforderten.

Jeder Proband durfte selbst entscheiden, wann er die Aufgabe für bearbeitet hält und musste dies durch Rückgabe der Karteikarte an den Versuchsleiter demonstrieren. Wenn alle Aufgaben mit einer Variante bearbeitet wurden, wurde in einem kurzen Interview nach den zwei besten und den zwei schlechtesten Aspekten der gerade getesteten Variante gefragt. Zudem sollte der Proband beschreiben, wie gut er sich allgemein mit der Variante zurecht gefunden hat und ob ihm manche Aspekte zu lange gedauert haben oder andere zu schnell waren. Verbal wurde hier noch kein Vergleich mit einer eventuellen vorangegangenen Variante erzwungen. Nach dem Interview wurde der selbe Prozess mit der nächsten Variante durchgeführt.

Nach Bearbeitung aller Aufgaben mit allen Varianten wurde ein abschließendes Interview geführt, in dem der Proband seinen Favoriten unter den drei Konzepten inklusive einer Begründung für seine Wahl nennen sollte. An dieser Stelle wurde er gebeten, die drei getesteten Konzepte in eine persönliche Rangfolge zu bringen. Ebenfalls wurde nach der Variante gefragt, die am angenehmsten zu bedienen war sowie nach jener, die am nachvollziehbarsten erschien. Der Proband sollte zusätzlich einschätzen, mit welcher Variante er die Aufgaben am schnellsten bearbeiten konnte. Darüber hinaus wurde jeder Proband gefragt, wie gut oder wenig gut ihm der Moduswechsel gefallen hat, der in zwei der drei Varianten eingebaut war. Die vorerst letzte Frage war, ob alle Funktionen vorhanden waren, die der Proband im Alltag nutzt oder ob er Funktionen vermisst hat.

Mit Hilfe der Antworten zu den Interviewfragen konnten die Forschungsfragen F_1 , F_2 und F_3 beantwortet werden. Ergebnisse der Beobachtung während der Bearbeitung der Aufgaben durch die Probanden lieferten zusätzliche Informationen zur Beantwortung von F_1 . Welche Aspekte hierbei beachtet wurden, wird im Nachfolgenden beschrieben.

Die Aufgabenblöcke, die die Probanden mit jeder Variante bearbeiten sollten, waren ähnlich und wurden so verteilt dass jeder Block mit jeder Variante gleich oft bearbeitet wurde. Auf diese Weise wurde ausgeschlossen, dass Probanden Probleme mit einer Aufgabe auf eine bestimmte Variante projizieren. Die Aufgaben wurden so ausgesucht, dass sie möglichst alle Funktionen des Browsers abdecken und jenen Aufgaben entsprechen, die in den Fokusgruppen als wichtig empfunden wurden. Eine weitere Basis für die Aufgaben waren die Anforderungen seitens der Volkswagen AG.

Es resultierten drei 15-teilige Aufgabenblöcke, bestehend aus drei verschiedenen Aufgabentypen. Der erste Typus waren Navigationsaufgaben, bei denen es um die Präzision der Mauszeigersteuerung und um die Interaktion mit Web2.0-Elementen ging. Hintergrund waren unter anderem Aussagen aus den

Fokusgruppen, in denen der Wunsch nach Bedienung von Web2.0-Anwendungen wie beispielsweise Videoclips auf YouTube angegeben wurde. Acht Teilaufgaben sollten den Bereich der Navigationsaufgaben abdecken. Die dazugehörigen Aufforderungen an den Probanden waren beispielsweise: „*Starten / Pausieren / Spulen Sie das Video.*“ oder „*Wählen Sie den Artikel aus.*“.

Der zweite Aufgabentyp bestand aus Orientierungs- und Wahrnehmungsaufgaben bei denen horizontales und vertikales Scrollen sowie Zoomen die Ziele waren. Da es für die Teilnehmer der Fokusgruppen eine wichtige Aufgabe darstellte, sich zu Informieren (vgl. Kapitel 4.1.1.) und das kleine Display in großem Abstand eventuelle Schwierigkeiten beim Lesen bergen könnte, wurden diesem Aufgabentyp neun Teilaufgaben gewidmet. Beispiele hierzu lauten wie folgt: „*Navigieren Sie zum ersten Artikel aus dem Bereich Politik.*“ oder „*Lesen Sie die erste Zeile der fett gedruckten Zusammenfassung vor. Vergrößern Sie den Artikel, wenn Sie ihn dann besser lesen können.*“. Mit der zweiten Frage sollte unter anderem überprüft werden, ob die Probanden zoomen müssen, um Texte angenehmer lesen zu können und welche Zoomstufe hierbei gewählt wird.

Der dritte Aufgabentyp waren Aufgaben zum Bookmarkmanagement, welches für die Teilnehmer der Fokusgruppen eine essentielle Funktion eines Browsers darstellt. Bei dieser Art von Aufgaben sollten die Probanden Bookmarks anlegen und wieder auffinden.

Nach jeder bearbeiteten Aufgabe wurde der Proband befragt, wie schwierig es seiner Meinung nach war die jeweilige Aufgabe mit der zu Grunde liegenden Browservariante zu lösen. Hierbei sollte er sich für eine Zahl auf einer Skala von 1 (sehr leicht) bis 5 (sehr schwer) entscheiden. Diese Benotung wurde anschließend jeweils einem oder zwei Aufgabentypen zugeordnet. Nach den Tests wurde für jede Variante bei jedem Probanden der Durchschnitt ausgerechnet und überprüft, ob sich das Ergebnis mit der Wahl des Favoriten deckt.

Bei den mobilen Tests nahmen fünf Probanden teil, die bereits beim stationären Test partizipiert hatten. Auch hier wurde ein Within-Subject Design gewählt, sodass jeder Proband erneut alle drei Browserkonzepte testen und auf ihre Gebrauchstauglichkeit in realer Umgebung überprüfen konnte. Die mobilen Tests fanden jeweils vier bis sechs Tage, im Durchschnitt fünf Tage, nach dem stationären Test statt. Um eine Vergleichbarkeit der Ergebnisse zu gewährleisten, wurde eine Teststrecke im Konstanzer Stadtverkehr gewählt, die mit jedem Probanden abgefahren wurde. Da die Probanden unterschiedlich lange für die Bearbeitung der Aufgabenblöcke brauchten, war die Teststrecke so konzipiert, dass ein Abschnitt gegebenenfalls mehrfach abgefahren wurde.

Die Reihenfolge der Browservarianten wurde auch hier bei jedem Test verändert, damit erneut eine Beeinflussung durch Lerneffekte vermieden wird. Zusätzlich wurde beim mobilen Test mit der Variante begonnen, die beim stationären Test als Favorit angegeben wurde. Sollten tatsächlich Lerneffekte bei der Wahl des Favoriten ausschlaggebend sein, hätte auf diese Weise beim mobilen Test eine der beiden anderen Varianten der neue Favorit werden müssen.

Auch in dieser Testreihe wurden Aufgaben zur Navigation, Orientierung und Wahrnehmung sowie zum Bookmarkmanagement bearbeitet, jedoch waren es hier nur jeweils sechs Aufgaben pro Variante. Während bei den stationären Tests aufgrund der sich stets aktualisierenden Webseite keine Fragen zu Inhalten gestellt werden konnten, war dies bei den mobilen Tests durch Verwendung der Webseiten aus dem Cachespeicher möglich. Um diese Art der Aufgaben bearbeiten zu können, mussten die Probanden während der Fahrt Texte lesen und bei der Auswertung konnte überprüft werden, wie gut dies bei Erschütterungen möglich war und welche Zoomstufe hierbei gewählt wurde. Wieder sollten die Probanden nach Bearbeitung jeder Aufgabe eine Bewertung abgeben, wie schwierig es mit der aktuellen Variante war, die jeweilige Aufgabe zu lösen.

Bei der Auswertung der erfassten Bildschirminhalte wurde darauf geachtet, auf welche Weise die Probanden scrollten, wenn sie mehrere Alternativen zur Auswahl hatten. Sie wurden in diesem Zusammenhang auch bereits während der Tests befragt, wie gut sie mit der Segmenttraversierung beim Scrollen zurechtkamen. Ebenfalls wurde überprüft, welche Variante der Linkauswahl gewählt wurde, wenn hier mehrere Möglichkeiten geboten wurden. Hatten die Probanden gezoomt, wurde die Zoomstufe ausgewertet um zu überprüfen, wie groß die Schrift sein muss und ob ein Zoomen überhaupt nötig war, um Texte angenehm lesen zu können. Dies wurde vor allem auch in Hinblick auf Sehschwächen und das Abhanden sein einer Sehhilfe bei einer solchen übergeprüft. Der Umgang der Probanden mit verschiedenen Interaktionsmodi wurde inspiziert und sie wurden auch während der Tests nach der Benutzerfreundlichkeit des Moduswechsels befragt. Es wurde ebenfalls untersucht, inwieweit sich das allgemeine Browserkonzept über einen Zeitraum von durchschnittlich fünf Tagen merken lässt und ob die Probanden sich spätestens nach dem Durchlauf mit der ersten Variante an den Aufbau und die Bedienung erinnern können.

Sowohl bei den stationären, als auch bei den mobilen Tests stellte sich heraus, dass das Anvisieren und Treffen von Links problematisch ist. Vor allem die Erschütterungen des Fahrzeugs, die standardmäßig während der Fahrt auftreten, sind für das präzise Arbeiten hinderlich. Obwohl die Linktraversierung in Variante B für die meisten Probanden als wenig benutzerfreundlich erschien, wurden ihre Vorteile vor allem bei den mobilen Tests erkannt: sie ermöglicht trotz weniger genauer Arbeitsweise und trotz Erschütterungen eine präzise Auswahl von Links. Die zeilenorientierte Reihenfolge der angewählten Elemente erschien jedoch wenig intuitiv.

Im nächsten Iterationsschritt musste also ein Weg gefunden werden, auch während der Autofahrt eine präzise Auswahl von Links zu ermöglichen. Die Reihenfolge darf hier jedoch nicht vorgegeben werden damit der Benutzer nicht das Gefühl erhält, fremdbestimmt zu sein. Eine Art Mauszeigermagnetismus sollte für dieses Problem eine adäquate Lösung bieten.

Ein Aspekt, der den Großteil der Probanden dazu geführt hat, sowohl im stationären, als auch im mobilen Test Variante C (Segmenttraversierung) zu bevorzugen, war die Möglichkeit des schnellen

Scrollens. Um sich einen Überblick über die Webseite zu verschaffen, erschien die Möglichkeit mit dem Mauszeiger über die Bildschirmkanten zu scrollen, als zu langwierig. Obwohl vier Probanden das gleichzeitige horizontale und vertikale Scrollen als verwirrend empfanden, wählten sieben der neun Probanden im stationären Test diese Variante zu ihrem Favoriten. Beim mobilen Test sprachen sich drei der fünf Probanden für diese Variante aus.

Um das Scrollen angenehmer zu gestalten, musste auch hier die Kontrolle, die durch die vorgegebene Reihenfolge der Segmente vorhanden war, an den Benutzer abgegeben werden. Es musste jedoch eine Möglichkeit gefunden werden, trotzdem ein schnelles Scrollen zu ermöglichen. Im nächsten Schritt wurde daher auf das horizontale Scrollen verzichtet. Ein Sachverhalt, der dies unterstützte, war die Änderung einer Rahmenbedingung durch die Volkswagen AG: Die Auflösung des Displays betrug nicht länger eine Breite von 800 Pixeln, sondern wurde auf 1024 Pixel erhöht. Da die meisten Webseiten, die für die Teilnehmer der Fokusgruppen relevant waren, diese Breite nicht überschreiten, konnte trotz Ausschluss des horizontalen Scrollens die gesamte Webseitenbreite angezeigt werden. Für jene Benutzer, die dennoch das Scrollen über die Bildschirmkante bevorzugen, sollte diese Möglichkeit weiterhin beibehalten werden.

Sieben der neun Probanden gaben bei den Benutzertests an, auch ohne Vergrößerung Texte angenehm lesen zu können. Zwei der Probanden tätigten diese Aussage, obwohl sie eine Sehschwäche haben und ihre Sehhilfe bei den Tests nicht verfügbar war. Zu Beachten gilt jedoch, dass bei den getesteten Varianten der Zoom nur zu einer verpixelten Vergrößerung des Textes führte. So war das Lesen hier bei einer etwas höheren Zoomstufe meist schwieriger, als bei der Basisstufe. Im finalen Browserkonzept sollte allerdings ohnehin nicht auf das Zoomen verzichtet werden, da es das Lesen vor Allem während der Fahrt angenehmer macht und es auch Benutzer gibt, die ansonsten die kleinen Schriftgrößen nicht entziffern könnten. Ob Zoomen nun tatsächlich nötig ist und welche Zoomstufe hierbei gewählt wird, musste in einem späteren Test mit dem finalen Prototypen überprüft werden.

Einigen Probanden fiel es schwer, mit verschiedenen Interaktionsmodi und dem dazugehörigen Moduswechsel umzugehen. Ein Aspekt, der diesen Umstand sicherlich erschwerte war, dass eine Mauszeigersteuerung um Zoommodus nicht möglich war – stattdessen konnte mittels Bewegung des Joysticks der gezoomte Bildschirmausschnitt bewegt werden, ohne dass der Proband hierfür an die Kanten fahren musste. Weiterhin erschien es problematisch, dass die gewählte Zoomstufe verlassen wurde, sobald der Proband den Modus wechselte. Aus diesem Grunde wurde die modusfreie Browservariante A von fünf Probanden als die nachvollziehbarste Variante deklariert.

Das Ziel der Modusfreiheit sollte zu Gunsten der Einfachheit und Nachvollziehbarkeit sicherlich angestrebt werden, doch sind es weiterhin sehr viele Funktionen, die der Benutzer eines automotiven Browsers benötigt. Diese in einer modusfreien Variante einzubetten ist bisher nicht gelungen, ohne dafür Einschränkungen bei einigen wichtigen Funktionen in Kauf zu nehmen. Daher war das Ziel für die weitere Entwicklung, die beiden Interaktionsmodi so zu optimieren, dass der spätere Benutzer seinen individuell favorisierten Modus beibehalten und dennoch alle wichtigen Aktionen durchführen

kann. Eine Bedingung hierfür war, dass die Funktion des Mauszeigers in beiden Modi vorhanden sein muss.

Ferner stellte sich heraus, dass das Bookmark-Konzept nicht intuitiv war. Statt die Favoriten in einem separaten Ordner in der Hauptansicht neben den geöffneten Webseiten zu platzieren, musste eine Möglichkeit gefunden werden, die Trennung visuell noch stärker zu hervorzuheben. Aus diesem Grunde wurde der Bookmark-Ordner entfernt und in einer separaten „Welt“ untergebracht, sodass der Benutzer nun zwischen der Bookmark-Welt und der Geöffnete-Webseiten-Welt wechseln kann. Verwirrungen aufgrund des ähnlichen Designs beider Welten sollten so vermieden werden.

Ebenfalls als das Design betreffend stellten sich die Probleme mit den Softkeys dar – einige Probanden versuchten diese zu aktivieren, indem sie mit dem Mauszeiger darüber fuhren und zu klicken versuchten. Grund hierfür war vermutlich die Button-ähnliche Anmutung, die durch ein Re-Design umgangen werden musste.

Sämtliche Usability-Probleme waren in dieser Testreihe unabhängig von der jeweiligen Zielgruppe. Lediglich bezüglich der Linktraversierung fiel auf, dass sich die gesamte Gruppe der Berufstätigen, und nur diese, negativ über die Reihenfolge der Elemente äußerte.

Die erste der anfangs aufgestellten Hypothesen wurde durch die Ergebnisse der Interviews und der Beobachtung des Nutzerverhaltens widerlegt. H_1 lautete: *„Probanden nutzen in allen Varianten hauptsächlich die Mauszeigersteuerung für das Scrollen durch Webseiten.“*. Dies stellte sich als falsch heraus, da das Scrollen über die Bildschirmkanten für die meisten Probanden zu langwierig erschien. Wurden schnellere Alternativen angeboten, wurden diese eher wahrgenommen. So bevorzugte der Großteil der Probanden das Scrollen mittels Segmenttraversierung. Ein anderer Teil scrollte mittels Joystickbewegung im Zoommodus, da hier der Mauszeiger nicht erst an den Bildschirmrand bewegt werden musste, um den Bildausschnitt zu verschieben.

Die anderen drei Hypothesen hingegen erwiesen sich als korrekt: Tatsächlich war *„Variante A (modusfreie Variante) am leichtesten nachvollziehbar und bietet daher das höchste Ease-of-Learning.“*. Obwohl Variante C mit ihrer Segmenttraversierung bei den Tests die meisten Anhänger fand, empfanden die meisten Probanden die modusfreie Variante A als besser nachvollziehbar. Die Identifikation der beiden Modi erforderte noch eine zu hohe mentale Beanspruchung, weshalb der Moduswechsel und die Modi selbst noch stark überarbeitet werden mussten.

Es fiel auf, dass die *„Probanden nach intensiver Nutzung Variante C (Segmenttraversierung) benutzen“*. Der Grund hierfür liegt in der Möglichkeit des schnellen Scrollens und des schnellen Überblicks über die Webseite, die diese Variante ihrem Benutzer bietet. Wie auch das Widerlegen von H_1 zeigt, muss bei der weiteren Entwicklung des Browsers eine Möglichkeit zum schnellen Navigieren innerhalb von Webseiten ermöglicht werden, um die subjektive Zufriedenheit des Nutzers zu gewährleisten.

Dass „*die Auswahl aktiver Elemente (Links) mittels Variante B (Linktraversierung) präziser erfolgt.*“, bestätigten einige Probanden vor allem beim mobilen Test. Die Positionierung mittels Joystick war bereits im Stand oft langwierig und dieser Sachverhalt wurde durch Erschütterungen während der Fahrt verstärkt. Da die Linktraversierung für die meisten Probanden unnatürlich und verwirrend erschien, musste in Zukunft eine alternative Variante der Optimierung gefunden werden.

Die ursprünglich gestellten Forschungsfragen konnten durch diese Evaluationsstudie ebenfalls beantwortet werden. Nach eigenen Aussagen kamen alle Probanden mit allen drei Varianten gut oder sehr gut zurecht. Das Surfen im Fahrzeug gefällt im Allgemeinen, doch gaben einige Probanden an, dass die Bedienung mittels MMI-Terminal gewöhnungsbedürftig sei. Dies spiegelt sich auch in den bereits genannten Problemen der präzisen Auswahl wieder.

Eine Funktion, die sich die Probanden bei einem automotiven Webbrowser wünschen, ist eine schnellere Scrollfunktion als das Scrollen über die Bildschirmkanten. In diesem Zusammenhang fiel auf, dass ein einfacher Zugriff auf eine schnelle Scrollfunktion wichtiger zu sein scheint, als ein einfacher Zugriff auf eine Zoomfunktion. Der Grund hierfür liegt vermutlich darin, dass das Scrollen bereits beim Surfen und Suchen benötigt wird, wohingegen es des Zoomens erst beim Lesen von Texten bedarf. Ein Proband wünschte sich eine Anzeige, die wie die Scrollingleiste am handelsüblichen Browser, die Länge der aktuellen Webseite anzeigt. Die Mehrzahl der Probanden hatte jedoch in den vorgestellten Prototypen alle Funktionen gefunden, die sie im alltäglichen Gebrauch benötigen.

Bezüglich des Vergleichs der Ergebnisse des stationären und des mobilen Tests ergaben sich keine gravierenden Unterschiede. Es stellte sich lediglich heraus, dass die Probleme, die bei den stationären Tests auftraten, während der Fahrt verstärkt wurden: die Mauszeigerpositionierung war ungenauer und Probanden, die den Moduswechsel oder die Segmenttraversierung bereits stationär als verwirrend empfanden, erschienen diese als noch befremdlicher. Hinsichtlich des Zoom-Verhaltens konnte bei dieser Testreihe kein Unterschied zwischen Stand und Fahrt festgestellt werden.

4.1.3. Herausforderungen und Einschränkungen

Sowohl die Anforderungsanalyse, als auch die Evaluationsstudie stellte das Projektteam vor diverse Herausforderungen, die zum Teil allgemeiner Natur waren und zum Teil insbesondere aufgrund des automotiven Kontextes entstanden.

Eine allgemeine Herausforderung gab es durch den Wechsel vom Audi A8 zum Audi Q7. Durch den neuen Anwendungskontext gab es mehrere Zielgruppen, deren Bedürfnisse es zu beachten galt. Während für die Entwicklung für den Audi A8 hauptsächlich Geschäftsleute hätten befragt werden müssen, sollten für den Audi Q7 die Bedürfnisse von Jugendlichen, jungen Erwachsenen, Berufstätigen und Rentnern berücksichtigt werden.

Bereits in der Phase der Anforderungsanalyse wurden daher die Personas so erstellt, dass sie die Mitglieder einer Familie repräsentierten, in welcher jeder den automotiven Browser einfach und gut bedienen können sollte. Die Teilnehmer der Fokusgruppen wurden schließlich so ausgewählt, dass sie einigen dieser Zielgruppen entstammten. Die erste Zielgruppe war die der jungen Erwachsenen, die mitten im Studium sind und sich für neue Technologien und Interaktionsmöglichkeiten interessieren. Eine zweite Zielgruppe umfasste technikaffine Berufseinsteiger. Eine dritte Zielgruppe bestand aus Berufstätigen, die sich weniger gut mit alternativen Eingabemöglichkeiten auskennen und ihnen daher eher kritisch gegenüberstehen. Die Teilnehmer der Fokusgruppen sollten sich unter anderem auch in Schüler und Rentner hineinversetzen um deren Bedürfnisse und Charakteristiken gemeinsam zu erschließen und zu diskutieren. So wurde versucht, mit Hilfe der bestehenden Personen alle Zielgruppen abzudecken. Idealerweise hätten jedoch auch noch unmittelbar diese Personen eingeladen werden sollen, um die jeweiligen Informationen aus erster Hand und ohne Mittelspersonen zu erhalten.

Auch bei der Evaluationsstudie mussten die diversen Zielgruppen weiterhin berücksichtigt werden. Hier wurden neben Studenten und Berufstätigen auch noch zwei Schüler eingeladen, um oben genanntes Problem zu umgehen. Es wurde zudem in jeder Zielgruppe auf eine möglichst ausgewogene Verteilung der Geschlechter geachtet. In der Gruppe der Berufstätigen wurde zudem darauf geachtet, sowohl Personen mit, als auch Personen ohne Hochschulabschluss mit einzubeziehen.

Eine zweite Herausforderung betraf die Anzahl der Probanden. Bei qualitativ orientierten Usability-Tests in der Mensch-Computer Interaktion wird typischerweise mit etwa fünf bis zwölf Probanden pro Benutzergruppe getestet. Aus zeitlichen und finanziellen Gründen war es dem Forschungsteam zwar nicht möglich mit so vielen Probanden zu testen. Aus diesem Grunde wurde darauf geachtet, eine repräsentative Zusammenstellung der Probanden zu gewinnen um zumindest alle Benutzergruppen abzudecken. In der hier beschriebenen Studie sollten zwei Generationen getestet werden: potentielle Käufer eines Audi Q7 sowie deren Partner und potentielle Kinder eines Käufers. Die Gruppe der potentiellen Käufer wurde hier mit vier Probanden abgedeckt. In der Gruppe der potentiellen Kinder, die sich in dieser Testreihe weiter in Schüler und Studenten aufteilte, wirkten fünf Personen mit: zwei Schüler und drei Studenten. Der Vorteil der heterogenen Probandengruppe liegt darin, dass verschiedene Zielgruppen eventuell verschiedene Problemstellen haben, die auf diese Weise innerhalb einer Evaluationsstudie aufgedeckt und im nächsten Iterationsschritt beseitigt werden können. Nachteilig ist jedoch die größere Varianz, die diese heterogene Gruppe zur Folge hat. Auffälligkeiten in den Ergebnissen könnten auf sehr individuellen Charakteristiken beruhen, die möglicherweise zufällig bei mehreren der ausgewählten Probanden einer Gruppe auftreten. Vor Allem bei geringer Personenzahl ist es daher heikel, gruppeninterne oder gruppenübergreifende Vergleiche zu ziehen oder generelle Aussagen über eine der Gruppen zu machen. Doch durch den vielfältigen Input, der mit einer heterogenen Gruppe gewonnen werden kann, war dieses Testdesign besser für die Evaluation des

Gesamtkonzepts des Browsers sowie der verschiedenen, tieferliegenden Interaktionskonzepte geeignet.

Ein wichtiger Sachverhalt, der für die traditionelle Mensch-Computer Interaktion untypisch ist und mit dem aus diesem Grunde ebenfalls umgegangen werden musste, betraf den automotiven Kontext der Studie. Bei typischen Benutzbarkeitstests wird zumeist in einem Labor mit stark kontrollierter Umgebung getestet. So sollen Unterbrechungen durch mögliche externe Besucher, Telefonate oder ähnliche alltägliche Gegebenheiten verhindert werden um das Leistungspotenzial des Nutzers nicht zu beeinträchtigen [Sharp et al. 07]. Ein solches Design war jedoch im Falle dieses Projekts wenig geeignet, da es um ein Nutzungsumfeld handelt, in welchem es ständig durch Telefonate oder die wechselnde Umgebung während Fahrt zu Unterbrechungen kommen kann. Um die ökologische Validität der Evaluationsstudie zu wahren wurde das Design in eine etwas weniger kontrollierte Variante umgewandelt. Da die Tests zudem nicht in einem Usability Labor, sondern in der tatsächlichen Nutzungsumgebung des Webbrowsers stattfanden, empfanden die Probanden auch die stationären Tests als realistische Situation der Benutzung. Diese realistische Testumgebung sollte das Gefühl des Getestet-Werdens minimieren und reale Probleme der Interaktionskonzepte aufzeigen. Zudem sollte es bewirken, dass die Probanden die Aufgaben ernst nehmen.

Da die mobilen Tests im Stadtverkehr stattfanden, konnte hier ebenfalls auf Ampeln, Staus, Passanten oder Ähnliches nicht eingewirkt werden. Der Vorteil dieses wenig kontrollierten Test-Designs lag darin, dass die Probanden weniger das Gefühl hatten, sich in einer Testumgebung zu befinden oder gar selbst die Testobjekte zu sein. Durch die Untersuchung im tatsächlichen Anwendungskontext sollten zudem validere Ergebnisse entstehen, die der realen Nutzung entsprechen.

Es wäre möglich gewesen, die Testreihen in einem Usability Labor oder in Fahrsimulator-ähnlichen Aufbauten durchzuführen, doch in diesen Umgebungen hätte es sowohl an physischer, als auch an funktionaler Wiedergabetreue gemangelt und die Situation hätte eher künstlich gewirkt. Die Umgebung hätte dem Probanden dadurch eher das Gefühl eines Usability Tests und statt dem einer natürlichen Benutzung vermittelt. Auch hätte der mobile Test durch die Abwesenheit kinästhetischer Signale zu anderen Ergebnissen geführt, da beispielsweise die Mauszeigerpositionierung oftmals erst durch die Erschütterungen während der Fahrt als größeres Problem angesehen wurde.

4.2. Fallstudie B: Benutzertest im automotiven Kontext

Die Ergebnisse der in Kapitel 4.1.2. beschriebenen Evaluation bestätigten die in den Fokusgruppen gefundenen Anforderungen und führten zu neuen Erkenntnissen in Bezug auf die Prioritäten der Benutzer beim automotiven Surfen. Entsprechend des in Abbildung 4.01 vorgestellten Lifecycles führten sie zu einer erneuten Diskussion der Anforderungen innerhalb des Projektteams und schließlich zu einem Neudesign des Prototyps. Es entstand ein neuartiges Browserkonzept, welches in einer zweiten Evaluationsstudie auf seine Gebrauchstauglichkeit hin überprüft werden sollte. Im Gegensatz zur in Kapitel 4.1.2. beschriebenen Studie lag der Fokus nicht mehr auf dem qualitativen Vergleich verschiedener Konzepte. Vielmehr galt es herauszufinden, inwiefern das neue und optimierte Interaktionskonzept den Problemen der automotiven Bedienung entgegenwirken konnte. Das Ziel der Studie war es daher, das Eingabegerät MMI-Terminal in Kombination mit einer optimierten Mauszeigersteuerung experimentell zu überprüfen.

Für die Untersuchung der Effektivität und Effizienz von stationär genutzten Eingabegeräten am Desktop-PC existiert eine Richtlinie in Form der ISO 9241-9 Annex B [ISO 9241-9]. Ein automotiver Nutzungskontext stellt jedoch andere Herausforderungen an die Evaluation, da sich die Ziele und Aufgaben von denen am Desktop-PC unterscheiden. Aus diesem Grund war es nicht praktikabel, die Richtlinien der ISO 9241-9 unverändert zu übernehmen. Die in den nachfolgenden Kapiteln vorgestellte Evaluation erforderte vielmehr eine alternative Herangehensweise, die das Entwicklerteam vor neue Herausforderungen stellte.

In Kapitel 4.2.1. wird das neue Browserkonzept kurz vorgestellt. Eine ausführliche und detailliertere Beschreibung des Systems sowie seiner Implementation ist in [Specht 10] zu finden. Die für die nachfolgende Evaluation neu definierten Ziele und die damit verbundenen Hypothesen und Aufgaben werden in Kapitel 4.2.2. erläutert. Im Anschluss werden das Testdesign sowie die Auswahl der Probanden für diese zweite Evaluationsstudie beschrieben, bevor in Kapitel 4.2.4. die Auswertung der gefundenen Daten und die Ergebnisse der Studie erörtert werden. Schließlich folgt in Kapitel 4.2.5. eine Zusammenstellung der Herausforderungen, die der alternative Evaluationsansatz mit sich brachte.

4.2.1. Das zu evaluierende System

Der finale Prototyp unterschied sich von der vorigen Generation in diversen Aspekten. Zum einen wurde die Gestaltung der Hauptansicht von einer V-förmigen Anordnung der Webseiten in die Darstellung eines Rondells übertragen (vgl. Abbildung 4.09 und 4.10). Das Bookmark-Konzept wurde hierbei ausgegliedert, um die mentale Trennung auch visuell stärker zu verdeutlichen. Die Anforderung seitens der Volkswagen AG, eine Bildschirmauflösung von 800 Pixel Breite zu berücksichtigen wurde auf 1024 Pixel erweitert, sodass nun die gesamte Breite einer typischen Webseite (entsprechend der in den Fokusgruppen genannten häufig genutzten Webseiten) auf dem Bildschirm dargestellt werden konnte.



Abbildung 4.09: Hauptansicht des Browsers der ersten Generation.



Abbildung 4.10: Hauptansicht des finalen Browsers.

Dadurch, dass nun die gesamte Breite einer Webseite auf dem Bildschirm dargestellt werden konnte, wurde das horizontale Scrollen, wie es in der ersten Browsergeneration mit Hilfe der Segmenttraversierung möglich war, hinfällig. Stattdessen konnte durch Drehen des MMI-Terminals ein vertikales Scrollen ermöglicht werden und dabei dennoch die gesamte Webseite betrachtet werden. Weiterhin wurden dem Benutzer zwei Interaktionsmodi zur Verfügung gestellt. Im Zoom-Modus wurde jedoch weiterhin die grundlegende Mauszeigersteuerung mittels Bewegen des MMI-Terminals ermöglicht. Zudem wurde die gewählte Zoomstufe beibehalten, wenn der Modus verlassen wurde.



Abbildung 4.11: Vertikales Scrollen einer Webseite auf Zoomstufe 1.9 mit halbtransparenter Vorschau im rechten Bildschirmbereich.

Das Scrollen in gezoomten Bereichen erfolgte ebenfalls nur in vertikaler Richtung, sodass beispielsweise vergrößerte Nachrichtenartikel problemlos gelesen werden konnten (vgl. Abbildung 4.11). Eine halbtransparente Vorschau der Webseite im rechten Bildschirmbereich zeigte dem Benutzer den angezeigten Ausschnitt im Kontext der gesamten Webseite an und sollte ihm zusätzlich als Orientierungshilfe dienen (vgl. rechter Bereich in Abbildung 4.11).

Zusätzlich sollte der Browser mit einer neuen Variante des MMI-Terminals bedient werden. Während dieses bei der ersten Browsergeneration aus einem Dreh-Drücksteller bestand, auf dessen Kopf ein Joystick für die Mauszeigersteuerung positioniert war, entfiel dieser Joystick bei der neuen Variante. Stattdessen wurde die Möglichkeit geboten, den Dreh-Drücksteller zusätzlich zu schieben um so den Mauszeiger zu positionieren (vgl. Abbildung 4.12 und 4.13). Da das MMI-Terminal jedoch nur auf die Richtung und nicht auf die Kraft des Drucks reagierte, wurde die präzise Mauszeigersteuerung erschwert.



Abbildung 4.12: Altes MMI-Terminal mit Joystick zur Mauszeigersteuerung.



Abbildung 4.13: Neues MMI-Terminal mit Schiebemöglichkeit zur Mauszeigersteuerung.

Dieses Problem sollte mit der bedeutendsten Veränderung im Vergleich zur ersten Browsergeneration behoben werden. Diese Veränderung betraf die Navigation auf Webseiten: Es wurden in diesem Zusammenhang zwei Mechanismen konzipiert und implementiert, welche eine Optimierung der Mauszeigersteuerung bewirken sollten. Mit Hilfe dieser Mechanismen sollten auch während der Fahrt eine einfache und schnelle Auswahl von Links und anderer aktiver Elemente ermöglicht werden. Ziel war es jedoch, dass die Mechanismen die Effektivität und Effizienz der Interaktion zwar erhöhten, aber dennoch unauffällig blieben. Der erste Mechanismus stellt eine intelligente Be- und Entschleunigung des Mauszeigers dar, die sowohl von der Dauer der Bewegung, als auch von den umgebenden Elementen bestimmt wurde (vgl. Abbildung 4.14). Der zweite Mechanismus ist eine von den umgebenden Elementen abhängige Ablenkung des Mauszeigers und wird im Nachfolgenden als „dynamische Kraftfelder“ bezeichnet (vgl. Abbildung 4.15) [Specht 10].

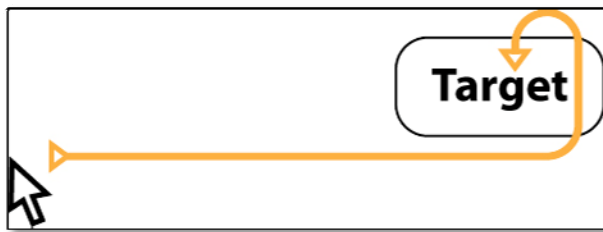


Abbildung 4.14: Von der Bewegungsdauer und Umgebung abhängige Be- und Entschleunigung des Mauszeigers.

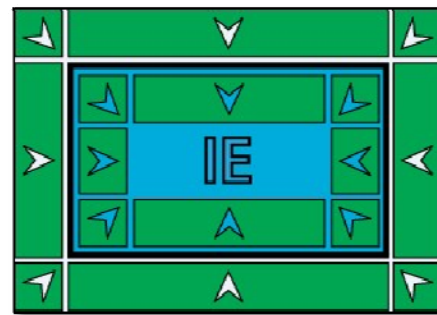


Abbildung 4.15: Dynamische Kraftfelder eines interaktiven Elements (IE).

Eine detaillierte Beschreibung der unterliegenden Konzepte des finalen Browserprototypen ist, wie bereits erwähnt, in [Specht 10] zu finden. Diese kurze Zusammenfassung sollte dazu dienen, die Gründe für die Herangehensweise der in den nachfolgenden Kapiteln beschriebenen Evaluationsstudie zu verdeutlichen und zugleich eine Grundlage für die gefundenen Ergebnisse bieten.

4.2.2. Ziele, Hypothesen und Aufgaben

Das übergeordnete Ziel dieser zweiten Evaluationsstudie war es, die Nützlichkeit der optimierten Mauszeigersteuerung zu überprüfen. Aus diesem Grund wurde eine vergleichende Studie durchgeführt, bei welcher der Unterschied zwischen der Performance zwischen einer Basisvariante und einer optimierten Variante überprüft werden sollte. Die Performance ist hierbei die Kombination von Effektivität und Effizienz bei der Ansteuerung und Auswahl interaktiver Elemente mit dem Mauszeiger. Insbesondere sollte der Einfluss der dynamischen Kraftfelder auf die Performance überprüft werden, da diese – im Gegensatz zu Variationen einer intelligenten Mauszeigerbeschleunigung – auf diese Weise noch nicht implementiert wurden. Aus diesem Grund beinhaltete die Basisvariante den Mechanismus der intelligenten Mauszeigersteuerung, wohingegen die optimierte Variante selbige mit den dynamischen Kraftfeldern kombinierte. Ein weiteres Ziel der Studie war es, die Performance beider Varianten in ihrem realen Nutzungsumfeld, dem fahrenden Fahrzeug, zu überprüfen. Darüber hinaus sollten etwaige Bedienungsprobleme aufgedeckt werden, was durch die zusätzliche Integration eines qualitativen Usability Test erreicht werden sollte.

Eine unabhängige Variable dieses Testszenarios ist demnach die Optimierung der Mauszeigersteuerung in Form der dynamischen Kraftfelder. Hier wurde zwischen „aktiviert“ und „deaktiviert“ variiert. Die zweite unabhängige Variable ist der Ort der Durchführung des Tests, bei dem Parkplatz und Stadtverkehr variiert wurden.

Basierend auf den Zielen der Studie und den Gestaltungsgrundlagen der optimierten Mauszeigersteuerung wurden zunächst einige Hypothesen aufgestellt, die mit Hilfe der

Evaluationsstudie überprüft werden sollten. Die Forschungshypothesen betrafen einerseits die Effektivität und Effizienz der optimierten Mauszeigersteuerung und andererseits die Auffälligkeit der selbigen. Die Effektivität der optimierten Mauszeigersteuerung wurde anhand der Fehler beim Anklicken interaktiver Elemente untersucht. Ihre Effizienz wurde unter Einbezug dreier Aspekte gemessen: der Zeit, die der Benutzer benötigt zum Ziel zu gelangen, der Overshoots die er dabei macht sowie der Strecke, die er auf dem Weg zum Ziel zurücklegt. Als Overshoot wird das Passieren des Zielelements definiert, ohne dass dieses dabei ausgewählt wird. Die zurückgelegte Strecke hängt in gewissem Maße mit der Anzahl an Overshoots zusammen, da bei einer hohen Anzahl an Overshoots längere Strecken zurückgelegt werden, als wenn ein Element sofort erfolgreich getroffen wird. Die vier genannten Aspekte sind die abhängigen Variablen dieser Studie, mittels denen die Performance der Mauszeigersteuerung gemessen wurde. Die Forschungshypothesen für diese Studie waren:

H₁: Ziele können mit Hilfe der optimierten Mauszeigersteuerung schneller getroffen werden, als mit der Basisvariante.

H₂: Mit Hilfe der optimierten Mauszeigersteuerung werden weniger Fehler gemacht, als mit der Basisvariante.

H₃: Mit Hilfe der optimierten Mauszeigersteuerung werden weniger Overshoots gemacht, als mit der Basisvariante.

H₄: Mit Hilfe der optimierten Mauszeigersteuerung werden kleinere Strecken zurückgelegt.

H₅: Es wird kein Unterschied zwischen der optimierten Mauszeigersteuerung und der Basisvariante wahrgenommen.

Die Effektivität der optimierten Mauszeigersteuerung sollte mit Hilfe der Hypothese H₂ überprüft werden, wohingegen die Hypothesen H₁, H₃ und H₄ Indikatoren für die Effizienz der Mauszeigersteuerung sind. Die Hypothese H₅ betrifft die Auffälligkeit der optimierten Mauszeigersteuerung, die wie bereits in Kapitel 4.2.1. erwähnt wurde, möglichst gering sein sollte.

Um die Effektivität und Effizienz zweier Eingabemöglichkeiten zu vergleichen, ist die Integration eines Tapping-Tests ein gebräuchliches Hilfsmittel. Im Kontext von Eingabegeräten für den Desktop-PC empfiehlt die [ISO 9241-9] einen ein- oder multidirektionalen Tapping-Test. Beim eindimensionalen Tapping-Test muss der Proband abwechselnd zwei Rechtecke mit einer Breite w

anklicken, die sich in einem Abstand d zueinander befinden (vgl. Abbildung 4.16). Dieser Test sollte mit unterschiedlichen Schwierigkeiten durchgeführt werden, was durch die Variation von w und d gewährleistet wird. Der multidirektionale Tapping-Test besteht aus einer kreisförmig angeordneten Menge an nummerierten Elementen der Breite w . Die Elemente sind dabei so nummeriert, dass sich das jeweils nachfolgende Element in etwa dem gleichen Abstand zum aktuellen Element befindet, wie das vorige (vgl. Abbildung 4.17). Der Proband muss die Elemente des Kreises bei einem solchen Test der Reihe nach auswählen, wobei das jeweilige Zielelement beispielsweise farblich hervorgehoben werden muss. Auch hier kann die Schwierigkeit von Testdurchlauf zu Testdurchlauf verändert werden, indem der Kreisdurchmesser und dadurch die Distanz zwischen den einzelnen Elementen verkleinert oder vergrößert wird [ISO 9241-9].

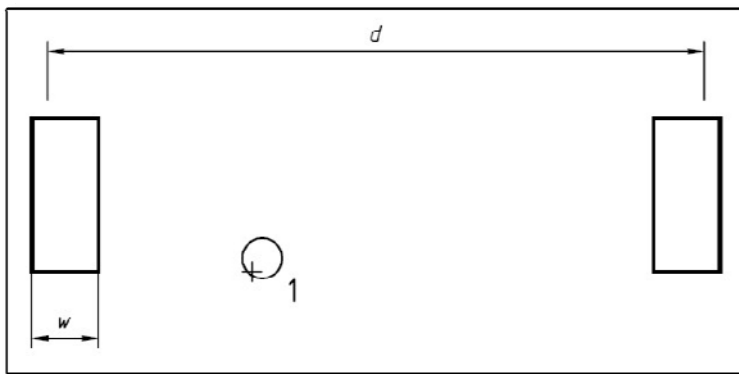


Abbildung 4.16: Eindirektionaler Tapping-Test (Quelle: [ISO 9241-9]).

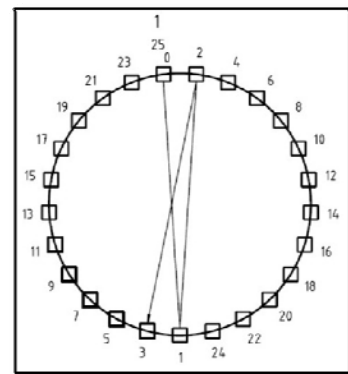


Abbildung 4.17: Multidirektionaler Tapping-Test (Quelle: [ISO 9241-9]).

Ein solcher Tapping-Test war jedoch bei der durchgeführten Evaluationsstudie nicht praktikabel, da er kein reales Setting für den vorliegenden Anwendungskontext darstellte und dadurch nur eine geringe ökologische Validität hätte gewährleisten können. Die Verwendung einer realen Webseite hätte ein weniger künstliches Setting geschaffen, doch wurde hierauf verzichtet, um den Fokus der Probanden auf der Aufgabe zu belassen. Das Vorhandensein von Webseiteninhalten hätte für die Probanden eine zusätzliche Herausforderung dargestellt, was die Effizienz der Bedienung beeinträchtigt hätte und so beim Vergleich der Interaktionskonzepte zu weniger zuverlässigen Ergebnissen geführt hätten. Darüber hinaus sind in einer einzelnen Webseite nicht alle Elemente vorhanden, mit welchen die Befragten der Fokusgruppen im Laufe einer typischen Browsing-Sitzung interagieren (vgl. [Söter 08]).

Um dennoch einen zum Anwendungskontext passenden Tapping-Test durchführen zu können, hat die Autorin dieser Arbeit eigens eine Webseite konstruiert, die aus den wichtigsten Elementen einer typischen Browsing-Sitzung bestand (vgl. Abbildung 4.18 und 4.19). Da der Tapping-Test auf diese Weise dennoch auf Elementen realer Webseiten durchgeführt werden konnte, hatte dies den Vorteil, dass die ökologische Validität des Tests gewährleistet war.

Da die Befragten der Fokusgruppen angaben, sich während der Autofahrt über aktuelle Geschehnisse informieren zu wollen, wurden horizontal und vertikal angeordnete Elemente mit einer Höhe von 14 bzw. 18 Pixeln (rot und orange in Abbildung 4.19) integriert, wie sie bei bekannten deutschen Nachrichtenwebseiten in der Navigationsleiste Verwendung finden. Große sowie lange, schmale Elemente (grün in Abbildung 4.19) repräsentieren Bilder und Links, wie sie bei einer Bildersuche vorhanden sein könnten. Da die Befragten zudem angaben, sich während der Fahrt mit Hilfe von Videoclips unterhalten zu wollen, wurden kleine, fast quadratische Elemente (blau in Abbildung 4.19) in die Webseite für den Tapping-Test eingebaut, welche Schaltflächen repräsentieren, wie sie bei Videoapplikationen auf Web2.0 Webseiten (hier: YouTube) zu finden sind.

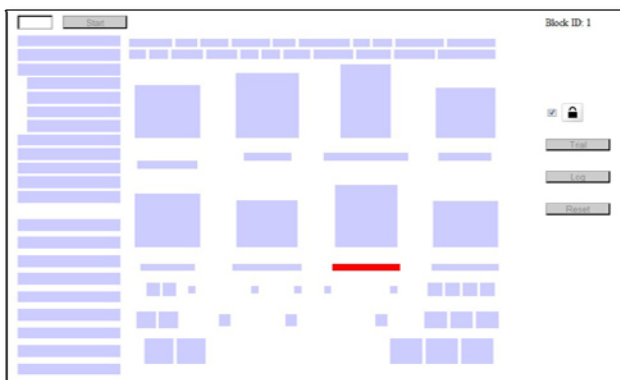


Abbildung 4.18: Webseite für den Tapping-Test.

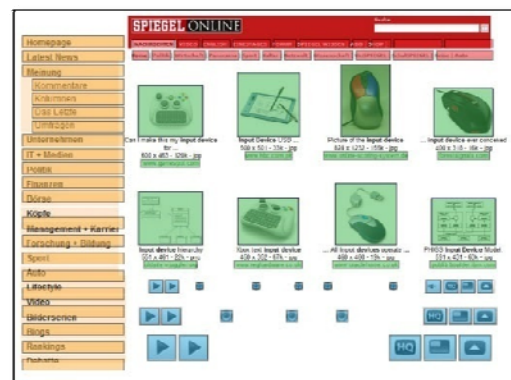


Abbildung 4.19: Herkunft der Elemente der Webseite für den Tapping-Test.

Beim Tapping-Test sollte der Proband nun ein rot aufleuchtendes Element (vgl. Abbildung 4.18) möglichst schnell und möglichst genau treffen. Bei einem erfolgreichen Klick wurde das nächste Element rot markiert. Beim stationären Test wurden dabei fünf Blöcke durchlaufen, von denen der erste und der letzte aus jeweils 20 Trials bestanden. Die Blöcke zwei, drei und vier bestanden aus jeweils 40 Trials. Zudem wurden gerade sowie schräge Strecken getrennt behandelt: die Reihenfolge in welcher die Elemente im ersten und letzten Block angeklickt werden mussten, wurde so gewählt, dass auch schräge Strecken zurückgelegt werden mussten. Dies erforderte von den Probanden, die 8-Wege-Steuerung des MMI-Terminals (Nord, NO, Ost, SO, Süd, SW, West, NW) einzusetzen, um das jeweilige Ziel schneller zu erreichen. Die Reihenfolge der Elemente in den mittleren Blöcken war hingegen so konzipiert, dass die Strecke zum jeweils nächsten Zielelement durch eine gerade Bewegung des MMI-Terminals (N, O, S, W) zurückgelegt werden konnte. Die Unterscheidung wurde hier gemacht, um zu überprüfen, ob das Lernen anhand von einfachen Aufgaben (hier: 4-Wege-Steuerung) einen Einfluss auf die Performance bei komplexeren Aufgaben (hier: 8-Wege-Steuerung) hat. Diese Fragestellung ist jedoch nicht Teil dieser Arbeit und wird an anderer Stelle diskutiert.

Beim mobilen Test wurden zwei Blöcke à 40 Trials durchlaufen, wobei hier gerade und schräge Strecken in gleicher Verteilung zurückgelegt werden mussten. Im Hintergrund der Tapping-Test Webseite wurden je Trial unter anderem die Zeit und die Strecke bis zum ersten Klick geloggt. Die Fehlerrate sowie die Anzahl der Overshoots pro Trial wurden ebenfalls aufgezeichnet. Um bei der Auswertung Vergleiche ziehen zu können, wurden die jeweiligen durchschnittlichen Werte pro Trial normalisiert.

Im Rahmen des integrierten qualitativen Usability Tests zur Erfassung der Bedienungsprobleme wurden für den stationären Test jeweils zwölf, für den mobilen Test jeweils elf Aufgaben vorbereitet. Diese entsprachen den in den Fokusgruppen geäußerten Aufgaben und Zielen. Beim stationären Test dienten sechs dieser Aufgaben der Orientierung auf Webseiten und der Wahrnehmung von Informationen. Beim mobilen Test waren es fünf Aufgaben, die sich mit dieser Thematik auseinandersetzten. Der Kernaspekt der Orientierungs- und Wahrnehmungsaufgaben waren Scrollen und Zoomen. Beispielsweise sollte zu bestimmten Artikeln auf einer Nachrichtenwebseite navigiert werden oder es sollten Texte gelesen und Fragen dazu beantwortet werden. Um Problemen mit der Verfügbarkeit des Internets aus dem Weg zu gehen, wurden die Webseiten gespeichert und aus dem Cache abgerufen. Auf diese Weise konnte auch sichergestellt werden, dass die Webseiten sich nicht während des Tests veränderten, wie das bei Nachrichtenwebseiten mehrmals täglich der Fall ist. Somit wurde die Vergleichbarkeit der Ergebnisse gewährleistet.

Jeweils zwei Aufgaben pro Aufgabenset behandelten sowohl in stationären, als auch im mobilen Test die Navigation. Hierbei ging es darum, einen speziellen Link anzuklicken. Drei weitere Aufgaben zielten jeweils auf das allgemeinen Darstellungs- und Bedienkonzept des Browsers ab. Beispielsweise sollte hier ein Softkey betätigt oder eine andere geöffnete Webseite besucht werden.

Während der Bearbeitung der Aufgaben sollten die Probanden die Methode des lauten Denkens anwenden. Der Versuchsleiter notierte während der Beobachtung zusätzliche Auffälligkeiten. Interviews im Anschluss des Tests sollten schließlich dabei helfen Problemstellen zu identifizieren und zu erörtern sowie die weiteren Forschungsfragen zu beantworten:

- F₁: Wird die optimierte Mauszeigersteuerung vom Benutzer wahrgenommen?
- F₂: Wie gefällt die optimierte Mauszeigersteuerung dem Benutzer?
- F₃: Wie wird das Scrollen mittels Drehrad empfunden und wann wird es angewandt?
- F₄: Wird Zoomen beim Lesen bevorzugt?
- F₅: Welche weiteren Bedienungsprobleme tauchen auf?

4.2.3. Testdesign und Auswahl der Probanden

Um das neue Interaktionskonzept auf seine Gebrauchstauglichkeit in realer Umgebung überprüfen zu können, wurden sowohl Tests im Stand, als auch während der Fahrt gemacht. Der mobile Testablauf war hierbei ähnlich wie im Stand (vgl. Abbildung 4.20). Es wurden lediglich die Aufgaben der jeweiligen Teilbereiche leicht abgewandelt, damit beim mobilen Test eine kürzere Dauer ermöglicht werden konnte.

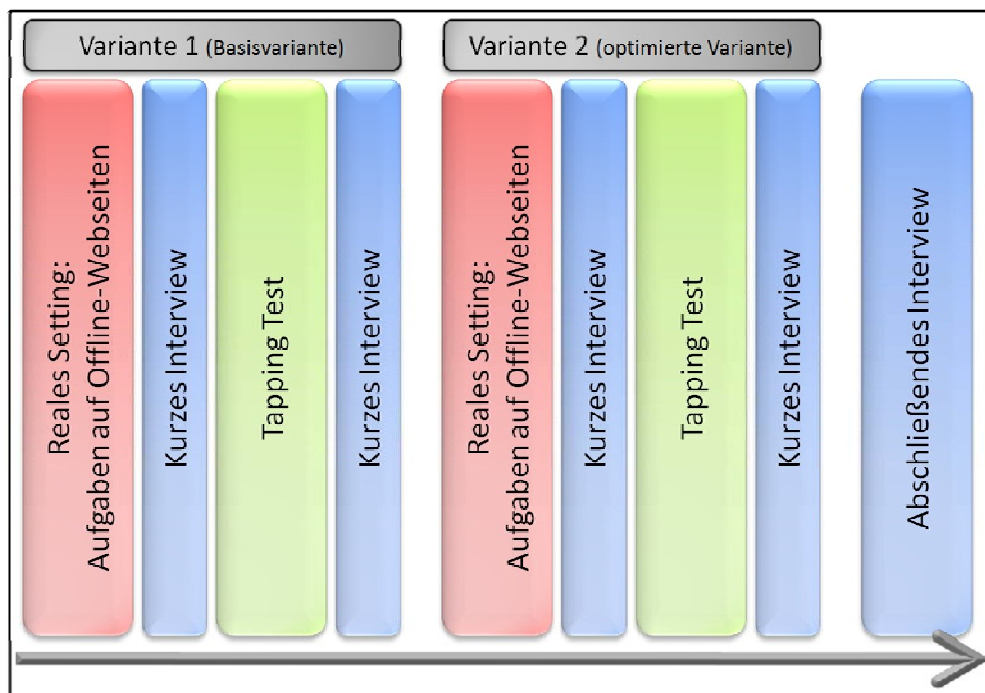


Abbildung 4.20: Testablauf bei der Evaluationsstudie im Juli / August 2009.

Die Reihenfolge, in welcher die optimierte und die Basisvariante getestet wurden, war ausbalanciert um die Beeinflussung der Performance durch Lerneffekte zu minimieren. Um ferner überprüfen zu können, ob die Wirkung der dynamischen Kraftfelder wahrgenommen wird, wurde den Probanden im Voraus nicht erzählt, dass zwei Varianten getestet wurden.

Zunächst wurden beispielsweise mit der Basisvariante Aufgaben auf realen Webseiten bearbeitet. Diese waren so konzipiert, dass der Proband nach der vierten Aufgabe bereits hätte zoomen und scrollen können und auch Softkeys hätte benutzen können. Wurden bis zu diesem Zeitpunkt nicht alle Funktionen zumindest ausprobiert, wurden diese vom Untersuchungsleiter erklärt.

Mit einem kurzen Interview nach Abschluss dieses Teilbereichs sollte herausgefunden werden, was bis dahin positiv oder negativ aufgefallen war. Anschließend sollte der Proband zur Hauptansicht des Browsers zurückkehren, die Webseite für den Tapping-Test aufrufen und diesen absolvieren. Nach Beendigung des letzten Blocks im Tapping-Test wurde erneut ein kurzes Interview geführt, in welchem der Proband erneut sagen sollte, was ihm bis dahin positiv oder negativ aufgefallen war.

Während er diese Fragen beantwortete, speicherte der Versuchsleiter die Logdatei des Probanden, schloss den Browser und startete die optimierte Variante. Keinem der Probanden ist hier aufgefallen, dass eine andere Variante des Browsers gestartet wurde, da bei Betreten des Fahrzeugs die zuerst zu testende Variante bereits gestartet war. Nach dem kurzen Interview wurden mit der optimierten Variante ebenfalls die vier Schritte Realwebseiten – Interview – Tapping-Test – Interview durchlaufen und abschließend wurde ein zusammenfassendes Interview durchgeführt (vgl. Abbildung 4.20).

Beim abschließenden Interview sollte der Proband nochmals zusammenfassen, welche Aspekte des Browsers er als besonders positiv und welche er als negativ wahrgenommen hat. Anschließend wurde ihm ein Fragebogen überreicht, in welchem er gefragt wurde, ob er eine Veränderung im Laufe des Tests feststellen konnte und welcher Art diese Veränderung war. Für den Fall, dass er eine Veränderung wahrgenommen hatte, wurden ihm einige Antwortmöglichkeiten zur Auswahl gegeben. Diese waren: „Ich habe mich verbessert.“, „Ich habe mich verschlechtert.“, „Die Bedienung ist anstrengender geworden.“, „Die Bedienung ist weniger anstrengend geworden.“, „Ich konnte die Links besser treffen.“, „Ich konnte die Links weniger gut treffen.“ und „Sonstiges“, wobei er letzteres noch genau spezifizieren konnte. Jeder Proband wurde darauf hingewiesen, dass er so viele Optionen ankreuzen könne, wie für ihn zutreffen.

Im Falle einer wahrgenommenen Veränderung wurde der Proband nach den Gründen hierfür befragt. Dabei wurden ihm je nach seinen Antworten im Fragebogen mögliche Gründe wie Ermüdungserscheinungen, verstärkte beziehungsweise nachlassende Konzentration oder ein Lerneffekt vorgeschlagen. Sollte bis zu diesem Zeitpunkt die veränderte Mauszeigersteuerung nicht erwähnt worden sein, wurde der Proband vom Untersuchungsleiter direkt gefragt, ob ihm eine Veränderung des Mauszeigerverhaltens aufgefallen sei. Anschließend wurde sie ihm konkret auf der Webseite des Tapping-Tests vorgeführt und ihm wurde angeboten diese Veränderung selbst noch einmal auszuprobieren. Im Anschluss wurde er gefragt, wie ihm die optimierte Mauszeigersteuerung gefällt.

Im Rahmen des stationären Teils wurden Beobachtungen und Kernaussagen des Probanden handschriftlich notiert. Zusätzlich wurden alle Tests mittels Videokamera aufgezeichnet, um nach den Tests eine detailliertere Auswertung durchführen zu können. Dies war insbesondere für die Erfassung der Daten im mobilen Test notwendig, da hier aufgrund der Erschütterungen durch die Fahrt kein handschriftliches Protokoll verfasst wurde. Die Bildschirminhalte wurden mit Hilfe der Aufzeichnungssoftware TechSmith MORAE® erfasst.

Der mobile Test fand durchschnittlich 7,125 Tage nach dem stationären Test statt. Der Testablauf war hierbei der selbe wie beim stationären Test (vgl. Abbildung 4.20). Allerdings wurde vor der Losfahrt einmal der Tapping-Test mit der Basisvariante im Stand durchgeführt. Mit Hilfe der hierbei geloggen Daten sollte später festgestellt werden können, ob die Performance persistent war oder ob durch die

einwöchige Pause ein Absacken der Performance festzustellen war. Damit der Vergleich der Performance-Daten der Fahrt mit denen des Standes legitim war, durfte es keine signifikante Verschlechterung der Performance nach der etwa einwöchigen Pause geben. Verglichen wurden hierbei die Daten dieses „Baseline“-Tests mit denen der stationären Basisvariante.

Die Strecke, auf der die mobilen Tests durchgeführt werden sollten, führte zunächst über eine Landstraße ins Konstanzer Industriegebiet und anschließend in die Innenstadt (vgl. Abbildung 4.21). Insbesondere der Abschnitt in der Innenstadt ist typischerweise stark befahren und enthält viele Ampeln, sodass vom Fahrer häufiges Beschleunigen und Bremsen erfordert wird. Die Strecke wurde dennoch ausgewählt, um durch die störenden Einflüsse eine typische Benutzungssituation zu modellieren und um so eine höhere ökologische Validität der Untersuchung zu gewährleisten. Da die mobilen Tests jedoch zu einem Zeitpunkt stattfanden, in welchem Schulferien in Baden-Württemberg sind, waren die Straßen in der Innenstadt stärker befahren als sonst, was zu sehr starken Verkehrsstauungen in einem Abschnitt führte. Dennoch wurde die Strecke bis auf einen Test beibehalten, um die interindividuelle Vergleichbarkeit sicherzustellen. Da die Alternativstrecke dieses einen Tests jedoch in gleichen Maße Stauungen enthielt, waren auch hier die Testbedingungen ähnlich und dadurch ein Vergleich mit den anderen Probanden legitim.

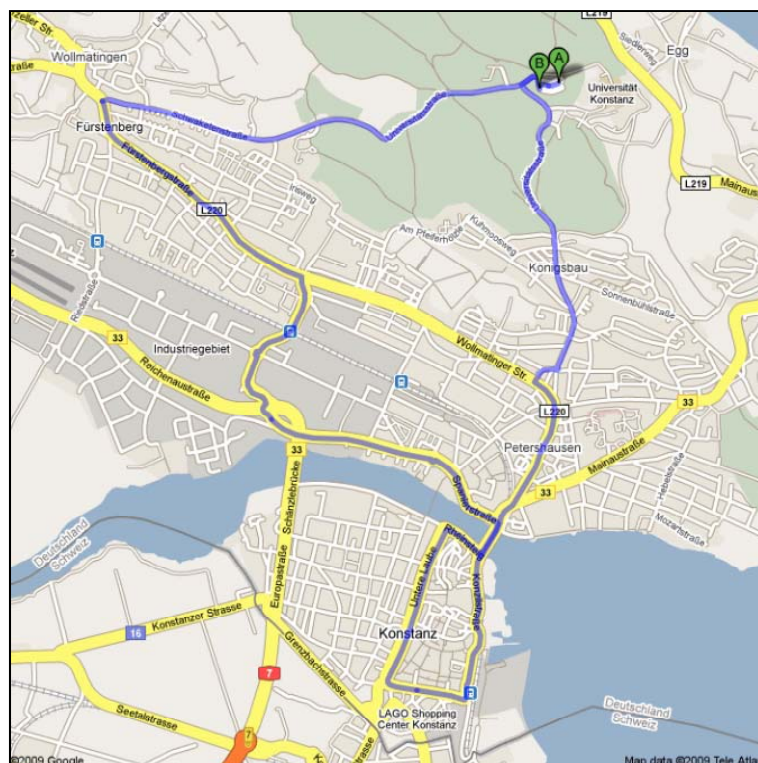


Abbildung 4.21: Versuchsstrecke für die mobilen Tests im August 2009.

Im Gegensatz zur ersten Evaluationsstudie wurden bei der Auswahl der Probanden andere Kriterien berücksichtigt. Da der Kernaspekt dieser Untersuchung der experimentelle Vergleich zwischen der optimierten und der Basisvariante war und da aus zeitlichen und finanziellen Gründen nur eine kleine Probandengruppe zur Verfügung stand, sollte diese möglichst homogen sein. Nur so können zuverlässige interindividuelle Vergleiche gezogen werden. Zudem sollten bei der Evaluation die Interaktionskonzepte und nicht die Handhabung des Eingabegerätes untersucht werden. Aus diesem Grund wurde die Probandengruppe aus acht technikaffinen Schülern im Alter von 16 bis 17 Jahren zusammengestellt. Sechs der Probanden waren männlich und spielten in ihrer Freizeit Computerspiele, sodass eine gewisse Affinität in Bezug auf neue beziehungsweise Joystick-ähnliche Eingabegeräte gewährleistet war. Die anderen beiden Probanden waren weiblich und spielten zwar keine Computerspiele, besaßen jedoch einen eigenen Computer und nutzten diesen und das Internet täglich. Zwei der Probanden hatten bereits bei der Evaluationsstudie im Januar 2008 teilgenommen.

Ein zusätzlicher Aspekt, der für die Auswahl dieser Probanden sprach, war die Tatsache dass der Einsatzort des zu testenden Webbrowsers der Fondbereich eines Audi Q7 war. Die Personen, die ihn hier benutzen würden, wären aller Wahrscheinlichkeit nach die Kinder des Käufers.

Da diese Probanden jedoch noch minderjährig waren und das Testdesign auch mobile Tests enthielt, die eine potentielle Gefahrenquelle darstellen könnten, wurde für den zweiten Teil das Einverständnis der Eltern gefordert.

Wie in Kapitel 4.2.2. erwähnt sollte auch ein qualitativer Usability Test integriert werden. Da Personen mit anderen Charakteristiken als die Schüler möglicherweise auf andere Problemstellen stoßen würden, wurden für den mobilen Test zusätzlich drei berufstätige Probanden eingeladen. Zwei der Probanden waren männlich, ein Proband war weiblich. Zwei dieser Probanden hatten ebenfalls bereits bei der Evaluationsstudie im Januar 2009 teilgenommen.

4.2.4. Aufbereitung der Daten, Auswertung und Ergebnisse

Zur Ermittlung der Effizienz der optimierten Mauszeigersteuerung und zur Berechnung des Unterschieds im Vergleich zur Basisvariante, wurden die Logdateien der Schüler in *Microsoft Excel* überführt. Hier wurden zunächst grundlegende Berechnungen durchgeführt. Signifikanztests zur Überprüfung der statistischen Hypothesen wurde mit Hilfe der Software *PASW Statistics 17* (früher: *SPSS Statistics*) durchgeführt. Um den Einfluss durch Lerneffekte zu minimieren, wurden bei der Auswertung der stationären Tests lediglich die letzten beiden Blöcke gewertet, bei den mobilen Tests jeweils der letzte Block.

Ein wichtiger Aspekt der Studie ist der Vergleich der Performance zwischen Basis- und optimierter Variante. Zunächst wurden hierfür die Ergebnisse des stationären Tests miteinander verglichen. Ein statistischer Signifikanztest sollte überprüfen, ob es einen Unterschied zwischen beiden Varianten

gibt, in welche Richtung dieser Unterschied geht und wie groß er ist. Zudem sollte überprüft werden, ob dieser Unterschied statistisch signifikant ist oder ob das gefundene Ergebnis zufällig zustande gekommen ist.

Beim statistischen Signifikanztest wird eine Nullhypothese aufgestellt, deren Richtigkeit überprüft wird. In der Regel geht diese Nullhypothese davon aus, dass es keinen Unterschied zwischen zwei zu vergleichenden Parametern gibt, wie sie hier in Form der beiden Interaktionskonzepte „optimierte Variante“ und „Basisvariante“ vorliegen. Die Forschungshypothese hingegen besagt, dass zwischen diesen Parametern ein unterscheidendes Merkmal existiert [Hussy & Jain 02]. Im Rahmen dieser Untersuchung lautet die Nullhypothese also:

H_0 : Die Basisvariante und die optimierte Variante unterscheiden sich nicht hinsichtlich ihrer Performance.

Diese wird im Folgenden in die Nullhypothesen H_{0a} und H_{0c} aufgespalten. Während sich H_{0a} mit den Unterschieden im stationären Test befasst, stellt H_{0c} die Nullhypothese für den mobilen Teil des Experiments dar. H_{0a} lautet in diesem Fall:

H_{0a} : Die Basisvariante und die optimierte Variante unterscheiden sich bei der stationären Bedienung nicht hinsichtlich ihrer Performance.

Um die Nullhypothese ablehnen und die Hypothesen aus Kapitel 4.2.2. annehmen zu können, muss ein signifikanter Unterschied zwischen den beiden Varianten gegeben sein. Das Signifikanzniveau wurde im Rahmen dieser Studie mit 5% spezifiziert.

Zusätzlich zum Signifikanztest wurde die Effektgröße berechnet. Diese gibt an, wie groß und somit auch wie bedeutungsvoll ein Unterschied oder ein Zusammenhang zwischen zwei Faktoren ist. Da die Effektgröße im Gegensatz zur statistischen Signifikanz unabhängig von der Stichprobengröße ist, kann sie auch als Maß für den Vergleich verschiedener Untersuchungen genutzt werden [Hussy & Jain 02]. Im Rahmen dieser Studie wurde die Effektgröße anhand des partiellen Eta-Quadrats (η_p^2) berechnet. Dieses ergibt sich aus dem Quotienten aus der Faktorquadratsumme des Effekts zur Summe dieses Wertes und der Fehlerquadratsumme [Rudolf & Kuhlich 08]:

$$\eta_p^2 = \frac{\text{Faktorquadratsumme des Effekts}}{\text{Faktorquadratsumme des Effekts} + \text{Fehlerquadratsumme}} = \frac{\text{QS (Effekt)}}{\text{QS (Effekt)} + \text{QS (Fehler)}}$$

Eine Klassifikation der Effektgröße wurde 1988 von Cohen aufgestellt. Demnach ist bis zu einem Wert von η^2 bis 0,1 von einem kleinen Effekt die Rede. Bis 0,6 kann man von einem mittleren Wert sprechen. Ab 0,14 kann der Effekt als groß bezeichnet werden [Hussy & Jain 02].

Bei der Auswertung der Daten zeigte sich, dass die Probanden mit der optimierten Variante durchschnittlich schneller waren, weniger Fehler und Overshoots machten und dabei eine kürzere Strecke zurücklegten (vgl. Tabelle 4.01). Es zeigte sich auch, dass die Unterschiede bezüglich der benötigten Zeit und der getätigten Overshoots signifikant waren. Dass die Unterschiede in den anderen Dimensionen nicht signifikant wurden, könnte die geringe Stichprobe zur Ursache haben.

<i>Stationärer Test</i>	Basisvariante	Optimierte Variante	Statistische Signifikanz
Ø Zeit / Trial	3,10 sec s = 0,50 sec	2,89 sec s = 0,45 sec	F = 7,746 p = 0,027
Ø Fehler / Trial	0,019 s = 0,012	0,006 s = 0,009	F = 3,500 p = 0,104
Ø Overshoots / Trial	0,439 s = 0,169	0,207 s = 0,103	F = 22,443 p = 0,002
Ø Strecke / Trial	317,15 Pixel s = 10,76 Pixel	316,0 Pixel s = 15,02 Pixel	F = 0,094 p = 0,769

Tabelle 4.01: Ergebnisse des Signifikanztests für den stationären Tapping-Test mit der Basis- und der optimierten Variante.

Auffällig war die geringe Fehlerrate, die im Laufe des Tests allgemein sehr niedrig war. Die durchschnittliche Anzahl an Fehlern pro Trial im vierten und fünften Block des stationären Tests betrug bei der Basisvariante 1,9%. Bei Interaktion mit der optimierten Variante lag dieser durchschnittliche Fehler 0,6%. Zum Vergleich lag die durchschnittliche Fehlerrate aller Blöcke bei der Basisvariante im stationären Test bei 2,5% (maximal 4,5%), bei der optimierten Variante bei 0,9% (maximal 1,5%). Zusätzlich war zum Teil eine relativ hohe Streuung erkennbar, sodass in Bezug auf die Fehlerrate keine signifikanten Unterschiede erkannt werden können.

Die verhältnismäßig geringe Streuung in Bezug auf Zeit und Overshoots (vgl. Abbildung 4.22 und 4.23) sowie der große Unterschied zwischen den Mittelwerten dieser Dimensionen spiegeln sich auch in der Effektgröße wieder. Insbesondere in Bezug auf die Overshoots pro Trial ist ein großer Effekt sichtbar, was sich auch im berechneten partiellen Eta-Quadrat niederschlägt. Dieser hat hier den Wert $\eta_p^2 = 0,762$. Auch der signifikante Unterschied in Bezug auf die Zeit pro Trial weist auf einen großen Effekt hin. Das η_p^2 hat hier den Wert 0,525.

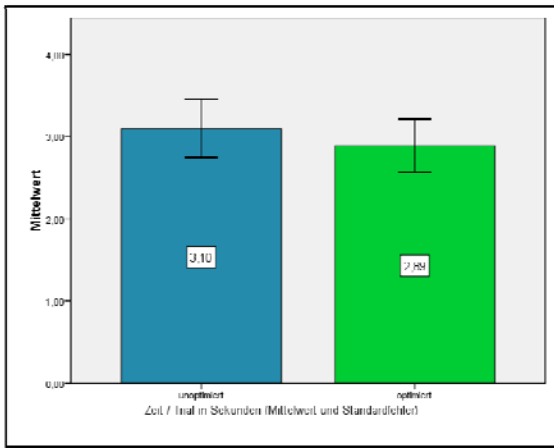


Abbildung 4.22: Mittelwerte und Streuung der benötigten Zeit pro Trial im stationären Test.

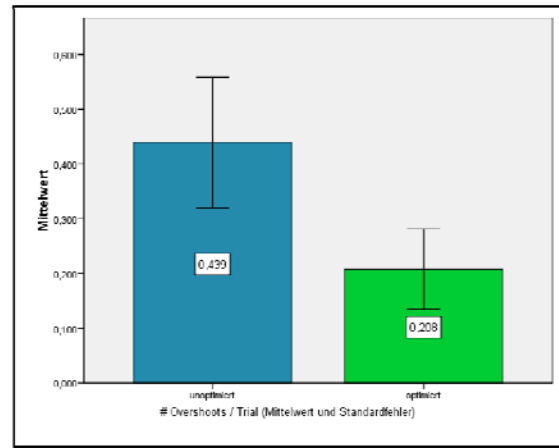


Abbildung 4.23: Mittelwerte und Streuung der vollführten Overshoots pro Trial im stationären Test.

Aufgrund der signifikanten und großen Unterschiede in Bezug auf Zeit und Overshoots wird die H_{0a} abgelehnt. Die Verbesserung der Performance mit der optimierten Variante konnte zumindest in zwei der vier Dimensionen als statistisch signifikant nachgewiesen werden.

Ein anderer wichtiger Aspekt der Untersuchung betraf den Unterschied in der Performance der Basisvariante in der Fahrt im Vergleich zum Stand. Aufgrund störender externer Einflüsse während der Fahrt, kann davon ausgegangen werden, dass ein schnelles und korrektes Treffen von Links hier schwieriger zu vollführen ist. Die Ergebnisse der ersten Evaluationsstudie, in welcher die Probanden die Mauszeigerpositionierung während der Fahrt subjektiv als schwieriger empfanden, bestätigen diese Theorie. Daher sollte ein statistischer Signifikanztest durchgeführt werden, der die Performance der stationären Bedienung mit der Performance während der Fahrt vergleicht und folgende Nullhypothese überprüfen sollte:

H_{0b} : Die Performance der Basisvariante im Stand unterscheidet sich nicht von ihrer Performance während der Fahrt.

Um diesen Vergleich jedoch ziehen zu können musste in einem vorausgehenden Schritt überprüft werden, ob die Performance auch nach einer Pause von durchschnittlich 7,125 Tagen persistent war. Hierfür mussten die Logdaten des stationären vierten und fünften Blocks mit der Basisvariante mit den Ergebnissen des Baseline-Tests verglichen werden.

Da jedoch in den jeweiligen Tests unterschiedliche Distanzen zurückgelegt wurden, mussten die Daten zunächst normalisiert werden. Eine erste Normalisierung betraf die Blöcke des stationären Tests, da hier im vierten Block 40 Trials mit geraden Strecken zurückgelegt wurden, während im fünften Block 20 Trials mit schrägen Strecken zu absolvieren waren. Bei der Normalisierung wurde daher zunächst die durchschnittliche ideale Distanz des vierten stationären Blocks mit der des fünften stationären Blocks verrechnet. Die ideale Distanz entspricht hierbei der Distanz zwischen dem Mittelpunkt des Ausgangselementes zum Mittelpunkt des Zielelementes (vgl. Abbildung 4.24).

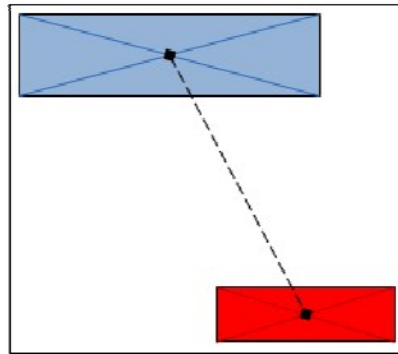


Abbildung 4.24: Ideale Distanz zwischen Start- und Zielelement eines Trials.

Während bei Block 4 des stationären Tests die durchschnittliche Distanz zwischen zwei Elementen 290,325 Pixel betrug, waren es bei Block 5 durchschnittlich 282,15 Pixel. Die Zeit, die durchschnittlich für einen Trial in Block 5 benötigt wurde, wurde also mit dem Faktor $\frac{290,325}{282,15} \approx 1,03$ multipliziert. Auf diese Weise wurde die Zeit berechnet, die der Proband vermutlich für einen Trial des fünften Blocks benötigt hätte, um eine ebenso lange Strecke, wie im vierten Block zurückzulegen. Mit den Werten für Strecke sowie Overshoots wurde ebenso verfahren (vgl. Tabelle 4.02). Die Fehlerrate wurde nicht mit diesem Faktor multipliziert, da dieser Wert weniger von der Strecke, als mehr vom Zielelement abhängt.

	Block 4	Block 5	Block 5 normalisiert	Normalisierter Ø Block 4 & 5
Ideale Distanz	290,325	282,15	1,03 x 282,15 = 290,325	$\frac{290,325 + 290,325}{2} = 290,325$
Distanz	300,425	343,2	1,03 x 343,2 = 353,144	$\frac{300,425 + 353,144}{2} = 326,784$
Zeit	2708,35	3114,1	1,03 x 3114,1 = 3204,328	$\frac{2708,35 + 3204,328}{2} = 2956,339$
Overshoots	0,275	0,35	1,03 x 0,35 = 0,360	$\frac{0,275 + 0,360}{2} = 0,318$
Fehler	0,25	0	0 (Keine Normalisierung)	$\frac{0,25 + 0}{2} = 0,125$

Tabelle 4.02: Beispielrechnung bei der Normalisierung von Block 4 und 5 des stationären Tests. (Basisvariante bei Proband #1, jeweils durchschnittliche Werte pro Trial)

Um die Ergebnisse des stationären Tests mit jenen der Baseline beziehungsweise mit jenen des mobilen Tests zu vergleichen, musste hier ebenfalls normalisiert werden. Da die Hälfte der Strecken im mobilen Test gerade und die andere Hälfte schräg waren, mussten zunächst bezüglich des stationären Tests gleiche Bedingungen geschaffen werden. Die normalisierte durchschnittliche Strecke von Block 4 und 5 des stationären Tests betrug 290,325. Im mobilen Test war die durchschnittliche Strecke pro Trial mit 307,05 Pixeln länger. Daher wurden die Werte für Zeit/Trial, Strecke/Trial und Overshoots/Trial mit dem Faktor $\frac{290,325}{307,05} \approx 0,95$ multipliziert, um eine Basis für den Vergleich zu generieren (vgl. Tabelle 4.03). Die normalisierten Werte des Baseline-Blocks beziehungsweise des zweiten mobilen Blocks konnten so mit dem Durchschnitt der normalisierten Werte des vierten und fünften stationären Blocks verglichen werden.

	Normalisierter Ø Block 4 & 5 (stationär)	Baseline	Normalisierte Baseline
Ideale Distanz	290,325	307,05	0,95 x 307,05 = 290,325
Zeit pro Trial	2956,339	2776,5	0,95 x 2776,5 = 2577,113

Tabelle 4.03: Beispielrechnung bei der Normalisierung der Zeit für den Vergleich der letzten beiden stationären Blöcke mit dem Baseline-Block (Daten von Proband #1).

Der Signifikanztest unter Verwendung der normalisierten Daten zeigte, dass trotz der einwöchigen Pause kein Absacken der Performance vorhanden war. Tatsächlich waren die Probanden beim Baseline-Test signifikant schneller (Baseline-Test: 2,87 sec. Stationär: 3,14 sec) ($F = 9,443$; $p = 0,018$) und machten signifikant weniger Overshoots (Baseline-Test: 0,278. Stationär: 0,445) ($F = 13,707$; $p = 0,008$) pro Trial. Sie machten zudem weniger Fehler und legten eine kürzere Strecke zurück, wobei sich die Unterschiede in diesen Dimensionen nicht als statistisch signifikant nachgewiesen werden können. Die Verbesserung der Probanden trotz der Pause ist ungewöhnlich und lässt sich anhand der gegebenen Datenbasis nicht erklären.

Da die Performance, wie beschrieben, nicht eingebrochen ist, war ein Vergleich des stationären Tests mit dem mobilen Test legitim. Auch hier wurden die Werte wie oben beschrieben normalisiert, damit ein Vergleich möglich war. Dieser wurde mit Hilfe eines statistischen Signifikanztest mittels *PASW Statistics 17* durchgeführt.

Abgesehen von der Zeit, die die Probanden durchschnittlich pro Trial brauchten, ist eine Verschlechterung der Performance im mobilen Test zu erkennen. Zum Teil sind jedoch die Unterschiede sehr klein oder die Streuung sehr groß, sodass die Signifikanz nicht nachgewiesen werden kann.

Basisvariante	Stationärer Test	Mobiler Test	Statistische Signifikanz
Ø Zeit / Trial	3,14 sec s = 0,51 sec	3,13 sec s = 0,58 sec	F = 0,008 p = 0,932
Ø Fehler / Trial	0,019 s = 0,012	0,050 s = 0,063	F = 2,108 p = 0,190
Ø Overshoots / Trial	0,445 s = 0,173	0,485 s = 0,269	F = 0,167 p = 0,695
Ø Strecke / Trial	321,95 Pixel s = 11,02 Pixel	331,44 Pixel s = 19,82 Pixel	F = 2,547 p = 0,155

Tabelle 4.04: Ergebnisse des Signifikanztests für den Tapping-Test mit der Basisvariante im Stand und während der Fahrt.

Ursache für die nicht signifikanten Ergebnisse könnte einerseits die geringe Stichprobengröße sein. Andererseits kann auch hier ein Effekt der Retention (= Aufrechterhaltung), der bereits beim Vergleich des stationären Tests mit dem Baseline-Test aufgefallen ist, eine Auswirkung auf die insgesamt eher geringe Verschlechterung gehabt haben. Die Vermutung, dass die Fahrt zu einer Verminderung der Performance bei Interaktion mit der Basisvariante führt, kann statistisch nicht nachgewiesen werden. Die H_{0b} kann aus diesem Grunde nicht zurückgewiesen werden und wird daher beibehalten.

Weitere Berechnungen wurden mittels *PASW Statistics 17* durchgeführt, um den Performance-Unterschied im mobilen Test zu untersuchen. Da im stationären Test ein signifikanter Unterschied in Bezug auf Zeit und Overshootrate erkennbar war, stellt sich nun die Frage, ob dieser Effekt auch im mobilen Test zu finden ist. Die Nullhypothese H_{0c} , die an dieser Stelle zu untersuchen ist, konstatiert dass während der Fahrt kein Performance-Unterschied zwischen der Basisvariante und der optimierten Variante zu erkennen ist.

H_{0c} : Die Basisvariante und die optimierte Variante unterscheiden sich bei der mobilen Interaktion nicht hinsichtlich ihrer Performance.

Für den Vergleich der vier Dimensionen Zeit, Fehler, Overshoots und Strecke war hier – wie beim Vergleich im stationären Rahmen – keine Normalisierung der Werte nötig. Die Ergebnisse der Berechnungen zeigen, dass die Richtung des Unterschieds gleich geblieben ist, wie beim stationären Test: Auch hier ist eine Verbesserung durch die optimierte Variante erkennbar.

Im Gegensatz zum stationären Test ist der Unterschied in Bezug auf die Dimension Zeit nicht mehr signifikant. Dies kann erneut an der geringen Stichprobengröße liegen. Im Gegensatz dazu ist jedoch im mobilen Test der Unterschied in Bezug auf die zurückgelegte Strecke signifikant geworden.

<i>Mobiler Test</i>	Basisvariante	Optimierte Variante	Statistische Signifikanz
Ø Zeit / Trial	3,37 sec s = 0,62 sec	3,07 sec s = 0,44 sec	F = 3,508 p = 0,103
Ø Fehler / Trial	0,050 s = 0,063	0,013 s = 0,019	F = 2,172 p = 0,184
Ø Overshoots / Trial	0,522 s = 0,290	0,188 s = 0,040	F = 11,178 p = 0,012
Ø Strecke / Trial	357,08 Pixel s = 21,36 Pixel	342,88 Pixel s = 21,52 Pixel	F = 10,753 p = 0,014

Tabelle 4.05: Ergebnisse des Signifikanztests für den mobilen Tapping-Test mit der Basis- und der optimierten Variante.

Bei der Betrachtung des Unterschiedes zwischen den Mittelwerten und unter Berücksichtigung der geringen Streuung (vgl. Abbildung 4.25), kann in Bezug auf die zurückgelegte Strecke ein großer Effekt erkannt werden. Tatsächlich ist auch die Abweichung von der idealen Distanz recht hoch: während die Probanden mit Hilfe der optimierten Mauszeigersteuerung durchschnittlich nur 35,83 Pixel mehr zurücklegten als die ideale Distanz beträgt, liegt die Abweichung mit der Basisvariante bei 50,03 Pixeln. Dieser große Unterschied spiegelt sich auch im η_p^2 wieder, der hier den Wert 0,606 annimmt.

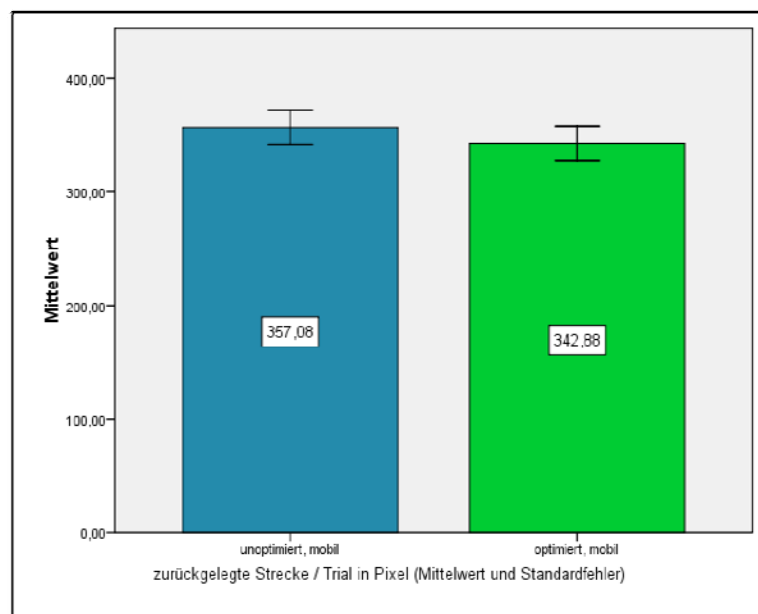


Abbildung 4.25: Mittelwerte und Streuung der zurückgelegten Strecke pro Trial im mobilen Test.

Auch bei der Betrachtung der Dimension Overshoots ist ein großer Unterschied zwischen den Mittelwerten der Basis- und der optimierten Variante zu erkennen (vgl. Abbildung 4.26). Darüber

hinaus ist die Streuung insbesondere bei Interaktion mit der optimierten Variante sehr klein. Dieser große Unterschied zeichnet sich auch in der Effektgröße $\eta_p^2 = 0,615$ ab.

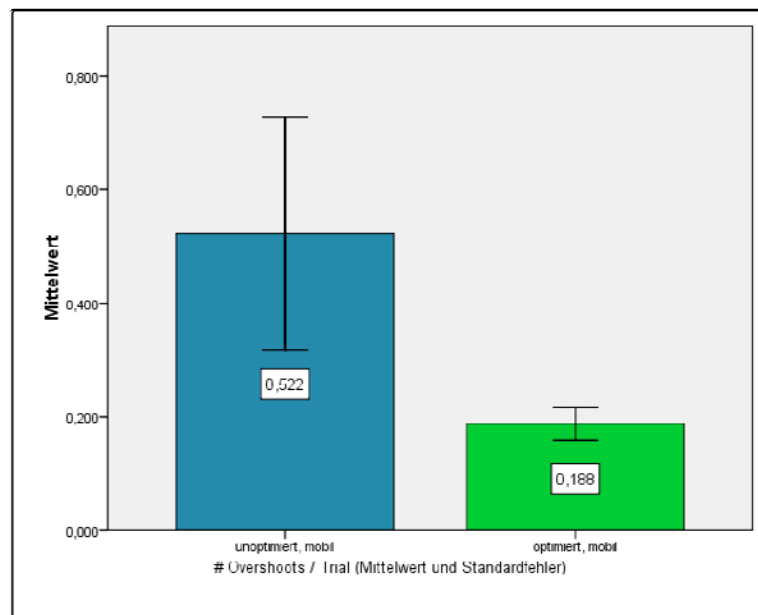


Abbildung 4.26: Mittelwerte und Streuung der Overshoots pro Trial im mobilen Test.

Der Grund für diese großen Effekte könnte sein, dass die Probanden während der Fahrt mehreren störenden Einflüssen ausgesetzt sind, als im Stand. Erschütterungen und weitere Störfaktoren schlagen sich im Fahrkontext negativ auf die Performance nieder. Daher sind die Unterschiede zwischen der Basis- und der optimierten Variante in allen vier Dimensionen beim mobilen Test deutlich größer, als beim stationären Test. Die optimierte Mauszeigersteuerung kann den Störfaktoren offensichtlich entgegenwirken und die Performance nachweislich signifikant verbessern. Aus diesem Grund wird die H_{0c} abgelehnt.

Die Forschungshypothese H_3 wurde sowohl im stationären, als auch im mobilen Test statistisch nachgewiesen, da „mit Hilfe der optimierten Mauszeigersteuerung weniger Overshoots gemacht werden, als mit der Basisvariante“. Die Hypothese H_1 besagte, dass „Ziele können mit Hilfe der optimierten Mauszeigersteuerung schneller getroffen werden, als mit der Basisvariante“ wurde zumindest für die stationäre Benutzung bestätigt. H_4 lautete, dass „mit Hilfe der optimierten Mauszeigersteuerung kleinere Strecken zurückgelegt werden“. Diese wurde für den mobilen Test statistisch nachgewiesen. Lediglich H_2 , die besagte dass „mit Hilfe der optimierten Mauszeigersteuerung weniger Fehler gemacht werden, als mit der Basisvariante“ konnte weder beim stationären, noch beim mobilen Test statistisch nachgewiesen werden und muss daher zurückgewiesen werden. Sowohl im stationären, als auch im mobilen Test ist eine Verbesserung in einigen Dimensionen der Performance deutlich erkennbar, sodass der Mehrwert der dynamischen Kraftfelder bestätigt wurde. Eine weitere Evaluationsstudie mit einer größeren Stichprobe und unter Einbezug

mehrerer Zielgruppen könnte ein umfassenderes Bild über den Vorteil der dynamischen Kraftfelder gegenüber einem unoptimierten Interaktionskonzept liefern.

Auswertung der qualitativen Daten

Auch die Beobachtungen und die Ergebnisse der Interviews lieferten weitere Hinweise auf die Nützlichkeit und Gebrauchstauglichkeit sowohl der optimierten Mauszeigersteuerung, als auch des gesamten finalen Interaktionskonzeptes. Insbesondere der am Ende des Tests überreichte Fragebogen und die darauffolgenden Fragen bezüglich einer eventuellen Veränderung im Laufe des Tests sollten die Forschungshypothese H_5 , dass „*kein Unterschied zwischen der optimierten Mauszeigersteuerung und der Basisvariante wahrgenommen wird*“ auf ihre Richtigkeit überprüfen. Dieser Fragebogen wurde sowohl nach dem stationären, als auch nach dem mobilen Test gereicht. Für die Beantwortung von H_5 wurden bei den Schülern jedoch nur die Antworten des ersten Testdurchgangs gewertet, da diese zum Zeitpunkt des mobilen Tests bereits von der optimierten Mauszeigersteuerung wussten.

Es stellte sich heraus, dass vier der elf Probanden die optimierte Mauszeigersteuerung wahrgenommen haben. Dabei handelte es sich um drei Schüler und einen Berufstätigen. Zwei der Schüler haben erst die optimierte Variante und anschließend die Basisvariante getestet, der dritte Schüler sowie der Berufstätige haben zunächst mit der Basisvariante und danach mit der optimierten Variante interagiert. Die dynamischen Kraftfelder wurden jedoch nur beim Tapping-Test bemerkt. Auf realen Webseiten konnte keiner der elf Probanden eine Veränderung des Mauszeigerverhaltens wahrnehmen. Auch beim mobilen Test, bei welchem die Schüler bereits wussten, dass sie zwei Varianten testeten, hat keiner die Wirkung der dynamischen Kraftfelder auf realen Webseiten wahrgenommen.

Die vier Probanden, die die Wirkung der Kraftfelder bemerkt hatten, äußerten sich diesbezüglich positiv und gaben an, dass ihnen diese beim Treffen der Links geholfen hat. Die anderen sieben Probanden gaben auf Nachfrage durch den Untersuchungsleiter an, keine Veränderung des Mauszeigerverhaltens wahrgenommen zu haben. Nach Demonstration der Veränderung durch den Untersuchungsleiter befanden fünf dieser sieben Probanden die dynamischen Kraftfelder als hilfreich, da sie trotz ungenauer Arbeitsweise eine präzise Auswahl ermöglicht. Die beiden anderen Probanden merkten an, dass die Ablenkung des Mauszeigers problematisch sein könnte, wenn ungewollte Strecken automatisch zurückgelegt würden. Da ihnen die Ablenkung jedoch nicht aufgefallen ist, wird diese Aussage nicht als Bedienungsproblem angesehen.

Sowohl die Schüler, als auch die Berufstätigen gaben an, im Laufe des Tests eine Veränderung wahrgenommen zu haben. Sieben Schüler und zwei Berufstätige nahmen eine allgemeine Verbesserung wahr. Sechs Schüler und zwei Berufstätige gaben an, im späteren Verlauf des Tests Links besser getroffen zu haben. Zum Großteil wurde ein gewisser Lerneffekt als Grund für diese Verbesserung genannt. Als Alternative wurden die dynamischen Kraftfelder genannt. Ein Schüler,

welcher zuerst die optimierte Variante getestet hat, hat eine allgemeine Verschlechterung und ein schlechteres Treffen der Links festgestellt und begründete diese mit dem Fehlen der dynamischen Kraftfelder. Ein Berufstätiger, welcher die Wirkung der dynamischen Kraftfelder nicht wahrgenommen hat, hat weder allgemein, noch in Bezug auf das Treffen der Links eine Veränderung feststellen können.

Die Auswertung der Logdaten hat bereits gezeigt, dass die dynamischen Kraftfelder beim Tapping-Test einen signifikant positiven Einfluss auf die Genauigkeit und Schnelligkeit beim Treffen von Links haben. Die Beobachtungen des Untersuchungsleiters beim Auswerten der Videoaufzeichnungen bestärken den Effekt der geringeren Overshootrate bei der optimierten Variante, sowohl beim Tapping-Test als auch auf realen Webseiten. Zudem hat keiner der Probanden die Wirkung der dynamischen Kraftfelder auf Webseiten wahrnehmen können, beim Tapping-Test wurde diese von der Mehrzahl der Probanden ebenfalls nicht bemerkt. Fiel den Probanden die Veränderung doch auf, wurde sie als positiv bezeichnet. Aus diesen Gründen wird die Hypothese H_5 angenommen.

Die erste Forschungsfrage F_1 , ob „*die optimierte Mauszeigersteuerung vom Benutzer wahrgenommen wird*“ kann verneint werden, da die dynamischen Kraftfelder von den meisten Probanden nicht wahrgenommen wurden. Die Frage F_2 „*wie dem Benutzer die optimierte Mauszeigersteuerung gefällt*“ kann damit beantwortet werden, dass die Optimierung vom Großteil der Probanden als positiv und hilfreich empfunden wird.

Um die Forschungsfrage F_3 , „*wie das Scrollen mittels Drehrad empfunden und wann wird es angewandt wird*“, beantworten zu können, wurden die Ergebnisse der Interviews und die Auswertung der Videoaufzeichnungen zusammengefasst. Es hat sich hier gezeigt, dass insbesondere zur Überbrückung längerer Strecken oder beim Navigieren zu bekannten Bereichen einer Webseite das Drehrad zum Scrollen verwendet wird. Auch um sich einen Überblick über eine Webseite zu verschaffen, beispielsweise bei der Suche nach einem bestimmten Artikel, wird diese Funktion angewandt. Neun der elf Probanden nutzten das Scrollen mittels Drehrad für diese Zwecke. Das Ziel des Projektteams war es, mit dieser Funktion eine schnelle Navigation zu gewährleisten, wie sie in der ersten Evaluationsstudie gewünscht wurde. Dieses Ziel wurde erreicht.

Vier der Probanden nutzten sowohl im Stand als auch in der Fahrt das Scrollen mittels Drehrad auch zum Lesen und empfanden die Schrittgröße als angenehm. Die anderen sieben benutzten hierfür das Scrollen mittels Bildschirmkanten. Vier der Probanden gaben an, die Schrittgröße sei zum Lesen zu groß. Eine dieser vier Personen scrollte beim Lesen dennoch weiterhin mittels Drehrad.

Bei der Auswertung der Videodaten fiel zusätzlich auf, dass sieben der Probanden an einer Stelle Informationen überlesen hatten und wieder nach oben scrollen mussten, um diese zu finden. Häufig passierte dies nur einmal und auch nur, wenn Texte nicht gelesen sondern gescannt wurden. Allerdings

kann das Übersehen von Informationen bei der visuellen Suche auch mit der Schrittgröße beim Scrollen mittels Drehrad in Zusammenhang stehen. Das Verkleinern der Schritte würde zwar eine bessere Unterstützung der Leseaktivitäten bieten, der Benutzer müsste hierfür jedoch ein langsames Navigieren in Kauf nehmen. Da jedoch eine schnelle Möglichkeit der Navigation in der Evaluationsstudie im Januar explizit gewünscht wurde, empfiehlt es sich in dieser Angelegenheit, die Schrittgröße beizubehalten, da eine langsamere Scrolling-Möglichkeit durch das Bewegen des Mauszeigers an die Bildschirmkanten gegeben ist.

Da bei der ersten Evaluationsstudie aufgrund der technischen Gegebenheiten das Zoomen zu einer verpixelten Darstellung des Textes führte, konnten hier keine verlässlichen Aussagen bezüglich des Zoomingverhaltens gegeben werden. Im finalen System wurde ein stufenloses Zoomen ermöglicht, den zehn der elf Probanden beim Lesen von Texten verwendeten. Nachdem diese Probanden die Zoomfunktion selbstständig gefunden hatten oder nach Bearbeitung der ersten Aufgaben vom Untersuchungsleiter auf diese Funktion aufmerksam gemacht wurden, konnten die Probanden diese Funktion problemlos aufrufen und verwenden. Die Forschungsfrage F₄, ob „*Zoomen beim Lesen bevorzugt wird*“, muss demnach bejaht werden, da die Mehrheit der Probanden diese Funktion oft und gerne nutzten.

Fünf der elf Probanden gaben zwar sowohl im stationären als auch im mobilen Test an, auch ohne Vergrößerung problemlos lesen zu können. Die Möglichkeit zu Zoomen sollte aber dennoch gegeben sein, da sie von der Probandengruppe häufig genutzt wurde und sie für andere Zielgruppen elementar sein könnte.

Zu Frage F₅, „*welche weiteren Bedienungsprobleme auftauchen*“, gibt es mehrere Antworten. Zunächst fielen während der Evaluationsstudie einige Probleme in Bezug auf die Benutzung der Softkeys auf. Um die maximale Bildschirmfläche nutzen zu können, wurden diese ausgeblendet, sobald der Proband begonnen hat, mit dem System zu interagieren. Ein Einblenden der Softkeys wurde durch einmaliges Drücken des „BACK“-Knopfes am MMI-Terminal ermöglicht (vgl. Abbildung 4.27, MMI-Terminal siehe Abbildung 4.13). Dieses Konzept wurde von der TV-Applikation des aktuellen MMI-Systems übernommen. Im Gegensatz dazu, waren die Softkeys jedoch stets aufrufbar, auch wenn sie nicht dargestellt wurden. Das einmalige Drücken des „BACK“-Knopfes sollte also eine Hilfestellung bieten, falls sich der Benutzer über die Anordnung der Funktionen nicht mehr bewusst ist.

Durch zweimaliges Drücken dieses Knopfes kann der Benutzer des Webbrowsers auf die Hauptansicht zurückkehren, in welcher er sämtliche geöffneten Webseiten („Tabs“ bei *Mozilla Firefox Version 3.5.3*, „Registerkarten“ im *Internet Explorer 8*) sieht.



Abbildung 4.27: Darstellung der Softkeys in den Bildschirmecken vor Interaktion durch Benutzer.

Für die Probanden erschien die Zuordnung der in den Bildschirmecken angezeigten Funktionen zu den um das MMI-Terminal angeordneten Knöpfen zunächst nicht intuitiv. So versuchten sechs Probanden anfangs die Funktionen mittels Mausclick auf die Softkeys aufzurufen. Da diese jedoch bei einer Interaktion sofort verschwinden, riefen einige dieser Probanden sie mehrfach wieder mittels „BACK“ auf, um anschließend erneut darauf klicken zu können. Dieses Problem tauchte jedoch nur beim ersten Durchgang auf. Im späteren Verlauf des Tests fiel auf, dass neun der Probanden die Funktionen direkt aufrufen, ohne die Softkeys zuvor mittels „BACK“ eingeblendet zu haben. Die beiden anderen Probanden gaben jedoch an, dass sie auch nach einer gewissen Nutzungszeit die Benutzung der Softkeys als verwirrend empfanden. So sei das Einblenden der Softkeys mittels „BACK“ nicht intuitiv und es musste auch länger überlegt werden, bevor Funktionen aufgerufen wurden.

Ein anderes Bedienungsproblem betrifft das Zurückkehren zur Hauptansicht. Der Grund hierfür ist vermutlich, dass der Webbrowser zwei verschiedene Möglichkeiten des „zurück“ offeriert. Einerseits kann durch Drücken des Softkeys „Seite zurück“ in der Chronologie der besuchten Webseiten eine Seite zurück navigiert werden. Andererseits kann durch zweimaliges Drücken des „BACK“-Knopfes zur Hauptansicht des Browsers zurückgekehrt werden. Sieben der Probanden versuchten nun, mehrfach „Seite zurück“ zu drücken, um zur Hauptansicht zu gelangen. Sie vermuteten, dass automatisch zu selbiger zurückgekehrt wird, wenn sie in der Chronologie bis vor die erste besuchte Seite zurück navigieren würden. Drei dieser Probanden fanden den Weg schließlich selbstständig, nachdem sie bemerkten, dass mehrfaches „Seite zurück“ nicht zum gewünschten Effekt führen würde. Einer der Probanden gab an, dass er den Weg zur Hauptansicht nur durch Glück gefunden hatte. Einem weiteren Probanden musste durch den Untersuchungsleiter Hilfestellung geboten werden, da er sonst den Weg zur Hauptansicht nicht gefunden hätte. Dieses Problem ist jedoch im späteren Verlauf des Tests nicht mehr aufgetreten. Erfahrene MMI-Nutzer stoßen zudem vermutlich nicht auf dieses Problem, da das Zurücknavigieren in der Menüstruktur des MMI-Systems ebenfalls mit Hilfe des

„BACK“-Knopfes funktioniert. Nichtsdestotrotz wäre eine intuitivere Möglichkeit des Zurücknavigierens hilfreich.

Ein drittes Bedienungsproblem, welches jedoch nur bei einem geringen Teil der Probanden aufgetreten ist, betrifft die Gestaltung der Hauptansicht des Browsers. Da die geöffneten Webseiten in einem Rondell angeordnet sind, kann es hier zu Schwierigkeiten bei der Auswahl der gewünschten Webseite kommen. In den Tat hatten drei Probanden diese Schwierigkeiten, die sich darin äußerten, dass sie zunächst versuchten, Softkeys zu drücken, um ein Element auszuwählen oder das MMI-Terminal zu schieben um einen Mauszeiger zu suchen. Diese Interaktionsformen wurden intuitiv genutzt, da sie von der Interaktion mit der Maus am handelsüblichen Desktop-PC bekannt sind. Dieses Problem tauchte jedoch nur bei der ersten Konfrontation mit dem Webbrowser auf. Zudem haben die anderen acht Probanden die gewünschten Webseiten intuitiv mittels Druck auf das MMI-Terminal ausgewählt.

Vier der Probanden gaben zusätzlich an, dass die Gestaltung der Hauptansicht sehr hilfreich sei, da die Webseiten sofort erkannt werden. Im Gegensatz dazu muss beim handelsüblichen Webbrowser der Webseitentitel gelesen oder ein Icon der Webseite zur Identifikation genutzt werden.

Insgesamt fällt auf, dass das Design der Hauptansicht hilfreich und die Bedienung des Rondells für den Großteil der Probanden intuitiv ist. Der Weg zur Hauptansicht wird jedoch nicht gleich gefunden und die Softkeys werden, wie schon in der Evaluationsstudie im Januar 2009 weiterhin angeklickt. Diese Probleme treten vermutlich insbesondere bei neuen Nutzern des MMI-Systems auf, da erfahrene Nutzer das unterliegende Interaktionskonzept mittlerweile kennen. Nichtsdestotrotz bedarf es hier einer Optimierung des Browsers, um auch diese Funktionen für alle Benutzer intuitiv zugänglich zu machen.

Darüber hinaus fällt auf, dass durch Integration der Mauszeigersteuerung im Zoom-Modus der Umgang mit beiden Modi leichter fällt und verständlicher ist, als bei der vorigen Browsergeneration. Insgesamt funktioniert das finale Konzept gut. Viele Bedienungsprobleme seiner Vorgänger wurden beseitigt und das Konzept wurde so optimiert, dass die Benutzung effektiv, effizient und subjektiv zufrieden stellend ist.

4.2.5. Herausforderungen und Einschränkungen

Nachdem bereits die erste Evaluationsstudie in diesem Kooperationsprojekt unter Berücksichtigung von klassischen Methoden der Mensch-Computer Interaktion stattfand, sollte die zweite Evaluationsstudie mit Hilfe von Methoden der Automobilindustrie durchgeführt werden. Wie sich jedoch zeigte, sind größere Unterschiede zwischen den Methoden der Automobilindustrie und denen der traditionellen Mensch-Computer Interaktion insbesondere dann zu finden, wenn es sich bei dem zu untersuchenden System um eines für den Fahrer handelt. In diesem Kontext war dies jedoch nicht der Fall, sodass bis auf die außergewöhnliche Testumgebung klassischen Methoden der HCI angewandt wurden. Dennoch stellte auch die zweite Evaluationsstudie das Entwicklerteam und den

Untersuchungsleiter vor neue Herausforderungen, da diese sich in ihrer Art und ihrer Zielsetzung stark von der ersten Studie unterschied. Daher mussten hier teilweise andere Methoden der klassischen Mensch-Computer Interaktion auf den automotiven Kontext übertragen werden.

Die erste große Herausforderung betraf die Auswahl und Akquise geeigneter Probanden. Wie in Kapitel 4.2.3 beschrieben, sollte eine Gruppe von Schülern das neue Interaktionskonzept mit der unoptimierten Variante vergleichen. Da diese allesamt minderjährig waren, mussten in Bezug auf die Durchführung der Tests einige Aspekte beachtet werden, die in der ersten Evaluationsstudie noch nicht von Belang waren. So mussten beispielsweise die Erziehungsberechtigten ihr Einverständnis für Fahrtests geben, bevor diese mit ihren Kindern durchgeführt werden konnten. Auch die Akquise der jungen Probanden gestaltete sich als schwieriger, als der Einbezug von Studenten während der ersten Evaluationsstudie, doch gelang es dem Projektteam letztlich doch, alle benötigten Probanden für die Studie zu gewinnen.

Zusätzlich zu den Tests mit Schülern wurden drei berufstätige Probanden zum Test eingeladen, die in verschiedensten Bereichen der Universität Konstanz angestellt sind. Diese sollten zusätzlichen subjektiven Input liefern und wurden daher nur zum mobilen Test eingeladen. Der Grund für die Wahl der mobilen Testumgebung ist, dass sich bereits in der ersten Evaluationsstudie zeigte, dass sich die Bedienungsprobleme in der Fahrt von denen im Stand nicht gänzlich unterscheiden, sondern sich lediglich verstärken oder erst hier auftreten. Zudem ist die Fahrt ohnehin der reale Anwendungskontext des automotiven Webbrowsers sodass Tests in dieser Umgebung ökologisch valider sind und die gefundenen Ergebnisse eher der tatsächlichen Nutzung entsprechen.

Alternativ hätte der experimentelle Vergleich der optimierten und der Basisvariante im Stand und während der Fahrt auch mit Hilfe von berufstätigen Probanden durchgeführt werden können, doch wäre hierfür die Integration mehrerer Probanden nötig gewesen. Im Rahmen dieser Untersuchung war dies aus organisatorischen und zeitlichen Gründen nicht möglich.

Eine weitere Herausforderung betraf die Auswahl einer geeigneten Pause zwischen den stationären und mobilen Tests, die die Schüler absolvieren sollten. Davon ausgehend, dass längere Fahrten, in welchen die Nutzung eines automotiven Webbrowsers in Frage käme, unter der Woche vermutlich eine Seltenheit sind und eher bei einem Wochenendausflug der Familie auftreten, wurden sieben Tage als geeignete Pause empfunden. Aus diesem Grunde wurden die Termine für die zweite, mobile Testreihe entsprechend der Verfügbarkeit der Probanden so gewählt, dass etwas eine Woche zwischen den beiden Tests lag.

Auch die Wahl einer geeigneten Teststrecke stellte eine Herausforderung dar. Da die Probanden nicht unnötig lang eingespannt werden sollten, musste die Strecke so gewählt werden, dass sie möglichst zentral lag und keine allzu langen Anfahrten erforderte. Daher wurde als Start- und Zielpunkt ein

Parkplatz im Eingangsbereich der Universität Konstanz gewählt. Die Strecke sollte ebenfalls in den Zeitrahmen der Evaluation passen und gegebenenfalls beliebig verlängert oder verkürzt werden können, falls manche Probanden für die Bearbeitung der Aufgaben mehr oder weniger Zeit benötigten. Deshalb wurde – wie bereits bei der ersten Evaluationsstudie im Januar 2009 – eine Strecke gewählt, die um beliebig viele Runden um die Konstanzer Innenstadt erweitert werden konnte. Obwohl die Tests an verschiedenen Wochentagen und zu verschiedenen Tageszeiten stattfanden, fanden sich der Fahrer, die Untersuchungsleiter sowie der jeweilige Proband stets in der Situation eines zäh fließenden Verkehrs wieder. Idealerweise hätte eine Strecke gewählt werden sollen, die mehr Landstraßen oder Autobahnen einschließt, da längere Strecken typischerweise über solche Abschnitte führen. Dies wäre jedoch möglicherweise in Bezug auf das Einverständnis der Eltern problematisch gewesen, da eine langsamere Fahrt im Stadtverkehr weniger Gefahren birgt als die schnellere Fahrt auf einer Autobahn.

In Bezug auf die Gestaltung der Aufgaben stellte sich auch die Untersuchung der Effektivität und Effizienz der Interaktionskonzepte als Herausforderung dar. Wie bereits in Kapitel 4.2.2. beschrieben, war die Verwendung eines simplen, ein- oder multidirektionalen Tapping-Tests nicht praktikabel. Der besondere Nutzungskontextes und die damit verbundenen besonderen Aufgaben erforderten eine Abwandlung dahingehend, dass reale Elemente realer Webseiten zu einer Tapping-Test Webseite kombiniert werden mussten, um Informationen über die tatsächliche Nützlichkeit der dynamischen Kraftfelder zu erhalten.

Eine letzte Herausforderung trat schließlich bei der Auswertung der Daten auf. Da während der Nutzung des automotiven Webbrowsers und insbesondere während der Bearbeitung des Tapping-Tests typischerweise Lerneffekte auftreten, sollte deren Einfluss weitestgehend eliminiert werden, um verlässlichere Vergleiche ziehen zu können. Aus diesem Grund wertete die Autorin dieser Arbeit zum Vergleich der beiden getesteten Varianten jeweils die letzten beiden Blöcke des Tapping-Tests aus. Auch die Performance-Daten der berufstätigen Probanden sollten zumindest im Rahmen der mobilen Tests zusätzlich ausgewertet werden, doch aufgrund technischer Probleme bei einem der drei Probanden konnten hier die Daten des einen Testdurchlaufs nicht gespeichert werden. Bei einem weiteren Probanden wurde der Tapping-Test frühzeitig abgebrochen, da der Proband zugab, diesen sabotieren zu wollen, da er keine Lust mehr auf die Aufgabe hatte. Somit waren nur die Daten des dritten Probanden vollständig, was der Auswertung den Sinn nahm. Anhand der verfügbaren Daten wurde jedoch kein allzu großer Unterschied zwischen der Performance der Schüler und der Performance der Berufstätigen festgestellt. Idealerweise hätten jedoch diesbezüglich weitere Tests mit berufstätigen Probanden durchgeführt werden müssen, was jedoch – wie bereits oben beschrieben – aus organisatorischen und zeitlichen Gründen nicht möglich war.

5. Zusammenfassung und Fazit

Der automotiv Kontext stellt bei der Konzeption, Entwicklung und Evaluation neuer Systeme einige Herausforderungen an die Entwickler, die sich von den Herausforderungen in der klassischen Mensch-Computer Interaktion, wie sie von Systemen für Desktop-PC bekannt sind, unterscheiden. Ziel dieser Arbeit war es, diese Unterschiede mit besonderem Fokus auf Datenerhebung und Evaluation herauszuarbeiten. Es konnten hierbei drei entscheidende Unterschiede zwischen automotivem und Desktop-Kontext festgestellt werden, die im nachfolgenden nochmals zusammengefasst werden. Diese betreffen die Auswahl der Probanden, Probanden also der Teilnehmer einer Studie, weiterhin die Forschungsfragen und die dadurch notwendigen Abwandlungen der klassischen Methoden der Mensch-Computer Interaktion sowie die Befragungs- und Testumgebung, die durch die besonderen Forschungsfragen vorgegeben wird.

Die Probanden, die im Rahmen von automotiven Entwicklungsprozessen zu Benutzerbefragungen, Benutzungsstudien und Benutzbarkeitstests eingeladen werden, unterscheiden sich wie folgt. Je nach Art und Anwendungskontext des zu entwickelnden Systems ist die Zielgruppe sehr genau definiert. Eine grobe Unterteilung findet hier zunächst nach Fahrer, Beifahrer und Passagier im Fondbereich statt. Bei manchen automotiven Forschungsfragen, muss jedoch genauer spezifiziert werden, zum Beispiel nach Fahrzeugbesitzern bestimmter Fahrzeugklassen.

Die Personen, die in der Automobilindustrie im Rahmen von Anforderungsanalysen und Evaluationen befragt werden, stammen typischerweise aus dem konzerninternen Probandenpool. Dies hat zwar organisatorische Vorteile wie Schnelligkeit, Kosteneffizienz und eine bessere Überwachung der Geheimhaltungsrichtlinien, doch wie in Kapitel 3.4.3. beispielhaft erläutert wurde, sind Probanden aus dem Probandenpool durch ihren Arbeitgeber und ihr Bild der Automarke voreingenommen. Ein Sachverhalt, der nicht immer zu validen Ergebnissen führt.

Bei universitären Forschungsprojekten mit automotivem Hintergrund besteht in der Regel kein Zugriff auf eine Probandendatenbank. Dies hat zwar einen höheren organisatorischen Aufwand zur Folge, doch kann durch die selbstständige Auswahl der Probanden eine höhere Vielfalt in Bezug auf die persönlichen Einstellungen zur Fahrzeugmarke gewährleistet werden. Dieser Umstand führt dazu, dass die Probanden unabhängig sind und die benötigten Funktionen eines Systems oder seine Gebrauchstauglichkeit valide bestimmen können. Allerdings kann dieser Umstand auch dazu führen, dass die Probanden nicht der exakten Zielgruppe entsprechen, da universitären Forschungsgruppen zumeist keine allzu detaillierten Informationen über Alter, Geschlecht, Gehaltsklasse oder Familienverhältnisse des typischen Käufers einer bestimmten Fahrzeugklasse vorliegen.

Dennoch empfiehlt sich auch in der Automobilindustrie die Integration von externen Probanden, wie es durch die Akquise beispielsweise über Fahrzeughändler oder an Tankstellen möglich ist. Richtlinien in Bezug auf die Geheimhaltung können dennoch durch das Unterschreiben einer Geheimhaltungsverpflichtung eingehalten werden.

Ein weiterer Unterschied betrifft die Forschungsfragen die der automotive Nutzungskontext im Vergleich zu Untersuchungen von Systemen am Desktop-PC aufwirft und die damit verknüpften notwendigen Änderungen der Methoden zur Datenerhebung und Evaluation. Während es bei der Entwicklung von Systemen für Passagiere im Fond zumeist ausreicht, klassische Methoden der Mensch-Computer Interaktion heranzuziehen, um die Gebrauchstauglichkeit der Systeme zu analysieren, spielen bei der Gebrauchstauglichkeit von Fahrersystemen einige weitere Aspekte, wie die Sicherheit sämtlicher Verkehrsteilnehmer eine gewichtige Rolle. Aus diesem Grunde ist es wichtig, Faktoren wie die Ablenkung des Fahrers, seinen Frustrations- oder Erregungszustand zu ermitteln. Hierfür bedarf es der Integration alternativer Messmethoden wie beispielsweise dem Eyetracking (vgl. [Bach et al. 08]), der Messung der Spurhaltung mittels Durchführung eines Lane Change Tasks (vgl. [Kuhn 05]) oder auch der Messung der Herzfrequenz (vgl. [Gruenstein et al. 09]). Teilweise erfordern diese Methoden einen großen Aufwand in Bezug auf die Auswertung. Videoanalysen zur Beschreibung der Blickdaten beispielsweise werden in der Automobilindustrie daher eher selten angewandt. Im Gegensatz dazu sind sie in der automotiven Forschung an Universitäten eine gebräuchliche Methode, um die Ablenkungswirkung von interaktiven Systemen zu untersuchen.

Ein wichtiger Aspekt ist in diesem Zusammenhang, dass die Interaktion mit einem System für den Fahrzeugführer stets die Sekundäraufgabe ist, mit der er sich zusätzlich zu seiner Primäraufgabe, dem Steuern eines Fahrzeugs, auseinandersetzen muss. Im Zentrum automotiver Untersuchungen befinden sich daher oftmals Aspekte des Multi-Tasking. Im Vergleich dazu ist das Interagieren mit Systemen am Desktop-PC, wie es bei klassischen Untersuchungen in der Mensch-Computer Interaktion der Fall ist, die Primäraufgabe, bei der potentielle Unterbrechungen durch den Einsatz einer kontrollierten Umgebung eines Usability Labors vermieden werden.

Im Rahmen der automotiven Anforderungsanalyse eignen sich hingegen altbewährte Methoden wie Fragebögen, Interviews oder Fokusgruppen sehr gut, um mit den potentiellen Benutzern in Kontakt zu treten. Aus diesem Grunde werden sie auch in der Automobilindustrie häufig eingesetzt. Der Unterschied zu konventionellen Methoden ist hier lediglich in der Akquise der Probanden und dem Ort der Befragung zu sehen.

Der Ort der Befragung ist demnach der dritte Aspekt, in welchem sich automotive Untersuchungen von jenen der klassischen Mensch-Computer Interaktion unterscheiden. Im Gegensatz zu Systemen am Desktop-PC ist die Umgebung, in welcher automotive Systeme bedient werden, veränderlichen Umweltbedingungen unterworfen (vgl. Kapitel 2.1.), was sich auch auf die Umgebung auswirken sollte, in welcher automotive Systeme getestet werden. Grundsätzlich erhält man bei einer Untersuchung, die im tatsächlichen Nutzungskontext eines Systems stattfindet ökologisch validere Ergebnisse. Der Grund dafür ist, dass die hier durchgeführten Interaktionen eher der natürlichen Bedienung entsprechen, als solche, die im kontrollierten Umfeld einer Laborumgebung getätigt

werden. Daher empfiehlt es sich, für Evaluationen im automotiven Bereich, das Fahrzeug als Testumgebung heranzuziehen.

Wie oben genannte Forschungsfragen jedoch zeigen, müssen im automotiven Kontext manchmal auch sicherheitskritische Situationen untersucht werden. In solchen Fällen ist es zum Schutz des Probanden und aller anderen Verkehrsteilnehmer sinnvoller, den Anwendungskontext zu simulieren. Aus diesem Grund finden sehr viele Untersuchungen in der Automobilindustrie in einem Fahrsimulator statt. Im Gegensatz zu universitären Forschungsteams hat die Automobilindustrie hier jedoch einen entscheidenden Vorteil: Das höhere Budget ermöglicht es, hochwertige Fahrsimulatoren zu entwickeln, die durch 360°-Ansichten, Motorengeräuschen und manchmal auch durch motorisches Feedback über eine hohe physische Wiedergabetreue verfügen. In einer solchen Umgebung fällt es dem Probanden leichter, sich in den realen Nutzungskontext hineinzusetzen, als es bei Lo- oder Medium-Fi Simulatoren möglich ist, bei welchen der Proband auf einem Stuhl vor einem etwas größeren Monitor sitzt und das virtuelle Fahrzeug mit Hilfe eines Spielekonsolen-Lenkrades steuert. Aus diesem Grunde werden Aufgaben in einem Hi-Fi Fahrsimulatoren eher ernst genommen, als es bei Lo-Fi Simulatoren der Fall ist [Bach et al. 08]. Fahrsimulatoren haben zudem den Vorteil, dass die benötigten Daten sehr genau erfasst werden können und die Auswertung daher ebenfalls mit einer großen Genauigkeit erfolgen können. Durch die von Test zu Test gleichbleibenden Rahmenbedingungen ist zudem eine bessere Vergleichbarkeit der Ergebnisse möglich. Darüber hinaus haben einige Untersuchungen gezeigt, dass die Ergebnisse aus Fahrsimulatorstudien grundsätzlich mit Fahrtests im realen Fahrzeug vergleichbar sind. Zumindest sind bei der Beantwortung der Forschungshypothesen ähnliche Trends erkennbar [Bach et al. 08] [Lange et al. 06]. Allerdings haben Untersuchungen auch gezeigt, dass Probanden in Simulatorstudien in manchen Situationen anders agieren und reagieren, als im realen Fahrzeug [Bach et al. 08] [Goodman et al. 97]. Grund hierfür ist einerseits, dass durch das fehlende sensorische Feedback die Simulatorsituation schwieriger zu kontrollieren ist und die Fahraufgabe daher mehr Aufmerksamkeit bedarf. Zum anderen werden jedoch auch Fahrfehler in diesem Kontext eher in Kauf genommen, da sie keine schwerwiegenden Konsequenzen für den Probanden haben.

Bei Untersuchungen universitärer Forschungsteams, bei denen weniger gut ausgestattete Simulatoren zum Einsatz kommen, verstärken sich diese Probleme. Da die Testumgebung bei Lo-Fi Simulatoren sehr artifiziell ist, ist das Verhalten der Probanden noch weiter von der Realität entfernt. Evaluationen in einem realen Fahrzeug können im universitären Umfeld jedoch auch nicht ohne Weiteres durchgeführt werden: diese sind nur dann möglich, wenn sie im Rahmen eines Kooperationsprojekts mit einem Automobilhersteller stattfinden und dieser das Fahrzeug zur Verfügung stellt. Andernfalls könnten die Anforderungen an die Sicherheit der Passagiere kaum erfüllt werden. Sollen Systeme für den Gebrauch durch den Fahrzeugführer getestet werden, verbieten zudem rechtliche und ethische Aspekte sicherheitsgefährdende Studien im realen Nutzungskontext [Bach et al. 08].

Bei sicherheitsunkritischen Untersuchungen von Systemen, wie es beispielsweise bei Systemen für Beifahrer oder Passagiere im Fond der Fall ist, sollten Evaluationen im realen Straßenverkehr jedoch durchaus in Betracht gezogen werden. Untersuchungen haben gezeigt, dass sich die Ergebnisse solcher Evaluationsstudien von jenen im Simulator in manchen Aspekten unterscheiden, wie beispielsweise dass im realen Fahrzeug der Blick bereitwilliger von der Straße abgewandt wird, als im Fahrsimulator (vgl. [Bach et al. 08], [Goodman et al. 97]). Ferner haben Untersuchungen gezeigt, dass manche Probleme, wie beispielsweise eine unpräzise Mauszeigersteuerung, erst bei der Fahrt auftreten (vgl. Kapitel 4.1.2.). Um diese Unterschiede und Bedienungsprobleme zu erkennen, ist nach Möglichkeit das reale Fahrzeug als Testsetting vorzuziehen. Gerade weil sich diese Umgebung durch veränderliche und nicht steuerbare Störeffekte auszeichnet, bilden hier stattfindende Untersuchungen die Realität eher nach und bewirken, dass der Proband seine Aufgaben ernst nimmt.

Fazit

Die Mitglieder der Projektteams, die für die Entwicklung des automotiven Webbrowsers verantwortlich waren, mussten sich sowohl bei der Konzeption, als auch bei der Evaluation des selbigen den Herausforderungen stellen, die der automotive Nutzungskontext mit sich brachte. Diese Arbeit hat gezeigt, wie diesen in einem universitären Umfeld begegnet werden kann, um einen gebrauchstauglichen finalen Prototyp zu entwickeln. Nicht zuletzt war es die iterative Herangehensweise bei der Entwicklung, die einen sehr großen Beitrag bei der Erreichung dieses Ziels leistete. Durch den benutzer- und benutzungsorientierten Ansatz und durch das Einbetten zweier Evaluationsstudien in den Entwicklungsprozess war es möglich, die Anforderungen mehrfach zu überarbeiten und so zu einem Konzept zu gelangen, welches nicht nur die benötigten Funktionen umfasst, sondern dessen Nutzung darüber hinaus unterhaltsam ist.

Die Probanden, die zu den Befragungen und Evaluationen eingeladen wurden, hatten verschiedene berufliche und edukative Hintergründe und gehörten zudem unterschiedlichen Altersklassen an, sodass verschiedene Zielgruppen abgedeckt wurden und dadurch vielfältiger Input zu Stande kam. Da sie zudem nicht unter dem Einfluss eines bestimmten Bildes der Fahrzeugmarke standen, konnten sachliche und valide Ergebnisse produziert werden. Dabei konnten dennoch die für die Automobilindustrie essentiellen Richtlinien der Geheimhaltung gewahrt werden, da die Probanden sorgfältig ausgewählt wurden und zudem vor den Tests Geheimhaltungsverpflichtungen unterschreiben mussten.

Dank der engen Zusammenarbeit der beiden Kooperationspartner, der Volkswagen AG und dem Lehrstuhl für Mensch-Computer Interaktion an der Universität Konstanz, konnten trotz des begrenzten zeitlichen und finanziellen Budgets des Projektteams zunächst Untersuchungen in einem Audi A8 und später in einem Audi Q7 durchgeführt werden. Diese Testumgebungen hatten nicht nur den Vorteil, dass durch ihre physische und funktionale Wiedergabetreue ein realistischeres Testsetting entstand, in

welchem, Aufgaben ernst genommen und in ihrem realen Anwendungskontext bearbeitet werden können. Zusätzlich war es auf diese Weise auch möglich, die wichtigste Forschungsfrage zu beantworten: wie gut sich das entwickelte Browserkonzept für eine Interaktion während der Fahrt eignet. Es konnten sowohl stationäre als auch mobile Untersuchungen durchgeführt werden, deren Ergebnisse verdeutlichten, wie wichtig eine Evaluation eines Systems in seinem tatsächlichen Nutzungskontext ist. Dies entspricht auch den Ergebnissen von [Bach et al. 08] und [Goodman et al. 97], die ebenfalls Unterschiede in der Bedienung zwischen realem und simuliertem Nutzungskontext gefunden haben. Insbesondere im sicherheitsunkritischen Kontext empfiehlt es sich daher im tatsächlichen Nutzungskontext zu testen. Nur so können die Bedienungsprobleme aufgespürt werden, die bei einer realen Nutzung auftreten und nur so können valide Ergebnisse erhalten werden, die für die Entwicklung eines gebrauchstauglichen und unterhaltsamen Systems erforderlich sind.

6. Literaturverzeichnis

- [AAM 02] AAM American Statement of Principles. 2002. In: Rosetta – Real Opportunities for Exploitation of Transport Telematics Applications: „HUMAN MACHINE INTERFACE - Guidelines and Legislation“. http://www.trg.soton.ac.uk/rosetta/workareas/2b_hmi/hmi_gl_aamsop.htm (Stand: 2009-12-01).
- [Ablassmeier et al. 06] Ablassmeier, Markus; Politschke, Tony & Rigoll, Gerhard. 2006. „A New Approach of a Context-Adaptive Search Agent for Automotive Environments“. CHI 2006 – Work-in-Progress. CHI 2009 in Montréal, Québec, Canada, 22. – 27. April 2006. <http://portal.acm.org/citation.cfm?id=1125451.1125745> (Stand: 2009-12-01).
- [Audi 08] AUDI AG. 2008. Webauftritt von Audi Deutschland. www.audi.de (Stand: 2008-11-30).
- [Audi 09a] AUDI AG. Produktbroschüre Q7. Gültig ab: Juni 2009.
- [Bach et al. 08] Bach, Kenneth M.; Jæger, Mads G.; Skov Mikael B. und Thomassen, Nils G. 2008. „Evaluating Driver Attention and Driving Behaviour: Comparing Controlled Driving and Simulated Driving“. In: Proceedings of the 22nd British HCI Group Annual Conference on HCI 2008: People and Computers XXII: Culture, Creativity, Interaction - Volume 1 (pp. 193-201). British Computer Society (Swinton, UK). ISBN: 978-1-906124-04-5. <http://portal.acm.org/citation.cfm?id=1531514.1531541> (Stand: 2009-12-01).
- [BMW 08] BMW Themenheft: „BMW ConnectedDrive“. 2008.
- [Card et al. 83] Card, Stuart K.; Moran, Thomas P. & Newell, Allen. 1983. „The Psychology of Human-Computer Interaction“. Lawrence Erlbaum Associates, Inc. Publishers (New Jersey). ISBN: 0-89859-243-7.
- [Constantine & Lockwood 99] Constantine, Larry L.; Lockwood, Lucy A.D. 1999. „Software for Use – A Practical Guide to the Models and Methods of Usage-Centered Design“. Addison-Wesley. ISBN: 0-201-92478-1.

- [Cheval 09] Chur Evaluation Laboratory (Cheval). 2009.
<http://www.cheval-lab.ch/cheval-wissensbasis/> (Stand: 2009-12-01).
- [DATech 09] Deutsche Akkreditierungsstelle Technik (2009): Leitfaden Usability, Version 1.2. <http://www.datech.de/share/files/Leitfaden-Usability.pdf> (Stand: 2009-12-01).
- [Diez 04] Diez, Willi. 2004. „Kundenzufriedenheit und Kundenbindung in der Automobilwirtschaft“. In: B. Ebel, M.B. Hofer, J. Al-Sibai: Automotive Management – Strategie und Marketing in der Automobilwirtschaft (pp. 673-694). Springer-Verlag (Berlin, Heidelberg). ISBN 3-540-00226-X.
- [Dix et al. 04] Dix, Alan; Finlay, Janet; Abowd, Gregory G. & Beale, Russel. 2004. „Human-Computer Interaction (3rd Edition). Pearson Education Limited (Harlow, Essex, England). ISBN: 0-13-046109-1.
- [Dubrowsky et al. 01] Dubrowsky, A.; Hüttner, J.; Warning, J.; Wandke, H. & Küting, H.-J. 2001. „CarE – ein Software-Tool zur Kognitiv-Ergonomischen Bewertung von Komponenten in Fahrzeugen“. In: Kongressband Deutscher Psychologentag 2001. ISBN: 3-931589-47-1.
- [Dumas & Redish 99] Dumas, Joseph S. & Redish, Janice C. 1999. „A practical Guide to Usability Testing“. Intellect Ltd (Exeter, England). ISBN: 1-84150-020-8.
- [Dzaack 08] Dr.-Ing. Dzaack, Jeronimo. 2008. „SimTrA: Die Analyse kognitiver Benutzermodelle für die Evaluation von Mensch-Maschine-Systemen“. In: VDI-Berichte Nr. 2041, 2008 (pp. 265-276). VDI Verlag GmbH (Düsseldorf). ISBN: 978-3-18-092041-2.
- [Dzida et al.01] Dzida, Wolfgang; Hofmann, Britta; Freitag, Regine; Redtenbacher, Wolfgang; Baggen, Robert; Geis, Thomas; Beimel, J.; Zurheiden, C.; Hampe-Neteler, W.; Hartwig, R. & Peters, H. 2001. „Gebrauchstauglichkeit von Software. ErgoNorm: Ein Verfahren zur Konformitätsprüfung von Software auf der Grundlage von DIN EN ISO 9241 Teile 10 und 11“. Hrsg.: Bundesanstalt für Arbeitsschutz und Arbeitsmedizin. Wirtschaftsverlag N. W. Verlag für neue Wissenschaft (Bremerhaven). ISBN-13: 978-3897016781.

- [Eilers et al. 86] Eilers, Karin; Nachreiner, Friedhelm & Hänecke Kerstin. 1986. „Entwicklung und Überprüfung einer Skala zur Erfassung subjektiv erlebter Anstrengung“. In: Zeitschrift für Arbeitswissenschaft, Band 40, 1986/4 (pp. 215-224). Ergonomia Verlag (Stuttgart).
- [ESoP 99] Europäische Kommission. 1999. European Statement of Principles. In: Rosetta – Real OpportunitieS for Exploitation of Transport Telematics Applications: „HUMAN MACHINE INTERFACE - Guidelines and Legislation“. http://www.trg.soton.ac.uk/rosetta/workareas/2b_hmi/hmi_gl_ecsop.htm (Stand: 2009-12-01).
- [Fragebogen 09] Söter, Andrea. 2009. Acht komplett ausgefüllte Fragebögen zum Thema „Benutzerbefragungen in der Automobilindustrie“ (erstellt bei www.questionpro.com).
- [Geis 06] Geis, Thomas. 2006. „Usability-Normung: Die neue DIN EN ISO 9241-110 "Grundsätze der Dialoggestaltung"“. 11. August 2006. <http://www.procontext.com/de/news/2006-08-11.html> (Stand: 2009-12-01).
- [Gerken & Reiterer 09] Gerken, Jens & Reiterer, Harald. 2009. „Eine Taxonomie für Längsschnittstudien in der MCI“. In: Mensch & Computer 2009: Grenzenlos frei? 9. Konferenz für interaktive und kooperative Medien, September 2009. Oldenbourg Verlag (München). http://hci.uni-konstanz.de/klinkh/gerken/Taxonomie_mc09.pdf (Stand: 2009-12-01).
- [Gespräche] Gespräche mit Mitarbeitern der Automobilindustrie, die die Autorin dieser Arbeit während der Kooperation geführt hat. Mitarbeiter, Doktoranden und Diplomanden haben dabei frei oder im Rahmen eines Interviews von ihren Erfahrungen und ihrer Arbeit berichtet.
- [Goodman et al. 97] Goodman, M.; Bents, F. D.; Tijerina, L.; Wierville, W.; Lerner, N. und Benel, D. 1997. „A Review of Human Factors Studies on Cellular Telephone Use While Driving“. In: An Investigation of the Safety Implications of Wireless Communications in Vehicles. U.S. Department of Transportation (November 2007). <http://www.nhtsa.dot.gov/people/injury/research/wireless/c5.htm> (Stand: 2009-12-01).

- [Gruenstein et al. 09] Gruenstein, Alexander; Orszulak, Jarrod; Liu, Sean; Roberts, Shannon; Zabel, Jeff; Reimer, Bryan; Mehler, Bruce; Seneff, Stephanie; Glass, James & Coughlin, Joseph. 2009. „City Browser: Developing a Conversational Automotive HMI“. CHI 2009 – Spotlight on Works in Progress, Session2. CHI 2009 in Boston, MA, USA, 4. – 9. April 2009.
<http://portal.acm.org/citation.cfm?doid=1520340.1520655>
 (Stand: 2009-12-01).
- [Hamacher et al. 02] Hamacher, Nico; Kraiss, Karl-Friedrich & Marrenbach, Jörg. 2002. „Einsatz formaler Methoden zur Evaluierung der Gebrauchsfähigkeit interaktiver Geräte“. In: it + ti – Informationstechnik und Technische Informatik 44, 1/2002 (pp. 49-55).
- [Hassenzahl et al. 09] Hassenzahl, Marc; Burmester, Michael & Koller, Franz. 2009. AttrakDiff – Ein Service der User Interface Design GmbH. <http://www.attrakdiff.de/>
 (Stand: 2009-12-01).
- [HCI KN 07] Lehrstuhl für Mensch-Computer Interaktion, Universität Konstanz. Projekt „ReSpECT“. <http://hci.uni-konstanz.de/index.php?a=research&b=projects&c=16078266>
 (Stand: 2009-12-01).
- [Hohler 07] Hohler, Bernd. 2007. „Qualitätsmanagement bei der Software-Entwicklung“. In: T. Pfeifer, R. Schmitt & W. Masing: Masing Handbuch Qualitätsmanagement (pp. 817-845). Carl Hanser Verlag (München). ISBN-13: 978-3-446-40752-7.
- [Hussy & Jain 02] Hussy, Walter & Jain, Anita. 2002. „Experimentelle Hypothesenprüfung in der Psychologie“. Hogrefe-Verlag (Göttingen). ISBN: 3-8017-1627-9.
- [ISO 9241-9] International Organization for Standardization. 2000. Ergonomic requirements for office work with visual display terminals. Part 9 – Requirements for non-keyboard input devices.
- [ISO 9241-11] International Organization for Standardization. 1998. Ergonomische Anforderungen für Bürotätigkeiten mit Bildschirmgeräten. Teil 11 – Anforderungen an die Gebrauchstauglichkeit.

- [ISO 9241-110] International Organization for Standardization. 2006. Ergonomische Anforderungen für Bürotätigkeiten mit Bildschirmgeräten. Teil 110 – Grundsätze der Dialoggestaltung.
- [Kern & Schmidt 07] Kern, Dagmar & Schmidt, Albrecht. 2007. „Gas station flash survey – A Method for interviewing drivers“. In: Mensch & Computer 2007 Workshopband. Workshop on Automotive UIs in Weimar, September 2007 (pp. 41-46). <http://www.pervasive.wiwi.uni-due.de/forschung/publikationen/gas-station-flash-survey-aeur-a-method-for-interviewing-drivers-940/> (Stand: 2009-12-01).
- [Kern et al. 07] Kern, Dagmar; Schmidt, Albrecht; Pitz, Michael & Bengler, Klaus. 2007. „Status- und Kontextinformationen für die Telekommunikation im Auto“. In: Mensch & Computer 2007 (pp. 119-128). Oldenburg-Verlag (Weimar). <http://www.pervasive.wiwi.uni-due.de/forschung/publikationen/status-und-kontextinformationen-fa-1-4-r-die-telekommunikation-im-auto-942/> (Stand: 2009-12-01).
- [Kern et al. 08] Kern, Dagmar; Müller, Marco; Schneegaß, Stefan; Wolejko-Wolejszo, Lukasz & Schmidt, Albrecht. 2008. „CARS – Configurable Automotive Research Simulator“. In: Automotive User Interfaces and Interactive Applications – AUIIA 08. Workshop at Mensch und Computer 2008, Lübeck, Germany. <http://www.pervasive.wiwi.uni-due.de/forschung/publikationen/cars-aeur-configurable-automotive-research-simulator-5379/> (Stand: 2009-12-01).
- [Kern et al. 09] Kern, Dagmar; Schmidt, Albrecht; Arnsmann, Jonas; Appelman, Thorsten; Pararasasegaran, Nillakshi & Piepiera, Benjamin. 2009. „Writing to Your Car: Handwritten Text Input While Driving“. CHI 2009 – Spotlight on Works in Progress, Session 2. CHI 2009 in Boston, MA, USA, 4. – 9. April 2009. <http://portal.acm.org/citation.cfm?doid=1520340.1520724> (Stand: 2009-12-01).
- [Kirakowski 00] Kirakowski, Jurek. 2000. “Questionnaires in Usability Engineering – A List of Frequently Asked Questions” (3rd Edition). 2. Juni 2000. <http://www.ucc.ie/hfrg/resources/qfaq1.html> (Stand: 2009-12-01).

- [Kitzinger 94] Kitzinger, Jenny. 1994. „The methodology of Focus Groups: the importance of interaction between research participants”. In: *Sociology of Health & Illness* Vol. 16 No. 1 (pp. 103-121). Basil Blackwell Ltd. (Oxford, England). ISSN: 0141-9889.
- [Kuhn 05] Kuhn, Friedemann. 2005. „Methode zur Bewertung der Fahrerablenkung durch Fahrerinformations-Systeme“. World Usability Day 2005 (München). http://www.gui-design.de/download/wud_LCT_2005-11_Stuttgart.pdf (Stand: 2009-12-01).
- [Lange et al. 06] Lange, Christian; Wohlfarter, Martin & Bubb, Heiner. 2006. „Vergleichbarkeit von Usability Labor und Realversuch zur Bestimmung der ergonomischen Güte und der Ablenkungswirkung von Nebenaufgaben im Kfz“. In: *Proceedings GfA Gesellschaft für Arbeitswissenschaften Kongress, Stuttgart 2006*.
- [Lindgaard 94] Lindgaard, Gitte. 1994. „Usability Testing and System Evaluation – A guide for designing useful computer systems”. Chapman & Hall (London, UK). ISBN: 0-412-46100-5.
- [Minin et al. 08] Minin, Luca; Montanari, Roberto; Corbelli, Cesare & Iani, Cristina. 2008. „Force Feedback: New Frontiers as the Innovative Driving Comfort Tool“. *CHI 2008 Proceedings – Case Studies*. CHI 2008 in Florenz, Italien, 5. – 10. April 2008. <http://portal.acm.org/citation.cfm?id=1358628.1358647> (Stand: 2009-12-01).
- [Morgan 97] Morgan, David L. 1997. „Focus Groups as Qualitative Research“ (2nd Edition) (pp. 42-44). Sage Publications, Inc (Thousand Oaks, California). ISBN: 0-7619-0343-7.
- [MT 08] Motor Talk. 2008. „BMW iDrive: Das Original setzt neue Maßstäbe“. 3. November 2008. <http://www.motor-talk.de/news/bmw-idrive-das-original-setzt-neue-massstaebe-t2033824.html> (Stand: 2009-12-01).
- [Müller & Fiedland 09] Müller, Christian & Fiedland, Gerald. 2009. „Multimodal Interfaces for Automotive Applications (MIAA)“. *IUI 2009 in Sanibel Island, Florida*,

USA, 8. – 11. Februar 2009. <http://portal.acm.org/citation.cfm?id=1502732>
(Stand: 2009-12-01).

- [Nielsen 90] Nielsen, Jakob. 1990. „Heuristic Evaluation of User Interfaces“. In: Proceedings of the SIGCHI conference on Human factors in computing systems: Empowering people (pp. 249-256). Conference on Human Factors in Computing Systems in Seattle, Washington, USA, April 1990. ISBN:0-201-50932-6. <http://portal.acm.org/citation.cfm?id=97281> (Stand: 2009-12-01).
- [Nielsen 94] Nielsen, Jakob. 1994. „Ten Usability Heuristics“. ISSN 1548-5552. http://www.useit.com/papers/heuristic/heuristic_list.html (Stand: 2009-12-01).
- [Nielsen 97] Nielsen, Jakob. 1997. „The Use and Misuse of Focus Groups“. <http://www.useit.com/papers/focusgroups.html> (Stand: 2009-12-01).
- [Nielsen 00] Nielsen, Jakob. 2000. „Why You Only Need to Test with 5 Users“. ISSN 1548-5552. <http://www.useit.com/alertbox/20000319.html>
(Stand: 2009-12-01).
- [Norman 05] Norman, Donald A. 2005. „Human-centered design considered harmful“. In: Interactions Vol.12 Issue 4 (pp. 14-19). ACM (New York, NY, USA). ISSN: 1072-5520. <http://portal.acm.org/citation.cfm?id=1070976>
(Stand: 2009-12-01).
- [Noyes & Baber 99] Noyes, Janet M. & Baber, Chris. 1999. „User-Centered Design of Systems“. Springer-Verlag (London, UK). ISBN: 3-540-76007-5.
- [Pataki et al. 05] Pataki, Krisztin; Schulze Kissing, Dirk; Mahlke, Sascha & Thüring, Manfred. 2005. „Anwendung von Usability-Maßen zur Nutzeneinschätzung von Fahrerassistenzsystemen“. In K. Karrer, B. Gauss, B. & Ch. Steffens: Beiträge zur Mensch-Maschine-Systemtechnik aus Forschung und Praxis (pp. 211-228). Symposium Publishing (Düsseldorf). ISBN-13: 978-3936608724.
- [QUIS 09] Human-Computer Interaction Lab (HCIL) der Universität von Maryland. „Questionnaire for User Interaction Satisfaction“. <http://lap.umd.edu/quis/>
(Stand: 2009-12-01).

- [Rudolf & Kuhlisch 08] Rudolf, Matthias & Kuhlisch, Wiltrud. 2008. „Biostatistik: Eine Einführung für Biowissenschaftler“ (1. Auflage). Pearson Studium. ISBN: 978-3-8273-7269-7.
- [Sharp et al. 07] Sharp, Helen; Rogers, Yvonne & Preece, Jenny. 2007. „Interaction Design: Beyond Human-Computer Interaction“ (2nd Edition). Wiley & Sons Ltd (England). ISBN: 978-0-470-01866-8.
- [Söter 08] Söter, Andrea. 2008. „Anforderungsanalyse und konzeptionelles Design für ein Rear-Seat Entertainment System der Volkswagen AG“. Hausarbeit zum Masterprojekt, Universität Konstanz.
- [Specht 10] Specht, Marcus. 2010. Arbeitstitel: „Benutzerorientierte Entwicklung eines Webbrowsers mit multimodalen Interaktionstechniken für Fahrzeuge der Volkswagen AG“. Bachelorarbeit, Universität Konstanz.
- [SUMI 09] Human Factors Research Group der Universität Cork. „Software Usability Measurement Inventory“. <http://sumi.ucc.ie/> (Stand: 2009-12-01).
- [StVO 09] Bundesministerium für Verkehr, Bau und Stadtentwicklung. 2009. „Straßenverkehrs-Ordnung“. <http://www.bmvbs.de/Verkehr/Strasse-1449/Strassenverkehrs-Ordnung.htm> (Stand: 2009-12-01).
- [Technofile 08] Bray, Jim. 2008. „Audi A6 V8 - Another Seductive Sedan“. In: Technofile. http://www.technofile.com/cars/audi_a6.asp (Stand: 2009-12-01).
- [Urbas & Leuchter 08] Urbas, Leon & Leuchter Sandro. 2008. „Usability Engineering of ‚In Vehicle Information Systems‘ with Multi-Tasking GOMS“. In: Proceedings Dachkongress Patientensicherheit und Medizintechnik. APS 2008 – 3. Jahrestagung des Aktionsbündnis Patientensicherheit e.V. und MEK 2008 – 5. Medizintechnik und Ergonomiekongress. VDE [CD-ROM].
- [Vöhringer-Kuhnt et al. 08] Vöhringer-Kuhnt,Thomas; Rößger, Peter & Kiss, Miklós. 2008. „CarUSE – praktikable Maße zur HM-Evaluation im Fahrzeug: Konzeptionelle Entwicklung eines Bewertungsinstrumentariums“. In: Useware 2008, VDI-Berichte Nr. 2041 (pp. 67-78). VDI-Verlag GmbH (Düsseldorf). ISBN: 978-3-18-092041-2.

- [VW 09] Volkswagen AG . 2009. „Mitmachen motiviert“. In: „autogramm“ – Die Zeitung für die Mitarbeiterinnen und Mitarbeiter der Marke Volkswagen. Ausgabe 8/2009.
http://autogramm.volkswagen.de/08_09/aktuell/aktuell_04.html
 (Stand: 2009-12-01).
- [Wäller et al. 08] Wäller, Christoph; Bendewald, Lennart & Dr. Oel, Peter. 2008. „Optimierung der automotiven Touchscreen-Bedienung durch expandierende Ziele“. In: Useware 2008, VDI-Berichte Nr. 2041 (pp. 29-42). VDI-Verlag GmbH (Düsseldorf). ISBN: 978-3-18-092041-2.
- [Wiklund 94] Wiklund, Michael E. 1994. „Usability In Practice – How Companies develop User-Friendly Products“ (pp. 1-19). Academic Press, Inc.
 ISBN: 0-12-751250-0.
- [Willumeit 97] Willumeit, Heinz. 1997. „IsoMetrics^S - Fragebogen zur Evaluation von graphischen Benutzungsschnittstellen“ (Version 2.01).
<http://www.isometrics.uni-osnabrueck.de/> (Stand: 2009-12-01).
- [Zimmermann et al. 07] Zimmermann, Dirk; Grötzbach, Lennart & Freymann, Maximiane. 2007. „Ansatz zur nutzerzentrierten Requirement-Analyse und Evaluation: Ein Framework-Entwurf“. In: K. Röse & H. Brau: Usability Professionals 2007: Berichtband des fünften Workshops des German Chapters der Usability Professionals Association e.V. Fraunhofer IRB Verlag. ISBN 978-3-8167-7414-3. http://www.lennartgroetzbach.de/downloads/UP07_DZLGMF.pdf
 (Stand: 2009-12-01).

7. Abbildungsverzeichnis

Abbildung 1.01: Das allgemeine Verhältnis zwischen Gebrauchstauglichkeit und Komplexität (Punkte entsprechen Produkten). (Quelle: [Wiklund 94]).....	7
Abbildung 1.02: Audi Q7 (Quelle: [Audi 09a]) mit Screen für das RSE (rot) und potentiellm Benutzer des RSE.....	9
Abbildung 2.01: Nutzungskontext Schreibtisch mit Desktop-PC.	11
Abbildung 2.02: Nutzungskontext Fahrzeug mit Infotainment-/Rear-Seat-Entertainmentsystem.	11
Abbildung 3.01: Beispielhafte, tabellarische Darstellung einer Persona (Quelle: [Söter 08]).	25
Abbildung 3.02: Entwicklung des Probandenpools seit seiner Gründung 2004 (Quelle: [VW 09])...	33
Abbildung 3.03: Teilnehmer einer Fokusgruppe bei der Diskussion über automotive Interaktionskonzepte.	36
Abbildung 3.04: Einfacher Lifecycle des Interaktionsdesign nach [Sharp et al. 07]	44
Abbildung 3.05: Lo-Fi Fahrsimulator mit geringer physischer und funktionaler Wiedergabetreue (Quelle: [Kern et al. 08])......	62
Abbildung 3.06: Medium-Fi Fahrsimulator mit mittlerer physischer Wiedergabetreue (Quelle: [Bach et al. 08]).....	62
Abbildung 3.07: Hi-Fi Fahrsimulator mit 360°-Ansicht. Blick durch die Frontscheibe (Quelle: [MT 08]).....	63
Abbildung 3.08: Hi-Fi Fahrsimulator mit 360°-Ansicht und Rückspiegel (Quelle: [MT 08]).....	63
Abbildung 3.09: Vier Videokameras erfassen im Simulatorkontext verschiedene Aspekte der Interaktion (Quelle: [Bach et al. 08]).	65
Abbildung 3.10: Gute Spurhaltung = Kleine Fläche (Quelle: [Kuhn 05]).	66
Abbildung 3.11: Schlechte Spurhaltung = große Fläche (Quelle: [Kuhn 05]).	66
Abbildung 3.12: Kontrolliertes Fahren auf einer Übungsstrecke (Quelle: [Bach et al. 08])......	68
Abbildung 4.01: Einfacher Lifecycle des Interaktionsdesign nach [Sharp et al. 07]	76
Abbildung 4.02: Modell des Usage-Centered Design nach Constantine und Lockwood [Constantine & Lockwood 99].....	77
Abbildung 4.03: Das MMI-Terminal als Eingabegerät für das Audi MMI-System (Quelle: [Technofile 08]).	79
Abbildung 4.04: Bildschirmaufbau des Audi MMI-Systems (Quelle: [Audi 08]).	79
Abbildung 4.05: Low-Fidelity Prototyp des grundlegenden Browserkonzeptes.....	79
Abbildung 4.06: High-Fidelity Prototyp von Browservariante B.....	79
Abbildung 4.07: Der Morae® Manager bei der Auswertung einer Aufzeichnung.	83
Abbildung 4.08: Ablauf der Benutzertests der explorativen Evaluationsstudie.	84
Abbildung 4.09: Hauptansicht des Browsers der ersten Generation.	94
Abbildung 4.10: Hauptansicht des finalen Browsers.	94

Abbildung 4.11: Vertikales Scrollen einer Webseite auf Zoomstufe 1.9 mit halbtransparenter Vorschau im rechten Bildschirmbereich.....	94
Abbildung 4.12: Altes MMI-Terminal mit Joystick zur Mauszeigersteuerung.....	95
Abbildung 4.13: Neues MMI-Terminal mit Schiebemöglichkeit zur Mauszeigersteuerung.	95
Abbildung 4.14: Von der Bewegungsdauer und Umgebung abhängige Be- und Entschleunigung des Mauszeigers.	96
Abbildung 4.15: Dynamische Kraftfelder eines interaktiven Elements (IE).....	96
Abbildung 4.16: Eindirektionaler Tapping-Test (Quelle: [ISO 9241-9]).....	98
Abbildung 4.17: Multidirektionaler Tapping-Test (Quelle: [ISO 9241-9]).	98
Abbildung 4.18: Webseite für den Tapping-Test.	99
Abbildung 4.19: Herkunft der Elemente der Webseite für den Tapping-Test.....	99
Abbildung 4.20: Testablauf bei der Evaluationsstudie im Juli / August 2009.	101
Abbildung 4.21: Versuchsstrecke für die mobilen Tests im August 2009.	103
Abbildung 4.22: Mittelwerte und Streuung der benötigten Zeit pro Trial im stationären Test.	107
Abbildung 4.23: Mittelwerte und Streuung der vollführten Overshoots pro Trial im stationären Test.	107
Abbildung 4.24: Ideale Distanz zwischen Start- und Zielelement eines Trials.....	108
Abbildung 4.25: Mittelwerte und Streuung der zurückgelegten Strecke pro Trial im mobilen Test.	111
Abbildung 4.26: Mittelwerte und Streuung der Overshoots pro Trial im mobilen Test.....	112
Abbildung 4.27: Darstellung der Softkeys in den Bildschirmecken vor Interaktion durch Benutzer.	116

8. Tabellenverzeichnis

Tabelle 2.01: Unterschiede der Nutzungskontexte Desktop-PC und Automobil	12
Tabelle 3.01: Qualitätsmerkmale für Software nach ISO/IEC 9126-1 (Quelle: [Hohler 07]).....	30
Tabelle 4.01: Ergebnisse des Signifikanztests für den stationären Tapping-Test mit der Basis- und der optimierten Variante.	106
Tabelle 4.02: Beispielrechnung bei der Normalisierung von Block 4 und 5 des stationären Tests. (Basisvariante bei Proband #1, jeweils durchschnittliche Werte pro Trial).....	108
Tabelle 4.03: Beispielrechnung bei der Normalisierung der Zeit für den Vergleich der letzten beiden stationären Blöcke mit dem Baseline-Block (Daten von Proband #1).	109
Tabelle 4.04: Ergebnisse des Signifikanztests für den Tapping-Test mit der Basisvariante im Stand und während der Fahrt.	110
Tabelle 4.05: Ergebnisse des Signifikanztests für den mobilen Tapping-Test mit der Basis- und der optimierten Variante.	111

9. Anhang

Auf der DVD befinden sich folgende Materialien:

- Fragebogen zum Thema Benutzerbefragungen, Benutzungsstudien und Benutzbarkeitstests für Mitglieder der Automobilindustrie.

- 1. Evaluationsstudie (Januar 2009)
 - o Testmaterial:
 - Willkommensschreiben, Einverständniserklärung, etc.
 - Pre Test Fragebogen (demografische Daten).
 - Testaufgaben (stationär & mobil).
 - Interviewfragen & Template für die Antworten.
 - o Auswertung:
 - Tabellarische Zusammenfassung der demografischen Daten der Probanden.
 - Zusammenfassung der Auffälligkeiten je Proband.
 - Tabellarische Zusammenfassung aller Auffälligkeiten & Usability Probleme.
 - o Beispielvideo eines Tests.

- 2. Evaluationsstudie (Juli / August 2009)
 - o Testmaterial:
 - Willkommensschreiben, Einverständniserklärung, Einverständniserklärung der Eltern, etc.
 - Pre Test Fragebogen (demografische Daten).
 - Testaufgaben (stationär & mobil) inklusive Antworten.
 - Interviewfragen.
 - Post Test Fragebogen.
 - o Auswertung:
 - Tabellarische Zusammenfassung der demografischen Daten der Probanden.
 - Logdaten.
 - Qualitative Auswertung: Auffälligkeiten bei jedem Probanden & Tabellarische Zusammenfassung der Auffälligkeiten und Probleme.
 - o Webseite für den Tapping Test.
 - o Beispielvideo eines Tests.

