

Fachgebiet Verteilte Messsysteme
Fakultät für Elektrotechnik und Informationstechnik
Technische Universität München

**Konzeption und Implementierung
eines adaptiven Touchscreen-Interfaces
für das Fahrzeug**

Katharina Susanne Bachfischer

Vollständiger Abdruck der von der Fakultät für
Elektrotechnik und Informationstechnik der Technischen
Universität München zur Erlangung des akademischen
Grades eines

Doktor-Ingenieurs

genehmigten Dissertation.

Vorsitzender: Univ.-Prof. Dr.-Ing. habil. G. Rigoll

Prüfer der Dissertation:

1. Univ.-Prof. Dr.-Ing. F. Puente León
2. Univ.-Prof. Dr. rer. nat. H. Bubb

Die Dissertation wurde am 23.04.2008 bei der Technischen
Universität München eingereicht und durch die Fakultät für
Elektrotechnik und Informationstechnik am 29.07.2008
angenommen.

Konzeption und
Implementierung eines
adaptiven
Touchscreen-Interfaces für das
Fahrzeug

Katharina Susanne Bachfischer

23.04.2008

Inhaltsverzeichnis

Vorwort	V
Kurzzusammenfassung	VII
Abstract	IX
1 Einleitung	1
1.1 Motivation	1
1.2 Zielsetzung	3
1.3 Aufbau der Arbeit	4
2 Thematische Einführung	5
2.1 Grundlegende Begriffsterminologie	5
2.2 Technische Kommunikation	9
2.3 Humankommunikation	11
2.4 Ein- und Ausgabekanäle	14
2.5 Multimodale Mensch-Maschine-Systeme	15
2.5.1 Potenzial multimodaler Schnittstellen	15
2.5.2 Limitierungen multimodaler Schnittstellen	15
2.5.3 Klassifizierung multimodaler Systeme	16
2.5.4 Informationsfusion in multimodalen Systemen	20
2.6 Adaptive und adaptierbare Systeme	22
2.6.1 Benutzermodelle	23
2.6.2 Mentale Modelle	25
2.7 Gestaltungsaspekte von Mensch-Maschine-Systemen	26
2.8 Mensch-Maschine-Kommunikation im Kraftfahrzeug	28
2.8.1 Bediensituation und Aufgabenklassifizierung im Fahrzeug	28
2.8.2 Spezifische Anforderungen an Kfz-Systeme	29
2.8.3 Serienstand	30
2.8.4 Infotainmentsysteme im Forschungsstadium	32

3	Herleitung eines erweiterten MMK-Modells für Anzeige und Bedienung im Kraftfahrzeug	35
3.1	Übertragung von Humankommunikationsprinzipien auf die Mensch-Maschine-Kommunikation	36
3.1.1	Emotionale Signalisierung	36
3.1.2	Multimodale Kommunikation	38
3.1.3	Adaptive Kommunikation	39
3.2	Dialogmodell eines synergistisch-multimodalen Systems .	40
3.3	Prototypisch implementiertes System	45
4	Bestimmung der Position einer Hand vor einem Display	47
4.1	Motivation zum Einsatz kapazitiver Sensorik	47
4.1.1	Anforderungen	47
4.1.2	Kamerabasierte Gestenerkennung	48
4.1.3	Infrarotbasierte Gestenerkennung	49
4.1.4	Kapazitive Gestenerkennung	50
4.2	Kapazitive Sensorik	50
4.2.1	Theorie der kapazitiven Abstandsmessung	50
4.2.2	Anwendungsmodi kapazitiver Sensoren	53
4.2.3	Grenzen kapazitiver Sensorik	56
4.2.4	Qualitative Beschreibung des kapazitiven Sensorsystems	57
4.3	Positionsermittlung der Hand vor einem Display	60
4.3.1	Statistische Modellierung der Sensoren	60
4.3.2	Maximum-Likelihood-Schätzung	61
5	Methodik und Verfahren zur Fusion multimodaler Eingaben	65
5.1	Motivation zum Einsatz einer Soft-Decision-Fusionsmethode	65
5.1.1	Anforderungen	65
5.1.2	Multimodale Fusionsmethoden	67
5.2	Integrationsmethodik im Modalitätsmanager auf semantischer Ebene	68
5.2.1	Integrationsverfahren	69
5.2.2	Hypothesengenerierung und -verarbeitung	70
5.2.3	Fehlermanagement	72

6	Realisierung eines fahrzeugtauglichen Systems zur synergistisch-multimodalen Eingabe	73
6.1	Systemaufbau	73
6.2	Implementierte Anwendungsszenarien	75
6.2.1	Einblendung und Vergrößerung von Schaltflächen bei Annäherung	75
6.2.2	Interaktion durch räumliche Spezifikationen	78
6.3	Handpositionserkennung	81
6.3.1	Verwendete Hardware	81
6.3.2	Personenspezifische Kennlinienanpassung	82
7	Evaluierung der Annäherungsfunktion	85
7.1	Positionsschätzung der Hand vor einem Display	85
7.1.1	Konzeption	85
7.1.2	Parametrierung und Durchführung der Messung	88
7.1.3	Ergebnisse	90
7.2	Studie zur subjektiven Bewertung der Annäherungsfunktion	91
7.2.1	Konzeption	91
7.2.2	Ergebnisse	95
7.2.3	Interpretation und Fazit	97
7.3	Expertenstudie zur Evaluierung der Bedienbarkeit	99
7.3.1	Ergonomische Betrachtungen	99
7.3.2	Fragestellung und Zielsetzung der Studie	102
7.3.3	Konzeption	103
7.3.4	Messgrößen und Datenaufbereitung	105
7.3.5	Ergebnisse	108
7.3.6	Interpretation und Fazit	120
7.4	Studie zur Blickbewegung bei der Bedienung eines annäherungssensitiven Touchscreens	122
7.4.1	Fragestellung und Zielsetzung der Studie	122
7.4.2	Konzeption	123
7.4.3	Messgrößen und Datenaufbereitung	124
7.4.4	Ergebnisse	125
7.4.5	Interpretation und Fazit	128
7.5	Bedienbarkeit für Erstnutzer	131
7.5.1	Konzeption	131
7.5.2	Ergebnisse	133
7.5.3	Fazit	134

7.6 Zusammenfassung der Ergebnisse aus den einzelnen Studien 134

8	Evaluierung der Fusion multimodaler Eingaben	137
8.1	Basisuntersuchung zur subjektiven Bewertung von multimodalen Eingaben	137
8.1.1	Konzeption	137
8.1.2	Ergebnisse	138
8.1.3	Interpretation und Fazit	139
8.2	Evaluierung der Bedienbarkeit	139
8.2.1	Fragestellung und Zielsetzung der Studie	139
8.2.2	Konzeption	140
8.2.3	Ergebnisse	143
8.2.4	Interpretation und Fazit	149
9	Zusammenfassung und Ausblick	153
	Tabellenverzeichnis	157
	Abbildungsverzeichnis	159
	Abkürzungsverzeichnis	165
	Literaturverzeichnis	166

Vorwort

Die vorliegende Arbeit entstand während meiner Tätigkeit in der Konzernforschung der Volkswagen AG in Wolfsburg¹. Sie wurde in enger Kooperation mit dem Fachgebiet Verteilte Messsysteme der Technischen Universität München angefertigt.

An dieser Stelle möchte ich mich ganz herzlich bei Herrn Prof. F. Puente León für die Unterstützung der Arbeit, die universitäre Betreuung und die vielen wertvollen Hinweise bei der Durchführung bedanken. Des Weiteren gilt mein besonderer Dank Jörg Lilienthal und Dr. Peter Oel für die fachliche Begleitung bei der Volkswagen AG und die Anregungen, die sie als Diskussionspartner in die Arbeit eingebracht haben.

Daneben bedanke ich mich recht herzlich bei allen Mitarbeitern des Teams *Bedienkonzepte und Fahrer* für die gute Zusammenarbeit. An dieser Stelle richte ich meinen Dank auch an die anderen Kollegen, Diplomanden, Praktikanten und alle weiteren Personen, die zum Gelingen dieser Arbeit beigetragen haben.

Schließlich danke ich ganz besonders Andre Peterwerth und meinen Eltern für ihre beständige Unterstützung. Sie haben dadurch maßgeblich zum Erfolg der Arbeit beigetragen.

Wolfsburg, im April 2008

Katharina Bachfischer

¹Die Ergebnisse, Meinungen und Schlüsse dieser Dissertation sind jedoch nicht notwendigerweise die der Volkswagen AG.

Kurzzusammenfassung

Die vorliegende Arbeit umfasst die Konzeptionierung, Implementierung und Evaluierung eines adaptiven Touchscreen-Interfaces im Kraftfahrzeug. Von besonderem Interesse ist hierbei die implizite Erkennung und Nutzung der Annäherung einer Hand an einen in der Mittelkonsole eines Fahrzeugs verbauten Touchscreen. Dieser wurde durch die Anbindung eines Spracherkenners zu einer multimodalen Bedieneinheit erweitert. Dies erfolgt in der Absicht, dem Fahrer (oder dem Beifahrer), die Bedienung – v. a. in Fahrzeugen der kleineren Klassen – trotz der stetig steigenden Anzahl an Fahrerinformations- und Fahrerassistenzfunktionen zu erleichtern oder überhaupt erst zu ermöglichen. Ein wichtiger Aspekt ist dabei die Orientierung an Prinzipien, die der Nutzer aus der Humankommunikation kennt.

Die zugrunde liegende Idee ist die Interpretation der Annäherung einer Hand als Bedienintention: Sobald eine Bedienintention erkannt wird, wechselt die Anzeige von einem Anzeige- in einen Bedienmodus. Für die multimodale Interaktion wird die Annäherung zur impliziten Aktivierung des Spracherkenners genutzt. Darauf aufbauend werden verschiedene Szenarien implementiert, die u. a. synergistisch-multimodale Eingaben erlauben.

Ein Schwerpunkt der Arbeit ist die Entwicklung eines Algorithmus zur dreidimensionalen Erkennung der Hand oder eines Bedienfingers vor dem Touchscreen-Display unter Berücksichtigung der restriktiven Randbedingungen für den Einsatz im Fahrzeug: berührungslose Detektion, Echtzeitfähigkeit, hohe Robustheit, designneutraler Verbau sowie nicht zuletzt Wirtschaftlichkeit in Bezug auf Teilekosten und Rechenleistung. Im Gegensatz zu vielen anderen Ansätzen, die meist videobasiert arbeiten, baut der hier entwickelte Ansatz auf einer kapazitiven Sensor-Lösung auf: Aus vier Sensoren, die an den Ecken des Displays angebracht sind, wird mittels einer Maximum-Likelihood-Schätzung die Position der Hand detektiert. Eine Überprüfung des Algorithmus zeigt, dass sich in einem Abstand von 4 cm ein Cursor auf dem Bildschirm noch immer so steuern

lässt, dass die Abweichung vom Schnittpunkt eines Lots von der Fingerspitze auf die Displayoberfläche maximal 1 cm beträgt.

Ein weiterer Schwerpunkt ist die Evaluierung des Systems, die viel versprechende Ergebnisse liefert: Das adaptive Touchscreen-Interface ist weitgehend intuitiv bedienbar und findet eine hohe Akzeptanz. Darüber hinaus unterscheiden sich die Bewegungszeiten und Blickabwendungszeiten bei der Bedienung im adaptiven Modus während der Fahrt im Simulator nicht von denen eines statischen Touchscreens. Es können jedoch besonders bei kleinen Displays geringere Fehlerraten erzielt werden.

Abstract

The work presented in this thesis comprises the design, implementation and evaluation of an adaptive in-car touchscreen interface. Special interest is dedicated to the implicit detection of the proximity of a hand to a touchscreen located in the center console of a vehicle. The setup of the user interface was expanded to a multimodal control unit by connecting a speech recognizer. That is motivated by the aim to facilitate the driver's or co-driver's interaction in the context of a rising number of driver information and assistance systems especially in smaller vehicles. An important aspect is the orientation towards principles of interaction which the user knows well from human to human communication.

The underlying idea is the interpretation of the proximity of a hand as an interaction intention: As soon as an interaction intention is detected, the display changes from an information mode to an interaction mode. Regarding the multimodal interaction, proximity is used to activate the speech recognizer. Building on that, different scenarios are implemented which allow for synergistic-multimodal input and also other variants of multimodal use.

One main focus is the design of an algorithm to three-dimensionally detect the hand or the interaction finger with respect to the restrictive constraints of an in-car system namely contactless detection, real-time capability, high robustness, design-neutral integration and cost effectiveness with regard to partial costs and processing power. In contrast to many other systems which work mostly camera-based, the approach presented here is based on a capacitive sensor system: the position of the hand is inferred via a maximum likelihood estimation by four sensors positioned near the edges of the display. The evaluation of the algorithm shows that in a distance of 4 cm a cursor on the display can be navigated in a way that the perpendicular from the finger tip to the display maximally differs 1 cm from the cursor.

Another focus is the evaluation of the whole system, which gave promising results: the adaptive touchscreen-interface is largely intuitive and

received positive feedback in user acceptance ratings. Moreover, movement duration and eyes-off-road times while interacting with the system in the adaptive mode do not differ from those of a static touchscreen. However, in comparative studies with smaller displays, the error rates of the adaptive mode could be reduced even further when compared to alternative setups.

1 Einleitung

„Was man im Großen nicht kann, soll man im
Kleinen nicht unversucht lassen.“
(Adolph Kolping)

1.1 Motivation

Die Mobilität der Menschen ist stark zunehmend – immer mehr Menschen sind immer länger unterwegs. Daraus lassen sich zwei gegensätzliche Forderungen ableiten: Einerseits soll der Fahrkomfort steigen und so ‚Unterwegs Sein‘ angenehm und erstrebenswert werden. In diesem Zusammenhang und im Kontext stärkerer Vernetzung werden zunehmend mehr Funktionen in das HMI (Human Machine Interface) im Kraftfahrzeug (Kfz) integriert. Andererseits soll jedoch der Fahrer, der aufgrund des steigenden Verkehrsaufkommens mit der Fahraufgabe stärker belastet wird, möglichst wenig durch Anzeige und Bedienung des HMIs abgelenkt werden. Die Bedienung des HMIs soll also möglichst einfach und verständlich sein, der unbewussten Erwartungshaltung des Nutzers entsprechen, kurz intuitiv sein.

Bild 1.1 zeigt den Anstieg der vom Fahrer bedienbaren Funktionen in einem Fahrzeug des Premiumsegments. In heutigen Modellen stehen damit fast viereinhalb Mal so viele Funktionen wie im ersten Modell zur Verfügung. Dies ist v. a. auf die zunehmende Anzahl an Sekundärfunktionen, also Funktionen, die nicht direkt der Fahrzeugstabilisierung dienen, zurückzuführen: Komfortfunktionen wie die automatische Sitzeinstellung oder die Klimaautomatik (mit zugehöriger Bedienung), aber auch Fahrerinformations- (FIS) oder Infotainmentsysteme¹, die dem Fahrer umfangreiche Kommunikations- und Unterhaltungsfunktionen zur Verfügung stellen, wurden eingeführt. Darüber hinaus wurden Fahrerassistenzsysteme (FAS) wie ACC (automatische Abstandshaltung), GRA (automatische Geschwindigkeitsregelung) oder Navigationssysteme, die

¹engl., Kunstwort aus ‚information‘ und ‚entertainment‘

den Fahrer bei der Fahrzeugführung und Navigation unterstützen, integriert.

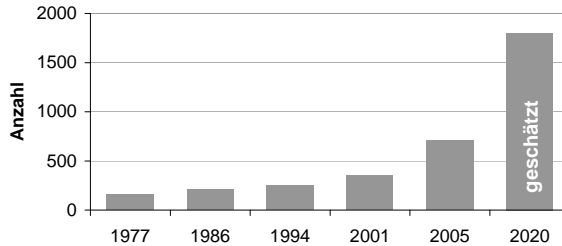


Bild 1.1: Anzahl der vom Fahrer bedienbaren Funktionen in einem Fahrzeug der Premium-Klasse (Quelle: Volkswagen).

Unter diesen Voraussetzungen gilt es, dem Fahrer eine sichere und intuitive Bedienung zur Verfügung zu stellen. Dies hat Einfluss auf die Anzahl und Art der Bedienelemente: vom klassischen komponentenorientierten Schalterkonzept wandelt sich das HMI zu einer integrierten Anzeige- und Bedienschnittstelle (siehe u.a. [71]). Informationen werden dabei auf einem zentral angebrachten hochauflösenden Display angezeigt, während ein Großteil der Bedienung über ein multifunktionales abgesetztes haptisches Element oder über Touchscreen/Touchpad erfolgt.

Die Anzahl der Bedienelemente ist dadurch zwar relativ zu den bedienbaren Funktionen stark zurückgegangen [104], solche Konzepte sind aber nicht immer einfacher zu bedienen, wie die Untersuchungen von Schatzenberg und Debus [129] zeigen. Der Bediendialog muss in der Umgebung Fahrzeug sehr genau auf die physische und kognitive Leistungsfähigkeit des Menschen in der jeweiligen Fahrsituation angepasst werden, da die durch Infotainment- und Kommunikationssysteme zusätzlich hervorgerufene sensorische und kognitive Belastung nicht über ein bestimmtes Maß hinaus gehen darf [86, 117, 145].

In einigen Studien wurde zur Unterstützung und Verbesserung des Bediendialogs die Nutzung zusätzlicher Ein- und Ausgabemodalitäten getestet. Dabei wurden unterschiedliche Kombinationen aus Gestenerkennung, Spracherkennung, Verfolgung der Kopfbewegungen, Eye-Tracking und klassischen haptischen Eingabeelementen wie Tastatur, Maus und Touchscreen untersucht. Sowohl Ergebnisse aus nicht-

automobiltechnischen Bereichen [6, 42, 68, 149, 151], als auch Studienergebnisse aus dem automobiltechnischen Umfeld [87] sind erfolgsversprechend. Auch die Ergebnisse von Schattenberg und Debus [129] (Sprache und Haptik), Geiger et al. [59] und Zobl [161] (Sprache und Gestik) sowie Neuss [103] (Sprache, Haptik und Gestik) deuten auf Vorteile bezüglich Fehleranfälligkeit, Bediengeschwindigkeit und Ablenkungseffekten von multimodalen gegenüber unimodalen Systemen hin. Niedermair [104] weist darauf hin, dass diese Vorteile aber nur dann zu Tage treten können, wenn die einzelnen Modalitäten in einem sinnvollen Verbund zusammenwirken.

Maglio et al. [91] stellen die These auf, dass zukünftige Benutzerschnittstellen sich mit Nutzeraktionen und -verhalten befassen müssen, um Bedürfnisse und zukünftige Aktionen des Nutzers antizipieren zu können und Informationen zu liefern, bevor der Nutzer sie explizit angefordert hat.

1.2 Zielsetzung

Ziel der vorliegenden Arbeit ist es, ein neues Interaktionskonzept für das Fahrzeug, basierend auf einem annäherungssensitiven Touchscreen-Interface, einzuführen und zu evaluieren. Dazu wurde zunächst ein umfassendes und skalierbares Kommunikationsmodell für das Fahrzeug entwickelt, das sich in vorteilhafter Weise an der Humankommunikation orientiert und wichtige Interaktionsprinzipien aufgreift. Im Folgenden wird das Modell am Beispiel eines adaptiven Touchscreen-Interfaces umgesetzt. Dazu wurde der Touchscreen mit Sensoren ausgestattet, um die Annäherung einer Hand an das Display und die damit verbundene Bedienintention detektieren zu können. Darüber hinaus wurde der Touchscreen mit einer Spracherkennung kombiniert, um multimodale Eingabeszenarien zu ermöglichen.

Im Gegensatz zu Forschungsarbeiten mit ähnlichen Inhalten besteht das Ziel dieser Arbeit darin, die Funktion der Annäherung bei Touchscreens als implizite Eingabemodalität einzuführen und deren Gebrauchstauglichkeit für das Fahrzeug zu evaluieren. Insbesondere der Einfluss auf Bedienzeiten und Blickabwendung wurde gezielt betrachtet.

Der Hauptschwerpunkt und die Innovation der Arbeit liegen wesentlich stärker auf der Optimierung eines annäherungssensitiven touchscreenba-

sierten Systems für das Fahrzeug als auf der grundsätzlichen Realisierung eines annäherungssensitiven Systems.

In diesem Rahmen ist die Arbeit als eine praxisorientierte, disziplinübergreifende Forschungsarbeit mit einem starken ingenieurwissenschaftlichen Anteil zu verstehen.

1.3 Aufbau der Arbeit

Zunächst wird in Kapitel 2 eine thematische Einführung gegeben, die die notwendigen Grundlagen und Voraussetzungen zum Verständnis der Arbeit legt. Nach einer Definition wichtiger Begriffe werden die Bereiche Humankommunikation und technische Kommunikation betrachtet. Darauf aufbauend werden wichtige Gesichtspunkte multimodaler Benutzerschnittstellen erörtert. Einige Aspekte zu adaptiven und adaptierbaren Systemen sowie eine Übersicht über grundlegende Gestaltungsaspekte von Mensch-Maschine-Systemen runden den Überblick ab. Abschließend werden die Serienstände von zwei Infotainment-Systemen in Fahrzeugen verglichen und bestehenden Systemen im Forschungsstadium gegenübergestellt.

In Kapitel 3 wird ein gegenüber heutigen Infotainment-Systemen erweitertes Modell einer Mensch-Maschine-Schnittstelle im Fahrzeug vorgestellt. Dieses bildet die Grundlage für die Entwicklung und Implementierung des adaptiven Touchscreen-Interfaces, für das in Kapitel 4 und 5 die Methoden zur Umsetzung dargestellt werden. Kapitel 4 beschäftigt sich dabei mit der Frage, wie eine positions aufgelöste Annäherungssensorik realisiert werden kann, und stellt eine robuste Lösung vor. Kapitel 5 hingegen zeigt ein Fusionsverfahren für Eingabebefehle, die aus Sprache und haptischer Bedienung zusammengesetzt sind, in einer speziellen Realisierung unter Einbeziehung der Annäherungsfunktion.

Kapitel 6 stellt den konkreten Aufbau und die Implementierung des adaptiven Touchscreen-Interfaces vor.

Kapitel 7 und 8 evaluieren das vorliegende System im Detail. Kapitel 7 legt den Schwerpunkt auf die Annäherungsfunktion und greift vor allem die Aspekte subjektive Bewertung, Bedienbarkeit sowie Blickabwendung und Bedienzeit auf, während Kapitel 8 die Ergebnisse einer Bewertung der multimodalen Integration vorstellt.

2 Thematische Einführung

„Wer die Vergangenheit nicht kennt, wird die
Zukunft nicht in den Griff bekommen.“
(Golo Mann)

2.1 Grundlegende Begriffsterminologie

Der Kommunikation zwischen technischen Systemen und dem Menschen als Nutzer dieser Systeme kommt aufgrund der kontinuierlichen Weiterentwicklung der Informationstechnologie und ihrer gleichzeitig zunehmenden Etablierung sowohl im beruflichen als auch im privaten Bereich eine wichtige Rolle zu. So hat sich für die Analyse, Konzeption und Evaluierung der entsprechenden Systeme und Systemschnittstellen ein eigenständiger Forschungsbereich entwickelt, an dem neben den klassischen Ingenieurwissenschaften auch die Arbeitswissenschaften, die Ergonomie, die Informatik, die Medizin und die Psychologie beteiligt sind: die Mensch-Maschine-Kommunikation. Damit eng verbunden sind die Begriffe Mensch-Maschine-System und Mensch-Maschine-Schnittstelle. Da die Ähnlichkeit der Begriffe oft zu Verwechslungen führt, soll im Folgenden eine Abgrenzung und Definition vorgenommen werden.

Mensch-Maschine-Kommunikation Analog zum Begriff Mensch-Maschine-Kommunikation (MMK) finden sich in der Literatur auch die Begriffe Mensch-Maschine-Interaktion (MMI) und Mensch-Computer-Interaktion (MCI), wenn die spezielle Bedeutung des Computers betont werden soll. Diese deutschen Bezeichnungen entsprechen den international etablierten Bezeichnungen Human-Machine-Communication (HMC), Human-Machine-Interaction (HMI¹) und Human-Computer-Interaction (HCI).

¹In dieser Arbeit wird die Abkürzung HMI jedoch für Human-Machine-Interface verwendet. Eine vollständige Übersicht der verwendeten Abkürzungen findet sich im Abkürzungsverzeichnis.

Kommunikation (lat. *communicare* „teilen, mitteilen, teilnehmen lassen; gemeinsam machen, vereinigen“) bezeichnet auf der menschlichen Alltagsebene den wechselseitigen Austausch von Gedanken in Sprache, Gestik, Mimik, Schrift oder Bild. Im erweiterten Sinn ist Kommunikation das wechselseitige Übermitteln von Daten oder von Signalen, die einen festgelegten Bedeutungsinhalt haben, auch zwischen Menschen und technischen Objekten oder Systemen. Der Begriff ist eng verwandt mit dem der Interaktion. In vielen Bereichen sind diese Begriffe sogar synonym, besonders dann, wenn Wechselseitigkeit für den Kommunikationsbegriff vorausgesetzt wird. Nach Paetau [114] ist unter dem formalen Aspekt die Interaktion gegenüber der Kommunikation der allgemeinere Begriff, dessen Bestimmungsstück die Bezogenheit des Handelns von zwei oder mehreren Personen ist. Die für diesen Wechselwirkungsprozess notwendigen Abstimmungen müssten nicht ausdrücklich wie in der Kommunikation geschehen. Bedeutsam sei allein, die Intention des Anderen zu erfassen, um das eigene Handeln daran auszurichten. Auch Seifert [134] verweist darauf, dass für koordinierte Handlungen zwischen Interaktionspartnern die Intentionen abzustimmen sind und eine Übereinstimmung über das Handlungsziel zu treffen ist. Allgemein besteht in der Literatur jedoch keine absolute Einigkeit bezüglich der genauen Definition und Abgrenzung der Begriffe Mensch-Maschine-Interaktion und -Kommunikation. In dieser Arbeit werden sie daher synonym verwendet.

Eine viel zitierte Definition für HCI stammt von der ACM Special Interest Group on Computer-Human Interaction [2]:

„Human-computer interaction is a discipline concerned with the design, evaluation and implementation of interactive computing systems for human use and with the study of major phenomena surrounding them.“

Mensch-Maschine-Kommunikation beruht also auf der synergetischen Kombination aus Technik, Design und Ergonomie [104] und ist keiner einzelnen Disziplin zuzuordnen (siehe Bild 2.1). Nur so können die technischen Rahmenbedingungen mit den Anforderungen des Designs und des Benutzers in einer Art und Weise fusioniert werden, dass ein funktionsfähiges, ästhetisch ansprechendes und gleichzeitig bedienbares Produkt entsteht [104].

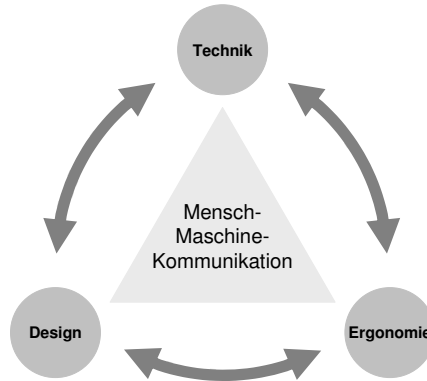


Bild 2.1: Interdisziplinärer Charakter der Mensch-Maschine-Kommunikation [104].

Mensch-Maschine-System Das Mensch-Maschine-System² unterscheidet sich von der übergeordneten Bezeichnung Mensch-Maschine-Interaktion durch den tatsächlichen Zusammenschluss eines oder mehrerer Benutzer mit der Intention, ein bestimmtes Ziel zu erreichen. Eine weit verbreitete Definition für ein Mensch-Maschine-System liefert Geiser [61]:

„Bei einem Mensch-Maschine-System wirkt der Mensch mit einer Maschine mit dem Ziel zusammen, eine selbst gewählte oder vorgegebene Aufgabe zu lösen.“

Der Begriff Maschine bezeichnet dabei allgemein ein technisches System, das nicht näher spezifiziert ist. Bild 2.2 zeigt die Struktur eines Mensch-Maschine-Systems in Form eines Regelkreises (siehe auch [76]). Der Mensch wird im oberen Teil des Bildes durch seine Sensoren zur Informationsaufnahme und seine Motorik zur Informationsausgabe sowie einer auf seinen Zielen, Regeln und Wissen beruhenden informationsverarbeitenden Komponente modelliert. Im unteren Teil des Modells wird die Maschine auf analoge Weise abgebildet: Informationen werden über Bedienelemente aufgenommen und über Anzeigen abgegeben; den

²Mit MMS wird in dieser Arbeit die Mensch-Maschine-Schnittstelle bezeichnet.

Zusammenhang zwischen den jeweiligen Ein- und Ausgaben stellt das Modul *Informationsverarbeitung* her. Sowohl Mensch als auch Maschine können jedoch auch Informationen aus der Umwelt aufnehmen. Ein geschlossener Regelkreis entsteht, wenn sich der Mensch mit seinen sensorischen und aktuatorischen Fähigkeiten an die entsprechenden Schnittstellen der Maschine anknüpft.

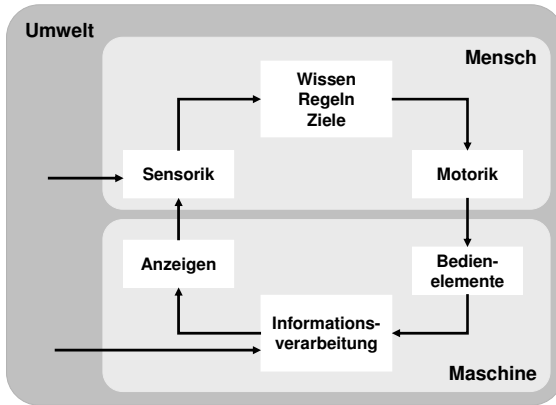


Bild 2.2: Struktur eines Mensch-Maschine-Systems nach [61] und [104].

Mensch-Maschine-Schnittstelle Ein weiterer wichtiger Begriff im Kontext der MMK ist die Bezeichnung Mensch-Maschine-Schnittstelle (MMS), auch Mensch-Maschine-Interface (MMI), Man-Machine-Interface (MMI) oder Human-Machine-Interface (HMI) genannt. Mit diesen Begriffen werden Arbeiten assoziiert, die sich mit der Schnittstelle zwischen Mensch und Maschine beschäftigen. Während die Termini Mensch-Maschine-System und Mensch-Maschine-Interaktion tendenziell eher Fragestellungen zur Konzeption und Struktur des Informationsaustausches in den Vordergrund stellen, behandeln Arbeiten zur Mensch-Maschine-Schnittstelle mehr den technischen Prozess der Kommunikation zwischen Mensch und Maschine. Eine klare Trennung der einzelnen Fachbegriffe und der damit verbundenen Thematiken wird jedoch auch hier in der Fachwelt nicht durchgehalten. Teilweise wird das HMI auf die in Bild 2.2 dargestellte Grenzfläche zwischen Mensch und Maschine

reduziert, in anderen Arbeiten werden hingegen über die verschiedenen Anzeige- und Bedienelemente, mit denen der reine Informationsaustausch stattfindet, „weiche“ Faktoren mit einbezogen [104]. „Weiche“ Faktoren, die u. a. nach Niedermaier [104] einen großen Einfluss auf die Effektivität des HMI haben, sind:

- die Leistungsfähigkeit des Menschen,
- die Motivation des Menschen,
- die Kenntnisse und Fähigkeiten des Menschen,
- die Umwelt (Verkehrssituation),
- die zu bearbeitende Aufgabe,
- die technischen Randbedingungen,
- die Antropometrie und
- der soziale und kulturelle Hintergrund.

Zur systematischen Erfassung dieser technischen und benutzerbezogenen Kriterien werden u. a. von Geiser [61] und Lang [85] entsprechende Modelle eingeführt. Dem Benutzer B werden dabei ein Aufgabenmodell $M(A)$ zur Durchführung der Aufgabe A und ein mentales Modell $M(S)$ zur Vorhersage des Systemverhaltes S zugeordnet. Seitens der Maschine M existieren ein Modell des Benutzers $M(B)$ und der Aufgabe $M(A)$. Dies ist für ein kooperatives System notwendig, um den Benutzer optimal unterstützen zu können. Bild 2.3 zeigt auf abstrakter Ebene das Spektrum der Mensch-Maschine-Schnittstelle mit den zentralen Instanzen Benutzer B, System S und Aufgabe A sowie die wechselseitig zwischen den einzelnen Komponenten existierenden Relationen und die zugeordneten Modelle.

2.2 Technische Kommunikation

Bereits im Jahr 1948 wies Claude E. Shannon [135] auf das grundlegende Problem der technischen Kommunikation hin:

„The fundamental problem of communication is that of reproducing at one point either exactly or approximately a message selected at another point.“

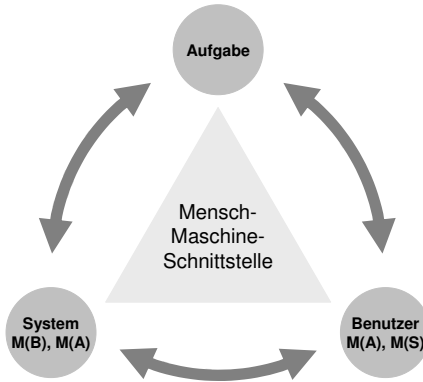


Bild 2.3: Das Spektrum der Mensch-Maschine-Schnittstelle [85].

Aus dieser Überlegung heraus veröffentlichte er in „A mathematical theory of communication“ die in Bild 2.4 gezeigte schematische Darstellung seines Kommunikationssystems.

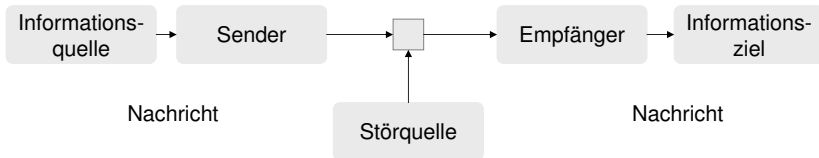


Bild 2.4: Kommunikationsmodell nach Shannon.

Information definiert Shannon [135], von der Nachrichtentechnik geprägt, als ein rein technisches Maß. Zur Untersuchung der Nachrichten auf einem u. U. gestörten Kanal zwischen Sender und Empfänger werden hauptsächlich statistische Methoden eingesetzt. Der Informationsgehalt eines übertragenen Zeichens hängt von der Wahrscheinlichkeit ab, mit der dieses Zeichen auftritt. Voraussetzung für eine korrekte Übertragung ist die Übereinstimmung des Zeichenvorrates von Sender und Empfänger. Auf einer rein technischen Ebene eignet sich das Modell zwar zur vollständigen Beschreibung eines Kommunikationsprozesses, unterliegt aber

einigen wichtigen Beschränkungen. Trogemann [147] beschreibt diese ausführlich. Shannon war es jedoch nie daran gelegen, ein allgemeingültiges Kommunikationsmodell zu schaffen [135]:

„Frequently the messages have meaning; that is they refer to or are correlated according to some system with certain physical or conceptual entities. These aspects of communication are irrelevant to the engineering problem.“

Für eine umfassende Beschreibung der Mensch-Maschine-Kommunikation ist der Informationsbegriff nach Shannon also nicht ausreichend. Daher wird in dieser Arbeit auf einen allgemeineren Informationsbegriff zurückgegriffen [154]:

„Information ist neues Wissen über ein Ereignis, einen Tatbestand oder einen Sachverhalt, Information ist Beseitigung von Ungewissheit.“

Information ist dabei alles, was potentiell zur Verminderung der vorhandenen Ungewissheit beitragen kann. Die Formalisierungen und Abstraktionsniveaus von Information können durchaus unterschiedlich sein. Als Maß für Information wird die durch ihr Hinzukommen beseitigte Ungewissheit definiert.

2.3 Humankommunikation

Paul Grice machte bereits Ende der sechziger Jahre in seinen *William-James-Lectures on Logic and Conversation* (veröffentlicht in [66]) darauf aufmerksam, dass die Machtverhältnisse in der Humankommunikation zwischen Sender und Empfänger anders gelagert sind als in der technischen Kommunikation (siehe auch [57]). Da es keinen für beide Kommunikationspartner verbindlichen Code gibt, ist es dem Empfänger auch nicht möglich, die empfangenen Zeichen des Senders gemäß dem gemeinsamen Code zu interpretieren. Vielmehr muss er ein eigenes Interpretationsmodell entwickeln und die ihm relevant erscheinenden Zeichen nach seinen eigenen Maßstäben deuten. Der Empfänger wird also zusätzlich noch zum Interpretieren, der seinen eigenen, unbewussten Regeln bei der Interpretation der Verhaltensweisen seines Gegenübers folgt [100]. Morris wies schon 1938 in „Foundations of the theory of signs“ [101] darauf hin, dass

die „Definitionshoheit“ für die semantische Beziehung zwischen Zeichen und Bezeichnetem in biologischen Systemen nicht auf Seiten des Senders, sondern auf Seiten des Empfängers liege. Kim et al. [81] stellen dies modellhaft, wie in Bild 2.5 gezeigt, dar. Ist Kommunikation erfolgreich, beinhaltet das Modell zwei informationsverarbeitende Regelkreise: Einerseits tauschen Menschen Informationen mit anderen durch Sprache und begleitendes Verhalten (Gestik, Mimik etc.) aus (Informationsaustausch-Kreis). Andererseits erfassen Menschen die Intention des Partners auf der Basis von Wissen bezüglich des Partners (Intentionsverständnis-Kreis).

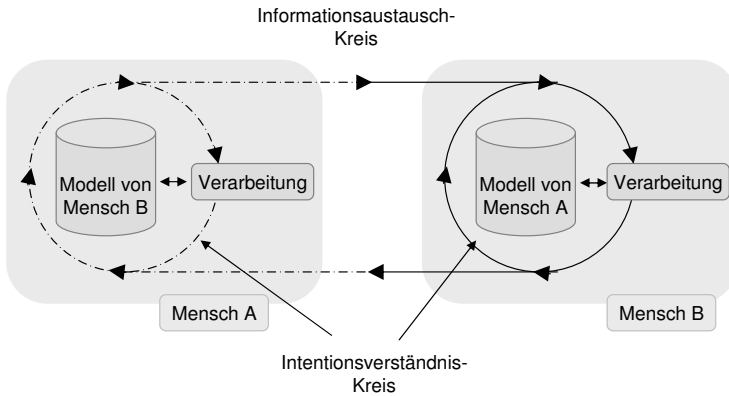


Bild 2.5: Intentionbasiertes Kommunikationsmodell nach Kim [81].

Frey [56, 58] zeigt des Weiteren, dass es sich bei menschlicher Kommunikation um einen inferentiellen Prozess handelt. Der Sender kann also darauf vertrauen, dass der Empfänger nicht nur die syntaktischen Defizite repariert, sondern darüber hinaus die empfangene Information um die zum Verständnis notwendigen Zusatzinformationen ergänzt, die nicht mitgeliefert wurden, da sie vom Sender als beim Empfänger bereits vorhanden vorausgesetzt wurden [100]. Nach Grice [66] orientiert sich der Empfänger dabei an vier Prinzipien: *quantity* (die Äußerungen des Senders enthalten weder mehr noch weniger Information als erforderlich), *quality* (sie sind weder unwahr noch unbewiesen), *relation* (sie sind nicht irrelevant für das betrachtete Thema) und *manner* (sie sind nicht in einer unhöflichen oder mehrdeutigen Weise vorgetragen). Damit können

nach Möller [100] weitere in der technischen Kommunikation unbekanntere Phänomene erklärt werden: das „kreative Verständnis“, das „kreative Missverständnis“ und die „Informationsabwehr“. Demnach steht es dem menschlichen Informationsempfänger frei [55]

„... die gehörten Äußerungen nach Belieben mit seinem Hintergrundwissen und Erfahrungsschatz in Beziehung zu bringen. Ein mögliches Ergebnis kann darin bestehen, dass die erhaltene Information Inferenzprozesse in Gang setzt, die dazu führen, dass der Hörer zu einem tieferen Verständnis der besprochenen Angelegenheit gelangt als der Sprecher.“

Umgekehrt könne sich der Rezipient allerdings auch mit seinen Schlussfolgerungen selber verwirren und die Sache werde ihm [im Sinne eines kreativen Missverständnisses] unklarer als sie zuvor war. Darüber hinaus stehe allerdings im Sinne der Informationsabwehr dem Empfänger die Option offen, die Information des Senders völlig abzuweisen, entweder weil er den Aussagen misstraut oder die Information als irrelevant erachtet [55].

Geht man aber davon aus, dass beide Kommunikationspartner an einer gelungenen Kommunikation interessiert sind, lassen sich daraus nach Trogemann [147] weiter reichende Konsequenzen ableiten: Wenn es für den Sender bedeutsam sei, welche Wirkung er beim Empfänger hinterlasse, werde er bemüht sein, Signale zu senden, die zu positiven Einschätzungen auf der Seite des Empfängers führen. Das Kommunikationsverhalten des menschlichen Senders steht damit unter dem heimlichen Diktat des Empfängers [147]. Vom Sender erfordert das die Fähigkeit der Antizipation von Interpretationsprozessen. Vom Empfänger hingegen verlangt es die Fähigkeit der Generierung von Feedback-Signalen während des Empfangs zur Steuerung des Senders. Beide Kommunikationspartner agierten nach Trogemann [147] zwar alternierend in der Sprecher- und der Hörerrolle, seien aber immer simultan Sender und Empfänger.

Ein weiterer wichtiger Aspekt in der Humankommunikation ist die – bei der interpersonellen Eindrucksbildung – teilweise größere Bedeutung des Gestus einer Person als die der sprachlichen Äußerungen. Der Einfluss dieser sogenannten nonverbalen Komponenten (auch: implizite Kommunikationsanteile) – Gestik, Mimik, Körperhaltung sowie die nicht-sprachlichen Aspekte des Sprechverhaltens (Klang, Rhythmus, Lautstärke etc.) – auf den Verlauf und das Ergebnis kommunikativer Bemühun-

gen ist bis heute nicht klar erforscht [100]. Man geht davon aus, dass für den größten Teil dieser Aspekte kein sozial autorisiertes Referenzsystem vorhanden ist. Damit gibt es auch keinen für Sender und Empfänger übereinstimmenden Code. Die Sinnstiftung in der nonverbalen Kommunikation erfolgt daher durch eine unwillkürlich vorgenommene, selektive und eigenwillige Interpretation des Empfängers [100].

2.4 Ein- und Ausgabekanäle

Der Mensch kann über verschiedene Kanäle Informationen abgeben und auch aufnehmen. So ist es ihm subjektiv einfach möglich, effektiv zu kommunizieren. Informationen werden dabei oft auf parallelen Kanälen übertragen. Die bereits angesprochenen Informationen, die zur interpersonellen Eindrucksbildung in Betracht gezogenen werden, werden oft parallel zum auditiven Kanal übertragen und empfangen. Dem Menschen stehen für den Empfang von Signalen, korreliert zu den sechs Sinnen (Sehsinn, Hörsinn, Geruchssinn, Geschmackssinn, Gleichgewichtssinn und Tastsinn/Haut) sechs Wahrnehmungskanäle zur Verfügung: der visuelle, der auditive, der olfaktorische, der gustatorische, der vestibuläre und der taktile Kanal. Allgemein werden diese Fähigkeiten als sensorische Fähigkeiten bezeichnet. Für die von Trogemann [147] angesprochenen Signale zur Beeinflussung eines Gesprächs ist das Spektrum etwas geringer: für eine intentionsbasierte Humankommunikation stehen nur der auditive, der visuelle und der taktile Kanal zur Verfügung [6].

Heutige technische Systeme sind sowohl bei der Aufnahme von Informationen als auch bei der Abgabe stärker eingeschränkt. Nach Geiser [61] können jedoch bei der Informationsabgabe auf Maschinenseite analog zum Menschen verschiedene Anzeigearten identifiziert werden, die sich mit den motorischen Fähigkeiten des Menschen decken: auf dem visuellen Kanal durch optische Anzeigen (Displays o. ä.), auf dem auditiven durch akustische (z. B. Sprachausgaben oder Warntöne) und auf dem taktilen Kanal durch haptische Anzeigen. Implizite Anteile der Kommunikation werden in Mensch-Maschine-Systemen jedoch meist nicht weiter betrachtet.

2.5 Multimodale Mensch-Maschine-Systeme

Multimodale Systeme nutzen die verschiedenen Kanäle des Menschen zur Abgabe und Aufnahme von Information. Der Begriff selbst leitet sich aus dem Lateinischen ab: „*multi*“ (lat.: viel, mehrere) und „*modus*“ (lat.: Art und Weise), d. h. es werden verschiedene Möglichkeiten zur Ein- und Ausgabe im HMI zur Verfügung gestellt. Multimodale Bedienung ist jedoch mehr als die parallele Verfügbarkeit verschiedener Eingabelemente. Vielmehr müssen die einzelnen Modalitäten schlüssig in ein Gesamtkonzept integriert sein, um so den Nutzer optimal bei der Interaktion zu unterstützen. Bei der Gestaltung multimodaler Schnittstellen sind daher nicht nur technische Aspekte, sondern auch ergonomische, ästhetische und psychologische Aspekte zu berücksichtigen. Besonders in der Domäne Fahrzeug spielen darüber hinaus auch ökonomische Aspekte eine große Rolle.

2.5.1 Potenzial multimodaler Schnittstellen

Multimodale Schnittstellen bieten ein großes Potenzial auf dem Weg zu einer intuitiven, effizienten, robusten und natürlichen Bedienung [112]. Dies wurde in vielen Studien, u. a. von [8, 40, 42, 68, 151] nachgewiesen: Die Geschwindigkeit der Aufgabenbearbeitung steigt, Benutzer machen weniger Fehler und tun in der jeweiligen Dialogsituation das Richtige. Nicht zuletzt werden natürliche Kommunikationsformen für eine intuitivere Bedienung und einen natürlicheren Dialog möglich. Bild 2.6 zeigt nach Niedermaier [104] das Potenzial von Multimodalität im Überblick.

2.5.2 Limitierungen multimodaler Schnittstellen

Die Gestaltung multimodaler Systeme geht jedoch auch immer mit erhöhtem technischen Aufwand einher. Die drei maßgeblichsten Punkte sind (siehe auch [104]):

- zusätzliche Hardwarekomponenten:
Sensoren zur Erfassung von Gestik, Mimik, Sprache etc.;
- erhöhte Anforderungen an die Infrastruktur:
CPU-Leistung, Arbeitsspeichergröße etc.;

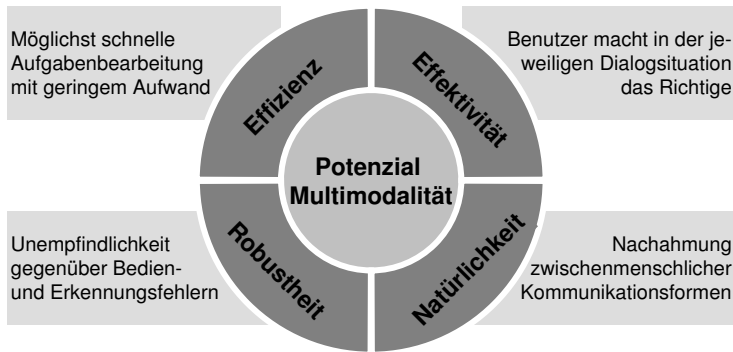


Bild 2.6: Potenzial Multimodalität [104].

- erhöhter Entwicklungsaufwand durch komplexere Software: Spezifikation umfangreicher Dialogmöglichkeiten, Verifikation der Softwarekomponenten.

Speziell in der Domäne Fahrzeug mit hohem Kostendruck sind dies stark limitierende Faktoren.

2.5.3 Klassifizierung multimodaler Systeme

Multimodale Systeme weisen eine große Anzahl an unterschiedlichen Realisierungen für das Zusammenspiel der Modalitäten bei der Bedienung auf. Um die Systeme klassifizieren und systematisieren zu können, wurde eine Reihe von Begriffen eingeführt, die teilweise deckungsgleich sind, sich teilweise aber auch in ihrer Definition unterscheiden. Im Folgenden sollen drei in der Literatur anerkannte Klassifikationsschemata vorgestellt werden.

Martin [93] definiert Modalität als den Prozess des Empfangens und Abgebens von Informationsbruchstücken. Darauf aufbauend definiert er sechs grundlegende „*types of cooperation*“ zwischen den Modalitäten: Transfer, Spezialisierung, Äquivalenz, Redundanz, Komplementarität und Simultanität.

- **Transfer** (*transfer*): Informationen aus einer Modalität werden

von einer anderen Modalität benutzt oder Informationen einer Modalität werden mit Informationen einer anderen fusioniert, um so die Erkennungsrate zu erhöhen.

- **Spezialisierung** (*specialization*): Eine bestimmte Art der Information wird immer mit Hilfe einer bestimmten Modalität übermittelt. Wenn eine Eins-zu-eins-Beziehung zwischen bestimmten Informationen und Modalitäten besteht, wird auch von absoluter Spezialisierung gesprochen. Dies hat den Vorteil, dass der Modalität semantische Information hinzugefügt wird und so der Interpretationsprozess erleichtert wird.
- **Äquivalenz** (*equivalence*): Informationen sind gleichwertig, d. h. sie können mit der einen oder der anderen Modalität übermittelt werden.
- **Redundanz** (*redundancy*): Dieselbe Information wird über mehrere Modalitäten übermittelt. Dies steigert die Erkennungsrate.
- **Komplementarität** (*complementarity*): Informationen werden mit verschiedenen Modalitäten übermittelt, müssen aber zunächst fusioniert werden, um interpretiert werden zu können. Wichtig ist hierbei das Fusionskriterium. Bei sequentieller Informationsübermittlung werden die einzelnen Informationsbruchstücke nacheinander übertragen, bei synergistischer Informationsübermittlung gleichzeitig. Auch hier können bestimmte Modalitäten auf bestimmte Ein- oder Ausgaben spezialisiert sein.
- **Simultanität** (*concurrency*): Verschiedene Informationen werden simultan übermittelt und verarbeitet, müssen aber nicht fusioniert werden. Vorteil ist eine schnellere Interaktion.

Neuss [103] unterteilt im Gegensatz zu Martin [93] Multimodalität in nur vier verschiedene Ausprägungen, die er hinsichtlich der Eingabe differenziert: seriell-redundant, seriell-exklusiv, parallel-ergänzend und parallel-verifizierend.

- **Seriell-redundant**: Systeme, die bezüglich ihrer Funktionalität die volle Redundanz der Eingabemöglichkeiten aufweisen und auf der Auswertungsseite voll verknüpft sind. Der Nutzer kann so während der Eingabe jederzeit die Modalität wechseln.

- **Seriell-exklusiv:** Systeme, bei denen die Eingabemodalitäten unterschiedlich mächtig oder die Modalitäten nicht verknüpft sind. Für die Umsetzung eines Plans muss sich der Nutzer unter Umständen vorab auf eine Modalität festlegen.
- **Parallel-ergänzend:** Systeme, die gleichzeitig getätigte Nutzereingaben mit verschiedenen Modalitäten und unterschiedlichen Bedeutungen korrelieren („*Put-That-There*“-Metapher [24]).
- **Parallel-verifizierend:** Verschiedene Modalitäten werden zur Disambiguierung oder Verifikation benutzt.

Einen weiteren Ansatz zur Klassifikation multimodaler Systeme liefern Nigay und Coutaz [106]. Die vorliegende Arbeit verwendet diese Definition, wie sie auch in abgewandelter Form in der Arbeit von Niedermaier [104] verwendet wird, die ebenfalls einen automobiltechnischen Fokus hat.

Zunächst werden bei Nigay und Coutaz [106] drei Aspekte unterschieden: der Gebrauch der Modalitäten (*use of modalities*), die Art der Informationsfusion (*fusion*) und das Abstraktionsniveau (*levels of abstraction*). Da die Fusion in Infotainmentsystemen aber meist auf der semantischen Ebene erfolgt, wird das Modell im Folgenden und in Anlehnung an Niedermaier [104] auf zwei Dimensionen reduziert. Die verbleibenden Aspekte „Gebrauch der Modalitäten“ und „Grad der Informationsfusion“ spannen eine zweidimensionale Matrix auf, in die multimodale Systeme eingeordnet werden können (siehe Bild 2.7).

- **Synergistisch** (*synergistic*): Teilinformationen werden parallel mit verschiedenen Modalitäten eingegeben und semantisch zu einer Gesamtinterpretation miteinander verbunden.
- **Simultan** (*concurrent*): Informationen können parallel eingegeben werden, werden aber unabhängig voneinander interpretiert. Eine parallele Bedienung verschiedener Funktionen ist dadurch möglich.
- **Exklusiv** (*exclusive*): Zu einem Zeitpunkt kann immer nur mit einer Modalität bedient werden, da Informationen sequentiell verarbeitet werden. Wechselt der Benutzer die Modalität, werden die Informationen unabhängig voneinander verarbeitet.

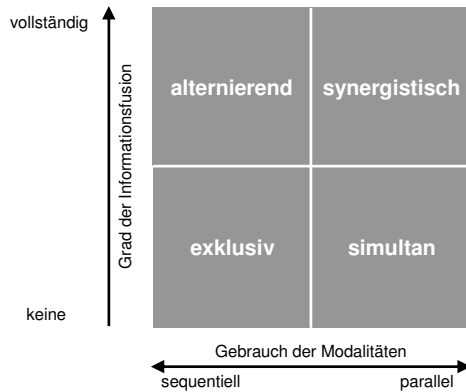


Bild 2.7: Klassifikationsschema für multimodale Systeme nach [104] und [106].

- **Alternierend** (*alternate*): Die verschiedenen Modalitäten können nur sequentiell verwendet werden, die Eingaben werden aber verbunden und als Gesamtheit interpretiert.

Niedermaier [104] analysiert eine größere Anzahl an multimodalen Mensch-Maschine-Systemen und kommt zum Ergebnis, dass die nicht für die Domäne Fahrzeug entwickelten Systeme meist die parallele Eingabe von Informationen über mehrere Modalitäten erlauben und diese dann im Verbund interpretieren. Es handelt sich also um synergistisch-multimodale Systeme. Hier sei u. a. auf das Projekt „SmartKom“ [152] verwiesen, in dem die gleichzeitige und komfortable Benutzung von Sprach-, Bild- und Datendiensten erprobt werden sollte. Oviatt et al. [113] weisen darauf hin, dass Befehle unter Einbeziehung räumlich-geometrischer Angaben, wie sie z. B. bei der Interaktion mit Karten vorkommen, der häufigste Aufgabentyp für den Einsatz synergistisch- oder alternierend-multimodaler Systeme sind.

Bei Systemen, wie sie derzeit im Fahrzeug erhältlich sind, werden Informationen unabhängig voneinander interpretiert, d. h. es findet keine Informationsfusion statt; der Nutzer entscheidet sich für eine Spracheingabe oder eine haptische Eingabe. Ein Wechsel z. B. während der Navigationszieleingabe ist nicht möglich. Die Systeme sind also dem Bereich der exklusiv-multimodalen Systeme zuzuordnen. Automobiltechni-

sche Systeme aus dem Forschungsbereich hingegen sind oft dem Bereich alternierend-multimodal zuzuordnen, d. h. hier können die verschiedenen Modalitäten im Verbund verwendet werden, wobei jedoch keine parallele Nutzung vorgesehen ist. So kann z. B. der erste Eingabeversuch mittels Sprache erfolgen, die Disambiguierung bei mehreren erkannten Begriffen durch haptische Listenauswahl. Die einzelnen Modalitäten fügen sich dann zu einem Gesamtkonzept zusammen und tragen dazu bei, dass eine Funktion effizient ausgeführt werden kann.

2.5.4 Informationsfusion in multimodalen Systemen

Die Zusammenführung (Fusion) verschiedener Informationen aus den jeweiligen Modalitäten auf ein gemeinsames semantisches Niveau sowie eine gesamtheitliche Interpretation des Fusionsergebnisses ist bei multimodalen Systemen von zentraler Bedeutung. Ein in der Praxis häufig eingesetztes Fusionsmodell ist das Omnibusmodell (siehe Bild 2.8), das die Mess- und Fusionsaufgabe als Regelkreis darstellt. Die vier Aufgaben

- Messen (Wahrnehmung, Signalverarbeitung),
- Ausloten (Merkmalsextraktion, Mustererkennung),
- Entscheiden (Kontextverarbeitung, Entscheidung) und
- Handeln (Steuerung, Ressourceneinsatz)

folgen aufeinander. Die Übergänge zwischen den Funktionen entsprechen den verschiedenen Fusionsstufen (Sensordatenfusion, *Soft-Decision*-Fusion, *Hard-Decision*-Fusion und Sensor-Management).

Für die eigentliche Sensorintegration und Datenfusion existieren verschiedene Konzepte (siehe auch [127]):

- **Konkurrierende Integration:** Zur Reduktion von Unsicherheit werden gleichartige Sensordaten mit gleicher Nutzinformation z. B. durch Mittelung fusioniert.
- **Komplementäre Integration:** Gleichartige Sensordaten mit unterschiedlicher Nutzinformation werden mit dem Ziel, Informationslücken zu schließen, fusioniert.
- **Kooperative Integration:** Aufgrund der verteilt vorliegenden Nutzinformation werden alle Messdaten zum Schließen auf die Messgröße ausgewertet.

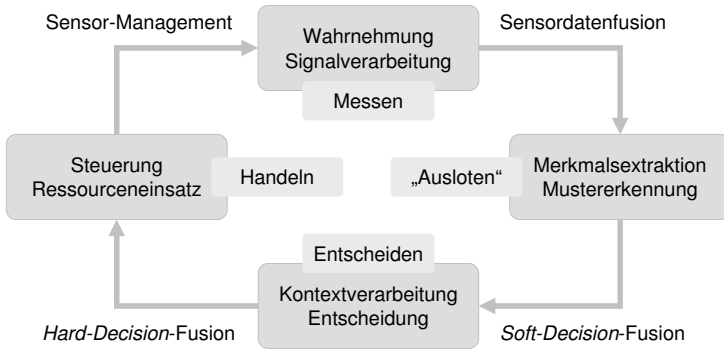


Bild 2.8: Omnibusmodell: Messen als Regelkreis.

Die eigentliche Ausführung der Informationsfusion kann – abhängig vom Abstraktionsniveau – auf drei Ebenen erfolgen (siehe auch Bild 2.9). Wird die Fusion auf höheren Ebenen durchgeführt, können Ergebnisse aus darunterliegenden Ebenen mit berücksichtigt werden. Nach [127] ist die Fusion auf einer höheren Abstraktionsebene meist effizienter, aber nicht unbedingt effektiver, da die Datenreduktion zu einem Informationsverlust führt.

- **Fusion auf der Signalebene:** Die Signale von Einzelsensoren werden unter der Voraussetzung von Vergleichbarkeit, Registrierung und Synchronisation, direkt kombiniert.
- **Fusion auf der Merkmalsebene:** Kann auf der Signalebene keine zeitliche oder räumliche Koheränz der Einzelsensordaten hergestellt werden, können auf der Merkmalsebene Signaldeskriptoren fusioniert werden, um verbesserte Schätzwerte bestimmter Signaleigenschaften zu erhalten.
- **Fusion auf der Symbolebene:** Zum Treffen von Entscheidungen werden auf dieser Ebene symbolische Signaldeskriptoren (Detektions- oder Klassifikationsergebnisse), denen z.B. Wahrscheinlichkeiten zugeordnet sind, kombiniert.

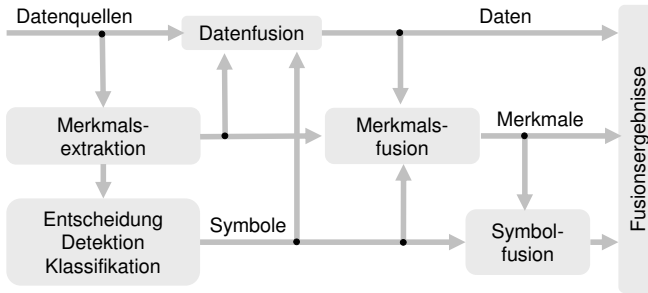


Bild 2.9: Abstraktionsebenen bei der Informationsfusion.

Bei multimodalen Systemen sind sowohl die Sensorsignale als auch die Information aus den verschiedenen Modalitäten fast immer sehr unterschiedlich, sodass unter dem Begriff *Informationsfusion* hier meist eine kooperative Fusion zu verstehen ist. Eine Ausnahme bilden z. B. audiovisuelle Erkennungssysteme, die Lippenbewegungen und Sprachsignale gemeinsam zur Erkennung einer Äußerung verwenden. Solche Systeme spielen jedoch im Kontext Fahrzeug eine untergeordnete Rolle und werden daher in dieser Arbeit nicht näher berücksichtigt.

Die Fusion der Daten in multimodalen Systemen erfolgt meist auf Symbolebene, da hier beliebige Modalitäten miteinander fusioniert werden können. Dazu wird zunächst für jedes Signal getrennt eine semantische Interpretation der jeweiligen Eingabe erzeugt. Diese Teilinformationen werden anschließend fusioniert, sodass eine gesamtheitliche Interpretation der Benutzerintention vorliegt.

2.6 Adaptive und adaptierbare Systeme

Adaptive Systeme sind Systeme, die sich an die spezifischen Bedürfnisse des Benutzers und dessen Eingaben individuell und automatisch anpassen und die Informationsabgabe adäquat und an den jeweiligen Benutzer angepasst gestalten. Da sich die Ziele und Eigenschaften eines Nutzers verändern, muss ein adaptives System das Modell seines Benutzers ständig aktualisieren. Dabei verfolgt es die Benutzerinteraktionen und trifft Annahmen über den Nutzer [21].

Ziel adaptiver Systeme ist es, den Benutzer kognitiv zu entlasten und die Interaktion zu vereinfachen. Davon unterschieden werden adaptierbare Systeme, die dem Benutzer die Möglichkeit bieten, bestimmte Aspekte, z. B. des Interfaces, nach seinen spezifischen Bedürfnissen selbst zu gestalten. Adaptive und adaptierbare Funktionen können in Systemen parallel implementiert werden. Tabelle 2.1 zeigt die Unterschiede zwischen adaptierbaren und adaptiven Systemen.

Adaptive Systeme im Kraftfahrzeug versuchen den Zustand des Benutzers zu erkennen und bei der Interaktion zu berücksichtigen. Zum Zustand des Benutzers gehören sowohl kurzfristige als auch langfristige Merkmale. Kurzfristige Merkmale sind z. B. die momentane Arbeitslast, das aktuelle Ziel des Benutzers und die Beanspruchung. Zu den langfristigen Merkmalen zählen z. B. persönliche Eigenschaften, Wissen und Interessen. Während die Belastung des Nutzers aus physiologischen Größen ermittelt werden kann, sind für die Einschätzung von individuellen Merkmalen im Kfz eher Beobachtungen der Interaktion des Benutzers mit dem System (Tastendrucke etc.) vorzuziehen. Individuelle Merkmale können Bedienfehler, Fehlerhäufigkeit, Interaktionsfrequenz, Fahrstil oder das aktuell verfolgte Ziel sein.

2.6.1 Benutzermodelle

Kooperative Systeme, die dem Benutzer die Erledigung seiner Aufgaben erleichtern, müssen die Ziele, Pläne, das Wissen sowie die Überzeugungen und gegebenenfalls die Einstellungen und Neigungen des Partners berücksichtigen. Daher benötigen sie ein Benutzermodell, das als Wissensbasis explizite Annahmen über alle Aspekte des Benutzers enthält, die für den Dialog zwischen Mensch und Maschine relevant sein können [61, 153].

Ein Ziel beschreibt dabei nach Blank [21]

„...einen Zustand, der unmittelbar nach der Ausführung eines Planes gültig ist. Ein Ziel kann auf verschiedenen Wegen erreicht werden, wobei jeder Weg einem Plan entspricht.“

Zur Charakterisierung von Benutzermodellen schlagen Kass und Finin [79] fünf Dimensionen vor (siehe Bild 2.10): den Grad der Spezialisierung, die Modifizierbarkeit, den Zeithorizont, die Anwendungsart und die Akquisitionsart. Bei den meisten bekannten Modellen liegen nicht die

Tabelle 2.1: Vergleich von adaptiven und adaptierbaren Systemen nach [62].

Aspekt	Adaptierbare Systeme	Adaptive Systeme
Definition	Benutzer passt die Funktionalität an (Unterstützung durch das System)	Dynamische Adaption des Systems auf die aktuelle Handlung und den Benutzer
Wissen	Wissen des Systems wird erweitert	Wissen ist im System enthalten
Stärken	Benutzer kontrolliert; Benutzer kennt Tätigkeit am besten; Systemwissen passt besser; erfolgreiche Modelle vorhanden	Wenig (oder kein) Aufwand für den Benutzer; Benutzer benötigt kein spezielles Wissen
Schwächen	Systeme werden inkompatibel; Benutzer müssen Anpassung durchführen; Komplexität wächst an, da der Benutzer die Adaptierbarkeitskomponente verstehen muss	Benutzer hat Probleme, ein konsistentes Modell der Anwendung zu entwickeln; Kontrollverlust
Erforderliche Mechanismen	Ebenenarchitektur; Anwendungsmodell und -orientierung; Feedback vom System; sinnvolle Gestaltung	Modelle des Benutzers, der Tätigkeit und der Dialoge; Wissensbasis der Ziele und Pläne; leistungsfähige Bewertungsfähigkeiten; dynamische Modellaktualisierung
Anwendungsgebiete	Information Retrieval; Modifizierung durch Endnutzer; Anpassbarkeit; Filtermöglichkeiten; gebrauchstaugliche Gestaltung	Aktives Hilfesystem; Kriteriensysteme; Benutzungsoberfläche anpassbar; Information Retrieval

extremen Ausprägungen vor, sondern es wird eine vorteilhafte Kombination angestrebt.



Bild 2.10: Dimensionen eines Benutzermodells [79].

Benutzermodelle stellen also den Menschen mit seinen individuellen Fähigkeiten ins Zentrum der Betrachtungen. Sie betrachten Benutzer und Aufgabe in einem einheitlichen Kontext und setzen daher im Gegensatz zu den Aufgabenmodellen an zwei Punkten des Dreiecks aus Bild 2.3 an.

Blank [21] unterscheidet vier Methoden der Benutzermodellierung: Meta-Dialog (der Benutzer beschreibt und klassifiziert sich selbst), Beobachtung des Benutzers, Annahmen über Benutzerklassen und Annahmen über alle Benutzer.

2.6.2 Mentale Modelle

Wahrnehmung bezeichnet im Allgemeinen den Vorgang der bewussten Aufnahme von Informationen eines Lebewesens über seine Sinne. Bei lernfähigen Lebewesen wird ein Teil dieser Wahrnehmungen im Gedächtnis gespeichert. In diesen Erfahrungen können Muster erkannt werden, aus denen wiederum Regeln abgeleitet werden. Im Laufe der Zeit entstehen so mentale Repräsentationen der individuell relevanten Ausschnitte der Welt, also gewissermaßen reduzierte und komprimierte Abbilder der Realität im Gedächtnis. Die abgebildeten Merkmale werden dabei stark von bereits vorhandenem Vorwissen und der evtl. vorhandenen Handlungsabsicht des Menschen beeinflusst. Nach Dutke [50] ist ein mentales Modell zum Verstehen von unbekanntem Sachinhalten dann besonders effektiv, wenn es neben dem Wissen auch die Problemstellung der spezi-

fischen Situation schematisch abbildet. Für die Steuerung von Routinehandlungen ist ein mentales Modell dagegen effektiver, wenn es Einzelergebnisse abstrahiert und Kontext, Handlungen und Handlungsfolgen schematisch abbildet.

Zu beachten ist jedoch, dass nicht nur die Komponenten des mentalen Modells bei jedem Menschen unterschiedlich sind, sondern auch ihre Gewichtung. Während einige Menschen eher bildhaft denken, orientieren sich andere eher an anderen Sinneseindrücken und Erfahrungen. Trotzdem lassen sich mit der Theorie der mentalen Modelle komplexe Prozesse der Informationsverarbeitung veranschaulichen und sich aus ihren Annahmen Rückschlüsse auf die Gestaltung von Benutzerschnittstellen ziehen [82, 137]. Norman [108] weist darauf hin, dass es bei der Gestaltung von Mensch-Maschine-Schnittstellen darauf ankommt, ein konsistentes, kohärentes Modell zu entwickeln, das es dem Nutzer ermöglicht, ein – mit dem der Systemschnittstelle zugrunde liegenden Modell – konsistentes mentales Modell zu entwickeln. Bei der Entwicklung von MMIs ist also sicherzustellen, dass die Differenz zwischen den mentalen Modellen der Systementwickler und denen der Endnutzer möglichst gering ist (siehe auch [110]), um eine für den Endnutzer einfach und intuitiv zu bedienende Schnittstelle zu erreichen. Nur so kann sich der Mensch ein mentales Bild komplexer Strukturen schaffen, um nachfolgende Bedienschritte bereits im Voraus planen zu können [69].

Man spricht auch von „innerer Konsistenz“, wenn das Dialogverhalten des Systems den Erwartungen entspricht [21], die der Nutzer an die Benutzung der Anwendung stellt und die er aus anderen Arbeitsabläufen mitbringt (entsprechend DIN EN ISO 9241 [48]). Mentale Modelle entziehen sich jedoch der unmittelbaren Beobachtung, sodass sie durch beobachtbares Verhalten validiert werden müssen.

2.7 Gestaltungsaspekte von Mensch-Maschine-Systemen

Neben technischen Aspekten spielen v. a. inhaltliche und gestalterische Aspekte bei der Realisierung multimodaler HMIs eine große Rolle. Dazu existieren für die verschiedenen Aufgaben und Domänen so genannte *Styleguides* (Gestaltungshinweise). Ein allgemein akzeptiertes und viel verwendetes Kriterium bei der Gestaltung von HMIs ist die Norm DIN

EN ISO 9241 [48]. Bild 2.11 stellt die sieben Aspekte für eine gebrauchstaugliche Gestaltung technischer Komponenten dar. Ergänzend dazu können die „Acht Goldenen Regeln“ des Schnittstellendesigns von Shneiderman sowie die zehn Usability-Heuristiken von Nielsen herangezogen werden [44, 105, 131].



Bild 2.11: Ergonomie der Mensch-System-Interaktion: Grundsätze der Dialoggestaltung nach ISO 9241 [48].

Nach Dahm [44] ergeben sich aufbauend auf die sieben Handlungsschritte von Norman [109] folgende Folgerungen und Richtlinien bei der Gestaltung von Mensch-Maschine-Schnittstellen:

- Es muss in jeder Situation erkennbar sein, welche Handlungen/Aktionen möglich sind.
- Der Zustand des Systems muss erkennbar sein.
- Die Transformation zwischen den Zielen des Benutzers in Bedienungshandlungen muss möglichst leicht erfolgen können.
- Das zugrunde liegende konzeptionelle Modell muss sich schnell erschließen lassen.
- Die Reaktion und der sich daraus ergebende Zustand müssen leicht interpretierbar sein.

- Der Benutzer muss den angezeigten neuen Systemzustand möglichst leicht mit seinem Ziel vergleichen können.
- Die Entwickler müssen sich über die Unterschiede in den eigentlichen Zielen der Benutzer und den von ihnen angebotenen Aktionen und Funktionen im Klaren sein. Funktionen sind immer nur Mittel zum Zweck, ein Nutzen für den Anwender ergibt sich nur aus seinen Zielen.

Ist der Nutzer jedoch nicht in der Lage, aus dem angestrebten Ziel eine Bedienung abzuleiten oder auszuführen, spricht Norman [111] vom „*Gulf of Execution*“. Ergeben sich hingegen bei der Interpretation der von der Maschine abgegebenen Informationen Schwierigkeiten oder Unstimmigkeiten, führt Dahm [44] den Begriff des „*Gulf of Evaluation*“ ein.

Für die Gestaltung im Fahrzeug sind v. a. eine effiziente Ablesbarkeit der dargestellten Informationen sowie eine schnelle und fehlerfreie Bedienung des Infotainmentsystems wichtig, da die Interaktion nur als Nebenaufgabe neben der Primäraufgabe ausgeführt werden kann. Die DIN EN ISO 15008 [47] beschäftigt sich daher mit den spezifischen Anforderungen an die Displaygestaltung, wie z. B. der Schriftart, dem Schriftschnitt oder der Farbwahl.

2.8 Mensch-Maschine-Kommunikation im Kraftfahrzeug

2.8.1 Bediensituation und Aufgabenklassifizierung im Fahrzeug

Die Anforderungen an den Fahrer im Fahrzeug lassen sich grob in zwei Bereiche einteilen. Die Fahrzeugführung stellt an den Fahrer sowohl kontinuierliche als auch diskontinuierliche Aufgaben und zieht den größten Teil der Aufmerksamkeit auf sich [64]. Die Bedienung von Infotainmentsystemen, die nur einen indirekten Bezug zur Fahraufgabe aufweist, hat dagegen lediglich intermittierenden Charakter.

Entsprechend ihrer Wichtigkeit erfolgt nach Rassl [121] und Bubb [32] eine hierarchische Aufteilung in primäre, sekundäre und tertiäre Aufgaben. Dabei beschreibt die primäre Fahraufgabe alle Tätigkeiten, die das sichere Bewegen des Fahrzeugs im Verkehrssystem betreffen und kann nach Bubb [31] darauf reduziert werden, dass jede Berührung mit ste-

henden oder sich bewegenden Objekten im Verkehrsraum zu vermeiden ist. Die Fahrzeugführung kann mit Hilfe eines hierarchisch verschachtelten Regelkreises beschrieben werden, dessen zentrale Elemente die Navigationsaufgabe, die Führungsaufgabe und die Stabilisierungsaufgabe darstellen [18, 33].

Sekundäre Aufgaben dienen nicht der direkten Führung des Fahrzeugs, fallen aber im Rahmen der Fahraufgabe an und können als Reaktion auf Verkehrssituationen, Umweltbedingungen oder Kommunikationsmaßnahmen auftreten. Dabei werden reaktiv bedingte Aufgaben, die aufgrund der Verkehrssituation auftreten und aktiv sekundäre Aufgaben, die der Kommunikation mit anderen Verkehrsteilnehmern dienen, unterschieden.

Tertiäre Aufgaben stehen nicht direkt mit der Primär- oder Sekundäraufgabe in Verbindung, sondern sind Aufgaben, die das Informations-, Sozial- und Komfortbedürfnis des Fahrers betreffen und dienen dem Zufriedenstellen von Komfort-, Unterhaltungs- und Informationsbedürfnissen [121]. Sie dürfen die primäre und sekundäre Aufgabenerfüllung nicht über ein bestimmtes Maß hinaus stören [32].

2.8.2 Spezifische Anforderungen an Kfz-Systeme

Mensch-Maschine-Interfaces, die im automobilen Kontext eingesetzt werden, müssen einer Reihe von kfz-spezifischen Anforderungen genügen. Entscheidend dabei ist, wie für alle Komponenten eines Kraftfahrzeugs, der Wirtschaftlichkeitsaspekt. Daher ist bei der Systementwicklung auf kostengünstige Hardware und einen geringen Ressourcenbedarf an Rechenleistung und Speicherkapazität zu achten.

Des Weiteren sind sog. ‚*Package*‘-Aspekte zu berücksichtigen. Für HMI-Komponenten ist bei der Fahrzeugentwicklung meist ein bestimmter Raum (z. B. ein DIN-Schacht) vorgesehen. Funktionen oder Technologien, die das heute bereits bestehende HMI erweitern, sollten sich in das heutige *Package* einfügen und möglichst wenig Platz in Anspruch nehmen.

Die Fahrzeugumgebung ist im Vergleich zu Desktop-Anwendungen aus technischer Sicht eine anspruchsvolle Umgebung. Wechselnde Lichtverhältnisse erschweren den Einsatz von bildgebenden und bildverarbeitenden Systemen, extreme Temperaturschwankungen (-40 °C bis $+80\text{ °C}$) stellen große Herausforderungen an Displays und andere Hardware-

Elemente.

Zuletzt sei die Systemreaktionszeit erwähnt. In nicht-automobilen Kontexten wirkt sich diese meist nur auf die Benutzerakzeptanz aus, während die Systemreaktionszeit bei Kfz-Systemen Einfluss auf die Verkehrssicherheit des Fahrzeugs haben kann. Geiger [60] zeigte in Untersuchungen, dass das Ausbleiben einer Systemreaktion bereits nach wenigen 100 ms zu Kontrollblicken des Fahrers und somit zur Ablenkung vom Verkehrsgeschehen führt. Die Verarbeitung jeglicher Eingaben im HMI muss somit in weniger als 100 ms erfolgen.

2.8.3 Serienstand

In heutigen Serienfahrzeugen existiert eine große Anzahl an unterschiedlichen Anzeige- und Bediensystemen. Um zwei grundsätzliche Tendenzen in den Konzepten aufzuzeigen, werden hier beispielhaft das *Multi Media Interface* von Audi und das *RNS 510* von VW gegenübergestellt (siehe Bild 2.12). Zur visuellen Informationsausgabe dient in beiden Fällen ein



(a) *Multi Media Interface* von Audi (Quelle: Volkswagen).



(b) *RNS 510* von VW (Quelle: Volkswagen).

Bild 2.12: Gegenüberstellung von zwei marktüblichen *High-End*-Anzeige- und Bediensystemen in der Mittelkonsole.

hochauflösendes Farbdisplay in der Mittelkonsole. Ziel beider Systeme ist eine Reduzierung auf wesentliche Bedienelemente und eine einheitliche Logik, um eine schnelle sowie intuitive Nutzung einer Vielzahl von Funktionen sicherzustellen.

Das Audi *Multi Media Interface* besteht aus zwei Elementen: der Be-

dieneinheit zwischen Schaltung und Mittelarmlehne und räumlich davon getrennt der Anzeigebereich (7" Display) oben in der Instrumententafel. Die ergonomisch günstige Platzierung des Displays im direkten Sichtbereich und die der Bedieneinheit in Handnähe sollen die Blickabwendung von der Straße reduzieren (siehe hierzu auch [17]). Das Display ist darüber hinaus in die Schalttafel versenkbar. In der Bedieneinheit lassen sich die Hauptfunktionen Entertainment, Kommunikation, Information und Steuerung direkt über Funktionstasten aufrufen. Innerhalb dieser Hauptfunktionen sind die jeweils wichtigsten Unterfunktionen über vier Steuerungstasten anwählbar, die kontextabhängig mit verschiedenen Funktionen belegt werden. Diese sind um den zentralen Dreh-Drück-Steuerungsknopf angeordnet [10]. Die aktuelle Belegung dieser *Softkeys* wird in den vier Ecken auf dem Display angezeigt.

Der Schwerpunkt der ergonomischen Betrachtung beim *RNS 510* liegt auf einer einfachen Hand-Auge-Koordination sowie einem leicht erfassbaren Zusammenhang zwischen Funktion und Bedienung, um den kognitiven Aufwand bei der Bedienung zu verringern (siehe auch [159]). Bedien- und Anzeigeelement bilden daher eine Einheit. Neben einem Touchscreen (6,5"), der sowohl der Ein- als auch der Ausgabe dient, sind acht feste Tasten seitlich des Touchscreens und zwei Dreh-Drücksteller seitlich-unterhalb angebracht. Über die Tasten lassen sich die (Hauptmenü-) Funktionen Radio, Media, Telefon, Kartendarstellung, Navigation, Verkehrsinformation, Setup und Lautstärkeinstellungen direkt aufrufen. Die Auswahl von Untermenüpunkten erfolgt über den Touchscreen oder die beiden Dreh-Drücksteller.

Beide Systeme können zumindest teilweise auch per Sprache³ oder mit Hilfe von Bedienelementen am Lenkrad bedient werden. Darüber hinaus werden die Displays in der Mittelkonsole durch ein wesentlich kleineres Display im Kombiinstrument sowie um weitere Bedienmöglichkeiten für Klimaanlage und Radio in der Mittelkonsole erweitert. Im Kombiinstrument wird ein gefilterter, verringerter Umfang möglicher Informationsanzeigen aus dem Mittelkonsolendisplay angezeigt.

Bild 2.13 zeigt im Gegensatz zu den beiden *High-End*-Geräten die entsprechenden Basisausstattungen. Es wird deutlich, dass Einbußen in Bezug auf Bedienkomfort und Wertigkeit in Kauf genommen werden

³Insbesondere für kurze, einfache und schnelle Aktionen ziehen Nutzer jedoch der Sprachbedienung die manuelle Interaktion vor [71, 103].



(a) Radioanlage *concert* von Audi (Quelle: Volkswagen).



(b) *RNS 300* von VW (Quelle: Volkswagen).

Bild 2.13: Gegenüberstellung von zwei marktüblichen Basisausstattungen für Anzeige- und Bediensysteme in der Mittelkonsole.

müssen. Anstelle der Farbdisplays wurde hier auf Monochromdisplays zurückgegriffen. Beide Systeme sind nicht touchfähig, sondern mit einem *Softkey*-Bedienkonzept ausgestattet und vereinen damit Anzeige und Bedienung örtlich. In zukünftigen Systemen wird sich die Größe der Displays nicht maßgeblich verändern, aber es ist davon auszugehen, dass auch hier hochauflösende Farbdisplays zum Einsatz kommen werden. Aufgrund zunehmender FAS- und FIS-Funktionen auch in der Basisausstattung werden die Anzahl der anzuzeigenden Funktionen und damit die Informationsdichte steigen.

2.8.4 Infotainmentsysteme im Forschungsstadium

In den letzten Jahren wurde eine Reihe von Infotainmentsystemen vorgestellt, die sich mit einer intuitiveren Bedienbarkeit beschäftigen. Hier soll eine Auswahl von Systemen vorgestellt werden, deren Schwerpunkt auf einer alternierend- oder synergistisch-multimodalen Bedienung durch haptisch-gestische Eingaben in Kombination mit Sprache liegt oder die sich mit dem Thema ‚Interaktion durch Annäherung‘ oder Zeigegesten befassen. Während im automobilen Bereich solche Systeme noch eher selten sind, haben sie in anderen Bereichen wie z. B. bei Kiosk-Touchscreens bereits eine weitere Verbreitung gefunden und sind kommerziell erhältlich. Hier sei beispielhaft auf den *Siemens Virtual Touchscreen* [51] oder

den *Fraunhofer PointScreen* [54] verwiesen.

Systeme basierend auf Zeigegesten oder Interaktion durch Annäherung Cairnie et al. [35] stellten bereits im Jahr 2000 ein Bedienkonzept zur gestischen Steuerung sekundärer Fahrzeugfunktionen vor. Der Fahrer aktiviert bestimmte Funktionen, indem er mit dem Zeigefinger auf Schaltflächen auf einem Display zeigt. Die Lage der Hand wird dabei mit einem *6DOF-Sensor* (*six-degrees-of-freedom*) ermittelt.

Geiger entwickelte in seiner Arbeit „Berührungslose Bedienung von Infotainment-Systemen im Fahrzeug“ [59] ein System zur gestischen Bedienung des Infotainments. Neben gestischen Eingaben sind in diesem System auch exklusiv-multimodale Spracheingaben vorgesehen.

Jaguar integrierte in seinem *Concept Car* 2007 einen Annäherungssensor, der auf eine sich nähernde Hand mit der Freigabe eines Infotainment-Monitors reagiert. Der mittlere Teil des Instrumentenbretts dreht sich um 90° und gibt einen *Dual View*-Monitor frei, der verschiedene Darstellungen für Fahrer und Beifahrer anzeigt.

Das Bose Media System ist mit einem Drehregler ausgestattet, der auf die Annäherung des Benutzers reagiert. Wird vom Infrarotsensor neben dem Drehsteller eine Hand detektiert, werden auf dem Interaktionsdisplay die Interaktionsmöglichkeiten angezeigt. Die Annäherungsdetektion ist jedoch nicht ortsspezifisch.

Multimodale Bediensysteme Althoff stellt in seiner Arbeit [6] „Ein generischer Ansatz zur Integration multimodaler Benutzereingaben“ einen Algorithmus vor, mit dem alternierend- sowie synergistisch-multimodale Sprache und haptische Bedienung integriert werden können. Allerdings muss der Spracherkenner explizit aktiviert werden.

McGlaun entwickelt den Ansatz weiter [95] und geht vor allem auf den Aspekt des kontextsensitiven Fehlermanagements ein.

Im Rahmen des *SmartKom*-Projekts entstanden mehrere Arbeiten, die sich ebenfalls mit der Integration von Sprache und Zeigegesten beschäftigten. Im Bereich der multimodalen Karteninteraktion wurde ein Prototyp implementiert, der sich bedingt für den Einsatz im Fahrzeug eignet [74], aber ursprünglich zur mobilen Navigation konzipiert wurde.

3 Herleitung eines erweiterten MMK-Modells für Anzeige und Bedienung im Kraftfahrzeug

„Was für eine vorzügliche Einrichtung, dass die Gedanken nicht als sichtbare Schrift über unsere Stirne laufen.“
(Christa Wolf)

Die steigende Anzahl von Fahrerassistenz- und Fahrerinformationsfunktionen veranlassen die Automobilindustrie (siehe auch Bild 1.1), neue HMI-Schnittstellen zu entwickeln, die mit dieser Komplexität umgehen können. Auf der einen Seite müssen dabei speziell automobiler Aspekte wie die Integration der wachsenden Zahl an Funktionen in limitiertem Raum und die Reduktion der Bedienkomplexität adressiert werden. Andererseits müssen aber auch eine möglichst geringe Ablenkung von der Fahraufgabe sichergestellt sowie ergonomische Aspekte beachtet werden.

Eines der Hauptthemen bei der Entwicklung von automobilen HMIs ist die Erweiterung der heutigen Mensch-Maschine-Kommunikation um Kommunikationsprinzipien und Modalitäten, die der Nutzer aus der Humankommunikation kennt.

Humankommunikation ist oft wesentlich effektiver und effizienter als Mensch-Maschine-Kommunikation in heutigen Fahrzeugen. Daher liegt es nahe, detaillierter zu betrachten, wie sich Kfz-Mensch-Maschine-Schnittstellen um Kommunikationsprinzipien und Arten der Modalitätennutzung aus der Humankommunikation sinnvoll erweitern lassen. Im Folgenden werden dafür geeignete Prinzipien identifiziert, und es wird daraus ein neues Kommunikationsmodell für das Fahrzeug abgeleitet.

3.1 Übertragung von Humankommunikationsprinzipien auf die Mensch-Maschine-Kommunikation

Wie in Kapitel 2 bereits dargelegt wurde, bestehen zwischen der Humankommunikation und der Mensch-Maschine-Kommunikation in heutigen Fahrzeugen große Unterschiede. Der Nutzer verhält sich jedoch in der Mensch-Maschine-Kommunikation ähnlich wie in der Kommunikation mit anderen Menschen, wie Reeves und Nass zeigen konnten [124]:

„In short, we have found that individuals’ interactions with computers, television, and new media are fundamentally social and natural, just like interactions in real life. The key word is ‘fundamentally’. Everyone expects media to obey a wide range of social and natural rules. All these rules come from the world of interpersonal interaction, and from studies about how people interact with the real world. But all of them apply equally well to media.“

Die zweite Erkenntnis von Reeves und Nass lautet:

„People respond socially and naturally to media even though they believe it is not reasonable to do so, and even though they don’t think that these responses characterize themselves.“

Aus diesen Erkenntnissen – Menschen interagieren mit technischen Systemen so, wie sie mit anderen Menschen interagieren – lässt sich ableiten, dass sich Prinzipien aus der Humankommunikation auf die Mensch-Maschine-Kommunikation übertragen lassen. Einige dafür besonders geeignete Prinzipien sollen im Folgenden vorgestellt werden.

3.1.1 Emotionale Signalisierung

Viele Benutzerschnittstellen kranken daran, dass untrainierte Nutzer ihr Verhalten an die impliziten funktionalen Regeln des technischen Systems anpassen müssen, um es verstehen und bedienen zu können. Speziell unerwartete Systemreaktionen lassen den Nutzer im Unklaren über Sinn, Relevanz und mögliche Konsequenzen der Ausgabe. Im Kontext automobiler HMIs kann dies nicht nur zu einem negativen Produkterlebnis führen, sondern auch zu erhöhtem Interaktionsaufwand und daraus resultie-

rend zu Sicherheitsrisiken in Folge abnehmender Aufmerksamkeit für die primäre Fahraufgabe. Speziell können inakurate Systemausgaben von Fahrerassistenzsystemen in einen Vertrauensverlust münden oder sogar zu falschen oder verzögerten Reaktionen in (zeit-)kritischen Situationen führen.

Dies ist bei automobilen Schnittstellen heute oftmals auf ihre generelle Restriktion auf reinen Informationsaustausch [144] unter Vernachlässigung anderer Kommunikationsniveaus zurückzuführen, die der Nutzer aus der sozialen Interaktion und interpersonellen Kommunikation kennt. Die meisten Nutzer tendieren dazu, wie bereits erwähnt, nicht-informationstragende Signale (nonverbale Komponenten der Kommunikation und nicht-sprachliche Aspekte des Sprechverhaltens) auch dann zu senden und v. a. zu empfangen, wenn sie mit technischen Systemen interagieren [124].

Nicht-informationstragende Signale (auch implizite Kommunikationsanteile genannt) sind eine wichtige Komponente jedes Kommunikationsprozesses. Sie beeinflussen den Kommunikationsprozess, indem sie Solidarität kreieren sowie Verständnis, Akzeptanz, Priorität, Dringlichkeit, Identifikation, Ablehnung oder Übereinstimmung signalisieren [9]. Für Mensch-Maschine-Systeme, die eine informationsbasierte Kommunikation um zusätzliche Ausdrucksmittel (emotionale Indikation) erweitern, wie wir sie aus der sozialen und interpersonellen Kommunikation kennen, wurde in [12] der Terminus „Emotionale Mensch-Maschine-Schnittstelle“ geprägt. Ein Aspekt ist dabei die Nutzung und Anpassung von bereits vorhandenen impliziten Kommunikationskomponenten. Darüber hinaus können einzelne ausgewählte Mittel, die explizit zur faktischen Information hinzugefügt werden, dazu dienen, Information in Hinblick auf die Bedeutung für den Nutzer zu bewerten: Die Maschine sendet spezifische emotionale Indikationen, die faktische Informationen bewertet, Bedeutung impliziert und implizit Entscheidungsempfehlungen ausspricht (Sicherheit, Dringlichkeit, etc.). Des Weiteren kann emotionale Signalisierung – über den funktionalen Nutzen in Bezug auf Sicherheit und Verständnis hinaus – die Interaktionsqualität aufwerten, in dem gezielt positive Interaktionserfahrungen kreiert werden [12, 67, 123].

Das technische System Fahrzeug und sein spezifischer Kontext bieten im Gegensatz zu anderen Anwendungsgebieten eine gute Basis zur emotionalen Indikation, die es dem Nutzer erlaubt, bewerteten Einblick in den Systemstatus und Verständnis für Systementscheidungen zu erlan-

gen. Darüber hinaus können wichtige Daten für eine implizite Bewertung von faktischer Information aus bereits vorhandenen Daten geschlussfolgert werden, da zukünftige Fahrerassistenzsysteme sowohl den Umfeldkontext als auch die Situation im Fahrzeuginnenraum im Detail überwachen werden.

Obwohl kein umfassend autorisiertes soziales Referenzsystem für eine emotionale Signalisierung existiert, lassen sich doch einige grundlegende Prinzipien bestimmen. Beispielhaft sei hier ein rotes blinkendes Licht im Instrumentenbrett in Kombination mit einem ‚Beep‘-Ton genannt, das einen großen Teil der Information implizit transportiert, sowohl durch die Art und Weise, wie das Licht blinkt als auch durch die Lautstärke des Tons. Auch wenn der Fahrer sicherlich einige Anteile der interpersonellen Kommunikation in der MMK unterdrücken kann und die Interpretation emotionaler Signalisierung interpersonellen Unterschieden unterliegt, empfinden die meisten Fahrer trotzdem ein schnell blinkendes rotes Licht und lauten Klang als aggressiv oder sogar bedrohlich. Zu beachten sind jedoch immer kulturelle Unterschiede, die eine Anpassung an die verschiedenen Märkte notwendig machen können.

In heutigen Fahrzeugen wird diesem Aspekt jedoch nur ungenügend Rechnung getragen, wie sich bei näherer Betrachtung einiger Warnmeldungen zeigt: Die Gurtwarnung bei laufendem Motor wird von einigen Nutzern als weniger aggressiv empfunden als z. B. die Scheibenwischer-Warnung.

3.1.2 Multimodale Kommunikation

In Kapitel 2.3 wurde bereits erwähnt, dass neben der Sprache (auditiver Kanal) Gestik und Blickverhalten (also der taktile und visuelle Kanal) wesentliche Aspekte der Kommunikation sind, die oft parallel zur sprachlichen Äußerung eingesetzt werden.

In der Humankommunikation erfolgt z. B. die Spezifikation eines Ortes, an dem sich ein Objekt befindet („DER Stuhl da drüben“ bei gleichzeitigem Zeigen auf den Stuhl) oft per Zeigegeste, die Spezifikation der Objektart jedoch per Sprache. Diese Kombination aus Sprache und Zeigegesten zur Spezifikation eines Ortes oder von Abmessungen ist sehr effizient. Die zeitliche Synchronisierung der beiden Modalitäten spielt jedoch für das Verständnis des Gegenübers eine wesentliche Rolle.

In heutigen Mensch-Maschine-Systemen, die aus nicht-automobilen

Kontexten bekannt sind, ist zwar die multimodale Eingabe durch Sprache und Gestik teilweise möglich, jedoch muss der Spracherkenner meist explizit aktiviert werden. Bei der Interaktion im Fahrzeug bietet es sich an, den Spracherkenner intentionsbasiert und situationsabhängig zu aktivieren, z. B. auf Annäherung der Hand an das Interaktionsdisplay.

3.1.3 Adaptive Kommunikation

In Kapitel 2.3 wurde bereits dargelegt, dass es für den Sender, sofern er an einer gelungenen Kommunikation interessiert ist, wichtig ist, beim Empfänger eine positive Einschätzung zu hinterlassen. Daher wird der Sender die Interpretationsprozesse des Empfängers antizipieren und entsprechend handeln. Im Gegenzug wird der Empfänger während des Empfangs Signale generieren, die es dem Sender erlauben, Rückschlüsse über die Interpretationsprozesse im Empfänger zu ziehen. Dieser Prozess läuft meist auf einer impliziten Ebene. Maglio et al. [90] bezeichnen technische Systeme, die nicht-informationstragende Elemente der Kommunikation berücksichtigen, als „*attentive user interfaces*“:

„...systems that pay attention to what users do so that they can attend to what users need. Such systems track user behaviour, model user interests, and anticipate user desires and actions.“

In heutigen Mensch-Maschine-Systemen besitzt zwar meist der Nutzer ein mentales Modell des Informationssystems, dem Dialogpartner Computer fehlt aber oft ein derartiges Partnermodell [21]. Ein adaptives System sollte aber, in Anlehnung an die Humankommunikation v. a. die drei folgenden Aspekte berücksichtigen: die Intention des Nutzers, seine Aufmerksamkeit und seinen emotionalen Zustand.

Intentionsbasierte Kommunikation Überträgt man den Interaktionsbegriff im Sinne Paetaus [114] auf das Zusammenwirken von Mensch und Maschine, wird deutlich, dass über eine kommandobasierte Kommunikation hinaus eine auf der Erkennung von Intentionen basierende Reaktion des Systems möglich ist. In dem Maße, in dem es gelingt, die vielfältigen Verhaltensäußerungen von Benutzern zuverlässig zu interpretieren, kann die Mensch-Maschine-Interaktion zu einer ‚intentionsbasierten‘ Kommunikation ausgebaut werden, die eine intuitivere Bedienung von und Kom-

munikation/Interaktion mit Maschinen erlaubt und Frustrationserlebnisse verringert. Im Vergleich zu anderen Kontexten besteht im Fahrzeug eine relativ geringe und klar definierte Anzahl an Interaktionsmöglichkeiten, wodurch eine systemseitige Antizipation der Bedienintention und des Bedienzels relativ gut erreicht werden kann.

Aufmerksamkeitsbasierte Kommunikation In der Humankommunikation wird sich der Sender mit der übermittelten Informationsdichte sowie -form an die jeweils vorhandene Aufmerksamkeit des Empfängers anpassen oder auch versuchen, seine Aufmerksamkeit auf sich oder andere für ihn wichtige Aspekte zu lenken. Daten zur Schätzung der verfügbaren Aufmerksamkeit können im Fahrzeug aus bereits vorhandenen Daten gewonnen werden.

Emotionale Kommunikation Wie bereits mehrfach dargelegt, übermitteln die meisten Nutzer auch in der Kommunikation mit Maschinen nicht-informationstragende Signale. Aufgrund dieser Signale ist es denkbar, dass die Maschine über die Art der Repräsentation von Informationen entscheidet sowie über deren Timing und Penetranz. Emotionen des Nutzers könnten dann nicht nur die Schnittstelle, sondern auch Funktionen selbst kontrollieren, wie z. B. die automatische Abstandsregelung zum Vordermann [11].

3.2 Dialogmodell eines synergistisch-multimodalen Systems

Ausgehend von den vorher genannten aus der Humankommunikation abgeleiteten Interaktionsprinzipien wird im Folgenden ein Dialogmodell für das Fahrzeug vorgestellt, das diese Prinzipien beinhaltet. Die Systemarchitektur erweitert bestehende Modelle, wie z. B. von Akyol et al. [4], und enthält zusätzliche Module zur separaten Auswertung von expliziten und impliziten Eingaben. Darüber hinaus bietet die Architektur erweiterte Eingabemöglichkeiten (Gestik, Sprache und Blickbewegung) und bindet Kontextdaten wie den Fahrzustand und die Fahrsituation in das Modell ein (siehe Bild 3.1).

Wesentliche Aspekte sind der Modalitätsmanager, der Kommunikationsinterpret und der Präsentationsmanager. Der Modalitätsmanager

koordiniert die Kombination zueinander in Beziehung stehender Eingaben über verschiedene Modalitäten zu einem gemeinsamen Befehl und leitet diesen an den Informationsmanager weiter. Der Kommunikationsinterpret wertet die impliziten Anteile aller vom System erfassten Eingaben aus. Die ermittelten Parameter der Kommunikationssituation werden vom Kommunikationsinterpret sowohl an den Informationsmanager über den Modalitätsmanager als auch an den Präsentationsmanager weiter geleitet.

Eingabemodalitäten Für den Fahrer bestehen verschiedene Möglichkeiten, Eingaben zu treffen. Die (klassischen) Eingabemodalitäten wie Taster, Dreh-/Drücksteller, Touchscreen und Sprache stehen nach wie vor zur Verfügung. Als neue Eingabemöglichkeiten sind im Modul ‚Touchscreen und Schalter‘ sog. Touchgesten vorgesehen. Darunter werden einerseits Gesten auf dem Touchscreen, wie z. B. Wischgesten, aber auch das Einzeichnen von Gebieten oder die Eingabe von bestimmten Symbolen zum Aufrufen von ‚*Shortcuts*‘ (direktes Wechseln in Funktionen) verstanden. Bild 3.2 zeigt Beispiele für eine Wischgeste sowie für das Einzeichnen eines Gebiets auf dem Display.

Das Modul ‚Gestik-Sensorik‘ ist die Erweiterung von bestehenden Systemen um die Möglichkeit von Handgesten und dient der Überwachung von Handbewegungen. Unter Handgesten werden hier dynamische Gesten, d. h. kontinuierliche Bewegungen wie Wischbewegungen, verstanden. Der Einsatz statischer Gesten wird nicht weiter betrachtet, da es sich hierbei nicht um eine natürliche Art der Kommunikation handelt. Die Überwachung der Handposition dient vorwiegend der impliziten Interpretation (siehe Kommunikationsinterpret).

Das Modul Spracherkennung weist gegenüber heutigen Systemen die Möglichkeit der Erfassung impliziter Anteile bei der Sprachsteuerung auf, bietet jedoch dem Nutzer darüber hinaus keine neuen Eingabemöglichkeiten.

Ein weiteres neues Eingabemodul ist die Blickerfassung. Hier werden nicht nur die Blickbewegung sondern auch weitere Blickparameter wie Sakkaden, Fixationen o. ä. erfasst. Ein Beispiel expliziter Interaktion ist in dieser Modalität die Auswahl von Menüpunkten aus einer Liste mittels Blicksteuerung, ein Beispiel für eine implizite Auswertung des Blicksignals ist die Schlussfolgerung auf die visuelle Aufmerksamkeit.

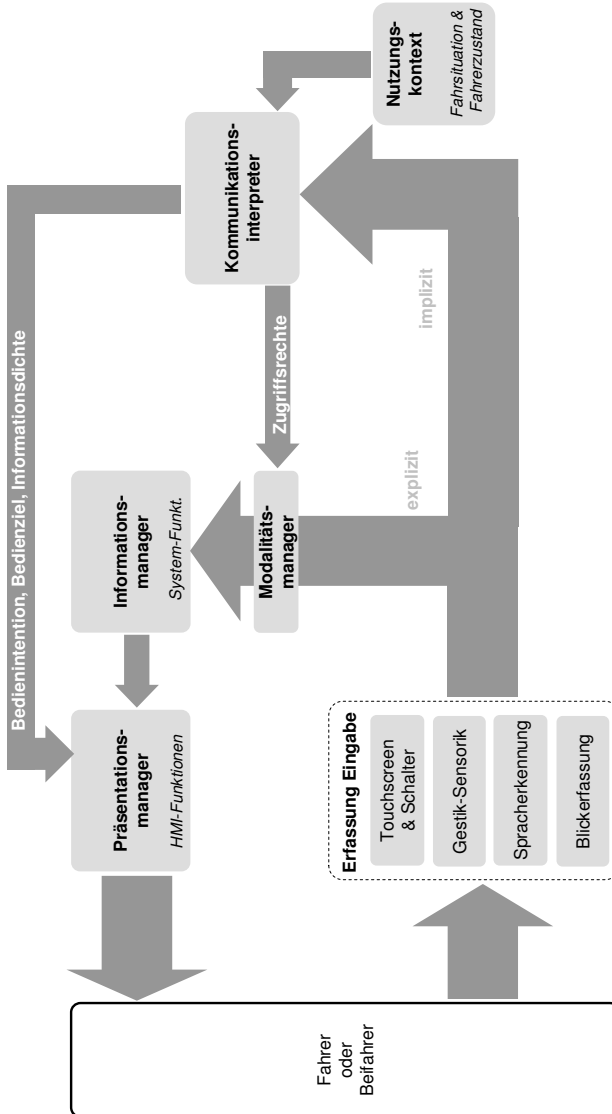


Bild 3.1: Architektur der erweiterten Benutzerschnittstellen im Kfz.



(a) Wischgesten auf dem Touchscreen zum Weiterschalten des Musiktitels.



(b) Einzeichnen eines Gebiets in einer Kartendarstellung.

Bild 3.2: Erweiterte Eingabemöglichkeiten auf dem Touchscreen.

Alle Eingabemodalitäten (haptisch, gestisch, akustisch, visuell) werden sowohl explizit als auch implizit ausgewertet. Explizite Eingaben sind dabei:

- Tastendrucke, Schalterbetätigungen, Drehstellerbetätigungen etc.,
- Touchgestik, Touchscreen-Tastendrucke,
- dynamische Gesten zur Steuerung einer Funktion,
- Eingabe von Sprachbefehlen,
- Auswahl von Objekten oder Menüpunkten durch Blickbewegung.

Als implizite Eingaben können ausgewertet werden:

- Unterscheidung von Fahrer und Beifahrer bei Schalter-/Taster- oder Drehstellerbedienung,
- Annäherung der Hand an bestimmte Bedienelemente (Koordinaten der Bedienhand, Bewegungsverlauf),
- Klang der Stimme, Prosodie, Schallrichtung, emotionale Färbung etc.,
- Blickbewegung, Gesichtsausdruck etc.

Aber auch nicht erkannte Eingaben können implizit interpretiert werden und so z. B. eine intuitive und adaptive Hilfe aufrufen.

Kommunikationsinterpretierer Der Kommunikationsinterpretierer wertet die impliziten Anteile der Nutzereingaben aus und verbindet diese mit

Daten aus dem Nutzungskontext, d. h. mit Daten über den Fahrerzustand oder die Fahrsituation. Anhand dieser Daten kann einerseits eine Systemreaktion veranlasst werden, ohne dass eine explizite Nutzereingabe erfolgt ist. Andererseits kann so die Ausführung von Funktionen eingeschränkt oder erweitert werden. Darüber hinaus kann das System Ausgaben an den Fahrer anpassen.

Für die Informationsfusion bieten sich verschiedene aus der Literatur bekannte Methoden wie z. B. Bayes'sche Netze oder Regelwerke an. Folgende Parameter können daraus abgeleitet werden:

- Grad der Aufmerksamkeit, Fokussierung der Aufmerksamkeit,
- Bedienintention (Will der Nutzer bedienen?),
- Bedienzziel (Welche Funktion will der Nutzer bedienen?),
- emotionaler Zustand des Nutzers,
- Zugriffsrechte auf bestimmte Funktionen,
- kognitive Belastung des Fahrers,
- Fahrer-/Beifahrer-Unterscheidung.

Informationsmanager Der Informationsmanager ist der klassische Kern des HMIs. An ihn sind die Funktionen angebunden. Er entscheidet mit den ihm zur Verfügung stehenden Informationen, welche Funktionen ausgelöst oder geschaltet werden.

Präsentationsmanager Der Präsentationsmanager entscheidet, in welcher Art und Weise Rückmeldung an den Fahrer gegeben wird. Klassisch passiert dies visuell, wobei die Ausgabe unabhängig von der Fahrer- und Fahrsituation wiederholbar gleich ist. In dem hier beschriebenen Ansatz kann die Präsentation aufbauend auf den Daten aus dem Kommunikationsinterpret und dem Informationsmanager variieren. Beispielfhaft sei hier die Anzeige von Informationen auf unterschiedlichen Displays je nach Blickrichtung oder Veränderung von Bildschirminhalten bei Annäherung einer Hand genannt. Die Ausgabe kann jedoch nicht nur visuell, sondern auch akustisch oder haptisch erfolgen. Hierzu wurden bereits in [11] verschiedene Ansätze aufgezeigt.

Informationsfusion im Modalitätsmanager Im Gegensatz zu heute verfügbaren Eingabesystemen sind die Modalitäten nicht mehr getrennt von-

einander zu betrachten. Während heute ein Bedienschnitt entweder haptisch oder sprachlich getätigt werden kann, sind in zukunftsweisenden Systemen synergistisch-multimodale Eingaben möglich. Als Beispiel seien hier *Point-to-Talk*-Funktionen (sprachliche Äußerung „Was bedeutet diese Taste da?“ in Kombination mit einer Zeigegeste) angeführt. Der Modalitätsmanager kombiniert die verschiedenen parallelen Eingaben zu einem Eingabebefehl, den er dann an den Informationsmanager weiterreicht (siehe auch Kapitel 5).

3.3 Prototypisch implementiertes System

Eine erste Implementierung einer adaptiven multimodalen Benutzerschnittstelle wurde am Beispiel eines Navigationssystems und eines Mediaplayers unter Einbeziehung einer Spracherkennung, einer Annäherungserkennung, einer Zeigegestenerkennung und eines Touchscreen-Interfaces realisiert (siehe Bild 3.3). Die implizite Nutzung von Eingabeinformationen wird anhand der Auswertung von Annäherungs- und Zeigegesten demonstriert, die Integration von synergistisch- und alternierend-multimodalen Eingaben erfolgt anhand eines *Point-to-Talk*-Szenarios sowie der ‚ortsspezifischen Suche‘ (näheres hierzu findet sich in Kapitel 6.2).

Folgende Aspekte wurden dabei speziell betrachtet:

- Erfassung der Position und Bewegungsrichtung der Bedienhand des Nutzers zur Bestimmung von Bedienintention und Bedienzziel;
- Fusion von multimodalen Befehlen zu einem kombinierten Eingabebefehl.

Zunächst wurde dafür ein Touchscreen einer Mittelkonsole in einem Fahrzeug mit einer Annäherungssensorik ausgestattet. Details dazu finden sich in Kapitel 4. Darauf aufbauend wurde ein System entwickelt, mit dem es möglich war, die Position einer Hand vor dem Touchscreen zu bestimmen.

In einem zweiten Schritt wurde dieser näherungsensitive Touchscreen mit einem Spracherkennungserweiterer erweitert, um die Szenarien zur multimodalen Integration implementieren zu können.

Alle implementierten Interaktionsverfahren wurden umfassend evaluiert. Die Ergebnisse finden sich in Kapitel 7.

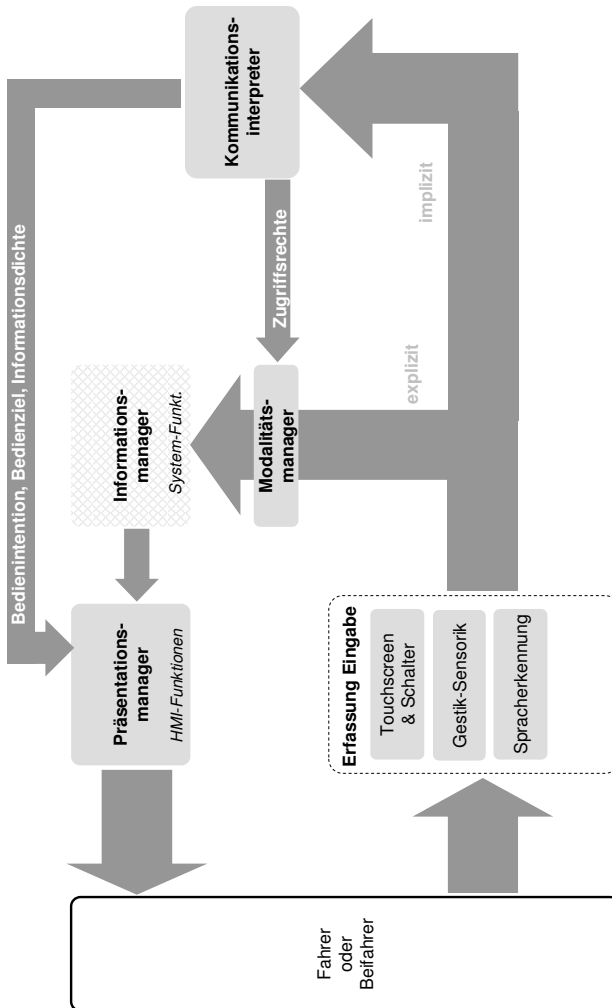


Bild 3.3: Architektur des implementierten Mensch-Maschine-Systems.

4 Bestimmung der Position einer Hand vor einem Display

„Der Wechsel allein ist das Beständige.“
(Arthur Schopenhauer)

Die implizite Nutzung von Eingabeinformationen, wie in Kapitel 3.2 dargestellt, wird in der vorliegenden Arbeit an der Auswertung von Annäherungs- und Zeigegesten vor einem Touchscreen demonstriert. Hierfür ist es notwendig, einen Touchscreen mit einer robusten Annäherungssensorik auszustatten, die darüber hinaus in der Lage ist, die Position eines Bedienfingers vor dem Display möglichst genau zu bestimmen.

4.1 Motivation zum Einsatz kapazitiver Sensorik

4.1.1 Anforderungen

Detektionsbereich Bereits in einem Abstand von 10 cm soll die Präsenz einer Hand vor dem Display detektiert werden. Der detektierte Raum soll sich dabei möglichst ausschließlich vor dem Display befinden. Die Unterscheidung, ob sich eine Bedienhand vor dem Display oder seitlich davon befindet, ist hardware- oder softwareseitig so genau wie möglich zu treffen. Innerhalb des Detektionsraumes soll die Position eines Bedienfingers ab einem Abstand von 4 cm einen Cursor auf dem Display so präzise steuern, dass die Abweichung vom Schnittpunkt eines Lots von der Fingerspitze auf die Displayoberfläche maximal 1 cm beträgt.

Hohe Robustheit Die Sensorik sollte möglichst robust und unempfindlich gegenüber Störeinflüssen arbeiten. Dabei sollte sichergestellt werden, dass die Sensoren möglichst ausschließlich auf eine Hand bzw. auf Körperteile des Nutzers reagieren und nicht auf andere im Detektionsbereich befindliche Gegenstände. Fehler und fehlerhaftes Auslösen der Sensoren,

z. B. durch Gegenstände im Detektionsbereich des Sensors, sollten also auszuschließen oder auf Sonderfälle reduzierbar sein. Darüber hinaus sollte die Handhabung der Bedienelemente mit einem Handschuh oder nassen Händen sowie nassen Handschuhen möglich sein. Unterschiedliche Handgrößen und Handgesten (Handhaltung) bei der Bedienung dürfen die Erkennungsleistung nicht maßgeblich beeinflussen.

Geringer systemtechnischer Aufwand Die Integration der Sensoren in bestehende Touchscreen-Systeme sollte mit möglichst geringem systemtechnischen Aufwand möglich sein. Es muss ein Kompromiss zwischen einer fehlerrobusten Lösung auf der einen Seite und den Mehrkosten an Rechenleistung und Zeit auf der anderen Seite gefunden werden.

Verbau der Sensoren Material und Design der Oberflächenwerkstoffe sollten durch die Sensoren nicht beeinträchtigt werden. Darüber hinaus dürfen die Sensorik-Bauteile im Interieur-Sichtbereich nicht erkennbar sein. Im Fall von optischen Sensoren bedeutet dies, dass sie z. B. im Dachhimmel verbaut werden müssten, wo sie für die Insassen nicht direkt sichtbar sind. Grundsätzlich sollten die Sensoren in einem Fahrzeug der Golf-Klasse innerhalb der dort vorherrschenden Baumaße einfach zu verbauen sein.

Kalibrierung Notwendige Kalibriermaßnahmen sollten bereits herstellerseitig durchführbar sein oder im Betrieb der Sensorik automatisiert und für den Nutzer nicht erkennbar durchgeführt werden.

Echtzeitfähigkeit Die Übertragung und Auswertung der Sensorwerte muss in Echtzeit erfolgen, sodass für den Nutzer bei der Bedienung keine spürbare Verzögerung entsteht.

4.1.2 Kamerabasierte Gestenerkennung

Viele aus dem Stand der Technik bekannte Systeme zur Gestenerkennung, die durchaus eine gewisse Verwandtheit mit dem hier entwickelten System aufweisen, arbeiten kamerabasiert. Dabei werden aus den Bilddaten Merkmale extrahiert, die für die Mustererkennung verwendet werden können. Zunächst muss aber (im Falle einer Handgestik-Erkennung) die

Hand vom Bildhintergrund getrennt werden (örtliche Segmentierung). Akyol et al. [3] beschreiben im Detail ein videobasiertes System zur Gestenbedienung im Fahrzeug. Üblicherweise werden bei Kamerasystemen, die für den Fahrzeugeinsatz entwickelt wurden, Beleuchtungen im nicht-sichtbaren Bereich eingesetzt. Sich verändernde Bildhintergründe, wie sie im Fahrzeug durch die stark wechselnden Lichtverhältnisse auftreten, stellen jedoch nach wie vor ein großes Problem dar. Die Folge ist zusätzlicher Rechen- und Hardwareaufwand mit den damit verbundenen Kosten. Darüber hinaus ist ein designneutraler Verbau von Kamerasystemen im Fahrzeug schwierig.

4.1.3 Infrarotbasierte Gestenerkennung

Geiger [59] stellt in seiner Arbeit „Berührungslose Bedienung von Infotainment-Systemen in Fahrzeug“ ein Verfahren zur infrarotbasierten Gestenerkennung vor. Dabei müssen die Sensoren in der Mittelkonsole unterhalb des Gestenraumes angebracht werden. Die dreidimensionale Detektion der Hand erfolgt über die Kombination mehrerer Sensoren zu einem zweidimensionalen Array, das die Abstände von den Sensoren zur Hand misst. Die Anzahl der eingesetzten Sensoren sowie deren Platzierung und Ausrichtung bestimmen den Informationsgehalt. Die örtliche Segmentierung ist bei diesem Verfahren leichter zu realisieren, da der Szenenhintergrund einen deutlich größeren Abstand zum Sensor-Array hat als die zu detektierende Hand. Für die Merkmalsgewinnung sind nach Geiger [59] keine aufwändigen Berechnungen zur Datenreduktion nötig, da jedes Abtastintervall einen Messwert pro Sensor liefert und diese Distanzmesswerte fast direkt als Merkmale verwendet werden können.

Vorteil gegenüber kamerabasierten Lösungen sind die Unempfindlichkeit der Sensoren gegenüber Fremdlichteinflüssen und die geringe Größe der Sensoren. Nachteilig ist hier jedoch, wie auch bei kamerabasierten Systemen, dass ein designneutraler Verbau im Fahrzeug nicht gewährleistet ist. Um die Funktionsfähigkeit der Sensoren zu erhalten, muss eine direkte Sichtverbindung zwischen dem Sensor und dem zu detektierendem Objekt – in diesem Fall der Hand – bestehen. Dies ist ohne einen Durchbruch in der Oberfläche nicht realisierbar.

4.1.4 Kapazitive Gestenerkennung

Kapazitive Sensoren sind ähnlich wie Infrarot-Sensoren unempfindlich gegenüber Fremdlichteinflüssen und lassen sich aufgrund der Baugröße gut in bestehende Systeme integrieren. Des Weiteren ist ein designneutraler Verbau hinter Kunststoffblenden, wie sie im Fahrzeug Verwendung finden, gut möglich. Die dreidimensionale Detektion der Hand erfolgt wie bei Infrarot-Systemen über die Kombination mehrerer Sensoren (siehe auch [139]). Auch hier ist keine aufwändige Berechnung zur Datenreduktion nötig, da jedes Abtastintervall einen Messwert pro Sensor liefert und diese Distanzmesswerte in die weitere Berechnung einfließen können. Ebenso können die weiteren der o. g. Anforderungen mit kapazitiven Messsystemen erfüllt werden.

Auch Pickering [116] weist darauf hin, dass sich für automobiler Gestenerkennungssysteme kapazitive Systeme gut eignen. Allerdings existieren keine weiteren Veröffentlichungen über die technische Umsetzung.

4.2 Kapazitive Sensorik

Der Begriff der Kapazität ist eng mit dem Bauteil Kondensator verwandt. Zum besseren Verständnis kapazitiver Detektoren ist es jedoch erforderlich, diesen Kapazitätsbegriff zu erweitern. Sobald mehrere Elektroden in einem System verwendet werden, lässt sich dieses nur mehr durch eine Kapazitätsmatrix beschreiben, nicht jedoch durch eine einzelne Kapazität. Ein kapazitiver Sensor wertet Änderungen in der Kapazitätsmatrix aus, die durch eine Person oder ein Objekt verursacht wurden und kann so z. B. auf die Annäherung einer Person schließen.

4.2.1 Theorie der kapazitiven Abstandsmessung

Werden zwei elektrische Leiter L_1 und L_2 , wie in Bild 4.1 dargestellt, durch ein Dielektrikum getrennt, spricht man von einem elektrischen Kondensator. Ist ein Leiter mit der Ladung $Q_1 = Q$ geladen und der andere Leiter ungeladen, so stellt sich zwischen beiden die Spannung $U_1 = U$ ein. Nach Maxwell gilt sowohl für Spannung als auch für Ladung das Prinzip der linearen Superposition. Dies bedeutet, dass ein Erhöhen der Ladung auf das m -fache zu einem m -fach größeren Potenzial zwischen den beiden Leitern führt. Dieser Zusammenhang wird durch die

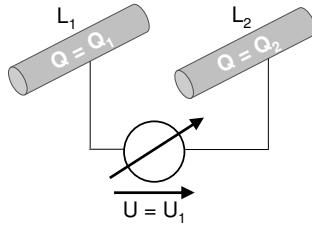


Bild 4.1: Schematischer Aufbau eines Kondensators [22].

Gleichung für die Kapazität C

$$C = \frac{Q}{U} \quad (4.1)$$

beschrieben. Einen Ansatz zur Berechnung von Kapazitäten bieten die Maxwell-Gleichungen. Da kapazitive Sensoren mit Frequenzen bis ca. 1 MHz, meist jedoch darunter arbeiten, liegen die Wellenlängen oberhalb 300 m. Im Fahrzeug ist die Sensoranordnung also immer wesentlich kleiner als die Wellenlänge. Daher treten keine Wellenlängenphänomene wie Brechung, Beugung oder Reflexion auf und die Maxwell-Gleichungen für quasistatische Felder vereinfachen sich zu [92]:

$$\nabla \cdot \mathbf{D} = \rho, \quad (4.2)$$

$$\nabla \times \mathbf{E} = 0. \quad (4.3)$$

Die Integration über ein Volumen V (nach dem Integralsatz von Gauß) zeigt, dass die Ladung innerhalb eines Volumens V gleich dem Fluss durch dessen Oberfläche ∂V ist:

$$\oint_{\partial V} \mathbf{D} \, d\mathbf{A} = \int_V \rho \, dV. \quad (4.4)$$

In den hier verwendeten kapazitiven Sensoren werden üblicherweise isotrope Dielektrika verwendet. Daher kann nach [22] die Dielektrizitätszahl ϵ als skalare Größe eingesetzt werden:

$$\mathbf{D} = \epsilon \mathbf{E}, \quad (4.5)$$

wobei $\epsilon = \epsilon_0 \cdot \epsilon_r$ ist. Betrachtet man das Dielektrikum als homogen, so folgt daraus:

$$\epsilon \oint_{\partial V} \mathbf{E} \, d\mathbf{A} = \int_V \rho \, dV. \quad (4.6)$$

Wird die Leitfähigkeit des Dielektrikums vernachlässigt, sodass es als ladungsfrei betrachtet werden kann, und wählt man das Volumen V , über das in Gleichung (4.6) zu integrieren ist so, dass es den gesamten Leiter L_1 umschließt, ergibt sich gemäß Gleichung (4.7) die auf diesem Leiter befindliche Ladung Q als Ergebnis des Volumenintegrals [22]:

$$\int_V \rho \, dV = Q. \quad (4.7)$$

Einsetzen von Gleichung (4.7) in Gleichung (4.6) führt zu:

$$\oint_{\partial V} \mathbf{E} \, d\mathbf{A} = \frac{Q}{\epsilon}. \quad (4.8)$$

Die elektrische Feldstärke \mathbf{E} kann also aus dem Quotienten Q/ϵ und einer Funktion $\mathbf{f}(x, y, z)$ berechnet werden:

$$\mathbf{E} = \frac{Q}{\epsilon} \mathbf{f}(x, y, z). \quad (4.9)$$

Hierbei beschreibt $\mathbf{f}(x, y, z)$ ausschließlich geometrische Eigenschaften.

Aufgrund der oben angeführten Anwendbarkeit der linearen Superposition kann die Gegenelektrode L_2 ohne Beschränkung der Allgemeinheit als ladungsfrei angenommen werden. Wegen ihrer hinreichend großen Leitfähigkeit können die Leiter als Äquipotenzialflächen angesehen werden. Daher kann die Spannung U zwischen den beiden Leitern als Integral über die elektrische Feldstärke \mathbf{E} entlang einer beliebigen Linie \mathbf{s} , die von einer zur anderen Elektrode verläuft, berechnet werden. Dieser Zusammenhang stellt sich wie folgt dar [22]:

$$U = \int \mathbf{E} \, d\mathbf{s}. \quad (4.10)$$

Damit ergibt sich für die Berechnung der Spannung U :

$$U = \frac{Q}{\epsilon} \int \mathbf{f}(x, y, z) \, d\mathbf{s}. \quad (4.11)$$

Vergleicht man Gleichung (4.11) mit Gleichung (4.1), folgt daraus für die Kapazität C zwischen den Leitern L_1 und L_2 :

$$C = \frac{\epsilon}{\int \mathbf{f}(x, y, z) \, d\mathbf{s}}. \quad (4.12)$$

Wie man daraus ersehen kann, hängt die Kapazität C eines Kondensators ausschließlich von der Dielektrizitätszahl ϵ sowie der geometrischen Anordnung und Form seiner Elektroden ab.

Ersetzt man einen oder beide Leiter durch ein anderes Objekt, kann man über die Kapazität, die der Leiter (ein Sensor) und das Objekt darstellen, auf die Distanz zwischen Sensor und Objekt schließen und das Objekt verfolgen. Das Objekt muss allerdings genügend freie Ladungen zur Verfügung stellen, um die negativen Ladungen auf der Kondensatorplatte ausgleichen zu können.

Auch die menschliche Hand und ein Leiter (eine Metallplatte) bilden einen Kondensator. Die Hand muss dabei allerdings nicht, wie in Bild 4.2 gezeigt, geerdet sein, da der menschliche Körper bereits ein genügend großes Ladungsreservoir darstellt. Bild 4.3 zeigt die dielektrische Konstante ϵ für Muskelgewebe in Abhängigkeit der Frequenz. Prinzipiell

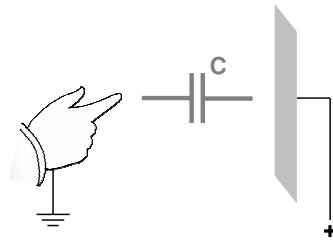


Bild 4.2: Die menschliche Hand und eine Metallplatte bilden einen Kondensator.

können kapazitive Detektoren zur Objektverfolgung auf größere Distanzen, bis etwa 3 m verwendet werden [158].

4.2.2 Anwendungsmodi kapazitiver Sensoren

Bild 4.4 zeigt das Ersatzschaltbild für eine Anordnung mit zwei Elektroden E_1 und E_2 und einem leitfähigen Objekt, z. B. einer Hand. $C_{E_1 E_2}$

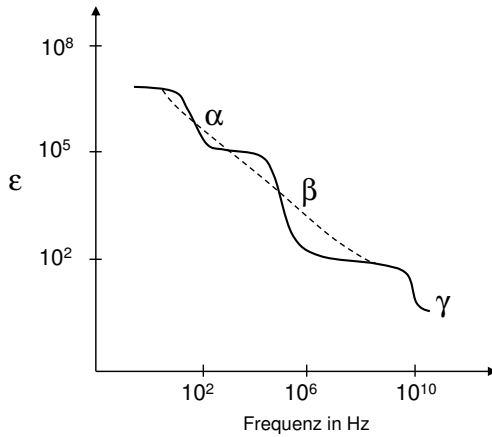


Bild 4.3: Verlauf der dielektrischen Konstante ϵ für Muskelgewebe in Abhängigkeit der Frequenz. Nach Daten von [132].

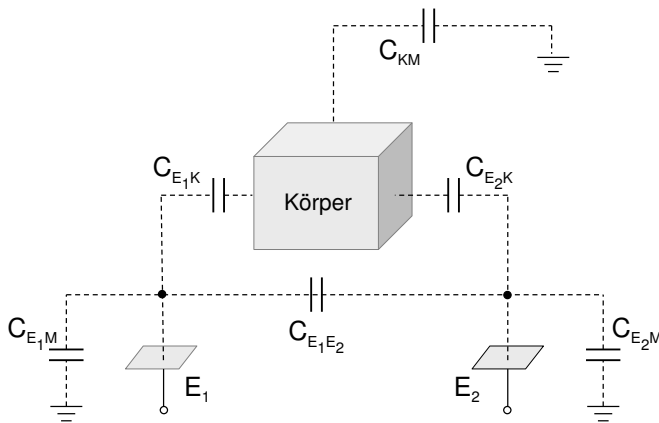


Bild 4.4: Ersatzschaltbild für eine kapazitive Detektoranordnung mit zwei Elektroden und einem leitfähigen Objekt oder Körper [92].

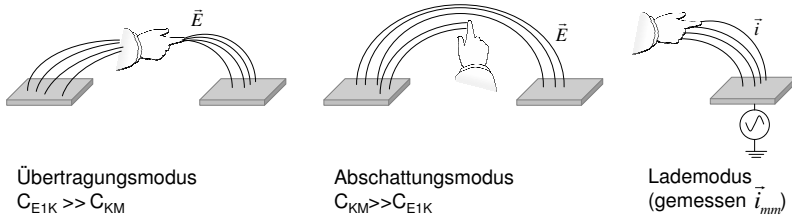


Bild 4.5: Drei Modi zur kapazitiven Messung [140].

bezeichnet die direkte Kopplung zwischen den Elektroden, C_{E_1M} und C_{E_2M} die direkte Kopplung der Elektroden an Masse. C_{E_1K} und C_{E_2K} stellen die Kapazitäten zwischen Elektroden und Körper dar, und C_{KM} ist die kapazitive Erdung des Körpers.

Der Körper beeinflusst sowohl die Kopplung der Elektroden untereinander als auch die Kopplung an Masse. Bei verhältnismäßig geringer Erdung des Körpers (C_{KM} verhältnismäßig klein) kann von einer Elektrode zur anderen gesendet werden. Die Gegenwart des Körpers erhöht also die Kopplung der Elektroden. Dieser Modus wird nach [140] als „*transmit mode*“ (Übertragungsmodus) bezeichnet.

Im umgekehrten Fall (große Erdung des Körpers) schirmt der Körper die Elektroden gegeneinander ab, die Kopplung der Elektroden sinkt (siehe auch [92]). Smith [140] bezeichnet diesen Modus als „*shunt mode*“ (Abschattungsmodus) in Anlehnung an die abschirmende Wirkung des Körpers.

Wird zur Detektion des Körpers jedoch nicht die Änderung der Kopplung gemessen, sondern die Strom-Matrix, wird die Messung des Sendestroms i_{mm} einer Elektrode E_m als Messung im „*loading mode*“ (Lademodus) bezeichnet.

Bild 4.5 zeigt die drei Modi schematisch. Besondere Beachtung verdient die Tatsache, dass der Mensch Teil des Sensorsystems ist und damit das Messergebnis maßgeblich durch seine physiologischen Eigenschaften beeinflusst, wie z. B.:

- die unterschiedliche Kopplung des Körpers zu den Elektroden,
- die Handgröße und -geometrie,

- die Handhaltung (Faust oder gestreckte Finger).

4.2.3 Grenzen kapazitiver Sensorik

Exponentielle Kapazitätsabnahme mit dem Abstand Eine der größten Hürden für die Integration kapazitiver Sensorik im Bereich ‚Detektion von Menschen‘ ist die begrenzte Reichweite. Die sich aus der Gleichung (4.12) ergebende allgemeine Gleichung für einen Plattenkondensator

$$C \approx \frac{\epsilon A}{d} \quad (4.13)$$

impliziert, dass die Kapazität eines Kondensators indirekt proportional zur Distanz zwischen den Platten ist. Dies gilt jedoch nur für geringe Distanzen zwischen den Platten. Je weiter die Platten auseinander liegen, desto geringer ist das überlappende Gebiet relativ zur Umgebung. Bild 4.6 verdeutlicht dies. Daher ist ein realistischeres Modell der Kapa-

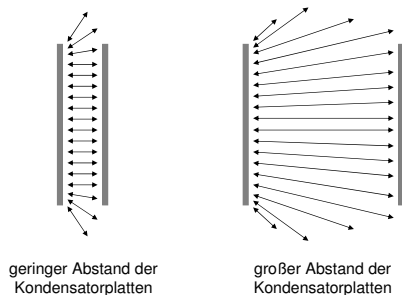


Bild 4.6: Mit steigender Distanz zwischen den Kondensatorplatten steigt die Streukapazität. Dieses Verhalten beschränkt die Reichweite kapazitiver Sensoren.

zität:

$$C \approx \frac{\epsilon A}{d^x}, \quad (4.14)$$

wobei die Konstante x , abhängig von der Umgebung, zwischen 1 und 3 liegt [158]. Dieser rapide Kapazitätsabfall mit steigender Distanz wirkt

das Problem einer Detektion in größeren Abständen auf. Die abnehmende Auflösung kann durch weitere Sensoren oder Filteralgorithmen zum Teil kompensiert werden.

Schließen von Objekten Kapazitive Sensoren sind anfällig für externe Störungen. Andere Objekte, die in die Nähe des Sensors kommen, können die Messung beeinflussen. Läuft z. B. eine Person in der Nähe des Sensors vorbei, kann dies mit der eigentlichen Messung interferieren. Daher ist es nicht einfach, verlässliche kapazitive Sensoren für stark frequentierte Umgebungen zu entwickeln. Im Fahrzeug, wo die Positionen der Insassen jedoch grundlegend bekannt sind, können diese Probleme zumindest teilweise gelöst werden. Trotzdem dämpfen Objekte in der Nähe der Sensoren die Signale und verringern die Präzision der Messung. Dies kann auch nur teilweise durch Schirmung verhindert werden. Wenn das zu detektierende Objekt bekannt ist, können jedoch vorab Muster für die Erkennung festgelegt werden. Nicht erwünschte Objekte können so von der Messung ausgeschlossen werden.

Begrenzte und uneindeutige Daten Die einzige Information, die ein kapazitiver Sensor zurückgibt, ist die Kapazität und die Änderung der Kapazität über die Zeit. Ein und dieselbe Kapazität kann von einer Person resultieren, die direkt vor dem Sensor steht oder von zwei Personen, die etwas weiter weg stehen. Kapazitive Sensoren können daher nicht zwischen Objekten unterscheiden, die dieselbe Kapazität verursachen. Hände verschiedener Nutzer zu unterscheiden ist nahezu unmöglich. Disambiguierung kann jedoch oft durch Verwendung weiterer Sensoren und anschließender Filterung der Daten sowie durch Kalibrierung (siehe auch Kapitel 6.3) erreicht werden.

4.2.4 Qualitative Beschreibung des kapazitiven Sensorsystems

Kapazitive Sensoren im Übertragungsmodus Neben der eigentlichen Positionsbestimmung der Hand sollte mit dem in dieser Arbeit eingesetzten kapazitiven System auch eine Unterscheidung zwischen Fahrer und Beifahrer getroffen werden. Daher wurde auf ein System im Übertragungsmodus zurückgegriffen, das durch die Möglichkeit der Signalübertragung über den Körper eine eindeutige Zuordnung zulässt. Dabei wird Information durch Modulation eines elektrischen Feldes, das

auf den Körper gekoppelt wird, codiert: Der Transmitter wird kapazitiv mit dem Empfänger durch den menschlichen Körper gekoppelt. Die Person kann dabei auf der Oszillatorelektrode sitzen, stehen, sie berühren oder anderweitig in der Nähe sein. Über den Körper wird das Signal im Pikoampere-Bereich zu einer oder mehreren Empfangselektroden übertragen, die das Signal demodulieren. Der Rückkanal ist über die Luft und die Erde gewährleistet. Bild 4.7 zeigt dies schematisch. Transmitter und Empfängerelektroden können als Kondensatorplatten mit einem Dielektrikum dazwischen modelliert werden. Solche Systeme sind – neben dem Vorteil der Möglichkeit der Datenübertragung – unempfindlicher gegenüber Störungen, wie sie sich z. B. durch Gegenstände in der Nähe oder nicht gewollte Bedienung z. B. von den hinteren Sitzen aus ergeben.

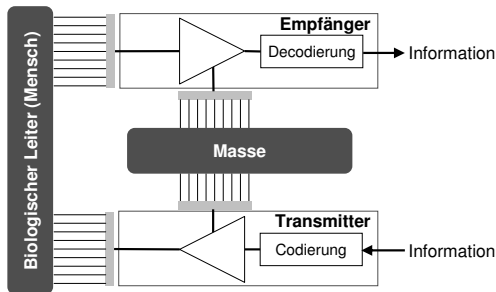


Bild 4.7: Prinzipschaltbild kapazitiver Sensorik im Übertragungsmodus nach [160].

Identifizierung der Sensorkennlinie Zur Identifizierung der Sensorkennlinien wurden Messreihen für die Annäherung einer Hand an ein Sensorelement in verschiedenen Raumrichtungen mit jeweils mehreren Benutzern aufgenommen. Anhand dieser Messungen konnte gezeigt werden, dass die Messwerte m für einen gegebenen Abstand r der Hand vom Sensor in guter Näherung mit einer Normalverteilung modelliert werden können. Die Breite der Streuung der Werte (Standardabweichung) in jedem Abstand r der Hand vom Sensor wird im Wesentlichen durch drei Parameter beeinflusst (siehe Bild 4.8):

1. Eigenrauschen der Sensoren,

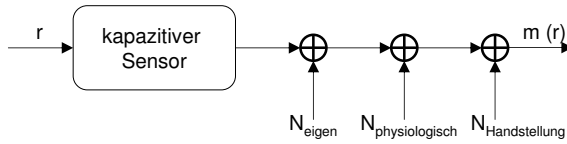


Bild 4.8: Rauschquellen im System.

2. Rauschen durch unterschiedliche Handgeometrie,
3. Rauschen durch unterschiedliche Handstellungen bei der Bedienung,

wobei der letzte Punkt den größten Einfluss auf die Standardabweichung hat.

Betrachtet man die Mittelwerte der Verteilungen an den einzelnen Punkten, zeigt sich, dass sich die Sensorkennlinie für alle Personen und in alle Richtungen mit geringem Fehler quadratisch approximieren lässt. Bild 4.9 zeigt sowohl die Kennlinie als auch die zugehörige Standardabweichung an fünf Messpunkten exemplarisch. Auffallend ist die Zu-

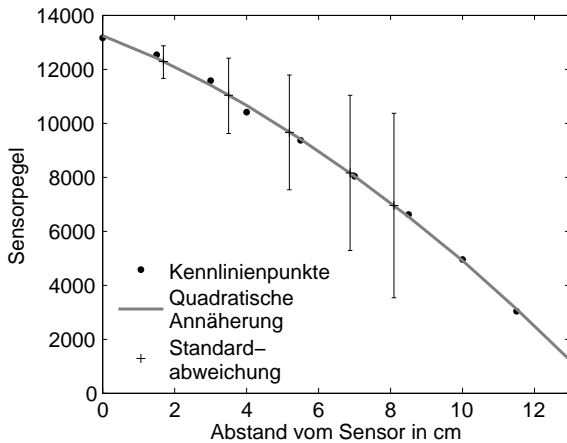


Bild 4.9: Kennlinie des Sensors.

nahme der Standardabweichung bei sinkendem Pegel (bedeutet größerer

Abstand zum Sensor). Das Rauschen ist demnach eine Funktion des zu bestimmenden Abstandes der Hand vom Sensor.

4.3 Positionsermittlung der Hand vor einem Display

Für die Positionsermittlung der Hand vor einem Display wurden vier Sensoren rings um das Display angeordnet. Die Positionsbestimmung aus den gemessenen Werten der einzelnen Sensoren erfolgt durch eine Maximum-Likelihood-Schätzung.

4.3.1 Statistische Modellierung der Sensoren

Im Folgenden bezeichne die Variable m den gemessenen Wert, d. h. den ausgegebenen Sensorwert/-pegel des kapazitiven Abstandssensors. Im Sinne der Statistik handelt es sich bei m um eine Zufallsvariable. r bezeichnet den wahren Wert, d. h. den Abstand des Messpunktes vom Sensor. Die Funktion $\sigma(r)$ beschreibt die (empirische) Standardabweichung in Abhängigkeit des wahren Wertes.

Die parabolische Kennlinie des Sensors lässt sich beschreiben durch:

$$m(r) = ar^2 + br + c. \quad (4.15)$$

Die bedingte Wahrscheinlichkeitsdichtefunktion unter Annahme von gaußverteilten Sensorwerten ergibt sich damit zu:

$$p(m|r) = \frac{1}{\sqrt{2\pi} \sigma(r)} \cdot e^{-0,5 \cdot \left(\frac{m-ar^2-br-c}{\sigma(r)}\right)^2}, \quad (4.16)$$

wobei die Standardabweichung eine Funktion von r ist und in guter Näherung linear mit dem Abstand r wächst:

$$\sigma(r) = t + kr. \quad (4.17)$$

Bild 4.10 zeigt beispielhaft die Likelihoodfunktion in Abhängigkeit des Radius.

Der Abstand r vom Sensor lässt sich im dreidimensionalen Raum als Radius einer Kugel um den Sensor an der Stelle (x_0, y_0, z_0) deuten:

$$r = \|\mathbf{x} - \mathbf{x}_0\| = \sqrt{(x - x_0)^2 + (y - y_0)^2 + (z - z_0)^2}. \quad (4.18)$$

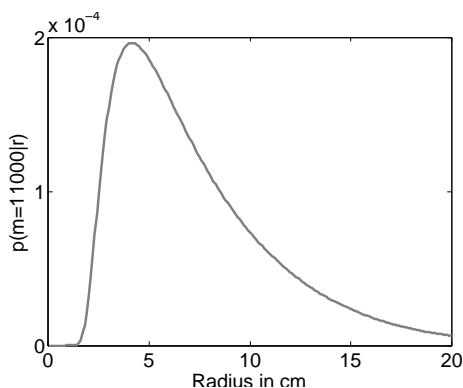


Bild 4.10: Likelihoodfunktion $p(m|r)$ für $m = 11000$.

Damit lassen sich die vorgestellten Gleichungen durch Substitution in ein euklidisches Koordinatensystem überführen. Für die distanzabhängige Standardabweichung gilt nun:

$$\sigma(\mathbf{x}) = t + k \cdot \|\mathbf{x} - \mathbf{x}_0\|. \quad (4.19)$$

Die Likelihoodfunktion lässt sich dann folgendermaßen angeben:

$$p(m|\mathbf{x}) = \frac{1}{\sqrt{2\pi} \sigma(\mathbf{x})} \cdot e^{-0,5 \cdot \left(\frac{m - a \cdot \|\mathbf{x} - \mathbf{x}_0\|^2 - b \|\mathbf{x} - \mathbf{x}_0\| - c}{\sigma(\mathbf{x})} \right)^2} \quad (4.20)$$

Zur besseren Veranschaulichung der Formeln zeigt Bild 4.11 beispielhaft die Likelihoodfunktion in Abhängigkeit von x und y für $z = 0$.

4.3.2 Maximum-Likelihood-Schätzung

Um eine sensorunabhängige Position der Hand zu schätzen, muss im Folgenden ein globales Koordinatensystem eingeführt werden. Dabei bietet es sich an, einen der vier Sensoren als Koordinatenursprung festzulegen. Abhängig von diesem Ursprung werden dann die Einbauorte der restlichen drei Sensoren als Offset verrechnet. Damit kann die resultierende Likelihoodfunktion als Fusion der vier Sensorwerte angegeben werden, wobei die experimentell ermittelten Likelihoodfunktionen der einzelnen

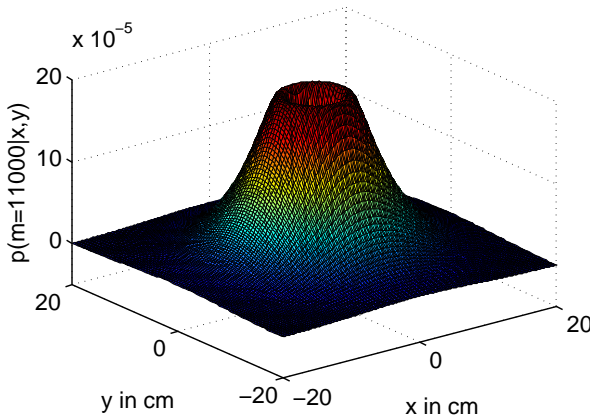


Bild 4.11: Likelihoodfunktion $p(m|x, y)$ für $m = 11000$.

Sensoren mit Hilfe des Offsetvektors \mathbf{x}_i parametrisiert werden:

$$L(\mathbf{x}) = \prod_{i=1}^4 p(m_i|\mathbf{x} - \mathbf{x}_i) \quad (4.21)$$

Wird diese Funktion in Abhängigkeit von \mathbf{x} maximiert, erhält man mit dem Argument des Maximums den Wert, an welchem die Funktion ihr Maximum annimmt (Maximum-Likelihood-Schätzung für \mathbf{x}).

Die Maximierung der Funktion erfolgt im Allgemeinen durch Differentiation nach \mathbf{x} und anschließender Nullpunktssuche. Da dies – wie allgemein bei Dichtefunktionen mit komplizierten Exponentenausdrücken – sehr aufwändig werden kann, wird die logarithmierte Likelihoodfunktion verwendet. Diese propagiert aufgrund der strengen Monotonie der Logarithmusfunktion die Extrempunkte der nicht-logarithmierten Likelihoodfunktion. Durch die Logarithmierung vereinfacht sich die Likelihoodfunktion zu:

$$l(\mathbf{x}) = \ln \left(\prod_{i=1}^4 p(m_i|\mathbf{x} - \mathbf{x}_i) \right) \quad (4.22)$$

$$= \sum_{i=1}^4 \ln(p(m_i|\mathbf{x} - \mathbf{x}_i)) . \quad (4.23)$$

Bild 4.12 zeigt beispielhaft für drei Sensoren die Likelihoodfunktionen in Abhängigkeit von x und y , Bild 4.13 zeigt die resultierende, fusionierte Likelihoodfunktion.

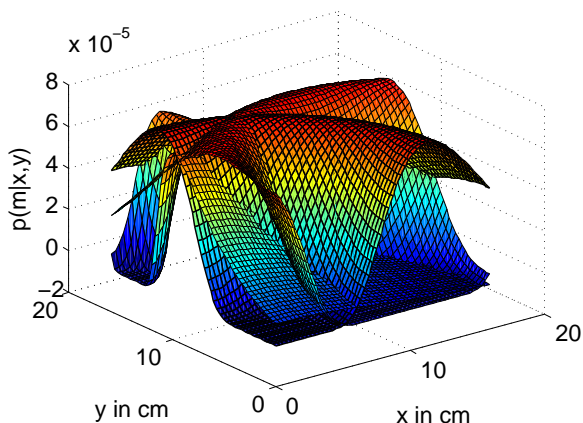


Bild 4.12: Likelihoodfunktionen für drei Sensoren an den Koordinaten $(0|0)$, $(10|0)$ und $(0|14)$.

Aufgrund der vom Radius abhängigen Standardabweichung zeigt sich (siehe Bild 4.13), dass die resultierende Likelihoodfunktion nicht mehr mit einer Normalverteilung angenähert werden kann.

Der Ort $\hat{\boldsymbol{x}}$ des Maximums der Likelihoodfunktion, der den geschätzten Ort eines Fingers vor einem Display beschreibt, lässt sich folgendermaßen schätzen:

$$\hat{\boldsymbol{x}} = \arg \max_{\boldsymbol{x}} (l(\boldsymbol{x})) \quad (4.24)$$

$$= \arg \max_{\boldsymbol{x}} \left(\sum_{i=1}^4 \ln(p(m_i | \boldsymbol{x} - \boldsymbol{x}_i)) \right). \quad (4.25)$$

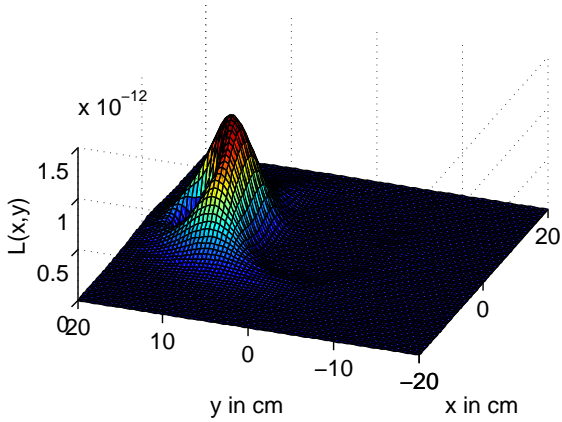
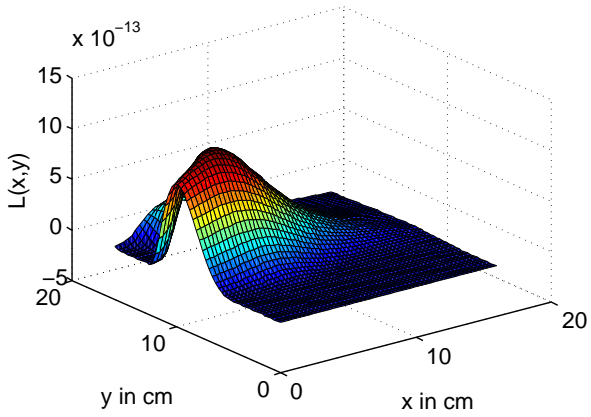


Bild 4.13: Fusionierte Likelihoodfunktion.

5 Methodik und Verfahren zur Fusion multimodaler Eingaben

„Der Mensch ist ein sprachbegabtes Tier und wird sich immer durch das Wort verführen lassen.“
(Simone de Beauvoir)

Für die Nutzung von Informationsteilstücken, die über verschiedene Modalitäten eingegeben wurden, ist es notwendig, diese zu fusionieren, um ein passendes Systemkommando zu erzeugen, das der Nutzerintention entspricht. Die einzelnen Informationsteile werden aus den verschiedenen Erkennern geliefert, die jeweils auf eine Modalität spezialisiert sind.

Zur Realisierung solcher multimodaler Interaktionssysteme wurde in der Vergangenheit eine Reihe von Konzepten entwickelt: Bekannte Systemansätze sind u. a. Quickset [41], MATIS [107], Quickfusion [142] und ICARE [26, 27] sowie die Arbeiten von Althoff [6]. Diese verbinden gemäß [106] multimodale Informationsanteile auf einer rein semantischen Ebene.

In dem in dieser Arbeit entwickelten System liegt der Schwerpunkt auf der Optimierung des Zusammenspiels von Spracheingabe und Touchscreen für die Anwendung im Fahrzeug sowie der Nutzung impliziter Informationsanteile (Annäherung) zur situationsadaptiven Aktivierung des Spracherkenners.

5.1 Motivation zum Einsatz einer Soft-Decision-Fusionsmethode

5.1.1 Anforderungen

Einbeziehung impliziter Kommunikationsanteile Der Modalitätsmanager muss in der Lage sein, implizite Kommunikationsanteile, wie z. B. die Annäherung der Hand an ein Display oder den Blickfokus, auszuwerten. Aufgrund dieser Daten können bestimmte Eingabemöglichkeiten freigegeben werden und Sensoren, wie z. B. der Spracherkennung, aktiviert

werden. Darüber hinaus ist es mittels impliziter Kommunikationsanteile möglich, das Konfidenzmaß für das Gesamterkennungsergebnis zu beeinflussen.

Wahl der möglichen Eingabemodalitäten Während [6] davon ausgeht, dass der Benutzer jederzeit die freie Wahl zwischen mehreren Modalitäten haben soll, wird die Wahl der Eingabemodalitäten im hier vorgestellten Modalitätsmanager vom Systementwickler nach ergonomischen Gesichtspunkten vorgegeben. Es gilt zwar, dass möglichst redundante Bedienmöglichkeiten vorgesehen werden sollten, jedoch ist eine bestimmte Systemeingabe selten über alle Modalitäten möglich. Die einzelnen Systemfunktionalitäten sind allerdings immer über eine unimodale Kommunikationsart sowie häufig über multimodale Interaktionsarten aktivierbar.

Zeitliche Synchronisation der Benutzereingaben und Verarbeitung unimodaler Benutzereingaben Generell muss der Modalitätsmanager in der Lage sein, zeitlich versetzte Eingaben, die zu einem gemeinsamen Eingabebefehl kombiniert werden sollen, als solche zu erkennen. Der zeitliche Versatz kann dabei einerseits vom Benutzer verursacht werden und andererseits durch unterschiedliche Erkennelaufzeiten der einzelnen Eingabekomponenten zu Stande kommen. Im Allgemeinen kann davon ausgegangen werden, dass die Erkennungsergebnisse zeitlich versetzt beim Modalitätsmanager eintreffen, weswegen eine Synchronisation aller Eingabekomponenten von fundamentaler Bedeutung ist.

Darüber hinaus müssen jedoch auch unimodale Benutzereingaben als solche erkannt werden und dürfen nicht weiter fusioniert werden.

Quantitative Beurteilung des Erkennungsergebnisses Aufgrund der komplexen Struktur möglicher Benutzereingaben sind im Allgemeinen mehrere Systemreaktionen impliziert. Um jedoch die vom Benutzer intendierte Systemreaktion auszugeben, muss zunächst ein Maß für die Qualität der einzelnen Erkennungsergebnisse erzeugt werden. Aus den einzelnen Erkennungsergebnissen und deren Konfidenzmaßen kann dann ein quantitativer Vertrauenswert für das Gesamterkennungsergebnis errechnet werden, der für die Bestimmung der Systemreaktion herangezogen werden kann.

Situationspezifisches Fehlermanagement Fehler bei der Interaktion mit komplexeren Infotainment-Systemen sind einerseits auf fehlerhafte Nutzereingaben, andererseits aber auf Systemfehler zurückzuführen. Systemfehler entstehen z. B. durch die Verwendung statistischer Klassifikatoren, die immer nur mit einer bestimmten Wahrscheinlichkeit auf die Korrektheit einer erkannten Nutzereingabe schließen können. Fehlerhafte Nutzereingaben entstehen z. B. durch mangelnde Kenntnis des Systems auf Seiten des Nutzers. Ein intuitiv bedienbares System sollte geeignete Strategien vorhalten, um beiden Fehlerquellen entgegenzuwirken oder um Maßnahmen zur Fehlerbehebung einzuleiten.

Erweiterbarkeit der Eingabemodule In Hinblick auf zukünftige Weiterentwicklungen sollte der Modalitätsmanager so ausgestaltet sein, dass weitere Eingabemodule einfach angebunden werden können. Daher sollte eine einheitliche Schnittstelle zwischen Modalitätsmanager und den einzelnen Eingabemodulen definiert werden. Darüber hinaus muss das Regelwerk für die in einer bestimmten Situation zugelassenen Eingaben erweiterbar sein.

Geringer systemtechnischer Aufwand Die Integration multimodaler Eingaben in bestehende Systeme sollte mit möglichst geringem systemtechnischen Aufwand verbunden sein, damit eine weitestgehend echtzeitfähige kfz-taugliche Anwendung sichergestellt ist. Daher gilt es, einen Kompromiss zwischen den Vorteilen einer intuitiveren und fehlerrobusteren Bedienung auf der einen Seite und den Mehrkosten an Zeit und Rechenleistung auf der anderen Seite abzuschätzen und realistisch zu bewerten [6].

5.1.2 Multimodale Fusionsmethoden

Early-Signal-Fusion Wie bereits in Kapitel 2.5.4 erwähnt, ist die frühest mögliche Ebene in der Informationsfusion die Sensordatenfusion. Dabei werden Rohdaten aus den Sensoren fusioniert. In multimodalen Systemen wird dies meist auf der Basis von Hidden-Markov-Modellen, neuronalen Netzen oder Bayes'schen Netzen erreicht. Diese Art der Fusion eignet sich besonders gut für zeitlich synchronisierte Werte. Nachteilig dabei ist, dass die Daten, die aus den unterschiedlichen Erkennernmodulen bereitgestellt werden, ein gemeinsames Datenformat aufweisen müssen.

Early-signal-Fusionsmethoden werden z. B. bei der Fusion von Kamerabilddaten mit Infrarotbilddaten oder bei der Spracherkennung in Kombination mit Lippenbewegungsdaten eingesetzt (detaillierte Ausführungen finden sich in [83]).

Late-Semantic-Fusion Multimodale Systeme, die mit *Late-Semantic*-Fusionsansätzen arbeiten, integrieren ebenfalls mehrere Einzelerkener. Für jeden Erkener werden jedoch separat Datenvorverarbeitung, Merkmalsextraktion und Klassifikation durchgeführt. Die Einzelergebnisse der Erkener werden dann zu einem Gesamtergebnis fusioniert. Jedes Erkennungsergebnis ist mit einem Konfidenzmaß aus dem Einzelerkener versehen. Diese Konfidenzmaße werden in einem weiteren Schritt zu einem Gesamtkonfidenzmaß, meist durch lineare Kombination, fusioniert. Der Vorteil dieses Verfahrens liegt v. a. darin, dass verschiedene Erkener unabhängig voneinander betrieben werden können. Darüber hinaus ist die Erweiterbarkeit des Systems einfacher möglich als bei *Early-Signal*-Fusionsansätzen. Nachteilig ist hier, dass die Entscheidung für ein bestimmtes Ergebnis im Erkener gefällt wird, unabhängig von den Ergebnissen der anderen Erkener.

Soft-Decision-Fusion *Soft-Decision*-Fusionsalgorithmen kombinieren Ansätze aus der *Early-Signal*-Fusion und der *Late-Semantic*-Fusion. Es werden nicht nur die Konfidenzmaße des eigentlichen Erkenerergebnisses in die Fusion mit einbezogen, sondern auch N -best-Listen aus jedem Erkener. So kann sichergestellt werden, dass bei der Fusion nicht unbedingt immer das Einzelerkenerergebnis mit der höchsten Konfidenz verwendet wird, sondern das Ergebnis, das zur höchsten Gesamtkonfidenz beiträgt.

5.2 Integrationsmethodik im Modalitätsmanager auf semantischer Ebene

Unabhängig von der Fusionsebene nach [127] stehen verschiedene Integrationsmethoden zur Verfügung [6], von denen zwei Verfahren im vorliegenden System zur Anwendung kommen.

5.2.1 Integrationsverfahren

Temporale Integration Hier werden die Eingabedaten in ihrer zeitlichen Relation betrachtet. Sie werden als zusammengehörig erachtet, falls sie sich überlappen oder innerhalb eines definierten zeitlichen Abstandes aufeinander folgen. Aufgrund der Aktivierung des Spracherkenners durch Annäherung der Hand an das Display steht im Gegensatz zu klassischen *Push-to-Talk*-Konzepten, bei denen zu Beginn der Spracheingabe eine Taste gedrückt wird, für die Fusion kein definierter Anfangszeitpunkt der Spracheingabe zur Verfügung. Die Auswertung aufgrund des Endpunktes der Spracheingabe gestaltet sich insofern schwierig, da die Sprachbefehle sehr unterschiedlich lang sind (vgl. Künstler „Beatles“ mit einem Songtitel „Ich werde mein Leben lang üben, dich so zu lieben, wie ich dich lieben will, wenn du gehst“). Daher ist eine temporale Integration im vorliegenden System nur schwer möglich.

Regelbasierte Integration Die Integration wird auf Basis eines Regelwerkes, wie z. B. einer kontextfreien Grammatik oder einem ‚*Slot-Filling*-Algorithmus‘ vorgenommen. Die in Kapitel 6.2 beschriebenen Anwendungsszenarien weisen einen eher niedrigen Komplexitätsgrad auf, was im Kontext Fahrzeug aufgrund der vorhandenen Primäraufgabe Fahren auch gewünscht ist. Aufgrund dieser Tatsache wird auf ein relativ einfaches Verfahren zur Integration zurückgegriffen, dass sich jedoch am *Slot-Filling* orientiert. Alle möglichen Eingabebefehle werden dafür in drei verschiedene Kategorien unterteilt:

- *Vollkommandos* bestehen aus einer einzigen, nicht integrierbaren Anweisung.
- *Funktionen* können sowohl mit Parametern integriert werden als auch alleine ausgeführt werden. In den vorliegenden Bedienszenarien ist einer Funktion maximal ein weiterer Parameter zugeordnet, der u. U. auch optional sein kann.
- *Parameter* können nur zusammen mit Funktionen ausgeführt werden. Steht keine passende Funktion zur Verfügung, wird der eingegebene Parameter ignoriert.

Alle Eingabebefehle können weitere Werte – wie z. B. die Identifikationsnummer eines POIs – mitführen. Diese Werte sind jedoch für die Integration irrelevant.

Im Gegensatz zum klassischen *Slot-Filling* kann die Integration hier effizienter ablaufen, da die Parameter erst beim Auftreten einer Funktion auf Integrierbarkeit untersucht werden. Vollkommandos und Funktionen brauchen nicht auf ihre Integrierbarkeit mit anderen Funktionen oder anderen Vollkommandos untersucht werden.

Statistische Integration Statistische oder stochastische Integration bezeichnet das Kombinieren von Eingabedaten, die entweder mit Unsicherheit behaftet sind oder bei denen der Erkenner mehrere Alternativen liefert. Im hier vorliegenden System sollen beide Alternativen betrachtet werden. Der Spracherkener liefert sowohl N -Best-Listen als auch Konfidenzmaße für die einzelnen Erkenergebnisse.

5.2.2 Hypothesengenerierung und -verarbeitung

Zunächst empfängt die Fusionseinheit die einzelnen Teilinformationen (klassifiziert als Vollkommando, Funktion oder Parameter) und erstellt daraus jeweils eine Grundhypothese. Jeder Hypothese ist dabei ein Konfidenzmaß $\mathcal{K}_{\text{hyp},m}$ zugeordnet, das vom Erkener mitgeliefert wurde. Touchscreen-Eingaben werden dabei mit der Konfidenz 100 % versehen, der Spracherkener liefert zusammen mit der N -Best-Liste für jedes Ergebnis der Liste ein korrespondierendes Konfidenzmaß.

Vollkommandos werden sofort der Menge der aktiven Hypothesen zugeordnet, da hier keine weitere Fusion stattfinden kann. Parameter und Funktionen werden nach den o. g. Regeln mit den aktiven Hypothesen integriert und ebenfalls der Menge der aktiven Hypothesen hinzugefügt. Die Ursprungshypothesen werden dabei unverändert beibehalten. Das Konfidenzmaß \mathcal{K}_{hyp} für eine zusammengesetzte Hypothese ergibt sich dabei wie folgt:

$$\mathcal{K}_{\text{hyp}} = \frac{1}{M} \sum_{m=1}^M \mathcal{K}_{\text{hyp},m} \cdot p + (M-1), \quad (5.1)$$

wobei $(M-1)$ die Anzahl der mit der Funktion integrierten Parameter ist. p ist ein Faktor, der dazu dient, zusammengesetzte Hypothesen nicht-zusammengesetzten Hypothesen gegenüber zu bevorzugen.

Alle neuen Hypothesen mit den neu zugeordneten Konfidenzwerten werden fortlaufen an die Menge der aktiven Hypothesen weitergeleitet

und dort zusammen mit allen aktiven Hypothesen auf eine mögliche Ausführung hin überprüft. Dabei müssen folgende Kriterien eingehalten werden:

1. Für jedes Vollkommando muss die bestmögliche Hypothese ausgeführt werden; jedes Vollkommando darf nur ein einziges Mal ausgeführt werden.
2. Aus jeder N -Best-Liste darf nur ein Eintrag ausgeführt werden.
3. Funktionen müssen alle erforderlichen Parameter besitzen, bevor sie ausgeführt werden. Parameter können nicht ohne Funktion ausgeführt werden.
4. Die minimale Lebensdauer muss erreicht sein. Einige Funktionen haben optionale Parameter, die u. U. zeitlich erst nach der Funktion eingegeben werden. Um eine Integration dieser Parameter zu ermöglichen, müssen die Funktionen eine bestimmte Zeit ‚warten‘ bevor sie ohne optionalen Parameter ausgeführt werden.
5. Die maximale Lebensdauer darf nicht überschritten werden. Funktionen, denen Parameter fehlen, ‚warten‘ eine bestimmte Zeit in der Fusionskomponente. Erst nach dieser Zeitspanne werden sie aus der Menge der aktiven Hypothesen gelöscht.
6. Das Gesamtkonfidenzmaß einer Hypothese muss einen bestimmen Wert überschreiten. Dafür werden eine untere und eine obere Konfidenzschranke definiert. Überschreitet eine Hypothese die obere Schranke, wird sie sofort ausgeführt, wenn sie die o. g. Kriterien erfüllt. Unterschreitet eine Hypothese die untere Schranke, wird sie aus der Menge der aktiven Hypothesen gelöscht. Liegt eine Hypothese zwischen der unteren und der oberen Schranke, bleibt sie in der Menge der aktiven Hypothesen, bis die maximale Lebensdauer erreicht wurde. Wurde bis dahin keine Integrationsmöglichkeit mit einer höheren Konfidenz gefunden, wird die Hypothese ausgeführt.

Es ist zu beachten, dass die Reihenfolge der Benutzereingaben nicht notwendigerweise mit der Reihenfolge der Systemreaktionen korreliert. Insbesondere wenn die Eingabereihenfolge Funktion \rightarrow Vollkommando \rightarrow Parameter ist, wird das Vollkommando vor der Funktion ausgeführt, die

erst mit dem Parameter vollständig wurde. Beispielhaft sei hier eine Eingabe ‚Touch auf Künstler‘ → ‚Touch auf *Play*‘ → ‚Beatles‘ genannt, die in der Reihenfolge ‚Musik starten‘ → ‚Wechsel des Künstlers‘ ausgeführt werden würde.

5.2.3 Fehlermanagement

Wie bereits Kapitel 5.1.1 beschrieben, soll das System einerseits geeignete Strategien vorhalten, um Nutzerfehlern entgegenzuwirken und um Maßnahmen zur Fehlerbehebung einzuleiten. Andererseits sollen Systemfehler weitestgehend vermieden werden. Systemfehler werden im vorliegenden System v. a. dadurch reduziert, dass eine zentrale Konfiguration des Systems vorgenommen wird und keine doppelten Listen oder Regelwerke vorliegen. Es existiert also z. B. nur eine Liste mit Musiktiteln, auf die alle Komponenten (Spracherkennung, Hypothesengenerierung) zugreifen.

Darüber hinaus werden fehlerhafte Nutzereingaben (z. B. Touch auf einen POI ohne weitere sprachliche oder haptische Interaktion) in der GUI signalisiert. Voraussetzung ist, dass eine Hypothese ihre maximale Lebensdauer erreicht hat und die einzelnen zu fusionierenden Teile nicht anderweitig verwendet werden konnten. Des Weiteren wird bei gegensätzlichen Eingaben, die nicht aufgelöst werden können, ein Hinweis ausgegeben.

6 Realisierung eines fahrzeugtauglichen Systems zur synergistisch-multimodalen Eingabe

„Die Idee ist das Absolute, und alles Wirkliche ist nur Realisierung der Idee.“
(Hegel)

6.1 Systemaufbau

Bild 6.1 enthält eine schematische Darstellung des im Rahmen der vorliegenden Arbeit entwickelten Gesamtsystems. Zentraler Bestandteil ist die Evaluierungsschicht. Diese umfasst die Komponenten Modalitätsmanager, Kommunikationsinterpretierer und Informationsmanager. Gemäß der Kommunikationsabhängigkeiten in kontextadaptiven FIS und in Anlehnung an das Seeheim-Modell verwenden wir eine einfache 3-Ebenen-Architektur für deren Realisierung, da die Kommunikationsabhängigkeiten anderer Muster über die Anforderungen kontextadaptiver Anwendungen im Fahrzeug hinausgehen [11, 63].

Die verwendete Schichten-Architektur (siehe auch [11] und Bild 6.1) umfasst drei funktionale Ebenen und eine Wissensbasis:

- Sensorsystemsicht: Hardware und Softwarealgorithmen zur Erfassung von elementaren Kontextinformationen und Benutzereingaben; Konvertierung der Kontextinformationen basierend auf einer kontextfreien Grammatik mittels *Look-Up*-Tabelle in ein erkennerunabhängiges Format (Module „CM“, nach [94]);
- Evaluierungsschicht: Interpretation der Dialogsituation sowie Interpretation von expliziten Eingaben;
- Anwendungsschicht: Entscheidung über Ort und Zeit der Ausgabe der relevanten Informationen für den Nutzer;

- Benutzermodell: Informationen bezüglich individueller Merkmale sowie Bedienhistorie der Interaktion des Benutzers mit dem Fahrzeug.

Die Kommunikation zwischen den einzelnen Schichten erfolgt durch einen bidirektionalen Austausch von String-Messages, die über TCP/IP übertragen werden. Das System ist als Client-Server-Struktur aufgebaut; die Evaluierungsschicht stellt dabei den zentralen Server dar, an dem sich die Sensorsystemkomponenten und die Anwendungskomponenten anmelden. Eine solche verteilte Architektur ermöglicht die flexible Anbindung weiterer Anwendungsmodule und bietet darüber hinaus die Möglichkeit einer Verteilung der Komponenten auf verschiedene Rechner.

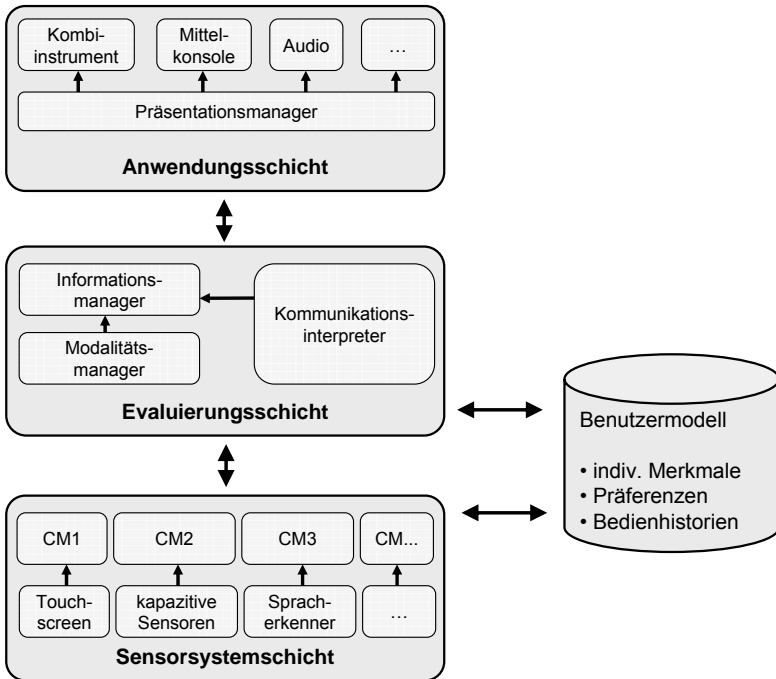


Bild 6.1: Schematischer Aufbau des Gesamtsystems.

6.2 Implementierte Anwendungsszenarien

Wie bereits in Kapitel 3.3 eingeführt, wird die Implementierung der adaptiven multimodalen Benutzerschnittstelle am Beispiel eines Navigationssystems und eines Mediaplayers demonstriert. Als Eingabemodalitäten stehen der Touchscreen sowie eine Spracherkennung zur Verfügung. Die implizite Nutzung von Interaktionsanteilen wird anhand der Auswertung der Position der Bedienhand gezeigt. Synergistisch-multimodale oder alternierend-multimodale Eingaben werden am Beispiel von *Point-to-Talk*-Anwendungen und räumlichen Spezifikationen demonstriert.

6.2.1 Einblendung und Vergrößerung von Schaltflächen bei Annäherung

Die Effizienz bei der Bedienung von Touchscreens hinsichtlich Bedienzeiten und Bedienfehlern steigt mit zunehmender Größe der Schaltflächen. Bei aktuellen touchscreen-basierten Benutzerschnittstellen im Fahrzeug sind die Dimensionen der bedienbaren Schaltflächen jedoch oft sehr stark reduziert, um bei limitierter Displaygröße Platz für die parallele Informationsanzeige zu schaffen. Durch den situativen Wechsel zwischen einem sog. ‚Anzeigemodus‘ und einem ‚Bedienmodus‘ kann die limitierte Anzeigefläche von Touchscreens im Fahrzeug unter Annahme einer Bedienintention effizient sowohl für die Informationsanzeige, als auch für die Darstellung von Bedienflächen genutzt werden (siehe Bild 6.2).

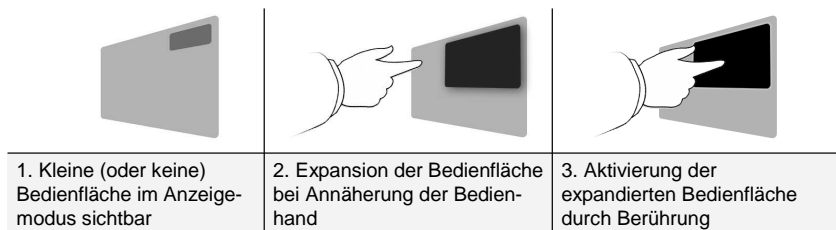


Bild 6.2: Grundsätzliches Bedienprinzip des adaptiven Touchscreens.

Im vorliegenden System wird die Annäherung einer Bedienhand, wie bereits mehrfach erwähnt, als Bedienintention interpretiert und es wird auf das Bedienziel des Nutzers geschlossen: Eine definierte Annäherung

an den Touchscreen wird ab einem Abstand von ca. 10 cm als generelle Bedienintention betrachtet. Daraufhin wird die Informationsdarstellung zugunsten einer verbesserten Bedienbarkeit der angebotenen Schaltflächen durch Skalierung oder Einblendung zusätzlicher Bedienoptionen optimiert. Ab einer Distanz von ca. 4 cm ergibt sich das konkrete operative Bedenziel aus der jeweils vom zeigenden Finger des Nutzers anvisierten Schaltfläche auf dem Touchscreen. Diese Schaltfläche kann dann entweder zusätzlich skaliert werden, sodass der Zielvorgang zum Betätigen der Taste nochmals vereinfacht wird oder es werden Bedienoptionen erst im Detail sichtbar. Je nach Gestaltungsvariante kann die verfügbare Anzeigefläche zum Zeitpunkt der Bedienung fast vollständig zur Darstellung von Bedienflächen und bei Nicht-Bedienung vollständig für die Informationsanzeige genutzt werden.

Maßnahmen zur Anpassung von Bedienoberflächen an die Bedienintention und das Bedenziel des Nutzers müssen hinsichtlich ihrer ergonomischen Qualität sehr sorgfältig gestaltet werden (siehe dazu auch Kapitel 2.7). Mögliche negative Effekte sind u. a. schlechte Vorhersagbarkeit, inkonsistentes Systemverhalten, Intransparenz, sowie falsches oder aufdringliches Systemverhalten. Bei der Implementierung sind daher über die Richtlinien zur Gestaltung interaktiver, bildschirmbasierter Anwendungen (siehe u. a. die Norm EN ISO 9241 [48]) hinaus folgende ergänzende Kriterien zu beachten:

- **Konsistenz der Position von angedeuteten Schaltflächen bei der Adaption:** Die Auge-Hand-Koordination darf während des Zielvorgangs nicht wechselnden Positionszielen folgen müssen.
- **Adäquates Feedback über die Systemaktivität:** Die jeweiligen Systemzustände und Modi müssen in direkter und für den Nutzer nachvollziehbarer Verbindung zu seinen Handlungen stehen.
- **Vermeidung von Fehlauflösungen:** Die Adaptionenmaßnahme darf nur bei tatsächlicher Bedienintention des Nutzers ausgeführt werden. Hierfür ist eine präzise sensorische Positionserkennung der Bedienhand notwendig.

Skalierung von POIs in einer Karte zur einfacheren manuellen Interaktion Am Beispiel der manuellen Interaktion mit POI-Icons in einer

Navigationskarte auf einem Touchscreen wurde die zusätzliche Skalierung von Bedienflächen exemplarisch implementiert (siehe Bild 6.3). Bei

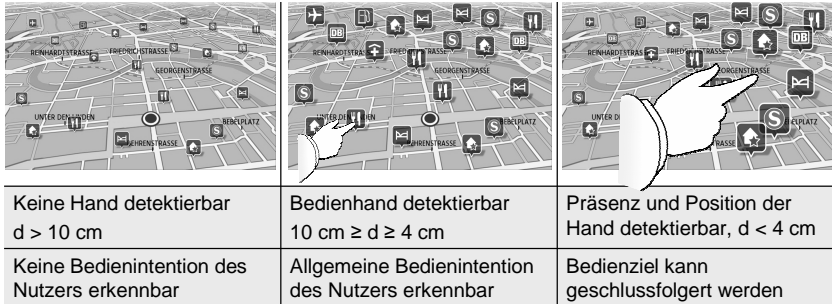


Bild 6.3: Situative Adaption von Bedienflächen.

allgemeiner Annäherung der Bedienhand wird die Größe der POI-Icons skaliert. Hierbei werden Straßennamen, Streckeninformationen und andere Details z. T. überdeckt, da sie für den Zeitraum der Bedienhandlung nicht relevant sind. Beim Anvisieren eines POIs ab einer Entfernung von ca. 4 cm können dieser POI sowie die umgebenden POIs weiter skaliert werden, um so die Treffsicherheit zu steigern. Wichtig ist, dass sich der geometrische Schwerpunkt der POIs dabei nicht maßgeblich verschiebt; es ändert sich nur ihre Größe.

Kapitel 7.2 evaluiert verschiedene Varianten der POI-Vergrößerung detailliert.

Einblenden von Bedienoptionen im Kontext Mediaplayer Die Einblendung von allgemeinen Bedienfeldern bei Annäherung sowie die Expansion der Bedienfelder zur Anzeige der einzelnen Bedienoptionen nach dem Erkennen des Bedienzies wurde am Beispiel eines Mediaplayers implementiert. Das grundlegende Bedienprinzip ist anschaulich in Bild 6.4 dargestellt. Bei Annäherung der Hand an das Display werden Bedienfelder eingeblendet (siehe Bild 6.4 rechts oben), die als Indikation für weitere Bedienoptionen dienen. Beim Anvisieren eines Indikators für die Bedienoptionen werden diese an der entsprechenden Stelle eingeblendet, ohne dass der Nutzer eine Schaltfläche betätigen muss (Bild 6.4 unten).

Kapitel 7.2 bis 7.4 evaluieren verschiedene Varianten des annäherungs-sensitiven touchscreen-basierten Mediaplayers im Detail.

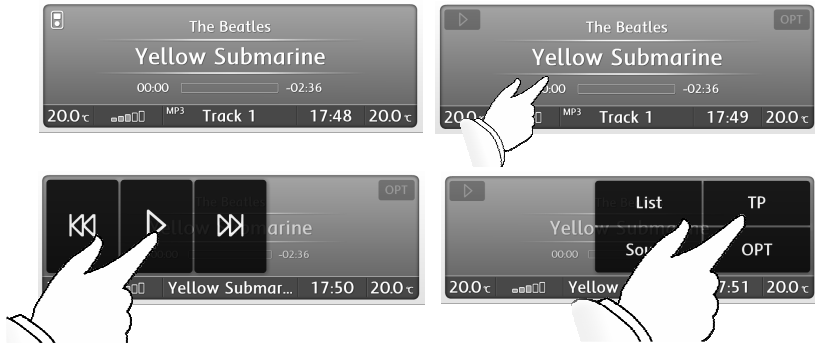


Bild 6.4: Einblendung von Indikatoren sowie Expansion von Bedienflächen.

6.2.2 Interaktion durch räumliche Spezifikationen

Point-to-Talk-Funktionen integrieren synergistisch-multimodal und alternierend-multimodal Touchscreen-Eingaben, Gestik und Sprache und nutzen dabei in idealer Weise die kommunikativen Leistungsmerkmale der unterschiedlichen Modalitäten. Von synergistischer Multimodalität sprechen wir hier (in Anlehnung an Nigay und Coutaz [106]), wenn Teilinformationen parallel über mehreren Modalitäten eingegeben werden und semantisch zu einer Gesamtinformation verbunden werden. Bei alternierender Multimodalität werden die Teilinformationen nacheinander eingegeben, trotzdem jedoch als Gesamtinformation interpretiert. Zur Spezifikation von Objekten, geometrischen Dimensionen und Positionen sowie zur Einstellung von Skalen (z. B. für die Lautstärkeregelung) eignen sich insbesondere Eingaben über den Touchscreen und manuelle Zeigegesten. Zur Eingabe von Namen, Begriffen und Befehlen ist meist Sprache geeigneter. Synergistische oder alternierende Multimodalität ermöglicht neben kürzeren Bedienzeiten durch (quasi-)parallele Informationseingaben auch eine Reduktion von Bedienfehlern, da die Eingaben über eine Bedienmodalität jeweils zur Verifizierung der Eingaben der jeweils parallel genutzten Bedienmodalität beitragen können. Zudem können Bedienele-

mente und Bedienschritte reduziert werden, indem z. B. bei entsprechender manueller Interaktion mit der Navigationskarte oder auf Annäherung der Hand an das Display der Spracherkenner automatisch (anstatt über eine *Push-to-Talk*-Taste) aktiviert wird.

In diesem System wurde dies – wie auch schon die Annäherungsfunktion – beispielhaft in zwei Hauptkontexten implementiert: Navigation und Mediaplayer.

Räumliche Spezifikation am Beispiel der Karteninteraktion im Menü Navigation Eine exemplarische Implementierung demonstriert eine ortsbezogene multimodale Suche nach POIs spezieller Themenkategorien. Hierbei spezifiziert der Nutzer die gesuchte POI-Kategorie (z. B. „Restaurants“) per Sprache und definiert gleichzeitig das spezielle Suchgebiet durch eine manuelle Zeigegeste (z. B. freies ‚Einkreisen‘ des Suchgebiets) auf dem Touchscreen. Der Spracherkenner wird automatisch aktiviert, sobald das System die Annäherung der Hand an das Display registriert. Der erkannte Suchbegriff wird auf dem Display eingeblendet und bei Abschluss der Zeigegeste werden auf der Karte innerhalb des angegebenen Suchgebiets die gefundenen POIs angezeigt (siehe Bild 6.5). Durch die Aktivierung der Spracherkennung aus dem definierten Anwendungskontext der Navigationskarte heraus wird nur das im jeweiligen Anwendungskontext sinnvolle Vokabular für die Erkennung zugrunde gelegt und damit die Fehlerrate gering gehalten.



Bild 6.5: Ortsspezifische Suche von POIs in einer Kartendarstellung.

Direkte multimodale Interaktion mit Bildschirmhalten Eine weitere Anwendung für synergistisch-multimodale Eingaben im Navigationskontext ist die direkte Interaktion mit einem POI. Anstatt den gewünschten POI per Sprache zu spezifizieren (problematisch bei mehreren POIs derselben Kategorie) oder die gewünschte Aktion (z. B. „Anrufen“) rein manuell über ein Menü auszuwählen (erfordert längere Bedienzeiten und Blickabwendungen) kann der Nutzer den POI über eine Zeigegeste auswählen und parallel dazu die gewünschte Aktion per Sprache spezifizieren (siehe auch Bild 6.6). Auch hier wird für die Spracherkennung nur das Vokabular der Begriffe zur Steuerung einer POI-Interaktion zugrunde gelegt, sodass die Wahrscheinlichkeit von Fehlerkennungen gering gehalten wird. Bei erkanntem Sprachbefehl wird die Aktion direkt ausgeführt. Nur bei nicht erkanntem oder nicht erfolgtem Sprachbefehl wird ein POI-Menü aufgerufen. Der Nutzer kann nun erneut per Sprache oder über den Touchscreen die gewünschte Option angeben.

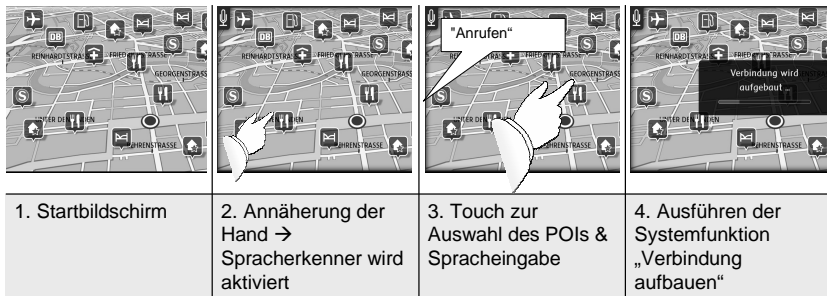


Bild 6.6: Direkte Interaktion mit POIs.

Im Kontext Mediaplayer wurde eine ähnliche Interaktionsmöglichkeit demonstriert. Dort kann der Nutzer zwischen Musik-Alben oder Interpreten wechseln, indem er per Sprache einen Befehl, z. B. „weiter“ oder „Beatles“ eingibt und den relevanten Kontext (Album oder Titel) per Touch auf dem Display spezifiziert. Die Eingabe ist dabei sowohl in der Reihenfolge Touch-Sprache als auch Sprache-Touch möglich.

Darüber hinaus wurde zu Evaluierungszwecken für alle genannten Szenarien die Möglichkeit implementiert, den Spracherkennung auf *Longpress* (400 ms) zu aktivieren. Genauereres findet sich hierzu in Kapitel 8.

6.3 Handpositionserkennung

6.3.1 Verwendete Hardware

Zur Handpositions- und Handgestikerkennung wurde aus bereits in Kapitel 4 dargelegten Gründen ein kapazitives Sensorsystem der Firma Ident Technology verwendet, das für die vorliegende Anwendung speziell weiter entwickelt wurde. Die Hardware trägt die Bezeichnung GestIC V5 und ist eine spezielle Anpassung der GestIC-Hardware für die vorliegende Applikation. Haupteinsatzbereich des ursprünglichen Systems ist die Anwesenheitsdetektion zum Einklemmschutz und zur Zugangskontrolle. Durch eine höhere Auflösung wird die grobe Unterscheidung Vorhandensein/Nichtvorhandensein einer Hand in eine Abstandserkennung überführt, die eine Positionsbestimmung ermöglicht. Der Sensor liefert dabei Werte zwischen 150 bei keinem detektierten Signal und 13000 bei unmittelbarem Kontakt der Hand mit dem Sensor. Die GestIC-Hardware arbeitet, gemäß der Differenzierung in Kapitel 6.3, im Übertragungsmodus. Der Benutzer wird durch Kontakt mit einer Sendeelektrode mit einem Signal beaufschlagt und dient als Sendeantenne für dieses Signal. Vier Empfängerelektroden registrieren die Stärke des ausgestrahlten Feldes und detektieren ein aufmoduliertes Signal. Durch eine zeitlich versetzte Beaufschlagung von Fahrer oder Beifahrer kann detektiert werden, ob Fahrer oder Beifahrer bedienen.

Die Signale der Sensoren werden auf der Microcontrollerplatine (siehe Bild 6.7 links) ausgewertet und in Pegelwerte umgewandelt. Auch die Benutzerunterscheidung und Nachpegelvorgänge werden an dieser Stelle realisiert. Die Platine wird über eine USB-Schnittstelle mit einem PC verbunden und über einen virtuellen COM-Port angesteuert.

Bild 6.7 rechts zeigt die Sendeelektrode aus leitfähigem Kunststoff, die das Signal weiterleitet, das von den vier mit der Microcontrollerplatine verkabelten Elektroden (Bild 6.7 links) empfangen wird. Die Sendelektroden wurden im vorliegenden Aufbau in Stühle bzw. in einen Autositz verbaut, wobei dort stattdessen die Sitzheizungsfolie verwendet werden konnte.

Die Empfindlichkeit der Sensoren wird durch die Amplitude des beaufschlagten Signals gesteuert. Sechs Dämpfungsstufen mit jeweils 3 dB Signalabschwächung sind möglich. Dadurch wird die Anpassung der Sensorik auf verschiedene Einbausituationen realisiert. Durch Veränderung

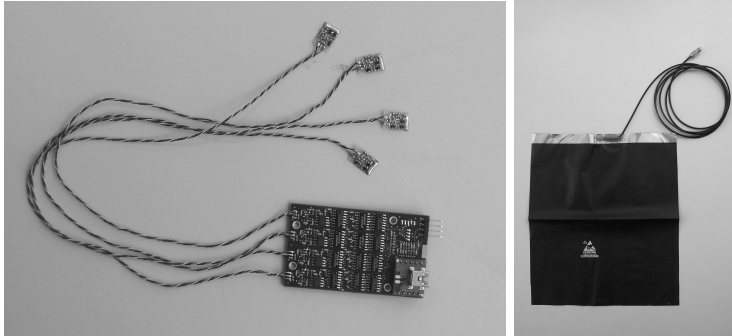


Bild 6.7: Links: Microcontrollerplatine mit vier Sensoren. Rechts: Sitzmatte zur Signalübertragung auf den Körper.

des Tastverhältnisses des PWM-modulierten Anregungssignals besteht eine zweite Möglichkeit des Eingriffs in die Signalpegel.

Bild 6.8 zeigt die genauen Positionierung der Sensoren in Bezug zum Display.

6.3.2 Personenspezifische Kennlinienanpassung

In Kapitel 4.2.4 wurde bereits dargelegt, dass physiologische Merkmale wie die Handgröße oder die Handhaltung einen Einfluss auf die Standardabweichung der Messwerte haben. Daher wurde im vorliegenden Aufbau eine benutzerspezifische Kalibrierung und Kennlinienanpassung implementiert. Bild 6.9 zeigt übertrieben die angenäherten Kennlinien für verschiedene Personen. Der Kennlinie zu entnehmen sind zwei spezielle Punkte: der Maximalwert (Messpunkt 1) des Sensorpegels und der Sensorpegel in der Displayecke (Messpunkt 2). Die Messwerte können an diesen Stellen für alle Benutzer als konstant angenommen werden. Für die benutzerspezifische Anpassung im laufenden Betrieb des Systems ist daher nur die Kenntnis eines weiteren Kennlinienpunktes nötig, der möglichst in der Mitte des Displays liegen sollte, damit eine gute Kennlinienmodellierung für alle vier Sensoren gleichzeitig erfolgen kann. Die Kennlinienanpassung wird dabei für alle vier Sensoren separat durchgeführt. Zur Verdeutlichung des Nutzens der Kennlinienanpassung sind in Bild 6.10 zwei Messergebnisse dargestellt. Bild 6.10 links zeigt das Mus-

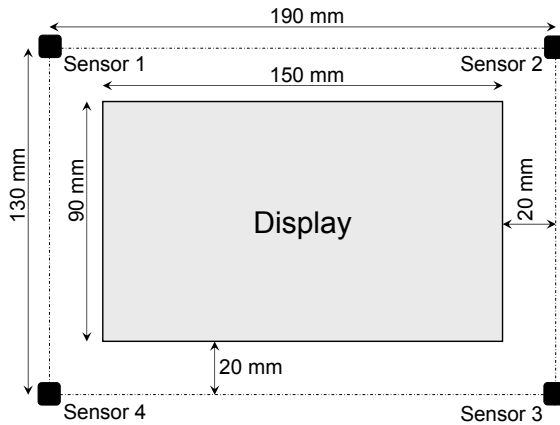


Bild 6.8: Anordnung der Sensoren um das Display mit Bemaßung.

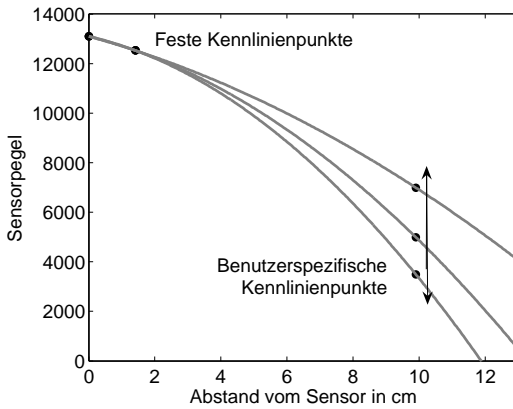


Bild 6.9: Kennlinie des Sensors und benutzerspezifische Verschiebung.

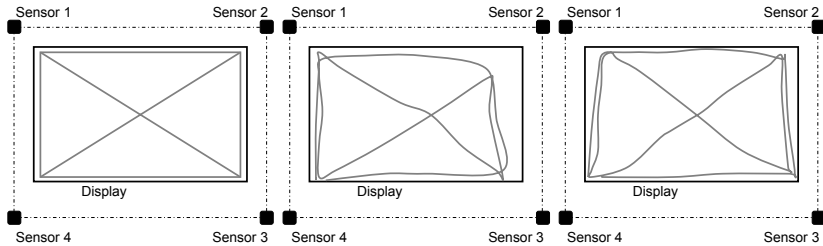


Bild 6.10: Einfluss der benutzerspezifischen Kennlinienanpassung: Testmuster (links) und rekonstruiertes Muster vor Kalibrierung (Mitte) und nach Kalibrierung (rechts).

ter, das von einer Testperson in einem Abstand von ca. 2 cm vor dem Display mit dem Finger nachgefahren werden sollte. Bild 6.10 Mitte zeigt das rekonstruierte Muster ohne Kennlinienanpassung. Es wird deutlich, dass die Abweichung der rekonstruierten Werte von den wahren Werten in der rechten oberen Ecke am größten ist. Dies ist in diesem Fall darauf zurückzuführen, dass der Nutzer als Linkshänder das Muster mit der linken Hand nachgefahren war. Der Sensor links unten wird daher besonders stark vom Handkörper beeinflusst und verschiebt das Ergebnis der Maximum-Likelihood-Schätzung in Richtung links unten. Bild 6.10 rechts zeigt das rekonstruierte Muster nach der Kennlinienanpassung. Die Verschiebung ist nun weitestgehend verschwunden.

In der Kategorisierung von Kass und Finin [79] (siehe auch Kapitel [79]) handelt es sich dabei um ein individuelles, dynamisches, beschreibendes und implizites Benutzermodell.

7 Evaluierung der Annäherungsfunktion

„Nicht alles was zählt, kann gezählt werden und
nicht alles, was gezählt werden kann, zählt.“
(Albert Einstein)

7.1 Positionsschätzung der Hand vor einem Display

In Kapitel 4.3 wurde der Algorithmus vorgestellt, mit dem die Positionsschätzung der Hand vor dem adaptiven Touchscreen erfolgt. Kapitel 6.3 beschreibt den Aufbau im Detail. Um eine Aussage über die Robustheit des Algorithmus und die Genauigkeit der Schätzung treffen zu können, wurde das System mit mehreren Nutzern getestet.

7.1.1 Konzeption

Bei der Konzeptionierung des Messaufbaus waren folgende Punkte zu adressieren:

- Der Nutzer sollte auch in der Testsituation möglichst natürlich mit dem System interagieren, d. h. so, wie er es außerhalb einer Testsituation auch tun würde. Es musste vermieden werden, dass der Nutzer seine grundsätzlich bevorzugte Handhaltung zu Gunsten einer besseren Funktionalität des Systems veränderte.
- Die Bestimmung der Genauigkeit der Maximum-Likelihood-Schätzung und daraus abgeleitet die minimale noch treffbare Größe von Schaltflächen auf dem Touchscreen sollte in mehreren definierten Ebenen vor dem Display erfolgen.
- Ermüdungserscheinungen der Testpersonen sollten weitestgehend vermieden werden. Die Dauer der Messung sollte also möglichst gering sein, und es sollte über die gesamte Messung einen gewissen Grad an Herausforderung für den Nutzer geben.

Mit diesen Rahmenbedingungen wurde ein Messaufbau entwickelt, der einen systemimmanenten Spielcharakter hat: Auf dem Display werden in einem festgelegten Raster zunächst gleich große Kreise angezeigt. Die Probanden werden aufgefordert, auf den jeweils per Zufallsgenerator ‚aktivierten‘ Kreis (visuell hervorgehoben) zu zeigen (siehe Bild 7.2 a). Während des gesamten Messverlaufs werden alle vom System ermittelten Koordinaten der Nutzerhand aufgezeichnet und mit einem Zeitstempel versehen. Während der Messung werden darüber hinaus zu verschiedenen Ereignissen Timer gestartet. Der erste Timer T_1 wird bei Aktivierung eines Kreises durch den Zufallsgenerator gestartet, der zweite Timer T_2 startet, sobald erstmals eine Koordinate aufgezeichnet wird, die innerhalb eines aktivierten Kreises liegt. Überschreitet der Timer T_1 ein vorher festgelegtes Maximum, ohne dass der Timer T_2 gestartet wird, wird der aktivierte Kreis als ‚nicht getroffen‘ markiert (visuelle Darstellung durch ein Kreuz im Kreis) und der Timer T_1 wird zurückgesetzt. Daraufhin wird vom Zufallsgenerator ein anderer Kreis aktiviert, visuell hervorgehoben und der Timer T_1 erneut gestartet. Wird jedoch der Timer T_2 rechtzeitig (also vor Überschreiten des Maximums von T_1) gestartet, werden alle – während der Zeit t_2 , in der der Timer T_2 läuft – ermittelten Koordinaten $\mathbf{x}_i = (x_i, y_i)$ in einen Speicher geschrieben und mit den Koordinaten $\mathcal{K}_{\text{aktiv}}$ der aktivierten Kreisfläche verglichen. Die Menge \mathcal{T} der ermittelten Koordinaten, die innerhalb von $\mathcal{K}_{\text{aktiv}}$ liegen, ergibt sich dabei folgendermaßen:

$$\mathcal{T} = \{\mathbf{x}_i | \mathbf{x}_i \in \mathcal{K}_{\text{aktiv}}\}. \quad (7.1)$$

Im Anschluss daran wird die Trefferrate h berechnet, ein Maß dafür, wie lange der Nutzer während des Timerlauf T_2 den Kreis getroffen hat:

$$h = \frac{|\mathcal{T}|}{|\mathbf{x}|}. \quad (7.2)$$

Die Trefferrate muss über einem bestimmten Schwellwert h_{min} liegen, damit der Kreis als getroffen zählt. Liegt die Trefferrate unterhalb des Schwellwerts h_{min} oder ist gleich dem Schwellwert, wird der Kreis als ‚nicht getroffen‘ gewertet. Aufgrund des natürlichen Zitterns der Hand ist es notwendig, eine solche Trefferrate zur Bewertung einzuführen, die mathematisch gesehen eine Filterung darstellt. Wird ein Kreis zweimal nacheinander als ‚nicht getroffen‘ markiert, wird er ‚gesperrt‘ (visuell dargestellt durch zwei Kreuze) und kann vom Zufallsgenerator nicht mehr

ausgewählt werden. Wurde ein Kreis dagegen als getroffen identifiziert, verkleinert sich sein Radius und es verschwinden evtl. vorhandene Markierungen und damit verbundene Kreuze. Der Schwerpunkt des angezeigten Kreises bleibt bei der Verkleinerung erhalten. Bild 7.1 zeigt in einem

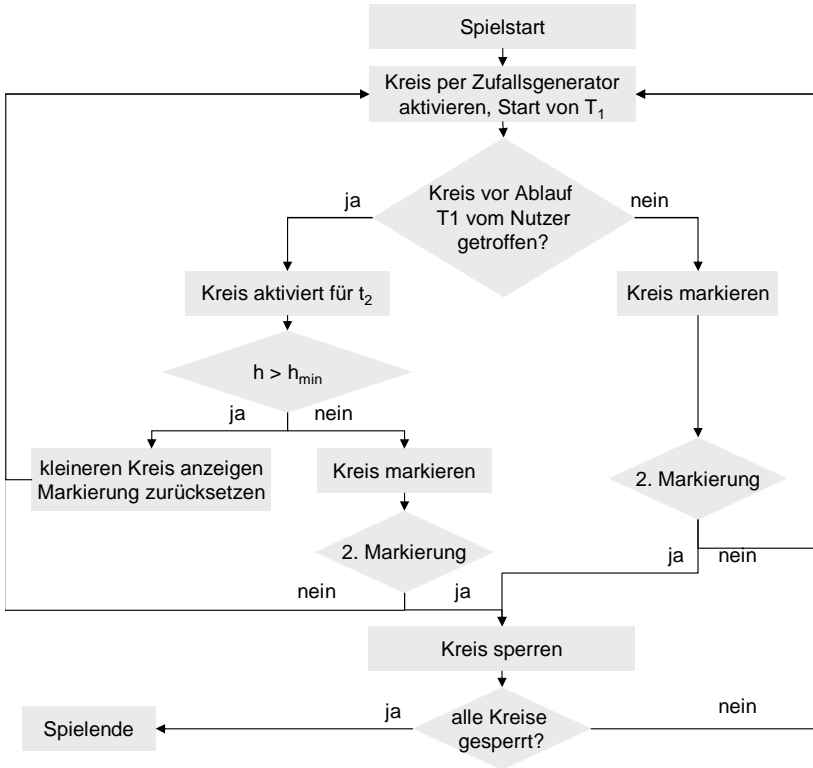


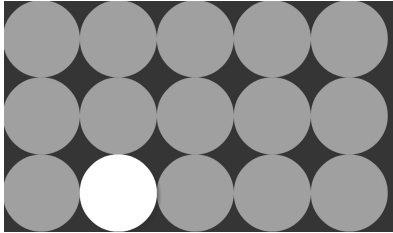
Bild 7.1: Ablauf der Messung im Flussdiagramm.

Flussdiagramm den Ablauf der Messung im Detail. Wichtig sind dabei zwei Restriktionen:

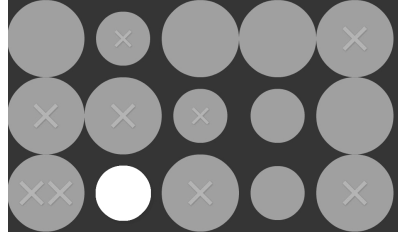
- Ein nicht getroffener Kreis kann ein weiteres Mal aktiviert werden, nach dem zweiten Fehlversuch wird er für die weitere Aktivierung gesperrt.

- Wird ein nicht getroffener Kreis bei der nächsten Aktivierung getroffen, verschwindet die Markierung und er wird auf den Ausgangszustand zurückgesetzt.

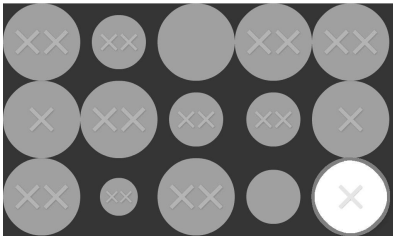
Bild 7.2 zeigt den Displayinhalt zu verschiedenen Zeitpunkten der Messung. Die Messung endet, wenn alle Kreise gesperrt sind. Die Radien der



(a) Anfangsbildschirm mit einem aktivierten Kreis.



(b) Bildschirm während des Spiels.



(c) Timerstart T_2 ist erfolgt.



(d) Bildschirm zum Ende der Messung.

Bild 7.2: Displayinhalte zu verschiedenen Zeitpunkten der Messung.

einzelnen Kreise werden als gerade nicht mehr treffbare Radien für eine bestimmte Person gespeichert. Die minimal treffbaren Radien sind die jeweils nächst größeren Radien, also die, bei denen der Kreis zum letzten Mal getroffen wurde.

7.1.2 Parametrierung und Durchführung der Messung

Das Spiel wurde in fünf Ebenen (0 cm, 2 cm, 4 cm, 6 cm und 8 cm) vor dem Display durchgeführt, wobei ein konstanter Abstand durch Anbringen einer Plexiglasplatte vor dem Display sichergestellt wurde. Für die

einzelnen Ebenen wurden in Vorstudien minimale Größen der Kreise bestimmt, sodass jeder Kreis in seiner Anfangsdarstellung in jedem Fall mindestens einmal von jedem Proband getroffen wurde.

Aufgrund der großen Varianz der Kreisgröße am Ende der Messung wurde die Anzahl der Kreise auf dem Display auf 15 begrenzt. So konnte eine nicht zu lange Dauer der Messung sichergestellt werden.

Tabelle 7.1 zeigt die Startwerte für die Kreise (Radien in mm) in der Reihenfolge von links oben nach rechts unten. Der Parameter für T_1

Tabelle 7.1: Startwerte: Radien der Kreise in den Ebenen 0 cm, 2 cm, 4 cm, 6 cm und 8 cm zu Beginn der Messung in mm.

Ebene 0 cm, 2 cm	15	15	15	15	15
	15	15	15	15	15
	15	15	15	15	15
Ebene 4 cm	30	20	20	20	30
	30	20	20	20	30
	30	20	20	20	30
Ebene 6 cm, 8 cm	30	30	30	30	30
	30	30	30	30	30
	30	30	30	30	30

wurde auf 500 ms festgelegt. Bei erwarteten Ereignissen liegt die Reaktionszeit des Menschen bei 100 ms bis 400 ms. Der Timer T_2 wurde mit einer Laufzeit von ebenfalls 500 ms eingestellt. Die Trefferrate wurde auf 0,9 festgelegt. Der Faktor, um den sich der Radius nach erfolgreichem Abläufen von T_2 verringerte, wurde auf 0,7 eingestellt. Damit ergaben sich die folgenden möglichen Kreisradien: 30 mm, 20 mm, 15 mm, 11 mm, 7 mm, 5 mm, 3 mm, 2 mm und 1 mm.

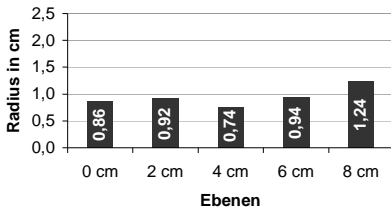
Die Messung wurde mit insgesamt 10 Personen durchgeführt. Darunter waren drei Frauen und sieben Männer. Es wurde darauf geachtet, dass gleichermaßen Links- und Rechtshänder in der Testgruppe vorhanden waren und die Handgrößen der Versuchspersonen möglichst stark variierten.

Nach einer kurzen Übungsphase und Kalibrierung des adaptiven Touchscreens wurden die Versuchspersonen aufgefordert, mit der Mes-

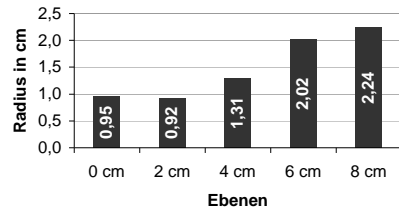
sung (,dem Spiel‘) zu beginnen. Es wurde zwar darauf hingewiesen, dass die Kreise möglichst klein werden sollten, ein Abweichen von der natürlichen Handhaltung würde aber Punktabzug bedeuten.

7.1.3 Ergebnisse

Zur besseren Darstellung der Ergebnisse wurde das Display in drei Regionen aufgeteilt. Die drei Kreise auf der linken Displayseite sowie die drei Kreise auf der rechten Displayseite wurden zu den ‚Randbereichen‘ zusammengefasst, die restlichen sechs Kreise in der Mitte des Displays wurden unter dem Bereich ‚Mitte‘. Es zeigte sich in der Messung, dass der minimal treffbare Radius in der Mitte deutlich niedriger liegt als in den Randbereichen. Bild 7.3 zeigt die gemittelten Werte über alle Messungen.



(a) In der Mitte des Displays.



(b) In den Randbereichen.

Bild 7.3: Minimaler durchschnittlich noch getroffener Kreisradius in cm.

Auffällig ist, dass in den Randbereichen in den Ebenen 0 cm und 2 cm der minimal treffbare Radius nahezu konstant ist, in den folgenden Ebenen allerdings stark ansteigt. Im mittleren Bereich hingegen ist der minimal treffbare Radius bis zu einem Abstand von 6 cm nahezu konstant und steigt erst in der Ebene 8 cm an. Dabei bleibt er allerdings deutlich unter dem Wert aus den Randbereichen.

Zu einem gewissen Teil ist das sicherlich auf den Verzerrungseffekt zurückzuführen, wie er auch in Bild 6.10 in Kapitel 6.3 sichtbar ist. Bild 7.4 zeigt die Messung für eine Person. Dabei wird deutlich, dass bei einer Bedienung mit der rechten Hand im linken Randbereich die minimal treffbaren Radien stärker ansteigen (Verzerrungseffekt) als im rechten

Randbereich. Aber auch im rechten Randbereich steigen die Radien wesentlich stärker an als in der Mitte. Dies ist auf die begrenzte Reichweite der Sensoren zurückzuführen. Bei Messungen im Randbereich deckt die Reichweite der Sensoren der gegenüberliegenden Seite den zu vermessenden Bereich u. U. nicht mehr ab. Daher kommt es zu Verzerrungen, da die Kennlinie nicht bis ins Unendliche mit einer Funktion zweiten Grades anzunähern ist, sondern in Sättigung geht. Ab einem gewissen Abstand der Hand vom Sensor kann also nicht mehr auf einen definierten Abstand rückgeschlossen werden, sondern nur noch festgestellt werden, dass der Messbereich überschritten ist. In diesem Fall wird jedoch trotzdem die maximale Reichweite des Sensors als Abstand der Hand vom Display angenommen. Grundsätzlich führte dies zu robusteren Messungen als die Vernachlässigung eines Sensors in diesem Fall. Ein größerer minimal erreichbarer Radius wird zu Gunsten einer höheren Robustheit also in Kauf genommen.

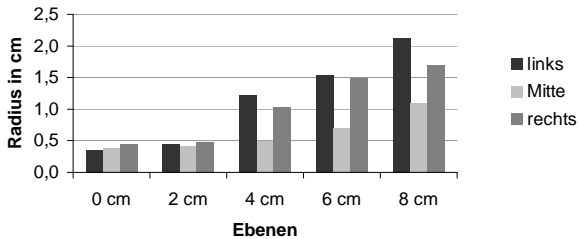


Bild 7.4: Minimaler durchschnittlich noch getroffener Kreisradius in cm für eine Person im linken Randbereich, der Mitte und dem rechten Randbereich.

7.2 Studie zur subjektiven Bewertung der Annäherungsfunktion

7.2.1 Konzeption

Testsystem In der vorliegenden Studie wurde aus subjektiver Sicht die Bedienbarkeit und Akzeptanz des adaptiven Touchinterfaces auf Basis verschiedener Gestaltungsvarianten evaluiert. Dabei wurden im ersten Teil der Studie drei adaptive Varianten der Darstellung einer statischen

Variante gegenübergestellt (siehe auch Bild 7.5). Als Kontext wurde der Mediaplayer gewählt, da er von den verfügbaren Schaltflächen den meisten Nutzern geläufig ist. Bei der Gestaltung der statischen Bedienvariante wurde die verfügbare Anzeigefläche zu je 50 % für die Informationsanzeige und zur Darstellung von Bedienflächen genutzt. Die adaptiven Bedienvarianten unterscheiden sich in ihren Ausprägungen hinsichtlich der eingesetzten ‚Indikatoren‘ der Bedienflächen im Anzeigemodus sowie dem Layout der Bedienflächen. ‚Indikatoren‘ werden zur Unterstützung der Bewegungsplanung sowie der Hand-Auge-Koordination vor und während der Bedienbewegung eingesetzt, um bereits im Anzeigemodus Informationen über Bedienflächen vermitteln zu können. Grundsätzlich lassen sich dabei semantische Indikation (Bedeutung und Funktion der Bedienflächen) und geometrische Indikation (finale Position und/oder Dimension der Bedienflächen) unterscheiden.

- **Bedienvariante 1: statisch, keine Indikation**

Die Anzeigefläche wird je zur Hälfte für die Informationsanzeige und die Bedienflächen genutzt. Aufgrund der fehlenden Adaption auf Annäherung ist der Informationsgehalt geringer (kein Fortschrittsbalken beim Abspielen des Titels). Die Schaltflächen sind wesentlich kleiner als in den adaptiven Varianten. Darüber hinaus ist der Abstand zwischen den einzelnen Schaltflächen geringer. Die Anforderungen in Bezug auf Schaltflächengröße und Textgröße aus DIN EN ISO 9241 [48] und DIN EN ISO 15008 [47] können hier nicht mehr erfüllt werden.

- **Bedienvariante 2: adaptiv, keine Indikation**

Die Anzeigefläche wird bei Annäherung fast vollständig zur Anzeige von Bedienflächen genutzt, nur der Titel des aktuellen Songs wird weiterhin in der Statuszeile angezeigt. Die Schaltflächen sind vertikal orientiert und wesentlich größer, was u. a. die Bewegungsgeschwindigkeit erhöhen und die Fehlerrate verringern könnte. Als nachteilig könnte sich jedoch erweisen, dass die Bedienhandlung in zwei Phasen geteilt wird, da die Bedienoptionen im Anzeigemodus nicht sichtbar sind. Eine spezielle Bedienhandlung ist also ohne gelerntes Wissen, wo sich welche Schaltfläche befindet, nicht planbar. Darüber hinaus verschwindet die Anzeige der Bedienflächen bei Unterbrechung der Eingabe vollständig.

- **Bedienvariante 3: adaptiv, geometrische und semantische Indikation**

Die Anzeigefläche wird bei Annäherung zum großen Teil für Bedienflächen genutzt, allerdings bleibt der Songtitel in der Mitte sichtbar. Im Anzeigemodus sind geometrische und semantische Indikatoren für die Bedienung am oberen und unteren Rand platziert, die Anzeigefläche wird also nur minimal belegt bei dennoch großen Bedienflächen. Nachteilig ist hier, dass ein Teil der Anzeigefläche für Indikatoren verwendet werden muss und die Beschriftung dieser nur sehr klein ist.

- **Bedienvariante 4: adaptiv, allgemeine geometrische und semantische Indikation**

Die Bedienflächen sind auf zwei funktionale Blöcke aufgeteilt, die durch Annäherung an den jeweiligen Indikator expandiert werden. Der mittlere Displaybereich wird durch diese Teilung doppelt genutzt, sodass eine größere Tastendarstellung für die einzelnen Blöcke möglich ist.

Für die Darstellung der Bedienvarianten wurde ein prototypischer Aufbau eines adaptiven Touchscreens (7" resistiver Touchscreen, 16:9, genutzte Anzeigefläche 12 cm × 4 cm) in einer abmessungsgetreuen, abstrakten Cockpit-Umgebung eingebaut. Die Darstellung der Interaktionsinhalte erfolgte jedoch nur auf einem verringerten Ausschnitt des 7"-Displays.

Im zweiten Teil der Studie wurden in einer Kartendarstellung verschiedene Varianten der Skalierung von POI-Icons bei Annäherung an das Display einer nicht adaptiven Variante gegenübergestellt (siehe Bild 7.6):

- **Variante 1: keine Skalierung**

Die POIs wurden nicht skaliert.

- **Variante 2: pauschale Skalierung**

Alle POIs wurden bei Annäherung an die Karte gleichmäßig skaliert.

- **Variante 3: positionsabhängige Skalierung**

Die POIs wurden abhängig von der Position des Bedienfingers skaliert.

Die Darstellung erfolgte in diesem Teil der Studie auf dem gesamten 7"-Display.

7 Evaluierung der Annäherungsfunktion



(a) Bedienvariante 1: statisch, keine Indikation.

(b) Bedienvariante 2: adaptiv, keine Indikation.



(c) Bedienvariante 3: adaptiv, geometrische und semantische Indikation.

(d) Bedienvariante 4: adaptiv, allgemeine geometrische und semantische Indikation.

Bild 7.5: Übersicht über die Bedienvarianten – die hintere Darstellung zeigt jeweils den Anzeigemodus, die vordere Darstellung zeigt die Einblendung der Bedienflächen bei Annäherung der Hand.



Bild 7.6: Varianten der POI-Skalierung: keine Skalierung, pauschale Skalierung und positionsabhängige Skalierung.

Testumgebung Die Studie wurde im Fahrsimulator durchgeführt. Die Probanden wurden aufgefordert, einem vorausfahrenden Fahrzeug in einem definierten Abstand zu folgen.

Aufgabenbeschreibung Die Versuchsteilnehmer wurden zunächst gebeten, sich getrennt sowohl mit adaptiven Touchinterface als auch dem Fahrsimulator vertraut zu machen. Im Anschluss wurden die ersten Fragen gestellt. Während des folgenden Versuchsteils mussten die Teilnehmer im Simulator einem Fahrzeug folgen und dabei einer ständigen Steuertätigkeit nachkommen. Währenddessen wurden die Probanden instruiert, parallel zur Fahraufgabe für jede der drei Bedienvarianten definierte Tasten (,Play‘ und ,TP‘) in separaten Bedienhandlungen so zügig wie möglich zu betätigen bzw. mit bestimmten POIs zu interagieren. Im Anschluss wurde wiederum eine Befragung durchgeführt.

Messgrößen Die abgefragten Kriterien wurden auf einer Ordinalskala von „sehr schlecht“ (--) , „schlecht“ (-), „neutral“ (0), „gut“ (+) bis „sehr gut“ (++) bewertet.

Stichprobenbeschreibung Insgesamt nahmen 29 Personen an der Untersuchung teil, davon 16 Männer (55 %) und 13 Frauen (45 %). 38 % der Versuchsteilnehmer haben in ihrem eigenen Fahrzeug ein Navigationssystem, 19 % wünschen sich immer die neueste Technik. Tabelle 7.2 zeigt die Altersverteilung.

Tabelle 7.2: Altersverteilung der Versuchsteilnehmer.

Alter in Jahren	25 – 30	31 – 40	41 – 50	51 – 60	> 61
Anzahl der Probanden	2	11	8	6	2

7.2.2 Ergebnisse

Die Annäherungsfunktion wird insgesamt sehr positiv beurteilt (siehe Bild 7.7). Es zeigt sich jedoch, dass Frauen die Annäherungsfunktion besser bewerten als Männer. Die Größe von Touchscreentasten ist ein

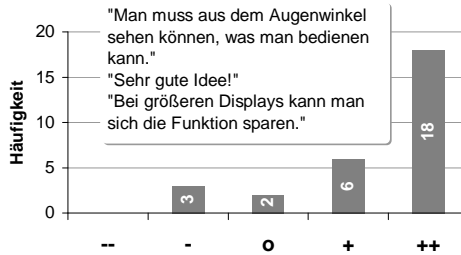


Bild 7.7: Wie gefällt Ihnen die Annäherungsfunktion allgemein?

subjektiv sehr wichtiges Kriterium, 11 Personen beurteilen sie als „wichtig“, 17 Personen als „sehr wichtig“. Bei der Frage nach der Sichtbarkeit der Bedienoptionen vor der Bedienung (Indikatoren) zeigt sich ein eher gemischtes Bild (siehe Bild 7.8). Insbesondere Männern ist dies jedoch „sehr wichtig“. Bild 7.9 zeigt in Box-Plots [25] die Bewertungen der ein-

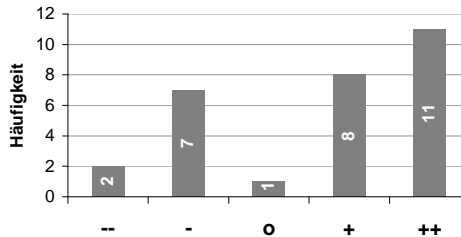


Bild 7.8: Wie wichtig ist es für Sie, die Bedienoptionen schon vor der Bedienung sehen zu können?

zelen Varianten. Im Hinblick auf die Frage, bei welcher Variante die subjektiv schnellste Bedienung möglich ist, zeigt sich, dass nur 10 Personen eine der adaptiven Varianten bevorzugten; 19 Versuchsteilnehmer waren hingegen der Meinung, die Bedienung wäre in der statischen Variante schneller möglich. Insgesamt würden jedoch mehr Versuchsteil-

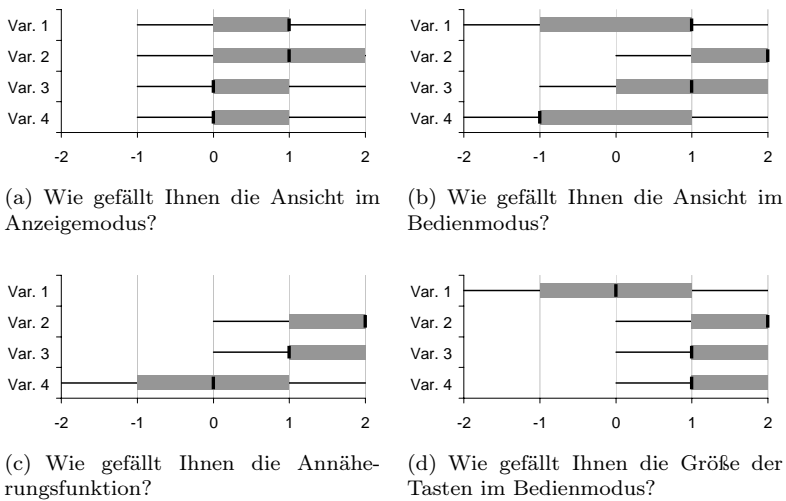


Bild 7.9: Bewertung der einzelnen Varianten.

nehmer eine der adaptiven Varianten bevorzugen. Die Unterschiede im Antwortverhalten vor und nach der Simulatorfahrt (siehe Bild 7.10) sind zwar nicht signifikant (χ^2 -Test: $p = 0,251$; siehe auch [25]), in der Tendenz gewinnen jedoch dynamische Varianten an Bedeutung. Zur Frage der POI-Skalierung äußerten sich die Versuchsteilnehmer überwiegend positiv (83% bewerteten mit „sehr gut“). Bei Bedienung ohne Skalierung kamen einige Versuchspersonen von der Straße ab. Bild 7.11 zeigt die Auswertung im Detail.

7.2.3 Interpretation und Fazit

Die Annäherungsfunktion wird insgesamt positiv bewertet, 72% der Versuchsteilnehmer bevorzugen im direkten Vergleich eine adaptive Variante. Dabei schneidet jedoch Variante 4 am schlechtesten ab und wird daher in den folgenden Studien nicht weiter betrachtet. Bezüglich des Einsatzes von Indikatoren konnte kein klares Bild gewonnen werden. Darüber hinaus führte die statische Variante zu den subjektiv geringsten Bedienzeiten. Daher werden diese Aspekte im Folgenden detailliert und objektiv

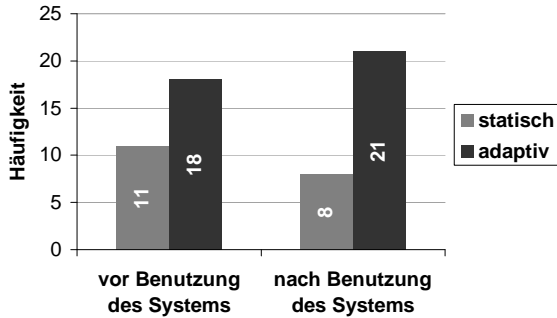
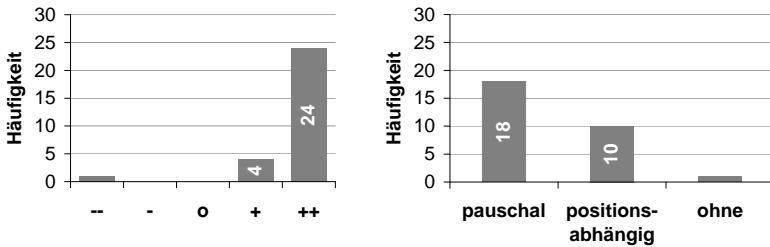


Bild 7.10: Welche der Varianten würden Sie insgesamt bevorzugen?



(a) Wie gefällt Ihnen die Expansion der POIs bei Annäherung?

(b) Welche Skalierungsvariante würden Sie bevorzugen?

Bild 7.11: Bewertung der POI-Skalierung.

untersucht. Auch liegt der Verdacht nahe, dass die Sensorik zu Verzögerungen führt, die eine durchgängige Bedienbewegung verhindern und so zu einer subjektiv und objektiv längeren Bedienzeit in den adaptiven Varianten führen. Die Vergrößerung von POIs, bei denen die zeitliche Synchronisierung nicht die Wichtigkeit erlangt, wie bei der Einblendung von Bedienflächen, wird durchgehend positiv beurteilt, wobei auch hier eine pauschale Skalierung bevorzugt wird.

7.3 Expertenstudie zur Evaluierung der Bedienbarkeit und zur optimalen Spezifikation der Annäherungssensorik

In Kapitel 7.2 wurde eine subjektive Bewertung der Annäherungsfunktion aufgrund von Probandenbefragungen durchgeführt. Für eine umfassende Bewertung, v. a. für den Einsatz im Automobil, ist dies jedoch nicht ausreichend. Daher wird in der folgenden Studie objektiv durch Bewegungs- und Zeitmessungen die Bedienbarkeit evaluiert und mit Erkenntnissen aus der Literatur verglichen. Dabei soll speziell der Einfluss der Sensorik überprüft werden, da die Probanden aus der vorherigen Studie eine subjektiv längere Bedienzeit in den adaptiven Varianten zu Protokoll gaben.

7.3.1 Ergonomische Betrachtungen

Die Interaktion mit Touchscreens kann im Fahrzeug während der Fahrt nur als Sekundäraufgabe neben der Fahrzeugführung ausgeführt werden. Daher sind eine schnelle Ablesbarkeit der dargestellten Informationen und eine effiziente und fehlerfreie Bedienbarkeit besonders wichtig.

Leistungsmerkmale einer Eingabe-Schnittstelle sind u. a. die nötige Bewegungszeit (MT – Movement Time) zum Aktivieren einer Schaltfläche und die Häufigkeit von Fehlertreffern (ER – Error Rate) bei mehrfacher Nutzung. Einen wesentlichen Einfluss hierauf hat der Durchmesser der dargestellten Schaltflächen (siehe auch [16, 133]), der auch für die Größe der Textbeschriftung maßgeblich ist. Die DIN EN ISO 9241-9 [48] empfiehlt unabhängig von spezifischen Bediensituationen für alle Tasten pauschal einen Diagonaldurchmesser von mindestens 2 cm. Stevens [141] empfiehlt bei einem durchschnittlichen Leseabstand von 700 mm

zwischen dem Auge des Fahrers und dem Display in der Mittelkonsole sowie einem mittleren Blickwinkel von ca. $0,33^\circ$ für Höhe eines Schriftzeichens mindestens 4 mm. Das mathematische Modell von Fitts, bekannt als *Fitts' Law* [53], sowie in erweiterter Form u.a. die Modelle von MacKenzie (siehe Gleichung 7.3 und Buxton [88, 89] beschreiben hingegen differenzierter den Einfluss des Durchmessers einer Schaltfläche und der Amplitude (Länge) der Bedienbewegung auf die resultierende Bewegungszeit. *Fitts' Law* ist eines der wenigen überprüfbareren Modelle im Schnittstellen-Design und wurde seit seiner Veröffentlichung 1954 in diversen Studien verifiziert [1, 44, 88, 98]:

$$MT = a + b \log_2 \left(\frac{A}{W} + 1 \right). \quad (7.3)$$

MT entspricht der mittleren Bewegungsdauer, A der Amplitude der Bedienbewegung (Distanz zwischen dem Ausgangspunkt der Bewegung und dem Mittelpunkt des Zielobjekts) und W dem Durchmesser des Zielobjekts in Richtung der Bedienbewegung. Die empirischen Konstanten a und b werden anhand von Messwerten durch lineare Regression ermittelt. Der Logarithmus $ID = \log_2 (A/W + 1)$ beschreibt den sog. ‚*Index of Difficulty*‘ der jeweiligen Bedienaufgabe. Da der Abstand zwischen Lenkrad und Display im Fahrzeug durch feste Einbaupositionen definiert ist, kann eine Reduktion von Bewegungszeiten vor allem durch eine Vergrößerung von W erreicht werden. Dem sich daraus ergebenden Konflikt zwischen der Forderung nach möglichst großen Bedienflächen und zugleich einer schnell erfassbaren und entsprechend großen Informationsdarstellung wirkt der automatische Wechsel zwischen Anzeige- und Bedienmodus entgegen. Zur Unterstützung der Bewegungsplanung sowie der Hand-Auge-Koordination vor und während der Bedienbewegung können, wie in Kapitel 7.2 bereits eingeführt, Indikatoren eingesetzt werden.

Nach Meyer [98, 99] bestehen zielgerichtete Bedienbewegungen immer aus einer, zwei oder n Teilbewegungen und lassen sich mit dem OIIM (*Optimized Initial Impulse Model*) beschreiben (siehe auch Bild 7.12). Danach beginnt eine gerichtete Bewegung immer mit einer ballistischen Teilbewegung, deren Verlauf bereits vor Bewegungsbeginn geplant wird und mit der ca. 90% der Entfernung zum Ziel zurückgelegt werden. Visuell wahrnehmbare Änderungen der Zielfläche können während dieser Bewegung die motorische Ausführung der Bewegungsplanung nicht beeinflussen. Bewegungsgeschwindigkeit und Genauigkeit verhalten sich

während dieser Phase indirekt proportional zueinander. Die verbleibende Abweichung zwischen Ziel und aktueller Position wird vom Menschen sensorisch erfasst und ist ausschlaggebend dafür, ob weitere Teilbewegungen zur Korrektur notwendig sind. Diese korrektiven Bewegungen finden unter geschlossenem visuellem Feedback statt [98, 130]. Bei effizienter Balance von Bewegungsgeschwindigkeit und Genauigkeit sind für die meisten Zielbewegungen zwei Teilbewegungen notwendig. Im Folgenden werden die nach der ballistischen Bewegungsphase einsetzenden korrektiven Bewegungen zusammenfassend als ‚Selektionsphase‘ bezeichnet.

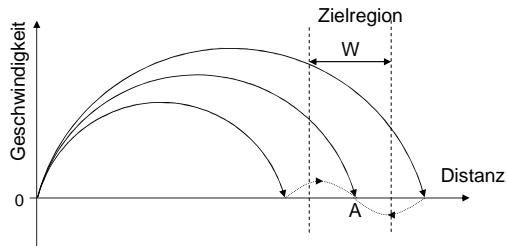


Bild 7.12: Schematische Darstellung von Sequenzen von Teilbewegungen zum Treffen eines Ziels nach dem OIIM [98].

In diesem Zusammenhang untersuchten McGuffin und Balakrishnan [97] den Effekt von ‚expandierenden Schaltflächen‘ auf die Bewegungsdauer bei Aktivierung von Schaltflächen mit einem mausgesteuerten Cursor. Die Ergebnisse zeigen, dass die Vergrößerung des Durchmessers von Schaltflächen zu einer deutlichen Reduktion der Bewegungsdauer führt, selbst wenn diese erst bei 90 % der Strecke zum Ziel einsetzt. Der finale Durchmesser des Zielobjekts hat also einen wesentlich höheren Einfluss auf MT als der Durchmesser zu Beginn der Zielbewegung.

Für die Adaption von Touchscreen-Bedienflächen lässt sich hieraus schließen, dass eine Vergrößerung des Zielobjekts entsprechend *Fitts' Law* zu einer Reduktion der Dauer der Bedienbewegung führt, auch wenn diese erst während der Zielbewegung einsetzt. Da visuelles Feedback keinen Einfluss auf den Verlauf der ballistischen Bewegungsphase hat [98], ist anzunehmen, dass die Bedienflächen erst zu Beginn der Selektionsphase in ihrer finalen Form dargestellt werden müssen.

Dabei ist jedoch zu bedenken, dass in den Arbeiten von McGuffin und Balakrishnan [97] lediglich die Dimension des Zielobjekts, nicht jedoch die Information über seine Platzierung oder Funktion verändert wurde. Fraglich ist daher für die Bedienbarkeit von adaptiven Schaltflächen auf einem Touchscreen, inwieweit die Qualität der ballistischen Initialbewegung sowie die Dauer weiterer Korrekturbewegungen von der zur Bewegungsplanung verfügbaren Information über das Ziel abhängen.

7.3.2 Fragestellung und Zielsetzung der Studie

Im Rahmen dieser Studie soll der Einfluss adaptiver Schaltflächen bei der Bedienung eines Touchscreens im Fahrzeug auf das Bewegungsverhalten des Nutzers untersucht und beschrieben werden. Von besonderem Interesse sind hierbei der Effekt einer abstrakten Indikation der Schaltflächen bereits im Anzeigemodus sowie der Vergleich des adaptiven mit einem nicht-adaptiven System.

Des Weiteren soll überprüft werden, ob adaptive Schaltflächen im Vergleich zu einer nicht-adaptiven Bedienvariante aufgrund der größeren Tasten zu reduzierten Fehltreffern bei ähnlichen Bewegungszeiten führen.

Die Ergebnisse sollen sowohl zur Bewertung entsprechender Bedienkonzepte als auch zur Überprüfung der Eignung des vorhandenen Sensorsystems sowie zur Ableitung von technischen Spezifikationen für ein optimales Sensorsystem zur Annäherungsdetektion genutzt werden.

Hierzu wurden folgende Hypothesen aufgestellt:

- **H1.** Die Bedienbewegungen zur Bedienung eines Touchscreens im Fahrzeug lassen sich sowohl für adaptive als auch für nicht-adaptive Bedienvarianten mit dem OIIM beschreiben.
- **H2.** Eine Adaption der Zielflächen während der Bedienbewegung führt im Vergleich zu einer statischen Bedienvariante zu geringeren Fehlerraten bei vergleichbaren Bewegungszeiten.
- **H3.** Der Einsatz von Indikatoren führt bei der Bedienung von adaptiven Zielen zu kürzeren Bewegungszeiten.
- **H4.** Das bisher verwendete Sensorsystem stört einen optimalen Bewegungsablauf und führt daher zu längeren Bewegungszeiten.

- **H5.** Aus den ermittelten Bewegungsabläufen lassen sich technische Spezifikationen für ein Sensorsystem zur Annäherungsdetektion ableiten.

7.3.3 Konzeption

Testsystem und Testumgebung Für die Durchführung der Studie wurde der prototypische Aufbau aus Kapitel 7.2.1 verwendet. Aufgrund der subjektiv eher negativ ausgefallenen Bewertung für Variante 4 wurde diese nicht weiter betrachtet. Im Folgenden werden also nur noch zwei adaptive Varianten einer statischen Variante gegenübergestellt. Zur Überprüfung der Eignung des Sensorsystems wurde die Möglichkeit implementiert, die Annäherungsfunktion durch einen Versuchsleiter („Wizard“) zu simulieren: Durch einen Tastendruck auf einer an das System angeschlossenen Tastatur konnte die Adaption gestartet werden. Damit standen für die Messung drei Modi zur Verfügung:

- **Statisch:** Die Bedienflächen waren von Anfang an komplett dargestellt, d. h. der Anzeigemodus wurde deaktiviert.
- **Wizard:** Der Bedienmodus wurde durch einen menschlichen Beobachter bei erkannter Bedienintention aktiviert.
- **Sensorik:** Der Bedienmodus wurde durch die in Abschnitt 6.3 beschriebene Sensorik aktiviert.

Die Messung der Bewegungszeiten vom Lenkrad bis zur Aktivierung einer Bedienfläche wurde durch Loslassen eines Tiptronic-Hebels (TT-Hebel) hinter dem Lenkrad gestartet und endete bei der ersten darauf folgenden Berührung des Touchscreens (auch bei einem Fehltreffer). Während der Bedienbewegung wurden Zwischenzeiten bei Auslösen der Adaption sowie bei Abschluss der Adaptionphase gemessen. Mit Hilfe eines *Lane Change Tests*¹ wurde parallel eine Primäraufgabe (Fahraufgabe) simuliert (siehe auch [64]).

Für die Bewegungsmessung stand darüber hinaus eine 3D-Cave zur Verfügung, in der die Position eines am Bedienfinger des Versuchsteilnehmers befindlichen Marker-Objekts gemessen wurde. Das Messsystem

¹Der *Lane Change Test* (LCT) ist ein Instrument zur Evaluation der Ablenkung des Fahrers durch Sekundäraufgaben und eignet sich speziell für Doppelaufgaben-Experimente. Derzeit befindet er sich im Prozess zur ISO-Standardisierung.

wurde zu Beginn auf die zu treffenden Schaltflächen auf dem Touchscreen kalibriert. Es wurde die direkte Distanz A des Bedienfingers zum Ziel (Luftlinie) gemessen und nicht die absolute Position im Koordinatenraum (in Bild 7.13 ist die Bewegung beispielhaft mit den Teilbewegungen $TB1$ und $TB2$ dargestellt).

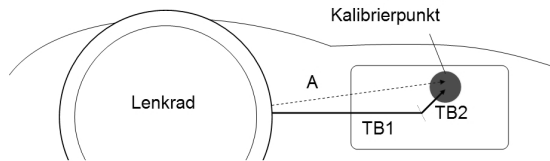


Bild 7.13: Messung der Distanz des Bedienfingers zum Touchscreen.

Versuchsdesign Bei der Untersuchung handelte es sich um ein ‚3×3-Within-Design‘ mit folgenden unabhängigen Variablen:

- **Modus:** statisch, Wizard, Sensorik;
- **Bedienvariante:** Variante 1 (statisch, keine Indikation), Variante 2 (adaptiv, keine Indikation), Variante 3 (adaptiv, geometrische und semantische Indikation).

Die Reihenfolge, in der die Bedienvarianten vorgegeben wurden, war genauso wie die Reihenfolge der Modi permutiert. Bezüglich des Tastendrucks musste immer zunächst die Taste ‚Play‘ und darauffolgend die Taste ‚TP‘ gedrückt werden.

Als abhängige Variablen wurden erhoben:

- **Handbewegungszeiten:** Zeit vom Start der Bewegung bis zum Treffen der Schaltfläche sowie Zwischenzeiten bei Start und Ende der Adaption in den adaptiven Varianten;
- **Handbewegungsverlauf:** Distanz zum Ziel im zeitlichen Verlauf.

Versuchsablauf Alle Probanden erhielten zunächst eine Einweisung in den LCT und die Funktionsweise der unterschiedlichen Bedienvarianten

und konnten beide Systeme ausgiebig testen. Anschließend wurden die Probanden instruiert, parallel zur Fahraufgabe für jede der drei Bedienvarianten jeweils zwei definierte Tasten (,Play‘ und ,TP‘) in separaten Bedienhandlungen so zügig wie möglich zu betätigen. Dies wurde für die Varianten 2 und 3 jeweils in den drei unterschiedlichen System-Modi durchgeführt. Die Variante 1 konnte nur in einem Modus (statisch) bedient werden. Bild 7.14 zeigt einen typischen Bedienablauf und die dazu aufgezeichneten Daten.






				
Ziehen des Lenkstockhebels.	Loslassen des Lenkstockhebels. Start der Zeitmessung.	Eintritt in den Detektionsbereich. Start der Adaption.	Abschluss der Adaption.	Berührung des Touchscreens. Ende der Zeitmessung
Zustand	start	entry	menu_complete	button play
Zeit seit Start (s)	0	0,328	0,594	0,75
Phasendauer (s)	0	0,328	0,266	0,156

Bild 7.14: Datenaufzeichnung einer typischen Bedienung.

Stichprobenbeschreibung Alle Versuchsteilnehmer waren Mitarbeiter der Volkswagen AG, die mit der Funktionsweise und dem Menüaufbau des adaptiven Touchscreens vertraut sind. Unter den 15 Versuchsteilnehmern waren 11 Männer und 4 Frauen. Das mittlere Alter lag bei 30,13 Jahren ($\sigma = 5,74$ Jahre).

7.3.4 Messgrößen und Datenaufbereitung

Insgesamt standen für die Auswertung 270 Aufzeichnungen von Bedienbewegungen zur Verfügung. Diese setzten sich aus je zwei Tastendrücken pro Versuchsperson für jede der drei Bedienvarianten in den System-Modi ,Sensor‘, ,Wizard‘ und ,Statisch‘ zusammen. Fehlerhaft aufgezeichnete Datensätze gingen nicht in die Analyse ein.

In allen System-Modi wurde für jede der drei Bedienvarianten jeweils die mittlere Bewegungsdauer MT vom Start der Messung bis zur Berührung des Touchscreens durch Mittelwertbildung über alle betreffenden individuellen Bewegungszeiten bestimmt. Die Fehlerrate ER errechnete sich aus der Relation der absoluten Anzahl der bei einer Variante aufgetretenen Bedienfehler aller Nutzer in Bezug auf die Anzahl aller insgesamt ausgeführten Bedienaufgaben innerhalb der betreffenden Variante.

Analog hierzu wurden die mittleren Bewegungszeiten für die separaten Bewegungsphasen (MT_1 , MT_2 , MT_3) durch Mittelwertbildung aller individuellen Bewegungszeiten für die entsprechende Phase innerhalb der betreffenden Variante bestimmt. MT_1 beschreibt die Phase von *start* bis *entry*, MT_2 von *entry* bis *menu_complete* und MT_3 von *menu_complete* bis *button_play* (siehe Bild 7.14).

Zur Überprüfung des OIIMs wurden Zeit-Weg-Diagramme aus den aufgezeichneten Bewegungsdaten erzeugt (siehe Bild 7.15 rechts oben) und hieraus Zeit-Geschwindigkeits-Diagramme (Bild 7.15 links oben stellt einen idealisierten Geschwindigkeitsverlauf dar) berechnet. Die Zeit-Weg-Diagramme wurden daraufhin abschnittsweise linearisiert. Die erste Phase der linearisierten Bewegung entspricht der ballistischen Phase des OIIM und ist durch einen steilen Verlauf gekennzeichnet. Als zeitliche Start- und Endpunkte dieser Phase wurden die Punkte T_0 bzw. T'_0 definiert, bei denen die gefilterten zugehörigen Geschwindigkeitswerte erstmals über bzw. unter einem Schwellwert von 0,25 m/s lagen. Anschließend wurde dieser Diagramm-Abschnitt extrahiert und es wurde durch die dazugehörigen Anfangs- und Endpunkte eine Gerade gelegt. Aus der Steigung dieser Geraden ergab sich unmittelbar die linearisierte Geschwindigkeit der ballistischen Phase, die im Folgenden als ‚Annäherungsgeschwindigkeit‘ bezeichnet wird. Aus Bild 7.15 rechts unten wird ersichtlich, dass eine solche Bestimmung des linearisierten Bewegungsverlaufs wenig vom idealisierten Verlauf abweicht. Hingegen führt die Linearisierung auf Basis der mittleren Geschwindigkeit im Intervall $[T_0, T'_0]$ des Zeit-Geschwindigkeits-Diagramms zu erheblichen Fehlern in der ersten Hälfte des Verlaufs.

Die zweite Phase der linearisierten Bewegung wurde durch die direkte Verbindung zwischen dem Endpunkt der ballistischen Phase (der verbleibende Restabstand zum Ziel bei Ende der ballistischen Phase) und dem Endpunkt der Gesamtbewegung beschrieben.

Aufgrund der Messung des direkten Abstandes D zwischen Kalibrier-

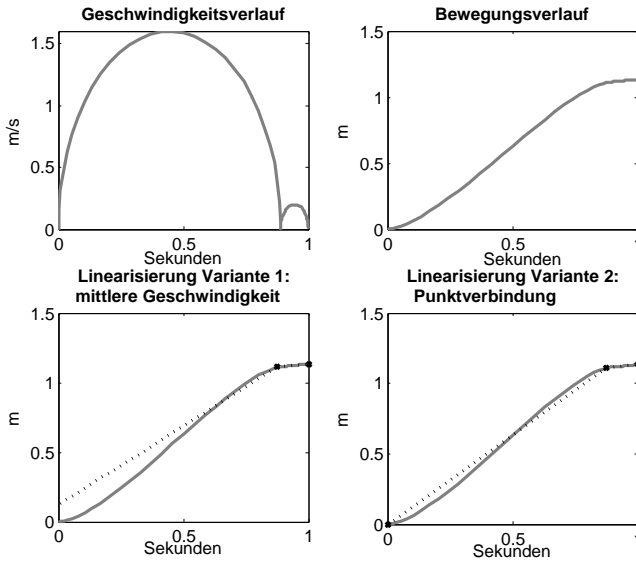


Bild 7.15: Idealisiertes Bewegungsmodell mit Linearisierung der Bewegungstrajektorie in der Selektions- und der ballistischen Phase.

punkt und Marker auf dem Bedienfinger ergab sich eine Verzerrung der Bewegungskurven, da die tatsächliche Bewegung zu Beginn nicht direkt auf den Touchscreen hin erfolgte, sondern zu einem virtuellen Punkt vor dem Display. Bild 7.16 zeigt jedoch, dass die Verzerrung für die Extraktion der Charakteristika vernachlässigt werden kann. Simulationen ergaben, dass der Fehler unter 2 % liegt.

Als charakteristische Werte zur Beschreibung der linearisierten Bewegungsdiagramme wurden somit die Annäherungsgeschwindigkeit sowie der mittlere Restabstand zum Ziel nach Ende der ballistischen Phase extrahiert.

Zur Spezifikation einer Sensorik zur Annäherungserkennung wurden diese beiden Charakteristika separat in einem ‚Worst-Case‘-Szenario berechnet. Hierbei wurden für die Sensorik eine sehr hohe Annäherungsgeschwindigkeit und ein sehr großer Restabstand zum Ziel als kritisch betrachtet, da diese die minimale Sensorreichweite bestimmen. Dementsprechend wurden pro Phase die 10 % diesbezüglich kritischsten Daten-

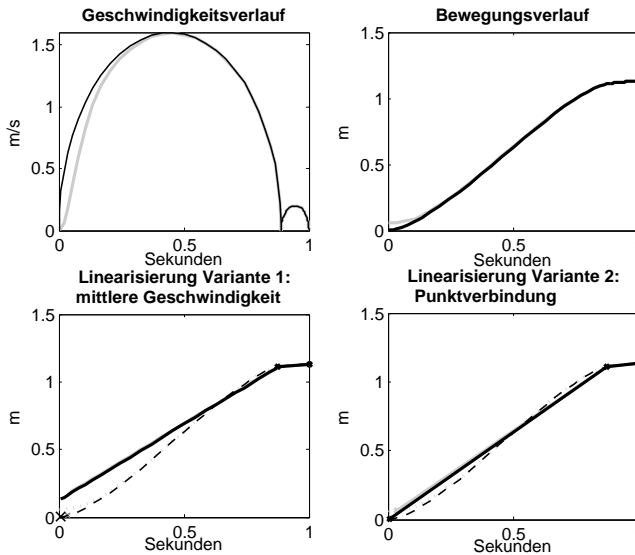


Bild 7.16: Idealisiertes Bewegungsmodell mit Linearisierung der Bewegungstrajektorie; schwarz: Idealisierte Modelldarstellung; grau: Verzerrung durch Messung wie in Bild 7.13 gezeigt.

sätze zum jeweiligen Charakteristikum ausgewertet, und es wurde hieraus ein zusammengesetzter, gemittelter Verlauf gebildet.

7.3.5 Ergebnisse

Mittlere Bewegungsdauer Die Ergebnisse zeigen, dass die adaptiven Bedienvarianten mit einer mittleren Bewegungszeit von 0,786 s (Standardabweichung $\sigma = 0,168$ s) und 0,816 s ($\sigma = 0,147$ s) in der statischen Ausprägung im Vergleich zur statischen Variante 1 signifikant schneller zu bedienen waren (siehe Bild 7.17). Dies ist nach Fitts auf die größeren Schaltflächen zurückzuführen.

Im Wizard-Modus ist die adaptive Variante mit Indikatoren im Mittel schneller bedienbar als die statische Variante 1, der Unterschied ist jedoch nicht signifikant. Die mittlere Bewegungsdauer unterscheidet sich zwischen den Modi Wizard und Statisch nur bei Variante 2 signifikant,

nicht jedoch bei Variante 3. Im Wizard-Modus zeigt sich außerdem, dass die Bewegungsdauer bei Variante 3 mit Indikation tendenziell kürzer ist als in Variante 2, allerdings ist der Unterschied nicht signifikant.

Im Sensor-Modus zeigt sich in allen Fällen, dass die Bewegungszeiten wesentlich und signifikant länger sind.

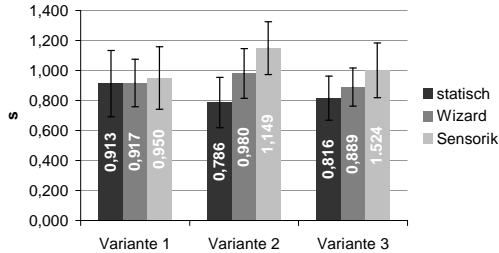


Bild 7.17: Mittlere Bewegungszeiten MT für die einzelnen Varianten in den einzelnen Modi.

Fehlerrate In der Variante 1 wurden zwei fehlerhafte Tastendrucke registriert, in den anderen Varianten jeweils keine. Die Fehlerrate von Variante 1 beläuft sich damit auf 7%.

Bewegungsdiagramme Bild 7.18 oben zeigt den gemessenen Bewegungsverlauf für alle Versuchspersonen in der Variante 2 im statischen Modus. Dabei wurden die Bewegungsverläufe für eine bessere Vergleichbarkeit im Gegensatz zur Darstellung in Bild 7.15 auf der negativen Zeitachse dargestellt. $t = 0\text{ s}$ entspricht dem Treffen der Taste. Bild 7.18 unten zeigt den aus dem Bewegungsverlauf berechneten Geschwindigkeitsverlauf. Zunächst kann bei allen Bewegungen eine schnelle Zielbewegung festgestellt werden. In der darauf folgenden Phase finden kleine Korrekturbewegungen sowohl mit negativen als auch positiven Geschwindigkeitsverläufen statt. Der Zeitpunkt $t = 0\text{ s}$ entspricht in dieser Darstellung dem Beginn der Bewegung. Im Vergleich dazu sind in Bild 7.19 und 7.20 der Bewegungs- und Geschwindigkeitsverlauf in der Variante 2 für den Wizard- und Sensor-Modus dargestellt. Im Vergleich zum statischen Modus zeigen sich eine größere Streuung der Werte sowie eine

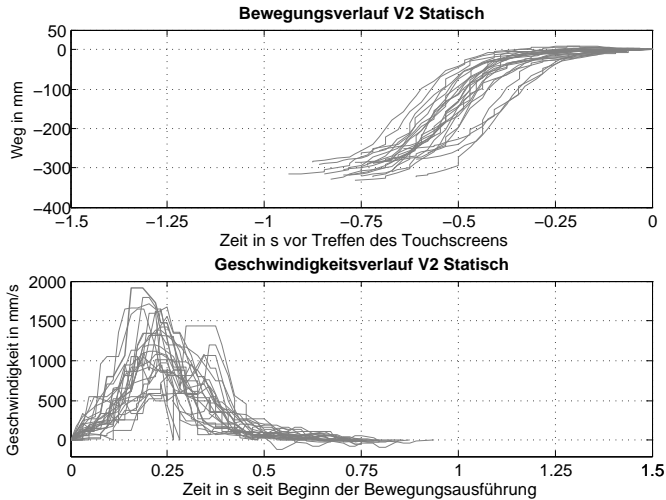


Bild 7.18: Bewegungs- und Geschwindigkeitsverlauf der einzelnen Versuchspersonen in Variante 2 im statischen Modus.

Verlängerung der Selektionsphase. Auch insgesamt verlängert sich, wie auch aus Bild 7.17 ersichtlich ist, die Bewegungsdauer. Nichtsdestotrotz sind auch hier mehrere Teilbewegungen erkennbar. Ähnliche Bewegungs- und Geschwindigkeitsverläufe zeigten sich auch in den Varianten 1 und 3.

Bedienphasen Bereits aus den Bildern 7.18 und 7.19 ist zu entnehmen, dass die ballistische Phase nach ca. 50% der Gesamtbedienzeit abgeschlossen ist. Dabei werden ca. 90% des Weges zurückgelegt. Bild 7.21 zeigt auf der Abszisse die mittlere Bewegungszeit der ballistischen Phase in Prozent der Gesamtbewegungszeit. Auf der Ordinate ist die mittlere zurückgelegte Wegstrecke in dieser Phase in Prozent der Gesamtwegstrecke dargestellt. Die eingetragenen Werte markieren das Ende der ballistischen Phase für die einzelnen Modi und Varianten. Im statischen Modus wird in allen Varianten in der ballistischen Phase ein größerer Anteil des Weges zurückgelegt als in den Modi Wizard oder Sensor. Trotzdem lässt sich auch in den Modi Wizard und Sensor (mit Ausnahme von Variante 2 im Modus Sensor) die Hypothese des OIIM (90% des Weges in 50%

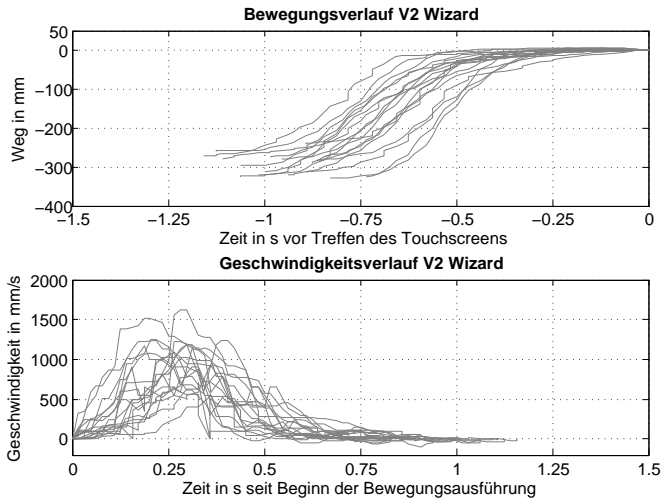


Bild 7.19: Bewegungs- und Geschwindigkeitsverlauf der einzelnen Versuchspersonen in Variante 2 im Wizard-Modus.

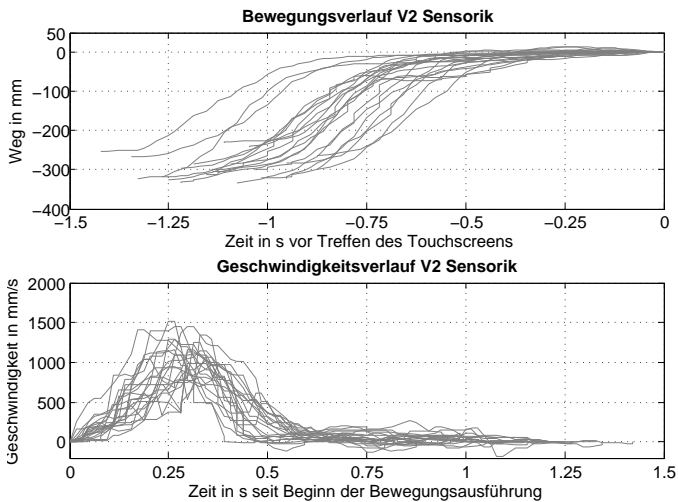


Bild 7.20: Bewegungs- und Geschwindigkeitsverlauf der einzelnen Versuchspersonen in Variante 2 im Sensor-Modus.

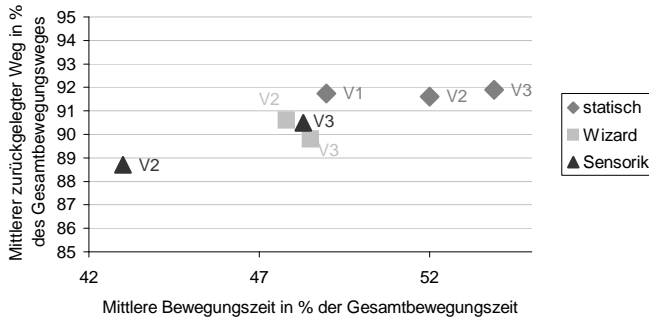


Bild 7.21: Zusammenhang zwischen dem mittleren zeitlichen Abschluss der ballistischen Phase in Prozent der Gesamtbewegung und dem mittleren zurückgelegten Weg in der ballistischen Phase in Prozent der Gesamtweglänge der Bewegung für die einzelnen Modi und Varianten.

der Zeit) in guter Näherung bestätigen. Variante 1 weist im Vergleich zu den Varianten 2 und 3 im statischen Modus eine prozentual kürzere ballistische Phase auf und damit eine längere Selektionsphase. Dies ist nach Fitts auf die kleinere Größe der Tasten zurückzuführen.

Der geringere Anteil an Wegstrecke während der ballistischen Phase in Variante 2 im Modus Sensorik ist wahrscheinlich auf eine zu langsame Reaktion der Sensorik in Kombination mit fehlender Planungsmöglichkeit aufgrund nicht vorhandener Indikatoren zurückzuführen. Es ist anzunehmen, dass die größere Unsicherheit über die Position der Bedienflächen zu einer Verkürzung des zurückgelegten Weges in der ballistischen Phase führt.

Bild 7.22 zeigt das mittlere prozentuale Ende der ballistischen Phase T'_0 sowie die geloggtten Events *entry* und *menu_complete*. Für den Wizard-Modus lässt sich feststellen, dass das Ende der ballistischen Phase nach der vollständigen Adaption des Menüs liegt (oder wie in Variante 3 mit diesem zusammenfällt), im Sensor-Modus ist dies genau umgekehrt. Dies stärkt die These, dass die Sensorik zu langsam reagiert, um eine kontinuierliche Bewegung zuzulassen. Der Nutzer muss nach dem Ende der ballistischen Phase zunächst warten, bis das Menü vollständig angezeigt wird und kann erst dann mit einer zielgerichteten Selektionsphase beginnen.

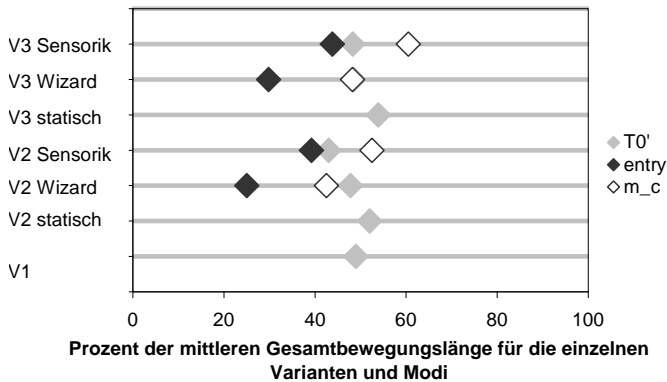


Bild 7.22: Mittleres Ende der ballistischen Phase sowie geloggte Events *entry* und *menu_complete* in Prozent der Gesamtbewegungszeit.

Absolut gesehen zeigt sich (siehe Bild 7.23), dass die Länge der ballistischen Phase fast unabhängig von Variante und Modus ist, die längere Bewegungszeit im Modus Sensorik kommt also vor allem durch die verlängerte Selektionsphase zu Stande.

Sowohl im Wizard-Modus als auch im Sensorik-Modus (Bild 7.24) zeigt sich bei näherer Betrachtung der einzelnen Bedienphasen MT_1 , MT_2 und MT_3 , dass die letzte Phase MT_3 , in der adaptiven Variante ohne Indikatoren wesentlich länger als in der Variante mit Indikatoren ausfällt. Die Bedienphasen MT_1 und MT_2 sind hingegen nahezu identisch.

Die längere Dauer von MT_3 ist wahrscheinlich auf die geringere Genauigkeit der ballistischen Phase sowie eine geringere Geschwindigkeit aufgrund von fehlender Information zur Bewegungsplanung zurückzuführen. Entsprechend sind häufigere und längere Korrekturbewegungen notwendig.

Bewegungsdiagramme aus charakteristischen Werten Tabelle 7.3 gibt einen Überblick über die aus den Bewegungsdiagrammen berechneten charakteristischen Werte für das *Worst-Case*-Szenario. Tabelle 7.4 zeigt die Werte analog, berechnet aus den Mittelwerten der Bewegungsdaten. $s_{m, \text{selektion}}$ bezeichnet dabei die mittlere noch zurückzulegende Teilstrecke nach Ende der ballistischen Phase für die Variante m ,

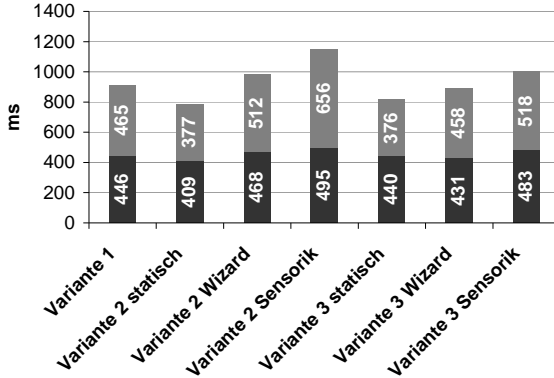


Bild 7.23: Mittlere absolute Bewegungszeiten aufgedgliedert nach ballistischer und Selektionsphase für die einzelnen Varianten und Modi.

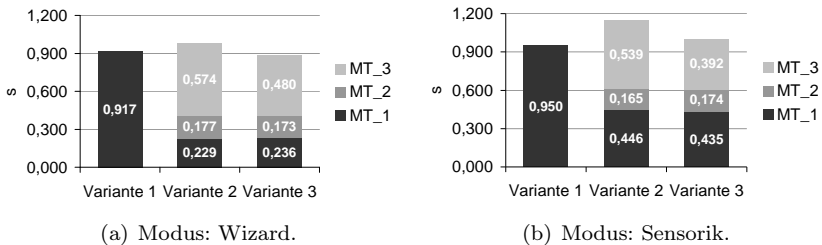


Bild 7.24: Mittlere Bewegungszeit für die einzelnen Varianten in den Bedienphasen MT_1 , MT_2 , MT_3 .

$v_{m,ballistisch}$ die angenäherte Steigung des Bewegungsverlaufs in der ballistischen Phase. Der Sensor-Modus ist hier nicht aufgeführt, da zur Festlegung der optimalen Sensorparameter nur charakteristische Werte herangezogen werden sollen, die aus dem für eine Variante optimalen Bewegungsverlauf für ein adaptives Touchinterface gewonnen wurden. Bild 7.25 bis Bild 7.28 zeigen die linearisierten Bewegungsdiagramme gemittelt über alle Bewegungstrajektorien sowie für die *Worst-Case*-Betrachtung für alle Varianten und Modi.

Tabelle 7.3: Charakteristische Werte für das *Worst-Case*-Szenario zur Festlegung der Sensorparameter.

Variante	Variante 1 nicht adaptiv		Variante 2 adaptiv ohne Ind.		Variante 3 adaptiv mit Ind.	
	Statisch	Wizard	Statisch	Wizard	Statisch	Wizard
$s_{m,selektion}$ in mm	40,60	44,21	62,07	38,11	36,36	
$v_{m,ballistisch}$ in mm/s	1003	1106	843,6	1007	771,3	

Tabelle 7.4: Charakteristische Werte aus den Mittelwerten (zur Festlegung der Sensorparameter).

Variante	Variante 1 nicht adaptiv		Variante 2 adaptiv ohne Ind.		Variante 3 adaptiv mit Ind.	
	Statisch	Wizard	Statisch	Wizard	Statisch	Wizard
$s_{m,selektion}$ in mm	18,34	20,67	24,25	21,18	16,80	
$v_{m,ballistisch}$ in mm/s	759,5	805,1	658,4	731,4	663,1	

Es zeigt sich, dass sowohl $s_{m,selektion}$ als auch $v_{m,ballistisch}$ im stati-

schen Modus in allen Varianten vergleichbar sind. Die adaptiven Varianten weisen im Wizard-Modus (wie auch im Sensor-Modus) geringere $v_{m,ballistisch}$ auf als im statischen Modus. Die mangelnde Information über das zu treffende Ziel führt anscheinend in den adaptiven Modi und Varianten zu einer Reduktion der Geschwindigkeit in der ballistischen Phase. Der Restabstand $s_{3,selektion}$ in Variante 3, Modus Wizard (mit sehr frühem Adaptionszeitpunkt und Indikation) ist im *Worst-Case*-Szenario vergleichbar mit den Restabständen im statischen Modus. Im Sensor-Modus (mit späterem Adaptionszeitpunkt) weist Variante 3 genauso wie Variante 2 in beiden adaptiven Modi einen deutlich größeren Restabstand auf. Vergleicht man die Varianten 2 und 3 miteinander in den Modi Wizard und Sensor, so zeigt sich, dass in allen Fällen der Restabstand in Variante 2 größer ist als in Variante 3, die Geschwindigkeiten in der ballistischen Phase sind jedoch vergleichbar.

Die verfügbare Information über das zu treffende Ziel während der ballistischen Phase scheint also einen Einfluss auf den Restabstand zum Ziel zu haben. Wichtig anzumerken hierbei ist jedoch, dass sich dies nicht unbedingt auf die Gesamtbewegungsdauer auswirkt, für die Spezifikation der Sensorik jedoch ein wichtiges Kriterium ist. Ein geringerer Restabstand zum Ziel am Ende der ballistischen Phase führt zu einer geringeren minimal erforderlichen Sensorreichweite.

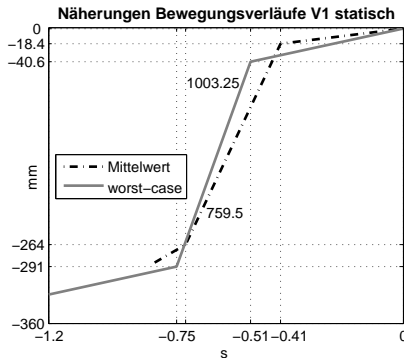


Bild 7.25: Linearisierter Bewegungsverlauf für die Variante 1.

Die minimal erforderliche Reichweite der Sensoren s_{sensor} (ausgehend

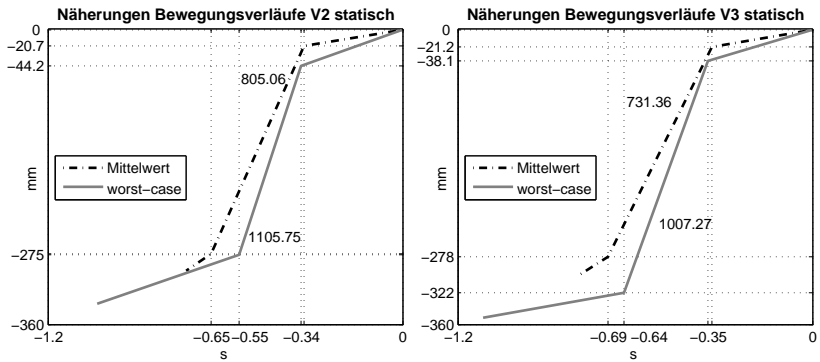


Bild 7.26: Linearisierter Bewegungsverlauf für die Varianten 2 (links) und 3 (rechts) im statischen Modus.

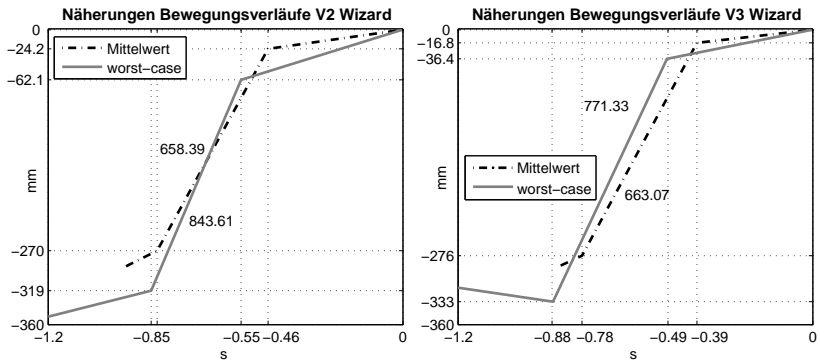


Bild 7.27: Linearisierter Bewegungsverlauf für die Varianten 2 (links) und 3 (rechts) im Wizard-Modus.

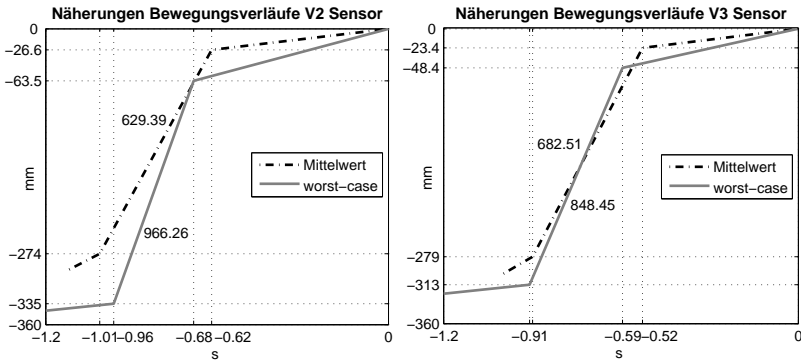


Bild 7.28: Linearisierter Bewegungsverlauf für die Varianten 2 (links) und 3 (rechts) im Sensorik-Modus.

vom Ziel in Bewegungsrichtung der Bedienhand) kann in Abhängigkeit der Adaptiondauer T_A und der Systemreaktionsdauer T_S folgendermaßen bestimmt werden:

$$s_{\text{sensor}} = s_{m,\text{selektion}} + v_{m,\text{ballistisch}} \cdot (T_A + T_S). \quad (7.4)$$

Dabei wird die Annahme zugrunde gelegt, dass die Adaption mit Beginn der Selektionsphase abgeschlossen sein muss.

Die sich für Adaptiondauern T_A von 50, 100 und 200 ms für die einzelnen Varianten aus den Wizard-Messungen ergebenden minimal erforderlichen Reichweiten der Sensoren vom Ziel aus betrachtet sind in Tabelle 7.5 dargestellt. Dabei wurde T_S zunächst vernachlässigt. Es zeigt sich, dass die Auswahl der Bedienvariante Einfluss auf die Definition der Mindestreichweite s_{sensor} der Sensoren hat.

Nimmt man jedoch eine Systemverzögerungszeit von 100 ms an, ergeben sich die in Tabelle 7.6 dargestellten Werte. Die Annahme einer Systemverzögerungszeit von $T_S = 100$ ms ist im vorliegenden System durchaus gerechtfertigt, wie Messungen ergeben hatten. Dies ist einerseits durch die Abtastrate des Systems und andererseits durch Verzögerungen in der Kommunikation der Sensorik mit der Macromedia-Flash-Anwendung bedingt.

Aus Tabelle 7.6 lässt sich also ablesen, dass bei einer Adaptionzeit

Tabelle 7.5: Minimal erforderliche Reichweite der Sensoren vom Ziel aus betrachtet für verschiedene Adaptionen für die einzelnen Varianten, ermittelt aus den Wizard-Messungen unter Vernachlässigung von T_S .

Szenario	Worst-Case			Mittelwert		
	50 ms	100 ms	200 ms	50 ms	100 ms	200 ms
T_A	50 ms	100 ms	200 ms	50 ms	100 ms	200 ms
$s_{\text{sensor},2}$ in mm	104	146	231	57	90	156
$s_{\text{sensor},3}$ in mm	75	113	191	50	83	149

von $T_A = 100$ ms und einer Sensorreichweite von 100 mm kein optimaler Bewegungsverlauf zu erzielen ist. Eine wesentlich größere Sensorreichweite ist jedoch im Fahrzeug nicht implementierbar, da ansonsten Bewegungen am Lenkrad oder Greifbewegungen zum Schalthebel zu einer Auslösung der Adaption führen würden. Geht man jedoch von einer verringerten Systemreaktionszeit von $T_S = 20$ ms und einer Adaptiondauer von $T_A = 100$ ms in Variante 3 aus, liegt die minimal erforderliche Sensorreichweite bei 96 mm, also unter dem maximal möglichen Wert. Für die *Worst-Case*-Berechnung würde jedoch auch hier keine optimale Bewegung erreicht werden können, hierfür wäre eine Adaptiondauer von 50 ms notwendig. Dabei ist jedoch zu beachten, dass für die *Worst-Case*-Betrachtungen nur die 10% ungünstigsten Datensätze herangezogen wurden, d. h. nur eine sehr geringe absolute Anzahl an Bewegungsdaten.

Tabelle 7.6: Minimal erforderliche Reichweite der Sensoren vom Ziel aus betrachtet für verschiedene Adaptionen für die einzelnen Varianten, ermittelt aus den Wizard-Messungen mit $T_S = 100$ ms.

Szenario	Worst-Case			Mittelwert		
	50 ms	100 ms	200 ms	50 ms	100 ms	200 ms
T_A	50 ms	100 ms	200 ms	50 ms	100 ms	200 ms
$s_{\text{sensor},2}$ in mm	189	231	315	123	156	222
$s_{\text{sensor},3}$ in mm	152	191	268	116	149	216

7.3.6 Interpretation und Fazit

Die Ergebnisse der Studie bestätigen, dass die Adaption der Bedienoberflächen bei Bedienintention des Nutzers den potenziellen Vorteil bietet, Bedienfehler bei gleichen oder sogar etwas kürzeren Bewegungszeiten deutlich zu reduzieren. Zur genaueren Bewertung der Ergebnisse findet im Folgenden ein Abgleich mit den zugrunde liegenden Hypothesen statt:

H1. Die Bedienbewegungen zur Bedienung eines Touchscreens im Fahrzeug lassen sich sowohl für adaptive als auch für nicht-adaptive Bedienvarianten mit dem OIIM beschreiben.

Diese Hypothese lässt sich auf Basis der Ergebnisse voll bestätigen. Die aufgezeichneten Bewegungsverläufe sowohl der statischen als auch der adaptiven Bedienvarianten zeigen eine Kurvenschar entsprechend dem OIIM. Mit der ballistischen Bewegungsphase werden ca. 90 % des Weges bis zum Ziel zurückgelegt und sie ist durchschnittlich nach ca. 50 % der Gesamtbewegungsdauer abgeschlossen. Darauf folgt die Selektionsphase, die sich aus unterschiedlich vielen Teilbewegungen zusammensetzt.

H2. Eine Adaption der Zielflächen während der Bedienbewegung führt im Vergleich zu einer statischen Bedienvariante zu geringeren Fehlerraten bei vergleichbaren Bewegungszeiten.

Diese Hypothese lässt sich bestätigen. Für die Bedienung der adaptiven Bedienvarianten lag die Fehlerrate bei 0 % – für die nicht adaptive Variante jedoch bei 7 %. Aufgrund der geringen Anzahl an Fehltreffern in dieser Studie sind die Zahlen jedoch nur eingeschränkt aussagefähig, konnten jedoch tendenziell in anderen Studien bestätigt werden. In der Praxis sind Fehlerraten von bis zu 5 % akzeptabel, diese Schwelle wurde von der nicht-adaptiven Variante jedoch bereits in dem für diese Studie genutzten stehenden Simulator deutlich überschritten. Eine Untersuchung der Bedienbarkeit in einem fahrenden Fahrzeug würde daher voraussichtlich zu noch höheren Fehlerraten für alle Bedienvarianten führen. Die mittleren Bewegungszeiten unterschieden sich entsprechend Bild 7.17 im Wizard-Modus nicht signifikant von den Bewegungszeiten in der statischen Variante 1 und können daher als vergleichbar betrachtet werden.

H3. Der Einsatz von Indikatoren führt bei der Bedienung von adaptiven Zielen zu kürzeren Bewegungszeiten.

Diese Hypothese lässt sich nur tendenziell bestätigen. Die Bewegungszeiten zur Bedienung von Variante 3 waren im Mittel zwar kürzer als die Bewegungszeiten zur Bedienung von Variante 2, jedoch nicht signifikant.

Sowohl die Selektionsphase als auch die Phase MT_3 waren in Variante 3 kürzer als in Variante 2. Die Selektionsphase in Variante 3 ist vergleichbar mit den Selektionsphasen im statischen Modus. Dies ist wahrscheinlich darauf zurückzuführen, dass die Indikation eine bessere Planung und somit genauere Ausführung der ballistischen Bewegungsphase ermöglicht, und somit am Ende entsprechend weniger Korrekturbewegungen nötig sind.

Im Rahmen der vorliegenden Untersuchung wurden Position und Funktion der finalen Bedienfläche per Indikation anhand einer hinsichtlich Dimension und Kontrast reduzierten Fläche im Anzeigemodus vermittelt. Einige Nutzer gaben an, die Indikation aufgrund ihrer entsprechend reduzierten visuellen Präsenz nicht vollständig nutzen zu können. In nachfolgenden Studien wäre daher zu untersuchen, welche grafische Präsenz der Indikation optimal ist und welchen Einfluss die Indikation weiterer Informationen über das Ziel auf die Bedienbewegung hat.

H4. Das bisher verwendete Sensorsystem stört einen optimalen Bewegungsablauf und führt daher zu längeren Bewegungszeiten. Diese Hypothese konnte bestätigt werden. Der verwendete prototypische Aufbau führt zu Systemverzögerungen von etwa 100 ms. Bei einer voreingestellten Adaptiondauer von ebenfalls 100 ms müsste die Sensorreichweite bei 231 mm bzw. 191 mm im *Worst-Case*-Szenario liegen, um einen optimalen Bewegungsverlauf zu erreichen. Eine Sensorreichweite dieser Größenordnung ist jedoch im Fahrzeug aufgrund der dann zahlreichen Fehlauflösungen nicht machbar. Es zeigte sich auch, dass die zu langsame Systemreaktion in Kombination mit der zu langen Adaptiondauer zwar die Geschwindigkeit der ballistischen Phase nicht beeinflusst, den Restabstand zum Ziel am Ende der ballistischen Phase jedoch maßgeblich.

H5. Aus den ermittelten Bewegungsabläufen lassen sich spezielle Leistungsmerkmale für ein technisches Sensorsystem zur Annäherungserkennung ableiten.

Eine Ableitung von Leistungsmerkmalen für ein Sensorsystem auf Basis theoretischer Grundlagen konnte gezeigt werden.

Bei der Bestimmung der Parameter zur Spezifikation einer Sensorik wurde entsprechend dem OIIM davon ausgegangen, dass die Adaption des Ziels erst bei Beginn der Selektionsphase vollständig abgeschlossen sein muss, da während der ballistischen Bewegungsphase der Bewegungsverlauf nicht auf Basis von visuellem Feedback verändert werden kann. Es

deutet sich jedoch an, dass visuelles Feedback auch bereits während der ballistischen Bewegungsphase vorteilhaft für die Planung der anschließenden Selektionsbewegung insofern genutzt werden kann, als der Restabstand zum Ziel nach Abschluss der ballistischen Phase bei Einsatz von Indikatoren deutlich geringer ausfällt als bei fehlender Indikation. Eine genaue Betrachtung der Blickzuwendung während der ballistischen Phase ist Gegenstand der nachfolgenden Studie.

Aus den gewonnenen Daten konnte jedoch abgeleitet werden, dass bei einer Systemreaktionsdauer von $T_S = 20$ ms, einer Adaptiondauer von $T_A = 100$ ms und dem Einsatz von Indikatoren die minimale Sensorreichweite bei 96 mm liegt, also unterhalb der maximal möglichen 10 cm.

7.4 Studie zur Blickbewegung bei der Bedienung eines annäherungssensitiven Touchscreens

7.4.1 Fragestellung und Zielsetzung der Studie

Im Rahmen dieser Studie sollte der Einfluss adaptiver Schaltflächen bei der Bedienung eines Touchscreens im Fahrzeug auf das Blickverhalten des Fahrers untersucht und beschrieben werden. Daneben ist es Ziel, die bereits in Kapitel 7.3 erhobenen Daten zu bestätigen. Besonderes Augenmerk lag auch hier auf dem Effekt von abstrakter Indikation der Schaltflächen im Anzeigemodus, sowie auf dem Vergleich von nicht-adaptivem mit adaptivem Verhalten.

Hierzu wurden folgende Hypothesen aufgestellt:

- **H1.** Durch die Adaption der Zielflächen während der Bedienbewegung ändern sich im Vergleich zu statischen Bedienvarianten die Blickabwendungszeiten von der Primäraufgabe nicht maßgeblich.
- **H2.** Der Einsatz von Indikatoren führt bei der Bedienung von adaptiven Zielen zu geringeren Blickabwendungszeiten von der Primäraufgabe.
- **H3.** Der Einsatz von Indikatoren führt bei der Bedienung von adaptiven Zielen zu geringeren Fixationszeiten des adaptiven Touchdisplays.
- **H4.** Eine Blickzuwendung findet nicht nur während der Selektionsphase, sondern bereits in der ballistischen Phase statt.

7.4.2 Konzeption

Testsystem und Testumgebung Die Konzeption (Testsystem und Testumgebung sowie Versuchsablauf) der Studie erfolgte analog der Konzeption der in Kapitel 7.3.3 beschriebenen Studie. Die Position des Bedienfingers wurde jedoch nicht über die gesamte Bewegung verfolgt.

Die Blickbewegungsmessung wurde mit dem iView X HED System der Firma SMI (Senso Motoric Instruments) durchgeführt. Dabei handelt es sich um ein kopfbasiertes System (HED = *Head-mounted Eye-tracking Device*) mit einer Abtastrate von 50 Hz. Neben der Erfassung der Blickbewegungen ermöglicht das System auch die Videoaufzeichnung des Blickareals, innerhalb dessen der Fokus der Blickbewegungen dargestellt werden kann.

Versuchsdesign Als abhängige Variablen wurden in dieser Studie erhoben:

- **Blickverhalten:** Blickdaten mit Angaben über das jeweilige Blickareal, Fixationsdauern und -häufigkeiten im zeitlichen Verlauf sowie in Bezug auf die Bedienphasen (*start, entry, menu_complete, button*);
- **Handbewegungszeiten:** Zeit vom Start der Bewegung bis zum Tastendruck.

Versuchsablauf Nach einer ausführlichen Trainingsphase mit dem adaptiven Touchinterface wurden die Versuchspersonen mit dem Blickbewegungshelm ausgestattet und für die Messung kalibriert. Erst nachdem die Kalibrierung erfolgreich war, wurde mit der eigentlichen Messung begonnen. Der Versuchsablauf gestaltete sich analog zu dem in Kapitel 7.3.3 beschriebenen.

Stichprobenbeschreibung An der Studie (Blickbewegungsmessung) nahmen 22 Personen teil. Für 14 Versuchsteilnehmer konnten zufriedenstellende Blickdaten gewonnen werden, davon waren 9 Teilnehmer männlich und 5 weiblich. Das mittlere Alter lag bei 28,86 Jahren ($\sigma = 3,44$ Jahre). Im Gegensatz zu den Teilnehmern aus der vorherigen Studie waren diese noch nicht geübt im Umgang mit dem System.

7.4.3 Messgrößen und Datenaufbereitung

Insgesamt standen für die Auswertung 252 Aufzeichnungen von Blickbewegungen während der Bedienung zur Verfügung. Diese setzten sich aus je zwei Tastendrücken für jede der drei Bedienvarianten in den System-Modi statisch, Wizard und Sensorik zusammen. Fehlerhaft aufgezeichnete Datensätze gingen nicht in die Analyse ein.

Die Blickrohdaten wurden zunächst mithilfe der Software BeGaze (Firma SMI) aufbereitet, die eine Festlegung von mehreren ‚*Areas of Interest*‘ (AOIs) erlaubt. Auf dieser Basis wurden daraufhin Fixationen, Sakkaden und Lidschläge extrahiert. In der vorliegenden Studie wurden AOIs für den LCT-Monitor sowie für das adaptive Touchinterface definiert (siehe Bild 7.29). Mit Hilfe der graphischen Darstellung der Fixationen wur-

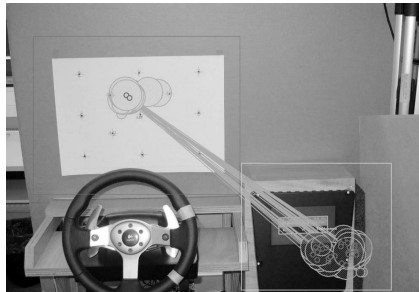


Bild 7.29: Darstellung des Blickverlaufs einer Versuchsperson und festgelegte AOIs in BeGaze; anstelle des LCTs ist die Kalibriervorlage zu sehen [148].

de hier bereits eine erste Kategorisierung der Daten im Sinne einer zufriedenstellenden Blickaufzeichnung vorgenommen. Darüber hinaus wurden die Videoaufnahmen mit dem übergelagerten Blickcursor kontrolliert. Fehlerhaft aufgezeichnete Datensätze, sowie Datensätze aus dem Wizard-Modus, bei denen die Zeit zwischen *start* und *menu_complete* über 400 ms lag (verspätete Reaktion des Versuchsleiters), gingen nicht in die weitere Analyse ein. Darüber hinaus wurden zunächst Datensätze mit einer Gesamtdauer von mehr als 5 Sekunden und im Folgenden (nach einer grafischen Analyse der Bewegungszeiten mittels Stengel-Blatt-Plot) Datensätze mit Bewegungszeiten über 2 Sekunden ausgeschlossen, da dies als Hinweis auf Probleme während der Datenerfassung gewertet wurde (siehe auch [148]).

Ein Großteil der untersuchten Variablen in dieser Studie weist eine schiefe Verteilung auf. Daher sind im Folgenden die Mediane und nicht die Mittelwerte angegeben. Als weitere Konsequenz wurde bei inferenzstatistischen Berechnungen auf nicht-parametrische Verfahren zurückgegriffen.

7.4.4 Ergebnisse

Mediane der Bewegungszeiten Im Gegensatz zu den Bewegungszeiten aus Kapitel 7.3 wurden hier deutlich längere Bewegungszeiten gemessen, was auf die geringere Übung der Versuchsteilnehmer in der Systembedienung zurückzuführen ist (siehe Bild 7.30). Es zeigt sich jedoch, dass auch hier die mittleren Bedienzeiten im Modus Sensor wesentlich höher sind als in den Modi statisch und Wizard; die mittleren Bedienzeiten im Wizard-Modus für die Varianten 2 und 3 sind vergleichbar mit der mittleren Bedienzeit in Variante 1. Ein genereller Vergleich der Bedienvarianten in den jeweiligen Modi zeigt keinen signifikanten Unterschied (Kruskal-Wallis-Test [25]). Ein Einfluss der Indikation in den Modi Wizard und Sensorik wird hier nicht sichtbar.

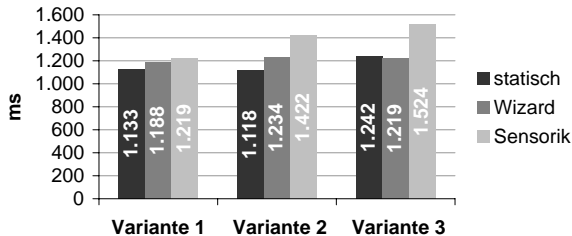


Bild 7.30: Mediane der Bewegungszeiten für die einzelnen Varianten.

Blickabwendung vom LCT und Blickzuwendung zum Touchscreen In Bild 7.31 sind die Mediane der Blickabwendungszeiten vom LCT bei Bedienung des Touchscreens dargestellt. Im Vergleich zu Bild 7.30 zeigen sich hier nur wenig Unterschiede, was darauf hindeutet, dass Blickabwendung und Bewegungszeit in hohem Maße korreliert sind ($r = 0,90$, $p < 0,001$, Spearman-Rang-Korrelation [25]). Ein Vergleich der drei Vari-

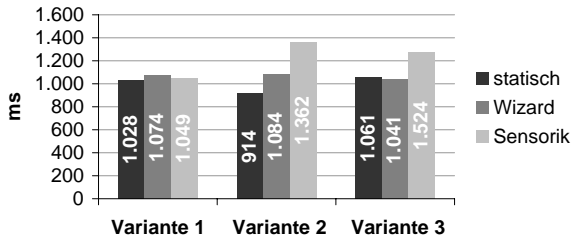


Bild 7.31: Blickabwendung vom LCT Bedienung.

anten bezüglich der Differenz zwischen Bedienzeit und Blickabwendungszeit zeigt ebenfalls keinen signifikanten Unterschied. Insgesamt ist jedoch bei jeder Variante die Blickabwendungszeit hochsignifikant kürzer als die Bedienzeit (alle $p < 0,001$).

Beim Vergleich der Bedienvarianten in Hinblick auf die Blickabwendungszeit unabhängig vom Modus zeigt sich kein signifikanter Unterschied. In einem weiteren Schritt wurde überprüft, ob ein Zusammenhang zwischen der jeweiligen Variante und der Dauer des Wechsels des Blicks vom LCT auf den Touchscreen (Ende letzte Fixation auf LCT – Beginn erste Fixation auf Touchscreen bei Bedienung) besteht. Dies ist jedoch nicht der Fall; die Varianten unterscheiden sich hinsichtlich der Wechseldauer nicht signifikant.

Bild 7.35 zeigt die Blickzuwendungszeiten auf den Touchscreen. Auch hier besteht kein signifikanter Unterschied zwischen den verschiedenen Varianten unabhängig vom Modus. Eine Analyse der Varianten hinsichtlich der Differenz zwischen der Bewegungszeit und der Blickzuwendung zeigt ebenfalls keinen signifikanten Unterschied zwischen den Varianten.

Fixationen In Bild 7.33 sind die Mediane des Beginns der ersten Fixation auf den Touchscreen dargestellt. Ein Vergleich der Bedienvarianten in den jeweiligen Modi zeigt jeweils keinen signifikanten Unterschied, d. h. die erste Anvisierung erfolgt in allen Bedienvarianten gleich schnell.

Aufgrund der kleinen Größe der Buttons war es in dieser Untersuchung nicht möglich, AOIs bezüglich der Buttons sinnvoll zu definieren. Da je-

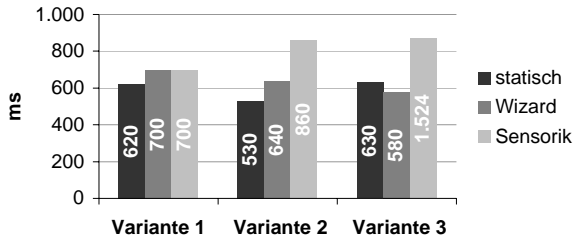


Bild 7.32: Blickzuwendung zum Touchscreen während der Bedienung.

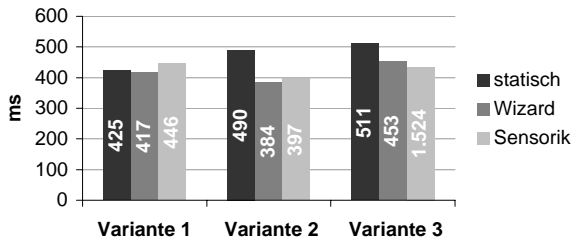


Bild 7.33: Beginn der ersten Fixation auf den Touchscreen während der Bedienung.

doch die Dauer einer Fixation Rückschlüsse auf die Intensität und Tiefe der Verarbeitung zulässt, erfolgte eine genauere Betrachtung der Dauer der ersten Fixation und der Dauer der letzten Fixation auf dem Display. Hierbei zeigte sich, dass die mediane Fixationsdauer bei der ersten Fixation bei 100 ms liegt, bei der letzten Fixation bei 360 ms. Dieser Unterschied ist hochsignifikant ($p < 0,001$, Wilcoxon-Test). Es ist also davon auszugehen, dass bei der letzten Fixation eine deutlich tiefere Verarbeitung stattfindet (das Ziel wurde gefunden), d. h. der Beginn der letzten Fixation ist als spätmöglicher Zeitpunkt der finalen Fixierung des Ziels zu sehen. Der Beginn der letzten Fixation ist in Bild 7.34 dargestellt. Bei der Betrachtung zeigt sich, dass dieser im Sensormodus wesentlich später liegt, was in Zusammenhang mit dem Wissen, dass die Adaption erst zu einem späteren Zeitpunkt vollständig erfolgt ist, erklärbar ist.

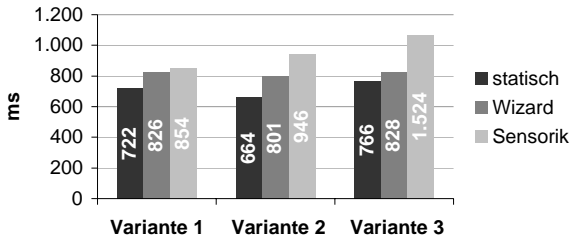


Bild 7.34: Beginn der letzten Fixation auf dem Display während der Bedienung.

Fixationshäufigkeit und -dauer Die Fixationshäufigkeit kann als Maß für die erforderlichen Verarbeitungsanforderungen dienen. Bei der Beschreibung von Suchprozessen deutet eine geringe Anzahl der Fixationen bis zum Erreichen des gewünschten Zieles auf ein ‚herausragendes‘ Ziel, eine trainierte Person oder eine einfache Suchaufgabe. Entsprechend ist eine hohe Zahl von Fixationen ein Hinweis auf eine Behinderung durch die Art der Repräsentation (Ebene der Informationsverarbeitung) oder des Layouts (Ebene der Informationsaufnahme).

Beim Vergleich der einzelnen Modi zeigt sich (siehe Bild 7.35), dass es im Sensorik-Modus im Vergleich zum statischen bzw. Wizard-Modus hochsignifikant mehr Fixationen gibt (alle $p < 0,001$). Beim Vergleich von statischem und Wizard-Modus zeichnet sich ein Vorteil für den statischen Modus mit weniger Fixationen ab ($p < 0,01$). Die Anzahl der Fixationen im Wizard-Modus in den adaptiven Varianten ist jedoch vergleichbar mit der Anzahl der Fixationen in der statischen Variante 1. Darüber hinaus ist die Anzahl der Fixationen im Wizard-Modus in Variante 3 mit Indikation tendenziell geringer als in Variante 2 ohne Indikation. Dies deutet darauf hin, dass aufgrund der Vorschau der Schaltflächen im Anzeigemodus eine bessere Orientierung ermöglicht wird.

Bezüglich der mittleren Fixationsdauern (siehe Bild 7.36) konnten jedoch keine signifikanten Unterschiede festgestellt werden.

7.4.5 Interpretation und Fazit

Nach dem von der *British Standards Institution* [30] veröffentlichten Richtlinien dokument zum Zwahlen-Diagramm [162] sollen Informatio-

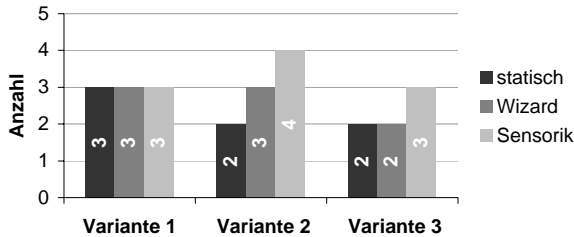


Bild 7.35: Mittlere Anzahl der Fixationen während der Bedienung.

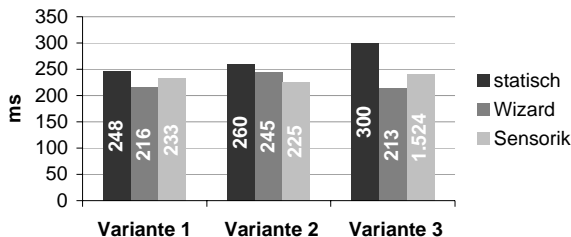


Bild 7.36: Median der Fixationsdauern auf den Touchscreen während der Bedienung.

nen, die auf einer graphischen Oberfläche präsentiert werden vom Fahrer mit einem ‚kurzen‘ Blick wahrgenommen werden können. Nach Ansicht von Zwahlen sind Blicke bis zu 1,2s noch akzeptabel, wenn die Anzahl der Blicke pro Informationseinheit den Wert drei nicht überschreitet. Das BSI weitet die Obergrenzen aufgrund weiterer Studien aus und legt als Obergrenze für die HMI-Blickzuwendung 2s sowie eine maximale Fixationsanzahl von vier fest (siehe hierzu auch [34, 143, 155, 156, 162]). Diese Festlegungen werden mit dem vorliegenden System sowohl im adaptiven als auch im statischen Modus in Bezug auf die Blickdauer und Fixationsanzahl erreicht.

Die Ergebnisse zeigen, dass ein adaptiver Touchscreen im Vergleich zu einem statischen Touchscreen keine wesentlichen Verschlechterungen in Bezug auf Blickabwendungszeiten mit sich bringt. Zur detaillierteren Bewertung werden im Folgenden die Ergebnisse der Studie mit den auf-

gestellten Hypothesen abgeglichen.

H1. Durch die Adaption der Zielflächen während der Bedienung ändern sich im Vergleich zu statischen Bedienvarianten die Blickabwendungszeiten von der Primäraufgabe nicht maßgeblich.

Diese Hypothese lässt sich bestätigen. Die Blickabwendungszeiten sind in hohem Maße mit den Bewegungszeiten korreliert. Da die Bewegungszeiten im Wizard-Modus und in der statischen Variante 1 nahezu gleich sind, sind auch die Blickabwendungszeiten vergleichbar.

H2. Der Einsatz von Indikatoren führt bei der Bedienung von adaptiven Zielen zu geringeren Blickabwendungszeiten von der Primäraufgabe.

Diese Hypothese konnte nicht bestätigt werden. Eine positive Wirkung von Indikatoren konnte in dieser Studie an keiner Stelle nachgewiesen werden.

H3. Der Einsatz von Indikatoren führt bei der Bedienung von adaptiven Zielen zu geringeren Fixationszeiten des adaptiven Touchdisplays.

Auch diese Hypothese kann nicht bestätigt werden. Die mittlere Fixationsdauer auf den Touchscreen ist im Wizard und statischen Modus in allen Varianten vergleichbar. Ein geringer Unterschied zeigt sich in der Anzahl der Fixationen: Diese liegt in Variante 2 ohne Indikation im Wizard und auch im Sensor-Modus höher als bei Variante 3. Dies kann in Hinblick auf eine einfachere Orientierung gedeutet werden. Eine subjektive Präferenz für Variante 2 äußerten die wenigsten Versuchspersonen, was als Bestätigung dieser Vermutung gewertet werden kann.

H4. Eine Blickzuwendung findet nicht nur während der Selektionsphase, sondern bereits in der ballistischen Phase statt.

Diese Hypothese kann bestätigt werden. Die erste Fixation auf das Touchdisplay startet im Median nach ca. 40 % der Zeit, was vor dem Beginn der Selektionsphase liegt, die nach 50 % der Zeit beginnt. Darüber hinaus findet fast während der gesamten Bewegung eine Blickabwendung vom LCT statt. Es ist denkbar, dass Information auch aus dem Augenwinkel für die weitere Bewegungsplanung genutzt wird. Der geringere Restabstand zum Ziel (siehe Kapitel 7.3) in der Variante mit Indikation ist daraus erklärbar.

7.5 Bedienbarkeit für Erstnutzer

7.5.1 Konzeption

Testsystem und Testumgebung Ziel dieser Benutzerstudie war es, herauszufinden, ob eine Erweiterung der Interaktionsmöglichkeiten eines Touchscreens in den dreidimensionalen Raum sich als sinnvoll für Erstnutzer erweist.

Aus Gründen der besseren Verständlichkeit für den Nutzer wurde in diesem Testsystem (analog zu dem in Kapitel 7.2 und Bild 7.5 vorgestellten System in Variante 2) eine durchgängige Benutzerführung integriert. In Anlehnung an heute übliche Fahrzeug-HMIs wurden alle simulierten Funktionen und Einstellungen jeweils einem Hauptmenü/einer Kategorie zugeordnet. Die Hauptmenüs wurden mit ‚Navigation‘, ‚Musik‘, ‚Auto‘, ‚Telefon‘ und ‚Setup‘ überschrieben. Folgende Schaltflächen wurden dabei annäherungsabhängig eingeblendet:

- **Navigation:** Standort speichern, Optionen, Stauumfahrung, Zielführung starten, Karte; in der Kartendarstellung: Vergrößerung der POIs (Points of Interest), Zurück-Taste;
- **Musik:** Titel vor und zurück, Play und Pause, TP Verkehrsfunk, Liste, Optionen;
- **Auto:** Einstellungen vornehmen, aktuelle Reiseinformationen;
- **Telefon:** Manuelle Wahl, SMS, SOS, Wahlwiederholung, Adressbuch, Optionen;
- **Setup:** Setup Navigation, Setup Audio, Setup Fahrzeug, Setup Telefon, Setup Allgemein.

Für die Gestaltung einer kraftfahrzeugähnlichen, aber dennoch möglichst ablenkungsfreien Umgebung, wurde das Testsystem in eine sog. Sitzkiste eingebaut. Die Fahrzeugkarosserie wurde hierzu aus Holz nachgebaut. In diese simulierte Karosserie wurden Seriensitze, ein Serienlenkrad und das Testsystem eingebaut. So konnten den Versuchspersonen ein reales Bild der Einbauposition und ein realistisches Bedingefühl gegeben werden.

Während Phasen, in denen die Versuchsperson das System frei bediente oder gestellte Aufgaben erfüllte, wurde eine Videoaufzeichnung durchgeführt. Dabei wurden die Bedienung sowie die Hand der Versuchsperson gefilmt. Kommentare und Anmerkungen der Versuchsperson wurden si-

multan aufgezeichnet. Eine Aufnahme des Gesichts war aus Datenschutzgründen nicht möglich.

Aufgabenbeschreibung Zunächst sollten die Probanden erste Erfahrungen mit dem System ohne Anleitung sammeln. Dabei sollten sie laut mitdenken, d. h. ihre Gedanken aussprechen. Der Versuchsleiter führte hierüber Protokoll. Spezielle Beachtung fand hierbei der Aspekt der intuitiven Bedienbarkeit. Im Anschluss an diese Erschließungsphase wurden die Probanden aufgefordert, im Audio-Menü ein anderes Lied auszuwählen.

Danach erklärte der Versuchsleiter kurz und stark vereinfacht die technische Funktionsweise: Der Touchscreen ist mit vier Sensoren ausgestattet, von denen jeder erfasst, wie weit die Hand vom Display entfernt ist. Es erfolgt eine Verrechnung der Sensorwerte. Daraus lässt sich die Position der Hand bestimmen.

Im Anschluss zeigte der Versuchsleiter verschiedene mögliche Funktionen. Allgemein aufgetretene Fragen der Versuchspersonen wurden protokolliert.

Das folgende Aufgabenset für die Versuchsteilnehmer umfasste mehrere Teilaufgaben:

- Aktuelle CD auf Pause setzen,
- in der Navigation die Kartendarstellung aufrufen,
- Titelliste der CD anzeigen lassen,
- Zielführung starten.

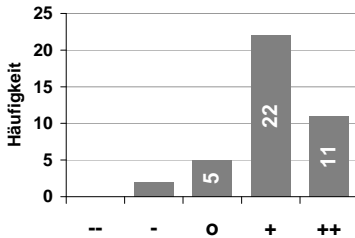
Messgrößen Bei allen Fragen wurden die erfolgreiche Bewältigung der Aufgabe oder evtl. auftretende Fehler protokolliert. Im Anschluss folgte eine Befragung der Probanden. Die abgefragten Kriterien wurden auf einer Ordinalskala von „sehr schlecht“ (– –), „schlecht“ (–), „neutral“ (0), „gut“ (+) bis „sehr gut“ (++) bewertet.

Stichprobenbeschreibung An der Untersuchung nahmen insgesamt 41 Personen teil. Bei den Teilnehmern handelte es sich ausschließlich um Fahrer der Golf-Klasse im Alter von 18 bis 56 Jahren. Das Durchschnittsalter der 30 teilnehmenden Männer lag bei 38,27 Jahren ($\sigma = 7,19$ Jahre), das Durchschnittsalter der 11 Frauen bei 40,18 Jahren ($\sigma = 10,15$ Jahre).

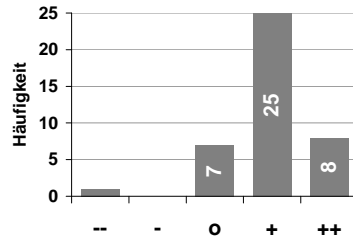
7.5.2 Ergebnisse

Die erste Aufgabe, die noch vor der Erklärung des Systems durch den Versuchsleiter gestellt wurde („Wählen Sie ein anderes Lied aus“), konnte von 38 Personen (93 %) ohne Probleme gelöst werden. Aus den Videoaufzeichnungen konnte ersehen werden, dass die meisten der Versuchsteilnehmer keine größeren Probleme bei der Bedienung hatten und sich zügig und selbstständig mit dem System vertraut machten.

Das nach der Erklärung durch den Versuchsleiter folgende Aufgabenset konnte dann von etwas mehr von 95 % der Versuchsteilnehmer vollständig gelöst werden, d. h. nur zwei Personen von 41 hatten noch Schwierigkeiten bei der Bedienung. Damit konnte ein Großteil der Probanden nach kurzer Einführung das System ohne Probleme bedienen. Aus Bild 7.37(a) wird ersichtlich, dass auch in dieser Studie eine überwiegende Zahl an Versuchspersonen vom Konzeptansatz des adaptiven Touchscreens überzeugt ist. Der Mann-Whitney U-Test [25] ($p = 0,146$) zu dieser Frage



(a) Bewertung des Infotainments.



(b) Bewertung der Funktion „Vergrößern von POIs bei Annäherung“.

Bild 7.37: Allgemeine Bewertung des Infotainments und der POI-Vergrößerung.

zeigt zwar keinen signifikanten Unterschied, allerdings ist tendenziell ersichtlich, dass Personen zwischen 30 und 40 Jahren das System im Allgemeinen besser bewerten als ältere Personen (41 bis 56 Jahre).

Die Funktion „Vergrößern von POIs in einer Karte bei Annäherung“ wurde fast durchweg positiv beurteilt. Bild 7.37(b) zeigt die Auswertung detaillierter.

Das Erscheinen des Sofortmenüs wurde sehr gut beurteilt, wobei der

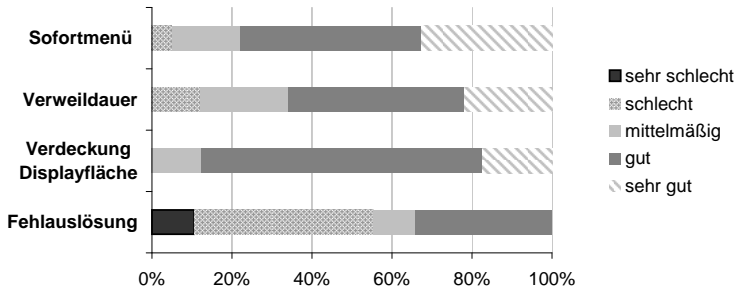


Bild 7.38: Bewertung des Sofortmenüs.

Mann-Whitney Test zeigt, dass Personen bis 40 Jahre diese Funktion wie auch schon das gesamte System tendenziell – jedoch nicht signifikant – besser beurteilen als Personen zwischen 41 und 56 Jahren ($p = 0,079$, $p > 0,05$). Der momentane Auslöseabstand von ca. 10 cm für das Sofortmenü und die POI-Vergrößerung wird von 83 % der Versuchsteilnehmer als ausreichend erachtet. Auch die Verweildauer des Sofortmenüs (3 s) wird vom Großteil mit „gut“ oder „sehr gut“ bewertet. Bei den Skeptikern zeigte sich die Tendenz zu einem schneller verschwindenden Sofortmenü. Die Verdeckung der Displayfläche durch das Sofortmenü stellt für die meisten Probanden kein Problem dar. Bild 7.38 zeigt die Auswertung im Detail.

7.5.3 Fazit

Auch Erstnutzer können das System ohne größere Probleme bedienen. Es wird schnell ein mit dem Systemmodell übereinstimmendes mentales Modell (siehe auch Kapitel 2.6.2) ausgebildet, das eine einfache Bedienung des Systems ermöglicht.

7.6 Zusammenfassung der Ergebnisse aus den einzelnen Studien

Die Ergebnisse der Studien zeigen, dass die Annäherungsfunktion überwiegend gut bis sehr gut bewertet wird. Im direkten Vergleich bevorzugen

beim Einsatz von kleinen Displays im Fahrzeug 72 % eine der adaptiven Varianten gegenüber einer statischen Bedienung. Prinzipiell wird eine pauschale Skalierung oder Einblendung von Schaltflächen bevorzugt, wie sich am Beispiel des Mediaplayers und der POI-Vergrößerung zeigte. Bei der Interaktion mit POIs in einer Karte zeigte sich jedoch kein klares Bild: Die ortsabhängige Skalierung wurde von 10 Fahrern bevorzugt, 18 Fahrer entschieden sich für die pauschale Skalierung. Es ist davon auszugehen, dass sich dieses Verhältnis im fahrenden Fahrzeug zugunsten der ortsabhängigen Skalierung verschiebt, da diese Variante eine stärkere Vergrößerung von Schaltflächen erlaubt als die pauschale Skalierung. Unter dem Einfluss von Fahrzeugvibrationen ist die Größe der Tasten meist ein subjektiv wichtigeres Kriterium als im stehenden Aufbau.

Die ortsabhängige Skalierung ist mit dem derzeitigen Aufbau und den verwendeten Algorithmen realisierbar. Der minimal erreichbare Durchmesser von Schaltflächen liegt in einer Ebene 2 cm vor dem Display bei 2 cm, in einer Ebene 8 cm vor dem Display im mittleren Displaybereich bei 2,5 cm, in den Randbereichen bei 5 cm. Die maximale Reichweite der Sensorik von 10 cm (nur Annäherungsdetektion) ist bei entsprechender Gestaltung der Anzeige- und Bedienoberfläche für den Einsatz im Fahrzeug ausreichend.

Messungen der Bewegungszeiten zum Treffen einer Taste beim adaptiven Touchinterface zeigten, dass bei Simulation einer optimalen Sensorik durch einen Wizard vergleichbare Bewegungszeiten wie beim Einsatz eines statischen Touchscreens erzielt wurden. Dies gilt sowohl für geübte als auch für ungeübte Nutzer. Bei Einsatz der derzeitigen Sensoren ergeben sich etwas höhere Bewegungszeiten, die aber dennoch absolut gesehen akzeptabel für den Einsatz im Fahrzeug sind. In allen Fällen wurden bei den adaptiven Varianten aufgrund der größeren Tasten wesentlich weniger Fehltreffer registriert als bei der statischen Variante.

Aus den linearisierten Bewegungsabläufen für geübte Nutzer konnten Parameter für die Spezifikation einer Sensorik abgeleitet werden, die die natürliche Bewegung möglichst wenig beeinflusst. Es wurde gezeigt, dass die Parameter prinzipiell mit der Verbausituation des adaptiven Touchscreen-Interfaces im Fahrzeug in Einklang stehen.

In allen Varianten und Modi konnte die Gültigkeit des *Optimized Initial Impulse Model* [98] in Bezug auf die Handbewegung nachgewiesen werden. Allerdings findet bereits in der ballistischen Phase eine Blickzuwendung zum Touchscreen-Interface statt. Darüber hinaus wurde gezeigt,

dass die Adaption keinen grundlegenden Einfluss auf die Blickbewegung des Nutzers im Vergleich zu einer statischen Variante hat. Vielmehr korreliert die Blickabwendung von der Straße stark mit der Bewegungszeit bei Bedienung des Touchinterfaces. Ziel muss es also sein, die Bewegungszeiten möglichst gering zu halten und möglichst wenig Interaktionsschritte zur Bedienung einer Funktion notwendig zu machen.

Dies kann einerseits durch eine größere Anzahl an Tasten auf der obersten Ebene (Direktzugriffe) und andererseits durch eine möglichst geringe Fehltrefferwahrscheinlichkeit (große Tastengröße) erreicht werden. Das adaptive Touchinterface ermöglicht eine wesentlich größere Darstellung von Tasten bei gleicher oder sogar mehr Informationsanzeige, als dies bei statischen Varianten der Fall ist, wodurch die Fehlerrate sinkt. Speziell bei Verwendung kleiner Display tritt dieser Vorteil zu Tage.

Neben dem direkten Vergleich einer statischen mit einer adaptiven Variante wurde der Einfluss von Indikatoren – kleine Vorschau-elemente für die Bedienung im Anzeigemodus – untersucht. In Hinblick auf Bewegungs- und Blickabwendungszeiten konnte sowohl mit der verwendeten Sensorik als auch bei Simulation einer optimalen Sensorik kein wesentlicher Unterschied festgestellt werden. Es ließ sich allenfalls bei geübten Nutzern eine leichte Tendenz zu sinkenden Bewegungszeiten beim Einsatz von Indikatoren feststellen. Auch bei der subjektiven Bewertung ergab sich keine klare Präferenz für oder gegen Indikatoren.

Der positive Effekt von Indikation tritt jedoch bei der Spezifikation des Sensorsystems zu Tage. Der Restabstand zum Ziel nach Ende der ballistischen Phase, der maßgeblich die minimale Reichweite der Sensoren beeinflusst, ist beim Einsatz von Indikatoren deutlich geringer. Dies ist wahrscheinlich auf die bessere Planbarkeit der Bewegung zurückzuführen.

8 Evaluierung der Fusion multimodaler Eingaben unter Einbeziehung der impliziten Modalität Annäherung

„Was wir wissen, ist ein Tropfen, was wir nicht wissen, ein Ozean.“
(Isaac Newton)

8.1 Basisuntersuchung zur subjektiven Bewertung von synergistisch- und alternierend-multimodalen Eingaben

In der vorliegenden Studie sollte die generelle subjektive Akzeptanz synergistisch- und alternierend-multimodaler Eingaben (Sprache und Touchscreen) im Fahrzeug evaluiert werden. Dies erfolgte einerseits am Beispiel der ‚ortsspezifischen Suche‘ und andererseits am Beispiel der ‚direkten Interaktion mit POIs‘.

8.1.1 Konzeption

Die Erhebung der Daten fand im Rahmen der in Kapitel 7.2 beschriebenen Studie statt.

Als Testsystem wurde der bereits beschriebene Aufbau (Kapitel 7.2) verwendet. Allerdings wurde für die Darstellung der Karte die gesamte Bildschirmgröße des 7"-Displays verwendet. Für die Interaktion standen jeweils verschiedene Möglichkeiten zur Verfügung. Es war sowohl eine rein haptische Bedienung als auch die Möglichkeit der multimodalen Bedienung implementiert. Die Spracheingabe erfolgte simuliert durch den Versuchsleiter und war daher über die gesamte Versuchsdauer möglich.

Die Versuchspersonen wurden zunächst gebeten, sich mit dem System im Stand vertraut zu machen und erst dann während der Fahrt im Simulator die Aufgaben, die durch den Versuchsleiter gestellt wurden, auszuführen.

8.1.2 Ergebnisse

Grundsätzlich wird die multimodale Interaktion per Sprache und Touchscreen sehr positiv gesehen. Es zeigte sich jedoch, dass die ‚ortsspezifische Suche‘ tendenziell besser beurteilt wird, als die direkte Interaktion mit POIs in einer Karte. Bild 8.1 zeigt die beiden Bewertungen im Vergleich. Auf die Aufforderung des Versuchsleiters – nach einer längeren Übungs-

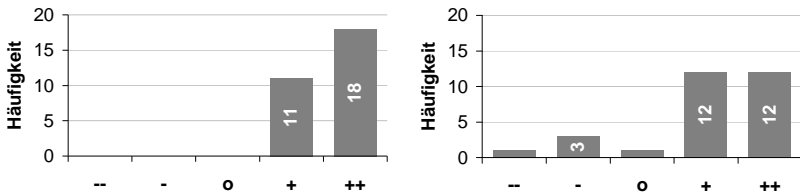
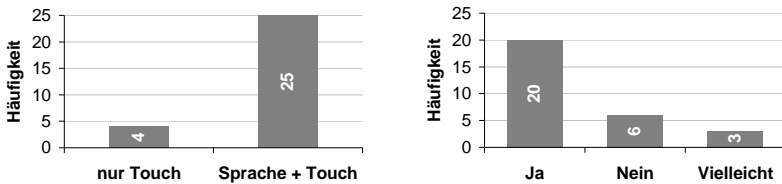


Bild 8.1: Bewertung der ‚ortsspezifischen Suche‘ und der ‚direkten POI-Interaktion‘ in Bezug auf Nützlichkeit.

phase – sich eine POI-Kategorie um einen gekennzeichneten Punkt in der Karte darstellen zu lassen, wählten weitaus mehr Personen eine multimodale Eingabemöglichkeit als die rein haptische Interaktion. Bild 8.2(a) zeigt das Ergebnis im Detail.



(a) Gewählte Variante zur Auswahl einer POI-Kategorie in der Kartendarstellung.

(b) Würden Sie die POI-Interaktion per *Point-to-Talk* auch in Ihrem Fahrzeug nutzen wollen?

Bild 8.2: ‚Ortsspezifische Suche‘ und ‚direkte POI-Interaktion‘.

Obwohl 24 Personen die POI-Interaktion nützlich oder sehr nützlich finden, würden diese jedoch nur 23 Personen in Ihrem Fahrzeug auch

nutzen wollen (Bild 8.2(b)).

8.1.3 Interpretation und Fazit

In Hinblick auf das Durchschnittsalter der Versuchspersonen und die durchschnittliche jährliche Fahrleistung, zeigen die Ergebnisse eine sehr positive Bewertung der Interaktionsverfahren. Oft angesprochen wurde eine gewisse Skepsis, ob sich diese Art der Bedienung im Fahrzeug überhaupt verwirklichen ließe. Darüber hinaus äußerten einige Versuchspersonen Bedenken, dass die ‚ortsspezifische Suche‘ zu sehr von der Fahraufgabe ablenken könnte, ein sicher ernst zu nehmender Einwand. Es ist jedoch anzumerken, dass bereits heute einige Eingaben, wie z. B. die Navigationszieleingabe in Fahrzeugen einiger Hersteller, nur im Stand getätigt werden können.

Eine Weiterentwicklung und weitere Evaluierung dieses Interaktionsverfahrens erscheint durchaus sinnvoll. Kapitel 8.2 zeigt die Ergebnisse der Studie, die mit dem funktionierenden System gewonnen wurden.

8.2 Evaluierung der Bedienbarkeit

8.2.1 Fragestellung und Zielsetzung der Studie

In der vorliegenden Studie sollte aus subjektiver und objektiver Sicht die Bedienbarkeit und Akzeptanz von *Point-to-Talk*-Szenarien sowie der ‚ortsspezifischen Suche‘ im Zusammenhang mit der Annäherungssensorik evaluiert werden. Als Vergleich wurde ein Bedienverfahren herangezogen, das Ähnlichkeiten mit dem heute serienmäßig vorhandenen *Push-to-Talk*-Verfahren (der Spracherkennung wird über eine Taste aktiviert) hat; dem Vergleichsverfahren liegt jedoch die Annahme zugrunde, dass bei Fahrzeugen der kleineren Klassen aus Kostengründen keine eigene *Push-to-Talk*-Taste verbaut werden kann. Daher erfolgt die Aktivierung des Spracherkenners durch einen längeren Druck auf den Touchscreen.

Für die Studie wurden folgende Hypothesen aufgestellt:

- **H1.** Die implizite Aktivierung des Spracherkenners über Annäherung bietet Vorteile gegenüber der expliziten Aktivierung des Spracherkenners hinsichtlich der Bedienzeiten.

- **H2.** Die implizite Aktivierung des Spracherkenners über Annäherung erhöht gegenüber der expliziten Aktivierung des Spracherkenners den Bedienkomfort.
- **H3.** H2 und H3 kommen besonders bei der ‚ortsspezifischen Suche‘ zum Tragen.
- **H4.** Die explizite Aktivierung des Spracherkenners führt zu weniger Bedienfehlern.

8.2.2 Konzeption

Testsystem Wie bereits angesprochen wurden für das Testsystem zwei verschiedene Verfahren zur Aktivierung des Spracherkenners implementiert und gegenübergestellt:

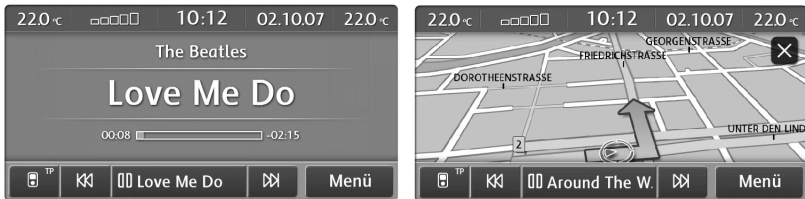
- **Variante 1: Implizite Aktivierung des Spracherkenners bei Annäherung an den Touchscreen**
Bei erkannter Annäherung wird der Spracherkennung aktiviert und ein Spracherkennung-Symbol eingeblendet, das dem Nutzer signalisiert, er könne jetzt per Sprache interagieren. Diese Variante wird im Folgenden mit *Point-to-Talk* (PTT) bezeichnet.
- **Variante 2: Explizite Aktivierung des Spracherkenners bei Longpress auf ein Element**
Nach erkanntem Longpress (400 ms) auf ein Interaktionselement wird der Spracherkennung aktiviert und das Spracherkennung-Symbol eingeblendet, das dem Nutzer die Möglichkeit der sprachlichen Interaktion mit dem aktivierten Interaktionselement signalisiert. Diese Variante wird im Folgenden mit *Talk-to-Object* (TTO) bezeichnet.

Als Kontext wurde zunächst der Mediaplayer gewählt, da er für die direkte Interaktion übersichtlich und einfach zu bedienen ist. Als Interaktionselemente standen hier der Titel und der Künstler zur Verfügung (siehe Bild 8.3(a)). Folgende Eingabemöglichkeiten existierten in Variante 1:

- Spezifikation von Titel oder Künstler per Sprache (Bsp.: „Beatles“).
- Spezifikation von Titel oder Künstler per Sprache mit dem Zusatz „Künstler“ oder „Titel“ (Bsp.: „Titel: Yesterday“).

- Spezifikation von Titel oder Künstler per Sprache und Einschränkung des Suchbereichs per Touch auf den Titel oder den Künstler, wobei die Reihenfolge zwischen Spracheingabe und Touch nicht maßgeblich war (Bsp.: „Dick Brave“ + Touch auf Künstler).

In Variante 2 war die Spracheingabe nur nach Longpress auf eines der Elemente (Titel oder Künstler) möglich.



(a) Mediaplayer mit den Interaktionselementen ‚Titel‘ und ‚Künstler‘.

(b) Navigationskarte ohne POIs.

Bild 8.3: Grundansichten von Mediaplayer und Navigationskarte.

Komplexere Interaktionen wurden anhand der ‚ortsspezifischen Suche‘ und der ‚direkten Interaktion mit POIs‘ im Szenario Navigation implementiert. Folgende Eingabemöglichkeiten standen dabei in Variante 1 zur Verfügung:

- Spezifikation einer POI-Interaktion („Hinfahren“, „Anrufen“ oder „Info“) per Sprache und Touch auf einen POI, wobei die Reihenfolge zwischen Spracheingabe und Touch nicht maßgeblich war.
- Spezifikation einer POI-Kategorie per Sprache (alle POIs dieser Kategorie wurden auf der Karte angezeigt) und Einschränken des Suchbereichs mittels Gebietspezifikation per Touchgestik.
- Spezifikation eines Gebietes auf dem Touchscreen und anschließende Angabe der POI-Kategorie per Sprache.

In Variante 2 wurde der Spracherkenner durch einen Longpress auf einen POI oder die eingezeichnete Region aktiviert. Eine Eingabe der POI-Kategorie ohne eingezeichnetes Gebiet war hier nicht möglich. Bild 8.3(b) zeigt die Grundansicht der Navigationskarte ohne POIs.

Der Bedienumfang wurde für diese Studie gegenüber bisher vorgestellten Konzepten stark reduziert, um eine Konzentration auf die alternierend- oder synergistisch-multimodale Bedienung mit Sprache und Touchscreen zu erreichen.

Für die Darstellung wurde ein prototypischer Aufbau des adaptiven Touchscreens (7" resistiver Touchscreen, 16:9), in eine abmessungsgetreuen, abstrakten Cockpit-Umgebung eingebaut. Die Darstellung der Interaktionsinhalte erfolgte jedoch nur auf einem verringerten Ausschnitt des 7"-Displays.

Die Messung der Bedienzeit vom Beginn bis zum Abschluss einer Aufgabe wurde durch Loslassen eines Tiptronic-Hebels (TT-Hebel) hinter dem Lenkrad gestartet und endete mit der Ausführung der entsprechenden Systemfunktion. Während der Bedienung wurden Zwischenzeiten bei Auslösen der Adaption sowie allen erkannten Teil-Eingaben protokolliert. Die Klassifikation, ob eine Aufgabe erfolgreich durchgeführt wurde, wurde vom Versuchsleiter vorgenommen.

Testumgebung Die Studie wurde im Fahrsimulator durchgeführt. Die Probanden wurden dabei aufgefordert, einem vorausfahrenden Fahrzeug in einem möglichst konstanten Abstand zu folgen.

Aufgabenbeschreibung Die Versuchsteilnehmer wurden zunächst gebeten, sich getrennt sowohl mit dem adaptiven Touchinterface als auch dem Fahrsimulator vertraut zu machen. Während des folgenden Versuchsteils mussten die Teilnehmer im Simulator einer Fahrzeug-Folgeaufgabe und der damit verbundenen ständigen Steuertätigkeit nachkommen. Während dessen wurden die Probanden instruiert, parallel zur Fahraufgabe für jede der beiden Varianten in separaten Bedienhandlungen so zügig wie möglich verschiedene Aufgaben auszuführen:

- Einen vorgegebenen Titel auswählen.
- Zu einem vorgegebenen Künstler wechseln.
- POIs einer vorgegebenen Kategorie in einem bestimmten Gebiet anzeigen lassen.
- Mit einem POI in einer vorgegebenen Art und Weise interagieren.

Im Anschluss wurde eine Befragung durchgeführt.

Versuchsablauf Nach einer kurzen Übungsphase im Simulator wurden die Versuchspersonen in zwei Gruppen aufgeteilt. Gruppe A begann mit Variante 1 der Interaktion (*Point-to-Talk*), Gruppe B mit Variante 2 (*Talk-to-Object*). Zunächst beschäftigten sich die Versuchspersonen jeweils mit dem Kontext Mediaplayer und erst im Anschluss mit dem Kontext Navigation.

In jedem Versuchsblock bestand für die Versuchspersonen zunächst die Möglichkeit, sich mit dem System vertraut zu machen. Erst im Folgenden wurden dann Bedienzeiten für die einzelnen Aufgaben gemessen. Im Anschluss erfolgte die Befragung. Die Reihenfolge der Aufgaben in den einzelnen Blöcken war genauso wie die Zuteilung zu den Gruppen vollständig randomisiert.

Stichprobenbeschreibung An der Studie nahmen insgesamt 19 Personen teil, davon waren 14 männlich und 5 weiblich. Das Durchschnittsalter lag bei 28,5 Jahren ($\sigma = 5,5$ Jahre).

Messgrößen und Datenaufbereitung Die abgefragten subjektiven Kriterien wurden auf einer Ordinalskala von „sehr schlecht“ (– –), „schlecht“ (–), „neutral“ (0), „gut“ (+) bis „sehr gut“ (++) bewertet. Darüber hinaus wurden die Bedienzeiten für die jeweilige Aufgabe protokolliert. Fehlerhaft aufgezeichnete Datensätze, sowie Datensätze, bei denen ein Nutzerfehler vorlag, gingen nicht in die quantitative Analyse der Bedienzeiten mit ein. Darüber hinaus wurden mittels Box-Whisker-Plot extreme Ausreißer entfernt.

Für einige Testverfahren wurden die Skalen der subjektiven Bewertung in Zahlenwerte umgewandelt. Die o. g. Ordinalskala wurde dabei mit –2 bis 2 belegt, die Antworten „ja“ mit 1, „nein“ mit –1 und „vielleicht“ mit Null.

8.2.3 Ergebnisse

Reihenfolgeeffekt Zunächst sollte ermittelt werden, ob es einen Reihenfolge-Effekt zwischen den beiden Bedienverfahren PTT und TTO gab. Es zeigte sich, dass bei *Point-to-Talk* kein Reihenfolge-Effekt festzustellen war, hingegen bei *Talk-to-Object* Gruppe A in zwei Aufgabentypen (‘Titel wählen‘ und ‘direkte POI-Interaktion‘) signifikant schneller bediente als Gruppe B, die mit dieser Interaktionsart begonnen hatte. Um die

Vergleichbarkeit der Daten zu gewährleisten, wurden für die folgenden Auswertungen für *Talk-to-Object* nur die Daten von Gruppe B ausgewertet.

Bedienzeiten Die weitere Betrachtung der Bedienzeiten ergab, dass nicht von einer Normalverteilung ausgegangen werden kann. Daher wurde – genauso wie auch schon in Kapitel 7.4 – auf eine ordinale Betrachtung zurückgegriffen. Bild 8.4 zeigt die Mediane der Bedienzeiten für die einzelnen Aufgaben sowie die 95 %-Konfidenzintervalle. Die Unterschiede

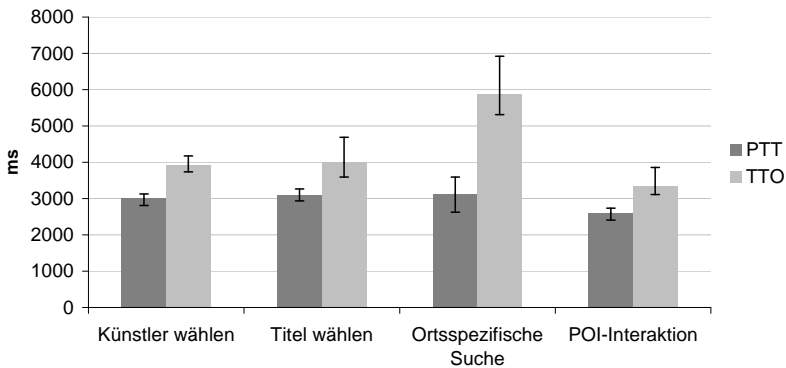


Bild 8.4: Mediane der Bedienzeiten im *Point-to-Talk*- und *Talk-to-Object*-Modus.

zwischen PTT und TTO sind in allen Aufgabenvarianten signifikant. Besonders auffällig ist der Unterschied bei der ‚ortsspezifischen Suche‘. Dies ist u. U. auf die dort im TTO-Modus komplexere Interaktion zurückzuführen. Zunächst musste eine Region eingezeichnet werden und erst im Folgenden konnte durch Longpress auf die Region der Spracherkenner aktiviert werden, um die POI-Kategorie einzusprechen. Im PTT-Modus war es möglich, schon während des Zeichnens die POI-Kategorie per Sprache zu spezifizieren. Die Unterschiede zwischen ‚Künstler wählen‘ und ‚Titel wählen‘ sind – wie zu erwarten – sowohl bei TTO als auch bei PTT nicht signifikant.

Die Bedienzeiten insgesamt sind relativ lang. Je nachdem, wie stark sich die tertiären Bedienhandlungen mit der primären Aufgabe Fahr-

zeugführung überlagern, führen diese zu einer Überlastung der kognitiven Ressourcen des Fahrers und beeinflussen die Leistungsfähigkeit. In verschiedenen Studien wurde jedoch beobachtet, dass Fahrer die Entscheidung eine Nebenaufgabe auszuführen von der Bewertung des situativen Umfelds abhängig machen und u. U., bevor sie eine Nebenaufgabe ausführen, entsprechend der Verkehrssituation die Geschwindigkeit verringern oder den Abstand vergrößern [118, 122, 146]. Aufgrund der vorgegeben Aufgabe („Folgen Sie dem vorausfahrenden Fahrzeug in möglichst konstantem Abstand“) war dies jedoch nur teilweise möglich. Im realen Einsatz könnten sich die Bedienzeiten daher verringern.

Das von der *Society of Automotive Engineers* veröffentlichte Verfahren zur Bewertung von Fahrerinformationssystemen [65] gibt als Obergrenze für die Gesamtbedienzeit einer Funktion im Stand 15 s an. Dieses Kriterium wird auch bei der ‚ortsspezifischen Suche‘ erfüllt.

Subjektive Bewertung Im Allgemeinen wird die Art der Bedienung, unabhängig ob PTT oder TTO, sehr positiv beurteilt. Bild 8.5 zeigt die Antworten auf die Frage, ob die Versuchspersonen diese Interaktionsmöglichkeit auch in ihrem Fahrzeug nutzen wollten.

Es zeigt sich allerdings auch, dass gerade die ‚ortsspezifische Suche‘, die auch die längste Bedienzeit im TTO-Modus hat, hier am schlechtesten beurteilt wird. Der Unterschied zwischen den Antworten zu PTT und TTO ist hier signifikant. Betrachtet man die Frage, wie komfortabel die

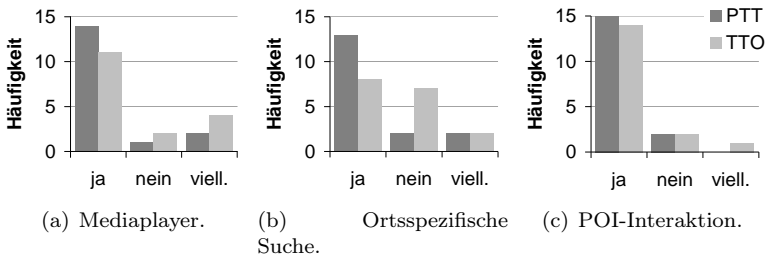


Bild 8.5: Würden Sie diese Interaktionsmöglichkeit auch in Ihrem Fahrzeug nutzen wollen?

Bedienung empfunden wird, wird PTT ebenfalls positiver beurteilt. Hier

ist hier der Unterschied in allen drei Szenarien signifikant. Bild 8.6 zeigt in Box-Plots die Aussagen der Versuchsteilnehmer zu dieser Frage.

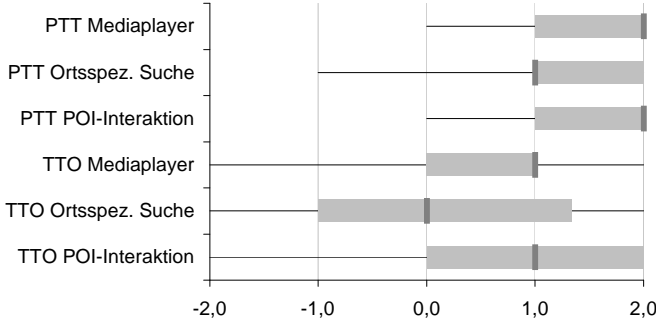
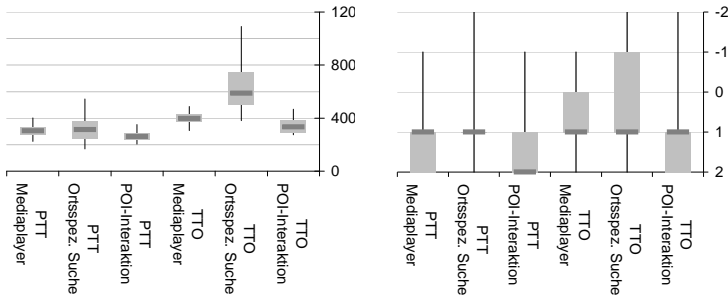


Bild 8.6: Box-Plots zur Frage: „Wie komfortabel war für Sie die Bedienung?“.

Bild 8.7 zeigt sowohl die gemessenen Bedienzeiten im Box-Plot als auch die Schätzungen der Versuchsteilnehmer auf einer Skala von -2 bis 2 . Die Bediengeschwindigkeit wird von den meisten Versuchspersonen im Vergleich zwischen den Aufgaben realistisch eingeschätzt. Die ‚ortsspezi-



(a) Gemessene Bedienzeiten.

(b) „Wie schnell konnten Sie die Aufgabe lösen?“

Bild 8.7: Vergleich der Bedienzeiten mit der subjektiven Einschätzung.

fische Suche‘, die in beiden Varianten jeweils die längsten Bedienzeiten

aufweist, wird auch in der subjektiven Befragung als solche eingeschätzt.

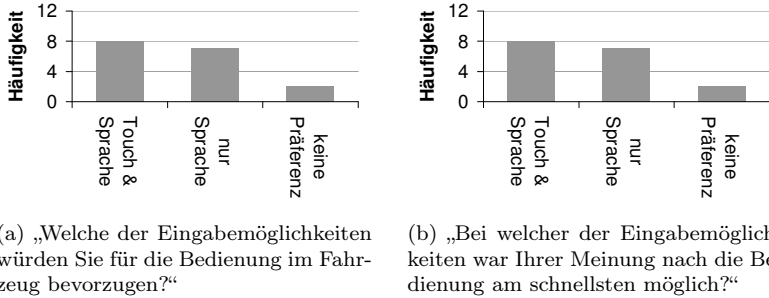


Bild 8.8: Gegenüberstellung der Bedienmöglichkeiten ‚Sprache‘ und ‚Touch und Sprache‘.

Betrachtet man die verschiedenen Eingabemöglichkeiten im Kontext Mediaplayer PTT (Bedienung nur per Sprache oder per Touch und Sprache) zeigt sich, dass die Präferenzen hier tendenziell ausgeglichen sind (siehe Bild 8.8). Während 7 Personen die Option nur per Sprache zu bedienen bevorzugen, entscheiden sich 8 Personen für die kombinierte Bedienung aus Sprache und Touch. Ein ähnliches Bild zeigte sich auf die Frage, bei welchem Verfahren die Bedienung schneller möglich wäre. Die Auswertung der Datensätze aus der Aufgabenbearbeitung zeigte, dass 59 Mal die Bedienung rein sprachlich erfolgte und 42 Mal in der Kombination mit Touch auf den Touchscreen. In den Bedienzeiten zeigt sich hier kein signifikanter Unterschied.

Fehlerrate Betrachtet man die Fehlerrate ER in den einzelnen Aufgaben (siehe Bild 8.9) zeigt sich, dass die Fehlerrate im Szenario Mediaplayer in beiden Varianten am größten ist. Hier fallen v. a. Fehler des Spracherkenners (ASR), insbesondere ‚ASR-Timeout‘-Fehler und ‚ASR-Fehler‘ ins Gewicht. Ein ‚ASR-Timeout‘-Fehler wurde dann protokolliert, wenn der Spracherkennung geschlossen wurde, bevor der Nutzer eine Spracheingabe beendet oder sogar begonnen hatte. Aufgrund des hier verwendeten Spracherkenners konnte der Beginn einer Spracheingabe nicht festgestellt werden, d. h. in einigen Fällen wurde der Spracher-

kenner geschlossen, obwohl die Nutzereingabe per Sprache noch nicht beendet war. Voraussetzung hierfür war allerdings, dass der Nutzer im PTT-Modus seine Hand nicht mehr vor dem Display hatte. Der Timeout des Spracherkenners wurde im PTT-Modus ab dem Verlassen des Detektionsbereichs der Annäherungssensorik gemessen, im TTO-Modus ab dem letzten erkannten Druck auf dem Touchscreen. Die etwas geringere Rate an ASR-Timeout-Fehlern im PTT-Modus ist wahrscheinlich auf diesen Umstand zurückzuführen.

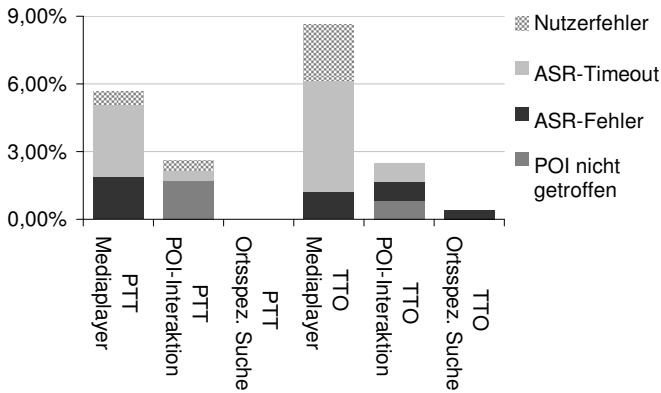


Bild 8.9: Fehlerrate bei der Bedienung in den einzelnen Szenarien.

Ein ASR-Fehler wurde dann protokolliert, wenn der Spracherkennung die Spracheingabe falsch oder nicht verstanden hatte, sodass auch durch geschickte Informationsfusion im Modalitätsmanager die Eingabe nicht korrekt ausgeführt werden konnte (Bsp.: es wurde „Beatles“ statt „Dick Brave“ erkannt). Die größere Anzahl an ASR-Fehlern im Szenario Mediaplayer als in den anderen Szenarien ist auf die größere dort geladene Grammatik zurückzuführen.

Die Fehlerrate bei der POI-Interaktion ist im Vergleich zum Mediaplayer in beiden Varianten insgesamt wesentlich geringer, obwohl hier die Komplexität der Interaktion und die Bedienzeiten vergleichbar sind. Allerdings wurden in allen Fällen zunächst Aufgaben zum Mediaplayer gestellt und erst in der Folge zur POI-Interaktion. Hier kann also von einem Lerneffekt ausgegangen werden, der sich zwar nicht in der Bedien-

zeit, jedoch deutlich in der Fehlerrate niederschlägt.

Es zeigt sich jedoch auch, dass die Fehlerrate im *Talk-to-Object*-Mediaplayer wesentlich höher ist als im *Point-to-Talk*-Modus, was auch aufgrund der großen Anzahl an Nutzerfehlern darauf hindeutet, dass die Bedienung im PTT-Modus einfacher als im TTO-Modus ist. Die wenigsten Fehler treten bei der ‚ortsspezifischen Suche‘ auf.

Vergleich der beiden Verfahren Die Abschlussbefragung am Schluss des Versuchs zielte darauf ab, die bevorzugte Bedienvariante (*Point-to-Talk* oder *Talk-to-Object*) klarer zu ermitteln. Bild 8.10 zeigt die Fragen „In welcher Variante passieren weniger Bedienfehler?“, „In welcher Variante ist die Bedienung schneller möglich?“, „In welcher Variante ist die Bedienung komfortabler?“ und „Welche Variante hat Ihnen besser gefallen?“ im Vergleich. Bei den letzten drei Fragen bekommt die implizite

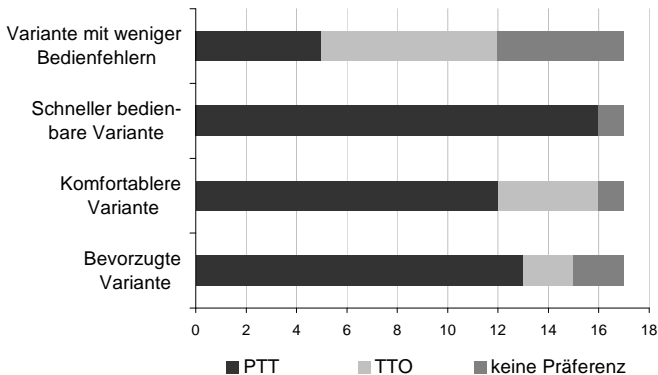


Bild 8.10: Subjektiver Vergleich der beiden Bedienvarianten PTT und TTO.

Aktivierung des Spracherkenners (PTT) weitaus mehr Nennungen als die explizite Aktivierung. Nur bei der Frage nach Fehlerwahrscheinlichkeiten zeigt sich ein eher gemischtes Bild.

8.2.4 Interpretation und Fazit

Die Ergebnisse zeigen, dass die implizite Aktivierung des Spracherkenners einerseits besser bewertet wird als die explizite Aktivierung und anderer-

seits auch zu geringeren Bedienzeiten führt. Zur detaillierten Bewertung werden im Folgenden die Ergebnisse der Studie mit den aufgestellten Hypothesen verglichen.

H1. Die implizite Aktivierung des Spracherkenners über Annäherung bietet Vorteile gegenüber der expliziten Aktivierung des Spracherkenners hinsichtlich der Bedienzeiten.

Die Hypothese kann voll bestätigt werden. Die Bedienzeiten für die einzelnen Aufgaben liegen im PTT-Modus signifikant niedriger als im TTO-Modus. Insgesamt sind die Bedienzeiten jedoch relativ lang. Dies ist auch darauf zurückzuführen, dass sich die meisten Personen zunächst an das Display annähern und erst nach einiger Zeit, wenn sie sicher sind, dass der Spracherkenner aktiv ist, auch wirklich sprechen. Im PTT-Modus entsteht dieser Zeitverzug noch vor der Touchscreen-Interaktion, im TTO-Modus während oder nach dem Longpress. In folgenden Studien wäre die Auswirkung auf die Blickabwendung zu prüfen.

Des Weiteren ist hier anzumerken, dass ein Reihenfolgeeffekt der Gruppen A und B nur bei der expliziten Aktivierung des Spracherkenners vorlag. Dies kann daraufhin gedeutet werden, dass die explizite Aktivierung der komplexere Bedienvorgang ist, der erst erlernt werden muss. PTT in Kombination mit Touch auf das Display kann als eine Übungsvorstufe für TTO betrachtet werden. Die Bedienung im Falle der impliziten Aktivierung des Spracherkenners ist hingegen so einfach, dass dort kein Reihenfolgeeffekt auftritt.

H2. Die implizite Aktivierung des Spracherkenners über Annäherung erhöht gegenüber der expliziten Aktivierung des Spracherkenners den Bedienkomfort.

Diese Hypothese lässt sich ebenfalls vollständig belegen. Der Mann-Whitney-Test zeigt, dass die Bewertung in Hinblick auf Komfort in allen Fällen für PTT signifikant besser ausfällt als für den TTO-Modus.

H3. H2 und H3 kommen besonders bei der ‚ortsspezifischen Suche‘ zum Tragen.

Auch diese Hypothese lässt sich vollständig bestätigen. Der Unterschied in den Bedienzeiten ist wahrscheinlich aufgrund des komplexen Interaktionsverfahrens im TTO-Modus bei der ‚ortsspezifischen Suche‘ (Region einzeichnen, Region per Longpress aktivieren, Sprechen) am deutlichsten. Hier zeigt sich besonders der Vorteil einer synergistisch-multimodalen Eingabe.

H4. Die explizite Aktivierung des Spracherkenners führt zu

weniger Bedienfehlern.

Diese Hypothese kann nur teilweise belegt werden. Auf die Frage, in welcher Bedienvariante weniger Bedienfehler entstehen, verschiebt sich das Bild im Gegensatz zu den anderen subjektiv bewerteten Fragen am meisten in Richtung TTO. Aus den gewonnenen quantitativen Daten ergab sich jedoch ein dazu gegensätzliches Bild. Evtl. könnte eine Verlängerung des Timeouts für den Spracherkenner die Fehlerrate maßgeblich nach unten korrigieren.

Abschließend lässt sich festhalten, dass *Point-to-Talk* mit der impliziten Aktivierung des Spracherkenners wesentlich besser für die Interaktion im Fahrzeug geeignet ist als *Talk-to-Object* mit der expliziten Aktivierung. Voraussetzung für diese Aussage ist allerdings die Annahme, dass keine *Push-to-Talk*-Taste zur Verfügung steht, mit der z.B. über eine Lenkrad-Taste der Spracherkenner aktiviert werden kann.

9 Zusammenfassung und Ausblick

„Aus den Steinen, die einem in den Weg gelegt werden, lassen sich Treppen bis in den Himmel bauen.“

(Erich Kästner)

Ausgehend von der ständig zunehmenden Anzahl an Funktionen und Bedienmöglichkeiten im Fahrzeug sollte ein umfassendes und skalierbares Modell entwickelt, technisch umgesetzt und evaluiert werden, das eine einfachere Interaktion mit dem Fahrzeug unterstützt und sich dabei vorteilhaft an Humankommunikationsprinzipien – also Kommunikationsprinzipien, die der Nutzer aus der Humankommunikation kennt – orientiert. Einerseits mussten dabei speziell automobiltechnische Aspekte wie die Integration der wachsenden Zahl an Funktionen in limitiertem Raum und die Reduktion der Bedienkomplexität adressiert werden. Andererseits mussten aber auch eine möglichst geringe Ablenkung von der Fahraufgabe sichergestellt sowie ergonomische Aspekte beachtet werden.

Die Arbeit gliedert sich hierzu in mehrere Bereiche. Zunächst wurden die theoretischen Grundlagen zur Entwicklung eines solchen Systems untersucht. Besonderer Schwerpunkt lag dabei auf der Beachtung impliziter Kommunikationsanteile sowie multimodaler Bedienung.

Das im Anschluss an die Analysephase entwickelte Dialogmodell erweitert bestehende Modelle, und enthält zusätzliche Module zur separaten Auswertung von expliziten und impliziten Eingaben. Darüber hinaus bietet die Architektur erweiterte Eingabemöglichkeiten (Gestik, Sprache und Blickbewegung) und bindet Kontextdaten wie den Fahrzustand und die Fahrsituation in das Modell ein.

Im nächsten Schritt erfolgte eine prototypische (Teil-)Implementierung des Modells, bei der der Schwerpunkt auf der impliziten Auswertung der Annäherung einer Hand an ein Mittelkonsolendisplay gelegt wurde. Die Grundphilosophie des hier verfolgten Ansatzes ist die Interpretation der Annäherung einer Hand als Bedienintention. Sobald eine Bedienintention erkannt wird, wechselt die Anzeige des Displays von einem Anzeige- in einen Bedienmodus: Schalt- oder Bedienflächen werden bei Annäherung

einer Hand vergrößert oder überhaupt erst eingeblendet. Durch die Anbindung eines Spracherkenners wurde das adaptive, annäherungssensitive Touchscreen-Interface zu einer multimodalen Bedieneinheit erweitert. Für die multimodale Interaktion wurde die Annäherung dahingehend interpretiert, dass der Spracherkenner aktiviert wird, sobald eine Bedienintention erkannt ist. Darauf aufbauend wurden verschiedene Szenarien implementiert, die u. a. synergistisch- und alternierend-multimodale Eingaben erlauben. Spezielles Augenmerk lag dabei auf der ‚ortsspezifischen Suche‘ sowie der ‚direkten POI-Interaktion‘.

Der Schwerpunkt bei der Umsetzung des Modells lag auf der Entwicklung eines Algorithmus zur dreidimensionalen Erkennung der Hand oder eines Bedienfingers unter Berücksichtigung der restriktiven Randbedingungen für den Einsatz im Fahrzeug: berührungslose Detektion, Echtzeitfähigkeit, hohe Robustheit, designneutraler Verbau sowie nicht zuletzt Wirtschaftlichkeit in Bezug auf Teilekosten und Rechenleistung. Im Gegensatz zu vielen anderen Ansätzen, die meist videobasiert arbeiten, baut der hier entwickelte Ansatz auf einer kapazitiven Sensor-Lösung auf: Aus vier Sensoren, die an den Ecken des Displays angebracht sind, wird mittels einer Maximum-Likelihood-Schätzung die Position der Hand geschätzt. Eine Überprüfung des Algorithmus zeigt, dass sich in einem Abstand von 4 cm ein Cursor auf dem Bildschirm noch immer so steuern lässt, dass die Abweichung vom Schnittpunkt eines Lots von der Fingerspitze auf die Displayoberfläche maximal 1 cm beträgt. Eine allgemeine Detektion der Annäherung kann jedoch schon bei Abständen von über 10 cm erreicht werden.

Des Weiteren wurde ein Verfahren entwickelt und implementiert, das alternierend- oder synergistisch-multimodale Eingaben auf Basis einer impliziten Aktivierung eines Spracherkenners und unter Einbeziehung von Touchgesten erlaubt.

Ein weiterer Schwerpunkt war die Evaluierung des Gesamtsystems in Bezug auf ergonomische und subjektive Kriterien. Dazu wurden mehrere Nutzerstudien durchgeführt. Es zeigte sich, dass sich Bewegungszeiten und Blickabwendungszeiten bei der Bedienung des adaptiven Touchscreens nur wenig von denen eines statischen Touchscreens unterscheiden. Einen nicht zu vernachlässigenden Einfluss hat hierbei allerdings die grafische Darstellung und dabei v. a. der Einsatz von Indikatoren. Indikatoren sind aber auch maßgeblich für die Festlegung des optimalen Detektionsbereichs, damit eine durchgängige Bedienbewegung auch

bei adaptiven Schaltflächen nicht gestört wird. Voraussetzung für vergleichbare Bedienzeiten im statischen und adaptiven Modus ist also der Einsatz einer geeigneten Sensorik zur impliziten Annäherungsdetektion. Diese konnte in dieser Arbeit zwar nicht aufgebaut werden, aber es konnten genaue Spezifikationen abgeleitet werden. Darüber hinaus konnte die Gültigkeit von gängigen Bewegungsmodellen für die Bedienung des adaptiven Touchscreens im Fahrzeug nachgewiesen werden. Das adaptive Touchscreen-Interface ist weitgehend intuitiv bedienbar und findet eine hohe Akzeptanz, wie die subjektiven Bewertungen zeigen.

Bei der Evaluierung der multimodalen Konzepte zeigte sich, dass die implizite Aktivierung des Spracherkenners auf Annäherung deutliche Vorteile gegenüber einem Konzept bietet, bei dem der Spracherkennung explizit über den Touchscreen aktiviert werden muss. Sowohl die subjektive Beurteilung als auch die objektiven Kriterien (Bedienzeiten) fallen deutlich positiver aus.

Abschließend betrachtet liefert das vorgestellte System einen grundlegenden Beitrag hin zu einer intuitiveren, natürlicheren Mensch-Maschine-Kommunikation im Fahrzeug. Die adaptive Vergrößerung oder Einblendung von Schaltflächen ermöglicht erst die Nutzung von kleineren Displays als Touchscreens. Neue Arten der multimodalen Integration werden durch die implizite Aktivierung des Spracherkenners eröffnet. Aufgrund der geringen Systemkosten und der Tatsache, dass kleine Displays als Touchscreens verwendet werden können, trägt das System dazu bei, bei der steigenden Anzahl an Fahrerinformations- und Fahrerassistenzfunktionen, komplexere Interaktionsmöglichkeiten auch in Fahrzeugen der kleineren Klassen zu implementieren. Der Erfolg des Systems wird jedoch maßgeblich davon abhängen, ob die Modalitäten zusammen mit den expandierenden Schaltflächen in ein überzeugendes Gesamtbedienkonzept integriert werden können.

Tabellenverzeichnis

2.1	Vergleich von adaptiven und adaptierbaren Systemen nach [62].	24
7.1	Startwerte: Radien der Kreise in den Ebenen 0 cm, 2 cm, 4 cm, 6 cm und 8 cm zu Beginn der Messung in mm. . . .	89
7.2	Altersverteilung der Versuchsteilnehmer.	95
7.3	Charakteristische Werte für das <i>Worst-Case</i> -Szenario zur Festlegung der Sensorparameter.	115
7.4	Charakteristische Werte aus den Mittelwerten (zur Festlegung der Sensorparameter).	115
7.5	Minimal erforderliche Reichweite der Sensoren vom Ziel aus betrachtet für verschiedene Adaptionen für die einzelnen Varianten, ermittelt aus den Wizard-Messungen unter Vernachlässigung von T_S	119
7.6	Minimal erforderliche Reichweite der Sensoren vom Ziel aus betrachtet für verschiedene Adaptionen für die einzelnen Varianten, ermittelt aus den Wizard-Messungen mit $T_S = 100$ ms.	119

Abbildungsverzeichnis

1.1	Anzahl der vom Fahrer bedienbaren Funktionen in einem Fahrzeug der Premium-Klasse (Quelle: Volkswagen). . . .	2
2.1	Interdisziplinärer Charakter der Mensch-Maschine-Kommunikation [104].	7
2.2	Struktur eines Mensch-Maschine-Systems nach [61] und [104].	8
2.3	Das Spektrum der Mensch-Maschine-Schnittstelle [85]. . .	10
2.4	Kommunikationsmodell nach Shannon.	10
2.5	Intentionsbasiertes Kommunikationsmodell nach Kim [81].	12
2.6	Potenzial Multimodalität [104].	16
2.7	Klassifikationsschema für multimodale Systeme nach [104] und [106].	19
2.8	Omnibusmodell: Messen als Regelkreis.	21
2.9	Abstraktionsebenen bei der Informationsfusion.	22
2.10	Dimensionen eines Benutzermodells [79].	25
2.11	Ergonomie der Mensch-System-Interaktion: Grundsätze der Dialoggestaltung nach ISO 9241 [48].	27
2.12	Gegenüberstellung von zwei marktüblichen <i>High-End</i> -Anzeige- und Bediensystemen in der Mittelkonsole.	30
2.13	Gegenüberstellung von zwei marktüblichen Basisausstattungen für Anzeige- und Bediensysteme in der Mittelkonsole.	32
3.1	Architektur der erweiterten Benutzerschnittstellen im Kfz.	42
3.2	Erweiterte Eingabemöglichkeiten auf dem Touchscreen. .	43
3.3	Architektur des implementierten Mensch-Maschine-Systems.	46
4.1	Schematischer Aufbau eines Kondensators [22].	51
4.2	Die menschliche Hand und eine Metallplatte bilden einen Kondensator.	53

4.3	Verlauf der dielektrischen Konstante ϵ für Muskelgewebe in Abhängigkeit der Frequenz. Nach Daten von [132].	54
4.4	Ersatzschaltbild für eine kapazitive Detektoranordnung mit zwei Elektroden und einem leitfähigen Objekt oder Körper [92].	54
4.5	Drei Modi zur kapazitiven Messung [140].	55
4.6	Mit steigender Distanz zwischen den Kondensatorplatten steigt die Streukapazität. Dieses Verhalten beschränkt die Reichweite kapazitiver Sensoren.	56
4.7	Prinzipschaltbild kapazitiver Sensorik im Übertragungsmodus nach [160].	58
4.8	Rauschquellen im System.	59
4.9	Kennlinie des Sensors.	59
4.10	Likelihoodfunktion $p(m r)$ für $m = 11000$	61
4.11	Likelihoodfunktion $p(m x, y)$ für $m = 11000$	62
4.12	Likelihoodfunktionen für drei Sensoren an den Koordinaten $(0 0)$, $(10 0)$ und $(0 14)$	63
4.13	Fusionierte Likelihoodfunktion.	64
6.1	Schematischer Aufbau des Gesamtsystems.	74
6.2	Grundsätzliches Bedienprinzip des adaptiven Touchscreens.	75
6.3	Situative Adaption von Bedienflächen.	77
6.4	Einblendung von Indikatoren sowie Expansion von Bedienflächen.	78
6.5	Ortsspezifische Suche von POIs in einer Kartendarstellung.	79
6.6	Direkte Interaktion mit POIs.	80
6.7	Links: Microcontrollerplatine mit vier Sensoren. Rechts: Sitzmatte zur Signalübertragung auf den Körper.	82
6.8	Anordnung der Sensoren um das Display mit Bemaßung.	83
6.9	Kennlinie des Sensors und benutzerspezifische Verschiebung.	83
6.10	Einfluss der benutzerspezifischen Kennlinienanpassung: Testmuster (links) und rekonstruiertes Muster vor Kalibrierung (Mitte) und nach Kalibrierung (rechts).	84
7.1	Ablauf der Messung im Flussdiagramm.	87
7.2	Displayinhalte zu verschiedenen Zeitpunkten der Messung.	88
7.3	Minimaler durchschnittlich noch getroffener Kreisradius in cm.	90

7.4	Minimaler durchschnittlich noch getroffener Kreisradius in cm für eine Person im linken Randbereich, der Mitte und dem rechten Randbereich.	91
7.5	Übersicht über die Bedienvarianten – die hintere Darstellung zeigt jeweils den Anzeigemodus, die vordere Darstellung zeigt die Einblendung der Bedienflächen bei Annäherung der Hand.	94
7.6	Varianten der POI-Skalierung: keine Skalierung, pauschale Skalierung und positionsabhängige Skalierung.	94
7.7	Wie gefällt Ihnen die Annäherungsfunktion allgemein? . .	96
7.8	Wie wichtig ist es für Sie, die Bedioptionen schon vor der Bedienung sehen zu können?	96
7.9	Bewertung der einzelnen Varianten.	97
7.10	Welche der Varianten würden Sie insgesamt bevorzugen? .	98
7.11	Bewertung der POI-Skalierung.	98
7.12	Schematische Darstellung von Sequenzen von Teilbewegungen zum Treffen eines Ziels nach dem OIIM [98]. . . .	101
7.13	Messung der Distanz des Bedienfingers zum Touchscreen.	104
7.14	Datenaufzeichnung einer typischen Bedienung.	105
7.15	Idealisiertes Bewegungsmodell mit Linearisierung der Bewegungstrajektorie in der Selektions- und der ballistischen Phase.	107
7.16	Idealisiertes Bewegungsmodell mit Linearisierung der Bewegungstrajektorie; schwarz: Idealisierte Modelldarstellung; grau: Verzerrung durch Messung wie in Bild 7.13 gezeigt.	108
7.17	Mittlere Bewegungszeiten MT für die einzelnen Varianten in den einzelnen Modi.	109
7.18	Bewegungs- und Geschwindigkeitsverlauf der einzelnen Versuchspersonen in Variante 2 im statischen Modus. . . .	110
7.19	Bewegungs- und Geschwindigkeitsverlauf der einzelnen Versuchspersonen in Variante 2 im Wizard-Modus.	111
7.20	Bewegungs- und Geschwindigkeitsverlauf der einzelnen Versuchspersonen in Variante 2 im Sensor-Modus.	111

7.21	Zusammenhang zwischen dem mittleren zeitlichen Abschluss der ballistischen Phase in Prozent der Gesamtbewegung und dem mittleren zurückgelegten Weg in der ballistischen Phase in Prozent der Gesamtweglänge der Bewegung für die einzelnen Modi und Varianten.	112
7.22	Mittleres Ende der ballistischen Phase sowie geloggte Events <i>entry</i> und <i>menu_complete</i> in Prozent der Gesamtbewegungszeit.	113
7.23	Mittlere absolute Bewegungszeiten aufgegliedert nach ballistischer und Selektionsphase für die einzelnen Varianten und Modi.	114
7.24	Mittlere Bewegungszeit für die einzelnen Varianten in den Bedienphasen MT_1, MT_2, MT_3	114
7.25	Linearisierter Bewegungsverlauf für die Variante 1.	116
7.26	Linearisierter Bewegungsverlauf für die Varianten 2 (links) und 3 (rechts) im statischen Modus.	117
7.27	Linearisierter Bewegungsverlauf für die Varianten 2 (links) und 3 (rechts) im Wizard-Modus.	117
7.28	Linearisierter Bewegungsverlauf für die Varianten 2 (links) und 3 (rechts) im Sensorik-Modus.	118
7.29	Darstellung des Blickverlaufs einer Versuchsperson und festgelegte AOIs in BeGaze; anstelle des LCTs ist die Kalibriervorlage zu sehen [148].	124
7.30	Mediane der Bewegungszeiten für die einzelnen Varianten.	125
7.31	Blickabwendung vom LCT Bedienung.	126
7.32	Blickzuwendung zum Touchscreen während der Bedienung.	127
7.33	Beginn der ersten Fixation auf den Touchscreen während der Bedienung.	127
7.34	Beginn der letzten Fixation auf dem Display während der Bedienung.	128
7.35	Mittlere Anzahl der Fixationen während der Bedienung.	129
7.36	Median der Fixationsdauern auf den Touchscreen während der Bedienung.	129
7.37	Allgemeine Bewertung des Infotainments und der POI-Vergrößerung.	133
7.38	Bewertung des Sofortmenüs.	134

8.1	Bewertung der ‚ortsspezifischen Suche‘ und der ‚direkten POI-Interaktion‘ in Bezug auf Nützlichkeit.	138
8.2	‚Ortsspezifische Suche‘ und ‚direkte POI-Interaktion‘.	138
8.3	Grundansichten von Mediaplayer und Navigationskarte.	141
8.4	Mediane der Bedienzeiten im <i>Point-to-Talk</i> - und <i>Talk-to-Object</i> -Modus.	144
8.5	Würden Sie diese Interaktionsmöglichkeit auch in Ihrem Fahrzeug nutzen wollen?	145
8.6	Box-Plots zur Frage: „Wie komfortabel war für Sie die Bedienung?“	146
8.7	Vergleich der Bedienzeiten mit der subjektiven Einschätzung.	146
8.8	Gegenüberstellung der Bedienmöglichkeiten ‚Sprache‘ und ‚Touch und Sprache‘.	147
8.9	Fehlerrate bei der Bedienung in den einzelnen Szenarien.	148
8.10	Subjektiver Vergleich der beiden Bedienvarianten PTT und TTO.	149

Abkürzungsverzeichnis

ACC	Adaptive Cruise Control
ASR	Automated Speech Recogniser
FAS	Fahrerassistenzsystem
GRA	Geschwindigkeitsregelanlage
HCI	Human-Computer-Interaction oder Human-Computer-Interface
HMC	Human-Machine-Interaction
HMI	Human-Machine-Interface oder Human-Machine-Interaction
LCT	Lane Change Test
MCI	Mensch-Computer-Interaktion
MMI	Mensch-Maschine-Interaktion oder Mensch-Maschine-Interface
MMK	Mensch-Maschine-Kommunikation
MMS	Mensch-Maschine-Schnittstelle
POI	Point of Interest
PTT	<i>Point-to-Talk</i> – Bedienverfahren, bei dem der Spracherkenner auf Annäherung aktiviert wird
TP	Traffic Program (Verkehrsfunk)
TTO	<i>Talk-to-Object</i> – Bedienverfahren, bei dem der Spracherkenner auf Longpress aktiviert wird

Literatur

- [1] ACCOT, J. ; ZHAI, S.: Refining Fitts' law models for bivariate pointing. In: *Conference on Human factors in computing systems*, 2003
- [2] ACM. *ACM SIGCHI Curricula for Human-Computer Interaction*. <http://www.acm.org/sigchi/cdg/cdg2.html> (Zugriff am 12.07.2006). 1996
- [3] AKYOL, S. ; CANZLER, U. ; BENGLER, K. ; HAHN, W.: Gesture Control for Use in Automobiles. In: *Workshop on Machine Vision Applications*, 2000
- [4] AKYOL, S. ; LIBUDA, L. ; KRAISS, K.-F.: Multimodale Benutzung adaptiver Kfz-Bordsysteme. In: *Kraftfahrzeugführung*. Springer, 2001
- [5] ALBRECHT, D. ; ZUKERMAN, I. ; NICHOLSON, A.: Bayesian Models for Keyhole Plan Recognition in an Adventure Game. In: *User Modeling and User-Adapted Interaction*, 1998, S. 5–47
- [6] ALTHOFF, F.: *Ein generischer Ansatz zur Integration multimodaler Benutzereingaben*, TU München, Diss., 2004
- [7] ALTHOFF, F. ; LINDL, R. ; WALCHSHÄUSL, L.: Robust Multimodal Hand- and Head Gesture Recognition for Controlling Automotive Systems. In: *VDI Berichte „Fahrer im 21. Jahrhundert“*, 2005
- [8] ALTHOFF, F. ; MCGLAUN, G. ; LANG, M.: Combining Multiple Input Modalities for Virtual Reality Navigation – A User study. In: *9th International Conference on Human Computer Interaction*, 2001
- [9] ARGYLE, M.: *Körpersprache und Kommunikation*. Junfermann, 1992

- [10] AUDI AG. *Beschreibung des Audi Multi Media Interface*. http://www.audi.de/audi/de/de2/neuwagen/a8/a8_limousine/audi_mmi.html (letzter Zugriff: 1.12.2007)
- [11] BACHFISCHER, K. ; BOHNENBERGER, T. ; HOFMANN, M. ; WÄLLER, C. ; WU, Y.: Kontext-adaptive Fahrerinformationssysteme am Beispiel eines Navigationssystems. In: *Künstliche Intelligenz* 3 (2007), S. 57–63
- [12] BACHFISCHER, K. ; HENZE, S. ; WÄLLER, C.: Emotional User Interfaces in the Car. In: *Emotion and Computing – Current Research and Future Impact*, 2006
- [13] BALLAGAS, R. ; BORCHERS, J.: Selexels: a Conceptual Framework for Pointing Devices with Low Expressiveness / RWTH Aachen. 2006. – Forschungsbericht
- [14] BEDWORTH, M. ; O’BRIEN, J.: The Omnibus Model: A New Model of Data Fusion? In: *IEEE Aerospace and Electronic Systems Magazine* Bd. 15, 2000
- [15] BEKKERING, H. ; SAILER, U.: Coordination of Eye and Hand in Time and Space. In: *Program Brain Research* 140 (2002), S. 365–373
- [16] BENDER, G.: *Touch Screen Performance as a Function of the Duration of Auditory Feedback and Target Size*, Wichita State University, Diss., 1999
- [17] BENGLER, K. ; HERRLER, M. ; KÜNZER, H.: Usability Engineering bei der Entwicklung von iDrive. In: *it - Information Technology* 44 (2002), S. 145
- [18] BERNOTAT, R.: Plenary Session: Operation Functions in Vehicle Control, Anthropotechnik in der Fahrzeugführung. In: *Ergonomics* 13 (1970), S. 353–377
- [19] Kap. Theoretische Grundlagen der Informationsfusion In: BEYERER, J. ; SANDER, J. ; WERLING, S.: *Fusion heterogener Informationsquellen*. Universitätsverlag Karlsruhe, 2006, S. 21–38

-
- [20] BILLINGHURST, M.: Put That Where? Voice and Gesture at the Graphics Interface. In: *SIGGRAPH Computer Graphics* 32 (1998), S. 60–63. – ISSN 0097–8930
- [21] BLANK, K.: *Benutzermodellierung für adaptive interaktive Systeme: Architektur, Methoden, Werkzeuge und Anwendungen*, Stuttgart, Diss., 1996
- [22] BLUTH, J.: *Modellierung und Realisierung eines digitalen Trägerfrequenzmesssystems zur Messung von kapazitiven Sensoren im Umfeld einer Ultrapräzisionsdrehmaschine*, Universität Kaiserslautern, Diss., 2006
- [23] BOHNENBERGER, T. ; BRANDHERM, B. ; GROSSMANN-HUTTER, B. ; HECKMANN, D. ; WITTIG, F.: Empirically Grounded Decision-Theoretic Adaptation to Situation-Dependent Resource Limitations. In: *Künstliche Intelligenz* 03 (2002), S. 10–16
- [24] BOLT, R.: Put-that-there: Voice and gesture at the graphics interface. In: *7th Annual Conference on Computer Graphics and Interactive Techniques*, 1980
- [25] BORTZ, J.: *Statistik für Sozialwissenschaftler*. Springer, 1999
- [26] BOUCHET, J. ; NIGAY, L.: ICARE: A Component-based Approach for the Design and Development of Multimodal Interfaces. In: *CHI '04 Extended Abstracts on Human Factors in Computing Systems*, 2004
- [27] BOUCHET, J. ; NIGAY, L. ; GANILLE, T.: ICARE: Software Components for Rapidly Developing Multimodal Interfaces. In: *6th International Conference on Multimodal Interfaces*, 2004
- [28] BOURGUET, M. ; ANDO, A.: Synchronization of Speech and Hand Gestures During Multimodal Human-Computer Interaction. In: *CHI 98 Conference Summary on Human Factors in Computing Systems*, 1998
- [29] BRANDHERM, B.: Verarbeitung von Sensordaten in Dynamischen Bayesschen Netzen. In: *Adaptivität und Benutzermodellierung in interaktiven Softwaresystemen*, 2001

- [30] BSI: *BSI DD 235:1996. Guide to in-vehicle information systems.* 1996. – British Standards Institution
- [31] BUBB, H. *Der Fahrprozess – Informationsverarbeitung durch den Fahrer.* VDA-Vortrag. 20./21. März 2002
- [32] Kap. Fahrerassistenz primär ein Beitrag zum Komfort oder für die Sicherheit? In: BUBB, H.: *VDI-Bericht Nr. 1768.* VDI-Verlag, 2003, S. 25–44
- [33] Kap. Systemstruktur In: BUBB, H. ; SCHMIDTKE, H.: *Ergonomie.* Hanser Verlag, 1993, S. 305–333
- [34] BURNS, P. ; LANSDOWN, T. *E-Distraktion: the challenges for safe and usable internet services in vehicles. Internet Forum on the Safety Impact of Driver Distraction When Using In-Vehicle Technologies.* <http://www.driverdistraction.org>. 2000
- [35] CAIRNIE, N. ; RICKETS, I. ; MCKENNA, S. ; MCALLISTER, G.: Using Finger-Pointing to Operate Secondary Controls in Automobiles. In: *IEEE Intelligent Vehicles Symposium*, 2000
- [36] CARD, S. ; MORAN, T. ; NEWELL, A.: *The Psychology of Human Computer Interaction.* Erlbaum, 1983
- [37] CAREY, D.: Eye-Hand Coordination: Eye to Hand or Hand to Eye? In: *Current Biology* 11 (2000), S. 416–419
- [38] CAVE, K. ; BICHOT, N.: Visospatial Attention: Beyond a Spotlight Model. In: *Psychonomic Bulletin and Review*, 1999
- [39] CHEN, L.: Incorporating Gesture and Gaze into Multimodal Models of Human-to-Human Communication. In: *Conference of the North American Chapter of the Association for Computational Linguistics on Human Language Technology*, 2006
- [40] COHEN, P. ; JOHNSTON, M. ; MCGEE, D. ; OVIATT, S. ; CLOW, J. ; SMITH, I.: The Efficiency of Multimodal Interaction: a Case Study. In: *6th Conference on Applied Natural Language Processing*, 1998

-
- [41] COHEN, P. ; JOHNSTON, M. ; MCGEE, D. ; OVIATT, S. ; PITTMAN, J. ; SMITH, I. ; CHEN, L. ; CLOW, J.: QuickSet: Multimodal Interaction for Distributed Applications. In: *5th ACM International Conference on Multimedia*, 1997
- [42] COHEN, P. ; MCGEE, D. ; CLOW, J.: The Efficiency of Multimodal Interaction for a Map-Based Task. In: *6th Conference on Applied Natural Language Processing*, 2000
- [43] CREED, C.: Using Emotion Simulation to Influence User Attitudes and Behaviour. In: *HCI 2005 - Doctoral Consortium*, 2005
- [44] DAHM, M.: *Grundlagen der Mensch-Computer-Interaktion*. Pearson Studium, 2006
- [45] DEY, A. ; ABOWD, G.: Towards a Better Understanding of Context and Context-Awareness / College of Computing, Georgia Institute of Technology. 1999. – Forschungsbericht
- [46] DIETZ, P. ; LEIGH, D.: DiamondTouch: A Multi-User Touch Technology. In: *14th Annual ACM symposium on User Interface Software and Technology*, 2001
- [47] DIN: *DIN EN ISO 15008:2003. Ergonomische Aspekte von Fahrerinformations- und Assistenzsystemen: Anforderungen und Bewertungsmethoden der visuellen Informationsdarstellung im Fahrzeug*. 2003. – Deutsches Institut für Normung
- [48] DIN: *DIN EN ISO 9241-110:2006. Ergonomie der Mensch-System-Interaktion - Teil 110: Grundsätze der Dialoggestaltung*. August 2006. – Deutsches Institut für Normung
- [49] DOUGLAS, S. ; MITHAL, A.: *The Ergonomics of Computer Pointing Devices*. Springer, 1997
- [50] DUTKE, S.: *Mentale Modelle: Konstrukte des Wissens und Verstehens: Kognitionspsychologische Grundlagen für die Software-Ergonomie*. Verlag für Angewandte Psychologie, 1994
- [51] EBERL, U. *Dialog mit einer Maschine*. Siemens Webzine, Ausgabe 2/1997 <http://w4.siemens.de/FuI/de/archiv/>

- zeitschrift/heft2_97/artikel103/index.html (letzter Zugriff am 19.03.2008). 1997
- [52] FAGERBERG, P. ; STAHL, A. ; HÖÖK, K.: eMoto: Emotionally Engaging Interaction. In: *Personal Ubiquitous Computing*, 2004
- [53] FITTS, P.: The Information Capacity of the Human Motor System in Controlling the Amplitude of Movement. In: *Journal of Experimental Psychology*, 1954
- [54] FRAUNHOFER: *PointScreen – Interaktion ohne Berührung*. http://www.imk.fraunhofer.de/sixcms/media.php/130/pointscreen_screen_deutsch02.pdf (letzter Zugriff am 19.03.2008). – Fraunhofer-Institut für Intelligente Analyse- und Informationssysteme IAIS
- [55] FRENZ, H. ; FREY, S. ; KEMPTER, G.: Theoretische Grundlagen der multimedialen Kommunikation. In: *Spektrum der Wissenschaft* 8 (1996), S. 32–38
- [56] FREY, S.: Prejudice and Inferential Communication: A New Look at an Old Problem. In: *Indoctrinability, Ideology and Warfare. Evolutionary Perspectives*, 1998
- [57] FREY, S.: *Die Macht des Bildes - Der Einfluß der nonverbalen Kommunikation auf Kultur und Politik*. Huber, 1999
- [58] FREY, S.: Neue Wege in der Kommunikationsforschung. In: *Gene, Neurone, Qubits & co. Unsere Welten der Information*, 1999
- [59] GEIGER, M.: *Berührungslose Bedienung von Infotainment-Systemen im Fahrzeug*, TU München, Diss., 2003
- [60] GEIGER, M. ; NIESCHULZ, R. ; ZOBL, M. ; LANG, M.: Bedienkonzept zur gestenbasierten Interaktion mit Geräten im Automobil. In: *Tagungsband VDI/VDE - GMA Fachtagung USEWARE 2002 Mensch-Maschine-Kommunikation/Design*, 2002
- [61] GEISER, G.: *Mensch-Maschine-Kommunikation*. Oldenbourg-Verlag, 1990

-
- [62] GERHARD, G.: User Modeling in Human-Computer-Interaction. In: *User Modeling and User-Adapted Interaction* 11 (2001), S. 65–86. – ISSN 0924–1868
- [63] GREEN, M.: Report on Dialogue Specification Tools. In: *User Interface Management Systems*, 1985
- [64] GREEN, P.: Measure and Methods Used to Assess the Safety and Usability of Driver Information Systems / University of Michigan, Transportation Research Institute. 1994. – Forschungsbericht
- [65] GREEN, P.: The 15-Second Rule for Driver Information Systems / Intelligent Transportation Society of America. 1999. – Forschungsbericht
- [66] GRICE, P.: Logic and Conversation. In: *Syntax and Semantics*, 1975
- [67] HASSENZAHL, M. ; BEU, A. ; BURMESTER, M.: Engineering Joy. In: *IEEE Software* 18 (2001), S. 2–8
- [68] HAUPTMANN, A.: Speech and Gestures for Graphic Image Manipulation. In: *SIGCHI Conference on Human Factors in Computing Systems*, 1989
- [69] HAUSCHILD, F. ; OEL, P.: Bedienung erlebbar machen. In: *Fachbeitrag der Carmeq*, 2006
- [70] HEINZE, C.: *Modelling Intention Recognition for Intelligent Agent Systems*, University of Melbourne, Diss., 2005
- [71] HOF, A.: *Entwicklung eines adaptiven Hilfesystems für multimodale Anzeige- und Bedienkonzepte im Fahrzeug*, Universität Regensburg, Diss., 2007
- [72] HOFMANN, M.: *Intentionsbasierte maschinelle Interpretation von Benutzeraktionen*, TU München, Diss., 2003
- [73] HORVITZ, E. ; BREESE, J. ; HECKERMAN, D. ; HOVEL, D. ; ROMMELSE, K.: The Lumiere Project: Bayesian User Modeling for Inferring the Goals and Needs of Software Users. In: *14th Conference on Uncertainty in Artificial Intelligence*, 1998

- [74] HÄUSSLER, J. ; ZIPF, A.: Multimodale Karteninteraktion und inkrementelle Zielführung zur integrierten Navigationsunterstützung für Fußgänger und Autofahrer. In: *Symposium für Angewandte Geographische Informationstechnologie*, 2003
- [75] HYRSKYKARI, A. ; MAJARANTA, P. ; RÄIHÄ, K.: From Gaze Control to Attentive Interfaces. In: *11th International Conference on Human-Computer Interaction*, 2005
- [76] JOHANNSON, G.: *Mensch-Maschine-Systeme*. Springer, 2003
- [77] JÜRGENSOHN, T.: *Hybride Fahrermodelle*. Pro Unniversitate Verlag, 1998
- [78] JÜRGENSOHN, T. ; TIMPE, K. (Hrsg.): *Kraftfahrzeugführung*. Springer, 2001
- [79] KASS, R. ; FININ, T.: The Role of User Models in Cooperative Interactive Systems. In: *International Journal of Intelligent Systems*, 1989
- [80] KERN, D. ; SCHMIDT, A.: Gefühlsorientierte Benutzerschnittstellen. In: *User Interface Update 1.0*, 2005
- [81] KIM, D. ; SONG, W. ; HAN, J. ; BIEN, Z.: Effective Intention Reading Technique As a Means of Human-robot Interaction for Human Centered Systems. In: *Knowledge-Based Systems*, 2001
- [82] KÜNZER, A.: *Handlungsprädiktion zur Gestaltung einer adaptiven Benutzungsunterstützung in autonomen Produktionszellen*, RWTH Aachen, Diss., 2005
- [83] KRAISS, K.: *Advanced Man-Machine Interaction*. Springer, 2006
- [84] KRÖSCHE, J. ; BOLL, S.: Kontextsensitive Points of Interest in mobilen Anwendungen, Schriftenreihe des Instituts für Geoinformatik. In: *Geoinformation und Mobilität – von der Forschung zur praktischen Anwendung, Beiträge zu den Münsteraner GI-Tagen*, 2004
- [85] LANG, M. *Mensch-Maschine-Kommunikation II*. Kurzmanuskript zur Vorlesung. 2000

-
- [86] LIBUDA, L. ; KRAISS, K.: Dialogassistentz im Kraftfahrzeug. In: *45. Fachausschusssitzung Anthropotechnik der DGLR „Entscheidungsunterstützung für die Fahrzeug- und Prozessführung“*. DGLR, 2003
- [87] LIU, Y.: Comparative Study of the Effects of Auditory, Visual and Multimodality Displays on Drivers' Performance in Advanced Traveller Information Systems. In: *Ergonomics* 44 (2001), S. 425 – 442
- [88] MACKENZIE, S.: Fitts' Law as a Research and Design Tool in Human-Computer Interaction. In: *Human-Computer Interaction* 7 (1992), S. 91–139
- [89] MACKENZIE, S. ; BUXTON, W.: Extending Fitts' Law to Two-Dimensional Tasks. In: *Proceedings of CHI-92*, 1992
- [90] MAGLIO, P. ; CAMPBELL, C.: Tradeoffs in Displaying Peripheral Information. In: *Conference on Human Factors in Computing Systems*, 2000
- [91] MAGLIO, P. ; MATLOCK, T. ; CAMPBELL, C. ; ZHAI, S. ; SMITH, B.: Gaze and Speech in Attentive User Interfaces. In: *International Conference on Multimodal Interfaces*, 2000
- [92] MARSCHNER, C.: *Messverfahren zur Eliminierung von Erdungseinflüssen bei kapazitiven Detektoren und ihre Anwendung zur Sitzbelegungserkennung in Kraftfahrzeugen*, TU München, Diss., 2003
- [93] MARTIN, J.: Towards Intelligent Cooperation Between Modalities. The Example of a System Enabling Multimodal Interaction with a Map. In: *Workshop on Intelligent Multimodal Systems*, 1997
- [94] MCGLAUN, G. ; ALTHOFF, F. ; LANG, M.: Ein neuer Systemansatz für die Integration multimodalen Inputs durch Late Semantic Fusion. In: *USEWARE 2002, VDI/VDE - GMA Fachtagung*, 2002
- [95] MCGLAUN, G. ; ALTHOFF, F. ; LANG, M. ; RIGOLL, G.: Kontextsensitives Fehlermanagement bei multimodaler Interaktion mit Infotainment- und Kommunikationseinrichtungen im Fahrzeug. In: *USEWARE 2004, VDI/VDE - GMA Fachtagung*, 2004

- [96] MCGLAUN, G. ; F. ALTHOFF, F. ; LANG, M. ; RIGOLL, G.: Development of a Generic Multimodal Framework for Handling Error Patterns during Human-Machine Interaction. In: *8th World Multi-Conference on Systems, Cybernetics, and Informatics*, 2004
- [97] MCGUFFIN, M. ; BALAKRISHNAN, R.: Fitts' Law and Expanding Targets: Experimental Studies and Designs for User Interfaces. In: *ACM Transactions on Computer-Human Interaction* 12 (2005), S. 388–422
- [98] MEYER, D. ; ABRAMS, R. ; KORNBLUM, S. ; WRIGHT, C. ; SMITH, J.: Optimality in Human Motor Performance: Ideal Control of Rapid Aimed Movements. In: *Psychological Review* 95 (1988), S. 340–370
- [99] MEYER, D. ; ABRAMS, R. ; WRIGHT, C.: Speed-Accuracy Tradeoffs in Aimed Movements: Toward a Theory of Rapid Voluntary Action. In: *Attention and Performance XIII, Motor Presentation and Control*, 1990
- [100] MÖLLER, C.: *Dreamteam Mensch-Computer – Neue Wege zur nutzergerechten Gestaltung des Mensch-Maschine-Dialogs*, Universität Duisburg-Essen, Diss., 2003
- [101] MORRIS, C.: Foundations of the Theory of Signs. In: *International Encyclopedia of Unified Science* 1 (1938)
- [102] NEGGERS, S.: *Oculomotor Behavior During Pointing*. Wissenschaftlicher Verlag Berlin, 2000
- [103] NEUSS, R.: *Usability Engineering als Ansatz zum Multimodalen Mensch-Maschine-Dialog*, TU München, Diss., 2000
- [104] NIEDERMAIER, F.: *Entwicklung und Bewertung eines Rapid-Prototyping Ansatzes zur multimodalen Mensch-Maschine-Interaktion im Kraftfahrzeug*, TU München, Diss., 2003
- [105] NIELSEN, J. ; MOLICH, R.: Heuristic Evaluation of User Interfaces. In: *SIGCHI conference on Human factors in computing systems*, 1990

-
- [106] NIGAY, L. ; COUTAZ, J.: A Design Space for Multimodal Systems: Concurrent Processing and Data Fusion. In: *Proceedings of the SIGCHI conference on Human factors in computing systems*, 1993
- [107] NIGAY, L. ; COUTAZ, J.: A Generic Platform for Addressing the Multimodal Challenge. In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 1995
- [108] NORMAN, D.: Cognitive Engineering. In: *User Centered System Design: New Perspectives on Human-Computer Interaction*. Erlbaum, 1986
- [109] NORMAN, D.: *The Psychology of Everyday Things*. Basic Books, 1988
- [110] NORMAN, D.: *The Invisible Computer*. MIT Press, 1998
- [111] NORMAN, D.: *The Design of Everyday Things*. Basic Books (reprint), 2002
- [112] OVIATT, S.: Ten Myths of Multimodal Interaction. In: *Communications of the ACM*, 1999. – ISSN 0001–0782
- [113] OVIATT, S. ; DEANGELI, A. ; KUHN, K.: Integration and Synchronization of Input Modes During Multimodal Human-Computer Interaction. In: *Proceedings of the SIGCHI conference on Human factors in computing systems*, 1997
- [114] PAETAU, M.: *Mensch-Maschine-Kommunikation. Software, Gestaltungspotentiale, Sozialverträglichkeit*. Frankfurt a. M. : Campus, 1990
- [115] PEARL, J.: *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. Morgan Kaufmann, 1988
- [116] PICKERING, C.: The Search for a Safer Driver Interface: A Review of Gesture Recognition Human Machine Interface. In: *Computing & Control Engineering Journal* 16 (2005), S. 34–40
- [117] PIECHULLA, W. ; MAYSER, C. ; GEHRKE, H. ; KÖNIG, W.: Echtzeit-Fahrerbeanspruchungsschätzung. In: *Information und Mobilität: 8ter internationales Symposium für Informationswissenschaft*, 2002

- [118] PIECHULLA, W. ; MAYSER, C. ; GEHRKE, H. ; KÖNIG, W.: Reducing Drivers' Mental Workload by Means of an Adaptive Man-Machine Interface. In: *Transportation Research Part F*, 2003
- [119] PO, B. ; FISHER, B. ; BOOTH, K.: Mouse and Touchscreen Selection in the Upper and Lower Visual Fields. In: *Proceedings of the SIGCHI conference on Human factors in computing systems*, 2004
- [120] POUPYREV, I. ; MARUYAMA, S. ; REKIMOTO, J.: Ambient Touch: Designing Tactile Interfaces for Handheld Devices. In: *15th annual ACM symposium on User interface software and technology*, 2002
- [121] RASSL, R.: *Ablenkungswirkung tertiärer Aufgaben im Pkw. System-ergonomische Analyse und Prognose*, TU München, Diss., 2004
- [122] RAUCH, N. ; GRADENEGGER, B. ; KRÜGER, H. *A three-level model of situation awareness for driving with in-vehicle devices*. Presentation on the Annual Meeting of the Human Factors and Ergonomic Society European Chapter. 2007
- [123] REEPS, I.: *Joy-of-Use – eine neue Qualität für interaktive Produkte*, Universität Konstanz, Diplomarbeit, 2004
- [124] REEVES, B. ; NASS, C.: *The Media Equation – How People Treat Computers, Television, and New Media Like Real People and Places*. CSLI Publ, 1996
- [125] REKIMOTO, J.: SmartSkin: An Infrastructure for Freehand Manipulation on Interactive Surfaces. In: *Proceedings of the SIGCHI conference on Human factors in computing systems*, 2002
- [126] ROSENBAUM, D.: *Human Motor Control*. San Diego: Academic Press, 1991
- [127] RUSER, H. ; PUENTE LEÓN, F: Informationsfusion – Eine Übersicht. In: *Technisches Messen* 74 (2007), S. 93–102
- [128] RUSSELL, S. ; NORVIG, P.: *Artificial Intelligence: A Modern Approach*. Englewood Cliffs, NJ: Prentice-Hall, 2003

-
- [129] Kap. Multimodale Anzeige- und Bedienkonzepte zur Steuerung technischer Systeme während der Fahrt im Kraftfahrzeug: Evaluationsbefunde zur Systemweiterentwicklung mit paralleler Sprachbedienung In: SCHATTENBERG, K. ; DEBUS, G.: *Kraftfahrzeugführung*. Springer, Berlin, 2001, S. 177–190
- [130] SCHEDLBAUER, M.: A Survey of Human Cognitive and Motor Performance Models / University of Massachusetts. 2007. – Forschungsbericht
- [131] SCHNEIDER, K.: *Softwarequalität. Grundlagen und Verfahren für Qualitätssicherung und Qualitätsmanagement*. Dpunkt Verlag, 2007
- [132] SCHWAN, H.: Dielectric Properties of tissues and Cell Suspensions. In: *Advances in Biological and Medical Physics*, 1957
- [133] SEARS, A. ; REVIS, D. ; SWATSKI, J. ; CRITTENDEN, R. ; SHNEIDERMAN, B.: Investigating Touchscreen Typing: The effect of keyboard size on typing speed. In: *Behaviour & Information Technology* 12 (1993), S. 17–22
- [134] SEIFERT, K.: *Evaluation multimodaler Computer-Systeme in frühen Entwicklungsphasen – Ein empirischer Ansatz zur Ableitung von Gestaltungshinweisen für multimodale Computer-Systeme*, TU Berlin, Diss., 2002
- [135] SHANNON, C.: A Mathematical Theory of Communications - Part I and II. In: *Bell System Technical Journal* Volume 27 (1948), S. 379–423 und 623–656
- [136] SHARMA, R. ; YEASIN, M. ; KRAHNSTOEVEER, N. ; RAUSCHERT, I. ; CAI, G. ; BREWER, I. ; MAC EACHREN, A. ; SENGUPTA, K.: Speech-Gesture Driven Multimodal Interfaces for Crisis Management. In: *Proceedings of the IEEE*, 2003
- [137] SHNEIDERMAN, B. ; PLAISANT, C.: *Designing the User Interface: Strategies for Effective Human-Computer-Interaction*. Addison-Wesley, 2004

- [138] SMITH, B. ; HO, J. ; ARK, W. ; ZHAI, S.: Hand Eye Coordination Patterns in Target Selection. In: *Symposium on Eye tracking Research & Applications*. New York, NY, USA : ACM Press, 2000. – ISBN 1-58113-280-8, S. 117–122
- [139] SMITH, J. *Toward Electric Field Tomography*. B.A., Computer Science, Philosophy, Massachusetts Institute of Technology. August 1995
- [140] SMITH, J.: *Electric Field Imaging*, MIT, Diss., 1999
- [141] STEVENS, A. ; QUIMBY, A. ; BOARD, A. ; KERSLOOT, T. ; BURNS, P.: Design Guidelines for Safety of In-Vehicle Information Systems / Transport Research Laboratory, UK. 2002. – Forschungsbericht
- [142] SUN, Y. ; CHEN, F. ; CHUNG, V.: QuickFusion: Multimodal Fusion Without Time Thresholds. In: *NICTA-HCSNet Multimodal User Interaction Workshop*, 2006
- [143] THEOFANOU, D.: *Maße zur Erfassung von visueller Ablenkung bei verschieden komplexen Streckenabschnitten*, Universität Regensburg, Diss., 2002
- [144] TIMPE, K.: Fahrzeugführung: Anmerkungen zum Thema. In: *Kraftfahrzeugführung*. Springer, 2001
- [145] TOTZKE, I.: Die Humanisierung multimodaler HMI – Die Bedeutung der Modalität und Situation der Mensch-Maschine-Interaktion. In: *IIR-Fachkongress: E-Car Infotainmentplattform und Telematikdienste für das Multimedia Auto 2001*, 2001
- [146] TOTZKE, I. ; SCHMIDT, G. ; KRÜGER, P.: Mentale Modelle von Menüsystemen – Bedeutung kognitiver Repräsentationen für den Kompetenzerwerb. In: *DGLR-Bericht: Entscheidungsfindung für die Fahrzeug- und Prozessführung*, 2003
- [147] TROGEMANN, G.: Erdbeersysteme – Über Humankommunikation und soziotechnische Interfaces. In: *Signale der Störung*. Wilhelm Fink Verlag, 2003

-
- [148] TRÖSTERER, S. ; LASCH, M.: *Abschlussbericht über die Durchführung einer Studie zur Analyse des Blickverhaltens bei der Betätigung unterschiedlicher Schaltflächen auf einem Touchscreen*. 2007. – Interner Bericht für die Konzernforschung Elektronik, Volkswagen AG
- [149] TSE, E. ; SHEN, C. ; GREENBERG, S. ; FORLINES, C.: Enabling Interaction With Single User Applications Through Speech and Gestures on a Multi-User Tabletop. In: *Working Conference on Advanced Visual Interfaces*, 2006
- [150] VERTEGAAL, R.: Designing Attentive Interfaces. In: *Symposium on Eye tracking research & applications*, 2002
- [151] VO, M. ; WAIBEL, A.: Multi-modal HCI: combination of gesture and speech recognition. In: *INTERACT and CHI Conference Companion on Human Factors in Computing Systems*, 1993
- [152] WAHLSTER, W. ; WAHLSTER, W. (Hrsg.): *SmartKom: Foundations of Multimodal Dialogue Systems*. Springer, 2006 (Cognitive Technologies)
- [153] WAHLSTER, W. ; KOBZA, A.: User Models in Dialog Systems. In: *User Models in Dialog Systems*. Springer, 1989
- [154] WERNER, D. ; SCHNEIDER, U.: *Taschenbuch der Informatik*. Fachbuchverlag Leipzig, 2004
- [155] Kap. Visual and manual demands of in-car controls and displays In: WIERWILLE, W.: *Automotive Ergonomics*. Taylor and Francis, 1993, S. 299–320
- [156] WIERWILLE, W. ; ANTIN, J. ; DINGUS, T. ; HULSE, M.: Visual attentional demand of an in-car navigation display system. In: *Vision in Vehicles 2* (1988), S. 307–316
- [157] WIMMER, R. ; HOLLEIS, P. ; KRANZ, M. ; SCHMIDT, A.: Thracker - Using Capacitive Sensing for Gesture Recognition. In: *6th International Workshop on Smart Appliances and Wearable Computing* Bd. 0. Los Alamitos, CA, USA : IEEE Computer Society, 2006. – ISSN 1545–0678, S. 64

- [158] WIMMER, R. ; KRANZ, M. ; SCHMIDT, A. ; BORING, S.: Pervasive Activity Detection with Capacitive Sensing. In: *5th Annual IEEE Conference on Pervasive Computing and Communications*, 2007
- [159] ZEILINGER, S.: *Aktive haptische Bedienelemente zur Interaktion mit Fahrerinformationssystemen*, Universität der Bundeswehr München, Diss., 2005
- [160] ZIMMERMAN, T.: *Personal Area Networks (PAN): Near-Field Intra-Body Communication*, MIT Media Lab, Diplomarbeit, 1995
- [161] ZOBL, M. ; GEIGER, M. ; SCHULLER, B. ; LANG, M. ; RIGOLL, G.: A Real-Time System for Hand Gesture Controlled Operation of In-Car Devices. In: *IEEE International Conference on Multimedia and Expo*, 2003
- [162] ZWAHLEN, H. ; ADAMS, C. ; DEBALD, D.: Safety aspects of CRT touch panel controls on automobiles. In: *Vision in Vehicles 2* (1988), S. 335–344