



Gestaltung avatarbasierter natürlichsprachlicher Hilfesysteme für den Einsatz in Fahrzeugen

Valentin Antonius Nicolescu



TECHNISCHE UNIVERSITÄT MÜNCHEN

Lehrstuhl für Wirtschaftsinformatik

Gestaltung avatarbasierter natürlichsprachlicher Hilfesysteme für den Einsatz in Fahrzeugen

Valentin Antonius Nicolescu

Vollständiger Abdruck der von der Fakultät für Informatik der Technischen Universität München zur Erlangung des akademischen Grades eines

Doktors der Naturwissenschaften (Dr. rer. nat.)

genehmigten Dissertation.

Vorsitzender: Univ.-Prof. Dr. Dr. h.c. Manfred Broy

Prüfer der Dissertation:

1. Univ.-Prof. Dr. Helmut Krcmar
2. Univ.-Prof. Gudrun J. Klinker, Ph.D.

Die Dissertation wurde am 27.01.2009 bei der Technischen Universität München eingereicht und durch die Fakultät für Informatik am 13.07.2009 angenommen.

Bibliografische Information der Deutschen Nationalbibliothek

Die Deutsche Nationalbibliothek verzeichnet diese Publikation in der Deutschen Nationalbibliografie; detaillierte bibliografische Daten sind im Internet über <http://dnb.ddb.de> abrufbar.

1. Aufl. - Göttingen : Cuvillier, 2009
Zugl.: (TU) München, Univ., Diss., 2009

978-3-86955-075-6

Audi Dissertationsreihe, Band 20

© CUVILLIER VERLAG, Göttingen 2009
Nonnenstieg 8, 37075 Göttingen
Telefon: 0551-54724-0
Telefax: 0551-54724-21
www.cuvillier.de

Alle Rechte vorbehalten. Ohne ausdrückliche Genehmigung des Verlages ist es nicht gestattet, das Buch oder Teile daraus auf fotomechanischem Weg (Fotokopie, Mikrokopie) zu vervielfältigen.

1. Auflage, 2009
Gedruckt auf säurefreiem Papier

978-3-86955-075-6

Geleitwort

Die Entwicklung von Automobilen wird heute mehr denn je von neuen Technologien getrieben, insbesondere software-basierte Funktionen spielen eine zunehmend wichtige Rolle. Sie ermöglichen die Bereitstellung neuartiger Assistenz- und Entertainmentsysteme und erlauben durch eine schnellere Erstellung zum einen das frühe Bereitstellen von Prototypen und zum anderen eine Verkürzung des traditionellen Entwicklungspfads. Vor allem die schnelle Bereitstellung von Ergebnissen ermöglicht frühe Usability- und Akzeptanztests mit den späteren Benutzern.

Die Vorteile dieser Entwicklung gehen jedoch auch mit neuen Herausforderungen einher. Durch die Software-Basis hält eine Vielfalt, Dynamik und damit Komplexität Einzug in das Fahrzeug, wie man sie bisher nur von Computersystemen gewohnt ist. Auch führt die vereinfachte Entwicklung zu vielen, teilweise gänzlich neuartigen Funktionen im Fahrzeug. Dieser Vielzahl an Funktionalitäten ist mit herkömmlichen Bedienkonzepten kaum noch zu begegnen, zumal nicht mehr jeder Funktion im Fahrzeug ein Bedienelement bzw. eine Bedienart zugewiesen werden kann. Zahlreiche Fahrzeughersteller sind bereits dazu übergegangen, eine zentrale Bedien- und Steuereinheit im Fahrzeug zu platzieren, die über ein universelles Menü Zugriff auf vielen Funktionen und Einstellungsmöglichkeiten bietet. Dieser Ansatz kommt den aktuellen Entwicklungen entgegen, verlagert das Problem der vielen Bedienelemente jedoch auf eine umfangreiche Menüstruktur, deren Bedienung dem Fahrer nicht weniger Kenntnis abverlangt.

Die zukünftige Entwicklung im Fahrzeug erfordert eine Schnittstelle, die intuitiv und ohne Schulungen verwendet werden kann und damit die Nutzung aller Optionen einfach macht. Ein solcher Ansatz ist eine sprachgesteuerte Schnittstelle zum Fahrzeug. Bei den heutigen Implementierungen kommen nur Sprachkommandos zum Einsatz, die auch erlernt werden müssen. Dagegen würde natürlichsprachliche und interaktive Kommunikation mit dem Auto auch ohne besondere Kenntnisse benutzt werden können. Zur weiteren Reduzierung der Hemmschwelle, sich mit einer Maschine natürlich sprachlich zu unterhalten, bietet sich zudem eine figürliche Darstellung des Interaktionspartners als Avatar an.

Der Mehrwert dieser Dissertation besteht in der vollständigen Umsetzung einer solchen, universellen, natürlichsprachlichen Schnittstelle für Funktionen im Fahrzeug. Neben den dabei entwickelten Architekturen und Methoden, bietet die Umsetzung in einem Fahrzeug die Basis für die detaillierte Untersuchung und Bewertung dieses Bedienparadigmas.

Ich wünsche der Arbeit die ihr gebührende Verbreitung und vor allem, daß die in ihr vorgelegte Vision frischen Wind in die Diskussion um die zukünftige Benutzung von Funktionen im Fahrzeug bringen wird.

Prof. Dr. Helmut Krcmar
Lehrstuhl für Wirtschaftsinformatik
Technische Universität München

Zusammenfassung

Moderne Fahrzeuge der Mittel- und Oberklasse verfügen über eine zunehmende Anzahl an Funktionen und Bedienelementen, die für die Insassen die Fahrt angenehmer und sicherer gestalten sollen. Diese Menge an Bedienoptionen stellt insbesondere den Fahrer vor die Herausforderung, sich mit allen offensichtlichen und versteckten Funktionen vertraut machen zu müssen. Das gedruckte Bordhandbuch, das als Hilfestellung in solchen Situationen gedacht ist, wird von den Benutzern kaum verwendet, weshalb alternative Ansätze den Zugriff auf die benötigten Informationen sichern müssen. Ein möglicher Ansatz zur einfachen Vermittlung von Benutzungsinformationen, der auch während der Fahrt genutzt werden kann, stellt ein interaktives, natürlichsprachliches Dialogsystem dar, das mithilfe eines Avatars die Hemmschwelle zur natürlichsprachlichen Interaktion senken soll. Da es sich dabei um einen bisher unerforschten Ansatz handelt, wird er im Rahmen dieser Forschungsarbeit umgesetzt und untersucht.

Die Grundlage der technischen Umsetzung des Avatarsystems im Fahrzeug stellt eine Referenzarchitektur für Avatarsysteme im Allgemeinen dar. Sie wird auf Basis der Analyse von Komponenten bestehender Avatar- und Sprachdialogsysteme erstellt. Die bereinigten Komponenten werden in die vier Bereiche der Eingabe, der Reaktionsermittlung, der Ausgabe und des Designs aufgeteilt und zu Teilsystemen zusammengestellt. Die daraus zusammengesetzte Gesamtreferenzarchitektur stellt die Ausgangsbasis für die konkrete Anwendung auf ein Avatarsystem im Fahrzeug dar, das insbesondere die eingeschränkten technischen Möglichkeiten und die Mobilität des Umfelds berücksichtigt.

Die Implementierung der Architektur des Avatarsystems erfolgt in einem Serienfahrzeug (Audi A4 Modell 2008), das zu Versuchszwecken erweitert wurde. Dies ermöglicht es, dass Videosignal des Multimedia-Interfaces im Fahrzeug durch die Ausgabe eines CarPC's zu überlagern, Informationen auf den Bussystemen des Fahrzeugs auszulesen und Sprachsignale zu erfassen und in Form einer Sprachsynthese wieder auszugeben.

Zur Umsetzung der aus der Architektur abgeleiteten Komponenten wird das Java Framework OSGi herangezogen, das die Entwicklung unabhängiger Bundles ermöglicht. So werden teilweise unter Integration bestehender Standardsoftware Bundles für die Spracherkennung, die Sprachsynthese, das Dialogmanagement, einen Chatbot, den Avatar und dessen dreidimensionaler Umwelt sowie zur Anbindung des CAN-Bus im Fahrzeug umgesetzt. Diese technische Plattform erlaubt die Verarbeitung multimedial aufbereiteten Inhalts, der zur Designzeit erstellt wird.

Bei der Erstellung der Inhalte des Avatarsystems in Dialogform wird das gedruckte Handbuch genutzt, das in seiner digitalen Form als strukturierte Datei vorliegt. Für die Anwendung im Versuchsfahrzeug werden ausgewählte Szenarien umgesetzt, die auch für die spätere Evaluation mit Probanden geeignet sind. Dabei werden Themen gewählt, die insbesondere neuartige Funktionen wie den Audi side assist, Audi lane assist oder Audi drive select betreffen. Aufgrund der Menge und der Struktur der Informationen wird ein systematisches

Vorgehen zur Konvertierung des Handbuchs in eine Dialogform erforderlich. Diese Konvertierung basiert dabei einerseits auf einem softwaregestützten Werkzeug, dem Dialogdesigner, und einem Vorgehensmodell.

Der Dialogdesigner erlaubt den Import des digitalen Handbuchs und die grafische Bearbeitung des daraus generierten Dialogmodells. Dabei können Sprachein- und -ausgaben, Ereignisse auf dem CAN-Bus des Fahrzeugs, die Reaktion des Avatars und seiner Umwelt, die Anzeige in der Benutzerschnittstelle und logische Elemente wie Verzweigungen, Variablen und Verweise in einem Modellierungstool integriert bearbeitet werden. Die Texte und Bilder des Handbuchs können dadurch mit weiterem multimedialem Material sowie der Mimik und Gestik eines Avatars ergänzt und zur Laufzeit verwendet werden. Die Übertragung der modellierten Dialogstrukturen erfolgt über den Export einer Regelbasis für den Chatbot und einer zugehörigen Sprachgrammatik für die Spracherkennung.

Der Dialogdesigner ermöglicht somit die Umsetzung des konzipierten Vorgehensmodells, das in sieben Phasen die notwendigen Arbeiten vom Handbuch zum Dialog detailliert. Es werden nach dem Import des Handbuchs die Phasen der Strukturierung, der Antwortgestaltung, der Eingabengestaltung, der Mediengestaltung, der Gestaltung der Umfeldfunktionen, der optionalen Integration von Emotionen und des Tests unterschieden. Die Aufteilung des Vorgehens in Phasen erlaubt damit die Zuweisung einzelnen Aufgaben an entsprechende Bearbeiter. Der mehrfache Durchlauf des zyklisch konzipierten Modells führt zur systematischen Erstellung und Verbesserung der Regelbasis für den Chatbot.

Die aufbereiteten Szenarien werden mit insgesamt 67 Probanden in drei Gruppen im Versuchsfahrzeug verwendet und analysiert. Dabei haben alle drei Gruppen die gleichen fünf Aufgaben zu lösen. Gruppe eins löst die Aufgaben nur mithilfe des gedruckten Handbuchs bei stehendem Fahrzeug und dient somit als Kontrollgruppe, während Gruppe zwei das Avatarsystem ebenfalls bei stehendem Fahrzeug verwendet. Gruppe drei löst die Aufgaben mit dem Avatarsystem während der Fahrt. Die Evaluation erfolgt durch einen prä- und einen postaktionalen Fragebogen sowie einer Videoaufzeichnung während der Aufgabenbearbeitung.

Dabei zeigt sich u. a. einer zielsicherere Verwendung von Funktionen und eine schnellere Aufgabenlösung von kurzen Aufgaben bei der Benutzung des Avatarsystems. Zudem wird das in den Szenarien vermittelte Wissen im Falle des Avatarsystems besser durch die Probanden verinnerlicht. Ebenso haben die Benutzer des Avatarsystems mehr Spaß bei der Aufgabenbearbeitung und bewerten das Avatarsystem als sehr nützliche Hilfe. Die abschließende Bewertung in Schulnoten zeigt das beste Ergebnis bei der Verwendung des Avatarsystems im Stand, während das Handbuch und das Avatarsystem während der Fahrt gleich gut an zweiter Stelle stehen.

Die Ergebnisse zeigen, dass ein Avatarsystem als Hilfe im Fahrzeug ein vielversprechender Ansatz ist, dessen optimale Gestaltung auf Basis des Systems dieser Forschungsarbeit weiter erforscht werden kann.

Vorwort

Die Faszination für Avatare und Avatarsysteme entwickelte sich bereits während meines Studiums, als ich einen Fernsehbericht über Avatarsysteme im Internet sah. Die Idee, mit technischen Mitteln virtuelle Menschen zu erschaffen, fesselte mich so sehr, dass ich bereits meine Diplomarbeit diesem Thema widmete und ein System für den Einsatz im Internet entwickelte. Während der Arbeit an meiner Diplomarbeit zeigten sich bereits viele Ansatzpunkte, die es weiter zu erforschen galt, dass ich den Vorsatz fasste, auch im Rahmen meiner Doktorarbeit diesen Bereich weiter zu beleuchten. Die technischen Fortschritte bei der Entwicklung von Standardsoftware eröffneten zusätzlich neue Aspekte in diesem Bereich.

Als sich die Möglichkeit ergab, zusammen mit Audi in einem spannenden Projekt ein Avatarsystem im Fahrzeug zu entwickeln, bekam ich die Möglichkeit, diesen Vorsatz in die Tat umzusetzen. Die technischen Herausforderungen und Gegebenheiten des Automotive Bereichs machten das Vorhaben umso interessanter. Das Ziel, das in Form eines Versuchsfahrzeugs auch erreicht wurde, war die Umsetzung eines Avatarsystems im Fahrzeug, das nach dem Motto „Give me something to hate“ die Ausgangsbasis für weitere Forschung in diesem Bereich darstellt.

Sowohl die technische als auch inhaltliche Vernetzung von Teilbereichen, ließen das Forschungsvorhaben zu einem Projekt werden, dass ich ohne Hilfe nicht in dieser Form hätte bewältigen können, weshalb ich mich bei allen Beteiligten bedanken möchte.

Zunächst möchte ich mich bei meinen Doktorvater Prof. Dr. Helmut Krcmar bedanken, der mich immer wieder ermutigt hat, einen Schritt weiter zu gehen, als ich vor hatte; bei Prof. Dr. Jan Marco Leimeister, der als Forschungsbereichsleiter den Fokus der Ergebnisse mit ruhiger Hand in die richtige Richtung geführt hat. Zudem möchte ich meinem Kollegen Holger Hoffmann danken, der zum einen durch seine inhaltliche Nähe oft als Gesprächspartner fungierte und zum anderen durch die Entwicklung seines Rapid Prototyping Projekts HIMEPP einige der verwendeten OSGi-Bundles bereitgestellt hat. Eine große Unterstützung insbesondere bei Fragen der Java Entwicklung sowie dem zeit- und nervenraubenden Reverse Engineering war mein Hiwi Sergej Trushin, der sehr souverän alle Aufgaben gelöst hat, die ich ihm gestellt habe. Ein weiterer herausragender Student, der den Grundstein für die grafische Oberfläche des Dialogdesigners gelegt hat, ist Gabriel Zavaczki-Bucur, dem ich an dieser Stelle für sein Engagement danken möchte.

Neben wichtigen Menschen im universitären Umfeld, spielte natürlich auch der Praxisbezug zu Audi eine sehr wichtige Rolle. Zunächst möchte ich Michael Faulbacher dafür danken, dass er das Projekt AViCoS in seiner Fachabteilung aufgenommen und geleitet hat. Stefan Sellschopp hat als Ansprechpartner in der Fachabteilung und inhaltlicher Gesprächspartner wichtige Impulse gegeben und mit seiner unerschöpflichen Energie bei allen Fachabteilungen Audis Stakeholder ausfindig gemacht und mit ins Boot genommen. Ohne Josef Schmidts Hilfe, der das Display Switch Module entwickelt, den Umbau des Versuchsfahrzeug koordiniert und auch sonst alle technischen Fragen beantwortet hat, würde das Testfahrzeug vermutlich auch optisch dem Titel „Versuchsfahrzeug“ entsprechen. Auch wenn sich bis

heute nur wenige Überschneidungen mit anderen INI.TUM- oder Audi-Projekten ergeben haben, danke ich Herrn Dr. Uwe Koser für seinen unermüdlichen Einsatz bei der Vernetzung meines Projekts.

Sebastian Schillinger, den ich in seiner Funktion als Koordinator der Bordliteratur besucht habe, hat mit seiner Begeisterung für das Thema immer wieder einen Energieschub vermittelt, der mich zur Umsetzung immer neuer Features beflügelt hat. Ohne Boris Blasinski, der mir die „Geheimnisse“ des CAN-Bus bei Audi erklärt hat, würde ich vermutlich heute noch vor dem CAN-Bus-Bundle sitzen und nicht begreifen, wieso es nicht geht. Stefan Christ, Kristin Dettmer und Tahar Bouaziz haben mich für die Vision von Audi hinsichtlich möglicher Bedienkonzepte sensibilisiert. Einige Monate an Wartezeit und dementsprechend viele graue Haare ersparte mit Herr Maca, der durch seinen Einsatz das Testfahrzeug sehr zügig aufbereiten ließ, so dass es schnell erweitert werden konnte. Abschließend will ich Christoph Frandrup für die Bereitstellung einiger 3D-Modelle, Wilfried Mader und Elfriede Erlewein für die Betreuung der Patentmeldungen und allen Audi-Mitarbeitern danken, die ich an dieser Stelle vergessen habe.

Durch die enge Verbindung zwischen Audi und VW, konnte ich mich auch mit der Konzernforschung zu meinem Projekt austauschen. In diesem Zusammenhang möchte ich Frau Dr. Yongmei Wu danken, welche innerhalb des Konzerns das Thema der Avatarsysteme im Fahrzeug angestoßen und mit der entsprechenden DVD für den Tiguan den ersten Schritt in die kommerzielle Nutzung gegangen ist. Darüber hinaus danke ich Dr. Peter Öl, Lorenz Bohrer, Moritz Neugebauer und Günter Horna zum konstruktiven Austausch zu diesem Thema. Abschließend möchte ich mich bei der gesamten Patent- und Rechtsabteilung von VW bedanken, die mich erst für das Thema der Patente sensibilisiert hat und ohne die ich nicht auf die Idee gekommen wäre, so viele Patentanmeldungen einzureichen.

Im Rahmen des Projekts wurde so weit möglich Standardsoftware verwendet. Zunächst ist da Helio Perroni Filho, der das Chatterbean entwickelt und unter GPL zur Verfügung gestellt hat. Weitere kritische Komponenten, die 3D-Engine, 3D-Objekte und der Avatar selbst, wurden von der Firma Charamel kostenlos zur Verfügung gestellt und mit Rat und Tat unterstützt. Aus diesem Grund möchte ich Norbert Helff, Alexander Bruno Reinecke und Lukas Übachs danken.

Abschließend möchte ich Rainer Rimmele und Stefanie Leimeister für das Korrekturlesen meiner Arbeit danken. Selbstverständlich geht mein Dank auch an meine Familie und alle anderen, die mich moralisch bei meinem Promotionsvorhaben unterstützt haben.

Inhaltsverzeichnis

Geleitwort	I
Zusammenfassung	III
Vorwort	V
Inhaltsverzeichnis	VII
Abbildungsverzeichnis	XII
Tabellenverzeichnis	XIV
Abkürzungsverzeichnis	XV
1. Problemstellung und Aufbau der Arbeit	1
1.1. Motivation	1
1.2. Vision des Lösungsansatzes	4
1.3. Forschungsfragen	5
1.4. Aufbau der Arbeit	6
2. Wissenschaftliche Grundposition und Methodologie	9
2.1. Wissenschaftstheoretische Strukturierungsansätze von Forschungsströmungen ..	9
2.2. Wissenschaftsverständnis dieser Arbeit.....	12
2.2.1. Konstruktivismus	12
2.2.2. Design Science	14
2.2.2.1. Phasen des Designzyklus	14
2.2.2.2. Erweiterung des Zyklus.....	16
2.2.3. Forschungsdesign	17
2.3. Zusammenfassung	19
3. Begriffliche Grundlagen und Stand der Technik	21
3.1. Sprachsynthese	22
3.1.1. Begriffliche Klärung	22
3.1.2. Verfahren zur Sprachsynthese.....	23
3.1.3. Text-to-Speech (TTS)	25
3.1.4. Optimierung der synthetisierten Sprache	27
3.1.4.1. Phonetische Optimierung	28
3.1.4.2. Prosodische Optimierung	29
3.1.5. Einsatzbereiche und Beispiele.....	31
3.2. Spracherkennung	32
3.2.1. Begriffliche Klärung	32
3.2.2. Funktionsweise der Spracherkennung.....	34
3.2.3. Sprachgrammatiken.....	36
3.2.4. Einsatzbereiche und Beispiele.....	38
3.3. Chatbots	39
3.3.1. Begriffliche Klärung	40

3.3.2.	Funktionsweise eines Chatbots	41
3.3.2.1.	Strategien zur Dialogsteuerung	41
3.3.2.2.	Sprachverarbeitung durch einen Chatbot	43
3.3.3.	Aufbau der Wissensbasis	44
3.3.3.1.	Artificial Intelligence Markup Language (AIML)	45
3.3.3.2.	VoiceXML	46
3.3.4.	Einsatzbereiche und Beispiele	48
3.4.	Avatare	49
3.4.1.	Begriffliche Klärung	49
3.4.2.	Technische Umsetzung	52
3.4.3.	Einsatzbereiche und Beispiele	54
3.4.3.1.	Strukturierung der Einsatzbereiche	54
3.4.3.2.	Wichtige Einsatzbereiche und Beispiele	55
3.5.	Avatarsysteme	57
3.5.1.	Begriffliche Klärung	57
3.5.2.	Technisches Zusammenspiel	59
3.5.2.1.	Abstimmung zweier Komponenten	60
3.5.2.2.	Abstimmung mehrerer Komponenten und ganzer Systeme	61
3.5.3.	Soziale Effekte	63
3.5.4.	Einsatzgebiete und Beispiele	65
3.6.	Fahrerassistenz- und Fahrerinformationssysteme	67
3.6.1.	Begriffliche Klärung	67
3.6.1.1.	Eindimensionale Unterscheidungsansätze	68
3.6.1.2.	Mehrdimensionale Unterscheidungsansätze	68
3.6.2.	Beispiele und deren Einordnung	70
3.6.2.1.	Assistenten aus einzelnen Bausteinen	70
3.6.2.2.	Kombinierte Assistenzsysteme	72
3.6.2.3.	Allgemeine Hilfesystemen im Fahrzeug	73
3.7.	Zusammenfassung	74
4.	Analyse und Anforderungserhebung	77
4.1.	Betrachtung des Forschungsumfelds	78
4.2.	Identifikation von Stakeholdern	79
4.2.1.	Stakeholder innerhalb des Volkswagen Konzerns	80
4.2.2.	Weitere Stakeholder	82
4.3.	Formulierung der operativen Ziele	83
4.4.	Ableitung der Anforderungen	88
4.4.1.	Anforderungen zur Umsetzung eines Avatarsystems (erstes Ziel)	88
4.4.2.	Anforderungen zur Benutzung ohne Lernaufwand (zweites Ziel)	89
4.4.3.	Anforderungen zur Minimierung der Ablenkung (drittes Ziel)	91
4.4.4.	Anforderungen zur Minimierung des Erstellungsaufwands (viertes Ziel)	93
4.4.5.	Anforderungen zur Modularität der Softwarekomponenten (fünftes Ziel)	94
4.5.	Zusammenfassung	95
5.	Konzeption	97
5.1.	Technische Konzeption	98
5.1.1.	Analyse bestehender Avatar- und Sprachdialogsysteme	98

5.1.1.1. Untersuchte Systeme und Architekturen.....	98
Sprachdialogsysteme.....	99
Avatarsysteme.....	100
5.1.1.2. Identifizierte Komponenten.....	101
Komponenten zur Eingabe.....	102
Komponenten zur Ausgabe.....	103
Datenkomponenten.....	105
Komponenten zur Reaktionsermittlung.....	107
Designkomponenten.....	111
5.1.2. Konzeption einer Referenzarchitektur für Avatarsysteme.....	112
5.1.2.1. Eingabesystem.....	113
5.1.2.2. Ausgabesystem.....	114
5.1.2.3. Designsystem.....	116
5.1.2.4. Reaktionsermittlungssystem.....	116
5.1.2.5. Übersicht der Referenzarchitektur.....	118
5.1.3. Anwendung der Referenzarchitektur auf den Einsatz im Fahrzeug.....	119
5.1.3.1. Verteilung der Systeme und Komponenten.....	120
5.1.3.2. Konkrete Systemarchitektur des Forschungsvorhabens.....	122
5.2. Inhaltliche Konzeption.....	123
5.2.1. Gestaltung der Kommunikation und des Domänenwissens.....	124
5.2.1.1. Struktureller Aufbau und Organisation der Inhalte.....	124
5.2.1.2. Bereitstellung von Funktionen zum Dialogablauf.....	126
5.2.1.3. Abdeckung inhaltlicher Bereiche.....	128
5.2.2. Inhaltserstellung und -pflege.....	131
5.2.2.1. Vorgehen zur Erstellung der Regelbasis.....	131
5.2.2.2. Werkzeuge zur Erstellung und Pflege.....	133
5.3. Konzeption der Benutzerschnittstelle.....	135
5.3.1. Gestaltung des Avatars.....	135
5.3.2. Gestaltung der Benutzeroberfläche.....	137
5.4. Zusammenfassung.....	139
6. Implementierung eines interaktiven Hilfesystems im Fahrzeug.....	141
6.1. Implementierung des technischen Systems.....	142
6.1.1. Technische Integration ins Fahrzeug.....	142
6.1.2. Wahl eines Entwicklungsframeworks.....	144
6.1.2.1. Übersicht relevanter Frameworks.....	144
6.1.2.2. Merkmale von OSGi.....	146
6.1.2.3. Anwendung von OSGi auf die Systemarchitektur.....	149
6.1.3. Umsetzung der Komponenten.....	150
6.1.3.1. Spracherkennung.....	151
6.1.3.2. CAN-Bus und Input Manager.....	152
Überblick wichtiger Bus-Systeme im Fahrzeug.....	152
Funktionsweise des CAN-Bundles.....	154
Designoptionen.....	156
6.1.3.3. Dialogmanager.....	157
6.1.3.4. Chatbot.....	158
AliceBot ChatterBean.....	159
Herausforderungen beim Einsatz von AliceBot im Avatarsystem.....	160

Umsetzung des Chatbot-Bundles	163
6.1.3.5. GUI-Framework mit Präsentationsmanager.....	166
6.1.3.6. Animationskomponente	169
Charamel CharActor	169
Funktionsumfang und Steuerung	171
6.1.3.7. Sprachsynthese	173
6.1.4. Entwicklung und Zusammenspiel der Komponenten	175
6.2. Erstellung von Inhalten	177
6.2.1. Gestaltung des Dialogeditors	177
6.2.1.1. Aufbau des Dialogeditors.....	178
6.2.1.2. Schritttypen	181
Grundlagen der Dialogeinteilung	181
Schritttypen des Dialogeditors	182
6.2.1.3. Synonymgruppen	185
6.2.1.4. Austauschfunktionen des Dialogeditors.....	187
Import von SGML.....	187
Export von AIML.....	188
6.2.2. Auswahl von Szenarien.....	190
6.2.3. Konvertierung von Inhalten in Dialoge.....	193
6.2.3.1. Phase 1: Strukturierung von Informationen	195
6.2.3.2. Phase 2: Gestaltung von Antworten	197
Schritte der Antwortphase	197
Richtlinien der Antwortphase.....	199
6.2.3.3. Phase 3: Gestaltung von Eingaben	200
6.2.3.4. Phase 4: Gestaltung von Medien	203
Gestaltung und Verwaltung der Medien	203
Integration von Avataramationen.....	205
6.2.3.5. Phase 5: Umsetzung von Umfeldfunktionen.....	206
6.2.3.6. Phase 6: Integration von Emotionen	209
6.2.3.7. Phase 7: Test der Dialoge.....	211
6.2.3.8. Besonderheiten bei der Konvertierung unstrukturierter Informationen....	213
6.3. Zusammenfassung	214
7. Evaluation des Hilfesystems	217
7.1. Planung und Durchführung der Evaluation.....	217
7.1.1. Konzeption der Evaluation.....	217
7.1.1.1. Vorbereitung des Versuchsfahrzeugs.....	219
7.1.1.2. Auswahl der Versuchspersonen	221
7.1.1.3. Beschreibung der Versuchsstrecke.....	223
7.1.2. Durchführung der Evaluation.....	225
7.2. Ergebnisse	227
7.2.1. Vorkenntnisse der Probanden.....	228
7.2.1.1. Ausbildung der Probanden	228
7.2.1.2. Technische Vorkenntnisse.....	230
7.2.1.3. Vorkenntnisse im Bereich Autofahren und Erfahrungen mit der Marke Audi	231
7.2.2. Bearbeitung der Testaufgaben.....	233
7.2.2.1. Bearbeitungserfolge	233

7.2.2.2.	Bearbeitungszeiten erfolgreich bearbeiteter Testaufgaben	233
7.2.2.3.	Zielsicherheit bei der Verwendung von Bedienelementen	236
7.2.2.4.	Wissen über die Bedienelemente	238
7.2.3.	Vergleich des Handbuchs mit dem Avatarsystem.....	239
7.2.3.1.	Benutzerfreundlichkeit	239
7.2.3.2.	Nutzen	240
7.2.3.3.	Gesamteindruck.....	242
7.2.4.	Bewertung des Avatarsystems mit geschlossenen Fragen	242
7.2.4.1.	Avatar	243
7.2.4.2.	Inhalt und Visualisierung	244
7.2.4.3.	Benutzeroberfläche.....	246
7.2.4.4.	Spracherkennung.....	247
	Bewertung der Spracherkennung auf Basis der postaktionalen Fragebögen	248
	Bewertung der Spracherkennung auf Basis der Videoanalyse.....	249
7.2.4.5.	Sprachausgabe	251
7.2.4.6.	Dialogverlauf bzw. Dialogmanager	252
7.2.4.7.	Gesamteindruck.....	253
7.2.5.	Offene Bewertung des Avatarsystems	255
7.2.5.1.	Positiv wahrgenommene Aspekte in den offenen Fragen.....	255
7.2.5.2.	Negativ wahrgenommene Aspekte in den offenen Fragen	256
7.3.	Ableitung von Hypothesen.....	258
7.4.	Zusammenfassung	260
8.	Schlüsse aus der Umsetzung des Hilfesystems.....	263
8.1.	Zusammenfassung der Ergebnisse	263
8.1.1.	Technische Entwicklung des Avatarsystems	263
8.1.2.	Gestaltung der Inhalte	265
8.1.3.	Wirkung auf Benutzer	266
8.2.	Fazit	267
8.3.	Ausblick.....	269
8.3.1.	Entwicklungspotenzial beim AViCoS-Avatarsystem	269
8.3.1.1.	Technische Weiterentwicklung	269
8.3.1.2.	Inhaltliche Weiterentwicklung	271
8.3.2.	Weiterentwicklung von Avatarsystemen im Fahrzeug	273
Literaturverzeichnis.....	277	
Anhang	303	
Anhang A: Mögliche Aufgaben für die Evaluation im Fahrzeug.....	303	
Anhang B: Präaktionaler Fragebogen der Evaluation.....	307	
Anhang C: Postaktionaler Fragebogen der Evaluation.....	315	

Abbildungsverzeichnis

Abbildung 1-1:Entwicklung der Funktionsmenge in Fahrzeugen	1
Abbildung 1-2:Bestehende Ansätze zur Informationsvermittlung im Fahrzeug	3
Abbildung 1-3:Ablauf der Arbeit	7
Abbildung 2-1:Entscheidungsparameter des Entwurfs eines Forschungsdesigns	11
Abbildung 2-2:Beschreibung der konstruktivistischen Grundposition	14
Abbildung 2-3:Information Systems Research Framework.....	17
Abbildung 3-1:Zusammenhänge der Grundlagenbereiche	21
Abbildung 3-2:Zusammenhänge der phonemischen und phonetischen Elemente.....	23
Abbildung 3-3:Phasen und Schritte von Text-to-Speech (TTS)	27
Abbildung 3-4:Stufen der Texteingabe zur Sprachsynthese	28
Abbildung 3-5:Auswirkung der Betonung auf die Bedeutung eines Satzes	29
Abbildung 3-6:Phasen und Schritte der linguistischen Spracherkennung	35
Abbildung 3-7:Beispiel eines Worthypothesengraphs	36
Abbildung 3-8:Beispiel einer BNF-Grammatik	37
Abbildung 3-9:Funktionen eines Chatbots in einem Sprachdialogsystem.....	41
Abbildung 3-10:Strategien zur Dialogsteuerung.....	42
Abbildung 3-11:Verarbeitungsschritte in einem Chatbots	44
Abbildung 3-12:Beispiel von AIML Regeln.....	46
Abbildung 3-13:Beispiel eines VoiceXML Menüs.....	47
Abbildung 3-14:Bekanntere Avatare: Robert T. Online, Lara Croft und Kyoko Date.....	50
Abbildung 3-15:Unterscheidung von Avatartypen nach dem Grad der Autonomie.....	50
Abbildung 3-16:Avatare als Repräsentation künstlicher und natürlicher Intelligenzen	51
Abbildung 3-17:Optionen der technischen Umsetzung von Avataren.....	53
Abbildung 3-18:Eigenschaften, Technologien und Betrachtungswinkel virtueller Menschen	54
Abbildung 3-19:Grundlegende Komponenten eines möglichen Avatarsystems	59
Abbildung 3-20:Notation eines Unterhaltungsausschnitts zweier Gesprächspartner	62
Abbildung 3-21:Beispiel eines kommerziellen Avatarsystems im Internet: Anna von IKEA	65
Abbildung 3-22:Interaktion mit dem virtuellen Immobilienmakler REA	66
Abbildung 3-23:Modell zur Identifizierung von Bausteinen eines Fahrerassistenzsystems....	69
Abbildung 4-1:Ableitung von Anforderung aus Zielen der Stakeholder	77
Abbildung 4-2:Organisationen und deren Zusammenhänge im Forschungsumfeld.....	78
Abbildung 4-3:Lebenszyklus von visuellen Daten für Handbücher bei der AUDI AG	85
Abbildung 5-1:Zusammenhang der einzelnen Konzeptionsteile	97
Abbildung 5-2:Input System der Referenzarchitektur für Avatarsysteme.....	114
Abbildung 5-3:Ausgabesystem der Referenzarchitektur für Avatarsysteme	115
Abbildung 5-4:Designsystem der Referenzarchitektur für Avatarsysteme.....	116
Abbildung 5-5:Reaktionsermittlungssystem der Referenzarchitektur für Avatarsysteme....	117
Abbildung 5-6:Gesamte Referenzarchitektur für Avatarsysteme	119
Abbildung 5-7:Verteilungsmöglichkeiten für Avatarsysteme	121
Abbildung 5-8:Architektur für Avatarsysteme im Fahrzeug	122
Abbildung 5-9:Designherausforderungen eines Embodied Conversational Agents (ECA) ..	123
Abbildung 5-10:Tiefe und Breite der inhaltlichen Bereiche.....	129
Abbildung 5-11:Verfeinerung des Inhaltsmodells	131
Abbildung 5-12:Aufwand zur Verbesserung der Beantwortungsrate	132
Abbildung 5-13:Screenshot des Regeleditors GaitoBot.....	134
Abbildung 5-14:Beispiele der konzeptionellen Avatargestaltung in Poser	137
Abbildung 5-15:Konzeptionelle Anordnung der Teile der Benutzeroberfläche	138

Abbildung 6-1:Zweiteilung der Implementierung in Technik und Inhalt.....	141
Abbildung 6-2:Integration des CarPC's in das Fahrzeug.....	143
Abbildung 6-3:Unterbringung des CarPC und des DSM im Handschuhfach.....	144
Abbildung 6-4:Architektur des OSGi Frameworks	147
Abbildung 6-5:Interaktion während des Bundle-Lebenszyklus.....	148
Abbildung 6-6:Aufteilung derArchitektur in OSGi-Bundles und weitere Komponenten	150
Abbildung 6-7:Verwendung der VoCon 3200 Bibliothek über JNI	151
Abbildung 6-8:Aufbau der XML-Datei zur Steuerung der CAN-Signalverarbeitung.....	155
Abbildung 6-9:Verarbeitungsschema des CAN-Bundles	156
Abbildung 6-10:Steuerung der Antwortermittlung im Dialogmanager	158
Abbildung 6-11:AIML Regeln zur Veranschaulichung des That-Problems.....	161
Abbildung 6-12:Initialisierung und der Antwortermittlung der Chatbot-Komponente	165
Abbildung 6-13:Befehlsverarbeitung im Output-System	166
Abbildung 6-14:Screenshot des GUI Frameworks	167
Abbildung 6-15:Avatar „Helen“ von Charamel.....	170
Abbildung 6-16:Zusammenspiel der Komponenten des Avatarsystems	177
Abbildung 6-17:Aufbau des Dialogeditors	179
Abbildung 6-18:Vereinfachtes Entity Relationship Modell des Dialogeditors	180
Abbildung 6-19:Verzweigungsschritt in Verbindung mit Sprachausgabeschritten.....	183
Abbildung 6-20:Verwendung eines Sprungschritts	184
Abbildung 6-21:Pflege einer Synonymgruppe im Synonymeditor.....	186
Abbildung 6-22:Multiplikation von Eingabeformulierungen durch Synonymgruppen.....	186
Abbildung 6-23:Umwandlung des PIM in ein AIML-PSM.....	189
Abbildung 6-24:Vorgehensmodells zur Konvertierung von Informationen in einen Dialog	193
Abbildung 6-25:Virtuelles Cockpit der Audi Handbuch CD.....	204
Abbildung 6-26:Komponenten eines emotionalen Modells für Avatar und Benutzer.....	210
Abbildung 6-27:Typischer Prozess zur Erstellung der technischen Dokumentation.....	214
Abbildung 7-1:Kameraperspektiven in Drauf - und Längssicht	220
Abbildung 7-2:Blinkwinkel und Aufbau der Kamera.....	221
Abbildung 7-3:Route der Versuchsstrecke und der Verlängerungsoptionen.....	224
Abbildung 7-4:Arbeitsplatz des Betreuers auf der Rückbank während der Versuchsfahrten	226
Abbildung 7-5:Schematischer Ablauf der Evaluation	227
Abbildung 7-6:Einstellung der Probanden gegenüber der Marke Audi.....	232
Abbildung 7-7:Verwendung von Handbuch bzw. Avatarsystem und Bearbeitungserfolg....	233
Abbildung 7-8:Korrekt beantwortete Wissensfragen.....	238
Abbildung 7-9:Bewertung der Benutzerfreundlichkeit.....	240
Abbildung 7-10:Bewertung des Nutzens	241
Abbildung 7-11:Bewertung der verwendeten Hilfe mit Schulnoten.....	242
Abbildung 7-12:Bewertung des Avatars	244
Abbildung 7-13:Bewertung der Visualisierung	245
Abbildung 7-14:Bewertung der Benutzeroberfläche	246
Abbildung 7-15:Verwendete Befehle der Benutzeroberfläche	247
Abbildung 7-16:Bewertung der Spracherkennung auf Basis der postaktionalen Fragebögen	248
Abbildung 7-17:Bewertung der Sprachausgabe.....	251
Abbildung 7-18:Bewertung des Dialogverlaufs.....	252
Abbildung 7-19:Bewertung des Gesamteindrucks.....	254
Abbildung 7-20:Gesamtbewertungsdurchschnitte als Indiz für den Verbesserungsbedarf ...	255
Abbildung 7-21:Anzahl der Nennungen positiver Aspekte nach Kategorien geordnet.....	256
Abbildung 7-22:Anzahl der Nennungen negativer Aspekte nach Kategorien geordnet.....	257
Abbildung 8-1:Mögliches Nutzungsverhalten nach inhaltlichen Bereichen.....	271

Tabellenverzeichnis

Tabelle 2-1: Strukturierungsansätze wissenschaftstheoretischer Grundpositionen	12
Tabelle 2-2: Anwendung des IS Research Frameworks auf die konkrete Forschung	18
Tabelle 3-1: Verfahren der Sprachsynthese	25
Tabelle 3-2: Weitere relevante AIML-Tags	46
Tabelle 4-1: Stakeholder und ihre Eigenschaften innerhalb des Volkswagen Konzerns.....	81
Tabelle 4-2: Weitere Stakeholder und ihre Eigenschaften.....	82
Tabelle 4-3: Unterstützung der Ziele durch die Stakeholder	87
Tabelle 4-4: Abgeleitete Anforderungen aus dem ersten operativen Ziel	89
Tabelle 4-5: Abgeleitete Anforderungen aus dem zweiten operativen Ziel.....	91
Tabelle 4-6: Abgeleitete Anforderungen aus dem dritten operativen Ziel.....	92
Tabelle 4-7: Abgeleitete Anforderungen aus dem vierten operativen Ziel.....	94
Tabelle 4-8: Abgeleitete Anforderungen aus dem fünften operativen Ziel	94
Tabelle 5-1: Untersuchte Sprachdialogsystemarchitekturen und ihre Merkmale	99
Tabelle 5-2: Untersuchte Avatarsystemarchitekturen und ihre Merkmale	101
Tabelle 5-3: Komponenten zur Eingabe.....	103
Tabelle 5-4: Komponenten zur Ausgabe.....	105
Tabelle 5-5: Datenkomponenten	107
Tabelle 5-6: Komponenten zur Reaktionsermittlung.....	110
Tabelle 5-7: Komponenten zum Design und der Pflege	112
Tabelle 5-8: Funktionen zur Dialog- und Umfeldsteuerung	128
Tabelle 6-1: CAN-Bussysteme im Audi A4 Testträger	153
Tabelle 6-2: Syntax für zusätzliche Befehle des Avatarsystems innerhalb von AIML	162
Tabelle 6-3: Konvertierung von Phonetik- und Prosodieanweisungen im Avatarsystem	163
Tabelle 6-4: Symbole des GUI-Frameworks	168
Tabelle 6-5: Werksfunktionen zur Steuerung des Avatars und seiner Umwelt.....	172
Tabelle 6-6: Tags der SVOX Markup Language	175
Tabelle 6-7: Schritttypen des Dialogeditors und ihre Merkmale	185
Tabelle 6-8: Wichtige SGML-Tags des Audi Benutzerhandbuchs und ihre Konvertierung	187
Tabelle 6-9: Bedienschritte und ihre Zusammenstellung zu Evaluationsszenarien.....	192
Tabelle 6-10: Mögliche Verteilung von Aufgaben an Personengruppen.....	195
Tabelle 6-11: Schritte und Richtlinien für die Phase der Strukturierung.....	197
Tabelle 6-12: Schritte und Richtlinien für die Phase der Antwortgestaltung	200
Tabelle 6-13: Schritte und Richtlinien für die Phase der Eingabengestaltung.....	203
Tabelle 6-14: Schritte und Richtlinien für die Phase der Mediengestaltung.....	206
Tabelle 6-15: Schritte für die Phase der Umfeldfunktionen	208
Tabelle 6-16: Schritte für die Phase der Emotionsintegration	211
Tabelle 6-17: Schritte für die Phase der Dialogtest.....	213
Tabelle 7-1: Einteilung und Wortlaut der Aufgaben während der Evaluation	219
Tabelle 7-2: Demographische Daten der Probanden.....	223
Tabelle 7-3: Bildungsabschlüsse und Ausbildungsfachrichtungen der Probanden	229
Tabelle 7-4: Technische Vorkenntnisse der Probanden.....	230
Tabelle 7-5: Fahrerfahrung der Probanden	232
Tabelle 7-6: Bearbeitungszeiten erfolgreich bearbeiteter Testaufgaben.....	235
Tabelle 7-7: Anteil der Verwendung korrekter Bedienelemente	237
Tabelle 7-8: Anzahl an Spracheingaben, Erkennungsraten und Kommandoanteile.....	250
Tabelle 7-9: Korrelationskoeffizienten der Spracheingabe und der Akzeptanz	260

Abkürzungsverzeichnis

ABNF	Augmented Backus-Naur-Form
ADB	ARD AusspracheDatenBank
AIML	Artificial Intelligence Markup Language
AMI-C	Automotive Multimedia Interface Collaboration
ASR	Automatic Speech Recognition
AUTOSAR	AUTomotive Open System ARchitecture
AViCoS	Avatar based Virtual Co-driver System
BLF/ BOSS LF	Bonn Open Synthesis System Label Format
BNF	Backus-Naur-Form / Backus-Normal-Form
BOMP	Bonn machine-readable Pronunciation Dictionary
CAD	Computer aided Design
CAE	Computer aided Engineering
CAN	Controller Area Network
CMUDict	Carnegie Mellon University Pronouncing Dictionary Format
CRM	Customer Relationship Management
CSTR	Center for Speech Technology Research Edinburgh
DSM	Display Switch Module
DTW	Dynamic Time Warping
ERL	Volkswagen Electronic Research Lab
TD-PSOLA	Time Domain Pitch Synchronous OverLap Add
EBNF	Extended Backus-Naur-Form
ECA	Embodied Conversational Agent
EMMA	Extensible MulitModal Annotation markup language
FAS	Fahrerassistenzsystem
FIS	Fahrerinformationssystem
GDML	Generic Dialogue Modeling Language
GSL	Nuance Grammar Specification Language
GToBi	German Tone and Break Indices
HIMEPP	Highly Integrated Modular Embedded Prototyping Platform
HIS	Herstellerinitiative Software
HMM	Hidden-Markov-Modelle
HNM	Harmonic plus Noise Model
INI.TUM	Ingolstadt Institute der TU München
IPA	Internationales Phonetische Alphabet
JNI	Java Native Interface
JSGF	Java Speech Grammar Format
JSML	Java Speech Markup Language
KI	Künstliche Intelligenz

KIMSAC	Kiosk-based Integrated Multimedia Service Access for Citizens
LIN	Local Interconnect Network
LPC	Linear Predictive Code
MARY	Modular Architecture for Research on speech sYnthesis
MDA	Model Driven Architecture
MMI	Multimedia Interface
MMORPG	Massively Multiplayer Online Role-Playing Game
MOF	Meta Object Facility
MOST	Media Oriented Systems Transport
NPC	Non Player Character
OCL	Object Constraint Language
ORM	Object-Relational Mapping
OSEK/VDX	Offene Systeme für die Elektronik im Kraftfahrzeug/ Vehicle Distributed Executive
OSGi	Open Services Gateway initiative
PDA	Personal Digital Assistant
PIM	Platform Independent Model
PNAMBIC	Pay No Attention to that Man Behing the Curtain
PSM	Platform Specific Model
REA	Real Estate Agent
SALT	Speech Application Language Tags
SAM	Speech Assessment Methods
SAMPA	Speech Assessment Methods Phonetic Alphabet
SAMPA-D	SAMPA für Deutsch
SAMPROSA	Speech Assessment Methods PROSsodic Alfabet
SAPI	Speech Application Programming Interface
SGML	Standard Generalized Markup Language
SIL	Semantic Interface Language
SPICE	Software Process Improvement and Capability Determination
SSML	Speech Synthesis Markup Language
STML	Spoken Text Markup Language
StVG	Straßenverkehrs-Gesetz
StVO	Straßenverkehrs-Ordnung
ToBi	Tone and Break Indices
TTS	Text-to-Speech
UML	Unified Modeling Language
WIMP	Window-Icon-Mouse-Pointer
XMI	XML Metadata Interchange
XSAMPA	Extended SAMPA

1. Problemstellung und Aufbau der Arbeit

„The reasonable man adapts himself to the world;
the unreasonable one persists in trying to adapt the world to himself.
Therefore all progress depends on the unreasonable man.“

George Bernard Shaw

Der Anstoß für die vorliegende Betrachtung avatarbasierter, natürlichsprachlicher Hilfesysteme für Fahrzeuge entstammt aus einer Problemstellung, mit der Fahrer von Mittel- und Oberklasse-Fahrzeugen zunehmend konfrontiert werden, die jedoch vielen nicht bewusst ist: Der Flut an Bedienoptionen im Fahrzeug. Um das notwendige Verständnis für den Untersuchungsgegenstand und die damit verbundenen Fragestellungen herzustellen, wird zunächst die Motivation für diese Arbeit veranschaulicht. Darauf aufbauend werden die drei emergenten Forschungsfragen abgeleitet. Abschließend steht der Aufbau der verbleibenden Kapitel dieser Arbeit im Vordergrund.

1.1. Motivation

Bei der Entwicklung von Automobilen wurde in den letzten Dekaden zunehmend auf Software-basierte Systeme gesetzt, was viele Schritte vereinfacht und neue Möglichkeiten eröffnet hat (Drews/Frees/Schellhase 2006; Koch 2004). Einerseits hat dies zur Erleichterung in der Konzeption und Umsetzung technischer Innovationen geführt, andererseits zu einer Zunahme an Funktionalität. Schäuffele und Zurawka (2006, 15) schätzen, dass die Anzahl an Funktionen von 1970 bis 2000 sich mehr als vervierfacht hat, was auch von Huber (2004) bestätigt wird (vgl. Abbildung 1-1). Dabei hat sich das Wachstum exponentiell entwickelt, so dass auch weiterhin mit einer kontinuierlichen Zunahme zu rechnen ist (Mercer/Fraunhofer IPA/Fraunhofer IML 2004, 15).

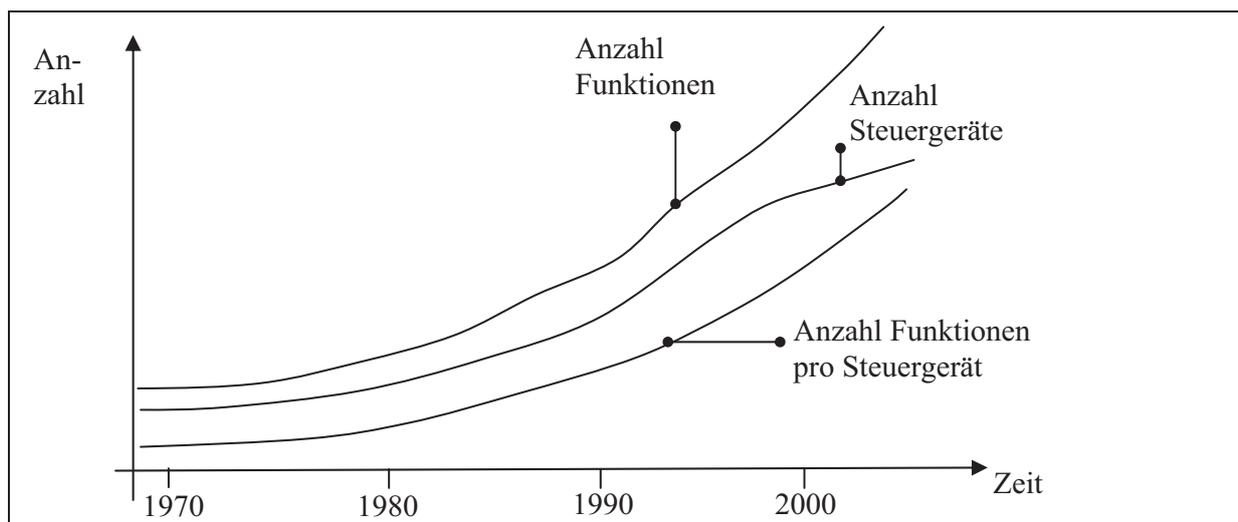


Abbildung 1-1: Entwicklung der Funktionsmenge in Fahrzeugen
Quelle: (Schäuffele/Zurawka 2006, S. 15)

Während die zunehmende Anzahl von Funktionen in Form neuer und erweiterter Steuergeräte dem Fahrer überwiegend verborgen geblieben ist, haben dennoch viele von ihnen

Auswirkungen auf das Bedienkonzept und die Bedienelemente. Diese technikgetriebene Zunahme an neuen Bedienmöglichkeiten führt bei vielen Fahrern zu einer Überforderung. Die Erweiterung des Funktionsumfangs innerhalb des Autos findet insbesondere im tertiären Funktionsbereich statt (vgl. dazu Tönnis/Broy/Klinker 2006), so dass sich die Fahrer nicht zwangsläufig mit den neuen Möglichkeiten beschäftigen müssen, um die Grundfunktionen des Autos - also das Fahren selbst - verwenden zu können (Sacher/Bubb 2006; Wäller et al. 2006, 4). Somit befindet sich in heutigen Automobilen eine Vielzahl an Funktionen, die nur von technikaffinen Benutzern voll ausgeschöpft wird.

Der Unkenntnis der Fahrer versuchen Automobilhersteller auf verschiedenen Wegen zu begegnen. Eine Möglichkeit ist, die Anzahl nicht offensichtlicher Funktionen, so genannter versteckter Funktionen, im Fahrzeug zu reduzieren. So hat DaimlerChrysler in den letzten Jahren etwa 600 Funktionen aus ihren Fahrzeugen wieder entfernt, nachdem Sie vom Nutzer nicht genutzt bzw. größtenteils nicht bekannt waren (Koch 2004, 25). Zum anderen können neben der Verbesserung gedruckter Handbücher hinsichtlich der Übersichtlichkeit und der Verständlichkeit auch neue Wege beschritten werden wie z.B. die Bereitstellung wichtiger Inhalte in Form einer Audio-CD (Grünweg 2007; Frankfurter Allgemeine Zeitung 2007), die Erstellung einer HTML-basierten Anleitung für den Heimcomputer bzw. für die Verwendung im Fahrzeug selbst zeigen. Obwohl diese Ansätze eine Verbesserung bringen können, sind sie nicht uneingeschränkt und zielgerichtet während der Fahrt verwendbar. Das Lesen des gedruckten Handbuchs ist während der Fahrt nicht ohne eine negative Beeinflussung des Fahrverhaltens möglich, während die Audio-CD nur Zugriff auf ausgewählte Informationen bietet. Die technische Infrastruktur in einem Serienfahrzeug bietet zudem nur eingeschränkte Möglichkeiten, HTML-basierte Inhalte darzustellen und damit digitale Anleitungen zu verwenden, die Informationen ohnehin nur in Textform bereitstellen können.

Aufbauend auf den oben vorgestellten Variationen von Hilfesystemen für den Einsatz im Fahrzeug, die jedoch die an sie gestellten Anforderungen aufgrund des jeweils verwendeten Mediums nicht voll erfüllen können, hat sich der Gedanke gebildet, einen virtuellen Beifahrer zu erstellen, der wichtige Bedienelemente erläutern kann. Eine erste serienmäßige Umsetzung dieses Gedankens erfolgte durch die Volkswagen AG, die für ihr neues Modell Tiguan eine Video-Anleitung präsentierte. Die entsprechende DVD kann im Fahrzeug abgespielt werden, wobei zu bestimmten Bedienelementen die virtuelle Assistentin „Carla“ in einer Videosequenz eine Erklärung liefert (Heise Autos 2007; vgl. auch Wäller et al. 2006). Diese Lösung kann jedoch ebenso wie die Audio-CD nur einer stark eingeschränkten Menge an Fragestellungen begegnen und erlaubt keine Interaktion zwischen Fahrer und Hilfesystem. Eine weiterführende Vision hat Daimler mit dem Konzeptauto F-700 vorgestellt, bei dem eine virtuelle Assistentin per Sprachein- und -ausgabe und der Darstellung eines Avatars bestimmte Aktionen im Kommunikationsumfeld des Fahrzeugs übernehmen kann (Mercedes Scene 2007).

Kritisch zu bewerten ist diese Lösung insofern, als dass sie bisher lediglich eine Vision ist und nur eine sehr eingeschränkte Menge an Funktionen unterstützt. Dabei liegt der Fokus nicht in der Erläuterung von Bedienelementen, deren Benutzung dem Fahrer nicht vertraut ist, sondern auf der Unterstützung bei der Bedienung des Telefons oder des Navigationssystems. Die Umsetzung eines interaktiven Handbuchs unter Verwendung eines virtuellen Assistenten

mit Sprachein und -ausgabe steht nach wie vor noch aus. Eine Gegenüberstellung der bisher genannten Lösungsansätze hinsichtlich der beiden kritischen Dimensionen Umfang der Informationen und Zugriff während der Fahrt ist in Abbildung 1-2 zu sehen. Dabei ist zu beachten, dass die Vision von Daimler nicht den Fokus der Informationsvermittlung hat, jedoch konzeptionell weit entwickelt ist und deshalb in dieser Betrachtung berücksichtigt werden muss.

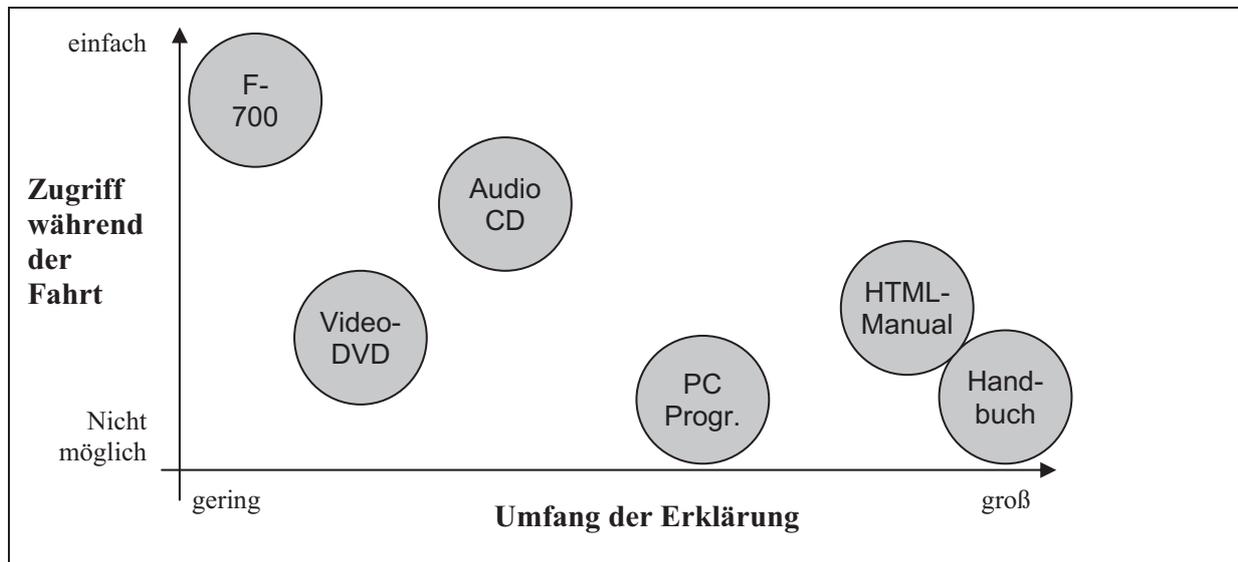


Abbildung 1-2: *Bestehende Ansätze zur Informationsvermittlung im Fahrzeug*
Quelle: (Eigene Darstellung)

Das Optimum der möglichen Ausprägungen verspricht somit einerseits einen einfachen Zugriff während der Fahrt und einen großen Umfang der angebotenen Informationen. Insbesondere die Herausforderung des Wissensumfangs stellt bei multimedialen Aufbereitungen bestehender Inhalte wie der genannten Video-DVD oder der Audio-CD einen großen Aufwand dar (vgl. dazu Rudolph/Krcmar 2004, 8ff). Daraus kann eine weitere Anforderung abgeleitet werden, welche die einfache Erweiterung und Pflege von Inhalten postuliert.

Der Anspruch, den diese Arbeit somit erfüllen muss, ist es, den bereits eingeschlagenen Pfad der virtuellen Assistenten im Fahrzeug weiterzugehen, um auch während der Fahrt eine möglichst große Menge an Wissen rund um ein Fahrzeug verfügbar zu machen. Dabei stellt sich jedoch die grundlegende Frage nach dem Sinn und den Auswirkungen einer solchen Lösung. Da sich dies nur anhand eines vollständigen Systems ermitteln lässt, ist das übergeordnete Ziel der Arbeit, die Erstellung eines technischen Systems, das sich hinsichtlich der Kriterien

- Zugriff während der Fahrt
- Umfang der Erläuterungen und damit verbunden
- einfache Pflege der Inhalte

aufbauend auf den bestehenden Lösungen dem möglichen Optimum annähern soll. Das Ergebnis versteht sich somit einerseits als Machbarkeitsstudie der angestrebten Lösung und gleichzeitig als Basis für spätere Untersuchungen hinsichtlich Sinn, Einfluss und Akzeptanz von Avatarsystemen im Fahrzeug.

1.2. Vision des Lösungsansatzes

Das Ziel der Entwicklungstätigkeit dieser Arbeit ist die Erstellung eines Systems im Auto, das die Vermittlung von Wissen zur Bedienung des Fahrzeugs und die Interpretation der Meldungen des Fahrzeugs ermöglichen soll. Somit wird ein hoher Grad der Interaktion implizit vorausgesetzt, der es dem Fahrer ermöglicht, sein Anliegen flexibel vorzutragen. Dies soll durch Spracherkennung und Sprachsynthese seitens dieses Systems ermöglicht werden. Eine Erweiterung um aktuell in der Forschung befindliche Konzepte wie Gesten- und Mimikerkennung können die Interaktion später gegebenenfalls ergänzen. Ein besonderer Aspekt im Vergleich zu aktuell bestehenden Systemen im Fahrzeug (vgl. dazu Heisterkamp 2001) ist die Natürlichsprachlichkeit der Konversation. Damit soll es sowohl dem Benutzer möglich sein, Fragen in prosaischer Form ohne die Berücksichtigung einer vorgegebenen Kommandostruktur zu stellen als auch dem System in vollständigen, sprachlich optimierten Sätzen zu antworten (vgl. Shneiderman/Plaisant 2005, 332). Die Ermittlung der Antwort auf die Frage des Fahrers soll mithilfe simulierter künstlicher Intelligenz bewerkstelligt werden (vgl. Winston 1987, 186ff). Dabei greift die Logikkomponente des Systems auf eine große Datenbank an kontextsensitiven Wenn-Dann-Kombinationen zu, um die passende Reaktion zu ermitteln. Dadurch soll in einem ersten Schritt der Zugriff auf das gespeicherte Wissen ermöglicht werden. Eine Herausforderung, die in diesem Zusammenhang offensichtlich ist, stellt die Aufarbeitung und Speicherung des Wissens in der notwendigen Form dar.

Die akustische Interaktion zwischen Fahrer und System soll in einem weiteren Schritt um eine visuelle Komponente erweitert werden. Durch einen Avatar soll das Fahrzeug personifiziert werden und somit psychische Hemmschwellen bei der natürlichsprachlichen Kommunikation entfallen. Während die Verwendung von einfachen Sprachkommandos beispielsweise im Auto oder bei der Handybenutzung zunimmt (Hagen/Said/Eckert 2004; Böcker/Holz auf der Heide 2005), ist die Verwendung vollständiger Sätze im Zusammenhang der Interaktion mit einer Maschine noch kaum üblich. Die bisher bestehenden Systeme mit Spracheingabe widersprechen damit dem menschlichen Verhalten, sich semantische Inhalte anstelle von spezifischen Sprachkommandos zu merken (Anderson 1985).

Durch die Anzeige einer virtuellen Person wird die Illusion eines Gesprächspartners geschaffen, mit dem man eher bereit ist, sich natürlich zu unterhalten. Die Einbettung einer virtuellen Gestalt in einen als Cyberspace bekannten dreidimensionalen grafischen Raum ermöglicht zudem die multimediale Vermittlung von Informationen. So kann der Avatar beispielsweise die korrekte Bedienung oder Funktionsweise eines Elementes im Fahrzeug visuell darstellen.

Die multimediale Plattform soll zudem um fahrzeugspezifische technische Komponenten ergänzt werden. Diese ermöglichen den Zugriff auf aktuelle Informationen im Fahrzeug wie z.B. die Geschwindigkeit, den Ölstand oder die Drehzahl. Darüber hinaus sollen wichtige

Ereignisse wie die Anzeige einer Fehler- oder Warnanzeige an das Avatarsystem weitergeleitet werden, um darauf mit einer entsprechenden Erklärung reagieren zu können. Dies soll den Nutzen und die Benutzerfreundlichkeit für den Fahrer erhöhen, da er Informationen darüber erhalten kann, wie er auf einen bestimmten Hinweis reagieren kann bzw. was die technischen Ursachen oder Konsequenzen eines Fehlers sein können.

Die Entwicklung des prototypischen Avatarsystems innerhalb dieses Dissertationsvorhabens verfolgt nicht die Absicht, eine serienreife Plattform bereitzustellen, sondern unter Laborbedingungen ein neues Technologiebündel zusammenzustellen und technisch zu erproben. Dies impliziert, dass keine Rücksicht auf Einschränkungen bei der Ressourcennutzung hinsichtlich der späteren Verwendung in einem Auto genommen werden muss. Da die geplante Plattform eine Vielzahl verschiedener Technologien verbindet, kann es innerhalb dieser Arbeit nicht Ziel sein, alle notwendigen Komponenten selbst zu entwickeln. Sofern es möglich ist, wird dabei auf bereits bestehende Produkte und Forschungsarbeiten zurückgegriffen. Der Mehrwert dieser Arbeit liegt somit in der Zusammenstellung und Ergänzung einzelner Technologien zu einem fahrzeugbasierten Avatarsystem zur interaktiven Vermittlung von Informationen über ein Fahrzeug.

1.3. Forschungsfragen

Aus der vorgestellten Vision lassen sich folgende zentrale Fragestellungen ableiten, die in ihrer logischen Reihenfolge beantwortet werden sollen:

1. Welche Architektur eignet sich für die Umsetzung eines Avatarsystems im Auto und welche domänenspezifischen Aspekte sind dabei zu berücksichtigen?

Die Bestimmung eines Rahmenwerks für das Vorgehen bei der Implementierung bzw. der Integration bestehender Komponenten stellt den Gegenstand der ersten Frage dar. Auf der Grundlage eines ausführlichen Literaturstudiums und einzelner Integrationsprototypen soll zunächst eine Architektur erarbeitet werden, die einerseits alle Anforderungen der Domäne berücksichtigt und andererseits der Betrachtung eines Avatarsystems im Allgemeinen Rechnung trägt. Aufbauend auf bestehenden Forschungsarbeiten in anderen Kontexten können einzelne Konzepte und Komponenten wieder verwendet werden. Eine zusätzliche Herausforderung stellt dabei die Einbettung in ein automobiles Umfeld dar. Dies impliziert die Verwendung von gebräuchlichen Standards und Produkten der Branche und muss somit den Aspekt der Kapselung und Integration in die Gesamtplattform Auto berücksichtigen. Auf der Grundlage der erarbeiteten Architektur soll ein entsprechendes System im Rahmen dieser Frage auch umgesetzt werden.

2. Wie lässt sich bestehendes, strukturiertes Wissen in Form eines technischen Handbuchs in die Wissensbasis eines Avatarsystems im Fahrzeug umsetzen?

Das fertige Technologiebündel des fahrzeugbasierten Avatarsystems kann alleine noch nicht zur Anwendung kommen. Die Erstellung bzw. Adaption multimedialen Inhalts haucht der Architektur erst Leben ein. Dabei müssen entsprechend der genau zu bestimmenden

Einsatzszenarien bestehende technische Handbücher des Autos in passende Formate konvertiert, erweitert und mit einer Art Drehbuch versehen werden. Ein möglicher Ansatz zur Automatisierung kann dabei die Verwendung einer Markup Language darstellen, mit der Texte und Grafiken markiert und mit Metadaten versehen werden. Um nachträgliche Änderungen der technischen Plattform zu diesem Zeitpunkt gering zu halten, muss die Struktur und die Koordination dieses Contents bereits zur Konzeptionsphase der Architektur berücksichtigt und festgelegt werden.

3. Welche Auswirkungen sind durch den Einsatz dieses Avatarsystems im Fahrzeug auf Fahrer zu beobachten?

Die letzte der drei Fragen befasst sich mit den Auswirkungen eines solchen Avatarsystems auf das Verhalten des Fahrers vor, während und nach einer Fahrt. Dabei soll der Fokus auf der Interaktion mit dem Avatar liegen und beispielsweise ein verändertes Bedienverhalten aufdecken können. Eine zentrale Frage dabei ist zudem, ob die Verwendung des Avatarsystems zur Erweiterung der Tiefe oder Breite bei der Benutzung von Bedienelementen führt, d.h. ob der Fahrer mehr Elemente bzw. bereits bekannte Elemente intensiver benutzt. Weitere Untersuchungsbereiche sind dabei die Auswirkungen der visuellen Komponenten (Multimedia-Inhalte und Avatar selbst) auf die Aufmerksamkeit des Fahrers. Neben diesen intersubjektiv messbaren Werten spielt auch die subjektive Wahrnehmung eine wichtige Rolle. Dabei ist neben der Akzeptanz des Systems auch die wahrgenommene Benutzerfreundlichkeit sowie das Bedienkonzept zu bewerten.

1.4. Aufbau der Arbeit

Nachdem in diesem Kapitel die Problemstellung und die aus ihr gewachsene grundlegende Idee vorgestellt und in Forschungsfragen konkretisiert wurde, wird sich **Kapitel 2** mit der Wissenschaftstheorie befassen. Dabei wird zunächst an die Notwendigkeit des Bezugs einer wissenschaftlichen Grundposition herangeführt. Nach der Vorstellung möglicher Strukturierungsansätze der Grundströmungen wird auf das Wissenschaftsverständnis dieser Arbeit eingegangen. Zunächst wird die Position des Konstruktivismus und der darin anwendbare Rahmen der Design Science erklärt. Den Abschluss bildet das konkrete Forschungsdesign dieser Arbeit mit den darin verwendeten Methoden und Zielen.

Kapitel 3 stellt die Grundlagen zum Verständnis der später folgenden technischen Entwicklung vor. Dabei wird auf die Bereiche Sprachsynthese, Spracherkennung, Chatbots als semantische Spracherkennung, Avatare, Avatarsysteme und schließlich Fahrerinformationssysteme eingegangen. Die einzelnen Themen werden mit ihren Bezügen zueinander und technischen Anknüpfungspunkten betrachtet, um ihren Zusammenhang zu verdeutlichen. Die Bereiche werden begrifflich definiert sowie mit Beispielen und Einsatzbereichen veranschaulicht.

Das Umfeld der Forschung sowie die Identifikation verschiedener Stakeholdergruppen und deren Ziele werden in **Kapitel 4** beleuchtet. Dabei dienen die operativen Ziele zur Konkretisierung der Anforderungen, die aus dem Umfeld und den Forschungsfragen

erwachsen. Die ermittelten Anforderungen werden dabei den operativen Zielen zugeordnet und dienen als Rahmen für die Konzeption bzw. Implementierung.

Die Konzeption des Avatarsystem in **Kapitel 5** erfolgt in drei Schritten. Die technische Konzeption analysiert bestehende Sprachdialog- und Avatarsysteme und beschreibt die Erstellung einer Referenzarchitektur für Avatarsysteme im Allgemeinen sowie deren Anwendung für den Einsatz im Automobil. Dieser erste Teil von Kapitel 5 liefert somit die Antwort auf die erste Forschungsfrage. Der zweite Teil stellt die Probleme bei der Gewinnung und Zusammenstellung von Inhalten für das Avatarsystem dar. Abschließend werden die Anforderungen und die daraus resultierenden Implikationen für die Benutzerschnittstelle als direkt wahrnehmbare Komponente des Avatarsystems betrachtet.

Die in Kapitel 5 gewonnenen Erkenntnisse werden im Rahmen der Implementierung in **Kapitel 6** sowohl auf die technische Plattform des Avatarsystems als auch auf die dafür erstellten Inhalte angewandt. Die technische Implementierung umfasst neben der Integration des Gesamtsystems ins Fahrzeug auch die Entwicklung und Zusammenstellung der einzelnen Softwarekomponenten. Unter Vorgabe der entwickelten Avatarplattform wird in einem zweiten Teil die Aufarbeitung bestehender Inhalte aus Handbuch und Konstruktionsdaten zur Logik des Avatarsystems betrachtet. Dabei wird neben der notwendigen Toolunterstützung eine Methode zur strukturierten Erstellung einer multimedialen Wissensbasis erarbeitet. Diese Methode gibt zusammen mit dem integrierten Softwaretool die Antwort auf die zweite Forschungsfrage.

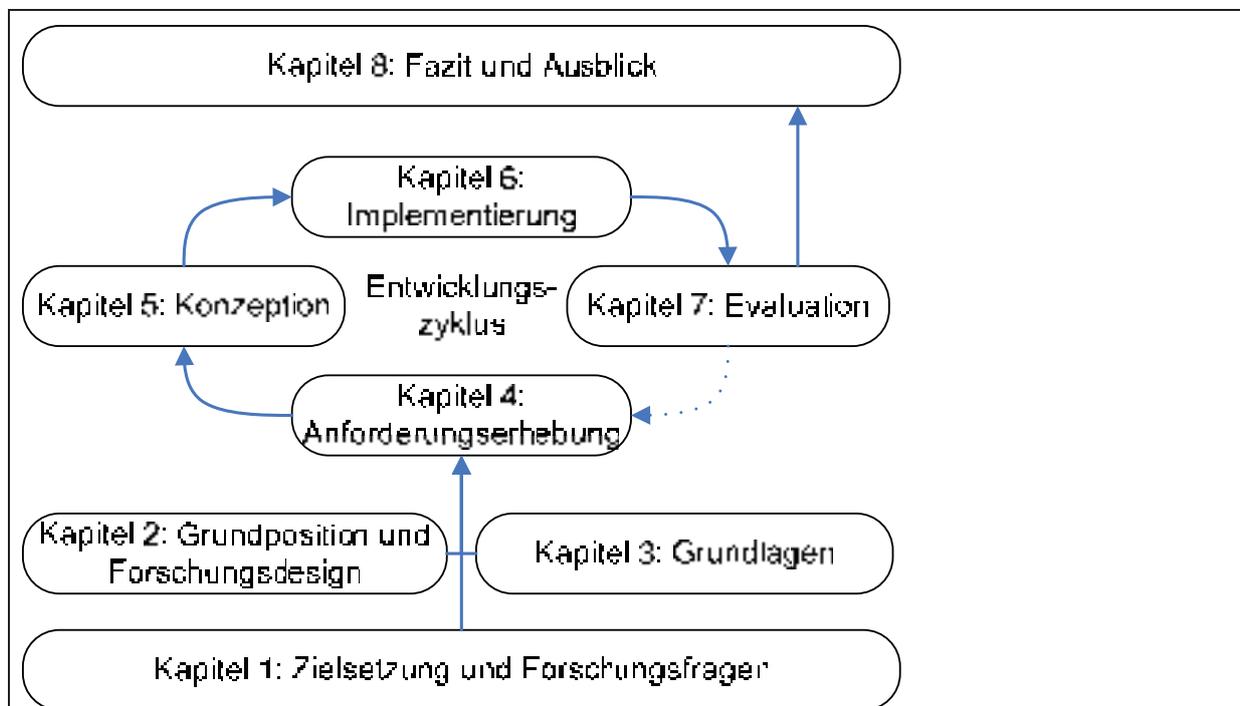


Abbildung 1-3: Ablauf der Arbeit
Quelle: (Eigene Darstellung)

Die Evaluation des in ein Versuchsfahrzeug übertragenen Avatarsystems stellt den Gegenstand von **Kapitel 7** dar. Neben der technischen und inhaltlichen Konzeption der Auswertung mit drei verschiedenen Versuchsgruppen werden dabei auch die Ergebnisse der Untersuchung dargestellt und interpretiert. Diese Darstellungen liefern damit zugleich die Beantwortung der dritten Forschungsfrage.

Das abschließende **Kapitel 8** fasst die Aktivitäten und Ergebnisse im Rahmen der Dissertation zusammen und bewertet sie kritisch. Ansatzpunkte zur Weiterentwicklung des konkreten Avatarsystems sowie allgemein zur Forschung im Bereich der Avatarsysteme in Fahrzeugen bilden den Abschluss dieser Arbeit. Eine Schematische Darstellung der Kapitel und ihrer Zusammenhänge ist in Abbildung 1-3 grafisch dargestellt.

2. Wissenschaftliche Grundposition und Methodologie

„... many arrows, loosed several ways,
Fly to one mark ...“

William Shakespeare, The Life of King Henry V

Wissenschaftliche Arbeit und ihre Ergebnisse entstehen auf der Grundlage impliziter oder expliziter Annahmen, welche die Nachvollziehbarkeit und das Verständnis durch andere als den Forscher selbst ermöglichen. Bereits Aristoteles hat dieses Verhältnis von Schlussfolgerung und Annahmen zu Syllogismen zusammengestellt, wobei zwei Prämissen die Voraussetzung einer Schlussfolgerung darstellen. Wissenschaftliche Beweise lassen sich demnach aus mehreren Syllogismen zusammenstellen (Büttemeyer 1995, 28ff).

Durch die Verbindung von Syllogismen stellt sich jedoch das Problem der Letztbegründung der ursprünglichen Annahmen solcher Beweisstrukturen. Diese als Münchhausen-Trilemma diskutierte Fragestellung kann prinzipiell auf drei Arten beantwortet werden. Die erste Möglichkeit stellt den Schluss ins Unendliche dar, der jedoch nicht praktikabel ist. Ein zweiter Ansatz wäre der Zirkelschluss, der nicht zulässig ist. Die dritte Möglichkeit, die uns schließlich zur Notwendigkeit wissenschaftlicher Grundpositionen bringt, ist die dogmatische Setzung einer Annahmestruktur als Beginn jeglichen Beweises (Albert 1968, 11). Auch wenn nach Albert (1968) all diese Ansätze nicht anwendbar sind, ebenso wie der Versuch des Baron Münchhausen, sich an den eigenen Haaren aus dem Sumpf zu ziehen, hat die Einführung von Axiomen, Hypothesen und Definitionen zur Formulierung von Grundannahmen durch Aristoteles Auswirkungen auf die moderne Forschung. Durch die Wahl einer wissenschaftlichen Grundposition und damit verbunden der Wahl von passenden Methoden wird eine Sammlung von Annahmen in die Forschungsarbeit als Grundlage integriert sowie Regeln zur Begründung der eigenen Forschungsarbeit durch andere festgelegt (Chmielewicz 1994, 6). Der Bezug einer wissenschaftlichen Grundposition ist zudem auch unter Berücksichtigung der zunehmenden neuen Forschungsbereiche und des globalen Wettbewerbs in der Forschung als ein Signal der ernsthaften Auseinandersetzung mit einem Forschungsgegenstand zu bewerten (Frank 2003).

Da diese „wissenschaftlichen Vorlagen“ auf verschiedenen Ebenen definiert wurden, sollen in diesem Kapitel zunächst in Kapitel 2.1 verschiedene Klassifizierungsansätze präsentiert und gegenübergestellt werden. Darauf aufbauend wird im Kapitel 2.2 auf den Konstruktivismus als relevante Grundposition für diese Arbeit eingegangen und der Bezug zur Forschungsarbeit dieses Buches deutlich gemacht. Als Ergebnis dieses Kapitels ist neben dem Bezug einer philosophischen Grundhaltung gegenüber dem Untersuchungsgegenstand auch die Bestimmung des Vorgehens und der dabei verwendeten Methoden im Rahmen dieser Arbeit anzusehen.

2.1. Wissenschaftstheoretische Strukturierungsansätze von Forschungsströmungen

Während die Unterscheidung verschiedener epistemologischer Standpunkte meist noch klar und disjunkt in philosophischen Fachbeiträgen vorgenommen wird, bleibt die Beschreibung

des Verhältnisses dieser Standpunkte zu passenden Methodologien und Methoden meist aus (Crotty 1998, 1). Um diesem Missstand entgegenzuwirken sind verschiedene Strukturierungsansätze mit mehreren Ebenen entstanden, die nun kurz vorgestellt werden, bevor anhand dieser Ordnung der wissenschaftliche Standpunkt dieser Arbeit bezogen wird. Es handelt sich dabei um die Sichtweisen von Crotty (1998), Guba und Lincoln (1994) sowie die speziell auf die Wirtschaftsinformatik gemünzte Betrachtung von Becker et al. (2003).

Crotty (1998, 2ff) schlägt eine vierstufige Ordnung eines wissenschaftlichen Vorhabens in die Bereiche Epistemologie, theoretische Perspektive, Methodologie und damit verbunden den konkreten Methoden vor. Dabei hängen die vier genannten Ebenen hierarchisch miteinander zusammen. Der epistemologische Standpunkt lässt sich gemäß der von Descartes propagierten Trennung von Untersuchendem (Subjekt) und Untersuchungsgegenstand (Objekt) nach Crotty in die drei Ausprägungen Objektivismus, Konstruktivismus und Subjektivismus unterscheiden¹.

In Abhängigkeit von diesem Standpunkt kann als theoretische Perspektive ein philosophischer Ansatz wie beispielsweise der Positivismus oder Interpretivismus gewählt werden. Zur Untersuchung eines bestimmten Gegenstandes oder Sachverhaltes kann im Rahmen der Grundposition und der Perspektive ein Werkzeugkasten in Form einer Methodologie bestimmt werden. Dabei kann eine Methodologie u.U. auch für verschiedene Kombinationen von epistemologischen Standpunkt und Perspektive eingesetzt werden. Die konkret angewandten Methoden ergeben demnach die Werkzeuge innerhalb dieses Kastens und können beispielsweise bei der Wahl der Methodologie „Action Research“ verschiedene Arten der Beobachtung beinhalten (siehe z. B. Baskerville 1999). Während dieser Ordnungsansatz wichtige Elemente eines Forschungsdesigns berücksichtigt und diese nach ihrer Zusammengehörigkeit gruppiert, ist keine absolute Trennschärfe bei der Definition der einzelnen Ebenen gegeben. Darüber hinaus wird der ontologische Aspekt nicht explizit berücksichtigt. Crotty (1998, 10) vertritt die Ansicht, dass die Frage nach der Ontologie untrennbar mit der Epistemologie zusammenhängt. Guba und Lincoln (1994, 108) führen als Beispiel dazu die ontologische Richtung des Realismus und den epistemologischen Standpunkt des Objektivismus an.

Einen weiteren, ähnlichen Klassifizierungsansatz stellen Guba und Lincoln (1994) vor, die jedoch explizit in Ontologie und Epistemologie unterscheiden. Sie stellen zum einen die Leitfrage nach der Natur der Realität (Ontologie) und zum anderen nach der Natur der Erkenntnis (Epistemologie). Ergänzt wird dies durch die Frage nach dem Ansatz zur Erkenntnisgewinnung und deren Verständnis (Methodologie) (Guba 1990). Eine Betrachtung von theoretischen Aspekten bzw. Paradigmen wie sie bei Crotty (1998, 2ff) erfolgt wird nicht explizit vorgenommen. Diese Aufteilung in Epistemologie, Ontologie und Methodologie wird auch von Tschamler (1996, 21f) geteilt. Weitere Strukturierungsansätze, die sich mit den Grundpositionen der Forschung insbesondere im Bereich der *Information Systems*

¹ Die Aufführung der konstruktivistischen Position als eigenständige Ausprägung in der Ebene der Epistemologie ist dabei kritisch zu betrachten, da er eine subjektive Betrachtung vertritt und somit keine grundlegend neue epistemologische Grundposition darstellt.

beschäftigen, werden bei Becker/ Niehaves (2006), Hirschheim/ Klein (1989) und Burrell/ Morgan (1979) besprochen.

Eine umfassendere und flexiblere Strukturierung nehmen Becker et al. (2003) vor, die neben der Grundposition auch die Forschungsziele als Einflussgrößen für die wählbaren Forschungsmethoden mit einbeziehen (vgl. Abbildung 2-1).

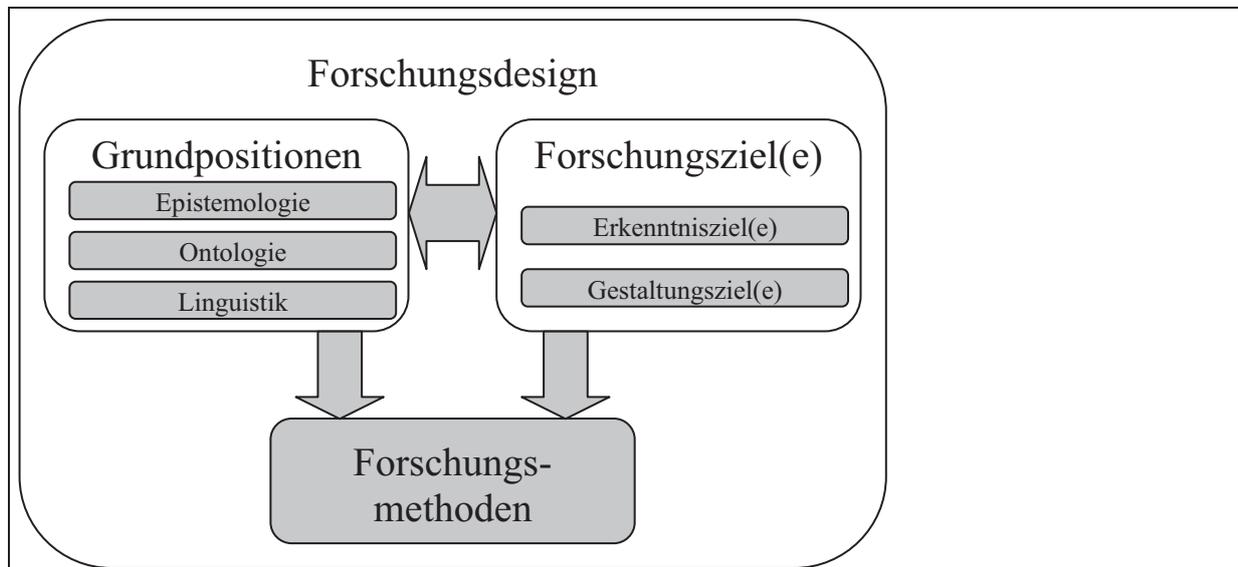


Abbildung 2-1: *Entscheidungsparameter des Entwurfs eines Forschungsdesigns*
Quelle: (Becker et al. 2003)

Sie wählen dabei zur Festlegung der wissenschaftlichen Grundposition das Konzept der Leitfragen. Sie identifizieren jedoch neben der Ontologie und der Epistemologie zusätzlich den Bereich der Linguistik. Die epistemologische Position kann dabei durch die Beantwortung der Fragen nach der Objektivität der Realität, den Quellen der Erkenntnisgewinnung und dem erlaubten Vorgehen zur Erkenntnisgewinnung (Induktion, Deduktion oder Abduktion) bezogen werden. Der Bereich der Linguistik wird neben Weiteren mit der zentralen Frage nach dem Einsatz der Sprache zur intersubjektiven Verständigung behandelt Becker et al. (2003, 6ff). Nach Tschamler (1996, 17) weist die Sprache für den Forscher eine doppelte Beziehung auf: „den Bezug zu sich selbst und die Beziehung zum Anderen mit dem gesprochen wird“ (Tschamler 1996, 17). Mögliche Forschungsziele werden spezifisch für die Wirtschaftsinformatik in Erkenntnis- und Gestaltungsziele unterteilt, wobei es bei Erstgenannten um das Verstehen und bei Zweitgenannten um die Entwicklung von Methoden, Techniken und Modellen oder Anwendungen geht. Die sich aufgrund dieser beiden Entscheidungsparameter ergebenden Methoden werden in einer für den Bereich der Wirtschaftsinformatik spezifischen Auflistung aufgeführt.

Eine Gegenüberstellung der vorgestellten Strukturierungsansätze erfolgt in Tabelle 2-1 nach teilweise gruppierten Kriterien. Aufgrund des Detailgrades und der Flexibilität der Klassifizierung von Becker et al. (2003) soll dieser Ansatz in den nachfolgenden Teilkapiteln verwendet werden, um das Wissenschaftsverständnis dieser Arbeit strukturiert aufzubauen und darzustellen.

Kriterium	Crotty	Becker et al.	Guba/Lincoln
Ontologie	Wird als fester Bestandteil der Epistemologie angesehen und deshalb nicht gesondert betrachtet	X	X
Epistemologie	X	Differenzierung in Objektivität der Realität, Quellen der Erkenntnisgewinnung und Vorgehen zur Erkenntnisgewinnung	X
Linguistik	-	X	-
Theoretische Perspektive/Paradigma	X	-	-
Methodologie	Differenzierung in spezifische Methoden	Differenzierung in spezifische Methoden der Wirtschaftsinformatik	X
Forschungsziel	-	Unterscheidung in Erkenntnis- und Gestaltungsziele	-

Tabelle 2-1: *Strukturierungsansätze wissenschaftstheoretischer Grundpositionen*
Quelle: (Eigene Darstellung)

2.2. Wissenschaftsverständnis dieser Arbeit

Mithilfe des Strukturierungsansatzes von Becker et al. (2003) als Orientierungshilfe soll nun die Grundposition dieser Arbeit, der Konstruktivismus, vorgestellt und eingeordnet werden². Dieser Ansatz ergibt sich durch den gestaltenden Charakter des in Kapitel 1.2 vorgestellten möglichen Lösungsansatzes der Problemstellung. Aufbauend auf dieser Grundposition wird zudem der Ansatz der Design Science genauer beleuchtet, der als Rahmenwerk für das artefaktgetriebene Vorgehen angesehen werden kann. Abschließend wird die konkrete Umsetzung dieser Vorgehensempfehlung für diese Arbeit erarbeitet und vorgestellt.

2.2.1. Konstruktivismus

Die Grundposition des Konstruktivismus ist Anfang des 20. Jahrhunderts aufgekommen und auch als Intuitionismus (Brouwer 1907; Heyting 1930), Operationalismus (Dingler 1931; Weyl 1990) oder Interpretivismus (Gregg/Kulkarni/Vinzé 2001, 170) bekannt. Ihre Etablierung erfuhr sie neben dem radikalen Konstruktivismus, der dem Menschen die Fähigkeit der objektiven Erkenntnis der Realität im Allgemeinen abspricht (von Glasersfeld 1992), in Wilhelm Kamlahs und Paul Lorenzens gemäßigttem Konstruktivismus der „Erlanger Schule“ (Kamlah/Lorenzen 1996; Lorenzen 1987).

Der Kerngedanke des Konstruktivismus lässt sich derart formulieren, dass jeder Wissenschaftler seine eigene, subjektive Abbildung der Realität hat. Dementsprechend müssen Erkenntnisse anderer Wissenschaftler auf Basis detaillierter Beschreibungen in der

² Ausführungen zu weiteren weit verbreiteter Grundpositionen können z.B. bei Tschamler (1996, 29ff), Chalmers (2001) oder Seiffert (1992) nachgelesen werden.

subjektiven Welt jedes Einzelnen nachvollzogen werden (Guba 1990, 25f; Büttemeyer 1995, 113f). Diese *Konstruktion* des eigenen subjektiven Verständnisses kann nur durch die Beschreibung der Erzeugung des eigentlichen Untersuchungsgegenstandes erreicht werden. Eine wichtige Voraussetzung für das Nachvollziehen einer Erkenntnis durch Andere ist dabei die Verwendung einer gemeinsamen Sprache, die zumindest häufig verwendete Ausdrücke ausreichend definiert (Lorenzen 1987, 162ff). Dadurch kann unter Bezugnahme auf eine gemeinsame Sprache die Beschreibung einer Erkenntnis genauer erfolgen. Im Bereich der Informatik sind so bereits zahlreiche konstruktive Sprachdefinitionen für verschiedene Detailbereiche entstanden (Luft 1981; Wedekind/Ortner 1980; Beckstein 1988).

Zwei grundlegende Unterschiede zu anderen wissenschaftstheoretischen Ansätzen liegen darin, dass die untersuchten Probleme aus der Gesellschaft bzw. der Lebenspraxis stammen (Janich/Kambartel/Mittelstraß 1974, 27) und das Ziel die wissenschaftliche Erarbeitung und Überführung einer Lösung in die Gesellschaft ist (Lorenzen 1987, 9). Somit haben sowohl Problemstellung als auch Erkenntnis einen Bezug zur Lebenspraxis, wodurch der Konstruktivismus sowohl eine kritische als auch eine normative Komponente enthält: Es werden reale Probleme beobachtet und Empfehlungen für deren Behebung ausgesprochen (Büttemeyer 1995, 111f).

Die Formulierung von Erkenntnissen unter Berufung auf die konstruktivistische Grundposition weist viele Parallelen zu Poppers kritischem Rationalismus auf, indem Aussagen nicht definitiv belegt sondern nur falsifiziert werden können (Popper 1994). Dieses hypothetisch-deduktive Vorgehen generiert auf der Grundlage einer Theorie singuläre Sätze, die falsifizierbar sind. Es liegt somit eine Zweiteilung der Erkenntnisgewinnung vor. Der erste Teil lässt sich auch als Abduktion bzw. Retroduktion bezeichnen, bei der eine Hypothese gebildet wird (Peirce 1976). Beim zweiten Teil handelt es sich um eine Deduktion, bei der auf Basis der gebildeten Hypothese ein Schluss auf Einzelfälle gezogen wird (Aristoteles, 20, 82 a 23-24). Während die Konstruktivisten dabei den Schwachpunkt der Theoriebildung durch den Bezug zur Lebenspraxis abzuschwächen versuchen, muss auch der Konstruktivismus damit kämpfen, dass die Theorie von den Einfällen der Forscher abhängt (Lorenzen 1987, 211). Die Erfahrung spielt an diesem Punkt in zweifacher Hinsicht eine Rolle: Sie ist der Ausgangspunkt einer konstruktiven Forschung und gleichzeitig dient sie der Überprüfung der hypothetisch angenommenen Allgemeinsätze (Büttemeyer 1995, 118). Der Erkenntnisprozess des Konstruktivismus wird von Tschamler (1996, 72) so zusammengefasst, dass die praktische Vernunft die Begründung für die reine Vernunft liefert.

Versucht man die Grundposition des Konstruktivismus der Erlanger Schule nun nach dem Strukturierungsansatz von Becker et al. (2003) mithilfe seiner zentralen Fragen einzuordnen, so ergibt sich daraus Abbildung 2-2. Aufgrund des konstruktivistischen Sprachaufbaus soll der objektive Austausch über einen Sachverhalt ermöglicht werden, auch, wenn die Rekonstruktion des beschriebenen Sachverhalts subjektiv erfolgt. Aufgrund der Ableitung singulärer Sätze auf der Grundlage einer Hypothese liegt ein hypothetisch-deduktives Vorgehen im Bereich der epistemologischen Frage vor. Dabei ist die Erfahrung, die sowohl das Problem stellt als es auch überprüft die einzige Quelle der Erkenntnis. Die Rekonstruktion beschriebener Sachverhalte für das subjektive Verständnis impliziert somit eine idealistische

Anschauung bezüglich der Epistemologie. Dies zieht sich auch in die Ontologie hinein, wonach es keine Realität unabhängig vom Erkennenden gibt (vgl. dazu auch Schütte 1999).

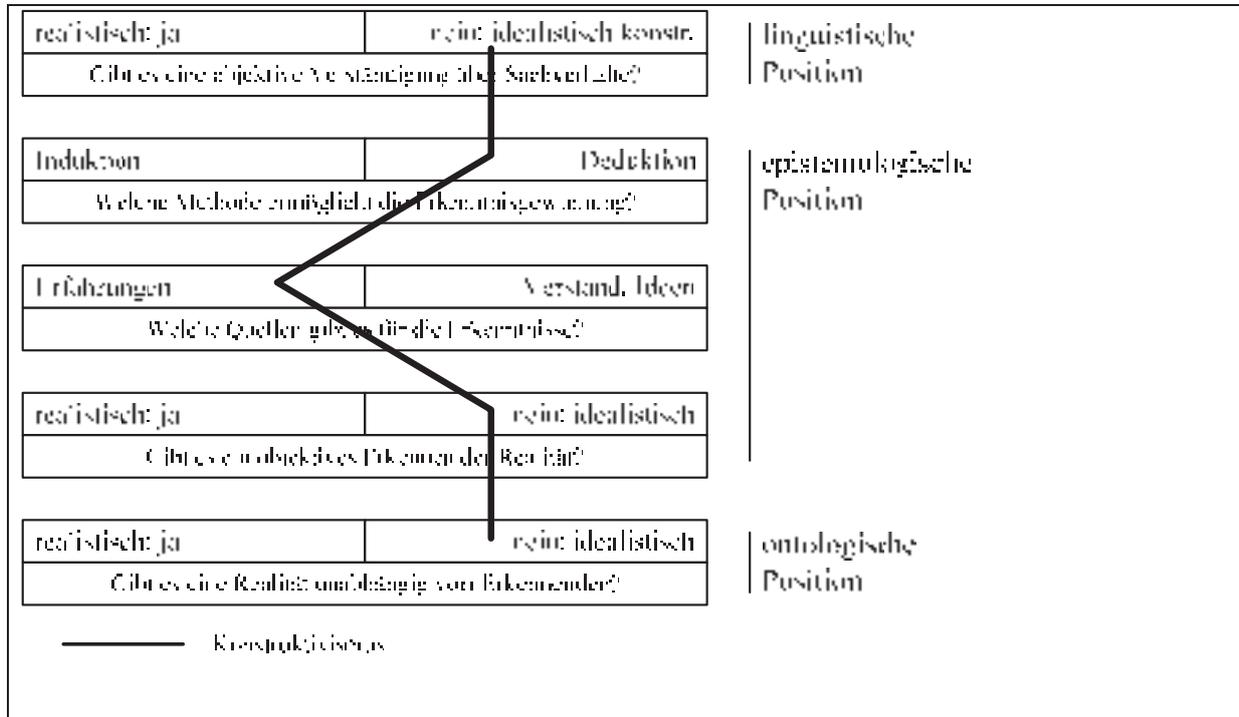


Abbildung 2-2: Beschreibung der konstruktivistischen Grundposition
Quelle: (In Anlehnung an Becker et al. 2003)

2.2.2. Design Science

Die Grundposition des Konstruktivismus gibt noch keine konkreten Empfehlungen für das Vorgehen bei der wissenschaftlichen Betrachtung eines Forschungsgegenstandes unter Bezug auf diese Grundposition. Als konkreten Rahmen für die konstruktive Forschung kann der als Design Science bekannte Ansatz verwendet werden. Auch wenn Gregg, Kulkarni und Vinzé (2001, 171ff) Design Science, welche sie als sozio-technischen Ansatz bezeichnen (Gregg/Kulkarni/Vinzé 2001, 173), auf eine Ebene mit Konstruktivismus und Positivismus stellen, handelt es sich dabei nicht um eine wissenschaftstheoretische Grundposition, zumal die reine Betrachtung von epistemologischen und ontologischen Ausrichtungen nicht ausreichend ist³.

2.2.2.1. Phasen des Designzyklus

Design Science beschäftigt sich mit dem Erstellungsprozess von künstlichen Artefakten zur Bewältigung einer Problemstellung und der Anwendung der daraus gewonnen Erkenntnisse (Simon 1996, 114f). Als Beispiel dafür lässt sich die Entwicklung der ersten Flugzeuge durch die Brüder Wright aufführen. Erst nach und durch den Bau des Flugapparats wurden die ihm zugrundeliegenden Gesetzmäßigkeiten in Form der Aerodynamik erkannt

³ Das gleiche Vorgehen ist auch bei Puroo (Puroo 2002, 8f) zu erkennen.

(Nunamaker/Chen/Purdin 1991, 632f). Grundlegende Erkenntnisse zur Design Science haben Takeda et al. (1990) beigetragen, die in einer Reihe von Designexperimenten zur Entwicklung von CAD-Software ein kognitives Designprozessmodell erarbeitet haben. Demnach werden fünf Phasen bzw. Subprozesse sequentiell durchlaufen, wobei Rücksprünge an bestimmten Stellen möglich sind. Die Phasen können in mehreren Zyklen durchlaufen werden, wenn während eines Zyklus ein neues Problem erwächst oder die Aufgabenstellung weiter fokussiert wird. Folgende Phasen wurden dabei identifiziert (Takeda et al. 1990, 12):

1. Zu Beginn des Designprozesses wird eine Problemstellung erkannt. Dies erfolgt meist durch die Identifizierung von Anforderungen, denen keine bekannten Lösungsmöglichkeiten gegenüber stehen.
2. In der Vorschlagsphase wird das erkannte Problem aufgegriffen und ein grober Rahmen möglicher Lösungswege und deren Untersuchung formuliert. Der Vorschlag dient als vorläufiges Planungsdokument für die weitere Forschung.
3. Der Kern des Designs erfolgt in der Entwicklungsphase, in der Lösungen erstellt werden, die als möglicher Lösungsansatz identifiziert wurden. Bei der Erstellung dieser Artefakte können neue Probleme erkannt werden, die ihrerseits einen neuen Designzyklus anstoßen.
4. In der Evaluationsphase wird der entwickelte Lösungsansatz bewertet, wodurch u.U. neue Probleme gefunden werden können, die wiederum eine neue Untersuchung anstoßen können. Ein wichtiger Aspekt bei der Bewertung ist das Auffinden von Kausalzusammenhängen und somit die Erklärung, wieso etwas funktioniert oder nicht.
5. In der letzten Phase wird aufgrund der Ergebnisse der Evaluation die Entscheidung getroffen, welche Aktivitäten diesem Zyklus folgen. Als Ergebnis kann ein neuer Designprozess initiiert oder das erarbeitete Artefakt für die ordentliche Verwendung aufgearbeitet werden.

Die Aufzählung der einzelnen Phasen zeigt insbesondere in der Evaluationsphase eine Problematik auf, welche durch die Anwendung der Design Science im Rahmen des Konstruktivismus entsteht. Die Evaluation für sich allein genommen vertritt die empirisch-analytische Grundhaltung, die sich auf die Beobachtung bestehender Objekte fokussiert (vgl. dazu Tschamler 1996, 47). Die Anwendung der damit verbundenen empirisch-quantitativen Methoden kann jedoch im Rahmen eines Methodenpluralismus auch mit anderen Grundhaltungen als der empirisch-analytischen erfolgen (Benbasat/Goldstein/Mead 1987; Yin 1994).

Während dieses Modell ausgeprägte Phasen für das mögliche Artefakt einer Software identifiziert, bleiben andere Ausprägungen von Artefakten eines Designprozesses außen vor. Eine Systematisierung möglicher Artefakte nehmen ergänzend March und Smith (1995, 255ff) vor, die als Ergebnis von Design Science Konstrukte, Modelle, Methoden und Instanzierungen sehen. Diese Auflistung kann zudem durch verbesserte Theorien ergänzt

werden (Purao 2002). Demnach wäre eine konkrete CAD-Software als Ergebnis des Designprozesses eine Instanzierung.

2.2.2.2. Erweiterung des Zyklus

Das ursprüngliche Modell von Takeda (1990) fokussiert sich vollkommen auf den eigentlichen Problemfindungsprozess und berücksichtigt dabei nicht, dass die gewonnene Erfahrung eines Designprozesszyklus nachhaltig verwahrt werden muss, damit sie auch bei ähnlichen Aufgabenstellungen herangezogen werden kann. March und Smith (1995) haben den eigentlichen Designprozess auf die beiden Phasen der Entwicklung und der Evaluierung beschränkt und stattdessen die Phasen der Theorisierung und der Rechtfertigung eingeführt, um der fehlenden Nachhaltigkeit des ursprünglichen Modells zu begegnen⁴. Während das Theoretisieren den Neuigkeitswert und den Einfluss auf die Domäne darstellen soll, wird im Rahmen der Rechtfertigung der Beweis versucht, dass die gewonnenen Erkenntnisse in mehreren Fällen gültig sind. Dieses Vorgehen entspricht nach der Abduktion oder Induktion während des eigentlichen Designprozesses einer Deduktion in der Theorisierung und Rechtfertigung (vgl. dazu auch Vaishnavi/Kuechler 2004). Dies kann auch als innere und äußere Umwelt der Design Science bezeichnet werden. Dabei ist die innere Umwelt der eigentliche Designprozess des Artefakts, während die äußere Umwelt die Problementstehung und die Aufbewahrung des gewonnenen Wissens beinhaltet. Der wissenschaftliche Mehrwert der Design Science entsteht somit an der Schnittstelle zwischen innerer und äußerer Umwelt (Simon 1996, 113).

Einen konsequenten Schritt weiter in dieser Ansicht gehen dabei Hevner et al. (2004), welche in ihrem Forschungsrahmen für Informationssysteme die äußere Umwelt weiter in die eigentliche Umwelt und die Wissensbasis aufteilen sowie deren Wechselwirkungen mit dem Designprozess im Mittelpunkt darstellen (vgl. Abbildung 2-3). Die Einflüsse durch das Umfeld und die Wissensbasis sind die Relevanz der Problemstellung und die Stringenz (engl.: rigor), mit der die Problemstellung wissenschaftlich untersucht werden muss (vgl. dazu auch Lee 1999)⁵. Sie unterteilen den eigentlichen Forschungsprozess in die beiden Bereiche entwickeln/bauen und rechtfertigen/evaluieren ein. Die Erkenntnisse aus dem Forschungsprozess fließen einerseits in Form der wiederverwendbaren Erfahrungen in die Wissensbasis ein, die ihrerseits bei weiteren Designprozessen herangezogen werden kann, um den Vorgang zu vereinfachen. Die Wissensbasis beinhaltet Grundlagenwissen und Methodologien. Bei der Verwendung dieser Elemente ist auf die strenge Berücksichtigung der wissenschaftlichen Stringenz zu achten. Auf der anderen Seite kommt das Artefakt der Forschung in seinem Umfeld zum Einsatz. Dieser Einsatz kann Auswirkungen auf Menschen, Organisationen und Technologie haben. Aus diesem Umfeld entstehen gleichzeitig die

⁴ Eine Erweiterung in Richtung der Phasen vor der eigentlichen Entwicklung nehmen Nunamaker und Chen (1991, 634ff) vor, die vor dem Bau und der Evaluierung eines Prototyps die Erstellung eines Konzepts, die Entwicklung einer Systemarchitektur sowie die Analyse und das Design des eigentlichen Systems sehen.

⁵ Eine kritische Bewertung der beiden Ansprüche der Relevanz und der Stringenz erfolgt bei Frank (2003), der die Gefahr von Drittmittel-finanzierter Einzelfallforschung und die Vermarktung methodisch richtiger aber wenig gehaltvoller wissenschaftlicher Erkenntnisse anprangert.

Probleme, die im Rahmen der Design Science untersucht werden sollen. Durch den Bezug der Problemstellung aus der äußeren Umwelt wird die Relevanz des Untersuchungsgegenstands sichergestellt. Obwohl Zelewski (2007, 114ff) an diesem Rahmenwerk an vielen Punkten Kritik an den Annahmen und der eingeschränkten Betrachtung der wissenschaftlichen Wirklichkeit äußert, befürwortet er dessen Verwendung dennoch als ausbaufähigen Ansatz zur Standardisierung der Forschung.

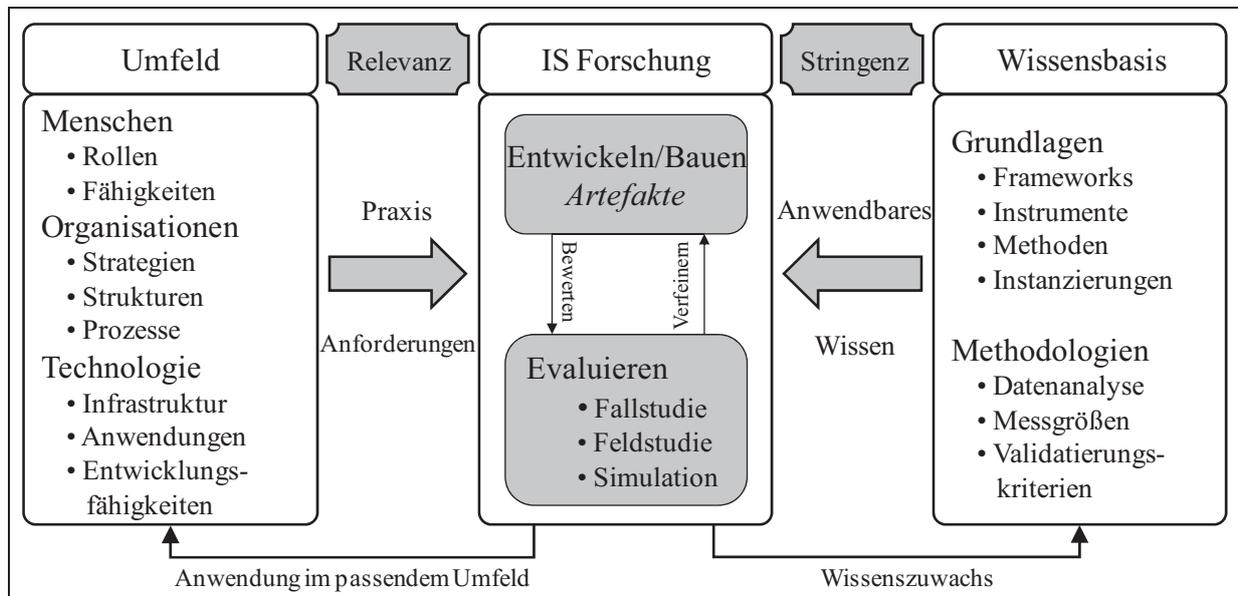


Abbildung 2-3: *Information Systems Research Framework*
Quelle: (Hevner et al. 2004, 80)

2.2.3. Forschungsdesign

Aufgrund der in Kapitel 1.3 formulierten Forschungsfragen ergeben sich zwei verschiedene Artefakte als angestrebtes Ziel dieser Forschungsarbeit. Zum einen soll ein Avatarsystem, also eine Instanzierung, erstellt werden, das nach seiner Fertigstellung evaluiert werden soll. Zum anderen soll bei der Entwicklung von Szenarien eine Methode gewonnen werden, wie bestehendes Wissen in Dialoge mit dem Avatarsystem überführt werden kann. Zur Gewinnung dieser beiden Artefakte muss nun der vorgestellte Forschungsrahmen durch die Einbindung angemessener Methoden für das konkrete Vorhaben angepasst werden. Obwohl dabei die Wahl der angewandten Methoden prinzipiell unabhängig von der wissenschaftstheoretischen Grundposition ist (Benbasat/Goldstein/Mead 1987; Yin 1994), eignen sich aufgrund des gestaltenden Charakters nur wenige Verfahren⁶. Die Verwendung des vorgestellten Frameworks von Hevner (2004, 80) soll dabei Bezugspunkte für die Identifizierung der zu berücksichtigenden Bereiche und bei der Wahl von Methoden geben. Der Einsatz der Design Science zur Entwicklung neuer Artefakte auf dem Feld der Human Computer Interaction wird dabei von Carroll und Kellogg (1989) empfohlen.

⁶ Ein Überblick über verfügbare Methoden zur Entwicklung und Evaluierung kann bei Tesch (1990, 77ff), Hevner et al. (2004, 86) oder Becker et al. (2003, 12) gewonnen werden.

Business need	Entwicklung eines prototypischen Avatarsystems als Proof of concept
Grundlagen	Verwendung von bestehenden Modellen, Rahmenwerken und Konstrukten zur Erstellung und zum Einsatz von Avatarsystemen
Entwicklungs- methode	Vertikales Prototyping: Das Avatarsystem wird als vertikaler Prototyp umgesetzt, der nach und nach um weitere Funktionen und Komponenten auf verschiedenen Ebenen erweitert wird.
Evaluations- methode	<p>Szenarioentwicklung: Zur Demonstration des Zusammenspiels aller Komponenten des Avatarsystems und als Voraussetzung für eine Evaluation des Einsatzes werden verschiedene Szenarien für den Einsatz im Fahrzeug entwickelt.</p> <p>Funktionaler Test (Black Box): Zur Sicherstellung der korrekten technischen Funktion werden bereits während der Entwicklung zahlreiche Black Box Tests durchgeführt.</p> <p>Labor- und Feldexperimente: Zur Bewertung des möglichen Einsatzes und des Verbesserungspotenzials des entwickelten Avatarsystems werden Versuche mit Fahrern in kontrollierter Umgebung sowie im Fahrzeug vorgenommen.</p>
Umfeld	<p>Menschen:</p> <p>Fahrer: Als spätere Endnutzer wird das Avatarsystem bei einem Serieneinsatz voraussichtlich Auswirkungen auf das Bedienungsverhalten haben</p> <p>Entwickler von Benutzerschnittstellen: Die Verwendung eines Avatars kann detaillierter analysiert und evtl. in die Benutzerschnittstellenentwicklung im Fahrzeug übernommen werden</p> <p>Redakteure von Chatbotdialogen: Die gewonnene Methode zur Wissenstransformation kann allgemein auch für Systeme außerhalb der Automotive Domäne verwendet werden</p> <p>Organisationen:</p> <p>Automobilhersteller und Zulieferer können neue Fahrzeuge und Bedienelemente flexibler konzipieren</p> <p>Technologie:</p> <p>Als Input: Einzelne, bestehende Technologien wie Spracherkennung, Sprachausgabe, Chatbot und Animationssoftware werden zu einem Avatarsystem zusammengestellt</p> <p>Als Output: Das Avatarsysteme mit allgemeinen und domänenspezifischen Erweiterungen fließt als potenzielle Anwendung in die Umwelt zurück</p>

Tabelle 2-2: *Anwendung des IS Research Frameworks auf die konkrete Forschung*
Quelle: (Eigene Darstellung)

Die Erstellung des Avatarsystems verwendet aus dem Bereich der Wissensbasis zunächst verschiedene Grundlagen aus bereits abgeschlossenen Arbeiten zu Avataren und benachbarten Themen. Dabei wurden verschiedene Konstrukte, Modelle und Rahmenwerke generiert, die kritisch geprüft und gegebenenfalls wieder verwendet werden können. Der eigentliche Designprozess basiert auf einem vertikalen Prototyp, der zunächst erstellt und nach und nach erweitert wird. Die Evaluation dieses Artefakts erfolgt dabei auf der Grundlage von Szenarien, die zur Demonstration des Zusammenspiels aller enthaltenen Komponenten entwickelt werden. Diese werden zunächst in einem funktionalen Test auf ihre Funktionsfähigkeit hin untersucht. Abschließend werden in Labor- und Feldexperimenten die potenziellen späteren Nutzer, also Fahrer von PKWs mit dem System und den Szenarien konfrontiert.

Eine Größe des Umfelds im Bereich der Menschen sind somit die Fahrer als potenzielle Endanwender, nach deren Anforderungen das Avatarsystem ausgelegt werden muss. Da das

Avatarsystem zunächst als Prototyp konzipiert ist, werden weitere Schritte notwendig sein, um das Konzept serienreif zu machen. Dies wird die Aufgabe der Entwickler von Benutzerschnittstellen im Automotive Bereich sein, die von der Forschung somit auch betroffen werden. Die letzte Gruppe der Menschen sind Redakteure von Dialogen für Chatbotssysteme, zumal die zu erwartende Methode ihre Arbeit besser strukturieren und damit die Wiederverwendbarkeit von bestehenden Inhalten ermöglichen kann. Aus Sicht der Organisationen sind die größtmöglichen Einflüsse bei Automobilherstellern und ihren Zulieferern zu erwarten, da die Verwendung eines Avatarsystems den Entwicklungsprozess von Fahrzeugen und einzelnen Komponenten darin verändern kann. Da das erwartete Artefakt des Avatarsystems selbst auch eine neue Technologie im automobilen Umfeld darstellt, ist im Bereich der Technologie in Technologien als Input und als Output zu unterscheiden. Der technologische Input erfolgt dabei durch bestehende Lösungen in Bereichen wie Spracherkennung und –synthese, Chatbots sowie Animation von 3D-Objekten. Das zu erstellende Avatarsystem stellt nach seiner Fertigstellung selbst wieder eine neue Technologie dar und kann somit als Output gewertet werden, der zu bestimmten Auswirkungen auf die Technologie und als Instanzierung auch auf die Wissensbasis Auswirkungen haben wird. Einen Überblick über die Eckpunkte des Forschungsdesigns gibt Tabelle 2-2.

2.3. Zusammenfassung

In diesem Kapitel wurde die wissenschaftstheoretische Betrachtung von Forschung erörtert und vom Allgemeinen zum Spezifischen hin auf diese Forschungsarbeit hin detailliert. Zunächst steht die Notwendigkeit einer wissenschaftstheoretischen Fundierung im Vordergrund, die zum einen die Grundhaltung des Wissenschaftlers charakterisiert und zum anderen die Nachvollziehbarkeit der erarbeiteten Ergebnisse durch Andere sicherstellt. Die Beschreibung verschiedener Strukturierungsansätze von Grundpositionen trägt dabei zum besseren Verständnis der relevanten Fragen und der möglichen Antworten daraus bei. Als besonders relevant stellt sich dabei der Ansatz von Becker et al. (2003) heraus, der zur weiteren Erklärung herangezogen wird. Aufgrund des gestalterischen Charakters des erwarteten Lösungsansatzes drängt sich die Grundhaltung des Konstruktivismus auf, der die subjektive Nachvollziehbarkeit von Erkenntnissen fokussiert. Als konkretisierter Forschungsrahmen auf der Grundlage des Konstruktivismus wird das Vorgehen von Takeda (1990) bzw. Hevner et al. (2004, 86) für die Design Science vorgestellt, das in einem weiteren Schritt konkret für die vorliegende Arbeit in Form eines Forschungsdesigns angepasst wird.

3. Begriffliche Grundlagen und Stand der Technik

„Wer hohe Türme bauen will, muß lange beim Fundament verweilen.“

Anton Bruckner (1824-1896), Österreichischer Komponist

Bevor der Entwicklungsprozess des Forschungsgegenstandes, also eines Avatarsystems zur Bereitstellung einer interaktiven Hilfe, detailliert beschrieben wird, sollen einige grundlegende Elemente betrachtet werden. In der Tradition der konstruktivistischen Weltanschauung kommt dabei der Definition von Begriffen eine elementare Rolle zu, da sie das gemeinsame Verständnis eines Sachverhalts erst ermöglicht (Büttemeyer 1995). Es sollen jedoch nicht nur die Definition, sondern auch der aktuelle Stand der Technik sowie Beispiele dargestellt werden. Zwischen den einzelnen Elementen Avatare, Chatbots, Avatarsysteme, Fahrerinformationssysteme, Spracherkennung und Sprachsynthese bestehen mit Blick auf das angestrebte Ziel eine Vielzahl von Zusammenhängen, die in Abbildung 3-1 visualisiert werden.

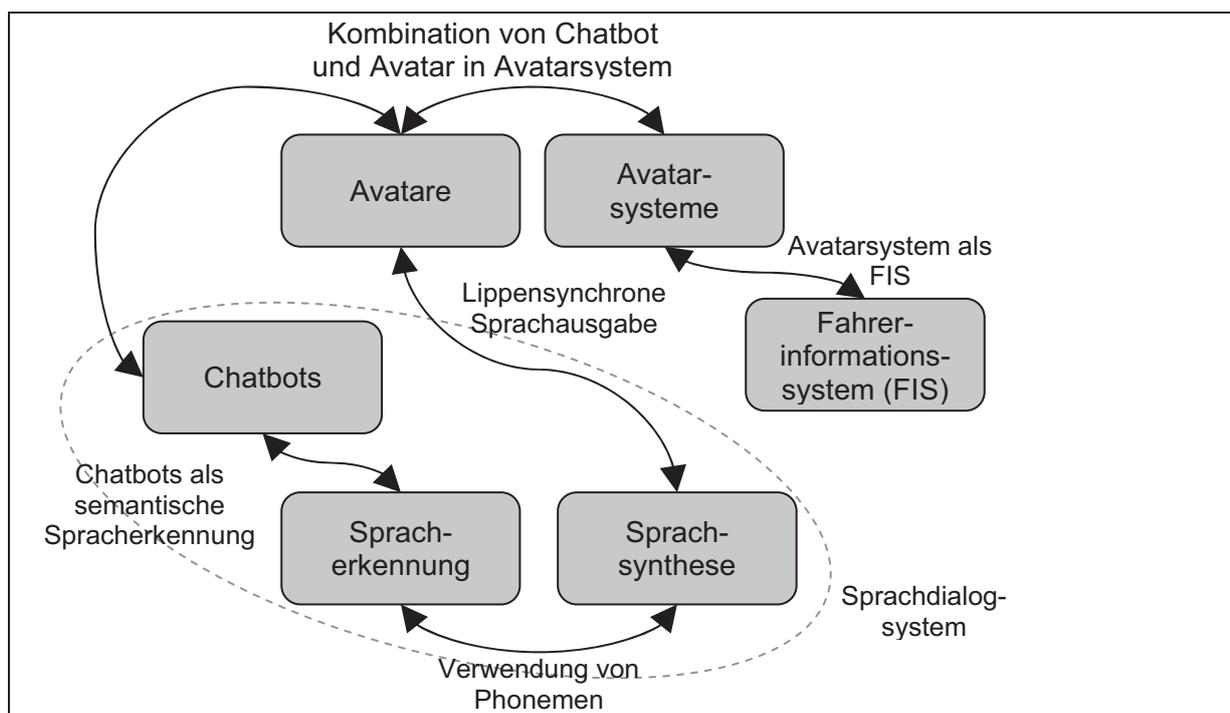


Abbildung 3-1: Zusammenhänge der Grundlagenbereiche
Quelle: (Eigene Darstellung)

So lässt sich die Verbindung von Chatbot und Avatar als Avatarsystem identifizieren, während die Kombination von Chatbot mit Spracherkennung und –synthese ein Sprachdialogsystem ergibt. Die Verwendung von Sprachsynthese zusammen mit einem Avatar beinhaltet zudem die Herausforderung einer lippensynchronen Sprachausgabe. Der Einsatz eines Avatarsystems in einem Fahrzeug lässt sich darüber hinaus dem Bereich der Fahrerinformationssysteme (FIS) zuordnen. Die Verbindungen dieser Bereiche sind somit genauso Gegenstand der Vorstellung wie die Bereiche selbst.

Da sich auf Basis der dargestellten Verknüpfungen keine eindeutige Reihenfolge der Präsentation ableiten lässt, werden zunächst die beiden Komponenten der Sprachsynthese und der Spracherkennung vorgestellt. Anschließend werden Chatbots als eine Möglichkeit der semantischen Spracherkennung und damit eng verbunden Avatare betrachtet. Den Abschluss bilden Avatarsysteme als Kombination aus Avataren und Chatbots und Fahrerinformationssysteme, zu denen sich ein Avatarsystem im Fahrzeug rechnen lässt.

3.1. Sprachsynthese

Die elektronische Synthese von Sprache kann den menschlichen Eindruck verstärken, den ein anthropomorpher Avatar allein aufgrund seines Aussehens bereits erweckt. Dabei lassen sich verschiedene Ausprägungen der Synthese identifizieren, die abhängig von den Eigenschaften der technischen Umgebung und der grafischen Ausprägung des Avatars zum Einsatz kommen können. Neben der Vorstellung verschiedener Syntheseverfahren spielt der Einsatz als Text-to-Speech-Komponente eine bedeutende Rolle, die weitere Problemfelder der Sprachsynthese aufzeigt. So kann die korrekte Aussprache, Betonung und das Sprechtempo den Effekt einer anthropomorphen Erscheinung weiter aufrecht erhalten, verbessern oder zunichtemachen. Da es nicht Ziel dieser Arbeit ist, eine neue Software für Sprachsynthese zu entwickeln, soll zunächst ein Überblick gängiger Verfahren mit dem Fokus auf Text-to-Speech gegeben und dabei insbesondere die Möglichkeiten der Optimierung der Aussprache und Betonung beleuchtet werden. Abschließend werden verschiedene Einsatzbereiche betrachtet, wobei auf das Automotive Umfeld näher eingegangen wird.

3.1.1. Begriffliche Klärung

Die künstliche Erstellung menschlicher Sprache ist bereits seit über 1000 Jahren ein Menschheitstraum, der erst 1779 durch van Kempelens mechanischen Sprachsynthesizer ansatzweise umgesetzt wurde. Mit diesem Gerät konnten 19 Konsonanten und 5 Vokale erzeugt werden. Die Elektrifizierung einer solchen Maschine erfolgte 1922 durch J.Q. Stewart. Darauf aufbauend entwickelte Homer Dudley 1939 den so genannten Voder, der schließlich in den Vocoder überging und damit starken Einfluss auf die frühe softwaregestützte Sprachsynthese ausübte. Sprachsynthese lässt sich somit als Verfahren zur künstlichen Erzeugung menschlicher Sprache verstehen (Furui 2001, 213ff). Zboril (1997) erweitert dieses Verständnis und hält folgende Definition fest:

Sprachsynthese beschäftigt sich mit der maschinellen Transformation der symbolischen Repräsentation einer Äußerung in ein akustisches Signal, welches von einem menschlichen Sprecher als der menschlichen Sprache hinreichend ähnlich anerkannt wird.

Zum Verständnis der in Kapitel 3.1.2 vorgestellten Verfahren der Sprachsynthese müssen zunächst einige Elemente und deren Zusammenhänge betrachtet werden. Das umfangreichste Element der Sprachsynthese ist ein Satz bzw. eine Phrase, welche aus mehreren Wörtern besteht. Diese wiederum werden aus Silben gebildet, die aus Lauten, so genannten Phonemen zusammengestellt sind. Dabei sind Phoneme die kleinsten lautlichen Segmente und werden in Vokale und Konsonanten unterteilt. Während Phoneme zur Betrachtung aus phonemischer

Sicht verwendet werden, lassen sich Phone unterscheiden, welche die phonetische Dimension symbolisieren. Phone sind damit die konkrete lautliche Realisierung eines Phonems (Sahel 2006, 1). Die phonemische Sequenz /papa/ (Phoneme) wird phonetisch [p^hap^ha] (Phone) notiert, was die die Aussprache des Wortes mit seinen langgezogenen Vokalen repräsentiert (Gibbon 1998). Die meisten Sprachen verwenden nicht mehr als 50 Phoneme (Furui 2001, 6). Diese geringe Zahl kommt dadurch zustande, dass Grundphoneme je nach Kontext leicht anders ausgesprochen werden. Die Variationen des Phonems werden als Allophone bezeichnet (Furui 2001, 228). Die Zusammenhänge der genannten Elemente sind in Abbildung 3-2 visualisiert.

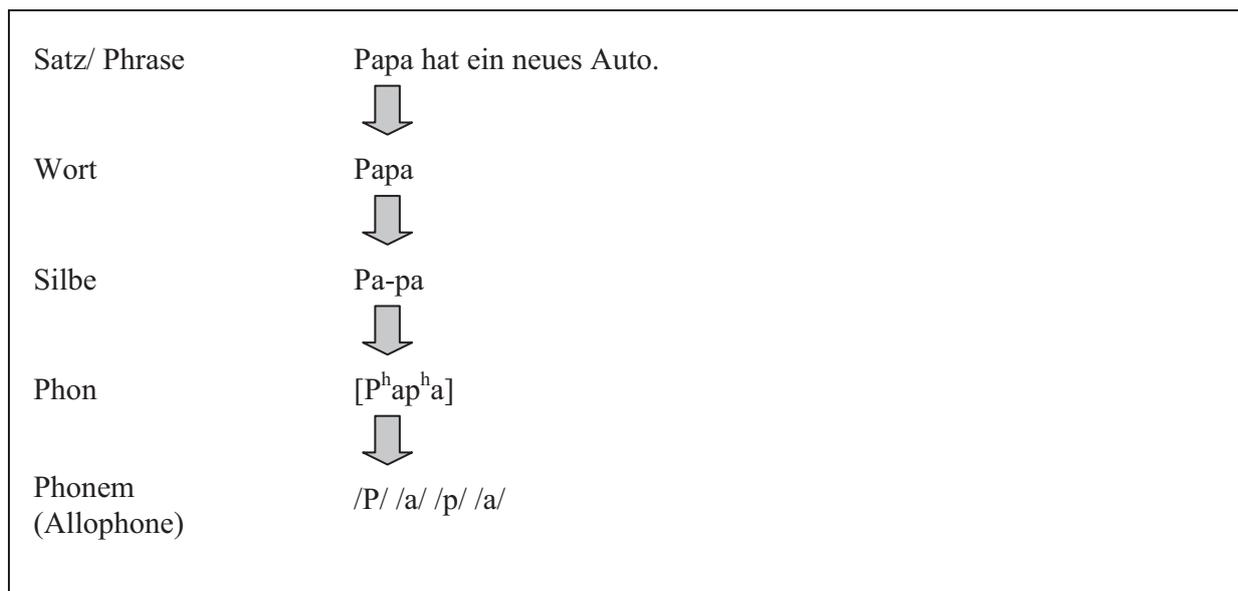


Abbildung 3-2: Zusammenhänge der phonemischen und phonetischen Elemente
Quelle: (Eigene Darstellung)

Neben diesen natürlichen Bestandteilen werden zu Zwecken der künstlichen Sprachsynthese zwei abstrakte Elemente definiert, welche die Aussprache verbessern sollen: Halbsilben und Diphone. Halbsilben stehen für den Bereich vom Beginn einer Silbe bis zu ihrer Mitte oder für den Bereich von der Mitte bis zum Ende der Silbe (Dettweiler 1981). Diphone erstrecken sich von der Mitte eines Phones zur Mitte des nächsten Phones und beinhalten somit den Übergangsbereich zwischen zwei Phonen (Schröter 2005).

3.1.2. Verfahren zur Sprachsynthese

Es lassen sich nach Dutoit (1997) und Schröter (2005) die regelbasierte und die concatenative Synthese unterscheiden. Furui (2001, 216ff) nimmt eine erweiterte Aufteilung vor und identifiziert die drei verschiedenen Verfahren des Waveform-Codings, der Analyse-Synthese und der regelbasierten Synthese. Dabei entspricht die concatenative Synthese der Analyse-Synthese bzw. in Teilen dem Waveform-Coding.

Das **Waveform-Coding** besteht nach Furui (2001, 217ff) in der analogen oder digitalen Aufzeichnung von gesprochenen Phrasen, Worten oder Silben, die zu einem späteren Zeitpunkt wiedergegeben werden können. Bei der Wiedergabe können zum Formen neuer

Sätze aufgenommene Bestandteile neu kombiniert werden. Bei dieser Konkatenation ergeben sich zwei Probleme, die mit diesem Ansatz nicht gelöst werden können (vgl. dazu auch Klatt 1987). Zum einen muss ein aufgezeichnetes Element je nach Einbettung in einem Satz anders ausgesprochen werden und zum anderen kann die Betonung der Worte nicht dem Inhalt entsprechend gesteuert werden. Diesem Nachteil versucht man mit der Aufzeichnung verschiedener Versionen von Worten und Silben entgegenzutreten. Dabei stellt sich jedoch die Frage nach der Auswahl und Verbindung der zueinander passenden Elemente. Wichtige Verfahren dabei sind z.B. TD-PSOLA (Time Domain Pitch Synchronous OverLap Add) (Moulines/Charpentier 1990) und HNM (Harmonic plus Noise Model) (Laroche/Stylianou/Moulines 1993). Diese Methoden arbeiten ohne umfangreiche Metadaten über die vorliegenden Elemente. Das Vorgehen beim Waveform-Coding entspricht der konkatenativen Synthese wie sie bei Dutoit (1997) und Schröter (2005) vorgestellt wird. Die Vorteile dieses Verfahrens liegen in der einfachen technischen Umsetzung und der hohen Sprachqualität. Aufgrund der geringen Anzahl von Berechnungen wird dabei vor allem der Speicher eines Rechners genutzt. Als Nachteil muss die geringe Flexibilität bei der Generierung neuer Phrasen angesehen werden.

Beim Vorgehen nach der **Analyse-Synthese** werden basierend auf dem Waveform-Coding bei der Aufzeichnung neben dem eigentlichen Klang weitere Parameter in Form von Metadaten aufgezeichnet. Diese Parameter ermöglichen es in begrenztem Rahmen, verschiedene Versionen von aufgezeichneten Elementen zu erzeugen und damit die Übergänge zwischen einzelnen Elementen beim Abspielen zu optimieren. Dabei kommen vergleichsweise einfache Verfahren zum Einsatz, die auf der Linear Predictive Code (LPC)-Methode basieren. Die flexiblere Handhabung von Übergängen und damit gespeicherten Elementen erlaubt auch die Verwendung von Phonemen anstelle von Worten und Silben. Die Qualität dieses Syntheseverfahrens ist hoch, wobei die Natürlichkeit unter der Parametersteuerung leidet. Die Flexibilität bei der Generierung neuer Phrasen ist höher als beim Waveform-Coding, was sich sowohl in der Komplexität des Verfahrens als auch in der Verwendung von Speicher und Prozessorleistung von Rechnern widerspiegelt (Furui 2001, 221f).

Die **regelbasierte Synthese** erlaubt die Erzeugung künstlicher Sprache nach phonetischen und linguistischen Regeln auf der Basis von Phonemen und daraus gebildeten Sequenzen. Dabei wird die Sprache aus zuvor aufgezeichneten Phonemen zusammengestellt, wobei die Möglichkeiten der Beeinflussung durch Regeln und damit Parametern aufgrund der kleineren Segmentierung der gespeicherten Elemente wesentlich größer sind. Sie ist somit die flexibelste Syntheseform, bereitet jedoch zugleich Probleme bezüglich der Qualität, da sie konzeptionell am weitesten von der reinen Sprachaufzeichnung entfernt ist. Um eine qualitativ hochwertige Sprachausgabe mit diesem Verfahren zu bewerkstelligen, müssen u.U. mehrere 10.000 Phoneme mit ihren Allophonen erfasst und mit Metainformationen versehen werden. Ebenso wie bei der Analyse-Synthese stellt auch hier der Übergang zwischen den einzelnen Phonemen eine Herausforderung dar. Dabei unterscheidet sich die Aussprache ein und desselben Phonems abhängig von umgebenden Phonemen und der Position innerhalb des Wortes. Ausgeprägte Systeme berücksichtigen dabei auch die Position innerhalb des Satzes und können dadurch teilweise automatisch Dauer, Lautstärke, Tonhöhe und Pausen für die Synthese ermitteln (Sagisaka/Tokhkura 1984). Bei der regelbasierten Steuerung der Synthese

können zwei verschiedene Paradigmen unterschieden werden: Die Formantsynthese und die artikulative Synthese. Bei der Formantsynthese wird versucht, den Klang des gesprochenen Satzes zu beeinflussen, damit er menschlich klingt. Die artikulative Synthese verfolgt das Ziel, die Synthese in Anlehnung an den menschlichen Vokaltrakt mit Parametern zu operationalisieren, um dadurch eine realistische Sprache zu erzeugen (McTear 2002, 127; Dutoit 1997) Durch die große Menge an Regeln ist dieses Verfahren sehr komplex und rechenintensiv. Die Qualität der damit erzeugten Sätze ist meist mittelmäßig. Die regelbasierte Synthese stellt jedoch eine wichtige Grundlagenforschung dar, die es in Zukunft ermöglichen kann, sehr flexibel natürliche und verständliche Sprache zu erzeugen.

Die oben vorgestellten Syntheseverfahren werden anhand der wichtigsten Kriterien in Tabelle 3-1 gegenübergestellt.

Kriterium	Waveform-Coding	Analyse-Synthese	Regelbasierte Synthese
Verständlichkeit	hoch	hoch	Mittel
Natürlichkeit	hoch	mittel	Mittel
Flexibilität der Synthese neuer Phrasen	gering	mittel	Hoch
Wortschatz	Gering (<500)	Groß (Tausende)	Unbegrenzt
Gespeicherte Elemente	Worte, Silben, Sätze	Worte, Silben, Sätze (Phoneme)	Phoneme, Silben
Komplexität	gering	mittel	Hoch
Hardwarebelastung	Speicher	Speicher und Prozessor	Speicher und Prozessor

Tabelle 3-1: *Verfahren der Sprachsynthese*
Quelle: (In Anlehnung an (Furui 2001, 218))

3.1.3. Text-to-Speech (TTS)

Die vorgestellten Syntheseverfahren können ohne ein Umfeld, das sie steuert nicht selbstständig operieren. Diese Steuerung kann entweder im Rahmen einer Anwendung erfolgen, die fest vorgegebene Phrasen ausgibt wie z.B. in einem Navigationssystem oder in flexiblerer Form durch die Übergabe eines beliebigen Inhaltes. Liegt dieser Inhalt in Form digitalen Textes vor, so spricht man von Text-to-Speech (TTS). Dutoit (1997) definiert Text-to-Speech als automatische Produktion von Sprache durch graphem-to-phonem Übertragung. Als Graphem sind die Buchstaben bzw. Silben der Wörter zu verstehen. Diese Anwendung der Sprachsynthese kommt in vielen Fällen zum Einsatz, da sie ein hohes Maß an Flexibilität ermöglicht und die Eingabe durch eine Textvariable erfolgt. Der übergebene Text kann dabei als reiner Text vorliegen oder bereits mit festgelegten Markierungen versehen sein, welche die Synthese und die Aussprache beeinflussen (Furui 2001, 236ff; Dutoit 1997). Auf diese Markierungen wird später im Rahmen der Sprachoptimierung und Prosodie in Kapitel 3.1.4 eingegangen.

Je flexibler die für TTS gewählte Synthese ist, desto mehr Informationen müssen aus dem übergebenen Text ermittelt werden, um die Parameter der Synthese zu befüllen. Somit sind verschiedene Schritte zur Normalisierung und Informationsgewinnung notwendig, bevor die eigentliche Synthese erfolgen kann. Auch wenn die Ausführung innerhalb der Phasen unterschiedlich erfolgen kann, lassen sich als Phasen die Text-, die phonetische und die prosodische Analyse identifizieren.

In der **Phase der Textanalyse** wird zunächst im Rahmen der semantischen Verarbeitung der Satzbau untersucht, wobei anhand von Satzzeichen erste Schlüsse gezogen werden können. Dabei birgt jedoch die eingeschränkte Sichtweise auf Punctuation die Gefahr, dass beispielsweise eine Kommazahl als Beginn eines Nebensatzes interpretiert wird. Ebenso ist Vorsicht walten zu lassen, wenn es um die Auflösung von Abkürzungen gilt (Taylor/Black 1997). Dies macht die Analyse des Umfeldes solcher Zeichen, u.U. also des gesamten Textes, notwendig. Zusammengesetzte Wörter werden mithilfe eines entsprechenden Wörterbuchs in ihre atomaren Worte zerlegt (Furui 2001, 235f; Schröter 2005, 16-2). Das Ergebnis dieser ersten Phase ist somit ein Text, der vollständige, atomare Wörter beinhaltet. War der übergebene Text bereits mit entsprechenden Markierungen für die spätere Aussprache versehen, werden diese auch bei der Textanalyse berücksichtigt und an die folgende Phase weitergegeben.

Die **Phase der phonetischen Analyse** hat zur Aufgabe, aus dem normalisierten Text eine entsprechende Sequenz von Phonen zu erstellen. Dabei spielt die Analyse der gesamten Phrase ebenfalls eine Rolle, da versucht wird, den Sinn des Satzes zu ermitteln. Insbesondere in Sprachen mit homographen Wörtern, d.h. Wörtern, die gleich geschrieben werden, jedoch verschiedene Bedeutungen haben, ist dies notwendig. So kann beispielsweise im Englischen „read“ die Bedeutung von „lesen“ oder „gelesen“ haben, ebenso wie „wind“ für „wickeln“ oder „Wind“ stehen kann. Nachdem die Aussprache in unklaren Fällen aufgrund des Kontextes ermittelt wurde, erfolgt die Text-to-Phonem Konvertierung, bei der auf Basis eines Aussprachewörterbuchs die passenden Phoneme für einzelne Silben und Zeichen bestimmt werden. Meist werden in einem zweiten Schritt die Übergänge zwischen Phonemen auf Basis einer gesonderten Datenbank geglättet (Allen et al. 1987; Coker 1985; Levinson/Olive/Tschirgi 1993). Alternativ kann auch die *graphem-to-phonem* bzw. *letter-to-sound* Konvertierung erfolgen, die nicht mit Wörterbüchern die Phoneme ermittelt, sondern dies direkt anhand der Buchstaben vornimmt. Dabei wird eine vergleichsweise kleine Sammlung an Ausnahmen gespeichert, die Fehler vermeiden soll. Das Ergebnis dieser Phase ist eine Sequenz von Phonemen bzw. Phonen (Hunnicut 1980; Dutoit 1997).

Die letzte Phase vor der eigentlichen Synthese ist die **prosodische Analyse**, bei der die Betonung, die Dauer, die Lautstärke und die Tonhöhe einzelner Silben und Phone bestimmt wird. Dabei dient dieser Schritt zum einen der Natürlichkeit der Sprache und zum anderen dem besseren Verständnis durch die Betonung wichtiger Wörter. Die Absicht, die in der Formulierung des geschriebenen Textes lag, wird meist erst durch die korrekte Betonung ersichtlich. Die Steuerung der Prosodie kann einerseits durch komplexe Algorithmen berechnet werden (Lieberman/Church 1992; Traber 1993; Hirschberg 1991) oder bereits mit dem übergebenen Text bestimmt werden. Verschiedene Eingriffsmöglichkeiten und damit verbundene Notationen sind Gegenstand von Kapitel 3.1.4. Das Ergebnis dieser Phase sind

die vollständigen Steuerdaten für die Synthese der künstlichen Sprache. Eine Übersicht aller genannten Phasen, ihrer einzelnen Schritte und der Zusammenhänge ist in Abbildung 3-3 zu sehen.

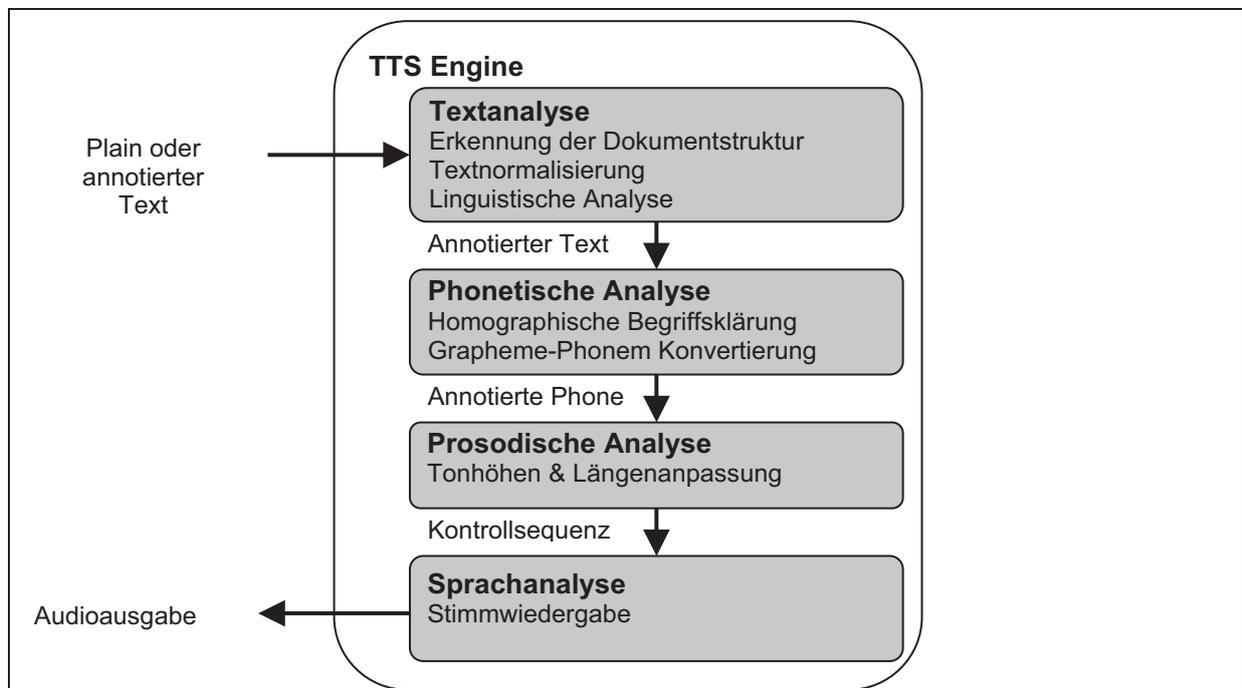


Abbildung 3-3: *Phasen und Schritte von Text-to-Speech (TTS)*
Quelle: (Schröter 2005, 16-2)

3.1.4. Optimierung der synthetisierten Sprache

Die mithilfe des beschriebenen Vorgehens synthetisierte Sprache enthält grundlegend die Informationen, die in dem eingegebenen Text enthalten waren. Dabei sind die Verständlichkeit und eine gegebenenfalls beabsichtigte Betonung einzelner Wörter meist nicht sichergestellt. Dies kann dazu führen, dass Informationen, die nur durch eine korrekte Aussprache und Betonung vermittelt werden können, verloren gehen. Um dies zu verhindern, sind verschiedene Annotationsmöglichkeiten entwickelt worden, die zum einen - beispielsweise im Falle von Eigennamen - die natürliche Aussprache und zum anderen die Betonung einzelner Wörter oder Wortsequenzen ermöglichen. Sie werden als phonetische bzw. prosodische Optimierung bezeichnet. Badzinski (1991) konnte beispielsweise nachweisen, dass synthetisierte Texte mit einer korrekten Betonung bei Kindern bessere Lernerfolge erzielt haben als nicht optimierte Sprachausgaben.

In Anlehnung an Taylor und Isard (1997, 2ff) lassen sich die möglichen Stufen, in denen ein zu synthetisierender Text eingegeben wird, nach dem Kriterium der Menschen- und Maschinenfreundlichkeit anordnen. Mit zunehmender Maschinenfreundlichkeit kann der Input einfacher und natürlicher in synthetisierte Sprache umgesetzt werden (vgl. Abbildung 3-4). Während der natürliche Text keiner Erklärung bedarf, wird im Folgenden auf die phonetische und prosodische Annotation von Texten eingegangen. Die Übermittlung von low

level Parametern wie Formant Tracks oder Linear Prediction Parameters soll dabei ausgeblendet werden.

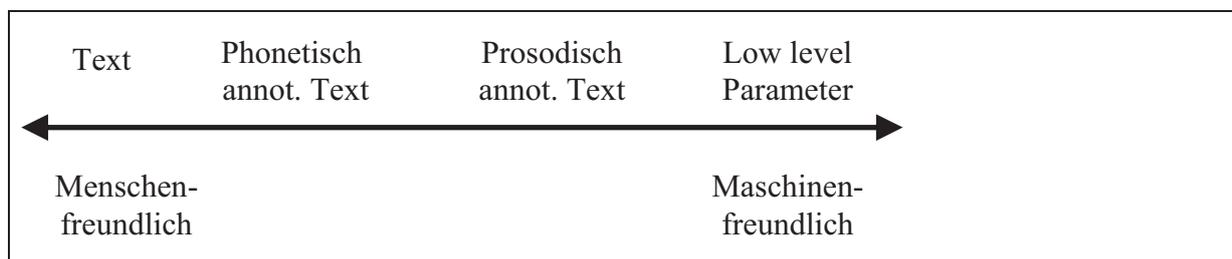


Abbildung 3-4: *Stufen der Texteingabe zur Sprachsynthese*
Quelle: (In Anlehnung an (Taylor/Isard 1997))

3.1.4.1. *Phonetische Optimierung*

Die am weitesten verbreitete Notationsform zur Transkription phonetischer Informationen von Wörtern ist das **internationale phonetische Alphabet (IPA)**. Dieses Alphabet wurde bereits 1886 definiert und stellt eine global einheitliche Lautsprache zur schriftlichen Formulierung jeglicher Sprachen dar. Sie wird vor allem in Lexika und Wörterbüchern zur Vermittlung der korrekten Aussprache eines fremden Wortes verwendet. Beispielsweise wird das griechische „Eureka“ in IPA als „Ευρηκα“ notiert (International Phonetic Association 2007). Dies zeigt zugleich einen Nachteil von IPA, der vor allem zu Beginn der computergestützten Sprachsynthese zu Problemen führte: Die IPA-Zeichen sind nicht in ASCII enthalten und können nur durch gesonderte Fonts oder mithilfe von Unicode dargestellt werden⁷. Eine Verwendung von IPA im Computerbereich war während der aufkommenden Sprachsynthese nur schwer umsetzbar, es wurde jedoch durch die Verwendung von Unicode möglich.

Ein ASCII-konformes Gegenstück zu IPA stellt das **Speech Assessment Methods Phonetic Alphabet (SAMPA)** dar, das im ESPRIT Projekt 1541 SAM (Speech Assessment Methods) in den Jahren 1987 bis 1989 sprachenabhängig entstanden ist. Dabei hat jedes IPA-Zeichen in SAMPA seine Entsprechung (Wells 1997). Im Laufe der Zeit hat SAMPA verschiedene Erweiterungen wie XSAMPA (Extended SAMPA), SAMPA-D (SAMPA für Deutsch) oder SAMPA-D-VMlex erfahren, die ein differenzierteres Zeicheninventar bereitstellen (Gibbon 1995). So wird das Wort „Eureka“ in X-SAMPA beispielsweise als „?OYRe:ka:“ geschrieben. Durch die kontinuierlich weitergeführte Entwicklung von SAMPA sowie die Maschinenlesbarkeit stellt diese Notation bzw. deren Weiterentwicklungen nach wie vor die wichtigste Notation für Ausspracheoptimierung im Computerbereich dar.

SAMPA wurde zur Basis bzw. zum Vorbild verschiedener weiterer Notationsvariationen wie etwa Phondat I und Phondat II im Rahmen des Verbmobil Projekts (Pompino-Marschall 1992; Draxler 1995), dem Bonn Open Synthesis System Label Format (BOSS LF bzw. BLF) (Breuer et al. 2001) oder dem Festival System (Taylor/Black/Caley 1998). Weitere

⁷ Die Problematik der Sonderzeichen in einer Lautschrift wurde bereits sehr früh erkannt und mit alternativen Vorschlägen zu lösen versucht (vgl. z.B. Scripture 1901).

Transkriptionen sind im Zusammenhang mit der Erstellung von Aussprachedatenbanken entstanden; so z.B. das Center for Speech Technology Research Edinburgh (CSTR) Lexicon Format oder das Carnegie Mellon University Pronouncing Dictionary Format (CMUDict) (Taylor/Black 1997). Aktuelle Aussprachedatenbanken im deutschsprachigen Raum werden z.B. von der ARD in der AusspracheDatenBank (ADB) gepflegt, wobei tagesaktuell insbesondere Eigennamen in akustischer als auch phonetisch transkribierter Version allen angeschlossenen Sendehäusern und Sprechern zur Verfügung gestellt werden (ARD 2007). Eine frei verfügbare Aussprachedatenbank ist das Bonn machine-readable Pronunciation Dictionary (BOMP), das im Rahmen der Entwicklung des Synthesystems HADIFIX entwickelt wurde und in mehreren Notationsformen wie z.B. SAMPA, BLF und Festival vorliegt (Institut für Kommunikationsforschung und Phonetik 2001). Darüber hinaus ist eine österreichische Aussprachedatenbank, die auch die deutsche Aussprache enthält, als kommerzielles Produkt im Jahr 2007 auf den Markt gekommen (ORF 2007; Muhr 2007). Zur Transkription der enthaltenen Wörter und Namen wird dabei IPA bzw. SAMPA verwendet.

3.1.4.2. Prosodische Optimierung

Die gesprochene Sprache lässt sich in den Inhalt und die Melodie eines Satzes unterteilen (Schröter 2005), wobei die zweite Eigenschaft durch die Betonung bestimmter Wörter zustande kommt. Bei der Sprachsynthese wird im Rahmen der prosodischen Optimierung versucht, die vom Autor eines Textes ursprünglich verfasste Bedeutung in die Melodie synthetisierter Sprache einzubringen. Die Prosodie bestimmt, wie ein Satz gesprochen wird, wobei neben der Melodie selbst der Rhythmus, die Betonung, die Lautstärke und Gefühle eine Rolle spielen (Childers et al. 1998; Kitahara/Tohkura 1989; Higuchi/Hirai/Sagisaka 1997; Maekawa 1999). Dabei wird zwischen dem so genannten lexical stress, der generellen Betonung eines Wortes und dem (pitch) accent bei der konkreten Äußerung unterschieden (Bolinger 1958). Die bewusste Anwendung der Betonung beim Sprechen kann einen bestimmten Teil eines Satzes hervorheben (vgl. Abbildung 3-5). Einmal wird dabei der *Zeitpunkt* betont, das zweite Mal steht das *Sehen* im Vordergrund, während das letzte Beispiel die Tatsache betont, *dass* er gesehen wurde.

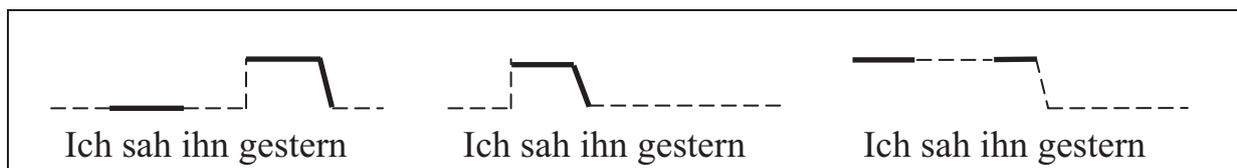


Abbildung 3-5: *Auswirkung der Betonung auf die Bedeutung eines Satzes*
 Quelle: (In Anlehnung an (Dutoit 1997))

Neben den individuellen Merkmalen natürlicher Sprecher lassen sich dabei auch allgemeine Beobachtungen machen. So beginnt ein Sprecher einen Satz meist mit einer etwas höheren Tonlage und Lautstärke, die gegen Ende des Satz abnimmt, was als Deklination bezeichnet wird (Furui 2001, 232).

Obwohl verschiedene Ansätze die automatische Generierung von Prosodie in einer Sprachsynthese versuchen, lässt sich dies nur mit hohem Aufwand erreichen (Sagisaka/Campbell/Higuch 1997). Während die Phonetik in den meisten Fällen auch ohne

Transkription verständlich umgesetzt werden kann, ist die Verwendung von Annotationen für die Prosodie wesentlich wichtiger, da die Bedeutung und damit die Betonung von Satzteilen meist nicht maschinell ermittelt werden kann (Schröter 2005, 16-3). Aus diesem Grund ist es nicht verwunderlich, dass sich im Gegensatz zu phonetischen Transkriptionen eine große Vielfalt von prosodischen Annotationen entwickelt hat, von denen nachfolgend die Wichtigsten vorgestellt werden sollen.

Die Annotation nach **Tone and Break Indices (ToBi)** wurde als eine der ersten bereits 1992 für die Prosodie des amerikanischen Englisch entwickelt und seitdem auch auf andere Sprachen angewandt (Silverman et al. 1992). Die deutsche Umsetzung von ToBi ist unter dem Namen GToBi (German ToBi) bekannt und bietet einen Zeichensatz für die Beschreibung der Tonhöhe von Zeichen und der Beschreibung von Unterbrechungen im Sprechfluss. Die Beschreibung der Tonhöhe einer Silbe wird beispielsweise mit „H“ für high pitch und „L“ für low pitch beschrieben, wobei verschiedene zusätzliche Zeichen symbolisieren, ob es sich um eine betonte Silbe handelt, eine betonte Silbe benachbart ist oder es sich dem Ende eines Satzes nähert (Reyelt et al. 1996; Grice/Baumann/Benzmüller 2007). Auch wenn ToBi als allgemein gültiger Standard entwickelt wurde und anfangs weite Verbreitung fand, ist es mit der Zeit zur Ausgangsbasis für andere Annotationen geworden (Wightman 2002). Ein Beispiel dafür, das den gesamten Symbolsatz von ToBi beinhaltet und um eigene Symbole erweitert, ist das **SAM PROSodic Alfabet (SAMPROSA)**, das ebenso wie SAMPA aus dem SAM-Projekt hervorgegangen ist. Dabei kann bei der Tonhöhe beispielsweise nicht nur in hoch und niedrig unterschieden werden, sondern es können Abstufungen eingefügt werden und die Übergänge zwischen den Silben detaillierter beschrieben werden. Darüber hinaus können Silben flexibler betont und die Länge von Silben angegeben werden (Wells 1995).

Ein Problem bei den ToBi-basierten Annotationen ist die Verwendung von kurzen, kryptischen Symbolen und damit die schlechte Lesbarkeit für Menschen. Ein alternativer Ansatz wurde ab 1997 vom Edinburgh Center for Speech Technology Research verfolgt, bei dem Text wie in HTML mit SGML-Tags versehen wurde, die für Menschen gut verständlich sind. Die erste dieser Sprachen war die **Speech Synthesis Markup Language (SSML)**, welche nur die Tags <phrase>, <emphasis>, <pronunciation>, <sound files> und <level> beinhaltet. Mit <phrase> wird dabei ein zu sprechender Satz angegeben. Die Betonung eines bestimmten Teils dieser Phrase kann mit <emphasis> erfolgen. Mit dem Tag <level> kann bestimmt werden, in welchem Maße die TTS-Software oder die SSML die prosodische Steuerung übernimmt. Mit <sound files> besteht die Möglichkeit, eine Sounddatei abzuspielen, wenn zu erwarten ist, dass die TTS das entsprechende Wort nicht korrekt synthetisieren kann. Das Tag <pronunciation> ermöglicht dabei, die Aussprache eines bestimmten Wortes im CSTR- oder CMUDict-Format anzugeben (Taylor/Isard 1997). Über den Zwischenschritt der **Spoken Text Markup Language (STML)**, die weitere Tags beinhaltet (Sproat et al. 1997), wurde die Annotation **SABLE** entwickelt, die neben den Tags auch Attribute beinhaltet, um die Prosodie zu verbessern. So wurden Tags eingeführt, um die Tonhöhe in vorgegebenen Stufen zu bestimmen, die Höhenvariation einzuschränken oder die Geschwindigkeit durch Veränderungen in Prozentwerten zu beeinflussen. Darüber hinaus können Attribute wie die Sprache, das Geschlecht und Alter des Sprechers angegeben werden (Sproat et al. 1998).

Ein ähnlicher Ansatz, der im Rahmen des Projekts Modular Architecture for Research on speech sYnthesis (MARY) entstanden ist, nennt sich **MaryXML** und unterstützt die gleichen Funktionen, wie sie durch SABLE gesteuert werden können. Dabei lassen sich jedoch die meisten Tags mit numerischen Werten versehen. Neben den Tags, die denen aus SABLE sehr ähnlich sind, können auch Symbole aus ToBi verwendet werden. Darüber hinaus können Fremdwörter markiert werden, so dass beispielsweise englische Wörter in einem deutschen Text gesondert behandelt werden können. Die Aussprache von Wörtern kann zudem mit SAMPA angegeben werden (DFKI 2007).

Abschließend sei der Ansatz von **VoiceXML** aufgeführt, der den Fokus auf die XML-basierte Beschreibung von natürlichsprachlichen Dialogen legt und dadurch auch die Prosodie thematisiert. Dabei werden vergleichbar wie in SABLE oder MaryXML Tags verwendet, mit deren Hilfe prosodische Informationen an die TTS übermittelt werden können (VoiceXML Forum 2000). Zu diesem Zweck finden Tags und Funktionen der **Java Speech Markup Language (JSML)** Verwendung (Sun Microsystems 1999).

3.1.5. Einsatzbereiche und Beispiele

Der Einsatz der Sprachsynthese bzw. insbesondere der TTS hat die Vorteile, dass der Empfänger der gesendeten Informationen kein besonderes Training zum Verständnis absolvieren muss und den Inhalt einer Phrase auch aufnehmen kann, während er mit anderen Aufgaben beschäftigt ist. Zudem ermöglicht die Sprache als Medium die Übertragung über bestehende Infrastrukturen wie Telefon oder Radio. Sie eignet sich generell jedoch nicht zur Übermittlung umfangreicher und komplexer Informationen (Furui 2001, 214). Neben dem Einsatz in der Forschung zur Verbesserung der Sprachqualität und der Algorithmen unterscheidet Dutoit (1997) folgende konkreten Einsatzbereiche (vgl. dazu auch Ngugi/Okelo-Odongo/Wagacha 2005, 81):

- Im Rahmen von Telekommunikationsdiensten wird die Sprachsynthese für automatische Ansagen, das Vorlesen von Nachrichten oder in Kombination mit Spracherkennung für komplette Sprachdialogsysteme eingesetzt (Aust 1998; Marshall 1998).
- Unter der Voraussetzung, dass die Synthese zumindest für den abgedeckten Bereich fehlerfrei funktioniert, kann sie auch für die Spracherziehung von Kindern oder die Vermittlung der korrekten Aussprache von Fremdsprachen zum Einsatz kommen (Sjölander et al. 1998; Garcia-Mateo et al. 1998; Ngugi/Okelo-Odongo/Wagacha 2005).
- Körperlich behinderten Menschen kann das Leben vereinfacht werden, indem blinden Menschen z.B. Texte vorgelesen werden oder Stumme mithilfe tragbarer Sprachsynthesysteme Texte synthetisieren lassen können (Evans/Belsey/Blenkhom 2000; Helal/Moore/Ramachandran 2001).
- Im Entertainmentbereich wird TTS für sprechende Bücher oder Spielzeuge verwendet.
- In kritischen Situationen ist die Sprachausgabe von Warnmeldungen (vocal monitoring) effektiver als eine blinkende Meldung auf einem Bildschirm darzustellen (Meng et al. 2004).

- Ein großer Einsatzbereich, der auch in dieser Arbeit thematisiert ist, stellt die Mensch-Maschine-Schnittstelle dar. So wird die Sprachsynthese beispielsweise für Navigationssysteme verwendet (Helal/Moore/Ramachandran 2001).

Die Reaktionen auf eine Sprachausgabe im Gegensatz zu einer äquivalenten Textausgabe sind nach Beobachtungen von Jensen et al. (2000) positiver bewertet worden. Ebenso wird von Hörern versucht, aus synthetisierter Sprache ähnlich viel soziale Informationen zu gewinnen wie aus realer Sprache (Nass/Lee 2001; Lee/Nass 2003). Insbesondere in der bereits zu Beginn von Kapitel 3 skizzierten Beziehung zwischen Sprachsynthese und Avataren lassen sich eine Vielzahl von Wechselwirkungen identifizieren. So lässt sich einerseits zeigen, dass die Ergänzung von Avataren durch Sprachsynthese die Erwartungen an die Menschlichkeit des virtuellen Charakters erhöht (Ball et al. 1997). Dies setzt jedoch das abgestimmte Verhalten von Avatar und qualitativ hochwertiger Sprachausgabe in Form einer Lippensynchronität voraus (Cassell et al. 1998; Lewis 1991; McGurk/MacDonald 1976; Aschenbner/Weiss 2005). Andererseits wird die sprachliche Interaktion auch in Kombination mit einer Spracherkennung bereitwilliger genutzt, wenn ein Avatar angezeigt wird (Krämer/Nitschke 2002).

3.2. Spracherkennung

Der Vorgang der Spracherkennung oder auch (Automatic) Speech Recognition (ASR) stellt den umgekehrten Prozess der TTS dar und ermöglicht dadurch die Nutzung von akustischen Sprachsignalen als Eingabekanal eines Systems. Im Folgenden wird der Begriff der Spracherkennung abgegrenzt und detailliert betrachtet. Darauf aufbauend werden das generelle technische Verfahren sowie die dafür eingesetzten Standards bzw. Grammatikformate beleuchtet. Den Abschluss der Betrachtung bildet ein Überblick über die Eignung und den Einsatz der Spracherkennung in verschiedenen Bereichen.

3.2.1. Begriffliche Klärung

Die Spracherkennung wie sie an dieser Stelle verstanden wird, beinhaltet die Konvertierung von kontinuierlichen akustischen Signalen eines Sprechers in eine Sequenz diskreter Einheiten wie Phoneme oder Wörter (McTear 2002, 103). Furui (2001, 243) detailliert diese knappe Definition folgendermaßen:

Spracherkennung ist der Prozess der automatischen Extraktion und Bestimmung linguistischer Informationen aus einer Sprachwelle unter Verwendung eines Computers oder anderer elektronischer Schaltungen.

Diese Formulierung lässt neben der Erkennung des linguistischen Inhalts auch die Erkennung des Sprechers zu (Furui 2001), wobei die Erkennung des Sprechers nicht im Fokus der weiteren Betrachtung stehen soll. Die ausschließliche Verarbeitung von Sprachwellen, also akustischen Signalen, beinhaltet zudem nicht die Spracherkennung durch Lippenlesen, wie sie beispielsweise von Potamianos et al. (2001) als Ergänzung oder von Newcomb (2008) als selbstständiger Eingabekanal vorgestellt wird. Als weiteres Abgrenzungskriterium lässt sich die Spracherkennung in einen linguistischen Bereich, also das Erkennen von Wörtern, und in

einen semantischen Bereich unterteilen, bei dem es um das Verstehen des Inhalts geht. Der zweite Bereich wird im Englischen als *Language Understanding* bzw. *Conversational Speech Recognition* bezeichnet und ist somit klar von der *Speech Recognition* unterscheidbar (McTear 2002, 103; Furui 2001, 246). Im deutschen Sprachgebrauch kann für beide Bereiche der Begriff „Spracherkennung“ zum Einsatz kommen, was zu Missverständnissen führen kann. Während in diesem Kapitel das linguistische Erkennen betrachtet wird, spielt die semantische Erkennung im Kapitel 3.3 im Rahmen der Chatbots eine Rolle.

Die erste dokumentierte Schaltung zur Erkennung von Sprache wurde 1952 von den Bell Telephone Laboratories entwickelt und konnte bereits einzelne gesprochene Zahlen mit einer Wahrscheinlichkeit von bis zu 99% erkennen (Davis/Buiddulph/Balashak 1952). Dieser erste Ansatz zur Erkennung einzelner Wörter wird in verschiedenen Bereichen nach wie vor verwendet (vgl. Kapitel 3.2.4). Im Vergleich zur Sprachsynthese, die bereits sehr früh auch ohne elektronische Geräte erforscht wurde, lässt sich erkennen, welche Komplexität beherrscht werden musste, um dieses erste System Jahrhunderte nach der ersten Sprachsynthese umzusetzen. Dies liegt auch am Umfang des Wortschatzes, den es u.U. zu erkennen gilt. So werden im Deutschen 8.000 bis 20.000 Wörter im täglichen Sprachgebrauch verwendet (Schukat-Talamazzini 1995, 6f). Die ersten Ansätze der Einzelworterkennung entwickelten sich mit der Zeit jedoch zur Erkennung von kontinuierlicher Sprache (Furui 2001, 256). Dabei lässt sich zwischen der Verbundworterkennung mit einem meist eingeschränkten Vokabular und der kontinuierlichen Spracherkennung mit beliebigen sprachlichen Äußerungen unterscheiden (Hess 2006, 143). Der Einsatz der kontinuierlichen Spracherkennung erlaubt die Anpassung der maschinellen Erkennung an den Sprecher, während sich zuvor der Sprecher an die Maschine anpassen musste (Krause 1992). Diese Entwicklung erlaubt somit die Umsetzung von Dialogsystemen in natürlicher Sprache für eine Vielzahl verschiedener Sprecher und kreiert für den Nutzer den Eindruck einer realen Gesprächssituation (Görtz 2006, 25).

Insgesamt lassen sich unabhängig davon, ob einzelne Wörter, Sequenzen bekannter Wörter oder beliebige Phrasen erkannt werden sollen, verschiedene Herausforderungen identifizieren, denen ein Spracherkennungsverfahren begegnen muss. Unter der linguistischen Variabilität lassen sich Aspekte wie die Koartikulation von Lauten, d.h. die gegenseitige Beeinflussung von Lauten in der Aussprache, erkennen (McTear 2002, 103). Neben diesem linguistischen Aspekt ist auch die Veränderung in der Sprache eines Sprechers zu betrachten. So kann ein und derselbe Sprecher eine Phrase in verschiedenen emotionalen und umweltbeeinflussten Zuständen oder mit einer veränderten Satzstellung äußern (Haas et al. 2004, 4; McTear 2002, 103). Die Berücksichtigung dieser Eigenheiten wird noch komplexer, wenn ein System sprecherunabhängig arbeiten kann. Dies ermöglicht es, die gesprochene Sprache beliebig vieler Sprecher zu erkennen, ohne das System auf die individuellen Sprecher zu trainieren (Haas et al. 2004, 3; McTear 2002, 104; Furui 2001, 246f). Die Erkennung kann jedoch auch durch technische Beschränkungen wie beispielsweise der Verwendung eines Telefons mit geringer akustischer Bandbreite oder eines schlechten Mikrofons beeinflusst werden (Haas et al. 2004, 5). Dabei muss zwischen den genannten konstanten Einschränkungen und zeitweisen Effekten wie einem kurzzeitigen Rauschen unterschieden werden (McTear 2002, 103f).

Spezifische Probleme ergeben sich zudem abhängig von der Art der Spracherkennung. So muss bei einer Spracherkennung von Wörtern je nach Einsatzzweck ein Wörterbuch vorliegen, das mehrere tausend Begriffe erkennen kann, wobei sich die Größe des Wörterbuchs negativ auf die Leistung der Erkennung auswirkt (McTear 2002, 104). Die uneingeschränkte kontinuierliche Spracherkennung setzt ein linguistisches Wissen voraus, das neben Wörterbüchern vor allem grammatikalische Regeln enthält, wie eine Phrase in der entsprechenden Sprache aufgebaut sein kann (Haas et al. 2004, 4). Eine weitere Herausforderung bei der kontinuierlichen Erkennung ist die Bestimmung des Anfangs und des Endes einer Phrase. Dieser als *Speech Period Detection* bezeichnete Vorgang wird vor allem durch Nichtsprachsignale wie Husten, dem Umblättern einer Seite oder „Ähms“ beim Sprechen erschwert (McTear 2002, 104; Furui 2001, 248f). Eine kontinuierliche Spracherkennung muss daher eine gewisse Fehlertoleranz aufweisen, um die linguistischen Inhalte identifizieren zu können.

3.2.2. Funktionsweise der Spracherkennung

Der technische Ablauf der Spracherkennung lässt sich unabhängig von Wort- oder kontinuierlicher Erkennung in die vier groben Phasen von Vorverarbeitung, Klassifikation, Sprachmodellierung und Suche unterteilen (Haas et al. 2004, 2). Da die Suche Teil der Sprachmodellierung ist, ergeben sich zusammengefasst drei Phasen. Hess (2006, 11) identifiziert bei der Verwendung eines sequentiellen bottom-up-Ansatzes weitere Phasen, die im Folgenden als Teilschritte der Phasen miteinbezogen werden (vgl. Abbildung 3-6)⁸. Unterschiede zwischen den verschiedenen Typen der Spracherkennung lassen sich insbesondere bei der Verarbeitung in der Klassifikations- und Suchphase erkennen, weshalb ihr besondere Beachtung geschenkt werden soll. Das Ziel der Betrachtung der einzelnen Phasen soll das Verständnis der Vor- und Nachteile der Erkennungstypen sein und somit den notwendigen Hintergrund für spätere Designentscheidungen liefern.

Die erste Phase der **Vorverarbeitung bzw. Merkmalsberechnung** beinhaltet neben der Digitalisierung von Schallwellen die Generierung von Metadaten über das erfasste Signal. Dazu wird das Signal bereits bei der Erfassung in Abständen von beispielsweise 10 Millisekunden abgetastet. Bei jeder Abtastung wird ein so genannter Merkmalsvektor bzw. Parametervektor erfasst, der mehrere Parameter wie z.B. die Amplitude oder Werte für die Modellierung des menschlichen Vokaltrakts beinhaltet. Die Berechnung kann dabei z.B. über eine Folge von Fouriertransformation, Integration und Transformation in den Cepstralbereich oder eine lineare Vorhersage im Rahmen eines autoregressiven Modells stattfinden (Schukat-Talamazzini 1995, 46ff). Diese Parameter ermöglichen eine schnellere Verarbeitung des Sprachsignals in den folgenden Phasen, da durch die Erhebung der Vektoren eine Datenreduktion mit dem Faktor 100 oder mehr stattfindet (Hess 2006, 12f). Durch die Erhebung von Zeitbereichsmerkmalen können zudem bereits erste Angaben für die folgende Segmentierung gesammelt werden (Schukat-Talamazzini 1995, 51f).

⁸ Der erwähnte bottom-up-Ansatz ist in der Praxis ungünstig, da Fehler in frühen Phasen nicht mehr korrigiert werden können. Er zeigt jedoch am deutlichsten die einzelnen Verarbeitungsschritte auf. Alternative Ansätze sind beispielsweise das hierarchische Modell, das Blackboard- und das Netzwerkmodell (Furui 2001, 308ff).

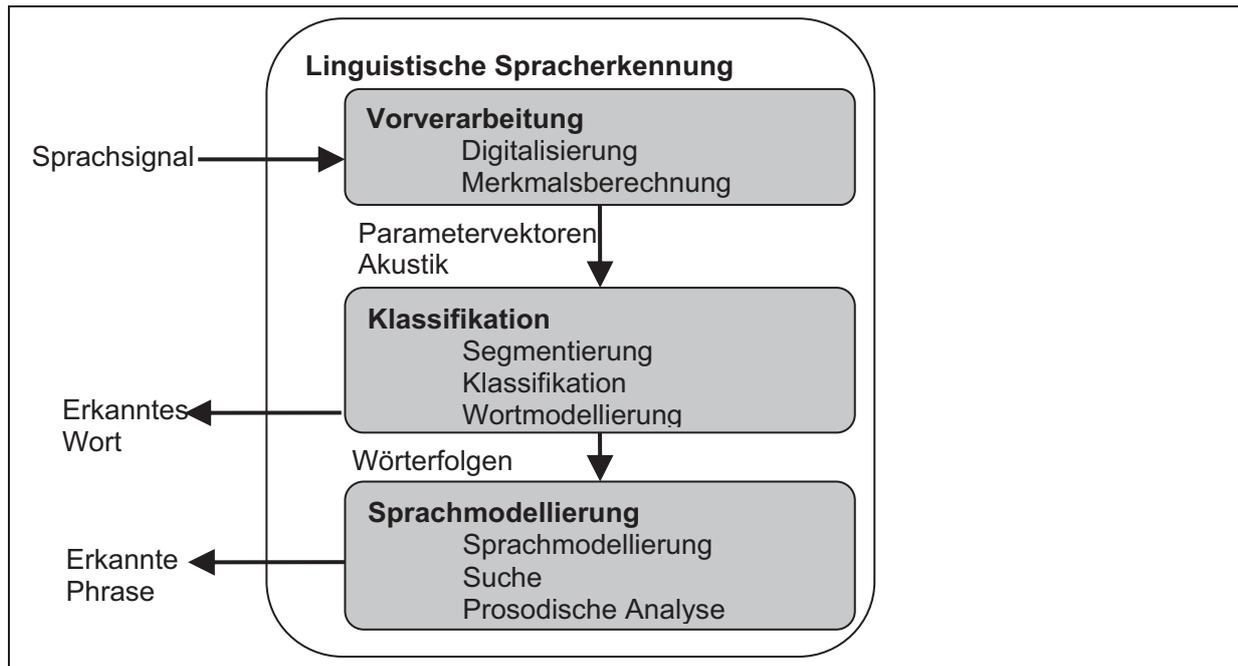


Abbildung 3-6: Phasen und Schritte der linguistischen Spracherkennung
Quelle: (Eigene Darstellung)

Der erste Schritt im Rahmen der **Klassifizierung** ist die Segmentierung der Parametervektoren, d.h. die Erkennung von einzelnen Einheiten wie Phone oder ganze Wörter. Dabei lässt sich eine explizite Segmentierung unterscheiden, die vor der eigentlichen Klassifizierung abläuft, und der häufiger verwendeten impliziten Segmentierung, die in die Klassifizierung integriert ist.

Die eigentliche Klassifizierung hat die Aufgabe, die erfassten Vektoren einzeln oder in Sequenzen einer bereits gespeicherten Klasse wie z.B. einer Silbe oder einem Wort zuzuordnen. Dies erfolgt je nach Wort- oder kontinuierlicher Erkennung meist durch Dynamic Time Warping (DTW) oder Hidden-Markov-Modelle (HMM) (Haas et al. 2004, 2). Beim erstgenannten Verfahren wird anhand des Abstands zwischen den Vektoren und den gespeicherten Mustern das Muster bestimmt, welches das geringste Distanzmaß hat. So lässt sich dem Problem der Längenunterschiede von Vektor und gespeichertem Muster begegnen. Dieses Verfahren wird heutzutage nur noch zur sprecherunabhängigen Einzelworterkennung verwendet (Schukat-Talamazzini 1995, 122ff). Die Verwendung eines Hidden-Markov-Modells ermöglicht eine flexiblere Art der Erkennung. Dabei werden beispielsweise einzelne Phoneme als Zustände festgelegt. Aufgrund der variablen Aussprechdauer eines Phonems wird für jeden Merkmalsvektor d.h. für jede erfolgte Abtastung die Wahrscheinlichkeit errechnet mit der es sich noch um das gleiche Phonem handelt oder ein anderes Phonem vorliegt (Schukat-Talamazzini 1995, 125f). Dieses Verfahren kommt insbesondere bei der kontinuierlichen Spracherkennung zum Tragen. Beiden Verfahren liegt ein statistisches Modell auf der Basis eines Lernprozesses zu Grunde.

Auf der Basis der mit bestimmten Wahrscheinlichkeiten ermittelten Phoneme wird im letzten Schritt dieser Phase die Identifizierung möglicher Wörter vorgenommen. Dieser Schritt ist

notwendig, da die Klassifizierung auf Basis des Sprachsignals mehrere passende Phoneme ermitteln kann. Da für eine sprecherunabhängige Suche nach ganzen Wörtern sehr viel Speicher- und Rechenressourcen aufgewendet werden müssen, ist man dazu übergegangen, Wortteile (Subwords) bzw. Triphone (Gruppierung von drei Phonemen) zu speichern (Furui 2001, 265). Die Auswahl der möglichen Wörter erfolgt auf Basis eines Wörterbuchs in der Wortmodellierung (Schukat-Talamazzini 1995, 167ff; Haas et al. 2004, 2). Im Falle einer Einzelworterkennung endet der Prozess an dieser Stelle.

Die **Sprachmodellierung** ist die letzte der drei Phasen und bestimmt anhand von Wahrscheinlichkeiten die korrekte Wortreihenfolge. Dadurch soll sichergestellt werden, dass beispielsweise statt der syntaktisch korrekten aber semantisch sinnlosen Wortfolge „Säge ehrte Frau schnitt“ die Phrase „Sehr geehrte Frau Schmidt“ erkannt wird (Haas et al. 2004, 2). Die Suche nach semantisch korrekten Wortfolgen nimmt mit der Länge der zu erkennenden Sequenz exponentiell zu. Dabei können verschiedene Suchverfahren wie Viterbi (Schukat-Talamazzini 1995, 132ff), die Strahlsuche (beam search) (Lowerre 1976; Schukat-Talamazzini 1995, 240ff) oder der A*-Algorithmus (Schukat-Talamazzini, 1995 #624@245ff) angewandt werden. Intern werden die möglichen, alternativen Sequenzen in Form eines Worthypothesenbaums abgebildet (vgl. Abbildung 3-7) (Wyrd et al. 1996, 196; vgl. dazu auch Pfau/Ruske 1996).

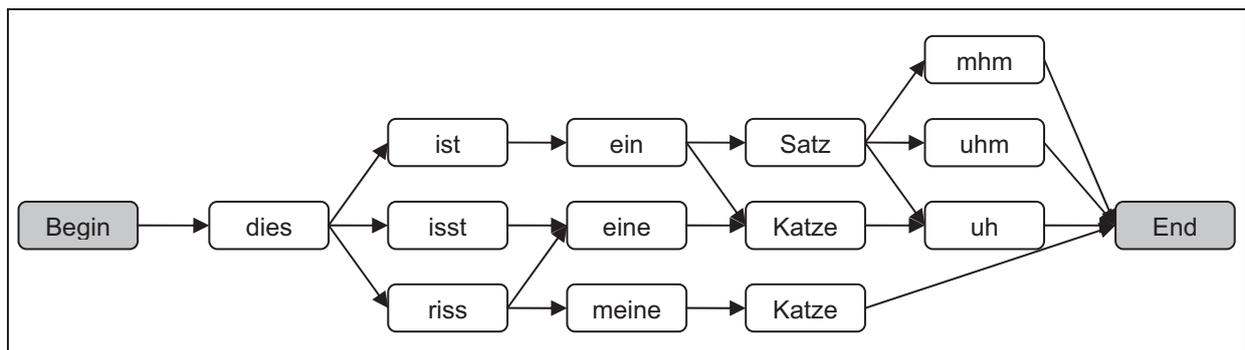


Abbildung 3-7: *Beispiel eines Worthypothesengraphs*

Quelle: (Carstensen et al. 2004, 580)

Die mit der höchsten Wahrscheinlichkeit ermittelte Wortfolge wird schließlich als Ergebnis der Suche zurückgegeben, wobei dies ohne Punkt und Komma erfolgt. In einer prosodischen Analyse wird schließlich durch die Identifizierung akzentuierter Silben sowie Phrasen und Satzgrenzen versucht, den ursprünglichen Inhalt des Sprachsignals mit Satzzeichen wieder herzustellen (Hess 2006, 196ff). Als Ergebnis der Spracherkennung wird im Falle der kontinuierlichen Erkennung somit eine Phrase mit Satzzeichen ausgegeben.

3.2.3. Sprachgrammatiken

Zur Erkennung von Wortfolgen werden Grammatiken herangezogen, die entweder bereits in einem ASR-System integriert sind oder von außerhalb eingespeist werden. Diese nachträglich geladenen Sprachgrammatiken können die bereits bestehenden Regeln ergänzen oder überblenden. Die Verwendung eigener spezifischer Grammatiken erhöht die Wahrscheinlichkeiten der Erkennung der zusätzlich eingespeisten Phrasen. Vor der

Verwendung einer Grammatik in einer Spracherkennungssoftware erfolgt die Konvertierung in ein internes, proprietäres Format in einer herstellerabhängigen Kompilierung. Zur Definition der Sprachgrammatiken können verschiedene standardisierte Formate zum Einsatz kommen. Im Folgenden sollen die Formate der Backus-Naur-Form (BNF) bzw. BNF+, des Java Speech Grammar Formats (JSGF) und der Nuance Grammar Specification Language (GSL), vorgestellt werden. Als weitere Formate mit geringerer Verbreitung können darüber hinaus beispielsweise die Whisper Speech Grammar (Ball et al. 1997) oder proprietäre Formate von Sprachdialogsoftwareherstellern genannt werden (Speech Experts 2003).

Die **Backus-Naur-Form (BNF)** bzw. Backus-Normalform bezeichnet eine formale Metasprache zur Definition kontextfreier Grammatiken. Sie wird zur Beschreibung der Syntax höherer Programmiersprachen genauso verwendet wie für die Notation von Befehlssätzen oder Kommunikationsprotokollen. Die ursprüngliche Konzeption dieser Sprache erfolgte für die Programmiersprache Algol 60 und erlebte seitdem zahlreiche Weiterentwicklungen wie z.B. in der Extended BNF (EBNF) oder der augmented BNF (ABNF) (Braun 2007). Ein wichtiges Element der Sprache ist die Definition von Mengen und deren Verschachtelung unter Verwendung logischer Operatoren (Knuth 1964). Abbildung 3-8 zeigt ein Beispiel, in dem zunächst die Gruppe der Ziffern definiert wird und in einer zweiten Regel für die Bildung einer Telefonnummer referenziert wird. Dabei werden zumindest eine führende Null, mehrere Ziffern, ein optionaler Schrägstrich und eine weitere Ziffernfolge erwartet. Dieses Beispiel kann bereits für die Definition einer Regel zur Spracherkennung von Telefonnummern zum Einsatz kommen.

```
<Ziffer> ::= 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9.
<Telefonnummer> ::= „0“ <Ziffer>+ [ „/“ ] <Ziffer>+
```

Abbildung 3-8: *Beispiel einer BNF-Grammatik*

Quelle: (Eigene Darstellung)

Eine spezifische Erweiterung von BNF erfolgte durch **BNF+**, das insbesondere vom Softwarehersteller Nuance (vormals Scansoft) eingesetzt wird. Dabei werden einige Vereinfachungen bei der Definition der Grammatik selbst vorgenommen, wie z.B. das Weglassen von Anführungszeichen bei Alternativen innerhalb einer Regel oder die Verwendung eines einfachen Gleichheitszeichens bei der Zuweisung von Alternativen zu einer Regel. Zudem können Grammatiken auf mehrere Dateien aufgeteilt und ganz oder in Teilen nach außen bereitgestellt (exportiert) und verwendet (importiert) werden. Auch lässt sich statt der Verwendung geschriebener Texte eine ASCII-Notation von IPA (vgl. Kapitel 3.1.4.1) einsetzen. Zwei weitere Neuerungen sind zum einen eine so genannte <NULL>-Regel, die aufgerufen wird, wenn kein Text gesprochen wurde sowie die interne Erweiterung des erkannten Inhalts. Dies ermöglicht es z.B. bei der Erkennung einer Telefonnummer das Signalwort „Telefonnummer:“ vor die eigentliche Nummer zu stellen und damit die spätere Verarbeitung der Eingabe zu vereinfachen (ScanSoft 2004).

Ein offenes Format, das alle Möglichkeiten von BNF+ beinhaltet und um weitere ergänzt, ist das **Java Speech Grammar Format (JSGF)**. Über die bereits für BNF+ genannten

Eigenschaften hinaus lassen sich Regeln als öffentlich oder nicht öffentlich deklarieren. Dies führt dazu, dass eine nicht öffentliche Regel nur durch andere Regeln referenziert werden kann, während öffentliche Regeln direkt zur Bestimmung der gesprochenen Wortfolge herangezogen werden können. Eine zweite Erweiterung erfolgte hinsichtlich der Gewichtung von Alternativen innerhalb einer Regel. So können Alternativen, die erwartungsgemäß häufiger vorkommen, entsprechend markiert werden und somit die Identifizierung der korrekten Wortfolge vereinfachen (Sun Microsystems 2007; Andersson et al. 2001, 45ff). Das JSGF wird beispielsweise vom Softwarehersteller Loquendo für seine Spracherkennung verwendet.

Die **Nuance Grammar Specification Language (GSL)** ist eine vollkommen objektorientierte Annotation von Grammatiken, welche die gleichen Möglichkeiten der bisher vorgestellten Formate bereitstellt. Dabei wird jedoch eine von BNF abweichende Verwendung von Symbolen verwendet (Andersson et al. 2001, 51ff). Zudem erfolgt die Erstellung der Grammatik aufgrund der Komplexität im objektorientierten Ansatz in einem gesonderten Softwaretool namens Nuance Grammar Builder. Dieses Programm erlaubt auch integrierte Tests der erstellten Grammatik (Nuance 2001). Sowohl JSGF als auch GSL können im Rahmen des in Kapitel 3.1.4.2 vorgestellten Formats VoiceXML eingebunden werden, um mögliche Benutzereingaben in Form von Sprachen oder das Drücken von Tasten am Telefon zu modellieren (Nuance 2001, 45ff; VoiceXML Forum 2000, 35ff).

3.2.4. Einsatzbereiche und Beispiele

Die Verwendung einer Spracheingabe als Eingabekanal bringt eine Vielzahl von Vorteilen mit sich. So ist zu ihrer Bedienung kein Training erforderlich und sie ist bis zu viermal schneller als die Benutzung einer Tastatur. Sie erlaubt zudem die Verrichtung einer weiteren Tätigkeit mit den Händen (Furui 2001, 244; Badler 1997, 8). Shneiderman (2000, 64) sieht die Verwendung von Spracherkennung überall dort als sinnvoll an, wo überwiegend mechanisch gearbeitet werden kann und nicht über die Tätigkeit selbst nachgedacht wird. Er sieht darin auch den Grund, dass sich die Spracheingabe für die Arbeit am PC bislang nicht durchsetzen konnte. Diesen Sachverhalt bestätigen auch Franz, Zelaznik und Smith (1992), die einen Zusammenhang zwischen dem Bewegungsrhythmus und der Sprechgeschwindigkeit nachgewiesen haben (vgl. dazu auch Bengler 1995; Färber 1987). Bartmann (1995) zeigt darüber hinaus, dass in Phasen in denen sich ein Benutzer auf ein bestimmtes Ereignis konzentrieren muss, die Fähigkeit zur Bedienung eines sprachgesteuerten Systems negativ beeinflusst wird. In einer Feldstudie wies sie dies mit Autofahrern in beanspruchenden Verkehrssituationen nach.

Die möglichen Anwendungsgebiete der Spracherkennung werden jedoch auch durch die Problemfelder der Technologie eingeschränkt. So führt der Einsatz in einer geräuschvollen Umgebung oft nicht zu den gewünschten Ergebnissen. Dem damit verbundenen Problem der Erkennung des Sprechbeginns kann entweder mit neuen Algorithmen (Singh/Boland 2007) oder mit der Verwendung eines Push-to-Talk-Buttons (Ball et al. 1997) begegnet werden. Ebenso stößt die Spracheingabe bei der Erkennung stark voneinander abweichender Sprechstile und Dialekte an ihre Grenzen (Furui 2001, 244f). Je nach Anwendungsfall ist eine Spracherkennung für Kommandos oder kontinuierliche Sprache angebrachter (Van

Compernelle 1997). So wird z.B. die Spracheingabe in ganzen Sätzen bei Sprachdialogsystemen als besonders wichtig angesehen (Peissner/Sell/Steimel 2006, 23).

Haas et al. (2004, 8f) identifizieren vier konkrete Einsatzbereiche für Spracherkennung: Diktiersysteme, Datenerfassung, Sprachdialogsysteme und Sprachsteuerung.

- Die Anwendung als **Diktiersystem** hat sich bisher in den Bereichen durchsetzen können, die häufig wiederkehrende Elemente aufweisen, wie beispielsweise beim Diktieren von Briefen oder Befunden. So wurden in Baden-Württemberg ca. 1000 Richterarbeitsplätze mit entsprechender Software zum Diktat von Urteilen eingerichtet {Haas, 2004 #562@8}. Darüber hinaus verwenden auch medizinische Einrichtungen, insbesondere in der Radiologie diese Technologie, um Befunde vereinfacht zu erfassen (Medizinische Informationstechnologie 2007; Teichgräber et al. 1999).
- Eine weniger anspruchsvolle Anwendung der Spracherkennung kann zur **Datenerfassung** erfolgen. Dies kann z.B. bei der Erstellung des Prüfberichts der KFZ-Hauptuntersuchung (Haas et al. 2004, 9) oder der Abwicklung einer Kommissionierung in der Lagerlogistik (Schulte 2003) der Fall sein.
- Den dritten Einsatzbereich stellen die **Sprachdialogsysteme** dar, die bereits im Rahmen der Sprachsynthese thematisiert wurden (vgl. Kapitel 3.1.5). Neben den dort bereits vorgestellten Beispielen (vgl. Aust 1998; Marshall 1998) werden solche Systeme als Vorstufe von Call-Centern eingesetzt. So kann z.B. bei der Autovermietung Sixt AG oder der Stadt Würzburg automatisch der richtige Ansprechpartner aufgrund einer konkreten Frage ermittelt werden (Haas et al. 2004, 9).
- Während Sprachdialogsysteme darauf ausgelegt sind, ein eingeschränktes Repertoire vollständiger natürlichsprachlicher Phrasen oder Teilen davon zu verarbeiten, wird im Rahmen der **Sprachsteuerung** überwiegend mit kurzen Kommandos gearbeitet. So können beispielsweise Einrichtungen im chirurgischen Bereich teilweise über Sprache gesteuert werden (Haas et al. 2004, 8). Funktionen, die alltäglicher sind und ebenfalls auf die gleiche Art angesprochen werden, sind im Mobiltelefon (Böcker/Holz auf der Heide 2005, 5) oder im Auto (Stenzel 1999) zu finden. Dabei zielen erste Prototypen aber bereits auf die Verarbeitung vollständiger, prosaischer Sätze ab (Westphal/Waibel 1999).

3.3. Chatbots

Die vorgestellte Spracherkennung fokussierte sich auf den linguistischen Teil der Verarbeitung, d.h. die Umwandlung des Sprachsignals in die darin enthaltenen Wörter. Die Ermittlung des Inhalts dieser Wörter, also die semantische Erkennung, erfolgt in einer nachgelagerten Komponente, die unter Verwendung verschiedener Paradigmen realisiert werden kann. Eine mögliche Ausprägung ist ein Chatbot, auf den im Folgenden genauer eingegangen wird. Nach der begrifflichen Klärung und Abgrenzung des Begriffs gegenüber anderen Ansätzen wird die typische Funktionsweise eines solchen Systems dargestellt. Darauf aufbauend soll der Fokus bei dieser Komponente auf der möglichen Gestaltung des Wissens mithilfe verschiedener Notationen liegen. Dabei werden insbesondere die Möglichkeiten und

Restriktionen aufgezeigt, welche mit dieser Technologie verbunden sind. Abschließend werden Anwendungsbereiche und Beispiele von Chatbots aufgezeigt.

3.3.1. Begriffliche Klärung

Chatterbots, Chatbots oder auch kurz Bots genannt, sind Softwareanwendungen, die ihren Namen aufgrund zweier elementarer Eigenschaften erhalten haben: Sie sind zum einen Roboter bzw. Automaten und sie verfolgen die Aufgabe des chattens. Nach de Angeli, Johnson und Coventry (2001) lässt sich der Begriff folgendermaßen definieren:

Chatbots sind Computerprogramme, die eine geschriebene Konversation mit einem Benutzer zu simulieren versuchen.

Der Einsatz von Chatbots ist somit ursprünglich auf geschriebene Dialoge beschränkt, was vor allem historische Gründe hat. Die Idee zur Umsetzung eines Chatbots basiert auf der überarbeiteten Version des Turing-Tests bzw. des Imitation Games⁹. Demnach sollte eine Versuchsperson mit zwei anderen visuell getrennten Entitäten chatten. Eine davon war ein Mensch und die andere ein Softwareprogramm, das einen Menschen simulieren sollte. Die Aufgabe der Versuchsperson war es, den Computer zu identifizieren (Turing 1950; vgl. auch Shah/Henry 2005). Die Idee von einem Computerprogramm, das einen Menschen imitiert, wurde erst Jahre später durch Joseph Weizenbaum umgesetzt. Er setzte 1966 das Computerprogramm Eliza¹⁰ um, das mithilfe einer Wissensbasis von ca. 400 Regeln einen Psychologen simulierte. Das Programm, das auch als „DOCTOR“ bekannt wurde, erfüllte seine Aufgabe sehr erfolgreich, indem es häufig Eingaben des realen Benutzers aufgriff und umformulierte. Dies gab den Patienten den Eindruck, dass das Programm sie verstanden habe (Weizenbaum 1966). Mit dem Internet Hype wurde diese Technologie wiederentdeckt und online eingesetzt, um Besucher auf Websites zu locken oder als virtuelle Berater Web-Shops zu bereichern (Bühler 2003, 112f; Buschmann 2003, 102ff). Meist kommen dabei Avatare als grafisches Frontend zum Einsatz, wobei sie nicht integraler Bestandteil eines Chatbots sind (vgl. Kapitel 3.5).

Chatbots übernehmen somit überwiegend im Internet die Funktion eines textbasierten Dialogsystems, u.U. mit Anbindung an andere Systeme wie z.B. Web-Shops oder Warenwirtschaftssysteme. Die Dialogfähigkeiten von Chatbots erlauben jedoch auch deren Verwendung als Logik- bzw. Dialogkomponente in einem Sprachdialogsystem. Wyard et al. (1996) identifizieren neben der Spracherkennung und der Sprachsynthese verschiedene dialoggestaltende Komponenten eines solchen Systems (vgl. dazu auch Keller 2004; Zue/Glass 2000, 1167ff). So spiegeln insbesondere die Elemente des Sprachverstehens, des Dialogmanagements, der Kommunikation mit externen Systemen und die Generierung einer Antwort die Fähigkeiten eines Chatbots wieder (vgl. Abbildung 3-9).

⁹ Die ursprüngliche Idee Turings hatte nicht das Ziel, einen Computer von einem Menschen, sondern einen Mann von einer Frau zu unterscheiden (vgl. dazu Bath 2001, 85; Hodges 1992, 415).

¹⁰ Der Name Eliza stammt von der Figur „Eliza Doolittle“ aus dem Schauspiel „My fair Lady“, in dem Eliza ebenso wie dem Computerprogramm das Sprechen beigebracht wird.

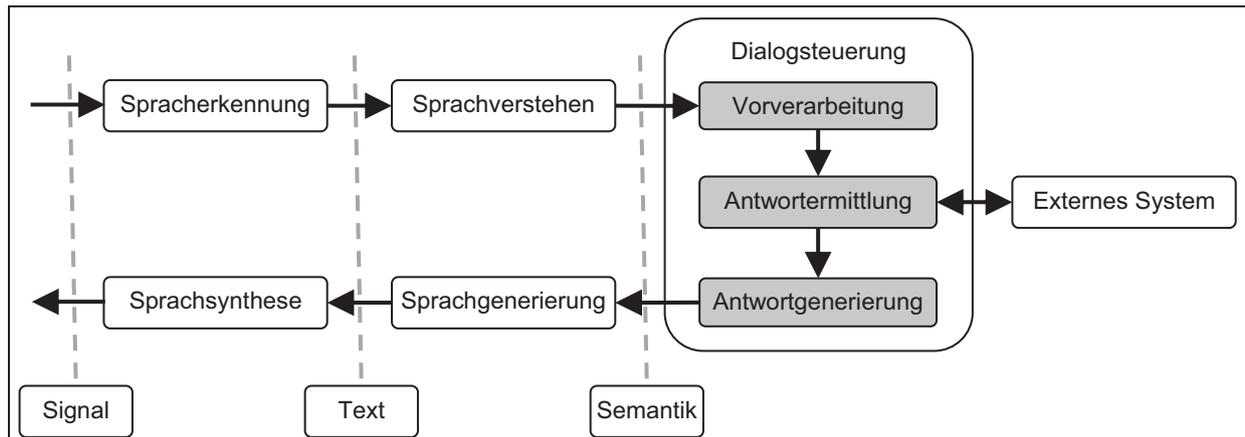


Abbildung 3-9: Funktionen eines Chatbots in einem Sprachdialogsystem
Quelle: (In Anlehnung an (Wyard et al. 1996))

3.3.2. Funktionsweise eines Chatbots

Die Ermittlung einer angemessenen Reaktion auf eine bestimmte Eingabe kann mit verschiedenen Ansätzen erreicht werden. Aufbauend auf der Vorstellung der drei möglichen Vorgehensweisen, sollen auf der Basis eines konkreten Chatbots die einzelnen internen Schritte der Verarbeitung beleuchtet werden.

3.3.2.1. Strategien zur Dialogsteuerung

McTear unterscheidet drei unterschiedliche Strategien zur Dialogkontrolle: Die state-basierte, die frame-basierte und die agenten-basierte Strategie (McTear 2002, 92ff). Die **state-basierte Strategie** führt Benutzer anhand eines fest vorgegebenen Musters durch Dialoge und erwartet an jeder Stelle eine vorgeschriebene Antwort in Form eines Wortes oder einer einfachen Phrase. Dieses restriktive Vorgehen erschwert die Eingabe durch den Benutzer, zumal die möglichen Eingabealternativen nicht immer intuitiv erkennbar sind und die Kontrolle des Dialogs beim System liegt (Zue/Glass 2000, 1166)¹¹. Diese Strategie ist robust und stellt sicher, dass alle notwendigen Daten für eine bestimmte Transaktion erhoben werden können (McTear 2002, 92).

Der **frame-basierte Ansatz** erlaubt eine flexiblere Informationsextraktion aus dem Gespräch mit einem Benutzer. So ist die Art und Reihenfolge der Informationseingabe nicht fest vorgegeben und sowohl das System als auch der Benutzer können die Initiative ergreifen (Glass/Seneff 2003). Informationen können zu jeder Zeit und in Kombination mit anderen Informationen geäußert und verarbeitet werden (McTear 2002, 92f).

Das **agenten-basierte Vorgehen** erlaubt komplexe Dialoge, da im Rahmen der Dialogkontrolle, die Absichten und Ziele des Benutzers erfasst und bei der Bildung einer Antwort berücksichtigt werden. So kann auf eine Frage nicht nur ein einfaches „ja“ oder

¹¹ Ball et al. (1997) sprechen sich beispielsweise als Alternative zu diesem Ansatz für die Verwendung eines Bayes'schen Netzes aus.

„nein“ geantwortet werden, sondern weiterführende Informationen im Rahmen der angenommenen Ziele geäußert werden (Sadek/de Mori 1998). Der Einsatz dieser Strategie setzt die Implementierung von künstlicher Intelligenz voraus (McTear 2002, 93). Der state- bzw. frame-basierte Ansatz entspricht der regelbasierten Umsetzung von künstlicher Intelligenz (Winston 1987, 186ff) während die Verwendung von Agenten dem deliberativen Verständnis von künstlicher Intelligenz zuzuordnen ist (Winston 1987, 320ff). Die Gegenüberstellung der drei Ansätze hinsichtlich der Verteilung der Initiative zwischen Mensch und Maschine sowie der Flexibilität in der Gesprächsführung erfolgt in Abbildung 3-10.

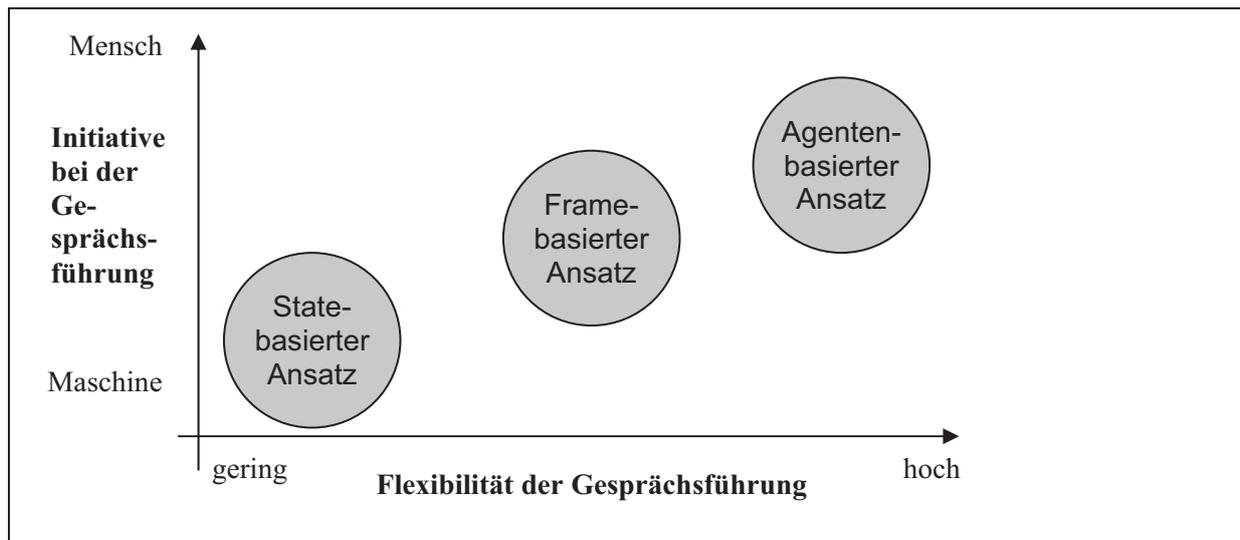


Abbildung 3-10: *Strategien zur Dialogsteuerung*
Quelle: (Eigene Darstellung)

Beim agenten-basierten Ansatz wird versucht, die eingegebene Sprache zu verstehen, d.h. die Wörter einer Phrase z.B. einer Wortart zuzuordnen und die Handlung sowie den Handelnden der Phrase zu ermitteln. Dabei werden meist ähnlich wie bei der linguistischen Spracherkennung auch Grammatiken herangezogen. Marcus (1995) berichtet bei dem Einsatz dieses Vorgehens von Fehlerraten von mehr als 60%, was einen zuverlässigen Einsatz kaum ermöglicht. Hingegen haben Kombinationen von semantischer Erkennung auf Basis von künstlicher Intelligenz (KI) und frame- oder state-basierten Ansätzen bessere Erkennungsraten erreicht (Moore 1995). Zue und Glass (2000, 1171) berichten in diesem Zusammenhang von Systemen, die Fehlerraten bei Worterkennung von 2,3% und bei Satzerkennung von nur 15,2% aufweisen, jedoch konnten diese Werte nur unter kontrollierten Bedingungen erreicht werden. In Fällen, in denen die semantische Analyse mit KI keine sicheren Ergebnisse finden konnte, wurde die Ermittlung einer Antwort auf Basis eines der beiden anderen Vorgehen durchgeführt (Stallard/Bobrow 1992; Dowding et al. 1993). Nach aktuellem Stand werden die besten Ergebnisse hinsichtlich der Simulation eines Gesprächspartners mit state- oder frame-basierten Strategien erreicht (Block 2004). Dies belegt der Wettbewerb um den Loebner Price, in dem basierend auf der Idee des Turing Tests Softwareanwendungen in einem Wettbewerb durch eine Jury begutachtet werden (vgl. dazu auch <http://www.loebner.net/Prizef/loebner-prize.html>, zugegriffen am 25.02.2008). Dabei erhält die beste textbasierte Simulation die silberne Loebner-Medaille (Vetter 2003, 73;

Shneiderman/Plaisant 2005, 339)¹². In diesem Wettbewerb haben bisher nur Chatbots gewonnen, die mit der frame-basierten Strategie operieren.

3.3.2.2. Sprachverarbeitung durch einen Chatbot

Zum besseren Verständnis der internen Abläufe in einem Chatbot soll exemplarisch dargestellt werden, wie die Verarbeitung mit dem Open Source Chatbot AliceBot (vgl. <http://www.alicebot.org>, zugegriffen am 28.07.2008) verwirklicht wird. Das interne Vorgehen lässt sich in drei Phasen unterteilen: Die Vorverarbeitung, die Antwortermittlung und die Antwortgenerierung. Dies entspricht einer weiter verdichteten Einteilung wie sie von Wyard et al. (1996) vorgenommen und in Kapitel 3.1.1 bereits grob dargestellt wurde.

In der **Vorverarbeitung** erfolgt zunächst eine Normalisierung des eingegebenen Textes. Dabei werden Umlaute aufgelöst, Ziffern in Zahlenworte umgewandelt und der gesamte Text in Großschreibung umgesetzt. Dieser Konvertierungsprozess auf Zeichen- oder Wortebene kann beliebig angepasst werden, so dass er auch zur Korrektur häufiger Rechtschreibfehler bei textbasierter Eingabe verwendet werden kann. Diese Korrektur ist ebenso wie die Transformation von Ziffern nach einer vorgelagerten linguistischen Spracherkennung meist nicht mehr notwendig. Abschließend wird auf der Grundlage vorgegebener Satzzeichen gegebenenfalls eine Trennung in mehrere Sätze vorgenommen, die in den folgenden Phasen gesondert verarbeitet werden. Ein weiterer Vorverarbeitungsschritt, der technisch jedoch meist bereits in der zweiten Phase abläuft, ist die semantische Vereinfachung einer Eingabe. Dabei werden zum einen gebräuchliche Floskeln entfernt, die nicht zur Antwortfindung oder semantischen Analyse beitragen. Zum anderen wird versucht, Wörter oder Phrasen mit gleicher Bedeutung in eine einzige Art der Formulierung umzuwandeln, um den Dialogentwicklungsprozess zu erleichtern (Ball et al. 1997)¹³.

Die zweite Phase, die **Antwortermittlung**, setzt die Übergabe des eingegebenen Textes in einer durch die erste Phase normalisierten Form voraus. Auf Basis dieses übergebenen Satzes wird in einer Datenstruktur ähnlich des Worthypothesengraphs in Abbildung 3-7 nach einem passenden Eintrag gesucht. Diese Datenstruktur enthält alle dem Chatbot bekannten möglichen Eingaben des Benutzers und wird beim Starten der Anwendung aus dessen Wissensbasis aufgebaut. Ein Vorgehen, das dabei im Gegensatz zur linguistischen Spracherkennung häufig bei einem Chatbot vorkommt, ist die Verwendung von Platzhaltern für einzelne oder mehrere Wörter. Diese Platzhalter werden genauso wie andere Wörter auch als „*“ in den Worthypothesengraph aufgenommen und bieten eine erhöhte Flexibilität bei der Gestaltung von Regeln. Die Enden der jeweiligen Äste im Graph sind nach zwei Kriterien sortiert: Dem Kontext und der letzten Antwort des Chatbots. Manche Antworten werden nur dann zurückgegeben, wenn ein bestimmtes Gesprächsthema vorliegt oder eine bestimmte Phrase zuvor vom Chatbot als Antwort geäußert wurde. Wird kein besonderer Kontext oder

¹² Die goldene Medaille wird nur für die Simulation auf audiovisueller Basis vergeben und wurde bisher noch von keiner Anwendung gewonnen.

¹³ Im Rahmen der symbolischen Reduktion wird dieses Vorgehen bei der Vorstellung von AIML in Kapitel 3.3.3.1 aufgegriffen.

eine vorausgegangene Aussage für eine Antwort festgelegt, kann diese immer auf eine bestimmte Eingabe erfolgen.

In der Phase der **Antwortgenerierung** wird die dem gefundenen Ast zugeordnete Antwort aufbereitet. Dabei können verschiedene Antworten ähnlich wie Funktionen oder Prozeduren ineinander verschachtelt werden. Variablen, die in der Antwort enthalten sind, werden durch deren Werte ersetzt. Darüber hinaus können Funktionsaufrufe von externen Systemen eingebunden werden. Auch diese Werte werden ermittelt und eingesetzt. Wurden in der Phase der Vorverarbeitung mehrere Sätze identifiziert, so werden die entsprechenden Antworten zusammengesetzt und zur Anzeige gebracht. Bei der Verwendung eines Chatbots in einem Avatarsystem können zudem auch Anweisungen zur Steuerung des Avatars oder anderer Komponenten enthalten sein. Diese Befehle werden von der Frontend-Komponente oder gegebenenfalls einem Dialogmanager verarbeitet und sind somit nicht Teil dieser dritten Phase (Reitter et al. 2001; Bush 2006).

Die drei Phasen mit den darin enthaltenen Schritten sind in Abbildung 3-11 visualisiert.

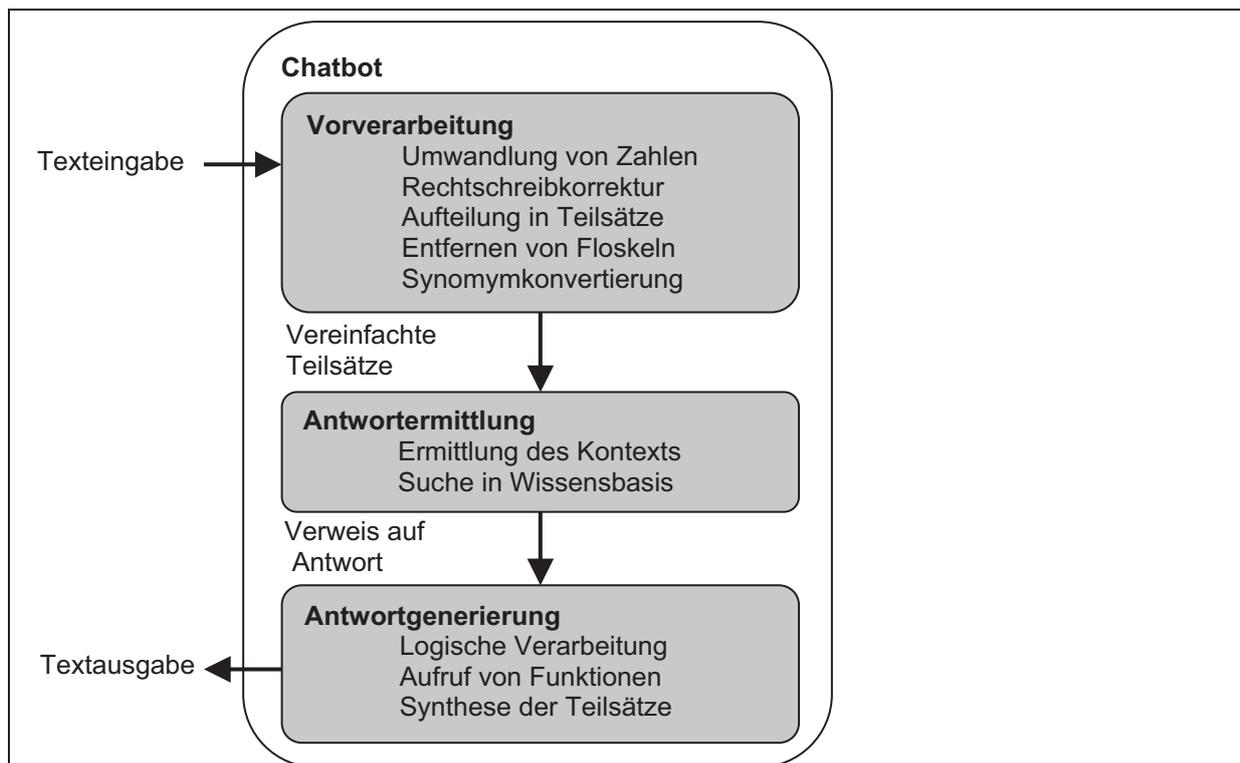


Abbildung 3-11: Verarbeitungsschritte in einem Chatbots
Quelle: (Eigene Darstellung)

3.3.3. Aufbau der Wissensbasis

Der Aufbau der Wissensbasis eines Chatbots variiert je nach verwendeter Software in Paradigma und Ausgestaltung. Mit den Einschränkungen der Wissensbasis gehen auch Einschränkungen bei der Flexibilität der Dialoggestaltung einher. Während die Verwendung einer Datenbank nahezuliegen scheint, haben sich insbesondere XML-basierte Ansätze zur

Repräsentation des Wissens aufgrund ihrer Offenheit verbreitet. Dazu zählt das in AliceBot verwendete Format Artificial Intelligence Markup Language (AIML) ebenso wie das bereits erwähnte VoiceXML. Neben diesen beiden offenen Formaten kann eine Vielzahl weiterer proprietärer Umsetzungen identifiziert werden, die meist mit syntaktisch sehr komplexen Schreibweisen arbeiten; so sind z.B. die Semantic Interface Language (SIL) aus dem Projekt SUNDial (Eckert 1996, 38ff), die Notation des Projekts Galaxy II (Seneff/Lau/Polifroni 1999), die Generic Dialogue Modeling Language (GDML) (Hanrieder/Hamerich 2004), die Notation von Kiwilogic (von Wendt 2003, 42f), die Waxholm Notation (Carlson/Hunnicut 1996), die Speech Application Language Tags (SALT) (Cisco Systems Inc. et al. 2002), die LOOM knowledge representation language (MacGregor/Bates 1987) oder die Extensible MultiModal Annotation markup language (EMMA) (W3C 2007) zu nennen. Unabhängig vom verwendeten Format stellt im Rahmen der Wissensbasis deren Erstellung auf der Grundlage strukturierten oder taziten Wissens die größte Herausforderung dar (vgl. Kapitel 6.2).

3.3.3.1. *Artificial Intelligence Markup Language (AIML)*

AIML verfügt über eine festgelegte Menge an Tags, mit deren Hilfe Regeln gestaltet werden können. Eine Regel besteht aus einer möglichen Benutzereingabe und einer Ausgabe bzw. Reaktion. In der Benutzereingabe können Platzhalter verwendet werden, die es ermöglichen, mit einer Regel viele verschiedene Eingaben zu verarbeiten. Dabei können die Inhalte der Platzhalter in der Antwort des Chatbots verwendet werden, was zu einem „Nachplappern“ des Benutzers genutzt werden kann ohne den Sinn zu verstehen.

Durch so genannte Kontexte kann der aktive Regelsatz eingeschränkt werden. Eine Regel kann einem bestimmten Kontext zugeordnet werden und kommt nur dann zum Einsatz, wenn dieser Kontext vorliegt. Die Information, welcher Kontext vorliegt, wird im Antwortteil einer Regel festgelegt. Wird einer Regel kein Kontext zugeordnet, so kann sie immer aufgerufen werden. Eine restriktivere Möglichkeit der Regelsatzbeschränkung ist die Vorgabe der zuvor ausgegebenen Antwort. Dadurch kommt eine Regel nur zum Einsatz, wenn als letzte Antwort des Chatbots eine bestimmte Phrase ausgegeben wurde.

Der Antwortteil einer Regel kann neben beliebigem Text auch verschiedene Funktionen verwenden. So können Variablen abgefragt und verändert werden, Verzweigungen eingefügt oder zufällig eine Ausgabe gewählt werden. Eine Besonderheit ist dabei der Aufruf anderer Regeln wie eine Funktion bzw. Prozedur durch die so genannte symbolische Reduktion. Dies ermöglicht neben der Verschachtelung von Regeln ineinander auch die Umleitung mehrerer gleichbedeutender Benutzereingaben zu ein und derselben Antwort. Je nach Implementierung des frame-basierten AIML-Interpreters können auch serverseitige Transaktionen wie das Aufrufen eines Befehls auf Betriebssystemebene zur Verfügung stehen.

Abbildung 3-12 zeigt ein sehr einfaches Beispiel, bei dem der Benutzer das Wort „Farben“ erwähnt und der Chatbot ihn daraufhin nach seiner Lieblingsfarbe fragt sowie den Kontext auf Farben einstellt. Eine bestimmte Regel, die nur im Anschluss an die Frage des Chatbots und in diesem Kontext zum Einsatz kommen kann, greift die erneute Eingabe des Benutzers auf und bestätigt, dass es sich dabei um eine schöne Farbe handelt. Dieses kleine Beispiel

macht bereits deutlich, dass AIML sehr flexibel ist, jedoch die Entwicklung umfangreicher Regelsätze aufgrund der fehlenden Gesamtstruktur schnell unübersichtlich wird. Insbesondere bei größeren Dialogstrukturen sind dadurch Fehler kaum zu vermeiden.

```

<category>
  <pattern>* FARBEN *</pattern>
  <template>
    Welche Lieblingsfarbe haben Sie?
    <think><set name="topic">Farben</set></think>
  </template>
</category>

<topic name="FARBEN">
  <category>
    <pattern>*</pattern>
    <that>WELCHE LIEBLINGSFARBE HABEN SIE</that>
    <template>
      <star/> ist wirklich eine schöne Farbe.
    </template>
  </category>
</topic>

```

Abbildung 3-12: *Beispiel von AIML Regeln*
Quelle: (Eigene Darstellung)

Weitere relevante Tags des AIML-Schemas sind in Tabelle 3-2 aufgeführt und beschrieben.

Tag	Bedeutung
<condition name="[Variable]" value="[Wert]" > ...</condition>	If-Then-Äquivalent; kann auch zur Fallunterscheidung eingesetzt werden
<random> </random>	Wählt zufällig eine der angegebenen Möglichkeiten aus
<set name="[Variable]"> [Wert] </set>	Weist einer Variable einen bestimmten Wert zu
<srai> ... </srai>	Lenkt die Eingabe auf eine andere Regel um bzw. ruft eine andere Regel auf
<think> ... </think>	Befehle und Texte, die zwischen den Tags stehen, werden nicht als Antwort angezeigt
<get name="[Variable]">	Gibt den Wert der entsprechenden Variable aus
<that/>	Enthält die letzte Antwort des Chatbots

Tabelle 3-2: *Weitere relevante AIML-Tags*
Quelle: (Eigene Darstellung)

3.3.3.2. VoiceXML

VoiceXML ist im Gegensatz zu AIML von Beginn an für den alleinigen Einsatz in Sprachdialogen konzipiert worden und weist dadurch weniger Flexibilität im Umgang mit der Benutzereingabe auf. Anwendungen, die dieses Format einsetzen, sind dadurch auf eine state-basierte Strategie beschränkt. Das verwendete XML-Schema beinhaltet eine bestimmte

Menge an XML-Tags, die sowohl zur Strukturierung als auch für Aktionen verwendet werden können. Es lassen sich insgesamt drei Strukturobjekte unterscheiden: Das Menü, welches Navigationszwecken dient, Subdialoge, die wie Funktionen aufgerufen werden können und zur aufrufenden Stelle zurückkehren und schließlich Formulare, welche Daten erfassen können. Die Definition von Formularen ermöglicht das Sammeln vorgegebener Daten und das Versenden des erstellten Formulars an einen herkömmlichen Web Server per Post- oder Get-Methode (Andersson et al. 2001, 9).

Im Gegensatz zu AIML erfolgt die Festlegung möglicher Eingaben nicht in VoiceXML selbst, sondern in referenzierten Grammatiken (Andersson et al. 2001, 11f) (vgl. dazu auch Kapitel 3.2.3). Neben der Integration der Spracheingabe kann auch die Sprachsynthese für die Ausgabe der so genannten Prompts gesteuert werden. Zur Gestaltung häufig auftretender Situationen wurden insgesamt drei Ereignisse vorgegeben, welche auch ohne eine Grammatik erkannt werden können: help, noinput und nomatch. Wenn der Benutzer nicht weiter weiß und explizit nach Hilfe verlangt, wird das erste Ereignis ausgelöst und eine entsprechende Aktion, meist eine Sprachausgabe, vorgenommen. Das Ereignis „noinput“ liegt vor, wenn von Seiten des Benutzers keine Eingabe erfolgt. Eine Eingabe, die nicht erkannt werden konnte, führt zum Ereignis „nomatch“ (VoiceXML Forum 2000, 13).

```

<menu>
  <prompt>Welches ist ihre Lieblingsfarbe
    <enumerate/>
  </prompt>
  <choice next="blau.vxml"> blau </choice>
  <choice next="rot.vxml"> rot </choice>

  <noinput>
    Welche ist denn nun Ihre Lieblingsfarbe?
  </noinput>

  <nomatch>
    Bitte sprechen Sie klar und deutlich.
    Ihre Optionen sind <enumerate/>
  </nomatch>
</menu>

```

Abbildung 3-13: *Beispiel eines VoiceXML Menüs*

Quelle: (In Anlehnung an (Andersson et al. 2001, 10))

Die Reaktionen auf bestimmte Eingaben können neben auszugebendem Text auch Variablen, Verzweigungen und Skripte auf Seiten des VoiceXML-Servers enthalten. Zudem sind Tags für das Unterbrechen und Weiterleiten einer Telefonverbindung enthalten. Darüber hinaus kann die Spracheingabe des Benutzers zur späteren Verwendung als Audiodatei aufgenommen werden oder eine bereits bestehende Audiodatei abgespielt werden. Bei Prompts kann angegeben werden, ob ein so genanntes Barge-In erlaubt ist. Beim Barge-In wird die Ansage des Dialogsystems durch die akustische Eingabe eines Benutzers unterbrochen. In manchen Fällen kann dies erlaubt werden, um den Dialog zu beschleunigen,

während andere Prompts vollständig vorgetragen werden müssen, bevor das System Eingaben akzeptiert (VoiceXML Forum 2000, 47).

Ein einfaches Beispiel für ein Menü, bei dem eine Farbe aus einer vorgetragenen Auswahl gewählt wird, ist in Abbildung 3-13 zu sehen. Je nachdem, welche Farbe der Benutzer wählt, wird die entsprechende VoiceXML-Datei geladen und mit dem darin befindlichen Menü oder Formular fortgefahren. Exemplarisch sind dabei auch die Ereignisse „noinput“ und „nomatch“ enthalten, die entsprechende Ausgaben bewirken. Es lässt sich erkennen, dass VoiceXML ein für Sprachdialoge optimiertes und stärker strukturiertes Format ist. Darüber hinaus vereinfacht die Integration von Sprachgrammatiken und Sprachsyntheseparametern die Übersicht. Gleichzeitig lässt es dabei jedoch weniger Flexibilität beim Verlauf von Dialogen.

3.3.4. Einsatzbereiche und Beispiele

Obwohl eine Studie von Gartner für das Jahr 2006 (Fenn et al. 2007, 20) die Technologie von Chatbots bzw. synthetischen Charakteren so einschätzt, dass sie erst in fünf bis zehn Jahren produktiv genutzt werden wird, lassen sich bereits jetzt zahlreiche Einsatzbereiche mit produktiven Systemen identifizieren.

- Das bedeutendste Anwendungsgebiet von Chatbots ist nach wie vor das **Internet**, in dem sie als Kaufassistenten, als Berater, zur Information der Besucher oder auch nur zur Unterhaltung auf Websites eingesetzt werden (Vetter 2003, 76). In diesem Bereich kommen die Möglichkeiten eines Chatbots optimal zum Einsatz, da durch die rein textbasierte Echtzeit-Kommunikation ein breites Spektrum möglicher Eingaben verarbeitet werden muss. Zum Zeitpunkt der Veröffentlichung eines solchen Systems können nicht alle Eventualitäten durch die vorhandenen Regeln abgedeckt werden, weshalb eine kontinuierliche Weiterentwicklung stattfindet. Als Ziel kann angestrebt werden, nach wenigen Monaten eine Erkennungsrate der Eingaben von bis zu 90% zu erreichen (Vetter 2003, 77).
- Der Einsatz im zweiten Bereich entstand aus den Bestrebungen heraus, nicht nur synchrone Anfragen im Internet, sondern auch asynchrone Anfrage wie E-Mails bearbeiten zu können. Im Rahmen der **Kundenbetreuung** werden Chatbots auch zur asynchronen Beantwortung von längeren Texten in Form von E-Mails eingesetzt (Buschmann 2003, 102).
- In **Auskunftssystemen**, die teilweise bereits bei der Sprachsynthese bzw. der Spracherkennung angesprochen wurden erfolgt ein Dialogmanagement, das Informationen über den Benutzer sammelt und gewünschte Daten wiedergibt. So erfolgt der Einsatz für Flugpläne (Peckham 1991; Price 1990), Bahnauskünfte (Aust et al. 1995; den Os et al. 1999; Allen et al. 1995) oder das Wetter (Zue/Glass 2000). Die Strategie solcher Systeme ist im Gegensatz zum frame-basierten Einsatz im Internet meist state-basiert angelegt, um eine höhere Robustheit bei der Erkennung und Verarbeitung zu gewährleisten.
- Da die Verwendung akustischer Bedienungselemente nach Vollrath und Totzke (2000) im einem Fahrzeug keine negativen Auswirkungen auf die primären Funktionen der Fahrzeugführung haben, kommen Dialogsysteme auch in **Fahrzeugen** zum Einsatz

(Hagen/Said/Eckert 2004; Haller 2003; Hamberger/Mauter 2003; Heisterkamp 2001). Dabei ist jedoch zu berücksichtigen, dass der Funktionsumfang solcher Systeme sehr gering ist und sich nur einfache Aktionen durchführen lassen (Hof 2007, 25). Meist kommen kurze akustische Kommandos in Form weniger Wörter zum Einsatz, so dass die Möglichkeiten der Natürlichsprachlichkeit nicht gegeben sind.

3.4. Avatare

Avatare sind durch den Einsatz im Zusammenhang mit Computern in einer Vielzahl von Aufgabenbereichen aufgetreten. Um ein klares Verständnis vom Terminus und dessen Bedeutung zu schaffen, werden zunächst der Begriff selbst sowie seine Herkunft und der heutige Gebrauch des Wortes beleuchtet. Auf der Grundlage dieser Definition, werden aktuelle Einsatzgebiete und die Rolle von Avataren darin betrachtet. Dies wird durch verschiedene Beispiele untermalt, um die bislang genutzten Potenziale und die weiteren Möglichkeiten zu skizzieren. Neben den Anwendungsmöglichkeiten von Avataren werden die technische Umsetzung und die dazu verwendeten Technologien einer genaueren Betrachtung ausgesetzt werden.

3.4.1. Begriffliche Klärung

Die ursprüngliche Bedeutung des Wortes „Avatar“ stammt aus dem Hinduismus vishnuitischer Ausprägung und bezeichnet dort eine der zehn Inkarnationen des Gottes Vishnu als „Avatara“¹⁴. Der Grundgedanke der religiösen Bedeutung wurde in die heutige technische Welt übernommen und symbolisiert die Verkörperung eines mächtigen Wesens bzw. einer Intelligenz durch eine unscheinbare Form. Der Wandel zur modernen Bedeutung des Wortes kann nicht eindeutig nachvollzogen werden, da zu verschiedenen Zeiten der Begriff im technischen Sinne verwendet wurde und über viele diese Verwendungen keine belastbaren Quellen vorliegen. So gibt es Anzeichen, dass der Begriff „Avatar“ anstelle des Namens „root“ für den Hauptbenutzer auf UNIX-Rechnern der Carnegie-Mellon benutzt worden ist. Zudem soll das amerikanische Militär in den 80er Jahren des 20. Jahrhunderts einen Begriff für Figuren in taktischen Computersimulationen gesucht und dafür „Avatar“ verwendet haben (Schmidt 1998; vgl. dazu auch Matthews/Miller 2003).

Die erste wissenschaftlich belastbare Erwähnung des Wortes im Zusammenhang mit Computern erfolgte in einem Projekt von LucasArts Entertainment. Dabei konnte eine Vielzahl von Benutzer im so genannten Habitat mithilfe von grafischen Figuren in einer Online-Community miteinander interagieren (Morningstar/Farmer 1991, 273ff). Die am häufigsten im modernen Zusammenhang erwähnte Herkunft des Begriffs geht auf Cyberpunk-Autor Neal Stephenson zurück, der in seinem Roman „Snowcrash“ den Begriff des Metaversums einführt. Dabei handelt es sich um eine Weiterentwicklung des heutigen Internets, in dem die Benutzer durch Avatare kommunizieren (Stephenson 1992, 8ff). Badler (1997) bezeichnet Avatare in diesem Zusammenhang auch als virtuelle Menschen (engl.: virtual humans).

¹⁴ In der Bhagavatapurana (Sanskrit: Das alte Buch von Gott), einer Sammlung von 12 Büchern werden die Geschichten der zehn Avatare beschrieben.

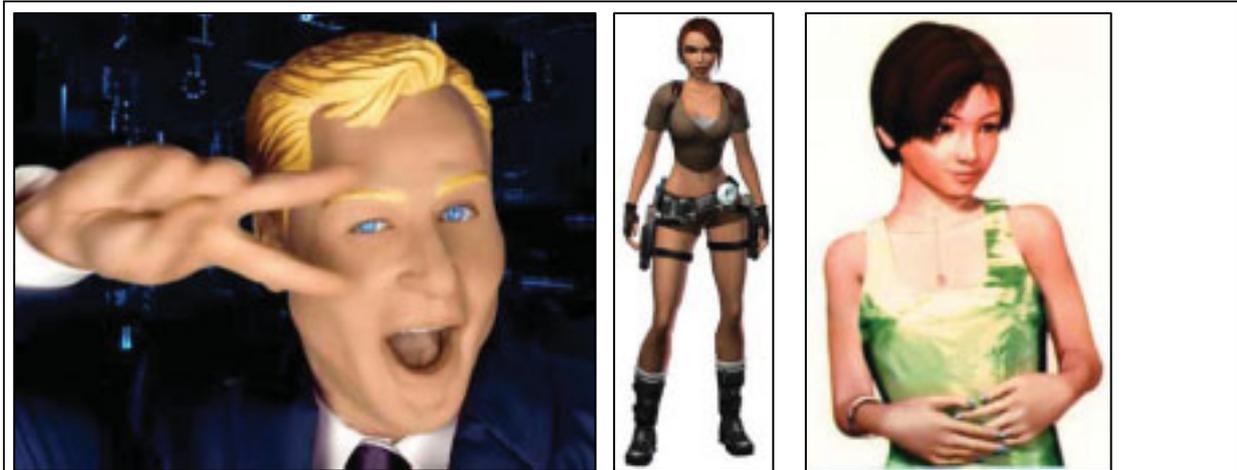


Abbildung 3-14: *Bekannte Avatare: Robert T. Online, Lara Croft und Kyoko Date*
 Quellen: (www.netzmafia.de/service/robert-t.html;
<http://www.larafantasy.de/28975/28984.html>;
www.opensorcery.net/Femalebob2.html, zugegriffen am 26.11.2007)

Stephensons Vision entspricht am ehesten der weit verbreiteten Vorstellung eines Avatars als die elektronische Repräsentation einer Person in einem virtuellen Raum (Wilcox 1998, 4ff) und hatte bisher entsprechend den größten Einfluss auf die heutige Kultur und Wissenschaft (Allbeck/Badler 1998, 19; Wilcox 1998, 23). Abbildung 3-14 zeigt einige bekannte Avatare: Robert T. Online, der als Werbemaskottchen des Internet Providers T-Online aufgetreten ist; Lara Croft, die Computerspieler im Spiel „Tomb Raider“ repräsentiert und Kyoko Date, die als virtuelle Sängerin die japanischen Musikcharts eroberte.

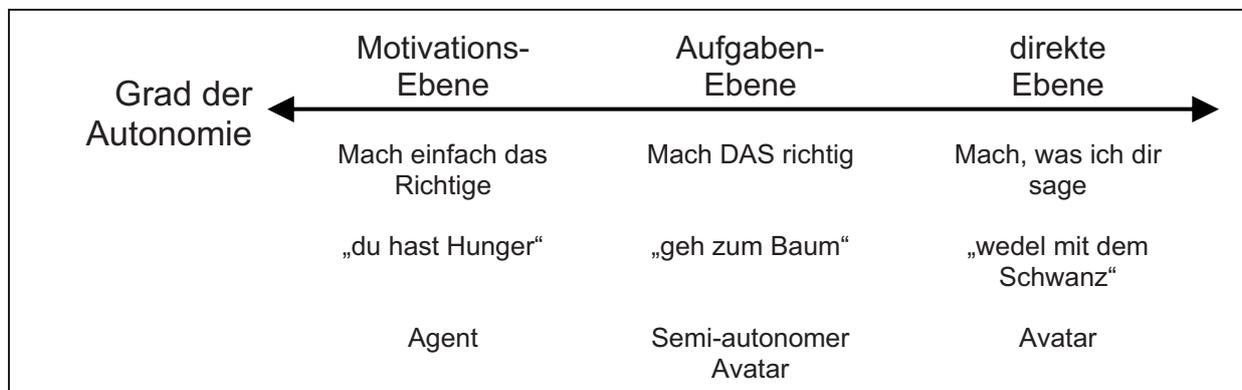


Abbildung 3-15: *Unterscheidung von Avatartypen nach dem Grad der Autonomie*
 Quelle: (In Anlehnung an (Blumberg/Galyean 1995, 48))

Döring (1999, 98) definiert einen Avatar als „Stellvertreter, virtuelle Repräsentanten der materiegebundenen Wesen, die sie ins Leben gerufen haben“. Diese Vorstellung wird dem Sachverhalt gerecht, das Menschen in einem virtuellen Raum durch grafische Abbildungen vertreten werden (vgl. dazu auch Sengers 2000, 10f). Diese Definition ist jedoch unzureichend, da sie nicht den Fall abdeckt, in dem ein Avatar eine künstliche Intelligenz vertritt. Bailenson und Blascovich (2004) sowie Badler (1997, 5f) unterscheiden deshalb zwischen Avataren als digitale Stellvertreter realer Menschen und verkörperten Agenten als Anzeige von Softwareagenten. Gemäß dieser Unterscheidung sind die Betrachtungsgegen-

tände dieses Kapitels Avatare, während die Kombination mit Logikkomponenten in Form von Agenten in dieser Arbeit als Avatarsysteme bezeichnet und im Kapitel 3.5 beleuchtet werden.

Neben der Unterscheidung dieser beiden möglichen Ausprägungen identifizieren Badler et al. (2000) darüber hinaus noch einen dritten Typ, den semi-autonomen Agenten, der auf Anweisung eines realen Benutzers hin eine bestimmte Handlung in einem eingeschränkten Rahmen selbstständig durchführen kann (Beispiele dazu sind zu finden bei Hayes-Roth/van Gent 1997; Johnson et al. 1999; Perlin/Goldberg 1996). Blumberg und Galyean (1995, 48f) unterscheiden diese drei Ausprägungen als Steuerung auf Motivations-, Aufgaben- und direkter Ebene (vgl. Abbildung 3-15). Badler (1997, 5) identifiziert neben dem Grad der Autonomie des Weiteren das Aussehen, die Funktion, den Zeitpunkt der Grafikberechnung und die Individualität als weitere Unterscheidungskriterien, wobei auch er die besondere Bedeutung der Autonomie erkennt. Unabhängig von der Verwendung von Avataren lässt sich somit folgende verallgemeinerte Definition festhalten:

Avatare repräsentieren als visuelle Stellvertreter natürliche oder künstliche Intelligenzen in einem virtuellen Raum.

Dieses Verständnis wird in Abbildung 3-16 für den Einsatz von Avataren verdeutlicht, wobei der Grundgedanke von Turk und Kölsch (2003, 4f) aufgegriffen wird, auf der Grundlage der Mensch-Mensch-Interaktion eine wahrnehmbare Mensch-Computer-Interaktion abzuleiten (vgl. dazu auch Maes 1994, 31ff). Sengers (2000, 5) sieht dabei den Avatar als Erweiterung des Menschen in die virtuelle Welt hinein (vgl. dazu auch Deray 2001, 129).

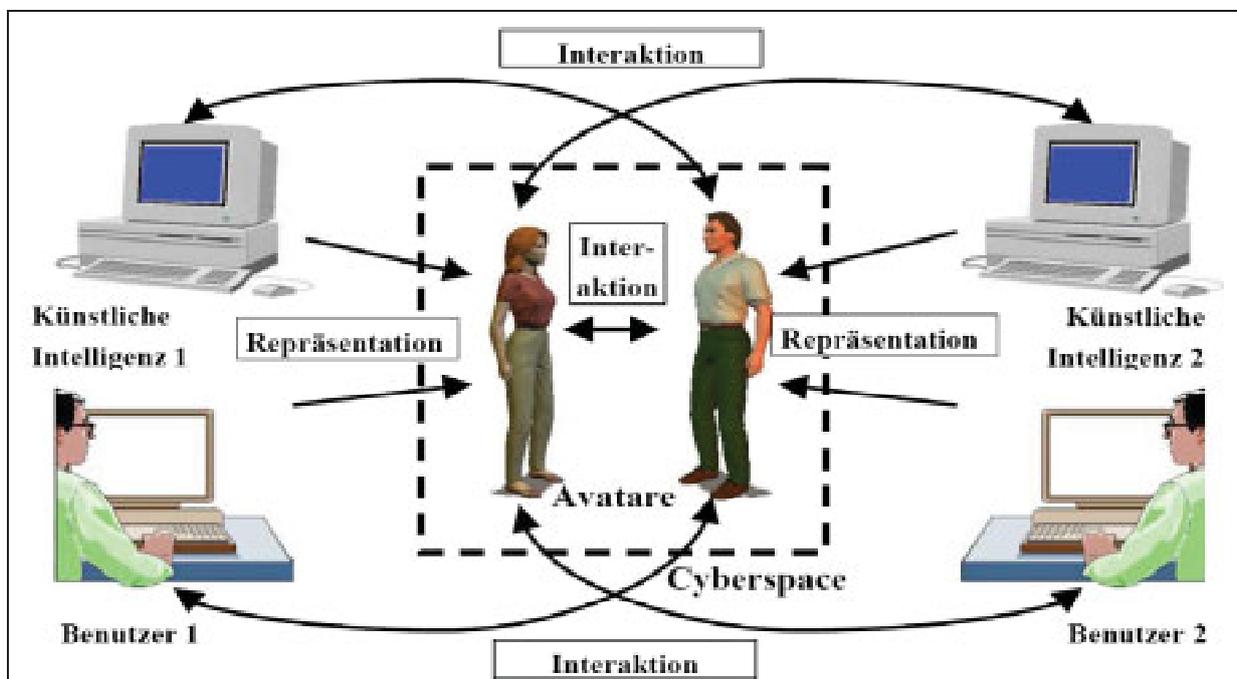


Abbildung 3-16: Avatare als Repräsentation künstlicher und natürlicher Intelligenzen
Quelle: (Eigene Darstellung)

3.4.2. Technische Umsetzung

Die technische Umsetzung eines Avatars lässt sich in die beiden Phasen des Designs und der Laufzeit unterteilen. In der Designphase fällt die Entscheidung über das grundsätzliche Aussehen eines Avatars, wobei insbesondere Kriterien wie Menschenähnlichkeit und davon abhängig Geschlecht, Alter und Aussehen eine Rolle spielen (Bath 2001, 84; Badler 1997, 5). Salem und Earle (2000) unterscheiden die drei möglichen Ausprägungen eines Avatars: Realistisch, abstrakt oder naturalistisch. Dabei stellen realistische Avatare eine reale Person dar, ein abstrakter Avatar eine fiktive Figur und ein naturalistischer Avatar eine meist menschliche an die Natur angelehnte Figur. Menschlich wirkende Figuren werden eher wie Menschen behandelt, während bei abstrakten Figuren auch unerwartete Gesten oder Verhaltensweisen akzeptiert werden. Dabei ist zu bemerken, dass besonders realistisch dargestellte, künstliche Menschen eine hohe Erwartungshaltung in ihrem realen Gegenüber aufbauen, der meist nicht entsprochen werden kann (Nowak/Biocca 2003). Ebenso spielt das Geschlecht eine Rolle, insbesondere beim Einsatz im Rahmen eines Avatarsystems (vgl. Kapitel 3.5), da beispielsweise weibliche Figuren eher anzüglichen Bemerkungen durch den Benutzer ausgesetzt sind (Lindner 2003, 11).

Die Überlegungen zur Gestaltung des Avatars sind neben dem Aussehen auch von der Erstellungs- und späteren Animationstechnologie abhängig. So bieten sich drei verschiedene Ansätze zur Erstellung an: Fotorealistisch, zweidimensional und dreidimensional (André/Rist/Müller 1998a, 25; Bath 2001, 84). Fotorealistische Avatare werden auf der Grundlage realer Menschen oder Tiere erstellt, wobei dies abhängig von der Art der späteren Animation in Form von Fotos oder Filmaufnahmen vorgenommen wird. Dieses Vorgehen ermöglicht es, schnell das entsprechende Material zu erstellen. Die spätere Erweiterung oder Änderung der erfassten Fotos oder Filme ist jedoch mit höherem Aufwand verbunden. Flexibler gestaltet sich hingegen die Erstellung als zweidimensionale Abbildung oder Trickfilm im Computer, da auch später noch unabhängig von einem Darsteller Änderungen und Erweiterungen vorgenommen werden können. Eine dreidimensionale Umsetzung eines Avatars im Computer bietet die flexibelste Option, da nach der Gestaltung der Figur die Veränderung und Bewegung des Avatars einfach realisiert werden können. Um realistisch wirkende Animationen mit einem dreidimensionalen Charakter zu realisieren kann die Motion-Capture-Technologie eingesetzt werden, bei der ein beliebiger Darsteller in einem besonders ausgerüsteten Anzug die späteren Bewegungen des Avatars durchführt.

Die Verwendung in einem Avatarsystem hat weitere Auswirkungen auf den Gestaltungsprozess, da von den geplanten Fähigkeiten des Avatars bereits frühe Entscheidungen abhängen. So kann die Wahl der Animationstechnologie davon abhängig sein, ob der Avatar nur auf bestimmte Situationen reagiert oder sich autonom bewegen kann. In manchen Fällen kann es auch Sinn machen, verschiedene Charaktere je nach Wahl des Benutzers zum Einsatz zu bringen, was zu einem Designprozess führt, der auch mit mehreren Figuren exakt reproduzierbar sein muss (Badler 1997, 5).

Zur Laufzeit eines Avatars, d.h. in der Phase, die er einem Benutzer ausgesetzt ist, lässt sich aus technischer Sicht ein zentrales Unterscheidungsmerkmal identifizieren: Die Art der Animation. Je nach dem, ob ein künstlich erschaffener oder fotorealistischer Avatar erstellt

wurde, stehen bis zu drei Möglichkeiten zur Verfügung: Statische Darstellung ohne Animationen, statisch offline generierte Animationen oder online generierte Animationen (Bath 2001, 84; Badler 1997, 5). Die statische Darstellung eines Avatars beinhaltet lediglich das Anzeigen von Bildern einer Figur, wobei diese Bilder je nach Situation wechseln können. Statische Animationen werden bereits in der Phase der Designzeit in Form von Filmen oder Animationssequenzen im Computer erstellt und situationsabhängig zur Laufzeit abgespielt. Dabei kann durch die übergangslose Kombination solcher Filmsequenzen der Eindruck eines „lebendigen“ Avatars erschaffen werden. Die dynamische Animation ermöglicht die Bewegung einer Figur in Echtzeit, d.h. zur Laufzeit. Dabei werden meist Animationskripte abgespielt, um die Figur zu steuern. Da diese Skripte in maschinenlesbarer Form vorliegen, lassen sich zudem Übergänge zwischen den einzelnen Animationskripten errechnen, so dass die Bewegungen fließend sind (Ball et al. 1997). Diese Animationsart wird jedoch nur für künstlich erschaffene Figuren genutzt und ist bei fotorealistischen Avataren nicht anzutreffen¹⁵. Die Zusammenhänge der beschriebenen Optionen für Design- und Laufzeit sind in Abbildung 3-17 dargestellt.

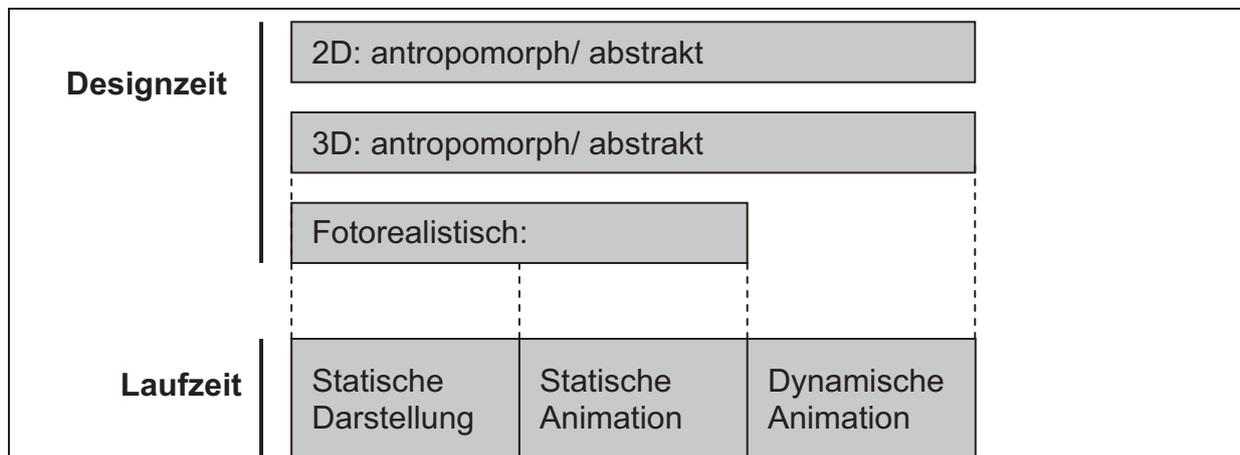


Abbildung 3-17: Optionen der technischen Umsetzung von Avataren
Quelle: (Eigene Darstellung)

Neben der Animation der Körper der Avatare im Ganzen sind auch Details von Bedeutung. So kommt beispielsweise der Bewegung des Gesichtes, insbesondere des Mundes, bei einer lippensynchronen Sprachausgabe eine wichtige Rolle zu (vgl. Lewis 1991; Waters/Levergood 1993; Albrecht/Haber/Seidel 2002; Cohen/Massaro 1993). Im Zweifelsfall bestimmt die Lippenbewegung eines Menschen die Interpretation des Gehörten. McGurk und MacDonald (1976) zeigten in einem Versuch, dass widersprüchliche visuelle und akustische Informationen beim Interaktionspartner zu neuen Inhalten verrechnet werden können, die nicht ursprünglich vermittelt werden sollten. Da ein Avatar häufig in einer virtuellen Umwelt zum Einsatz kommt, ist zudem die Steuerung dieser Szenerie eine Aufgabe, die in Abstimmung mit dem Verhalten des Avatars zu erfolgen hat (Ball et al. 1994).

¹⁵ Es lassen sich zwar Ansätze dahingehend erkennen, dass auch fotorealistische Animationen zur Echtzeit durch Viseme-Morphing übergangsfrei gesteuert werden können (Poggi/Pelachaud 2000; Ezzat/Poggio 1998), jedoch findet diese Technologie bei Avataren bzw. Avatarsystemen noch keine Verwendung.

3.4.3. Einsatzbereiche und Beispiele

Der in Kapitel 3.4.1 vorgenommenen Definition zufolge kann der Einsatz von Avataren in verschiedenen Bereichen identifiziert werden. Zunächst wird ein mögliches Klassifizierungsmodell für den Einsatz und den Fokus auf die Eigenschaften von Avataren vorgestellt. Anschließend werden die häufigsten Anwendungsbereiche mit wichtigen Vertretern und ihren spezifischen Charakteristika vorgestellt.

3.4.3.1. Strukturierung der Einsatzbereiche

Aufgrund der Vielseitigkeit ihrer Verwendung hat Badler (1997) versucht, ein Modell möglicher Einsatzbereiche für Avatare zu erstellen (vgl. Abbildung 3-18). Dabei unterscheidet er drei Dimensionen, deren Ausprägungen beliebig mit denen der jeweils anderen Dimensionen kombiniert werden können.

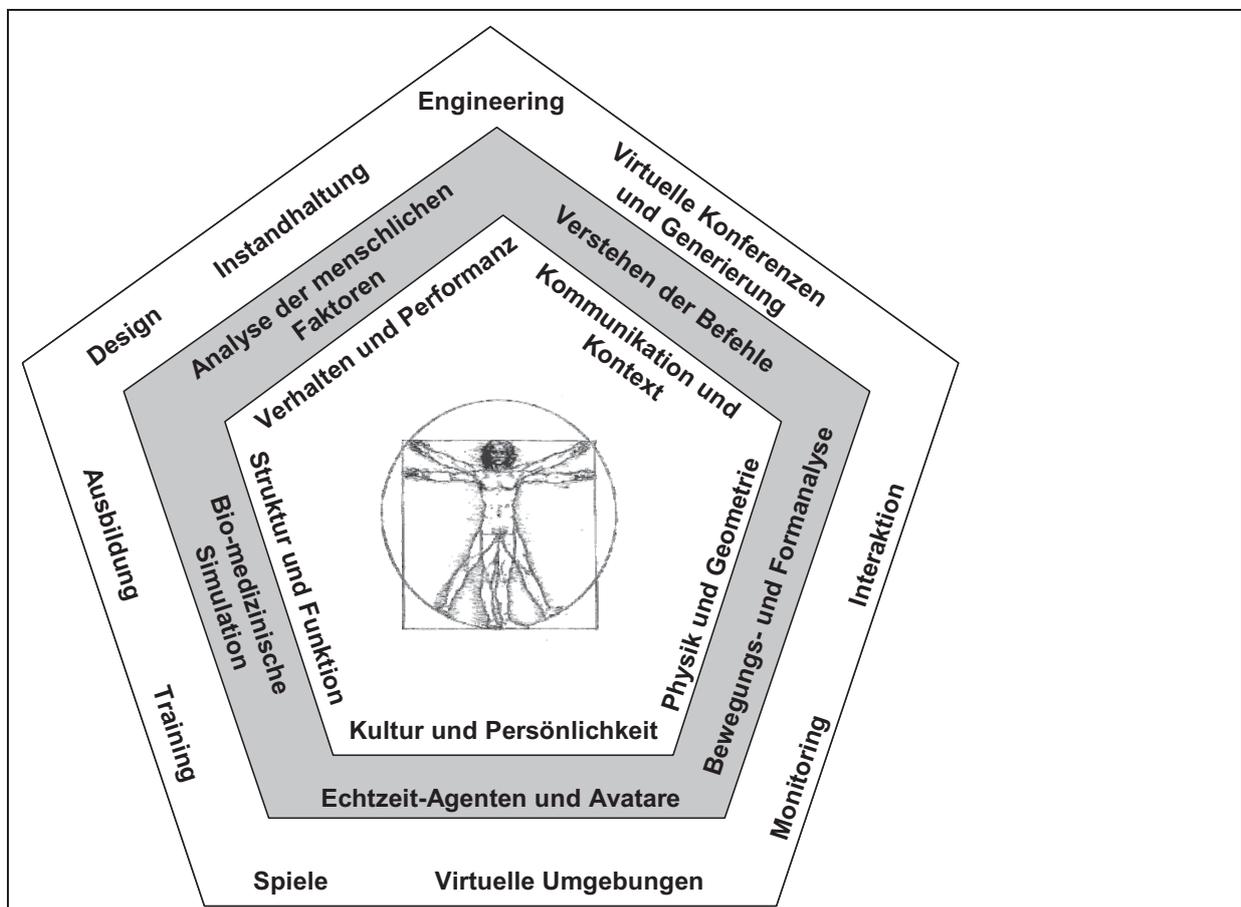


Abbildung 3-18: *Eigenschaften, Technologien und Betrachtungswinkel virtueller Menschen*
Quelle: (Badler 1997, 5)

Der innere Ring unterscheidet verschiedene Eigenschaften eines virtuellen Menschen, die untersucht oder angewandt werden können. Der mittlere Ring steht für die Technologien, die bei dessen Untersuchung oder Anwendung betrachtet werden können. Der äußere Ring schließlich gibt den Betrachtungswinkel bzw. den Anwendungsbereich an. So kann beispielsweise die Eigenschaft der Physik und Geometrie eines Avatars in einer Bewegungs-

und Gestaltanalyse zum Design neuer Gebrauchsgegenstände oder Maschinen verwendet werden. Die Zielvorstellung dieses Dissertationsvorhabens lässt sich demnach als Untersuchung von Kommunikation und Kontext mit Hilfe des Verständnisses von Anweisungen im Bereich der Ausbildung bzw. des Trainings ansehen. Kritisch anzumerken ist jedoch, dass dieses Modell mit dem Fokus auf die technische und soziale Entwicklung die ökonomische Perspektive außer Acht lässt. Die Anwendung von Avataren im Bereich der Beratung des Verkaufs oder der Werbung ist jedoch häufig der Fall. Darüber hinaus ist auch der Betrachtungswinkel des Entertainments nur durch Spiele gegeben, was der Realität nicht gerecht wird (Vetter 2003, 76).

Ein Gegengewicht zu dieser Betrachtung stellt die eindimensionale Aufführung der möglichen Anwendungsgebiete von Böcker und Holz auf der Heide (2005) dar. Sie identifizieren folgende Funktionen:

- Maskottchen in der Werbung
- Assistent bei Fragen der Bedienung
- Agent zur (teil-)autonomen Ausführung einer Aufgabe
- Stellvertreter in virtuellen Welten
- Moderator bei der Kommunikation zwischen Menschen
- Verkäufer
- Entertainer.

Diese Auflistung hat im Gegensatz zum Strukturierungsansatz von Badler (1997, 5) seinen Fokus auf den ökonomisch motivierten Bereichen sowie Unterhaltungsbereichen. Im Folgenden wird versucht, beide Modelle zu berücksichtigen und wichtige Einsatzbereiche beispielhaft vorzustellen.

3.4.3.2. *Wichtige Einsatzbereiche und Beispiele*

Das Potenzial von Avataren liegt in der Natürlichkeit der Metapher eines lebenden Wesens, mit der jeder Benutzer vertraut ist (Laurel 1997, 68). Ein Bereich der maßgeblich für den breiten Einsatz von Avataren in heutiger Software verantwortlich ist, lässt sich im **Entertainment** finden, zu dem vor allem Computerspiele, Filme, Chatanwendungen sowie Diskussionsforen zählen. In Computerspielen kam bereits früh die Notwendigkeit auf, den Spieler mit seiner virtuellen Umwelt interagieren zu lassen. Um den Spieler besser in das Spiel zu integrieren, wurden dazu bereits früh Avatare als Stellvertreter zur Interaktion mit der Spielwelt verwendet. Während die ersten Avatare nur aus wenigen Pixeln bestanden und in zweidimensionaler Form vorlagen, werden in aktuellen Spielen hochauflösende 3D-Figuren in Echtzeit berechnet. Ein solcher Avatar, der auch aufgrund seiner Verwendung zu Werbezwecken bekannt wurde, ist Lara Croft, der im Spiel „Tomb Raider“ den Spieler vertritt. Neben der Spielervertretung nehmen Avatare auch die Rolle der Darstellung von Nicht-Spielercharakteren (engl.: Non Player Characters (NPC)) wahr, wie sie in Online Mehrspieler-Welten (engl.: Massively Multiplayer Online Role-Playing Game (MMORPG)) wie „World of Warcraft“ oder in Simulationsspielen wie „Die Sims“ vorkommen. Einige

Arten von Computerspielen sind erst durch den Einsatz von solch (teil-) autonomen Avataren bzw. Agenten erst möglich geworden.

Weitere Einsatzbereiche, die insbesondere zu Zeiten des Internet- und E-Commerce-Hypes großen Zuspruch fanden, sind **Vertrieb und Marketing**. Im Rahmen von Websites werden dabei Avatare mithilfe von Chatbots zu virtuellen Beratern, die dem Besucher wichtige Fragen zum Sortiment und zu Produkteigenschaften beantworten. Je nach Ausgestaltung dieser Funktion werden Unterstützung bei der Navigation innerhalb des Internetauftritts oder auch ausgeklügelte Beratungsgespräche angeboten. Die Gespräche mit virtuellen Beratern können dabei auch unmerklich für den Besucher in Gespräche mit realen Beratern übergehen (Samuelson 2003, 33). Als Beispiele solcher Internetseiten, die aufgrund ihres Avatars an Bekanntheit gewonnen haben, sind www.yellostrom.de mit Eve, www.reifen.com mit Fritz Felge oder www.ikea.de mit Anna zu nennen. Während bei diesen Avataren die Aufgabe in der Beratung von Besuchern und damit der Generierung von Umsatz liegt, sind im Rahmen des Marketings auch andere Modelle denkbar. So werden Avatare in Kombination mit Chatbots auch zur reinen Gesprächsführung eingesetzt, um im Sinne des Customer Relationship Management (CRM) das Image eines Unternehmens zu verbessern oder Daten über die Benutzer zu sammeln (Samuelson 2003, 32f). Neben dem interaktiven Einsatz von Avataren werden sie auch als reine Werbeikonen bzw. Maskottchen eingesetzt. Bekannte Beispiele im deutschsprachigen Raum sind Robert T. Online, der zeitweise für T-Online warb oder der Bausparfuchs von Schwäbisch-Hall, der später auf <http://www.schwaebisch-hall.de> auch zum virtuellen Berater wurde (von Wendt 2003).

Die ursprüngliche Idee bei der Entstehung von Avataren war die Repräsentation von Menschen in virtuellen Welten. Dieser Idee wird der Einsatz in der **Kommunikation** zwischen Menschen im virtuellen Raum gerecht. Dieses auch als „Networked Virtual Environments“ bezeichnete Umfeld ermöglicht räumlich getrennten Teilnehmern zu kommunizieren und mithilfe ihrer Avatare auch zu interagieren. Dabei erfolgt der Austausch nicht nur auf Basis von geschriebener oder gesprochener Sprache, sondern kann auch durch Gesten und Mimik unterstützt werden (Pandzic 1998; Salem/Earle 2000). Neben eigens zur virtuellen Zusammenarbeit erschaffenen Systemen wie „Blaxxun Contact“ (<http://www.blaxxun.com>, zugegriffen am 29.07.2008), „Meeting3D“ und „WorkSpace3D“ von Tixeo (<http://www.tixeo.com>, zugegriffen am 29.07.2008), sind auch Plattformen entstanden, die den unterhaltsamen Charakter dieser Anwendung in den Vordergrund stellen. So dienen Angebote wie „ActiveWorld“ oder „Second Life“ der Möglichkeit, eigene Lebensräume zu erschaffen oder im Falle des Zweitgenannten sogar Einkäufe virtueller Objekte mit realem Geld zu tätigen (vgl. <http://www.activeworlds.com>; <http://www.secondlife.com>, zugegriffen am 29.07.2008).

Avatare haben auch Einzug in den Bereich **E-Learning und Präsentation** von Informationen erhalten. So haben Projekte wie „Talking Avatar“ (Suhendra 2001) oder „Presentation Jack“ (Noma/Badler 1997) Avatare als virtuelle Präsentatoren umgesetzt, die durch Skripte gesteuert mit Mimik, Gestik und Sprachausgabe Sachverhalte vorstellen können. Einen Schritt weiter ging das Vorhaben „BEAT“ (Cassell/Vilhjálmsón/Bickmore 2001), bei dem die Betonung der Sprachsynthese auf Basis des vorgestellten Inhalts gesteuert wurde. Eine interaktive Komponente kam zudem bei „Cyberella“ hinzu, die auf den Zuschauer bzw.

Benutzer eingehen kann (Rist et al. 2002). Als eine besondere Ausgabe dieses Avatareinsatzes ist zudem das ViSiCast Projekt zu nennen, das einen Avatar Inhalte in Gebärdensprache für Gehörlose vortragen lässt (Verlinden/Tijsseling/Frowein 2001; Kennaway 2001; IBM 2007). Als virtuelle Mentoren wirken Steve, der die Bedienung von Maschinen veranschaulicht (Rickel/Johnson/Thiebaut 2000), und Cosmo, der die grundlegende Funktionsweise von Computernetzwerken erklärt (Lester et al. 1997a). Neben diesen hauptsächlich vor wissenschaftlichem Hintergrund durchgeführten Projekten lässt sich jedoch auch die kommerzielle Verwendung in Trainingssoftware wie dem SAP Tutor erkennen (Hohage 2000). Dabei ist der Einfluss eines virtuellen Vortragenden nicht zu unterschätzen, wie Lester et al. (1997b) mit Herman dem Käfer oder Bernsen et al. (2005) mit einem virtuellen Hans Christian Andersen als Tutor für Kinder nachweisen konnten.

Die letzten Einsatzbereiche, die hier vorgestellt werden soll, sind **Simulation und Ergonomieuntersuchungen**. Dabei werden Avatare im Bereich des Computer Aided Engineerings (CAE) und Computer Aided Designs (CAD) genutzt, um neue Geräte und Abläufe menschengerecht zu gestalten. So helfen Anwendungen wie „CATIA Human Builder“ (Brückmann 2007) oder „Jack/Jill“ (Badler/Phillips/Webber 1993) beim virtuellen Entwurf, den Anforderungen durch die menschliche Bedienung gerecht zu werden. Neben diesen gestalterischen Aufgaben werden Avatare jedoch auch in militärischen Simulationen eingesetzt, um Bewegungsabläufe und Krisenverhalten zu optimieren (vgl. dazu Loftin et al. 2004; Matthews/Miller 2003).

3.5. Avatarsysteme

Nach dem in den Kapiteln 3.1 bis 3.4 die einzelnen technischen Komponenten eines Avatarsystems beschrieben wurden, wird nach einer Klärung des Begriffs des Avatarsystems ein Blick auf das System als Ganzes geworfen werden. Dabei sind insbesondere Aspekte der technischen Integration der verschiedenen Technologien von Bedeutung. Das Zusammenspiel und die Synchronisation der einzelnen Komponenten stehen bei dieser Betrachtung im Vordergrund. Der hauptsächliche Beweggrund für die Zusammenstellung dieser Technologien liegt jedoch im Synergiepotenzial auf psychologischer Ebene. Um einen virtuellen Charakter zu erschaffen, tragen die Eigenschaften der bereits vorgestellten Komponenten ihren Teil bei. Abschließend werden bereits umgesetzte Avatarsysteme beleuchtet, um Probleme und weiteren Forschungsbedarf zu identifizieren.

3.5.1. Begriffliche Klärung

Im Laufe der Zeit haben sich die Paradigmen zur Gestaltung der Benutzerschnittstelle von Computersystemen kontinuierlich verändert. Während in den 1950er Jahren noch Schalter, Lochkarten und Lampen die Mensch-Maschine-Interaktion bestimmten, kam in den 1970ern die Kommandozeile auf, die eine Benutzung ähnlich einer Schreibmaschine zuließ. In den 1980ern kam der Window-Icon-Mouse-Pointer (WIMP)-Ansatz auf, der bis heute den Großteil der softwarebasierten Anwendungen beherrscht. Diese historischen bzw. teilweise noch aktuellen Paradigmen sind dadurch gekennzeichnet, dass der Benutzer Bedienmuster erlernt hat, die jedoch nicht seinem natürlichen Verhalten entsprechen und somit nicht ohne Vorkenntnisse intuitiv bedienbar sind. Nach Turk und Robertson (2003, 2ff) wird sich in

Zukunft das Bedienmuster vieler Anwendungen hin zu einer natürlichen Interaktion entwickeln. Diese so genannten Post-WIMP- oder Anti-WIMP-Schnittstellen sind dabei durch neue Technologien wie 3D-Interfaces gekennzeichnet, die überall verfügbar sind, multimodal interagieren und bewusst mit Emotionen arbeiten (van Dam 1997; vgl. dazu auch Gentner/Nielson 1996).

Eine mögliche Ausprägung dieser neuen Schnittstellen ist der Embodied Conversational Agent (ECA), der nach Cassell (2000b, 2) verschiedene Eigenschaften eines realen Menschen aufweist. So ist er in der Lage, Konversationen zu führen, Emotionen und Persönlichkeit zu zeigen und er verfügt über einen (virtuellen) Körper. Dabei ist die Forderung nach einem Körper nicht in physischem Sinne zu verstehen, sondern er kann auch durch eine figürliche Darstellung auf einem Bildschirm realisiert werden. All diese Eigenschaften sollen dem Benutzer die Bedienung eines softwarebasierten Systems erleichtern, indem er Muster anwendet, die er aus der Mensch-Mensch-Kommunikation heraus bereits kennt (Blens/Krämer/Bente 2003, 297). Es ist somit nicht mehr Aufgabe des Menschen, sich an die Maschine anzupassen, sondern die Herausforderung an die Maschine, sich an den Menschen anzupassen.

Die geforderte Verkörperung der Schnittstelle legt die Verwendung von Avataren nahe, weshalb bei Embodied Conversational Agents (ECA) auch von Avatarsystemen gesprochen werden kann. Sie weisen neben der figürlichen Darstellung die Eigenschaften eines Dialogsystems auf, wie es ein Chatbot oder ein um die Sprachkomponente erweitertes Sprachdialogsystem ist (Churchill et al. 2000). Durch die Verwendung eines Avatars in Verbindung mit einer entsprechenden Softwaresteuerung liegt nach der in Kapitel 3.4.1 vorgestellten Definition von Bailenson und Blascovich (2004) ein Agent vor. Diese Kombination von Avatar und Softwaresteuerung, welche das Ziel verfolgt, nicht nur eine künstliche Intelligenz sondern ein künstliches Wesen zu schaffen, ist durchaus umstritten. Neben ethischen Fragestellungen steht insbesondere die technische Notwendigkeit und die soziale Sinnhaftigkeit im Vordergrund dieser Diskussion (Koda/Maes 1996; Shneiderman 1997, 98ff).

Gleichbedeutend mit dem Begriff des ECA lassen sich auch anthropomorphe Interface-Agenten (Krämer/Bente 2002, 205ff), virtuelle Menschen (Badler 1997, 4ff) oder weitere unternehmensabhängige Bezeichnungen von Avatarsystemen wie Lingubot, IQ-Agent oder Imigos identifizieren (von Wendt 2003, 39; Samuelson 2003, 29ff; Bühler 2003, 111). In Anlehnung an die Definition von ECA (Cassell et al. 1998, 21) lassen sich für Avatarsysteme wie folgt definieren:

Avatarsysteme sind softwarebasierte Systeme, die durch einen Avatar verkörpert werden und konversationale Fähigkeiten für die Kommunikation mit realen Menschen aufweisen. Sie können verbale und non-verbale Eingaben erkennen und verarbeiten und darauf mit verbalen und non-verbalen Ausgaben antworten. Sie können während eines Dialogs die Initiative ergreifen, dem Benutzer Feedback geben und den Gesprächsverlauf durch eigene Vorschläge bereichern.

Während diese Definition bereits explizite Anforderungen an die Konversation mit einem Avatarsystem stellt, gehen Cassell und Thórisson (1999b, 520) weiter, indem sie fordern, dass solche Systeme mit allen Fähigkeiten eines menschlichen Gesprächspartners ausgestattet werden sollen. Nach ihrem Verständnis soll es dadurch dem Avatarsystem möglich sein, Gestik, Mimik und Prosodie so zu benutzen, dass die Illusion eines Gesprächspartners für den Benutzer noch stärker ausgeprägt ist. Neben den konversationalen Fähigkeiten lassen sich auch die Eigenschaften eines Agenten in diesem Zusammenhang weiter fassen. So sieht Ball (1997, 192) ein solches System nicht nur als Werkzeug, sondern als Assistent an, der selbstständig Aufgaben erfüllen kann. Dies beinhaltet die Klärung ungenau gestellter Aufgaben sowie das Vorschlagen von Alternativen wie es auch Cassell et al. (1998, 21) vorschlagen.

3.5.2. Technisches Zusammenspiel

Während in den vorangegangenen Kapiteln die einzelnen Komponenten eines Avatarsystem vorgestellt wurden, sollen nun die Herausforderungen und möglichen Lösungsansätze bei deren Zusammenspiel betrachtet werden. Dabei liegt der Fokus auf der Sensibilisierung für die verschiedenen Arten der Probleme. Ein strukturiertes Vorgehen zur Gestaltung und damit die Lösung der aufgeführten Aspekte erfolgt im Rahmen der Entwicklung einer Referenzarchitektur in Kapitel 5.1.2.

Ein Avatarsystem besteht je nach Ausführung aus zwei bis vier bzw. fünf der bereits vorgestellten technischen Komponenten, wobei die Modalität der Textein- und -ausgabe außer Acht gelassen wird. Das kleinste System verwendet nur einen Avatar und eine Logikkomponente wie z.B. einen Chatbot. Gegebenenfalls lässt sich neben der Darstellung eines Avatars auch die Integration des Avatars in eine virtuelle Umwelt erkennen, die auch als eigenständige Komponente agieren kann. Dieser Grundaufbau kann für verschiedene Situationen um eine Spracheingabe oder Sprachsynthese erweitert werden (vgl. Abbildung 3-19). Im Rahmen der Logikkomponente können auch externe Systeme wie Warenwirtschaftsanwendungen oder Datenbanken eingebunden werden, die jedoch nicht als Bestandteile des Avatarsystems zu sehen sind.

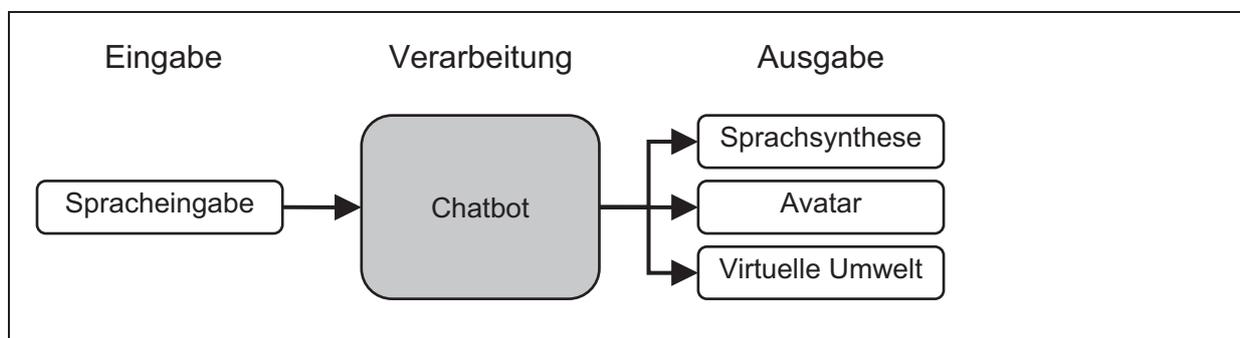


Abbildung 3-19: Grundlegende Komponenten eines möglichen Avatarsystems
Quelle: (Eigene Darstellung)

Neben technischen Herausforderungen wie den Betrieb von Komponenten auf verschiedenen Hardwareplattformen oder Betriebssystemen, sowie der Verschiedenartigkeit der verwendeten

Programmiersprachen und Schnittstellen, spielt insbesondere die synchronisierte Kontrolle aller Komponenten eine wichtige Rolle. Ein Avatarsystem verfolgt das Ziel, einen realen Menschen zu simulieren und muss deshalb als ein Ganzes agieren, um diesen Eindruck beim Benutzer zu erwecken bzw. aufrecht zu erhalten. Dies kann z.B. die Untermalung von geschriebenen oder sprachlich synthetisierten Inhalten mit einer passenden Bewegung des Avatars sein oder die Aufmerksamkeit des Avatars beim Erkennen eines Sprachsignals von Seiten des Benutzers. Im Folgenden sollen nun diese Aspekte aufgeführt werden. Dabei werden zunächst nur jeweils zwei Komponenten betrachtet und die Anzahl involvierter Teile schrittweise erhöht bis das gesamte System aus dem Blickwinkel der Integration beleuchtet wird.

3.5.2.1. Abstimmung zweier Komponenten

Ein elementares Problem ist die Abstimmung der Avatardarstellung auf die vermittelten Inhalte (Cassell et al. 1998, 22). Je nachdem, ob eine statische Darstellung bzw. Animation oder eine dynamische Animation verwendet wird (vgl. Kapitel 3.4.2), kann dies erfolgen, indem bestimmte Bilder oder Bildsequenzen zu einer Aussage des Chatbots angezeigt werden. Dabei können durch festgelegte Übergangspfade zwischen einzelnen Avatarpositionen und -gesten flüssige Bewegungsabläufe erreicht werden (Stone/Lester 1997; vgl. dazu auch André/Rist/Müller 1998b, 19). Flexibler erfolgt die Kombination verschiedener Bewegungen und Positionen durch die dynamische Animation, da durch die automatische Berechnung von Übergängen der Parameter immer flüssige Übergänge zwischen Bewegungen erzielt werden können (Noma/Badler 1997). Dabei erfolgt die Steuerung der Animationen über Befehle die in die Wissensbasis des Chatbots integriert sind. Möglichkeiten der Einbettung solcher Avatarkommandos zeigen Ball et al. (1997) bzw. Perlin und Goldberg (1996) mit ihrer Skriptsprache IMPROV. Für eine domänenspezifische Anwendung und bei bestimmten Architekturen ist es zudem möglich, Animationen autonom vom Avatar bestimmen und ausführen zu lassen (Kopp/Tepper/Cassell 2004; Cassell et al. 1998).

Lässt sich eine eigenständige virtuelle Umwelt identifizieren, in der der Avatar agieren kann, so erfordert dies eine Abstimmung des Verhaltens von Chatbot und Umwelt sowie des Avatars mit der Umwelt. Zum einen muss dabei durch die Regeln des Chatbots sichergestellt werden, dass sich die Umwelt entsprechend des ausgegebenen Inhalts anpasst, d.h. dass beispielsweise bestimmte Objekte sichtbar werden oder sich verändern. Zum anderen muss sich das Verhalten des Avatars der Umwelt anpassen, da erwartet wird, dass er sich seiner Umwelt bewusst ist. Das kann sich darin äußern, dass der Avatar mit Teilen seiner Umwelt interagiert. Dies kann in Chatbotregeln beispielsweise mit abstrakten Kommandos wie „Schau zum Auto“ oder „Bewege den Hebel X“ erreicht werden, wobei die einzigartige Benennung der Umweltobjekte vorausgesetzt wird (Rickel/Johnson 1997; Theune 2001).

Die Verwendung einer Sprachsynthese macht die Betrachtung des Zusammenspiels von Avatar und der Sprachsynthese erforderlich. Zur Unterstützung des menschlichen Eindrucks ist die lippen-synchrone Sprachausgabe ein wichtiger Punkt, dem nur bei dynamischer Animation des Avatars ausreichend begegnet werden kann. Bei der Synchronisation dieser rechenintensiven Komponenten müssen die Latenzzeiten berücksichtigt werden (Lewis 1991;

Waters/Levergood 1993; Albrecht/Haber/Seidel 2002; Cohen/Massaro 1993). Im Gegensatz zur Verwendung von Animationen ohne Sprachsynthese, wo sie meist symbolisch Emotionen und Absichten untermalen, muss die Sprachausgabe besser durch Animationen unterstützt werden. Beim Sprechen wird wie bei einem realen Menschen auch eine Körpersprache erwartet, die dem gesprochenen Inhalt entspricht. Dies kann meist automatisch durch den Avatar bewerkstelligt werden (Deray 2001; Cassell/Stone 1999; Beskow/McGlashan 1997; Cassell/Thórisson 1999a). Beim Einsatz einer prosodischen Notation in der Sprachsynthese zur Unterstreichung bestimmter Emotionen des Avatars kann dies zu einer Herausforderung werden, da die korrekten Emotionen ermittelt werden müssen. Mögliche Ansätze zur Integration von Animations- und Sprachsynthesebefehlen in die Regeln eines Chatbots zeigen Ball et al. (1997), Badler (1997, 10f) oder Kshirsagar et al. (2002) mit der Avatar Markup Language.

Beim Einsatz einer Spracherkennung erfordert dies für die Interaktion zwischen ihr und dem Chatbot, dass das so genannte *turn taking* abgestimmt ist. Darunter ist zu verstehen, dass ähnlich wie zwischen zwei Menschen eindeutig ist, welcher der beiden Gesprächspartner (Mensch oder Avatarsystem) sprechen darf (Cassell/Torres/Prevost 1999). Dies kann zum einen durch einen Push-to-Talk-Button geregelt werden, mit dem der Benutzer ankündigt, dass er etwas sagen möchte (Ball et al. 1997) oder durch Voice Activity Detection, bei der automatisch der Beginn eines Sprachsignals erkannt wird (Singh/Boland 2007).

Eng mit dem Turn taking verbunden ist das Signalisieren des Avatars, dass er dem Benutzer zuhört. Dies erfolgt in erster Linie dadurch, dass der Avatar den Benutzer anschaut, wenn ein Sprachsignal erkannt wird oder der Push-to-Talk-Button betätigt wurde (Lee/Badler/Badler 2002; Bilvi/Pelachaud 2003). Darüber hinaus erwarten Menschen eine Bestätigung dafür, dass ihnen ihr Gegenüber zuhört (Picard 2000, 15). Dies wird meist durch leichtes Nicken, Heben einer Augenbraue oder andere Mimik erreicht (Cassell et al. 1998, 22).

3.5.2.2. Abstimmung mehrerer Komponenten und ganzer Systeme

Die Kombination mehrerer Komponenten im Rahmen eines Avatarsystems lässt sich der Einfachheit halber in die Komponenten zur multimodalen Ausgabe sowie die Komponenten zur multimodalen Eingabe unterscheiden. Bei jedem dieser Blickwinkel spielt der Chatbot als Bindeglied eine wichtige Rolle.

Von den durch einen Benutzer wahrnehmbaren Teilen eines Avatarsystems wurden bisher nur der Avatar selbst, seine Umwelt und die Sprachausgabe betrachtet, da sie in vielen Systemen enthalten sind. Turk und Robertson (2000) erweitern diese Vorstellung jedoch sowohl hinsichtlich Ein- als auch Ausgabe um den Aspekt der Multimodalität und fassen dies zu einem Perceptual User Interface zusammen. So können neben den beiden bereits beschriebenen Ausgabekanälen Weitere hinzukommen, wenn beispielsweise die Steuerung von Geräten eine mögliche Reaktion eines Avatarsystems sein kann. Unabhängig von der Anzahl zu steuernder Komponenten, muss eine konzertierte Reaktion aller Teile erfolgen, um wie aus einem Guss zu erscheinen (Krämer/Nitschke 2002). Dies kann dazu führen, dass ein zeitlich genau abgestimmtes „Drehbuch“ die Reaktion eines Avatarsystems steuert. Mit der Teilautomatisierung von Reaktionen und der damit verbundenen Verteilung der Steuerung

kann dieses Drehbuch kurz gehalten werden. Dies reduziert den Aufwand, der bei der Gestaltung der Chatbotregeln entsteht. Cassel et al. (1998, 21) zeigen den Aufwand einer konzertierten Reaktion am Beispiel ihres Avatarsystems „Rea“ auf, bei dem lediglich der Avatar und die Sprachsynthese zentral gesteuert werden. Abbildung 3-20 zeigt die technische Notation einer ca. drei Sekunden langen Interaktion zwischen zwei realen Gesprächspartnern. Die Granularität der einzelnen Aktionen lässt erahnen, welcher Aufwand in der Abstimmung lediglich dieser zwei Ausgabemodalitäten für ein Avatarsystem stecken kann.

Die multimodale Eingabe eines Avatarsystems kann neben der Spracheingabe je nach Anwendungsgebiet auch gesonderte Eingabegeräte wie Tastatur, Maus, Touchscreen oder andere Bedienelemente beinhalten. Diese vom Computer her bekannten Eingabemodalitäten lassen sich jedoch auch durch Gesten, Mimik oder die Tonalität der Sprache ergänzen (Cassell et al. 1998). Während die Bindung an ein Eingabegerät eine ubiquitäre Kommunikation mit einem Avatarsystem einschränken kann, stehen die personengebundenen Modalitäten immer zur Verfügung. Dabei sehen Badler (1997, 8) sowie Ball et al. (1997) jedoch die Sprache als wichtigste Eingabemodalität eines Avatarsystems (vgl. dazu auch Cohen 1992).

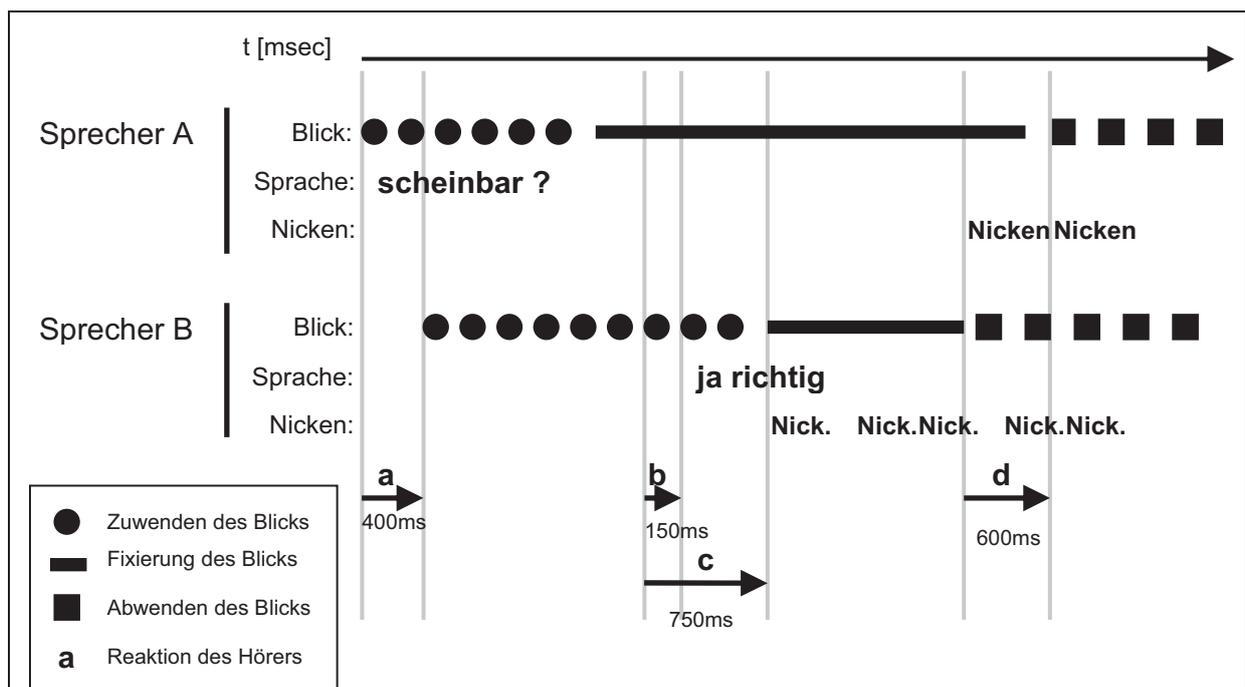


Abbildung 3-20: Notation eines Unterhaltungsausschnitts zweier Gesprächspartner
Quelle: (In Anlehnung an (Thórisson 1998, 20))

Bei der Vielzahl der Eingabemöglichkeiten stellt die korrekte Verarbeitung der teilweise gleichzeitig geäußerten Eingaben eine Herausforderung dar. Insbesondere die Priorität einer Eingabemodalität steht dabei im Vordergrund. Somit stellt sich die Frage, ob der übermittelte Inhalt einer Eingabemodalität besonderen Vorrang hat, durch eine andere Modalität unterstützt werden kann und wie mit widersprüchlichen Informationen verschiedener Modalitäten umgegangen werden soll (Bernsen/Buisine/Martin 2005; Bolt 1980; Lefebvre/Duncan/Poirier 1993; Althoff et al. 2001).

3.5.3. Soziale Effekte

In einer Welt, in der Computer als intelligent bezeichnet werden und der Durchbruch zur echten künstlichen Intelligenz greifbar nahe scheint, ist es nicht verwunderlich, dass in Rechnern und der darauf betriebenen Software mehr als nur eine Maschine gesehen wird. Auf diesem weit verbreiteten Denken sowie dem Verhalten bei natürlichsprachlicher Interaktion basiert die psychologische Wirkung von Avatarsystemen, die auch ohne einen Avatar, nur mit einem Chatbot oder sogar weniger Aufwand überprüft werden kann (Winograd/Flores 1989, 228). Reeves und Nass (1998) zeigten in einer Vielzahl von Versuchen, dass bei einer natürlichsprachlichen, meist textbasierten, Interaktion mit Rechnern diese durch deren Benutzer als soziale Akteure wahrgenommen wurden (vgl. dazu auch Erickson 1997; Nass et al. 1995; Nass/Steuer/Tauber 1994). Diese Beobachtung wird von den beiden Forschern mit dem Phänomen der „old minds“ begründet. Danach hat sich bei Menschen die Auffassung eingeprägt, dass Entitäten, die sozial agieren, d.h. auf Äußerungen anderer Entitäten gegebenenfalls auch emotional reagieren, menschlich wirken. Da wir nur gewohnt sind, mit anderen Menschen sozial zu interagieren, wird ein Rechner, der dieses Verhalten imitieren kann, als gleichwertig im sozialen und emotionalen Sinne angesehen (Reeves/Nass 1998, 3ff). Caporeal und Heyes (1996) begründen diesen Sachverhalt damit, dass Menschen alles Unerklärliche mit menschlichem Verhalten zu erklären versuchen, so dass Tieren, Ereignissen oder Objekten Menschlichkeit angedichtet wird. Die gedankliche Übertragung menschlicher Eigenschaften erfolgt somit auch auf Computer (de Angeli et al. 1999). Dieses Verhalten konnte in verschiedenen Versuchen auch für Power User nachgewiesen werden, die sich durch ihre tägliche Arbeit der Natur eines Computers bewusst sind (Reeves/Nass 1998, 26f).

Dieser beobachtete Zusammenhang zwischen sozialer Interaktion und Wahrnehmung durch den menschlichen Benutzer, der durch textbasierte Kommunikation auftritt, lässt sich bei der Darstellung von Bildern oder Bildsequenzen ebenfalls beobachten. Walker et al. (1994a) zeigen den Einfluss von Bildern, welche auf einem Rechner angezeigt werden, auf Stimmung und auf Leistungsfähigkeit menschlicher Probanden (vgl. dazu auch Sproull et al. 1997). Den gleichen Einfluss konnten Slater und Steed (2001) aufzeigen, indem sie animierte Personen als virtuelles Publikum für den Vortrag von Probanden verwendeten. Je nachdem, wie interessiert die Zuhörerschaft zu sein schien, bewerteten die Vortragenden sich selbst besser, wenn die Avatare interessiert wirkten und schlechter, wenn sie gelangweilt erschienen. Computer können somit unabhängig davon, ob die Kommunikation auf schriftlichem oder visuellem Wege stattfindet als ebenbürtig in sozialem Sinne angesehen werden¹⁶.

Während die getrennte Untersuchung von kommunikativem und visuellem Austausch zwischen Mensch und Maschine eine gleichrangige Behandlung offenbart, lassen sich diese Effekte auch für die Kombination aus beiden Komponenten, also einem Avatarsystem, erkennen (André/Rist/Müller 1998b; Lester et al. 1997b; Takeuchi/Nagao 1993; Walker/Sproull/Subramani 1994b). Die Menschlichkeit wird durch die Darstellung einer

¹⁶ Bei der visuellen Kommunikation ist anzumerken, dass Kiesler und Sproull (1997, 195) den geschilderten sozialen Einfluss insbesondere bei der Darstellung von menschenähnlichen Gesichtern im Gegensatz zu fiktiven oder tierischen Erscheinungen beobachtet haben.

interaktiven virtuellen Figur verstärkt, da der übliche Blickkontakt während eines Gesprächs gewahrt werden kann (Cassell/Bickmore 2000, 50). Zudem kann durch Bewegungen und Gesten des Avatars in bestimmten Gesprächssituationen eine nonverbale Körpersprache eingebracht werden, auf die Menschen reagieren und auf Basis der Konsistenz von verbaler und nonverbaler Sprache ihre Meinung über die Glaubwürdigkeit des Gegenübers bilden (Cassell/Bickmore 2000, 55; Biocca 1997). Die Bedeutung der nonverbalen Kommunikation wird insbesondere bei Bente und Krämer (2000, 39f) deutlich, die verschiedene Ansätze zur Modellierung entsprechenden Verhaltens für anthropomorphe Schnittstellen gegenüberstellen.

Der positive soziale Effekt solcher Avatarsysteme führt dazu, dass anthropomorphe Schnittstellen als besonders effizient und akzeptiert propagiert werden und von ihnen erwartet wird, dass sie die Mensch-Maschine-Schnittstelle benutzerfreundlicher, d.h. menschenfreundlicher, gestalten sollen (Bolt 1987; Brennan 1990; Laurel 1997; Takeuchi/Naito 1995). Bei der Akzeptanz solcher Systeme spielen jedoch weitere Faktoren wie etwa die Ethnizität des Avatars (Lee/Nass 1998) oder die Ähnlichkeit zum Benutzer (Churchill et al. 2000, 75) eine Rolle. Koda und Maes (1996) konnten jedoch zeigen, dass unabhängig von der grafischen Gestaltung des Avatars die wahrgenommene Intelligenz eines Systems nur von der Logikkomponente, also dem Agenten bzw. Chatbot, allein abhängt. So ergab auch eine Untersuchung von Andre, Rist und Müller (1998b) beispielsweise keine signifikante Verbesserung des Lernerfolgs durch den bloßen Einsatz eines Avatars in einem Softwaresystem.

Durch das Zusprechen sozialer Kompetenz gegenüber einem Avatar, lässt sich insbesondere beim Einsatz eines Avatarsystems unter Verwendung einer automatischen Spracherkennung als Eingabemodalität ein weiterer Effekt erkennen. Eine Spracheingabe wird dabei signifikant häufiger verwendet, wenn ein menschliches Gesicht auf dem Computerbildschirm angezeigt wird (Krämer/Nitschke 2002, 231). Dieses Verhalten lässt sich dadurch erklären, dass Menschen nur bei der Kommunikation mit anderen sozialen Akteuren gewohnt sind, natürliche Sprache zu verwenden. Die Anzeige eines anthropomorphen Wesens oder zumindest eines Gesichts ermöglicht es dem menschlichen Bewusstsein, den Rechner als sozialen Akteur zu akzeptieren. Dies führt dazu, dass nach dem Denkschema des Benutzers nicht mehr mit einer Maschine, sondern mit einem anderen sozialen Akteur interagiert wird. Cheyer und Julia (1999, 1) beschreiben diesen Effekt so, dass das Reden mit einem Computer ohne virtuelle Person wie das Reden gegen eine Wand sei.

Auch wenn die Anzeige eines Avatars die Hemmschwelle für die Verwendung der Spracheingabe aus sozialer Sicht reduziert, sind die meisten Benutzer gewohnt, alle wählbaren Optionen eines Softwaresystems visuell zu erkennen. Redet ein Benutzer nur mit einem Avatar, so birgt dies die Gefahr, dass der eingeschränkte Funktionsumfang des dahinterliegenden Chatbots nicht bewusst ist und sich schnell Frustration darüber einstellt, dass das Avatarsystem nicht alle Spracheingaben verarbeiten kann (André/Rist/Müller 1998b, 26).

Während in diesem Kapitel die sozialen Effekte der reinen Präsenz eines Avatarsystems auf den Benutzer beleuchtet wurden, ist der Einfluss der Gestaltung des Avatars hinsichtlich emotionaler Aspekte bislang nicht betrachtet worden. Ein bewusstes Design des

Avatarsysteme im Sinne des Affective Computing (vgl. Picard 2000) zur emotionalen Gestaltung des Gesprächsflusses kommt im Rahmen der Erstellung der Wissensbasis in Kapitel 6.2 zum Tragen.

3.5.4. Einsatzgebiete und Beispiele

Avatarsysteme nach dem Verständnis dieser Arbeit lassen sich in zwei Einsatzbereiche unterteilen. Zum einen werden sie in Webauftritten von Unternehmen für Entertainment oder Verkaufszwecke eingesetzt. Zum anderen wurden viele Systeme für die Grundlagenforschung in Form von Prototypen entwickelt. Dabei sind Letztere hinsichtlich aller relevanten Aspekte weiter entwickelt als die kommerziellen Systeme und weisen teilweise domänenspezifische Erweiterungen auf.

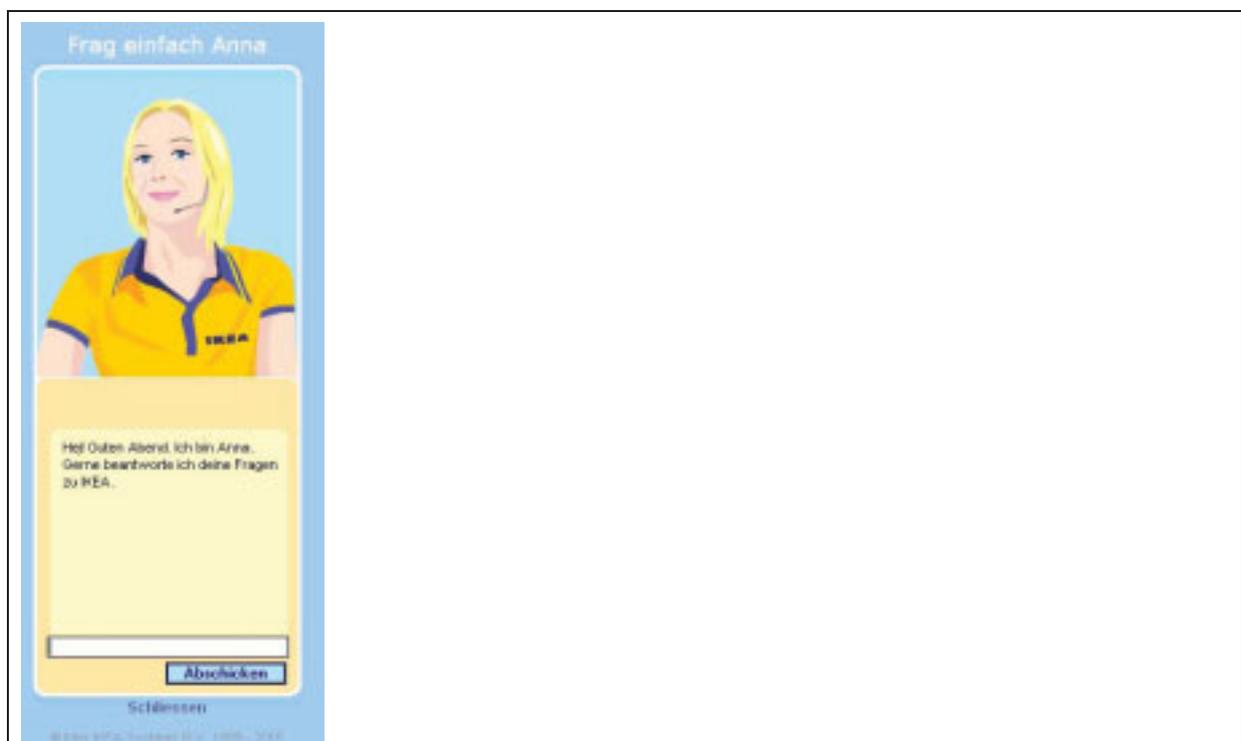


Abbildung 3-21: *Beispiel eines kommerziellen Avatarsystems im Internet: Anna von IKEA*
Quelle: (<http://www.ikea.de>, zugegriffen am 16.02.2008)

Wie in Ansätzen bereits in Kapitel 3.4.3.2 beschrieben, werden Avatarsysteme kommerziell überwiegend im Internet eingesetzt. Während dabei zu Beginn überwiegend die Aufgabe des Entertainment und Marketing im Vordergrund stand, sind heute nur noch wenige Systeme mit diesem Zweck anzutreffen (Bühler 2003, 112f; Buschmann 2003, 102ff). So dienen Avatare wie PUT (<http://www.schwarzepute.at>, zugegriffen am 29.07.2008) oder Chabba (<http://www.play.de/index.php>, zugegriffen am 29.07.2008) lediglich dem Chat mit den Besuchern der Webseiten ohne durch eine aktive Gesprächsführung auf Produkte oder Informationen hinzuweisen. Neuere Avatarsysteme im Internet dienen nicht mehr nur dem Zweck der Werbung, sondern haben meist die Aufgabe, Besucher aktiv zu relevanten Informationen zu führen oder bei der Auswahl von Produkten beratend zur Seite zu stehen. Dadurch sollen neben der Kundenakquise auch häufig wiederkehrende Fragen geklärt werden

(Samuelson 2003, 30ff). Als Beispiele lassen sich z.B. Anna (<http://www.ikea.de>, zugegriffen am 29.07.2008) oder Eve (<http://www.yellostrom.de>, zugegriffen am 29.07.2008) aufführen, die versuchen, Gesprächspartner durch entsprechende Fragen gezielt zu Informationen zu leiten (vgl. Abbildung 3-21). Bei der Verwendung von Avatarsystemen als Assistenten einer Website konnte gezeigt werden, dass die Benutzer beim Besuch der Seite mehr Spaß empfunden haben, mehr Vertrauen in die Seiten hatten und die Atmosphäre der Seite insgesamt verbessert werden konnte (Aberg/Shahmehri 2000, 2001, 2003).

Im Bereich der wissenschaftlichen Prototypen lassen sich bereits heute Technologien und Trends im Zusammenhang mit Avatarsystemen absehen, die erst in einigen Jahren den Sprung in den kommerziellen Einsatz schaffen können. Als besondere Meilensteine der Avatarsysteme sollen die beiden Systeme Peedy und REA vorgestellt werden. Weitere Systeme werden aus systemarchitektonischen Gesichtspunkten im Kapitel 5.1.1.1 näher betrachtet.



Abbildung 3-22: *Interaktion mit dem virtuellen Immobilienmakler REA*
Quelle: (Cassell 2001, 74)

Das Avatarsystem Peedy wurde in den 1990er Jahren von Microsoft entwickelt und verwendete als Avatar den Microsoft Agent Peedy, den Papageien. Das System konnte Lieder von Audio-CDs abspielen und wurde per Spracheingabe gesteuert. Die Ausgabe des Avatarsystems erfolgte durch Sprachsynthese. Der Agent konnte konkrete Interpreten und Titel aus der akustischen Eingabe erkennen und gegebenenfalls nach Rückfragen die gewünschten Lieder abspielen. Die Spracheingabe wurde aufgrund des eingegrenzten Einsatzbereiches semantisch analysiert, so dass auch unbekannte Satzformulierungen

verarbeitet werden konnten. Bei der grafischen Ausgabe wurde zwar darauf geachtet, dass die Animation des Papageien zum Inhalt passte, jedoch konnte dabei nicht von einer lippenynchronen Koordination gesprochen werden (Ball et al. 1997).

Während die Interaktion zwischen Peedy und dem Benutzer auf akustischer bzw. einseitig auch auf visueller Ebene erfolgte, ging das Forschungsteam um Justine Cassell (Cassell 2000a; Cassell et al. 1998; Cassell et al. 2001; Cassell/Vilhjálmsson/Bickmore 2001) am MIT mit dem Real Estate Agent (REA) einige Schritte weiter. So wurde die visuelle Interaktion in beide Richtungen hin erweitert, eine ausgeprägte Synchronisation von Lippen, Gesten und Sprache des Avatars umgesetzt, sowie Emotionen auf beiden Seiten berücksichtigt. Das Avatarsystem hatte die Aufgabe, dem Benutzer als virtueller Immobilienmakler zur Seite zu stehen und passende virtuelle Häuser zu präsentieren. Der Benutzer des Avatarsystems wurde in einem Labor durch Kameras erfasst, so dass seine Blickrichtung und Gesten erkannt und interpretiert werden konnten. Darüber hinaus wurde REA zusammen mit ihrer virtuellen Umwelt auf eine Leinwand projiziert, so dass sie in Lebensgröße erschien (vgl. Abbildung 3-22). Die Synchronisation der audiovisuellen Ausgabe wurde auf Buchstaben- bzw. Silbenebene genau mit Gestik, Mimik und Sprachsynthese abgestimmt, um einen möglichst menschlichen Eindruck zu erwecken. Die Emotionen des Avatarsystems wurden zum einen in einem internen Gefühlsmodell als auch in einem Gefühlsmodell des Benutzers erfasst und verarbeitet. Obwohl das Avatarsystem des MIT nur unter kontrollierten Laborbedingungen funktionierte und sich die Funktionsfähigkeit als Externer kaum nachvollziehen lässt, schlagen zumindest die Konzepte und Modelle, die dabei erstellt wurden, eine wegweisende Richtung ein.

3.6. Fahrerassistenz- und Fahrerinformationssysteme

Nach dem sich die vorangegangenen Grundlagenkapitel mit einzelnen Technologien bzw. im Falle des Avatarsystems mit deren Verbindung beschäftigt hat, fehlt nach wie vor der Bezug zum Einsatzbereich im Fahrzeug. Diese Lücke soll im Folgenden durch die strukturierte Betrachtung von Fahrerassistenz- (FAS) bzw. Fahrerinformationssystemen (FIS) geschlossen werden. Dies spielt insofern eine Rolle, als dass ein Avatarsystem beim Einsatz im Auto als FIS wahrgenommen wird. Da eine trennscharfe Unterscheidung zwischen FAS und FIS die Grundlage für die korrekte Einordnung des Avatarsystems darstellt, erfolgt dies in einem ersten Schritt. Darauf aufbauend werden verschiedene Beispiele vorgestellt und auf Basis der vorgestellten Definition eingeordnet.

3.6.1. Begriffliche Klärung

Die Unterscheidung zwischen Fahrerassistenz- und Fahrerinformationssystemen wird in vielen Fällen gar nicht oder nur oberflächlich gemacht, wobei aufgrund dieser Ungenauigkeit häufig von falschen Eigenschaften eines konkreten Systems ausgegangen wird (Bubb 2003). Zunächst soll aus diesem Grund der verwandte Begriff des FAS geklärt werden, um den des FIS eingrenzen und definieren zu können.

3.6.1.1. Eindimensionale Unterscheidungsansätze

Ein häufig verwendetes Differenzierungskriterium für Systeme im Fahrzeug ist die Verwendung für einen der drei Funktionsbereiche im Fahrzeug¹⁷. Der primäre Funktionsbereich deckt die Aufgaben des Fahrers zur Quer- und Längssteuerung des Fahrzeugs ab, wie z.B. das Beschleunigen oder Lenken. Der sekundäre Funktionsbereich beinhaltet Tätigkeiten zur Wahrung der Sicherheit für den Fahrer, das Fahrzeug und die Umwelt wie. z.B. die Verwendung des Scheibenwischers oder des Abblendlichts. Tertiäre Aufgaben dienen der Information oder Unterhaltung des Fahrers und beinhalten beispielsweise die Bedienung von Radio und Navigationssystem (Geiser 1985; Tönnis/Broy/Klinker 2006, 128f; Bubb 2003, 28). FAS arbeiten im primären und sekundären Bereich und greifen dabei teil-autonom oder vollkommen selbstständig ein, um die Sicherheit des Fahrzeugs, seiner Insassen und der Umwelt sicherzustellen. Dies schließt auch das Notbremsen oder Ändern der Lenkrichtung ein, welche im primären Bereich angesiedelt sind. FIS sind überwiegend im tertiären Funktionsbereich angesiedelt (Görtz 2006, 12).

Vergleicht man diese klare Zuordnung von FAS und FIS zu Funktionsbereichen, so erkennt man, dass die Zuordnung von FAS implizit annimmt, dass solche Systeme autonom ohne die Interaktion mit dem Fahrer agieren. Betrachtet man hingegen die allgemeine Definition von Assistenzsystemen aus Sicht der Mensch-Maschine-Interaktion, so sollen sie Teilaufgaben zur Erfüllung einer Gesamtaufgabe lösen sowie über Zustände informieren und zusätzliche Handlungsvorschläge unterbreiten (Timpe 1998). Dadurch wird auch Assistenzsystemen die bidirektionale Kommunikation mit dem Benutzer zugesprochen. Zudem lässt die Betrachtung eines FIS als informatorisches FAS (Meier-Arendt/Abel 2003, 76) die oben genannte trennscharfe Unterteilung unzulänglich erscheinen. Diese Ansicht unterstützt auch Stiller (2005), der FIS neben Komfortsystemen, effizienzsteigernden Systemen und autonomen Systemen als eine mögliche Ausprägung der FAS sieht.

3.6.1.2. Mehrdimensionale Unterscheidungsansätze

Einen ganzheitlichen Ansatz zur Klassifikation von FAS und damit implizit auch von FIS stellen Wandke, Wetzenstein und Polkehn (2005) vor. Sie entwickeln dabei ein dreidimensionales Modell entlang der Kriterien der unterstützten Handlungsphasen, der Initiative und der Anpassbarkeit (vgl. Abbildung 3-23) Dabei ist zu berücksichtigen, dass einzelne Kombinationen von Merkmalsausprägungen als Bausteine komplexer Systeme aus mehreren Bausteinen dienen können.

Sie unterscheiden dabei sechs **Handlungsphasen** während des Autofahrens, die durch ein Assistenzsystem unterstützt werden können. Zugunsten einer operationalisierbaren Abstraktion werden dabei nicht alle Teile menschlichen Handelns betrachtet. Bei einer Aktivität des Fahrers werden nicht zwangsläufig alle Phasen durchlaufen. Aufgrund zeitlicher Restriktionen können beispielsweise in kritischen Situationen auch nur wenige Phasen zum Tragen kommen. Sie identifizieren dabei folgende Phasen:

¹⁷ Alternative Ansätze unterscheiden beispielsweise nach der Verteilung der erbrachten Funktion eines Assistenzsystems zwischen den beiden Extrempunkten Mensch und Maschine (Sheridan 1988; Endsley/Kaber 1999).

1. Motiv- und Zielbildung
2. Informationsaufnahme
3. Informationsanalyse und –integration
4. Entscheidung
5. Aktionsausführung
6. Effektkontrolle (Wandke/Wetzenstein/Polkehn 2005, 47ff).

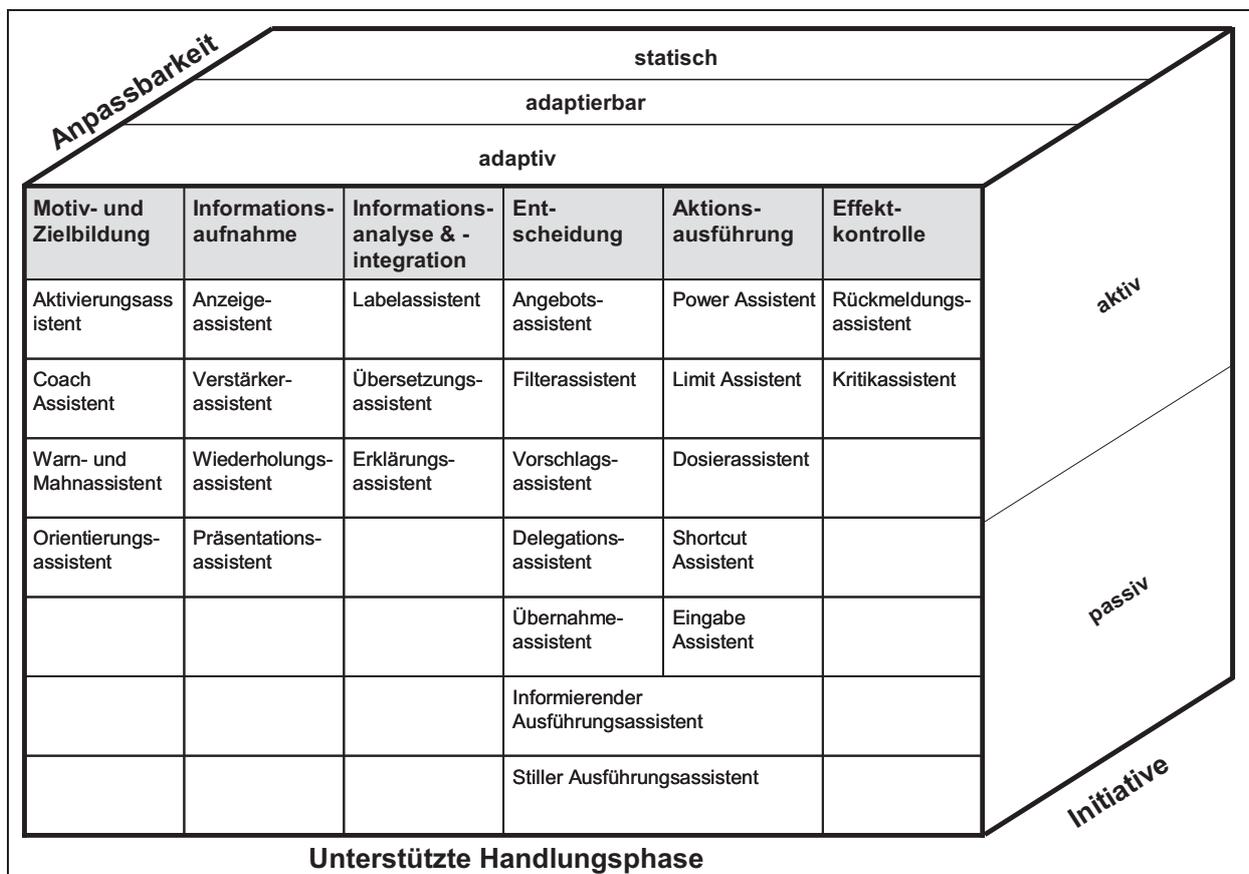


Abbildung 3-23: Modell zur Identifizierung von Bausteinen eines Fahrerassistenzsystems
 Quelle: (Wandke/Wetzenstein/Polkehn 2005, 58)

Diese sechs Phasen lassen sich aufgrund ihrer Eigenschaften grob in zwei Bereiche unterteilen: Den Informations- und den Aktionsteil. In den ersten vier Phasen steht der Aspekt der Informationsgewinnung im Vordergrund. Die letzten beiden Phasen haben die Aktionsausführung zur Aufgabe, wobei der letzten Phase zu einem gewissen Teil ebenfalls informatorische Aufgaben zufallen.

Prinzipiell kann jede der genannten Phasen dabei durch ein Assistenzsystem auf passive oder aktive Weise unterstützt werden, was eine Vereinfachung des von Endsley und Kaber (1999)

erstellten zehnstufigen Modells ist und die **Initiative** repräsentiert. Als passiv gilt ein Assistenzsystem demnach, wenn es nur durch Aufforderung des Fahrers aktiviert wird. Ein aktives System hingegen nimmt seine Funktion ohne menschliches Zutun auf, sobald bestimmte äußere Einflüsse es erforderlich machen. Es sind auch Kombinationen von aktiv und passiv denkbar, wenn beispielsweise nur die initiale Aktivierung durch den Benutzer erfolgt und das System danach selbstständig arbeitet (Wandke/Wetzenstein/Polkehn 2005, 55f).

Die dritte Dimension der **Anpassbarkeit** gibt an, inwieweit der Benutzer auf die Funktion des Assistenzsystems Einfluss nehmen kann. Ein statisches Assistenzsystem kann dabei nach Auslieferung nicht mehr angepasst werden, was sowohl für den Benutzer als auch für den Hersteller einen festgelegten Funktionsumfang sicherstellt. Dieser Typ von Systemen bietet sich bei sicherheitskritischen Systemen oder bei Systemen an, die durch eine Anpassung keinen Mehrwert für den Kunden liefern würden. Im Gegensatz dazu sind adaptierbare Systeme durch den Benutzer an seine Präferenzen oder Bedürfnisse anpassbar. Adaptive Systeme passen sich selbstständig ohne den Eingriff des Benutzers an. Dabei kann zwischen einer einmaligen, initialen und einer kontinuierlichen Anpassung unterschieden werden (Wandke/Wetzenstein/Polkehn 2005, 56ff).

Aus diesem detaillierten mehrdimensionalen Unterscheidungsansatz lässt sich folgende Definition für ein FIS zusammenfassen:

Ein Fahrerinformationssystem ist ein Fahrerassistenzsystem, das auf aktive oder passive Weise die Handlungsphasen, insbesondere jedoch die Phasen der Entscheidungsvorbereitung und –findung mit informatorischem Charakter unterstützt.

3.6.2. Beispiele und deren Einordnung

Basierend auf der detaillierten Unterscheidung verschiedener Assistenzsystembausteine lassen sich einfache Systeme aus einzelnen Bausteinen, komplexe Systeme unter Verwendung mehrerer Bausteine sowie allgemeine Hilfesysteme unterscheiden.

3.6.2.1. Assistenten aus einzelnen Bausteinen

Die Phasen des vorgestellten mehrdimensionalen Modells von Wandke, Wetzenstein und Polkehn (2005) lassen sich als Orientierungspunkte für die Vorstellung verschiedener Beispiele von FAS und FIS verwenden. Dabei lässt sich in bestimmten Phasen eine Häufung weit verbreiteter Assistenzsysteme identifizieren, während für andere Phasen bisher überwiegend nur Prototypen erstellt wurden.

In der Phase der **Motiv- und Zielbildung** lassen sich Aktivierungs-, Coach, Warn- und Mahn- sowie Orientierungsassistenten identifizieren (Wandke/Wetzenstein/Polkehn 2005, 49f). Aktivierungsassistenten verfolgen die Aufgabe, das Aktivierungs- bzw. Wahrnehmungsniveau des Fahrers aufrecht zu erhalten. So lassen sich an dieser Stelle Einschlafwarner nennen, die jedoch erst in einem prototypischen Stadium existieren. Das gleiche gilt für die Meldung einer Geschwindigkeitsüberschreitung, welche als Warnassistent

den Fahrer auf eine korrekte Fahrweise hinweist. Andere Beispiele sind Warnassistenten für Hindernisse auf der Fahrbahn oder rote Ampeln (Schierge 2005, 209f). Ein weit verbreiteter Orientierungsassistent ist beispielsweise das Navigationssystem, das Ziel- bzw. Richtungswechsel anregen kann. Die genannten Beispiele dieser Phase sind aktiv und statisch, während das Navigationssystem initial passiv ist und durch den Benutzer angepasst werden kann.

Die Phase der **Informationsaufnahme** ist durch Anzeige-, Verstärker-, Wiederholungs- sowie Präsentationsassistenten gekennzeichnet (Wandke/Wetzenstein/Polkehn 2005, 50f). Anzeigeassistenten wie die Erweiterung eines Head-up Displays um Elemente einer *Augmented Reality* können aktiv zusätzliche Informationen in das Blickfeld des Fahrers einblenden. Dies kann im Rahmen eines Verstärkerassistenten genutzt werden, um Signale technisch zu verstärken und dadurch beispielsweise Nachtfahrten durch Infrarotsicht und Radardaten zu ergänzen. Ein Wiederholungsassistent kann zusätzlich Verkehrszeichen wie Geschwindigkeitsbegrenzungen oder Überholverbote während ihrer Gültigkeit auf einem Display im Fahrzeug anzeigen und somit als Gedächtnisstütze fungieren. Die Transformation einer Ausgabemodalität in eine angemessenere Modalität übernimmt der Präsentationsassistent, der beispielsweise implizit in einem Navigationssystem enthalten ist und visuelle Informationen in Akustische überführt. Die oben genannten Systeme sind nach einer initialen Aktivierung und Anpassung durch den Fahrer aktiv.

Im Rahmen der **Informationsanalyse und -integration** lassen sich Label-, Übersetzungs- und Erklärungsassistenten unterscheiden (Wandke/Wetzenstein/Polkehn 2005, 51f). Während der Labelassistent Beschreibungen zu Ein- und Ausgabeelementen zur Verfügung stellt, bietet der Erklärungsassistent Hilfe zur Bedienung von Systemfunktionen und Handlungsoptionen an. Beide Assistenzsysteme sind statisch und passiver Natur. Sie sind in Ansätzen bereits umgesetzt, so z.B. im Demomodus von Navigationssystemen und der Ausgabe von akustischen Hilfetexten nach der Nennung bestimmter Sprachkommandos wie „Hilfe“. Übersetzungsassistenten, die ebenfalls passiv, aber adaptierbar arbeiten, können den Bezug zu externen Systemen wie beispielsweise einem iPod oder dem Handy herstellen.

Während der **Entscheidungsphase** bedarf der Fahrer besonders vieler Systeme, die sich im Angebots-, Filter-, Vorschlags-, Delegations-, Übernahme- und im informierenden oder stillen Ausführungsassistenten äußern (Wandke/Wetzenstein/Polkehn 2005, 52f). Ein Angebotsassistent kann ähnlich wie der Zielspeicher im Navigationssystem alle bereits gespeicherten Ziele präsentieren und dadurch den Eingabevorgang vereinfachen. Eine ähnliche Aufgabe übernimmt der Filterassistent, der z.B. beim Auswählen von Radiofrequenzen aus einer bekannten Menge an Optionen eine Auswahl präsentiert. Im Gegensatz dazu wählt der Vorschlagsassistent bereits die passendste Alternative (z.B. die kürzeste Route im Navigationsgerät) aus. Über den Vorschlag einer Auswahl hinaus, führt ein Delegationsassistent diese auch aus, wie dies beispielsweise bei einem automatischen Einparksystem der Fall ist. Ausführungsassistenten führen Aktionen durch und informieren den Fahrer darüber im Fall der informierenden Systeme. Stille Systeme informieren den Fahrer nicht über ihre Handlung. Der Großteil der Assistenzsysteme in dieser Phase ist passiv und weist je nach Anwendungsgebiet alle Formen der Anpassbarkeit auf.

Die **Ausführung einer Aktion** wird nach Wandke, Wetzenstein und Polkehn (2005, 53f) durch Power-, Limit-, Dosier-, Shortcut- oder Eingabeassistenten unterstützt. Während der Power-Assistent eine Aktion des Fahrers verstärkt (z.B. Bremskraftverstärker), beschränken Limit-Assistenten eine Aktion, wie dies beispielsweise bei einem Tempomat der Fall ist. Ein Dosier-Assistent verteilt die Aktion eines Benutzers wie z.B. beim ABS, welches die Bremsaktion zeitlich verteilt. Shortcut-Assistenten erleichtern den Zugriff auf bestimmte Funktionen oder Einstellungen, wie z.B. die tiptronic zum Schalten der Gänge per Wippen am Lenkrad oder das Speichern von Sitzeinstellungen in Profilen. Eingabeassistenten vereinfachen die Eingabe, indem sie beispielsweise zusätzliche Eingabemodalitäten wie Spracheingabe, Touchscreens oder Gestenerkennung anbieten. Bis auf die Shortcut-Assistenten sind die Systeme dieser Phase aktiv und teils statisch wie im Falle von Bremsassistent, ABS und tiptronic oder anpassbar im Falle von Tempobegrenzern und Speicherfunktionen.

Die **Rückmeldung** zu erfolgten oder geplanten Aktionen während der Effektkontrolle lässt sich in Rückmeldungs- und Kritikassistenten unterteilen (Wandke/Wetzenstein/Polkehn 2005, 54). Rückmeldungsassistenten wie eine Lenkwinkelanzeige beim Rückwärtsfahren oder die aktuelle Verbrauchsanzeige informieren während oder nach einer Aktion über die Auswirkungen. Ein Kritikassistent, der den Fahrer lobt oder tadelt, existiert zwar noch nicht, könnte jedoch den Fahrer zu rücksichtsvollem oder umweltschonendem Fahren erziehen.

3.6.2.2. *Kombinierte Assistenzsysteme*

Häufig verwenden Assistenzsysteme mehr als einen der beschriebenen Bausteine und lassen sich somit nicht mehr nur in eine Handlungsphase einordnen. Ein komplexes Beispiel ist die Adaptive Cruise Control (ACC), welche eine Kombination aus Geschwindigkeits- und Abstandsregelanlage darstellt. Sie ermöglicht es einem Fahrzeug, in einer bestimmten Geschwindigkeit zu fahren, wobei ein vorgegebener Abstand zum davor fahrenden Fahrzeug eingehalten wird (Bubb 2003; Sacher/Bubb 2006). Dieses System stellt somit eine Kombination aus Power- und Limit- sowie eines informierenden Ausführungsassistenten dar. Durch Power- und Limit-Assistenten werden Geschwindigkeit und Abstand zum Vorfahrenden gesteuert, während der Ausführungsassistent den Fahrer zusätzlich über den aktuellen Status informiert. Dies gibt ihm die Möglichkeit, selbst einzugreifen, wenn das vordere Fahrzeug beispielsweise eine Notbremsung macht. Das System als Ganzes ist als passiv und adaptierbar einzustufen, da es erst auf die Eingabe des Fahrers hin aktiv wird und sowohl Geschwindigkeit als auch Abstand vom Fahrer eingestellt werden können. Der einzelne Baustein des Ausführungsassistenten hingegen kann in diesem Zusammenhang als aktiv und statisch betrachtet werden. Ein über das ACC hinausgehender Stauassistent bzw. Stop-and-Go-Assistent (Hipp/Schaller 2003, 55ff) soll dem Fahrer beim Fahren und Abstandhalten in Stausituationen helfen (Schierge 2005, 213).

Als weiteres Beispiel lässt sich ein Einparksystem anführen, das zum einen in einer passiven und zum anderen in einer aktiven Variante möglich ist. Das passive System beinhaltet nur Bausteine der Informationsaufnahmephase und erleichtert das Einparken durch einen Präsentations- und einen Verstärkerassistenten. Der Präsentationsassistent wandelt die Information über den Abstand zu Hindernissen in akustische Signale, die mit abnehmendem

Abstand an Intensität zunehmen oder deren Intervall sich verkürzt. Der Präsentationsassistent kann beim Rückwärtsfahren das Bild einer Rückfahrkamera einblenden und den aufgrund des eingeschlagenen Lenkwinkels errechneten Weg grafisch anzeigen. Das aktive System kann zudem mithilfe eines stillen Ausführungsassistenten auch automatisch das Parken durchführen (Airaksinen et al. 2004).

3.6.2.3. Allgemeine Hilfesystemen im Fahrzeug

Als eine besondere Ausprägung kombinierter Assistenzsysteme mit dem Fokus auf Fahrerinformation können Hilfesysteme im Fahrzeug gewertet werden. Als Hilfesystem bezeichnet Bauer (1988) Computerprogramme, die eine explizite Erklärung geben und sich auf die interaktive Benutzung eines Computerprogramms beziehen. Je nach Art der Hilfe unterscheidet Krause (1988) tutorielle Systeme, die eine Anleitung für den strukturierten Wissenserwerb geben, und Auskunftssysteme, die als Nachschlagewerk in einer bestimmten Situation dienen. Der Bezug auf Computerprogramme bei der Definition des Begriffs zeigt zudem, dass sich diese Art der Systeme im Auto aufgrund der Vergleichbarkeit an den Fortschritten im Computerbereich orientiert (vgl. auch Drews/Frees/Schellhase 2006, 129).

Die Umsetzung der Hilfesysteme im Fahrzeug gestaltet sich technisch sehr verschieden, wobei in Premiumfahrzeugen sprachgesteuerte Systeme, wie sie von Vollrath und Totzke (2003) vorgeschlagen werden, dominieren. Die Sprachausgabe solcher Systeme, wie sie von Salmen (2002, 111ff) untersucht wurde, wird dabei jedoch nur in eingeschränktem Umfang für akustische Feedbacks genutzt (Hamberger/Mauter 2003, 225). Auch wenn durch audiovisuelle Ausgaben eine rudimentäre Modalität bei Ein- und Ausgabe der Hilfesysteme vorliegt, ist diese ausbaufähig und durch den Aspekt der Multimedialität erweiterbar (vgl. Marrenbach et al. 2001).

Eine detaillierte Gegenüberstellung bestehender, sprachdialogbasierter Hilfesysteme in Premiumfahrzeugen hat Hof (2007, 49ff) durchgeführt. Dabei wurden die Hilfesysteme von Audi, BMW, Ford, Honda, Infiniti, Lexus und Mercedes betrachtet. Er fand heraus, dass alle genannten Systeme passiv sind und nur durch die Aktivierung des Benutzers Informationen ausgeben. Die geleistete Hilfestellung erfolgt hinsichtlich des Hilfezeitpunkts bei allen Systemen synchron zur aktuellen Situation des Fahrers und dabei hoch integriert in das gesamte Bediensystem. Es lässt sich jedoch kein System individuell an den Fahrer anpassen. Unterschiede ergaben sich hinsichtlich des Kontextbezugs und der Multimodalität. So erfolgt mit Ausnahme von Infiniti und Lexus bei allen Hilfesystemen der Kontextbezug dynamisch. Hinsichtlich der Multimodalität erlaubt nur das System von Mercedes eine haptische Bedienung während der Sprachausgabe des Systems (Hof 2007, 54f). Neben den grundlegenden Merkmalen der Systeme bewertet Hof (2007, 60f) zudem die inhaltliche Struktur der Systeme anhand folgender Kriterien:

- Selbstbeschreibungsfähigkeit
- Aufgabenangemessenheit

- Lernförderlichkeit
- Steuerbarkeit
- Individualisierbarkeit
- Erwartungskonformität
- Fehlertoleranz
- Geringe Belastung des Kurzzeitgedächtnisses
- Reversibilität
- Unterteilung einer Sprachausgabe in Anfang, Mitte und Ende
- Vermeidung von Redundanz

Auch wenn dabei die Systeme von Audi, BMW und Mercedes sehr gut bewertet werden, sind insbesondere Handlungsbedarfe in den Bereichen Individualisierbarkeit, Fehlerbehandlung und Reversibilität zu erkennen (Hof 2007, 64). Diese Lücken können in der zukünftigen Weiterentwicklung der Systeme beispielsweise mit Ansätzen von Christ und Bauer (2003) oder Lilienthal (2003) geschlossen werden.

3.7. Zusammenfassung

Die Komplexität der Umsetzung eines Avatarsystems als interaktive Hilfe in einem Fahrzeug macht die Betrachtung vieler Basistechnologien notwendig. Die Sprachsynthese in etwa als einfach während der Fahrt wahrzunehmende Ausgabe hat im Laufe der letzten Jahrzehnte eine deutliche Entwicklung erfahren und erlaubt mittlerweile sogar die Reproduktion komplexer Phrasen und der Satzmelodie. Dazu werden verschiedene Beschreibungssprachen für phonetische und prosodische Annotation verwendet, welche die künstliche Sprache menschlicher klingen lassen.

Für eine vollkommen sprachbasierte Interaktion bedarf es zudem der automatischen Spracherkennung, die mit komplexen Verfahren aus einem akustischen Signal mit phonetischer und linguistischer Analyse die gesprochenen Wörter und Phrasen erkennt. Dazu werden Lexika und Sprachgrammatiken verwendet, die zum einen bekannte Wörter und zum anderen Informationen über mögliche Wortfolgen enthalten.

Die Bildung eines Sprachdialogsystems verlangt neben den Sprachkomponenten zur Ein- und Ausgabe einen verarbeitenden Teil, die Logikkomponente. Da der Einsatz echter künstlicher Intelligenz bisher keine akzeptablen Ergebnisse liefern konnte, beschränkt man sich dabei meist auf regelbasierte Verarbeitungselemente. Ein Chatbot ist solch ein Element, das für die Simulation eines Gesprächspartners speziell im Internet eingesetzt wird. Die Wissensbasis, welche die notwendigen Regeln enthält, stellt dabei die Grundlage der Verarbeitung dar und kann in verschiedenen Datenformaten erstellt werden.

Der Einsatz im Internet und dort insbesondere für den Bereich des e-Commerce führte dazu, dass Chatbots mit Abbildungen virtueller Wesen kombiniert wurden, den Avataren. Sie versuchen den Benutzern die Existenz eines sozialen Akteurs zu suggerieren, um die Hemmschwelle zu reduzieren, sich mit einer Maschine natürlichsprachlich zu unterhalten. Durch technische Fortschritte, die insbesondere auch durch den Bereich der Computerspiele initiiert wurden, lassen sich heutzutage bereits dreidimensionale Wesen einsetzen, die in Echtzeit animiert werden.

Die Verbindung von Chatbots und Avataren zu einem virtuellen Wesen mit vermeintlich eigener Intelligenz führt zur Bildung von Avatarsystemen. Dabei werden die sozialen Effekte der beiden einzelnen Komponenten genutzt und synergetisch verwendet. Während Chatbots und Avatare bereits einzeln als soziale Akteure betrachtet werden, stellt ein Avatarsystem einen noch menschlicheren Gesprächspartner dar. Dabei stellen sich aus technischer Sicht neue Herausforderungen, die bei der Verwendung von Sprache als Medium mit der Abstimmung der verschiedenen technischen Komponenten verbunden sind. So ist beispielsweise auf eine lippensynchrone Sprachausgabe oder zusätzlich auf die Synchronisation von Inhalten, Gesten und Tonfall zu achten.

Die Übertragung solch eines Avatarsystems in ein Fahrzeug dient der Unterstützung des Fahrers und lässt sich unter den Oberbegriff der Fahrerassistenzsysteme fassen. Während gemeinhin zwischen Fahrerassistenz- und Fahrerinformationssystemen unterschieden wird, lässt sich zeigen, dass FIS bestimmte Aufgabenbereiche bzw. Handlungsphasen des Fahrers im Rahmen von FAS unterstützen. Die Betrachtung von Hilfesystemen als Ausprägungen kombinierter Assistenzsysteme zeigt dabei insbesondere Handlungsbedarfe in den Bereichen der Individualisierbarkeit, der Fehlerbehandlung und der Reversibilität.

4. Analyse und Anforderungserhebung

„Der Mensch ist das Maß aller Dinge.“

Protagoras (ca. 485 - ca. 415 v. Chr.), griechischer Philosoph

Gemäß dem kognitiven Modell des Designzyklus von Takeda (1990, 12), erfolgt nach der Erkennung der Problemstellung, was bereits in Kapitel 1 erfolgt ist, die Vorschlagsphase (im Englischen: Suggestion). Diese ist dadurch gekennzeichnet, dass ein grober Rahmen für mögliche Lösungswege erstellt wird. Dieser Rahmen muss sich dabei an den Eigenschaften und dem Umfeld des Problems orientieren, um den Lösungskorridor abzustecken. Es lässt sich somit ein analysierender und ein konzeptioneller Teil dieser Phase unterscheiden, welche getrennt in diesem und dem nachfolgenden Kapitel 5 betrachtet werden. Gegenstand dieses Kapitels ist zunächst die Analyse des bestehenden Umfelds.

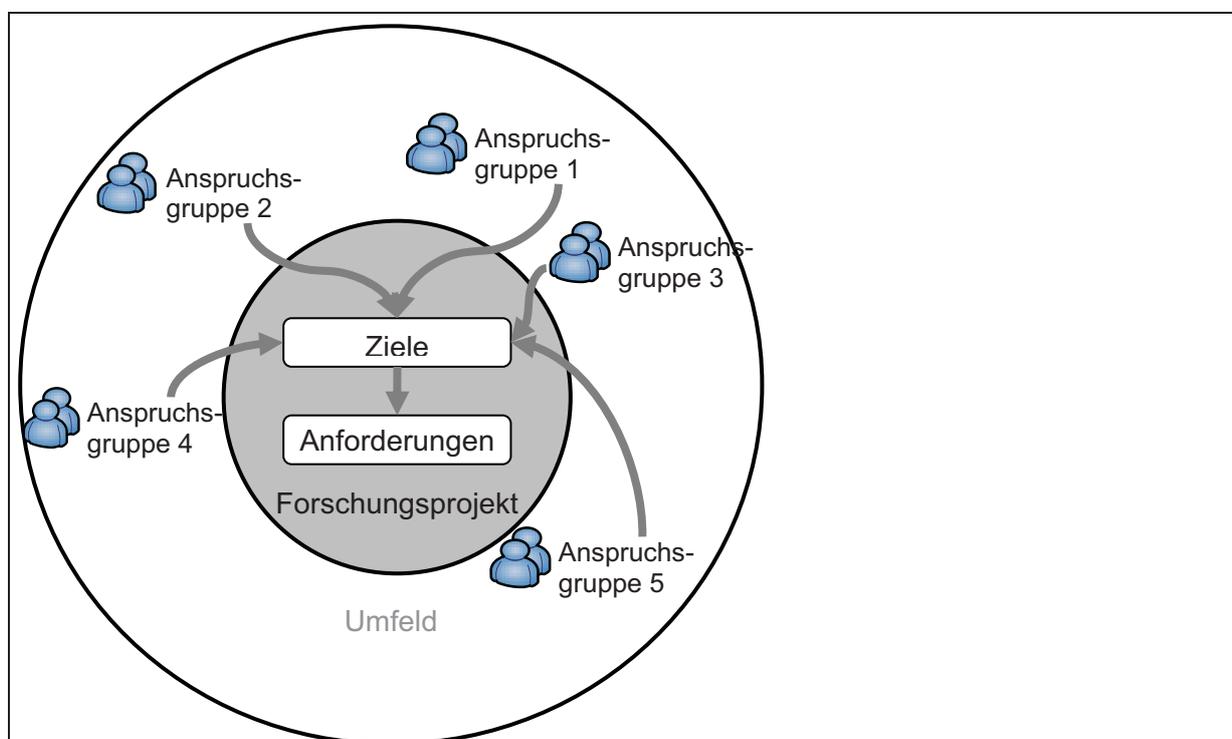


Abbildung 4-1: *Ableitung von Anforderung aus Zielen der Stakeholder*
Quelle: (Eigene Darstellung)

Der angestrebte Prototyp stellt die Umsetzung einer Vision und nicht die Aktualisierung oder Verbesserung eines bestehenden Systems dar, was die Fokussierung auf das Umfeld der Forschung bedingt. Darüber hinaus können die bereits in Kapitel 3.6.2.3 vorgestellten Systeme als verbesserungswürdige Grundlage herangezogen werden. Nach der Betrachtung des Forschungsumfeldes sollen die Anspruchsgruppen darin identifiziert und deren Ansprüche erhoben werden. Basierend auf der Sammlung dieser Äußerungen können die Ziele extrahiert und daraus Anforderungen abgeleitet werden (vgl. Rupp/die SOPHISTen 2007, 86f), was in Abbildung 4-1 visualisiert ist. Die Inhalte dieses Kapitels decken somit die Inhalte eines Lastenheftes ab, wie es z.B. bei Balzert (2001, 62ff) vorgestellt wird.

4.1. Betrachtung des Forschungsumfelds

Die strukturierte Betrachtung des Umfelds der in diesem Buch beschriebenen Forschung erlaubt es, beteiligte Organisationen, deren Untereinheiten sowie Beziehungen zwischen ihnen zu erkennen. Dies wird die Grundlage für das systematische Auffinden der Anspruchsgruppen in Kapitel 4.2 darstellen. Zudem können dadurch Entscheidungen im späteren Implementierungsprozess, die nicht explizit als Anforderungen formuliert wurden, nachvollzogen werden.

Die Erstellung des Prototyps erfolgt im Rahmen des Projektes „Avatar based Virtual Co-driver System“ (AViCoS), das in Zusammenarbeit der Technischen Universität München (TUM) mit dem Automobilhersteller AUDI AG aus Ingolstadt entstanden ist. Die Kooperation zwischen diesen beiden Organisationen findet im Rahmen der Public-Private-Partnership „Ingolstadt Institute der TUM“ (INI.TUM) statt. Vor diesem Hintergrund wird die Erforschung verschiedener Aspekte vorangetrieben, wobei jedes einzelne Forschungsprojekt individuell von einer Fachabteilung und einem Lehrstuhl der TUM betreut wird. Die finanziellen Mittel für die Bearbeitung sowie die Themen werden dabei von der AUDI AG gestellt. Die Betreuung von AViCoS wird durch die Fachabteilung für Prozess- und Systemintegration in der Elektrik und Elektronik auf Seiten von AUDI (I/FP-34) und durch den Lehrstuhl für Wirtschaftsinformatik I17 auf Seiten der TUM sicher gestellt. Parallel zu dem hier betrachteten Projekt AViCoS wird durch die gleichen Organisationseinheiten das Projekt „Highly Integrated Modular Embedded Prototyping Platform“ (HIMEPP) bearbeitet. Da verschiedene Überschneidungen zwischen den beiden Projekten bestehen, erfolgt ein reger inhaltlicher Austausch, der später in Kapitel 6.1 detaillierter beschrieben wird.

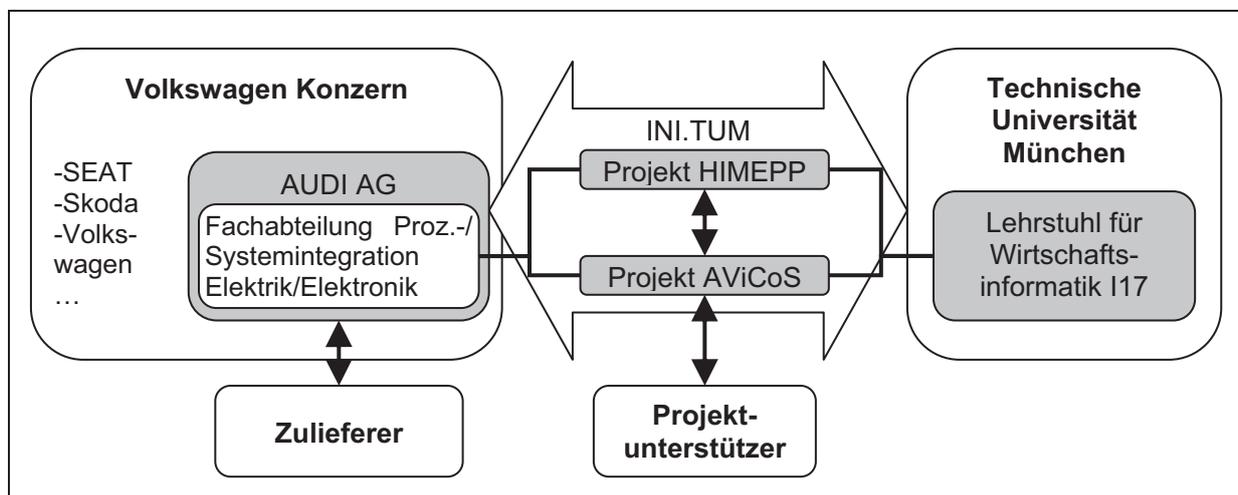


Abbildung 4-2: Organisationen und deren Zusammenhänge im Forschungsumfeld
Quelle: (Eigene Darstellung)

Neben den direkt am Forschungsprojekt beteiligten Einheiten sind weitere Organisationen zu berücksichtigen. So ist die AUDI AG in den Volkswagen Konzern eingegliedert, in dem auch weitere Automobilmarken anzutreffen sind (vgl. dazu auch http://www.volkswagenag.com/vwag/vwcorp/content/de/brands_and_companies.html, zugegriffen am 9.3.2008). Die Nähe der verschiedenen Marken ermöglicht einen Austausch

auf Forschungsebene. Die Bestrebungen aller Automobilbauer, ihre Fertigungstiefe zu reduzieren und damit viele Tätigkeiten an andere Unternehmen auszulagern, führen auch bei AUDI dazu, dass ein ausgeprägtes Netzwerk an Zulieferunternehmen in die Entwicklungsprozesse integriert ist. Dazu gehören auch ausgegründete Teile der Forschung und Entwicklung. Im Gegensatz zu Zulieferern, die bereits eine Geschäftsbeziehung mit AUDI oder dem Volkswagen Konzern haben, streben andere Unternehmen diesen Status an. Einige dieser Unternehmen unterstützen gezielt Forschungsprojekte im Rahmen von INI.TUM mit der Erwartung dadurch eine vertraglich abgesicherte Geschäftsbeziehung zu AUDI aufbauen zu können. Die genannten Organisationen und ihre Zusammenhänge sind in Abbildung 4-2 dargestellt.

Im Rahmen dieses Forschungsumfeldes ist das Ziel des Vorhabens „AViCoS“ ein prototypisches Avatarsystem im Fahrzeug umzusetzen. Als Anwendung auf diesem System soll ein interaktives, multimediales Handbuch umgesetzt werden, das per Spracheingabe und –synthese einen natürlichsprachlichen Dialog zu Aspekten der Fahrzeugbedienung ermöglicht. Auf Grundlage dieses Systems sollen erste Aussagen getroffen werden können, welche Auswirkungen beim Einsatz eines solchen Systems in einem Fahrzeug erwartet werden können. Die beispielhafte Integration in ein Serienfahrzeug wird durch die Bereitstellung eines entsprechenden Testträgers sowie zusätzlicher elektronischer Elemente durch die AUDI AG ermöglicht.

4.2. Identifikation von Stakeholdern

Gemäß DIN 69905, welche die Begrifflichkeiten im Rahmen einer Projektabwicklung regelt, werden Stakeholder im Sinne eines Projektes als Personen, Institutionen und Dokumente definiert, die in einem direkten oder indirekten Bezug zu einem Projekt stehen. Dies schließt neben Projektbeteiligten, die direkt die Gestaltung beeinflussen auch diejenigen ein, welche das Ergebnis des Projektes bedienen, warten oder in Schulungen weiter vermitteln müssen und somit nur indirekt mit dem Projektablauf in Verbindung stehen (vgl. dazu auch Rupp/die SOPHISTen 2007, 92). Somit dienen Stakeholder als primäre Quelle für Anforderungen an ein Projekt.

Im Folgenden werden auf Basis des vorgestellten Forschungsumfeldes systematisch die möglichen Stakeholder identifiziert und beschrieben. Dazu wird die von Rupp et al. (2007, 93) vorgeschlagene Systematik von Personen und Institutionen herangezogen, um in jeder organisatorischen Einheit entsprechende Beteiligte auszumachen. Darüber hinaus werden umfeldspezifische Dokumente sowie die Kritikpunkte aus anderen Avatarprojekten und aus den in Kapitel 3.6.2.3 vorgestellten Hilfesystemen verwendet. Die Betrachtung der Stakeholder erfolgt zweigeteilt zum einen innerhalb und zum anderen außerhalb des Volkswagen Konzerns. Dabei wurden mit Ausnahme der Dokumente die Ansprüche der Stakeholder in zahlreichen offenen Gesprächen mit Vertretern der Organisationseinheiten bzw. Unternehmen ermittelt.

4.2.1. Stakeholder innerhalb des Volkswagen Konzerns

Die meisten Stakeholder des Forschungsvorhabens AViCoS innerhalb des Volkswagen Konzerns sind bei der AUDI AG in Ingolstadt zu finden. Der wichtigste Stakeholder innerhalb der AUDI AG ist die Fachabteilung **Prozess- und Systemintegration Elektrik/Elektronik (I/FP-34)**, welche als Projektinitiator innerhalb des Unternehmens das direkte Management des Projekts übernimmt. Sie übernimmt innerhalb des Unternehmens die Koordination und Realisierung von Querschnittsaufgaben bei der Vorserienentwicklung neuer elektronischer Systeme bzw. der Erstellung von Rahmenwerken und Konzepten zum einheitlichen Umgang mit den Funktionen der Bordelektronik. Neben der organisatorischen Betreuung erfolgt somit auf inhaltlicher Ebene die Begleitung bei der Einbindung in die bestehende technische Fahrzeuginfrastruktur.

In der gleichen funktionalen Einheit angesiedelt, stellt die Abteilung **Virtuelle Realität (I/FP-32)** dreidimensionale, grafische Daten zur Verfügung, die auf Basis von Konstruktionszeichnungen erstellt werden. Analog dazu übernimmt die **Verkaufsabteilung Bordliteratur (I/VK-37)** die Bereitstellung von Inhalten aus den Bordhandbüchern in digitaler Form. Zudem würde ihr die Aufgabe zufallen, die Inhalte des avatarbasierten Handbuchs zu pflegen, sofern es zum Serieneinsatz kommen sollte.

Die Abteilung **Entwicklung Bedienkonzept (EE-71)** spielt für das Projekt eine besondere Rolle. Auch wenn sie keinen direkten Beitrag zum Projekt leistet, verfügt sie aufgrund der Erfahrung in der Serienentwicklung von Bedienelementen über umfangreiches Wissen. Dieses kann genutzt werden, um den Prototypen des avatarbasierten Handbuchs bereits in einem frühen Stadium nach Kriterien der Serienentwicklung zu gestalten und damit die Akzeptanz des neuen Systems innerhalb der AUDI AG zu erhöhen.

Die Aufgabe eines Projektponsors innerhalb der AUDI AG kommt der Abteilung **Wissenschaftsprojekte (I/SY)** zu. Sie ist für die Werbung für und die Vernetzung von INI.TUM-Projekten zuständig. Durch verschiedene Veranstaltungen und Informationsmaterial wird ermöglicht, dass der innerbetriebliche Informationsaustausch zwischen einzelnen Fachabteilungen und INI.TUM-Projekten gezielt vorgenommen werden kann.

Die AUDI Electronics Venture GmbH (AEV) ist ein Spin-Off der AUDI AG und mit der Erforschung und Entwicklung von Technologien beschäftigt, die nicht nur im eigenen Unternehmen genutzt werden. Die Abteilung **Elektronik Vorentwicklung (AEV-3)** ist vorwiegend mit der Umsetzung und Bewertung von prototypischen Umsetzungen befasst und steht im engen inhaltlichen Austausch mit der Abteilung I/FP-34. Dadurch kann auf bestehende Softwarekomponenten der AEV-3 zugegriffen werden, wobei die im Rahmen des Avatarprojekts entstehenden Komponenten auch der AEV-3 bereitgestellt werden. Während der Arbeiten an AViCoS ist zudem ein reger Austausch zur Umsetzung von Avatarsystemen im Fahrzeug aufgekommen.

Organisation	Untereinheit	Rolle	Grund für Aufnahme als Stakeholder	Entscheidungsbefugnis	Grad der Beteiligung
AUDI AG	I/FP-34, Prozess- und Systemintegration Elektrik, Elektronik	Experten für das Systemumfeld	Bereitstellung von Integrationshardware für den Prototypen	Gesamte organisatorische und inhaltliche Ausgestaltung auf AUDI-Seite	Hoch
		Management	Steuerung des Projektes		
	I/FP-32, Virtuelle Realität	Interner Zulieferer	Bereitstellung visueller Fahrzeugdaten	Aufarbeitung und Bereitstellung von 3D-Daten	Niedrig
	I/VK-37, Bordliteratur	Interner Zulieferer	Bereitstellung bestehender Handbuchinhalte	Aufarbeitung, Bereitstellung und Darstellung von Inhalten des Handbuchs	Mittel
		Betreiber	Pflege des avatargestützten Handbuchs bei Serieneinsatz		
	I/EE-71, Entwicklung Bedienkonzept	F+E Repräsentant	Know-how zur Überführung in die Serie	Inhaltliche Ausgestaltung auf AUDI-Seite	Mittel
I/SY, Wissenschaftsprojekte	Projektsponsor	Organisation von INI.TUM	Organisatorische Ausgestaltung auf AUDI-Seite	Niedrig	
AUDI Electronic Venture GmbH	I/AEV-3, Elektronik Vorentwicklung	Entwickler	Bereitstellung von Softwarekomponenten	Keine	Niedrig
		Anwender	Nutzer von Softwarekomponenten des Avatarprojekts		
VW Konzern	K-GEFE/H, Konzernforschung	Technische Experten	Seriennahe Erforschung von Avatarsystemen	Entscheidung über Projektfortführung	Niedrig

Tabelle 4-1: *Stakeholder und ihre Eigenschaften innerhalb des Volkswagen Konzerns*
Quelle: (Eigene Darstellung)

Der letzte Stakeholder innerhalb des VW Konzerns ist schließlich die **Konzernforschung (K-EFE/H)**, die bereits seit einigen Jahren an der Umsetzung eines Avatars im Fahrzeug forscht und mit der DVD-Anleitung für das Radionavigationssystem RNS 510 für den VW Tiguan den ersten serienmäßigen Einsatz eines Avatars im Fahrzeug erreicht hat (vgl. Heise Autos 2007). Wegen der thematischen Überschneidungen erfolgt dabei ein inhaltlicher Austausch und eine Aufteilung der beforschten Aspekte. Aufgrund der übergeordneten Stellung der Konzernforschung obliegt ihr in gewissem Rahmen die politische Einflussnahme auf das AViCoS-Projekt. Einen Überblick aller identifizierten Stakeholder und ihrer wichtigsten Merkmale innerhalb des Volkswagen Konzerns gibt Tabelle 4-1.

4.2.2. Weitere Stakeholder

Die Verankerung des Forschungsprojekts auf Seiten der TU München erfolgt über den **Lehrstuhl für Wirtschaftsinformatik I17**, der den Zugriff auf verschiedene technische und wissenschaftliche Ressourcen und den Austausch mit anderen Wissenschaftlern ermöglicht. Darüber hinaus lassen sich Studenten durch verschiedene Studienarbeiten in das Projekt integrieren. Wie das INI.TUM-Projekt AViCoS ist auch das Projekt **HIMEPP** im gleichen Umfeld angesiedelt und hat die Erstellung eines modularen Komponenten-Werkzeugkastens für das Rapid Prototyping im Automobilumfeld zum Ziel. Dadurch besteht eine Vielzahl von Anknüpfungspunkten zwischen den beiden Projekten.

Organisation	Untereinheit	Rolle	Beschreibung	Entscheidungs-befugnis	Grad der Beteiligung
TU München	Lehrstuhl für Wirtschaftsinformatik I17	Management	Koordination auf TUM-Seite und Einbettung in universitäres Umfeld	Organisatorische Ausgestaltung auf Seiten der TU München	Hoch
	Projekt HIMEPP	Entwickler	Angebundenes INI.TUM-Projekt	Keine	Hoch
Zulieferer	Nuance, SVOX, BFFT	Entwickler	Bereitstellung und Support von technischen Komponenten	Keine	niedrig
Projektunterstützer	Charamel, Simigon, Loquendo	Entwickler	Bereitstellung und Support von technischen Komponenten	Keine	mittel
Sonstige	Anwender	Anwender bzw. Prüfer / Auditor	Benutzen avatarbasiertes Hilfesystem im Fahrzeug	Entscheidungen über Usability und Ergonomie	niedrig
Dokumente	Ergonomische Richtlinien	Prüfer / Auditor		Unterstützt Akzeptanz und Benutzerfreundlichkeit	mittel
	DIN EN 9241-10, OSGi	Standardisierungsgremium		Akzeptanz durch Interoperabilität	niedrig

Tabelle 4-2: *Weitere Stakeholder und ihre Eigenschaften*
Quelle: (Eigene Darstellung)

Zulieferer der AUDI AG, deren Produkte genutzt werden, stellen neben ihren Produkten Know-how bereit, die zur optimalen Nutzung verwendet werden kann. Dabei ist zu berücksichtigen, dass die Bereitschaft, Produkte, Leistungen oder Wissen kostenfrei bereitzustellen geringer ist als bei Projektunterstützern. Als konkrete Unternehmen dieser

Gruppe sind beispielsweise Nuance und SVOX zu nennen, die Spracherkennungs- und Sprachsynthese-Software herstellen und sich während der Implementierungsphase als geeignete Zulieferer herausgestellt haben (vgl. Kapitel 6.1). BFFT stellt verschiedene Produkte und Dienstleistungen im Bereich der Fahrzeugelektronik her, so dass ein Brettaufbau mit wichtigen Steuergeräten vor der Verwendung eines Testfahrzeugs zum Testen von Fahrzeugfunktionen herangezogen wurde.

Im Gegensatz zu Zulieferern sind **Projektunterstützer** bereit, durch umfangreiche Tätigkeiten in Vorleistungen zu gehen, um in das Zulieferer-Netzwerk der AUDI AG aufgenommen zu werden. So haben sich im Laufe des Forschungsvorhabens die Unternehmen Charamel, Simigon und Loquendo aktiv für eine Zusammenarbeit angeboten. Während Charamel eine Software zur Echtzeit-Darstellung von Avataren herstellt, entwickelt Simigon interaktive Handbücher für Flugzeuge und Autos, die auf PC's in einzelnen Lektionen durchgearbeitet werden können. Loquendo bietet Lösungen zur Spracherkennung und Sprachsynthese an.

Über den gesamten Lebenszyklus eines neuen Systems im Auto hinweg betrachtet, stellt die Gruppe der **Anwender** die wichtigsten Stakeholder dar, zumal sie für die neue Lösung bereit sein müssen, einen entsprechenden Mehrpreis zu entrichten. Da im Rahmen dieses Forschungsprojekts zunächst die Realisierbarkeit und erst in einem zweiten Schritt die Ermittlung grundlegender Reaktionen erfolgen soll, kommt den Anwendern diese bedeutende Rolle nur bedingt zu.

Auch wenn sie keine Personen sind, stellen die Stakeholder in Form von **Dokumenten** einen wichtigen Teil dar. So sollen bei der Konzeption einer neuen Benutzeroberfläche Richtlinien zur Ergonomie berücksichtigt werden (vgl. Williams/Helbig 2007; Shneiderman/Plaisant 2005; Bernsen/Dybkjær/Dybkjær 1998; Kraemer/Landsbergen/Pouteau 1997). Ein weiterer Aspekt, der die Entwicklung und die Interoperabilität der technischen Lösung vereinfacht, ist die Orientierung an bestehenden Standards ISO und DIN-Normen {Comité Européen de Normalisation, 1996 #823) oder Industriestandards (OSGi Alliance 2007a). Eine Übersicht weiterer Stakeholder mit ihren wichtigsten Merkmalen gibt Tabelle 4-2.

4.3. Formulierung der operativen Ziele

Während die Ziele, die aus den Situationen der Stakeholder abgeleitet werden können, auf operativer Ebene angesiedelt sind, sind die Ziele, die implizit durch die Forschungsfragen in Kapitel 1.3 formuliert wurden, auf strategischer Ebene zu sehen. Die operativen Ziele lassen sich im Rahmen der Forschungsfragen auf Grundlage des eigentlichen Forschungsvorhabens mit seinen Stakeholdern ableiten. Die nachfolgende Betrachtung fasst somit Ziele zusammen, die zur Fertigstellung des angestrebten Artefakts in der notwendigen Qualität führen. Dabei werden sowohl die bisher genannten Stakeholder als auch die drei formulierten Forschungsfragen als Begründung der Ziele herangezogen.

1. Ein sprachgesteuertes Avatarsystem soll in Echtzeit multimediale Informationen zur Fahrzeugbedienung bereitstellen können.

Das Primärziel des Forschungsvorhabens ist die Planung und Umsetzung eines Avatarsystems, das per Sprachein- und -ausgabe Informationen zur Bedienung eines Fahrzeugs wiedergeben kann. Die Sprachkommunikation soll auf natürlichsprachlicher Ebene mit prosaischen Ein- und Ausgaben erfolgen können. Dabei sollen die Inhalte neben der visuellen und akustischen Präsentation durch einen Avatar auch durch Abbildungen und Animationssequenzen des erklärten Bereichs ergänzt werden.

Dieses Ziel ist zum einen Voraussetzung für die Untersuchung der dritten Forschungsfrage und zum anderen Grund für das Forschungsvorhaben aus Sicht der beauftragenden Fachabteilung und des Lehrstuhls. Während die Fachabteilung neue Technologien für den Einsatz im Fahrzeug erschließen will, können für den Lehrstuhl durch den Abschluss eines solch innovativen Projektes Folgeprojekte aufgrund des bestehenden Know-hows entstehen. Die Abteilung für Wissenschaftsprojekte verfolgt ebenfalls das Ziel, den Innovationscharakter des Projekts zu betonen, um der AUDI AG neue Impulse geben zu können. Projektunterstützer erhoffen sich durch die Notwendigkeit neuer Technologien eine Geschäftsbeziehung mit der AUDI AG aufbauen zu können. Eine Unterstützung dieses Ziels erfolgt durch die Forschungsfrage 1 und 2, da sie sich mit dem Bau eines solchen Systems beschäftigen, sowie der Fachabteilung für Bordliteratur, da es zu ihren Aufgaben gehört, neue Ansätze zur Verbreitung von Benutzerinformationen zu finden und einzusetzen.

2. Das Avatarsystem muss im Fahrzeug ohne durch den Benutzer wahrnehmbare Zusatzgeräte und ohne Lernaufwand einsetzbar sein.

Die Umsetzung des softwarebasierten Avatarsystems erfordert die Verwendung von zusätzlichen Hardwarekomponenten, die in einem Serienfahrzeug nicht in der notwendigen Form vorhanden sind. Diese zusätzlichen Geräte sollen für den Benutzer nicht wahrnehmbar sein. Er soll lediglich die bereits serienmäßig vorhandenen Bedienelemente zur Auswahl haben. Dies dient der Unterstützung der Forderung nach dem Wegfall eines Lernaufwands. Der Fahrer soll nicht zunächst lernen müssen, wie er das Avatarsystem oder gegebenenfalls weitere Bedienelemente zu bedienen hat, um die multimediale Anleitung verwenden zu können.

Dieses Ziel wird neben der Forschungsfrage 3 durch die Gruppe der Anwender voll unterstützt, da der initiale Lernaufwand neuer Elemente negative Auswirkungen auf den Ablauf der späteren Evaluation haben würde. Für die Anwender hingegen soll der Bedienaufwand in einem Fahrzeug durch das Avatarsystem nicht erhöht, sondern reduziert werden, was auch ergonomischen Richtlinien und Standardisierungsbestrebungen entspricht.

3. Die Ablenkung des Fahrers durch das Avatarsystem muss während der Fahrt minimiert werden, ohne die vermittelten Inhalte in ihrer Qualität einzuschränken.

Auch wenn bereits im zweiten Ziel gefordert wird, dass die Bedienung des neuen Systems einfach sein soll, so kann durch die Ausgabe des Systems dennoch eine Ablenkung des Fahrers erfolgen. Diesem Umstand soll das dritte Ziel vorbeugen, indem die Belastung des Fahrers situationsabhängig auf einem Minimum des Zumutbaren gehalten wird. Dies umfasst

insbesondere die Qualität und den Zeitpunkt einer visuellen Anzeige als auch der akustischen Ausgabe.

Dieses Ziel basiert primär auf den Richtlinien der Ergonomie, die eine Überbelastung des Benutzers vermeiden sollen. Dies ist ganz im Sinne des Fahrers, der durch die Benutzung neuer Systeme weder sich noch sein Eigentum oder andere Verkehrsteilnehmer gefährden will. Darüber hinaus ist dieses Ziel eng mit der dritten Forschungsfrage verknüpft, zumal sie auf der einen Seite die Aufmerksamkeit des Fahrers für die Evaluation benötigt, jedoch auf der anderen Seite keine zusätzliche Belastung oder gar ein Sicherheitsrisiko verursachen soll. Die Betrachtung der Ablenkung durch das Avatarsystem fällt auch in den Aufgabenbereich der Fachabteilung EE-71, da sie für die Integration eines solchen Systems in das bestehende Bedienkonzept verantwortlich wäre, wenn es zu einer Umsetzung in der Serie kommen sollte.

4. Der Aufwand für die Bereitstellung und Pflege der Inhalte des Avatarsystems soll minimiert werden.

Als Bezugspunkt für die Inhalte, die durch Avatarsystem vermittelt werden sollen, dient das bestehende Benutzerhandbuch. Die darin enthaltenen Informationen müssen dazu jedoch in eine Form gebracht werden, die einem Chatbot als Grundlage für Dialoge dienen kann. Darüber hinaus bietet die Präsentation der Inhalte auf einem bewegten, visuellen Medium im Gegensatz zu Printmedium die Möglichkeit, die Inhalte mit zusätzlichen Bildern oder Animationen zu versehen, die im ursprünglichen Handbuch nicht enthalten sind.

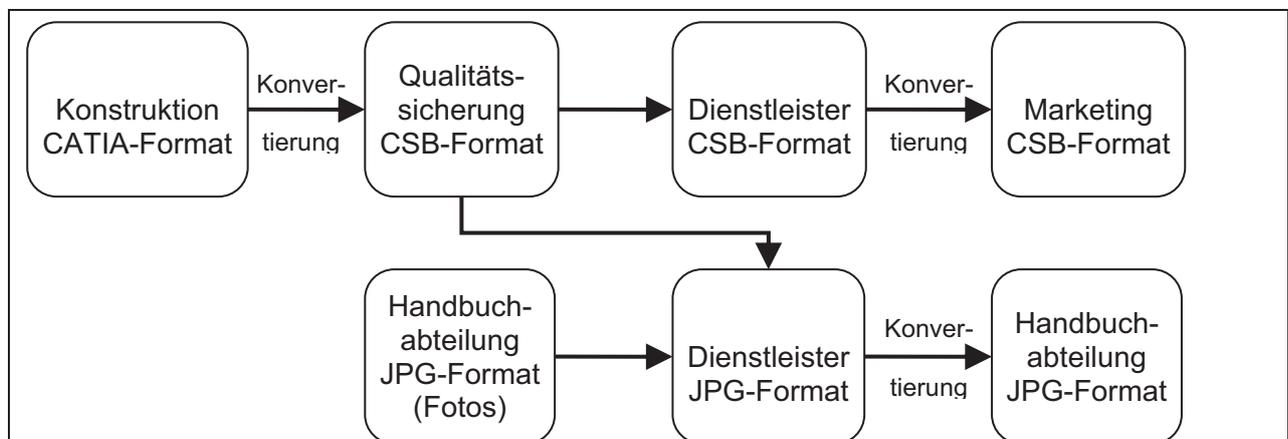


Abbildung 4-3: *Lebenszyklus von visuellen Daten für Handbücher bei der AUDI AG*
Quelle: (Eigene Darstellung)

Neben der Forschungsfrage 2, die sich die Bereitstellung der Inhalte zum Gegenstand macht, unterstützen insbesondere die internen Zulieferer in Form der Fachabteilungen FP-32 und VK-37 dieses Ziel, da es bei einem späteren Serieneinsatz an ihnen liegt, die Inhalte des gesamten Handbuchs bereitzustellen bzw. aufzuarbeiten. Insbesondere die Bereitstellung von Bildmaterial und Animationen stellt auf Basis der bestehenden Wege innerhalb des Unternehmens eine Herausforderung dar. So werden die Konstruktionsdaten zur Qualitätssicherung in eine virtuelle Realität konvertiert und von dort zum einen dem Marketing und zum anderen der Abteilung VK-37 bereitgestellt. Auf dem Weg zum

Marketing erfolgt dabei die Aufarbeitung und die Inszenierung durch einen externen Dienstleister, so dass als Ergebnis fotorealistische Umgebungen für virtuelle Showrooms oder so genannte „Beauty Shots“ entstehen, die ein Fahrzeug in einem Bild besonders gut zur Geltung kommen lassen. Die Integration von Bildmaterial in das Handbuch beginnt entweder mit Daten aus der Qualitätssicherung oder mit Fotos, die von bestimmten Bedienelementen gemacht wurden. Diese beiden Quellen werden von einem externen Dienstleister in ein einheitliches zweidimensionales Aussehen konvertiert und schließlich durch die Abteilung VK-37 in das Handbuch integriert (vgl. Abbildung 4-3).

Neben der Integration von Bildern in das Handbuch sieht sich die Fachabteilung VK-37 auch den Herausforderung verschiedener Sprachen sowie der Verwendung der Inhalte in softwarebasierten Formaten wie einem elektronischen Handbuch für den PC oder einer rein textbasierten On-board Variante des Handbuchs gegenüber. Diese neuen Betätigungsfelder werden dadurch bewältigt, dass die Handbuchinhalte modular und mit entsprechenden Metadaten versehen in einem Content Management System vorgehalten werden, aus dem sich die verschiedenen Formate exportieren lassen. Wenn auch nicht inhaltlich betroffen, so unterstützt die beauftragende Fachabteilung I/FP-34 die Aufwandsminimierung bei der Bereitstellung des Inhalts, da dadurch die Akzeptanz des neuen Systems innerhalb des Unternehmens gesteigert werden kann.

5. Die für das Avatarsystem entwickelten Softwarekomponenten sollen einfach wieder verwendbar sein.

Auch wenn das primäre Ziel der Forschung die Erstellung und Bewertung eines Avatarsystems im Fahrzeug ist, so sollen einzelnen Komponenten unabhängig vom Ergebnis der Evaluation gegebenenfalls in anderen Projekten verwendet werden können. Dies impliziert die Modularisierung des Systems in einzelne Komponenten sowie die Verwendung eines standardisierten Frameworks zur Vereinfachung und Wiederverwendung der einzelnen Module.

Dieses technische Thema wird zum einen von der beauftragenden Fachabteilung FP-37 vorangetrieben, da sie u.a. für die redundanzfreie Entwicklung von Prototypen in diesem Bereich mit verantwortlich ist. Zum anderen sind am modularen Aufbau alle Stakeholder interessiert, die selbst Komponenten bereitstellen oder verwenden wollen. Dazu lassen sich AEV-3, die Projektunterstützer sowie in abgeschwächtem Maße die Zulieferer, die Konzernforschung sowie die Fachabteilung EE-71 zählen. Während die Projektunterstützer durch die Verwendung eines Standards Austauschbarkeit bestehender Komponenten und damit die Chancen ihres Einstiegs bei Audi als Zulieferer erhöhen, ist dies gleichzeitig ein Risiko für bestehende Zulieferer. Bei der Konzernforschung stand zum Zeitpunkt der Anforderungserhebung die Entscheidung über die Verwendung von Standards noch nicht fest. Als weitere wichtige Stakeholder in diesem Zusammenhang ist einerseits die Forschungsfrage 1 zu nennen, die sich auch mit der Modularisierung des Systems beschäftigen muss. Darüber hinaus steht der innerbetriebliche Technologietransfer auch im Fokus der Abteilung für Wissenschaftsprojekte, so dass der Austausch von Erkenntnissen in Form von Komponenten ein wichtiger Punkt ist.

Organisation	Untereinheit	Ziel 1: Avatar- system	Ziel 2: Lern- aufwand	Ziel 3: Ablenkungs- minimierung	Ziel 4: Aufwands- minimierung	Ziel 5: Modularität
Forschungs-fragen	Forschungsfrage 1	○	-	-	-	●
	Forschungsfrage 2	○	-	-	●	-
	Forschungsfrage 3	●	●	●	-	-
AUDI AG	I/FP-34, Prozess- und Systemintegration Elektrik, Elektronik	●	-	-	○	●
	I/FP-32, Virtuelle Realität	-	-	-	●	-
	I/VK-37, Bordliteratur	○	-	-	●	-
	I/EE-71, Entwicklung Bedienkonzept	-	-	○	-	○
	I/SY, Wissenschaftsprojekte	●	-	-	-	●
AUDI Electronic Venture GmbH	I/AEV-3, Elektronik Vorentwicklung	-	-	-	-	●
VW Konzern	K-GEFE/H, Konzernforschung	-	-	-	-	○
TU München	Lehrstuhl für Wirtschaftsinformatik I17	●	-	-	-	-
Zulieferer	Nuance, SVOX, BFFT	-	-	-	-	○
Projektunterstützer	Charamel, Simigon, Loquendo	●	-	-	-	●
Sonstige	Anwender	-	●	●	-	-
Dokumente	Ergonomische Richtlinien	-	●	●	-	-
	DIN EN 9241-10, OSGi	-	○	-	-	●

● = unterstützt Ziel voll ○ = unterstützt Ziel - = steht Ziel neutral gegenüber

Tabelle 4-3: **Unterstützung der Ziele durch die Stakeholder**
Quelle: (Eigene Darstellung)

Ein Überblick der Unterstützung verschiedener Ziele durch die identifizierten Stakeholder ist in Tabelle 4-3 zu sehen. Dabei ist eine Gruppierung der Ziele in die drei Bereiche Gestaltung der Benutzerschnittstelle, Inhalte und technische Umsetzung entlang der drei Forschungsfragen zu beobachten. Die Ziele 1 bis 3 entsprechen dabei der Forschungsfrage 3 und den Stakeholdern, die mit der Außenwirkung des Systems befasst sind. Ziel 4 entspricht der Forschungsfrage 2 und den inhaltlich relevanten Stakeholdern, während Ziel 5 die Forschungsfrage 1 abdeckt und die entwickelnden Stakeholder berücksichtigt.

4.4. Ableitung der Anforderungen

Auf der Grundlage der zusammengefassten operativen Ziele werden im Folgenden konkrete Anforderungen an den Prototyp abgeleitet. Als Erhebungsmethode dienten überwiegend Gespräche mit Vertretern der einzelnen Stakeholdergruppen, aus denen die einzelnen Anforderungen explizit oder implizit hervorgingen und nachträglich durch die genannten Ziele kategorisiert wurden. Die Integration von Dokumenten erfolgte über deren Studium. Entsprechend der Bedeutung einer Anforderung zur Erreichung des Ziels bzw. aus Sicht des Stakeholders, lassen sich Muss-, Soll- und Kann-Anforderungen unterscheiden. Zur besseren Umsetzungskontrolle der Anforderungen sind diese jeweils für jedes Ziel fortlaufend nummeriert.

4.4.1. Anforderungen zur Umsetzung eines Avatarsystems (erstes Ziel)

Das grundlegende operative Ziel, ein Avatarsystem umzusetzen, ist durch folgende vier Eckpunkte gekennzeichnet: Natürlichsprachliche Sprachein- und -ausgabe, die Anzeige eines Avatars und dessen Umwelt. Bereits diese Eckpunkte stellen gleichzeitig die ersten Anforderungen dar, welche zur Operationalisierung weiter detailliert werden. Die Anforderungen zum ersten Ziel umfassen lediglich die grundlegenden Punkte, die notwendig sind, um ein rudimentäres Avatarsystem für den Einsatz im Fahrzeug umzusetzen. Weitere Aspekte, die durch andere Ziele in den Vordergrund gestellt werden, sind bei den jeweiligen Zielen formuliert.

Zunächst leiten sich Anforderungen an das Verständnis von Natürlichsprachlichkeit, der Landessprache sowie der technischen Qualität der Spracherkennung ab. Die Antwort des Avatarsystems soll ebenfalls mit einer bestimmten Qualität als vollständiger Satz und angemessen als Sprache synthetisiert werden. Die Qualität der Sprachsynthese kann dabei zum einen absolut anhand realer Stimmen und zum anderen relativ durch den Vergleich mit bestehenden Synthesystemen bestimmt werden. Da keine Sprachsynthese derzeit an eine reale Stimme heranreicht (vgl. dazu: Beutnagel et al. 1999; Nass/Lee 2001), bleibt nur die relative Bestimmung der Qualität. Neben der Verständlichkeit und Realitätsnähe muss die Synthese auch in Abstimmung mit der Avatardarstellung erfolgen. Dies beinhaltet beispielsweise Kriterien wie Geschlecht, Alter, Sprechgeschwindigkeit und Tonlage (vgl. dazu Nass/Steuer/Tauber 1994; Heidmann/Biesterfeldt 2004, 63). Der Bedarf einer stimmigen Sprachsynthese ist durch die Anzeige des Avatars sowie seiner virtuellen Umwelt bedingt. Diese beiden Elemente müssen für eine maximale Flexibilität bei der Dialoggestaltung in Echtzeit animierbar sein. Die Abstimmung zwischen dem Avatar und seiner Stimme muss sich auch in der Beziehung zwischen Avatar und Umwelt widerspiegeln (vgl. Theune 2001). Obwohl die Darstellung der Avataramwelt in dreidimensionaler Form zu bevorzugen ist, kann die Notwendigkeit aufkommen, bereits bestehendes zweidimensionales Bildmaterial in Form von Einzelbildern oder Bildsequenzen in Dialoge einzubinden.

Neben diesen zwingenden Muss-Anforderungen lassen sich weitere Soll- bzw. Kann-Anforderungen ableiten. So soll die Sprachsynthese neben einer passenden Stimme auch eine menschliche Sprechmelodie und Betonung aufweisen, um einen realistischeren Eindruck des Avatars zu bewirken. Dies beinhaltet die korrekte Aussprache komplexer Ausdrücke und von

Fremdwörtern, die im Automobilumfeld gebräuchlich sind. Als Kann-Anforderung lässt sich die Mehrsprachigkeit des Systems nennen, da dies bei einem späteren Serieneinsatz zu einer Muss-Anforderung werden kann und somit bereits beim Prototypenbau in der Konzeption berücksichtigt werden sollte. Einen Überblick über die Anforderungen, die sich direkt aus dem ersten operativen Ziel ableiten lassen, gibt Tabelle 4-4.

Nr.	Bedeutung	Anforderung
1.1	Muss	Verstehen natürlichsprachlicher Spracheingaben auf Deutsch
1.2		Spracherkennungsrate für ausgewählte Szenarien von 95%
1.3		Sicherstellung einer angemessenen Sprachsynthesequalität
1.4		Antworten des Chatbots in vollständigen Phrasen
1.5		Angemessene Stimme zum Avatar
1.6		Darstellung eines Avatars und seiner virtuellen Umwelt
1.7		Echtzeit-Animation des Avatars und seiner Umwelt
1.8		Abstimmung/Synchronisation des Avatars mit virtueller Umwelt
1.9		Integration von zweidimensionalen Abbildungen und Animationen
1.10	Soll	Korrekte Aussprache von gebräuchlichen Fremdwörtern
1.11		Verwendung einer menschlichen Sprachmelodie
1.12	Kann	Mehrsprachigkeit des Systems

Tabelle 4-4: *Abgeleitete Anforderungen aus dem ersten operativen Ziel*
Quelle: (Eigene Darstellung)

4.4.2. Anforderungen zur Benutzung ohne Lernaufwand (zweites Ziel)

Das zweite Ziel, die Verwendung des Systems ohne das Erlernen neuer Bedienmuster oder Geräte, beinhaltet aus der Zielformulierung heraus bereits die beiden Anforderungen, keine zusätzlichen, wahrnehmbaren Geräte zu verwenden und das System ohne das Erlernen von Kommandos benutzen zu können.

Ein wichtiger Punkt bei der Verwendung neuer Geräte ist die Wahrnehmung des Anwenders. Für die Umsetzung neuer Funktionen im Fahrzeug, insbesondere bei einer prototypischen Implementierung, sind weitere Geräte notwendig. Diese sind jedoch derart einzubauen, dass sie der Benutzer nicht sieht. Dementsprechend muss beispielsweise zur Anzeige des Avatars und seiner Umwelt das bestehende Display des Audi Multimedia Interfaces (MMI) genutzt werden, das als Ausstattungsmerkmal bereits serienmäßig verfügbar ist. Dieser Sachverhalt widerspricht zwar der Forderung Williams und Helbig (2007, 287) nach einem möglichst hochauflösenden Display und dessen optimaler Platzierung im Fahrzeug, erlaubt jedoch die Verwendung eines äußerlich unveränderten Fahrzeugs. Diese Anforderung wird durch Untersuchungen von Färber (1987) bestätigt, in denen er in einem Fahrsimulator mit Testpersonen Versuche hinsichtlich der Wirksamkeit von Warnmeldungen durchführte. Durch die ungewohnte Umgebung des Fahrsimulators verschlechterten sich die Reaktionszeiten und das Reaktionsverhalten der Testpersonen.

Die Verwendung des Systems ohne Training resultiert zum einen in der Forderung nach einer sprecherunabhängigen Spracherkennung, die ohne vorheriges Training der Software korrekte Ergebnisse liefert. Zum anderen muss dem Sprachgebrauch und dem Wortschatz verschiedener Benutzer auch im Chatbot Rechnung getragen werden, indem für die Ausführung einer Aktion durch das Avatarsystem verschiedene Formulierungen des Benutzers verstanden werden.

Um dem Benutzer trotz fehlender Einweisung in das Avatarsystem alle Möglichkeiten des Systems nutzen lassen zu können, muss eine überwiegend aktive Dialogführung durch den Chatbot erfolgen. Dabei wird an geeigneten Stellen auf weiterführende, logisch sequenzielle Inhalte verwiesen. Diese Dialogführung soll jedoch lediglich ein Angebot sein und nicht einen zwingenden Gesprächsablauf bezeichnen. Gemäß Shneiderman und Plaisant (2005, 75) muss der Benutzer einer Schnittstelle jederzeit den Eindruck haben, die Kontrolle zu haben (vgl. auch Williams/Helbig 2007, 301). Dieser Maxime folgend muss der Anwender die Möglichkeit haben, das Gesprächsthema und damit den Kontext zu wechseln. Um den Zugriff auf die Inhalte des interaktiven Handbuchs einfach zu erlangen, müssen Bedienelemente, deren Funktion erklärt werden soll, zum einen verbal beschrieben und zum anderen betätigt werden können. Der Aspekt der Betätigung ist dabei jedoch nur bei Funktionen des tertiären Bereichs sinnvoll, um die Sicherheit während der Fahrt nicht zu gefährden (vgl. dazu Geiser 1985). Für kritische Bedienelemente ist der Hinweis auf verfügbare Informationen nach einer planmäßigen Benutzung durch den Fahrer denkbar.

Ein bedeutender Punkt, der unerfahrenen Anwendern die Benutzung erlaubt, ist die Robustheit der Dialoge. Dies impliziert, dass das System fehlerhafte Eingaben erkennt, darauf hinweist und Möglichkeiten zur Korrektur bzw. Wiederaufnahme des Dialogs enthält (vgl. dazu Comité Européen de Normalisation 1996; Shneiderman/Plaisant 2005, 75; Williams/Helbig 2007, 300).

Die genannten Muss-Anforderungen lassen sich durch Soll-Vorgaben sinnvoll erweitern. So sollen im Fehlerfall durch das System Hinweise gegeben werden, wie fortgefahren werden kann, um den Dialog fortzuführen (Krahmer/Landsbergen/Pouteau 1997, 83). Eine Anforderung, um diesen Fehlerfall zu vermeiden, ist die automatische Korrektur von fehlerhaften Spracheingaben, was zum Teil bereits automatisch durch die Spracherkennung erfolgt oder gezielt in einem Chatbot konfiguriert werden kann (vgl. Dowding et al. 1993; Heeman/Allen 1997; Heidmann/Biesterfeldt 2004, 63). Um Fälle analysieren zu können, in denen trotz dieser Maßnahmen Fehler bei der Erkennung und damit im Dialogfluss aufgetreten sind, soll ein Logging-Mechanismus integriert werden. Die Analyse der Log-Dateien ist dabei nur die Vorarbeit zur späteren Erweiterung des Avatarsystems, um bekannte Fehler in Zukunft auffangen oder vermeiden zu können (vgl. dazu auch Williams/Helbig 2007, 295). Da Anwender nicht immer visuell erkennen können, ob ein erfolgter Sprachbefehl ausgeführt wird, muss die Mensch-Maschine-Kommunikation durch Feedback auf eine Eingabe ergänzt werden (Shneiderman/Plaisant 2005, 74; Comité Européen de Normalisation 1996). Dieses Feedback soll nach Williams und Helbig (2007, 301) maximal 250 Millisekunden nach der Eingabe des Benutzers gegeben werden. Ebenso macht die Möglichkeit, eine Aktion ungeschehen zu machen, den Anwender sicherer im Umgang mit dem System (Shneiderman/Plaisant 2005, 75). Die Unterstützung eines unerfahrenen

Benutzers kann auch durch die konsistente Verwendung von Farben, Symbolen und Begriffen (Krahmer/Landsbergen/Pouteau 1997, 83; Williams/Helbig 2007, 293), der Anlehnung an bekannte Farbmuster (Williams/Helbig 2007, 290), und durch das gezielte Angebot einer Einführung in die Bedienung des Systems gegeben werden (Comité Européen de Normalisation 1996).

Nr.	Bedeutung	Anforderung
2.1	Muss	Keine baulichen, sichtbaren Erweiterungen
2.2		Gestaltung des Avatars entsprechend dem bestehenden Display des Audi Multimedia Interfaces (MMI)
2.3		Benutzung ohne Training
2.4		Sprecherunabhängigkeit der Spracherkennung
2.5		Verstehen verschiedener Formulierungen und verschiedener Wortschätze der Anwender
2.6		Aktive, gerichtete Gesprächsführung durch Avatarsystem
2.7		Benutzer soll den Eindruck haben, das System zu kontrollieren
2.8		Themenwechsel durch Benutzer jederzeit möglich
2.9		Ermittlung zu erklärender Bedienelemente durch Sprache oder deren Bedienung
2.10		Robustheit der Dialoge bei falscher Bedienung
2.11	Soll	Bei fehlerhafter Spracherkennung soll Anleitung zum Fortfahren gegeben werden
2.12		Automatische Korrektur bei fehlerhafter Spracheingabe
2.13		Logging von Gesprächsverläufen, insbesondere nicht verstandene Spracheingaben, zur Verbesserung der Wissensbasis
2.14		Nachträgliche Erweiterung des Avatarsystems
2.15		Feedback auf jede Aktion
2.16		Aktionen sollen reversibel sein
2.17		Konsistente Farben, Symbole und Bezeichnungen
2.18		Unterstützung und Anleitung beim Erlernen des Dialogsystems
2.19	Kann	Aktionssequenzen sollen aus Anfang, Mitte und Ende bestehen

Tabelle 4-5: *Abgeleitete Anforderungen aus dem zweiten operativen Ziel*
Quelle: (Eigene Darstellung)

Als eine Kann-Anforderung dieses Ziels soll die Einteilung von Aktionssequenzen in die drei Teile Anfang, Mitte und Ende genannt werden (Shneiderman/Plaisant 2005, 75). Während diese Forderung durchaus angebracht ist, kann an dieser Stelle noch nicht abgeschätzt werden, ob dies während der Kommunikation in einem Fahrzeug sinnvoll ist oder eher störend wirkt. Die genannten Anforderungen des zweiten operativen Ziels sind in Tabelle 4-5 aufgeführt.

4.4.3. Anforderungen zur Minimierung der Ablenkung (drittes Ziel)

Die Minimierung der Ablenkung durch das Avatarsystem ist insbesondere aus Gründen der Sicherheit sowie der Usability während der Fahrt bedeutsam (vgl. dazu auch Williams/Helbig

2007, 284ff). Um das Kurzzeitgedächtnis des Anwenders nicht mit zu vielen Informationen zu belasten, müssen die Antworten des Avatarsystems so kurz sein, dass sie während der Fahrt wahrgenommen, verarbeitet und darauf gegebenenfalls reagiert werden kann (Shneiderman/Plaisant 2005, 75; vgl. dazu auch Miller 1956; Labiale 1992). Diese kurzen Antworten sollen die notwendigen Informationen einer Aussage oder Frage des Avatarsystems enthalten, um einen möglichst effizienten Dialog zu realisieren und für das Erreichen eines Kommunikationszieles überflüssige Interaktion zu vermeiden (Williams/Helbig 2007, 286f; Heidmann/Biesterfeldt 2004, 63).

Damit der Benutzer einfach nachvollziehen kann, ob eine Aktion von ihm durch das Avatarsystem erwartet wird oder nicht, muss bei jeder Antwort des Systems erkennbar sein, ob es sich um eine Frage oder eine Aussage bzw. Anleitung handelt (Krahmer/Landsbergen/Pouteau 1997, 83). Die Transparenz und das Vermeiden von Überraschungseffekten helfen, die mentale Belastung des Fahrers durch das Avatarsystem zu reduzieren (Burnett/Joyner 1997). Da das Avatarsystem nicht die Eigenschaft haben soll, die Belastung des Fahrers im Rahmen seiner Fahrzeugführungsaufgabe zu ermitteln, muss es dem Benutzer jederzeit selbst möglich sein, die Antwort des Avatarsystems und somit den Dialog zu steuern und gegebenenfalls in einer gefährlichen Situation zu unterbrechen (Williams/Helbig 2007, 289f). Diese Steuerung wird dabei nicht nur kurzfristig ausgeübt, sondern kann auch generell verwendet werden, indem beispielsweise bestimmte Anzeigeelemente wie der Avatar oder seine virtuelle Umwelt ganz oder teilweise durch den Anwender abgeschaltet werden können (vgl. dazu Williams/Helbig 2007, 302).

Nr.	Bedeutung	Anforderung
3.1	Muss	Kurze Antworten des Avatarsystems/Vermeidung größerer Belastungen für KZG
3.2		Vermeidung unnötiger oder überflüssiger Interaktion
3.3		Antworten müssen erkennen lassen, ob sie Fragen oder Anleitungen sind
3.4		Unterbrechung der Antworten des Avatars jederzeit möglich/ Steuerung der Wiedergabe des Avatarsystems durch Benutzer
3.5		Selektive Abschaltung von grafischen Anzeigeelementen
3.6	Soll	Speicherung und Pflege von Benutzeroptionen
3.7		Vermeidung von Wiederholung bereits vorgetragener Erklärungen
3.8		Rücksichtnahme auf Eigenschaften des Benutzers (z.B. Alter oder Kenntnisse aus anderen Domänen)
3.9	Kann	Shortcuts für bestimmte Kommandos und Inhalte

Tabelle 4-6: *Abgeleitete Anforderungen aus dem dritten operativen Ziel*
Quelle: (Eigene Darstellung)

Über diese Muss-Kriterien hinaus, sollten benutzerbezogene Optionen speicherbar sein, welche für die individuelle Anpassung der Dialoge oder der Anzeige genutzt werden können (Krahmer/Landsbergen/Pouteau 1997, 84; Williams/Helbig 2007, 304f). Diese Speicherung ermöglicht die Vermeidung von Hinweisen oder optionalen Erklärungen, die bereits vorgetragen wurden und kann bei weiteren Aufrufen der Informationen in einer

zielführenderen Interaktion resultieren. Diese Individualisierung der Dialoge soll abhängig von den Merkmalen eines Fahrers wie z.B. Alter oder Kenntnissen aus anderen Domänen wie dem Umgang mit Computern bereits automatisch angepasst sein (Williams/Helbig 2007, 283; Comité Européen de Normalisation 1996).

Als eine Kann-Anforderung dieses Ziels ist die Verwendung von Shortcuts zu nennen, welche den Zugriff auf bestimmte Informationen beschleunigen sollen (Shneiderman/Plaisant 2005, 75). Ein Überblick der genannten Anforderungen zum dritten Ziel gibt Tabelle 4-6 zu finden.

4.4.4. Anforderungen zur Minimierung des Erstellungsaufwands (viertes Ziel)

Der wichtigste Eckpunkt des vierten operativen Ziels ist die Minimierung des Erstellungsaufwands für die Inhalte des Avatarsystems. Dies ist insbesondere hinsichtlich einer möglichen Seriennutzung von Bedeutung. Bei den Anforderungen, die sich aus diesem Ziel ergeben, muss jedoch berücksichtigt werden, dass einige grundlegende Anforderungen an die Art und Qualität der Inhalt einschränkend auf die Zielerreichung wirken. Diese einschränkenden Faktoren werden ebenso berücksichtigt wie die direkt aus dem Ziel abgeleiteten Ansprüche.

Zur Vereinfachung des Erstellungsprozesses sollen die bereits bestehenden Inhalte aus den Handbüchern verwendet werden können. Das gleiche trifft auch auf die bestehenden dreidimensionalen Daten des Fahrzeugs, insbesondere der Bedienelemente, zu. Insgesamt muss der Prozess der Dialogerstellung auf der Basis des genannten Materials intuitiv ablaufen können, so dass keine besonderen technischen Kenntnisse oder eine Programmierung nötig sind. Dennoch muss der dabei erstellte Inhalt einigen Anforderungen genügen. So müssen Antworten des Systems variieren können, was insbesondere bei häufig gegebenen Antworten zu realistischeren Dialogen führt. Darüber hinaus müssen der aktuelle Dialogkontext und die Gesprächshistorie beim Dialogdesign berücksichtigt bzw. genutzt werden können. Die Forderung nach der Bestimmung des zu erklärenden Bedienelements durch dessen Bedienung, die bereits in den Anforderungen zum ersten Ziel vorkam, erfordert den Zugriff auf die technischen Bussysteme des Fahrzeugs. Dies ermöglicht die Integration von Ereignissen dieser Busse als Auslöser für bestimmte Inhalte. Eng mit dieser Anforderung verbunden, muss der Zugriff auf Stammdaten des Fahrzeugs wie z.B. Modellnummer und Ausstattungsmerkmale möglich sein. Dadurch können relevante Inhalte für den Benutzer bestimmt werden.

Über diese Muss-Anforderungen hinausgehend soll die Integration der bestehenden Inhalte aus dem Handbuch und der Konstruktion (teil-)automatisch zumindest in Grundstrukturen für Dialoge konvertiert werden können. Diese Anforderung ist trotz ihrer Bedeutung nur als Soll-Anforderung formuliert, da bisher keine Berichte über die erfolgreiche Umsetzung solch eines Vorgehens vorliegen und somit keine Aussagen über die Machbarkeit getroffen werden können. Da zu erwarten ist, dass die Verwendung der bestehenden Inhalte in Dialogen auch Wechselwirkungen auf die bestehenden Prozesse zu deren Erstellung und Verwaltung haben wird, soll eine enge Abstimmung mit den bestehenden Prozessen im Unternehmen erfolgen. Dies soll die Grundlage für eine spätere Integration in die Serienumsetzung legen. Die Inhalte des Avatarsystems werden auch die Steuerung der Darstellungsebene übernehmen können.

Da die Einbindung der passenden Animationen in die Dialoge einen zusätzlichen Aufwand erzeugt, die Möglichkeit der Animation jedoch möglichst optimal ausgenutzt werden soll, sollen zumindest Teile der Animationen des Avatars automatisch erfolgen können. Dabei können gewohnte Bewegungen beim Sprechvorgang generell oder auf Basis bestimmter Inhalte eingefügt werden. Die Generierung automatischer Animationen soll dabei nur stattfinden, wenn nicht explizit andere Animationen angegeben wurden. Eine Übersicht aller Anforderungen des vierten operativen Ziels ist in Tabelle 4-7 zu finden.

Nr.	Bedeutung	Anforderung
4.1	Muss	Verwendung bestehender Inhalte aus Handbüchern
4.2		Verwendung bestehenden dreidimensionalen, grafischen Materials
4.3		Intuitive Dialogerstellung
4.4		Variation von Antworten
4.5		Wissen über aktuellen Kontext und Gesprächsverlauf in Avatarsystem
4.6		Verbindung zu bestehenden Bordsysteme zur Gewährleistung von Interaktivität
4.7		Wissen über das eigene System, bzw. das eigene Fahrzeug
4.8	Soll	(Teil-)automatische Konvertierung und Integration bestehender Inhalte in Dialoge
4.9		Integration des Dialogerstellungsprozesses in bestehende Prozesse für Handbuch und 3D-Daten
4.10		Automatische Animationen von Kopf, Händen und Armen des Avatars

Tabelle 4-7: *Abgeleitete Anforderungen aus dem vierten operativen Ziel*
Quelle: (Eigene Darstellung)

4.4.5. Anforderungen zur Modularität der Softwarekomponenten (fünftes Ziel)

Das fünfte Ziel der Modularität fasst die Anforderungen an die technische Umsetzung zusammen. So muss diese auf Basis einer erweiterbaren, modularen Architektur implementiert werden, damit spätere Erweiterungen vorgenommen werden bzw. Komponenten in anderen Projekten wiederverwendet werden können. Dieser Austausch setzt die Verwendung eines standardisierten technischen Frameworks voraus, um eine optimale Interoperabilität zu gewährleisten. Die ausführliche Dokumentation der einzelnen Module ist dabei ebenfalls unabdingbar.

Nr.	Bedeutung	Anforderung
5.1	Muss	Umsetzung einer erweiterbaren, modularen Architektur
5.2		Verwendung eines standardisierten Entwicklungs-Frameworks
5.3		Dokumentation von Komponenten zur einfachen Wiederverwendung
5.4	Soll	Technische Konzeption darf die spätere Serienumsetzung nicht ausschließen

Tabelle 4-8: *Abgeleitete Anforderungen aus dem fünften operativen Ziel*
Quelle: (Eigene Darstellung)

Die Wahl der Architektur bzw. des Frameworks sollen dabei eine spätere Serienumsetzung nicht verhindern, so dass bereits bei der Erstellung des Prototyps eine Abstimmung mit der Serienentwicklung stattfinden soll. Die Anforderungen des letzten operativen Ziels sind in Tabelle 4-8 zusammengefasst.

4.5. Zusammenfassung

Die Betrachtung des Umfelds des Forschungsvorhabens, in dem der Prototyp entwickelt wird dient dem Abstecken der Rahmenbedingungen, innerhalb derer die Implementierung stattfindet. Als Ergebnis dieses Kapitels können die konkreten Anforderungen an die Umsetzung genannt werden. Im Rahmen der Umfeldanalyse wurden insbesondere der Volkswagen Konzern, vertreten durch die AUDI AG, sowie die Technische Universität München, vertreten durch den Lehrstuhl für Wirtschaftsinformatik I17, identifiziert. Das wissenschaftliche Projekt „Avatar based Virtual Co-driver System“ (AViCoS) verfolgt dabei die Umsetzung und Untersuchung des Avatarsystems. Das Projekt findet im Rahmen des Ingolstadt Institut der TU München (INI.TUM) statt, einer Kooperation, die zwischen den beiden Einheiten den Wissenstransfer fördert.

Das klar abgesteckte Umfeld des Projektes AViCoS erlaubt die systematische Suche nach Anspruchsgruppen und ihren Motivationslagen. Dabei lassen sich nicht nur Stakeholder innerhalb, sondern auch außerhalb der beiden genannten Organisationen mit ihren betroffenen Abteilungen finden. So spielt auch das Projekt „HIMEPP“ im Rahmen von INI.TUM eine Rolle als Austauschpartner. Ebenso lassen sich Verbindungen zu bestehenden Zulieferern der AUDI AG bzw. weiteren Unternehmen identifizieren, welche sich durch die Zusammenarbeit in einem wissenschaftlichen Projekt den Zugang als Lieferant erhoffen.

Die verschiedenen Stakeholdern mit ihren Zielen sowie die im Rahmen dieses Dissertationsvorhabens zu klärenden Forschungsfragen dienen als Quelle für die Formulierung operativer Ziele. So konnten fünf Ziele ausgemacht werden, die über die drei Ebenen der Benutzerschnittstelle, der Inhalte und der Technik verteilt sind. Die ersten drei Ziele postulieren die Forderungen nach einem interaktiven, sprachgesteuerten Avatarsystem, das ohne Lernaufwand dem Fahrer die Benutzung einzelner Bedienelemente im Fahrzeug erklären kann. Dabei soll zudem die Ablenkung des Fahrers gering gehalten werden. Das vierte Ziel besteht in der Minimierung des Erstellungsaufwands für die Inhalte des Avatarsystems. Das letzte Ziel umfasst schließlich die Modularität und Wiederverwendbarkeit der technischen Komponenten des Avatarsystems.

In einem letzten entscheidenden Schritt wurden zu den genannten Zielen konkrete Anforderungen aus Interviews und dem Dokumentenstudium bestimmter Richtlinien und Normen ermittelt. Diese Anforderungen wurden für jedes Ziel gesondert aufgestellt und nach ihrer Bedeutung in Muss-, Soll- und Kann-Anforderungen unterteilt.

5. Konzeption

„Gebraucht die Zeit, sie geht so schnell von hinnen, doch Ordnung lehrt euch Zeit gewinnen.“

Johann Wolfgang von Goethe (1749–1832), deutscher Dichter

Die Konzeption stellt die von der konkreten Implementierung abstrahierte Planung eines Vorgehens dar. In diesem Fall werden im Rahmen der Konzeption auf den drei relevanten Ebenen der Technik, des Inhalts und der Benutzerschnittstelle Überlegungen zur konkreten Umsetzung angestellt (vgl. Abbildung 5-1). Diese Vorstufe soll auf einem abstrakten Niveau wichtige Aspekte zusammenfassen und somit die spätere Implementierung vereinfachen.

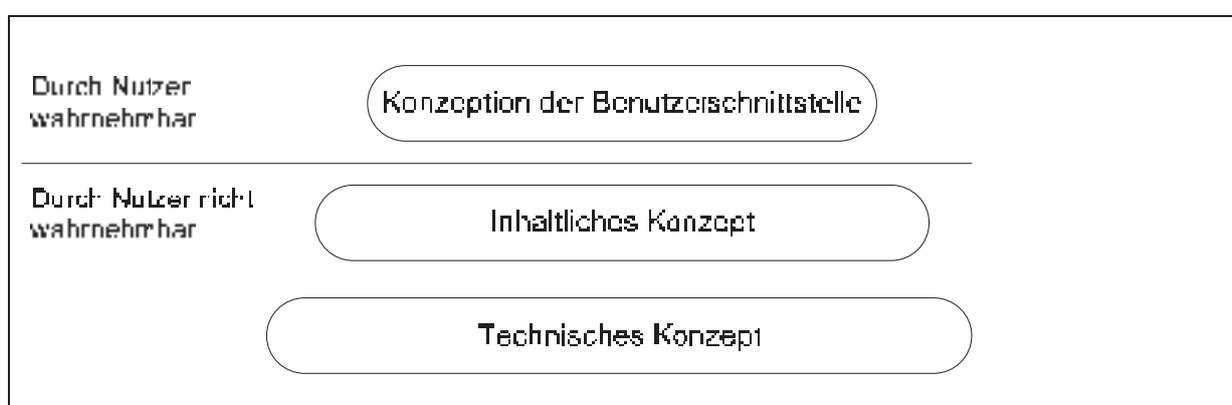


Abbildung 5-1: Zusammenhang der einzelnen Konzeptionsteile
Quelle: (Eigene Darstellung)

Die technische Ebene repräsentiert die Softwarekomponenten, die für die Umsetzung des Avatarsystems notwendig sind und gemäß dem fünften Ziel der Anforderungen in wiederverwendbaren Modulen vorliegen soll. Im Rahmen der Planung der Komponenten soll nicht nur eine Architektur für diesen Prototyp entstehen. Vielmehr soll auch auf Basis von Erfahrungen aus anderen Implementierungen eine Referenzarchitektur für natürlichsprachliche Avatarsysteme zusammengestellt werden, die zugleich ihre Anwendung im Rahmen dieser Forschungsarbeit erfährt. Damit wird gleichzeitig eine Antwort auf die erste Forschungsfrage nach einer geeigneten Architektur und domänenspezifischen Besonderheiten beantwortet.

Ein bedeutender Teil des Avatarsystems ist neben der technischen Plattform der Inhalt, der das System zum Leben erweckt. Deshalb soll der Aufbau des Inhalts betrachtet werden und Funktionen ermittelt werden, welche für die korrekte Verarbeitung und die Bereitstellung menschlich wirkender Dialoge notwendig sind. Dabei werden ebenso Mechanismen zur Verwaltung von Daten als auch zur Fehlervermeidung beleuchtet. Anschließend wird ein grober Rahmen für die möglichen Inhalte abgesteckt, um die Abhängigkeiten verschiedener Themen zu verdeutlichen. Zur Erstellung der Inhalte in Form einer Regelbasis bestehen verschiedene Vorgehensmodelle, die dargestellt werden. Einhergehend mit dem Vorgehen lassen sich auch verschiedene Paradigmen bzw. konkrete Ausprägungen von Werkzeugen zur Inhaltserzeugung und –verwaltung unterscheiden.

Den Abschluss des Kapitels bildet die Konzeption der Benutzeroberfläche, die sich zum einen mit der Gestaltung des Avatars auf Grundlage psychologischer und sozialer Erkenntnisse beschäftigt. Da der Avatar nicht alleiniger Bestandteil der visuellen Komponente ist, wird schließlich ein Rahmenwerk für die Integration des Avatars sowie weiterer Anzeigeelemente zusammengestellt. Dabei werden ebenso Besonderheiten eines Sprachdialogsystems als auch des Umfeldes im Fahrzeug berücksichtigt.

5.1. Technische Konzeption

Die technische Ebene des Prototyps umfasst überwiegend Softwarekomponente, deren Zusammenspiel in Rahmen einer Softwarearchitektur geregelt werden muss. Somit ist das Ziel dieses Teilkapitels die Ermittlung einer geeigneten Architektur, um die Module bzw. Komponenten des Avatarsystems identifizieren und beschreiben zu können. Auf Basis bereits abgeschlossener und dokumentierter Avatar- und Sprachdialogsysteme sollen zunächst bisher verwendete Komponenten gefunden werden. Dadurch kann einerseits eine Vereinheitlichung im Sprachgebrauch für die Bezeichnungen einzelnen Komponenten und andererseits eine breit angelegte Sammlung aller relevanten Komponenten erstellt werden. Darauf aufbauend wird im zweiten Schritt eine Referenzarchitektur aufgebaut, welche alle Komponenten berücksichtigt und sinnvoll miteinander in Verbindung setzt. Dieser domänenunabhängige Plan wird schließlich auf die konkreten Anforderungen dieses Forschungsvorhabens angewandt und somit einer ersten Anwendung unterzogen.

5.1.1. Analyse bestehender Avatar- und Sprachdialogsysteme

Da Avatarsysteme in Kombination mit natürlichsprachlicher Sprachein- und -ausgabe nur in wenigen Fällen ausreichend untersucht und beschreiben wurden, bietet es sich zur Sammlung relevanter Komponenten an, auch Avatarsysteme ohne akustische Sprachsignale sowie Sprachdialogsysteme ohne Avatare zu betrachten. Dabei spielen zunächst weder die Unterscheidung nach der Domäne noch nach dem Einsatzbereich wie internetbasierte oder stationäre Anwendungen eine Rolle. Die Sammlung relevanter Komponenten soll dabei Schritt für Schritt aufgebaut, verbreitert und schließlich inhaltlich differenziert werden, zumal gleiche Ausdrücke für verschiedene Funktionen verwendet werden. Diese Dekomposition bestehender Systeme dient der Identifizierung relevanter Einzelteile. Ein Vergleich der gesamten Architekturen ist aufgrund der Verschiedenheit der Entwicklungsparadigmen, des Fokus ihrer Beschreibungen sowie der verschiedenen Gestaltungsziele nicht sinnvoll möglich. Auf der Basis dieses induktiven Vorgehens erfolgt die Zusammenstellung einer Referenzarchitektur aus den identifizierten Komponenten. Diese wird schließlich deduktiv auf einen konkreten Fall, dem Forschungsvorhaben dieser Dissertation, in Kapitel 5.1.3 angewandt.

5.1.1.1. Untersuchte Systeme und Architekturen

Bei der Suche nach Komponenten werden Sprachdialogsysteme und Avatarsysteme betrachtet, wobei neben konkreten Anwendungen auch konzeptionelle Ansätze betrachtet werden. Die Auswahl der Beispiele erfolgt nach dem Umfang und der Tiefe der verfügbaren Dokumentation, wobei nur die Darstellung der technischen Architektur berücksichtigt wird. Die Untersuchung stellt nicht den Anspruch, eine vollständige Aufführung veröffentlichter

Architekturen aufzuweisen; vielmehr ist die Ermittlung aller relevanten Komponenten von Bedeutung. Die Domäne des Systems spielt bei der Findung von Einzelkomponenten keine Rolle, wird jedoch später bei der konkreten Anwendung auf die Architektur des Prototyps berücksichtigt. Sofern eine verarbeitende, logische Komponente in einem System oder einer Architektur vorkommt, wird zunächst nicht unterschieden, ob es sich um einen state-, frame- oder agenten-basierten Ansatz handelt (vgl. dazu McTear 2002, 92ff).

Sprachdialogsysteme

Das erste betrachtete Sprachdialogsystem stellt die Anwendung BusCat der British Telecom dar, welches die Navigation und Auswahl von Produkten auf einer Website für Telefone und Zubehör ermöglicht. Aus technischer Sicht werden wenige Komponenten unterschieden (Wyrd et al. 1996), weshalb sie als provisorische Grundlage der weiteren Untersuchung dient. Die drei Systeme Circuit-Fix-It Shop, SUNDIAL und TRAINS zeichnen sich dadurch aus, dass sie von McTear (2002) als Beispiele dreier Archetypen von Verarbeitungsmöglichkeiten benannt werden: Anwendungsdatenbank, Wissensbasis und Planungssystem.

Nr.	Art der Architektur	Domäne	Projektbeschreibung	Quellen
1	Konkret	E-Commerce	BusCat: Steuerung einer Website mit Produktkatalog	(Wyrd et al. 1996)
2		Support	Circuit-Fix-It Shop: Unterstützung bei Reparatur von Schaltungen	(Smith/Hipp 1994; McTear 2002)
3		Auskunftssystem	SUNDIAL: Auskunftssystem für Reiseinformationen	(Peckham 1993; McTear 2002)
4			TRAINS: Auskunftssystem für Bahnfahrplan	(Allen et al. 1995; McTear 2002)
5		Datenbankzugriff	LOQUI: Natürlichsprachliche Schnittstelle zu Datenbank	(Watchel 1989)
6		Automobil	Alternative Menüsteuerung im Fahrzeug	(Salmen 2002, 111ff)
7			VODIS: Sprachsteuerung für CD-Player mit Telefonanbindung im Fahrzeug	(Krahmer/Landsbergen/Pouteau 1997)
8	Konzeptionell	Automobil	Sprachgesteuerte Benutzeroberfläche für ein Bord-Dialogsystem	(Akyol/Libuda/Kraiss 2001)
9		-	Grundlegende Identifizierung wichtiger Komponenten	(Zue/Glass 2000)
10	Konzeptionell/ Konkret	-	EMBASSI: Elektronische Multimediale Bedien- und Service-Assistenz	(Elting et al. 2003; Hildebrand/Sá 2000; Buld et al. 2002)

Tabelle 5-1: *Untersuchte Sprachdialogsystemarchitekturen und ihre Merkmale*
Quelle: (Eigene Darstellung)

Einen allgemeinen Ansatz zum dialoggesteuerten Zugriff auf Informationen in einer Datenbank zeigt das LOQUI-System (Watchel 1989). Das System von Salmen (2002, 111ff) beschreibt eine konkrete, prototypische Anwendung zur sprachgesteuerten Menüsteuerung im Fahrzeug. Im Rahmen des Projekts VODIS beschreiben Kraemer, Landsbergen und Pouteau (1997, 84f) die Steuerung eines Radios mit CD-Player und Telefonanbindung im Fahrzeug per Sprache. Akyol (2001) stellt einen konzeptionellen Ansatz für die gleiche Domäne vor. Zue und Glass (2000) betrachten die Komponenten von Sprachdialogsystemen im Allgemeinen, während im Rahmen von EMBASSI sowohl eine umfangreiche Architektur als auch zahlreiche auf ihr basierende Anwendungen erstellt wurden (vgl. Elting et al. 2003). Einen Überblick der betrachteten Sprachdialogsysteme und ihrer wichtigsten Merkmale gibt Tabelle 5-1.

Avatarsysteme

Da manche Avatarsysteme auch die Funktionalitäten von Sprachdialogsystemen abdecken, lassen sich einige Systeme nicht nur einem Bereich zuordnen. Sofern ein System die Nutzung eines Avatars aufweist, wird es in dieser Untersuchung als Avatarsystem geführt, was zu einer größeren Zahl an betrachteten Avatarsystemen führt. Die meisten Beispiele sind in der Domäne des Internets zu finden, wo sie meist nur über textbasierte Kommunikation mit dem Benutzer interagieren können. In wenigen Fällen erfolgt auch eine Sprachausgabe (vgl. Kießling et al. 2001; Sölbrandt 2006). Im Bereich des E-Learning liegt nur in wenigen Fällen der Fokus der Darstellung auf der technischen Architektur, jedoch lassen sich die beiden Avatare Cosmo und Olga in diesen Bereich einordnen, welche die Funktion von Netzwerken bzw. eines Mikrowellenofens erklären (vgl. Lester et al. 1997a; Beskow/McGlashan 1997).

Viele Avatarsysteme werden nicht für eine praktische Anwendung, sondern als Laborprototypen zur Untersuchung eines bestimmten Teilaspekts entwickelt. Ein herausragendes System dieser Art ist der Real Estate Assistant (REA), der neben Sprachein- und -ausgabe auch Gesten- und Mimikerkennung integriert (Cassell et al. 1998) und sich somit als das bisher umfangreichste realisierte Avatarsystem darstellt. Neben der konkreten Anwendung liegt bei diesem System der Fokus auch auf der Ableitung einer Architektur für solche Systeme. Eine rein konzeptionelle Betrachtung von Avatarsystemen nehmen schließlich Arafa, Charlton und Mamdani (1999) mit ihrer KIMSAC (Kiosk-based Integrated Multimedia Service Access for Citizens)-Architektur vor. Die untersuchten Avatarsysteme und ihre wichtigsten Merkmale sind in Tabelle 5-2 aufgeführt.

Nr.	Art der Architektur	Domäne	Projektbeschreibung	Quelle
11	Konkret	Internet	EcoAvatar: Assistent zu Website	(Braun et al. 2003)
12			Cosima: Shopping-Assistent im Internet	(Kießling et al. 2001; Fischer et al. 2002)
13			BotCom: Erweiterung von Websites um Avatarsystem	(Laufer/Tatai/Nemeth 2005)
14			Erweiterung von Websites um Avatarsystem	(Sölbrandt 2006; Möbus et al. 2006)
15			Erweiterung einer Website mit Reiseinformationen	(André/Rist/Müller 1998b)
16		E-Learning	Cosmo: Lernassistent für die Funktion von Netzwerken	(Lester et al. 1997a; Lester et al. 1997b; Lester et al. 2000)
17			Olga erklärt die Funktion eines Mikrowellenofens	(Beskow/McGlashan 1997)
18		Auskunftssystem	Mobiles touristisches Informationssystem	(Cavalluzzi et al. 2004)
19		-	Framework zur Avataranimation	(Alexa et al. 2001)
20			FXPAL: virtueller Assistent	(Churchill et al. 2000)
21	Animation eines Gesichts als Gesprächspartner		(Nagao/Takeuchi 1994)	
22	Peedy: Assistent zum Abspielen von Musikstücken		(Ball et al. 1997)	
23	Konkret / Konzeptionell	Immobilien	REA: Real Estate Assistant, Virtueller Makler	(Cassell et al. 1998)
24	Konzeptionell	-	KIMSAC Architektur: Kiosk-based Integrated Multimedia Service Access for Citizens	(Arafa/Charlton/Mamdani 1999)

Tabelle 5-2: *Untersuchte Avatarsystemarchitekturen und ihre Merkmale*
Quelle: (Eigene Darstellung)

5.1.1.2. Identifizierte Komponenten

In den genannten Beispielen für Sprachdialog- und Avatarsysteme lassen sich über die in Kapitel 3.5 identifizierten Komponenten hinaus eine Vielzahl von Modulen finden, die sich unterschiedlichen Bereich zuordnen zu lassen. Um die Übersichtlichkeit zu erhöhen, soll bei der folgenden Aufzählung gefundener Komponenten eine Unterteilung in die Bereiche Ein- und Ausgabe, Datenkomponenten, Reaktionsermittlung und Designkomponenten erfolgen. Während die Ein- und Ausgabe alle Elemente zur Abdeckung multimodaler Verarbeitung und Steuerung beinhaltet, umschreiben die Datenkomponenten die Datencontainer, welche zur Reaktionsermittlung verwendet werden können. Die Antwortfindung selbst umfasst alle Elemente, um Information zu verarbeiten, zu vervollständigen und zu einer Antwort zusammenzustellen. Designkomponenten werden zur Gestaltung und Anpassung des Avatarsystems verwendet und sind somit nicht zwangsweise Teil der Laufzeitumgebung eines Avatarsystems. Im Folgenden wird keine vollständige Übersicht aller Fundstellen einzelner

Komponenten in der Literatur, sondern nur eine Übersicht der Komponenten sowie eine kurze Funktionsbeschreibung geben.

Komponenten zur Eingabe

Ein Eingabemodus, der sowohl bei Sprachdialogsystemen als auch bei manchen Avatarsystemen eine wichtige Rolle spielt, ist die Spracherkennung (McTear 2002, 103ff; Wyard et al. 1996, 188ff). Darüber hinaus werden bei internetbasierten Avatarsystemen zur Ergänzung einer Website textbasierte Eingabe sowie Ereignisse durch die Interaktion mit der Website selbst wie z.B. das Anklicken eines Links erkannt (Laufer/Tatai/Nemeth 2005; André/Rist/Müller 1998b). Zudem kann die Website auch proaktiv Ereignisse als Eingabe übergeben. Dieses Vorgehen kann prinzipiell nicht nur bei Websites, sondern bei allen angebundenen Softwaresystemen der Fall sein. Das System REA verwendet in seiner Laborumgebung mehrere Kameras, um die Präsenz, die Gestik und die Mimik des Benutzers zu erkennen (Cassell et al. 1998; Cassell et al. 2001; vgl. dazu auch Thórisson 1996). Neben universellen Eingabegeräten wie Maus und Tastatur können auch gesondert bereitgestellte haptische Eingabegeräte verwendet werden, wie sie beispielsweise im Fahrzeuginnenraum vorhanden sind (Akyol/Libuda/Kraiss 2001; Salmen 2002, 111ff)¹⁸. Neben diesen genannten Eingabemodi, die bereits in Sprachdialog- oder Avatarsystemen verwendet werden, sind aufgrund aktueller technischer Entwicklungen weitere Möglichkeiten wie Touchscreens, Temperatur- und Pulssensoren sowie Systeme zur Erkennung der Stimmung eines Benutzers auf der Basis seines Tonfalls oder der Wortwahl denkbar.

Elting et al. (2003, 93ff) verwenden in ihrer EMBASSI-Architektur weitere Komponenten für jede Eingabemodalität, welche jede Eingabe in einer Art Vorstufe aufarbeitet, bevor ein Ereignis weitergegeben wird. So sehen sie die Sprachanalyse bzw. das *language understanding* zur Auswahl der zutreffenden Erkennungsalternative aus einer Liste möglicher erkannter Formulierungen als Teil der Spracherkennung (vgl. dazu auch McTear 2002, 200ff; Zue/Glass 2000, 1168), die Bestimmung der Einheit, auf die per Gestenerkennung verwiesen wurde oder die Analyse der Oberflächeneingabe sind ebenfalls Ausprägungen dieser Vorstufe.

Da beim Einsatz multimodaler Eingaben die verschiedenen Modi unabhängig, synergetisch oder gegensätzlich zueinander stehen können, erfordert die Interpretation der Gesamtheit aller Eingabesignale eine weitere Komponente. Sie kommt unter dem Namen Interpreter (Akyol/Libuda/Kraiss 2001, 144), Input Manager (Cassell et al. 1998, 24f) oder polymodaler Input Manager (Elting et al. 2003, 94f) in verschiedenen Systemen vor und hat die Aufgabe, die Eingaben zu interpretieren und eine eindeutige Eingabe weiterzuleiten. Laufer, Tatai und Nemeth (2005) sehen in diesem Input Manager zudem das Element eines Multiplexers verankert, der gegebenenfalls die konsolidierte Eingabe an verschiedene Instanzen der Verarbeitungseinheit weiterleitet. Einen Überblick der identifizierten Eingabekomponenten gibt Tabelle 5-3.

¹⁸ Ein weiteres universelles Eingabegerät, das jedoch nur bei telefonbasierten Sprachdialogsystemen und somit nicht bei Avatarsystemen verwendet wird, ist das Telefon selbst.

Komponente	Kanal	Beschreibung	System
Spracherkennung	Akustisch	Umwandlung von Sprachsignalen in Text	(McTear 2002, 103ff; Wyard et al. 1996, 188ff)
Texteingabe	Haptisch	Eingabe von Text per Tastatur	(Laufer/Tatai/Nemeth 2005; André/Rist/Müller 1998b)
Externes System (Website)	Haptisch	Auslösen von Ereignissen als Eingabe	
Präsenzermittlung	Visuell	Ermittlung der physischen Präsenz eines Benutzers	(Cassell et al. 1998; Cassell et al. 2001; Thórisson 1996)
Gestenerkennung	Visuell	Visuelle Erkennung von Gesten	
Mimikerkennung	Visuell	Visuelle Erkennung von Mimik und Blickrichtung	
Eingabegeräte (Touchscreen)	Haptisch	Gesonderte Eingabegeräte in besonderen Umfeldern	(Akyol/Libuda/Kraiss 2001; Salmen 2002, 111ff)
Puls-/Temperaturmessung	Haptisch/ visuell	Messung der Aufregung eines Benutzers	-
Erkennung des Stimmung	Akustisch	Akustische Analyse des Sprachsignals	-
Vorstufe (Sprachanalyse, Gui-Analyse, Gestenanalyse)	-	Vorverarbeitung der Eingangssignale vor ihrer Weiterleitung	(Elting et al. 2003, 93ff)
Interpreter (Input manager, polymodaler Input manager)	-	Konsolidierung und Verrechnung multimodaler Eingangssignale	(Akyol/Libuda/Kraiss 2001, 144; Cassell et al. 1998, 24f; Elting et al. 2003, 94f)
Multiplexer	-	Weiterleitung der konsolidierte Eingabe an Verarbeitungseinheiten	(Laufer/Tatai/Nemeth 2005)

Tabelle 5-3: Komponenten zur Eingabe
Quelle: (Eigene Darstellung)

Komponenten zur Ausgabe

Auf der Ausgabeseite lassen sich die gleichen Modalitäten erkennen, mit denen auch die Eingabe erfolgt. So übernimmt die Sprachausgabe den akustischen Teil (McTear 2002, 126ff), während angebundene Displays den visuellen Teil repräsentieren (Salmen 2002, 111ff; Wyard et al. 1996; Kraemer/Landsbergen/Pouteau 1997). Cavalluzzi et al. (2004, 241) sehen die Darstellung bzw. Animation eines Avatars, der auch das Display zur Anzeige nutzt, als eine eigene Komponente innerhalb der Systemarchitektur. Sie stellen dieser Ausgabekomponente zudem den Body Generator voran, der die Planung der Avatarbewegung vornimmt. Als weitere visuelle Komponente, die bisher jedoch noch keine Anwendung in Avatarsystemen findet, kann ein emotionsabhängiges Licht ähnlich dem von Phillips verwendeten Ambilight-System eingesetzt werden (vgl. <http://www.aurea.philips.com>, zugegriffen am 1.4.2008). Zur Verwendung haptischer Ausgaben schlagen Akyol, Libuda und Kraiss (2001) den Einsatz von Vibration und Rückstellkraft vor, was jedoch nur durch besondere Geräte bewirkt werden kann. Je nach Anwendungsgebiet des Avatarsystems lassen sich auch sehr spezifische Ausgabegeräte identifizieren. So steuern Ball et al. (1997) einen

CD-Player (vgl. auch Kraemer/Landsbergen/Pouteau 1997), während die EMBASSI-Architektur die Integration von Videorecordern ermöglicht (Elting et al. 2003, 94). Als weitere Modalität, die bisher noch nicht Einzug in Avatarsysteme erhalten hat, ist die Verwendung äolischer Kräfte, d.h. die Erzeugung von Windstößen. Was bereits für Computerspiele eingesetzt wird (heise online 2006), könnte auch für Avatarsysteme eine Erweiterung darstellen.

Unabhängig von der eingesetzten Ausgabemodalität lassen sich verschiedene Komponenten zur Vorverarbeitung der auszugebenden Signale identifizieren. Zum einen lassen sich so genannte Wrapper erkennen, die als Stellvertreter eines realen Ausgabegerätes verwendet werden können. Salmen (2002, 111ff) verwendet dafür allgemein den Begriff der Devices und unterscheidet je nach dem, ob eine Vorverarbeitung durch einen Wrapper stattfindet, in reale und virtuelle Devices. Während reale Devices direkt steuerbare Geräte darstellen, verwenden virtuelle Devices eine Zwischenkomponente in Form eines Wrappers. In der EMBASSI-Architektur werden solche Zwischenkomponenten als X-Component bezeichnet (Elting et al. 2003, 94). Neben der Vorverarbeitung übernehmen sie zum einen die Aufgabe der Standardisierung der Schnittstelle zu Ausgabegeräten. Zum anderen verwenden Cavalluzzi et al. (2004, 241) solche Komponenten, um das angeschlossene Ausgabegerät auswechselbar zu machen. Sie setzen einen Wrapper ein, um die Verwendung verschiedener Avatarengines zur Avataramimation möglich zu machen.

Ebenso wie auf der Eingabeseite die Informationen, die im Zusammenspiel verschiedener Modalitäten verborgen sein können, durch den Input Manager dekodiert werden, müssen Informationen vor der Ausgabe erneut kodiert werden. Dieser Schritt erfolgt im Präsentationsmanager (Akyol/Libuda/Kraiss 2001, 144; Alexa et al. 2001, 157; André/Rist 1996), Character Composer (André/Rist/Müller 1998b) bzw. Action Scheduler (Cassell et al. 1998, 26f). Unabhängig von der Bezeichnung wird in dieser Komponente die über alle Modalitäten hinweg abgestimmte Ausgabe berechnet. Dabei spielen neben dem korrekten Timing auch gegenseitige Einflüsse der verschiedenen Modalitäten aufeinander eine Rolle. Lediglich Watchel (1989) trennt die Aufgaben des Präsentationsmanagers in die Aufgaben zur sprachlichen und grafischen Ausgabe auf. Akyol, Libuda und Kraiss (2001, 144) haben zudem als Vorstufe des Präsentationsmanagers einen Informationsmanager identifiziert, der auf der Basis des vorliegenden Belastungsgrades des Benutzers die verwendeten Modalitäten und ihren Inhalt gegebenenfalls beschränkt. Dieses Vorgehen ist sinnvoll, wenn zu erwarten ist, dass der Benutzer das Avatarsystem parallel zu einer wichtigen Aufgabe wie dem Autofahren verwendet. Durch die bewusste Reduktion der media richness kann die Ablenkung durch das Avatarsystem verringert werden (vgl. dazu Daft/Lengel 1984; Daft/Lengel/Trevino 1987; Rice 1992). Als Zwischenstufe zwischen dem Präsentationsmanager und den X-Components bzw. Devices sehen Elting et al. (2003, 94) verschiedene Assistentenkomponenten, die abhängig von der zu erwartenden Dauer der Ausgabedauer bzw. -art die Steuerung der Ausgabe übernehmen. Die genannten Ausgabekomponenten sind in der Übersicht in Tabelle 5-4 aufgeführt.

Komponente	Kanal	Beschreibung	System
Sprachausgabe	Akustisch	Text-to-Speech oder Abspielen von Sounddateien	(McTear 2002, 126ff)
Grafisches Display	Visuell	Anzeige von Text und grafischen Elemente	(Salmen 2002, 111ff; Wyard et al. 1996)
Avatardarstellung (Body Generator)	Visuell	Anzeige und Bewegung von Avataren	(Cavalluzzi et al. 2004, 241)
Vibration und Rückstellkraft	Haptisch	Haptische Rückmeldung an Benutzer	(Akyol/Libuda/Kraiss 2001)
Ausgabegerät (CD-Player, Videorecorder, öolische Einrichtung, Umgebungsbeleuchtung)	-	Steuerung von angeschlossenen Geräten	(Ball et al. 1997; Elting et al. 2003, 94)
Devices (X-Components, Wrapper)	-	Virtuelle Stellvertreter für angeschlossene Geräte	(Salmen 2002, 111ff; Elting et al. 2003, 94; Cavalluzzi et al. 2004, 241)
Präsentationsmanager (Character Composer, Action Scheduler)	-	Synchronisation der verschiedenen Ausgabemodalitäten	(Akyol/Libuda/Kraiss 2001, 144; Alexa et al. 2001, 157; André/Rist/Müller 1998b; Cassell et al. 1998, 26f),
Informationsmanager	-	Kontrolle von Ausgabemenge und –zeitpunkt	(Akyol/Libuda/Kraiss 2001, 144)
Assistenten	-	Verteilung der Ausgabeaufgaben	(Elting et al. 2003, 94)

Tabelle 5-4: *Komponenten zur Ausgabe*
Quelle: (Eigene Darstellung)

Datenkomponenten

Neben weiteren Datenbanken und –strukturen, die gegebenenfalls im Rahmen der spezifischen Aufgabe eines Avatarsystems zum Tragen kommen, lassen sich einige Datenkomponenten identifizieren, die unabhängig von der Aufgabe integrale Bestandteile sind. Aufgabenabhängige Datencontainer werden der Komponentenategorie der Reaktionsermittlung zugerechnet. In den untersuchten Systemen sind die einzelnen Datenstrukturen verschieden partitioniert, so dass es beim Vergleich von Strukturen aus verschiedenen Systemen zu Überschneidungen kommen kann.

Der erste Datencontainer wird bereits bei der Eingabe, in der Spracherkennung verwendet. Auch wenn aktuelle Spracherkennungssoftware bereits eine integrierte Sprachgrammatik beinhaltet, kann diese erweitert werden. Ebenso erfolgt die Verwendung einer Phonemdatenbank bei der Sprachsynthese, wobei diese meist nur durch den Hersteller angepasst werden kann. Eine Möglichkeit von der im Zusammenhang mit Avatarsystemen noch nicht berichtet wurde, ist die Verwendung eines Aussprachewörterbuchs zur korrekten Aussprache von Eigennamen oder Fremdwörtern. Avatarsysteme, die keine Sprachsynthese sondern vorgefertigte Audiosignale ausgeben, benötigen stattdessen einen entsprechenden

Datencontainer mit aufgezeichneten Audiosignalen (vgl. Ball et al. 1997). Ein weiterer Datencontainer, der direkt in Verbindung mit der Ausgabe steht, ist die Animationsdatenbank. Diese kann entweder komplette Filmsequenzen, Steuerdaten zur Echtzeitanimation (Ball et al. 1997) oder einen abstrakten Zugang zum Bewegungsapparat des Avatars anbieten. Alexa et al. (2001, 158) beschreiben eine entsprechende Komponente, welche die Animation des Avatars aufgrund dessen Verhalten und seiner „Stimmung“ automatisch berechnet.

Neben den Komponenten der Ein- und Ausgabe, benötigen insbesondere die Verarbeitungskomponenten umfangreiche Informationen. So lässt sich zum einen der Datencontainer zur Dialoghistorie nennen, in dem die gesamten Ein- und Ausgaben im Rahmen der aktuellen Dialogsession verzeichnet sind (Elting et al. 2003, 94; Watchel 1989). Zue und Glass erweitern diesen Datencontainer zum Diskurskontext, der neben den Daten über die Interaktion mit dem Benutzer noch weitere Statusinformationen enthalten kann, die sich während der Antwortermittlung ergeben (Zue/Glass 2000, 1168).

Zur korrekten Interpretation einer Benutzereingabe sind Informationen über den Benutzer hilfreich (Stegmann 2006, 36ff). Diese können in Form von Fakten über den Benutzer aber auch als Modell des emotionalen Zustands des Benutzers vorliegen (Laufer/Tatai/Nemeth 2005; Cavalluzzi et al. 2004, 241; Elting et al. 2003, 94). Neben dem emotionalen Modell des Benutzers ist auch ein emotionales Modell des Avatars denkbar, wie es Picard (2000) propagiert.

Für die korrekte Interpretation einer Eingabe bzw. die Umsetzung der Reaktion sind neben Daten über den Benutzer auch Informationen über das Umfeld des Avatarsystems von Bedeutung. Diese können im Falle statischer Informationen als Weltmodell vorliegen (Lester et al. 1997a, 65) oder für den Zugriff auf aktuelle Daten des Umfelds in Form der Dialogsituation organisiert werden (Akyol/Libuda/Kraiss 2001, 144). Während die statischen Umfeldinformationen der Orientierung in der Verarbeitungsphase dienen, kann die Integration aktueller Informationen in den Entscheidungsprozess insbesondere in Umgebungen mit hohem Ablenkungspotenzial durch das Avatarsystem gefährliche Situationen vermeiden. Die Datencontainer zum Benutzer und zum Umfeld werden in der EMBASSI-Architektur im Context Manager zusammengefasst. Neben diesen beiden Containern werden zudem die Daten der konkreten Anwendung sowie über die verfügbaren Ressourcen vorgehalten (Elting et al. 2003, 94f). Insgesamt lassen sich diese drei Datencontainer als Umfeldinformationen mit einem statischen sowie einem dynamischen Teil zusammenfassen.

Zue und Glass (2000) fassen die inhaltliche Interpretation der gesamten Kontextdaten als semantischen Rahmen zusammen, der die Bestimmung der Reaktion beeinflusst. In einer agenten-basierten Verarbeitung der Eingabe müssen zudem Ziele des Agenten und deren Beziehungen zueinander gespeichert sein, damit diese bei der Entscheidungsfindung berücksichtigt werden können (Cavalluzzi et al. 2004, 241). Im Falle von state- oder frame-basierten Systemen lässt sich die Wissensbasis mit den Regeln zur Antwortfindung als entsprechendes Äquivalent identifizieren (Churchill et al. 2000, 86; vgl. dazu auch Lindner 2003, 8f). Sofern dieser Datencontainer nicht extern als eigene Anwendung vorliegt, ist er ebenfalls als integraler Bestandteil des Systems zu sehen.

Ein Datencontainer, der in den untersuchten Architekturen kaum explizit genannt wird, jedoch ein wichtiger Bestandteil ist, sind die Logdaten (Laufer/Tatai/Nemeth 2005). In ihnen werden alle Ein- und Ausgaben sowie die interne Verarbeitung beschrieben, um den Entwicklern und Redakteuren das Erkennen von Fehlverhalten zu ermöglichen und es zu korrigieren. Die genannten Datenkomponenten sind in Tabelle 5-5 aufgelistet.

Komponente	Beschreibung	System
Sprachgrammatik	Grammatikregeln zur Spracherkennung und –synthese	(McTear 2002)
Aussprachewörterbuch	Verwendung zur korrekten Aussprache von Fremdwörtern und Eigennamen	-
Sound-Datenbank	Sammlung aufgenommener Klangdateien	(Ball et al. 1997)
Animationsdatenbank	Sammlung von Bildsequenzen oder Steuerdaten zur Avataranimation	(Ball et al. 1997; Alexa et al. 2001, 158)
Dialoghistorie (Diskurskontext)	Gespeicherte Ein- und Ausgaben des Benutzerdialogs	(Elting et al. 2003, 94; Zue/Glass 2000, 1168)
Benutzermodell	Modellhafte Darstellung der Eigenschaften und Absichten des Benutzers	(Laufer/Tatai/Nemeth 2005; Cavalluzzi et al. 2004, 241; Elting et al. 2003, 94)
Avatarmodell	Modellhafte Darstellung der Eigenschaften und Absichten des Avatarsystems	-
Umfeldinformationen (Weltmodell, Dialogsituation, Ressourcen)	Statische und dynamische Abbildung des Systemumfelds	(Lester et al. 1997a, 65; Akyol/Libuda/Kraiss 2001, 144; Elting et al. 2003, 94f)
Semantischer Rahmen	Interpretationsrahmen zur Reaktionsermittlung	(Zue/Glass 2000)
Agenten-Zielsystem	Ziele des Agenten und deren Zusammenhänge	(Cavalluzzi et al. 2004, 241)
Integrierte Knowledge Base	Integrierte Wissensbasis regelbasierter Systeme	(Churchill et al. 2000, 86; Lindner 2003, 8f)
Logdaten	Speicherung von Informationen aller Ein-, Ausgabe- und Verarbeitungsschritte	(Laufer/Tatai/Nemeth 2005)

Tabelle 5-5: Datenkomponenten
Quelle: (Eigene Darstellung)

Komponenten zur Reaktionsermittlung

Der gesamte Prozess der Antwortfindung bzw. Reaktionsermittlung lässt sich grob in zwei Bereiche unterteilen: Einen steuernden Teil und einen antwortermittelnden Teil. Der steuernde Teil übernimmt dabei die Ermittlung des Arbeitsablaufs und die Verteilung von Einzelaufgaben an Komponenten. Die zentrale Komponente dieser Workflowsteuerung ist der Dialogmanager, der in fast allen untersuchten Systemen vorkommt und die zentrale Kommunikationsschnittstelle für alle weiteren Komponenten dieser Phase darstellt (McTear 2002, 139ff; Zue/Glass 2000, 1168; Cavalluzzi et al. 2004, 240f). Der Dialogmanager, der

vereinzelt auch als *explanation planner* (vgl. Lester et al. 1997a, 65) oder Eventhandler (André/Rist/Müller 1998b) bezeichnet wird, erhält die konsolidierten Eingabedaten, verteilt sie in einer bestimmten Reihenfolge an angebundene Komponenten bis schließlich die fertige Antwort ermittelt wurde und übermittelt sie an den Präsentationsmanager. Dabei übernimmt er im Falle einer Spracherkennung bzw. Texteingabe auch die Kompensation von falsch formatierten oder unvollständigen Eingaben (McTear 2002, 139ff), um das System robust zu gestalten.

Im Zusammenhang mit dieser zentralen Stelle lassen sich noch weitere Komponenten identifizieren, die gesondert oder als Bestandteil des Dialogmanagers ihre Aufgaben verrichten. So lässt sich das *task module* identifizieren, welches Informationen über den Ablauf der Antwortermittlung beinhaltet und somit die Steuerung der Schritte im Dialogmanager übernimmt (McGlashan et al. 1990; McTear 2002, 139ff).

Eine weitere ähnlich gelagerte Komponente ist das *reaction module*, welches Informationen über den Verlauf und den aktuellen Status des Dialogs verwendet, um beispielsweise ohne die Abarbeitung aller relevanten Schritte sofort eine Reaktion auf eine Benutzereingabe auszugeben (Cassell et al. 1998, 26; Churchill et al. 2000, 85f). Dies ermöglicht es, den Avatar in eine zuhörende Haltung zu versetzen, sobald ein erstes akustisches Signal aufgenommen wird, noch bevor der Sinn des Gesagten ermittelt wurde. Diese Umleitung einer Eingabe an den vorgegebenen Schritten vorbei kann mit Reflexen des menschlichen Körpers verglichen werden. Salmen (2002, 111ff) bezeichnet diese Komponente als *message handler*, der hoch priorisierte Eingangsnachrichten sofort verarbeitet, während andere Nachrichten in eine Queue gelegt und nacheinander abgearbeitet werden.

Die Komponenten der eigentlichen Antwortfindung werden von Cassell et al. (1998, 26) als *deliberative module* zusammengefasst, während Churchill et al. (2000, 85f) eine Differenzierung dieses Moduls in die Bereiche *understanding module*, *response planner* und *generation module* vornimmt. Diese Unterteilung entspricht den wichtigsten Schritten der Antwortfindung, kann jedoch insbesondere im Bereich des *response planners* weiter differenziert werden.

Das *understanding module*, wie es Churchill et al. (2000, 86) verstehen, entspricht in dieser Untersuchung den Funktionen des Input Managers, der eingehende Signale miteinander verrechnen kann. Dennoch spielt dieser Schritt insofern eine Rolle, da die konsolidierte Eingabe einer semantischen Analyse unterzogen werden kann. Zum einen können an dieser Stelle die Eingabesignale bereinigt und Verweise auf vergangene Äußerungen mithilfe der Dialoghistorie aufgelöst werden (Hirst 1981, 4ff; Wyard et al. 1996, 199). Darüber hinaus kann auf der Grundlage der ermittelten semantischen Informationen das Benutzermodell aktualisiert werden, was von Nagao und Takeuchi (1994, 106) als *plan recognition* bzw. von McTear (2002, 139ff) als *belief module* bezeichnet wird. Dabei werden die aktuelle Absichten und damit das Zielsystem des Benutzers angepasst, um im nächsten Schritt die passende Reaktion ermitteln zu können.

Der nächste Schritt erfolgt nach Churchill et al. (2000, 86) im *response planner*, der auf der Grundlage des Ergebnisses aus dem vorangegangenen Schritt eine entsprechende Reaktion

ermittelt. Je nachdem, ob für die eigentliche Reaktionsermittlung eine agenten-basierte Komponente zum Tragen kommt oder nicht, lassen sich bei den untersuchten Systemen drei verschiedene Ansätze identifizieren. Der erste Ansatz greift auf den Datencontainer der Knowledge Base zurück, die bereits vorgestellt wurde. Dabei wird eine Abfrage nach einer passenden Antwort mit dem Ergebnis des *understanding modules* durchgeführt. In den hinterlegten Regeln wird eine passende Reaktion ermittelt und zurückgegeben (Zue/Glass 2000, 1168; Braun et al. 2003).

Der zweite Ansatz ist eine Mischung aus statischen Regeln und der dynamischen Zielverfolgung eines Agenten. Dabei wird während der Laufzeit auf ein Regelwerk zugegriffen, das zur Designzeit kompiliert wurde. Diese Kompilierung setzt alle möglichen Verhaltensmuster des Avatarsystems in statische Regeln um (André/Rist/Müller 1998b; Lester et al. 1997a). Dieser Zwischenschritt ist notwendig, wenn das Verhaltensmuster des Avatarsystems so komplex ist, dass es zur Laufzeit nicht in angemessener Zeit verarbeitet werden kann. Lester et al. (1997a, 65) nennen dieses kompilierte Regelwerk *behavior space*, während Andre/Rist/Müller (1998b) es als *behavior monitor* bezeichnen.

Der dritte, agenten-basierte Ansatz ermittelt die Reaktion auf Basis des Avatar- bzw. Agentenmodells und seines Zielsystems. Arafa, Charlton und Mamdani (1999, 6ff) unterscheiden innerhalb ihres *personality models* die drei Bereiche der *emotion selection*, *behavior description* und *behavior generation*. Zunächst wird der aktuelle Gefühlszustand sowohl des Agenten als auch des Benutzers ermittelt und dazu ein entsprechendes Verhalten generiert. Auf Basis der Verhaltensbeschreibungen werden dazu die wahrnehmbaren Veränderungen in Aussehen, Bewegung und Sprache des Avatars ermittelt und als Antwort des *response planners* zurückgegeben. Alexa et al. (2001, 158) sehen die Umwandlung von Absichten bzw. geplantem Verhalten in konkrete Handlungen als Teil einer gesonderten Komponente, der *behavior control*. Laufer, Tatai und Nemeth (2005) bilden die Beeinflussung der Avatarreaktion durch das emotionale Modell GALA ab, indem 24 diskrete emotionale Zustände abgebildet sind.

McTear (2002, 113) identifiziert weitere Komponenten, die er als *external communication* zusammenfasst, und die nach der Unterscheidung von Churchill et al. (2000, 86) zum Bereich des *response planners* gehören. Die einzelnen Komponenten sind die Platzhalter verschiedener externer Systeme, die zur Ergänzung der bereits zuvor ermittelten Reaktion des Avatarsystems verwendet werden. Diese externen Systeme sind nicht Bestandteil des Avatarsystems, bieten jedoch Zugriff auf wichtige Daten. Er unterscheidet die Anwendungsdatenbank (vgl. dazu auch Wyard et al. 1996, 201), die externe Knowledge Base (vgl. dazu auch Smith/Hipp 1994) und das Planning System (vgl. dazu auch Allen et al. 1995). Für den Fall der Anwendungsdatenbank wird als vorgelagerte Stufe der *information manager* verwendet, der die Abfrage an die Datenbank gegebenenfalls anpasst und fehlende Daten für die Anfrage nach Möglichkeit ergänzt (Whittaker/Attwater 1996). Ein ähnliches Vorgehen verwenden auch Kießling et al. (2001, 3) bei ihrer PreferenceSQL, die vage Suchkriterien in SQL-Anfragen an eine Datenbank umsetzt. Als weitere Komponente lässt sich ein Suchagent, wie ihn Kießling et al. (2001, 3) vorschlagen, in den Bereich der *external communication* subsumieren. Er kann im Internet oder in anderen angebotenen Quellen nach erforderlichen Informationen suchen.

Bereich	Komponente	Beschreibung	System
Workflow	Dialogmanager	Zentraler Kommunikationsknoten während der Antwortermittlung	(McTear 2002, 139ff; Zue/Glass 2000, 1168; Cavalluzzi et al. 2004, 240f)
	Task module	Steuerung des Dialogmanagers	(McGlashan et al. 1990; McTear 2002, 139ff)
	Reaction module (Message Handler)	Behandlung von Eingangsnachrichten abhängig von ihrer Priorität	(Cassell et al. 1998, 26; Churchill et al. 2000, 85f; Salmen 2002, 111ff)
Deliberative module	Understanding module (plan recognition, belief module)	Semantische Analyse der Eingabe zur Aktualisierung des Benutzermodells	(Nagao/Takeuchi 1994)@106; McTear, 2002 #544@139ff}
	Knowledge Base	Zugriff auf statisches Regelwerk	(Zue/Glass 2000, 1168; Braun et al. 2003)
	Behavior monitor/ behavior space	Zugriff auf statische, zur Designzeit kompiliertes Regelwerk	(André/Rist/Müller 1998b; Lester et al. 1997a)
	Personality model (Emotion Module/ behavior control)	Ermittlung bzw. Beeinflussung der Reaktion durch emotionales Model	(Arafa/Charlton/Mamdani 1999; Laufer/Tatai/Nemeth 2005; Alexa et al. 2001, 158)
	External communication: Application Database	Zugriff auf externe Anwendungsdatenbank	(McTear 2002, 113; Wyard et al. 1996, 201)
	External communication: Knowledge Base	Zugriff auf externe Wissensdatenbank	(McTear 2002, 113; Smith/Hipp 1994)
	External communication: Planning System	Zugriff auf externes Planungssystem	(McTear 2002, 113; Allen et al. 1995)
	Information manager (PreferenceSQL)	Konvertierung und Vervollständigung von Datenbankabfragen	(Whittaker/Attwater 1996; Kießling et al. 2001, 3)
	Suchagent	Sucht im Internet oder anderen Quellen nach Informationen	(Kießling et al. 2001, 3)
	Response generation (language generation)	Zusammenstellung des Drehbuch für die Reaktion des Systems	(Nagao/Takeuchi 1994, 104; Wyard et al. 1996, 202f)

Tabelle 5-6: Komponenten zur Reaktionsermittlung
Quelle: (Eigene Darstellung)

Den letzten Schritt nach der Einteilung von Churchill et al. (2000, 86) stellt die *response generation* (Nagao/Takeuchi 1994, 104; Wyard et al. 1996, 202f) oder auch *language generation* (McTear 2002, 124) im Fall von reinen Sprachdialogsystemen dar. Dabei wird die auszugebende Antwort hinsichtlich der auszugebenden Informationen, der Wortwahl bzw. mit Bezug auf die Benutzereingabe vorgenommen. Zudem werden Bewegungen des Avatars und seiner Umwelt an den entsprechenden Stellen eingebunden. Dieser Vorgang lässt sich als Generierung eines Drehbuchs für die Reaktion verstehen und ist der letzte Schritt der Reaktionsermittlung. Alle Komponenten der Reaktionsermittlung sind als Übersicht in Tabelle 5-6 aufgeführt.

Designkomponenten

Die initiale Erstellung verschiedener Daten und ihre fortlaufende Pflege werden mithilfe von Designkomponenten realisiert. Die Bearbeitung von Datencontainern ist dabei auf statische Daten beschränkt, die sich zur Laufzeit nicht ändern. Zu ihnen gehören beispielsweise die Avatare selbst, deren Bewegungsabläufe, sowie im Falle von Chatbots deren Regelbasis bzw. das Zielsystem von Agenten. Darüber hinaus kann auch das Umfeld- bzw. Weltmodell angepasst werden.

Die Anpassung des Avatars und seiner Bewegungen erfolgt im Avatar Editor (Sölbrandt 2006, 120f), wodurch Äußerlichkeiten des Avatars verändert und neue Animationsskripte angelegt werden können. Laufer, Tatai und Nemeth (2005) verwenden einen Editor zur Anpassung der Regeln eines Chatbots, der die Erweiterung und Korrektur der Wissensbasis ermöglicht (vgl. dazu auch Lindner 2003, 8f). Bei Verwendung eines *behavior space* (vgl. André/Rist/Müller 1998b), muss im Editor zudem eine Kompilierung des Verhaltens stattfinden. Der Editor kann durch weitere Hilfsmöglichkeiten erweitert werden. So lassen sich mithilfe des *log analyzers* Log-Dateien, die während der Laufzeit erstellt wurden, analysieren und daraus Rückschlüsse auf Benutzereingaben ziehen, die mit der aktuellen Wissensbasis nicht ausreichend verarbeitet werden können (Laufer/Tatai/Nemeth 2005). Braun et al. (2003, 98ff) schlagen zur automatischen Erweiterung des Regelwerks die Komponente des *content analyzers* vor, der auf Basis von Dokumenten neue Regeln hinzufügt. Dieses Vorgehen lässt sich jedoch nur dann sinnvoll realisieren, wenn der Inhalt, über den das Avatarsystem Auskunft geben kann, eben diese Dokumente sind und sie über einen bestimmten Zeitraum hinweg unverändert bleiben.

Neben diesen gängigen Komponenten, erfordert die Anpassung des Weltmodells sowie des Zielsystems bei einem agenten-basierten System die Existenz weiterer Editoren. Trotz ihrer Notwendigkeit wurde deren Existenz bzw. Funktion bisher nicht beschrieben.

Die untersuchten Architekturen gehen bei der Aktualisierung aller statischen Datencontainer zudem davon aus, dass sie auf einer zentralen, einfach zugänglichen Hardwarekomponente betrieben werden und jederzeit verändert werden können. Diese implizite Annahme lässt sich bei verteilten oder mobilen Avatarsystemen nicht gewährleisten. Cavalluzzi et al. (2004) haben in ihrem mobilen Avatarsystem dieses Problem dermaßen gelöst, dass lediglich die Ein- und Ausgabe mobil ist und die Verarbeitung zentral erfolgt. Dabei wird jedoch vorausgesetzt, dass eine konstante Verbindung zur zentralen Einheit aufrechterhalten werden kann. In Fällen, in denen dies nicht möglich ist, muss auch der Verarbeitungsteil des Avatarsystems auf dem mobilen Gerät untergebracht werden. Um bei einer solchen Konstellation eine Aktualisierung der Daten zu realisieren, muss zumindest sporadisch eine Verbindung zu einer zentralen Update-Komponente aufgebaut werden können (Glass/Seneff 2003, 21). Eine Übersicht aller identifizierten Komponenten des Designs und der Pflege sind in Tabelle 5-7 aufgeführt.

Komponente	Beschreibung	System
Avatar Editor	Erstellung und Anpassung des Avatars und seiner Bewegungen	(Sölbrandt 2006, 120f)
Dialogeditor	Erstellung und Anpassung des Regelwerks von Chatbots	(Laufer/Tatai/Nemeth 2005)
Log Analyser	Analyse von Logdateien zur Ergänzung fehlender Regeln	(Laufer/Tatai/Nemeth 2005)
Content Analyzer	Automatische Erstellung von Regeln zu Dokumenten im Avatarsystem	(Braun et al. 2003, 98ff)
Weltmodell Editor	Erstellung und Anpassung des Weltmodells	-
Zielsystem Editor	Erstellung und Anpassung des Zielsystems bei agenten-basierten Avatarsystemen	-
Update Manager	Verteilung und Synchronisation aktualisierter Regeln an mobile Avatarsysteme	-

Tabelle 5-7: Komponenten zum Design und der Pflege
Quelle: (Eigene Darstellung)

5.1.2. Konzeption einer Referenzarchitektur für Avatarsysteme

Der Begriff der Architektur fasst seit dem Mittelalter die Bedeutungen von Baustil und Baukunst zusammen. Der Baustil beinhaltet die konzeptionelle Betrachtung des Bauens, während die Baukunst die korrekte Umsetzung zum Gegenstand hat (Horn/Reinke 2002, 15). Im modernen Kontext der IT lassen sich Architekturen auf verschiedenen Ebenen finden wie z.B. einer IT-Landschaftsarchitektur, die eine Vielzahl von Systemen umfasst (vgl. Lankes/Matthes/Wittenburg 2005), oder Software-Architekturen, welche das Zusammenspiel verschiedener Elemente innerhalb einer Anwendung beschreiben (Witt/Baker/Meritt 1994).

Horn und Reinke (2002, 15) unterscheiden neben der Architektur, die eine konkrete Ausprägung bezeichnet, das Architekturmuster, welche als Vorlage für die Architektur einer bestimmten Art oder einzelner Bauteile dient. Eine Referenzarchitektur lässt sich demnach als Architekturmuster beschreiben, das als Vorlage einer Vielzahl konkreter Architekturen dienen kann. Im Rahmen dieser Forschungsarbeit dient die zu erstellende Referenzarchitektur als Vorlage für den Bau eines Avatarsystems. Da für den Bau eines konkreten Avatarsystems spezifische Anforderungen vorliegen, muss die Referenzarchitektur allen möglichen Anforderungen Rechnung tragen und beinhaltet somit eine Vielzahl von Bestandteilen, die nicht in jedem Fall verwendet werden.

Eine IT-Architektur bzw. ein –muster beinhaltet drei elementare Typen von Bestandteilen: Komponenten, Konnektoren und Systeme (Horn/Reinke 2002, 18f):

- Komponenten repräsentieren Verarbeitungs- oder Speichereinheiten eines Systems und können ihrerseits aus mehreren Komponenten zusammengesetzt sein. Dementsprechend ist eine solch zusammengesetzte bzw. komposite Komponente selbst wiederum ein System.

- Ein Konnektor stellt die Verbindung zwischen Komponenten dar und beeinflusst damit die Interaktion zwischen ihnen. Ebenso wie Komponenten können auch Konnektoren zu kompositen Strukturen zusammengestellt werden.
- Ein System beinhaltet einzelne Komponenten, Konnektoren und deren Zusammenstellung in Form von Systemen.

Der Bedarf einer Referenzarchitektur für Avatarsysteme erwächst aus der Beobachtung, dass keine Erfahrungen über die Umsetzung eines solchen Systems im Fahrzeug veröffentlicht und somit zugänglich sind. Da das automobiler Umfeld neue Möglichkeiten und Restriktionen für die Implementierung eines solchen Systems aufweist, ist die strukturierte Analyse bestehender, vergleichbarer Systeme unabdingbar, um unerwarteten Problemen während der Implementierung bestmöglich vorzubeugen. Ein Aspekt, der den Einsatz im Fahrzeug bereits auf den ersten Blick von den meisten untersuchten Systemen unterscheidet, ist die Mobilität des Systems. Eine Betrachtung dieses Aspekts haben unter den untersuchten Systemen nur Cavalluzzi et al. (2004) vorgenommen, wobei deren Domäne eine andere war. Die nachfolgende Referenzarchitektur versteht sich somit als erster Schritt zu einer konkreten Architektur eines Avatarsystems im Auto.

Gemäß den unterscheidbaren Systemen sollen im Folgenden zunächst die einzelnen Systeme sowie das Zusammenspiel ihrer Komponenten betrachtet werden, bevor die gesamte Referenzarchitektur zusammengesetzt wird. Dabei lassen sich die bereits als Komponentenbereiche vorgestellten Teile als Systeme identifizieren. Lediglich die Datenkomponenten sind über alle Systeme verteilt.

5.1.2.1. *Eingabesystem*

Das Eingabesystem ist primär durch die verschiedenen Eingabekanäle und die nachgelagerte Vorverarbeitung gekennzeichnet, in der die eingehenden Signale derart konvertiert werden, dass sie einfach maschinell weiterverarbeitet werden können. Im Rahmen der Vorstufe wird dabei auch im Falle der Sprachverarbeitung auf Sprachgrammatiken zugegriffen. Die Interpretation von Gesten und Mimiken erfordert in der Vorstufe das statische Wissen über das Umfeld, um das Objekt identifizieren zu können, auf das gezeigt oder geblickt wird. Diese Umfeldinformationen werden auch im Reaktionsermittlungssystem verwendet und können je nach Verteilung der Systeme auch gemeinsam genutzt werden. Die aufbereiteten Signale gehen in den Input Manager ein, der sie an das Reaktionsermittlungssystem weiterleitet. Diese Weiterleitung kann über einen Multiplexer erfolgen, sofern verschiedene Reaktionsermittlungssysteme parallel mit den eingehenden Informationen versorgt werden können. Je nach Verteilung und konkreter Umsetzung kann der Input Manager entweder Teil des Input Systems oder des Reaktionsermittlungssystems sein. Ein wichtiges Entscheidungskriterium bei dieser Option kann die Bandbreite zwischen den beiden Systemen sein. Liegt der Input Manager im Input System, so wird nur die konsolidierte Eingabe übermittelt, während ansonsten mit mehr Übertragungssignalen zu rechnen ist. Die Zusammenhänge zwischen den Komponenten des Input Systems sind in Abbildung 5-2 dargestellt.

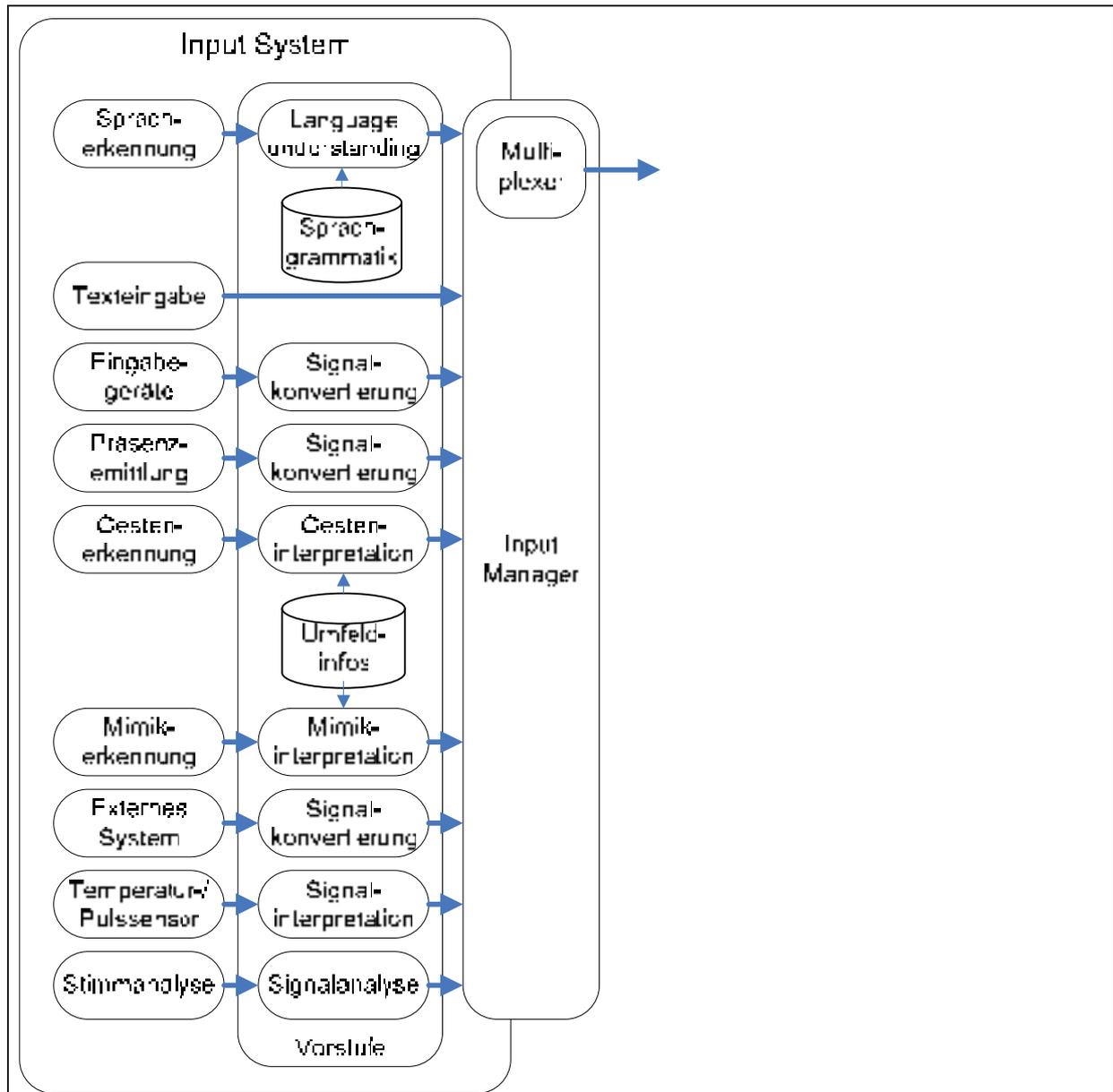


Abbildung 5-2: *Input System der Referenzarchitektur für Avatarsysteme*
Quelle: (Eigene Darstellung)

5.1.2.2. *Ausgabesystem*

Das Ausgabesystem beginnt seine Verarbeitung, sobald der Präsentationsmanager ein entsprechendes Signal erhalten hat. Dieses kann aufgrund einer ordentlichen Verarbeitung oder durch das *reaction module* im Reaktionsermittlungssystem als Reflex generiert worden sein. Ebenso wie der Input Manager kann auch der Präsentationsmanager ein Teil des Output Systems oder des vorgelagerten Reaktionsermittlungssystems sein. Die Planung der synchronen Ausgabe hängt zu einem großen Teil von den verwendeten Modalitäten ab. Insbesondere bei der Verwendung von Sounddateien und Animationen stellen die Länge der einzelnen Dateien oder Animationssequenzen einen entscheidenden Faktor dar. So wird bereits während der Planung im Präsentationsmanager auf die entsprechenden Datenbanken zugegriffen. Die Entscheidung, welche Modalitäten bei einer bestimmten Ausgabe zum

Tragen kommen, hängt auch vom Informationsmanager ab. Er ermittelt mithilfe der dynamischen Umfeldinformationen die Situation und verwendet je nach Belastung des Benutzers die zumutbaren Ausgabekanäle. Je nach Verteilung der Systeme können auch diese Umfeldinformationen zentral vorliegen. Bei der Verwendung von Assistenten zur Steuerung der Ausgabe können diese als gesonderte Komponenten vorliegen oder verschieden parametrisierte Instanzen des Präsentationsmanagers selbst sein. In jedem Fall muss die Ansteuerung der Ausgabekanäle durch den Präsentationsmanager bzw. durch die Superinstanz der Assistenten synchronisiert werden. Die weitergeleiteten Ausgabesignale werden durch Devices bzw. X-Components unter Einbezug entsprechender Datencontainer vorverarbeitet und ausgegeben. Die Sprachausgabe verwendet dabei entweder eine Sound-Datenbank oder eine Sprachgrammatik zur Synthese, während die Animationsaufbereitung auf die Animationsdatenbank zugreift. Alle Komponenten des Ausgabesystems sind in Abbildung 5-3 zu sehen.

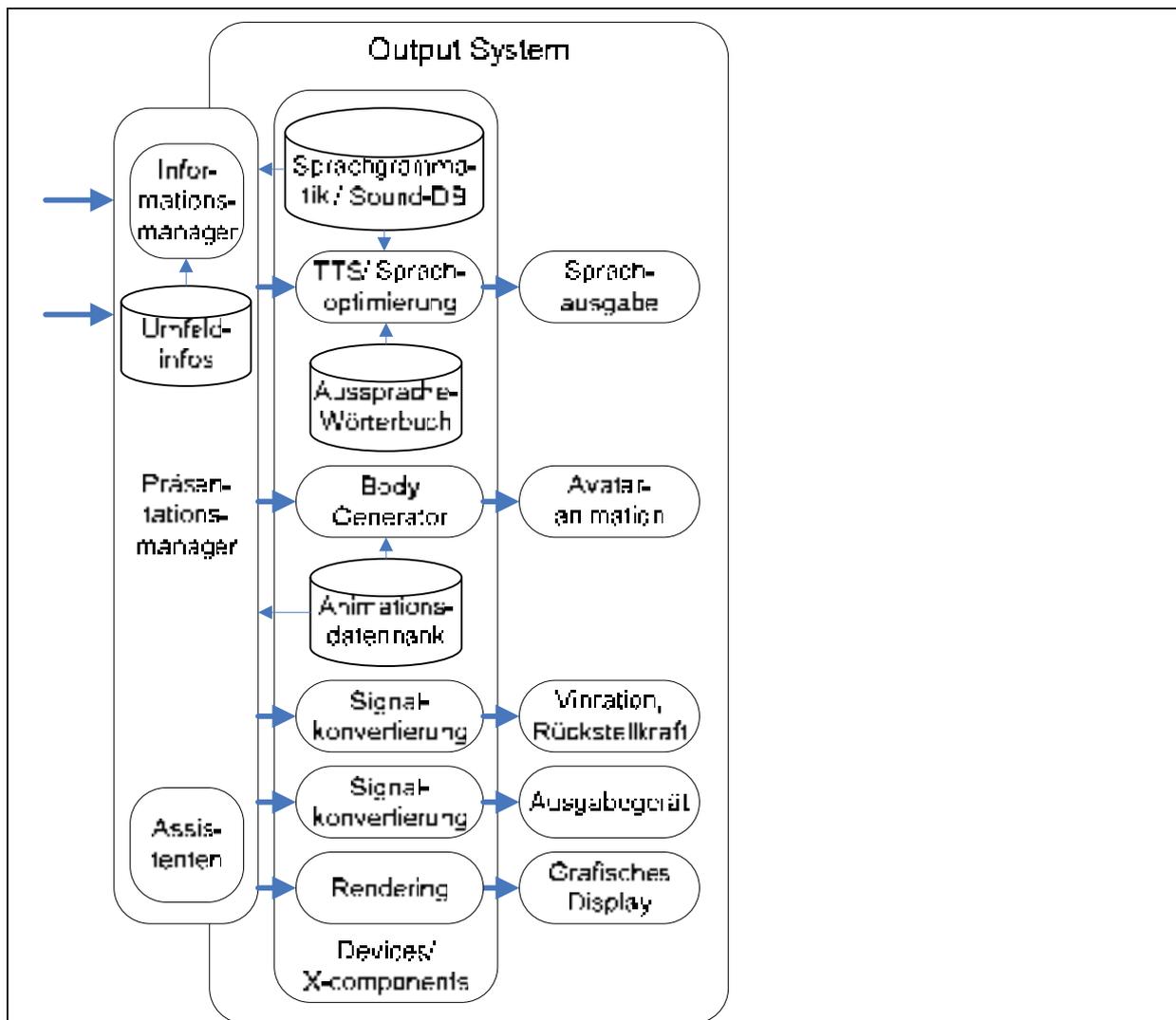


Abbildung 5-3: Ausgabesystem der Referenzarchitektur für Avatarsysteme
Quelle: (Eigene Darstellung)

5.1.2.3. Designsystem

Das Designsystem beinhaltet insgesamt vier Editoren für die Bereiche Dialog, Avatar, Weltmodell und Zielsystem. Der Dialogeditor ist aufgrund der Integration der beiden Komponenten des Log und des Content Analyzers selbst wiederum ein System. Der Log Analyzer wertet die Dialoghistorie und Logdaten aus, während der Content Analyzer angebundene Datenquellen regelmäßig auf Änderungen hin überprüft und daraus Regeln generiert. Der Avatareditor ermöglicht die Pflege des Avatars bzw. weiterer 3D-Objekte in dessen Umfeld. Der Weltmodelleditor kann für die Anpassung des statischen Teils der Umfeldinformationen verwendet werden. Der Zielsystemeditor erlaubt die Anpassung des Agentenzielsystems und kann je nach gewähltem Ansatz auf Grundlage dieser Angaben eine Regelbasis kompilieren.

Die Verbindung des Designsystems mit dem Reaktionsermittlungssystem erfolgt über den Update Manager, der sowohl den synchronen als auch den asynchronen bidirektionalen Datenaustausch ermöglicht. Dabei greifen alle Editoren schreibend in das Reaktionsermittlungssystem ein, während nur das Dialogeditorsystem auch Daten liest, die während der Laufzeit geändert wurden. Der Update Manager kann bei einer räumlichen Verteilung der beiden angeschlossenen Systeme sowohl im Reaktionsermittlungs- als auch im Designsystem vertreten sein. Dies ermöglicht es, zum einen Änderungen während der Laufzeit zu verzeichnen und zum anderen geänderte Daten der Designzeit für einen asynchronen Datenaustausch bereitzuhalten. Im Falle des Avatareditors, der auf einen Datencontainer zugreift, der dem Ausgabesystem zugeordnet ist, erfolgt eine gesonderte Verbindung des Update Managers zur Animationsdatenbank. Die Zusammenhänge der Komponenten des Designsystems sind in Abbildung 5-4 zu sehen, wobei die angebundene Datencontainer auf Seiten des Reaktionsermittlungssystems aus Gründen der Übersicht nicht dargestellt sind.

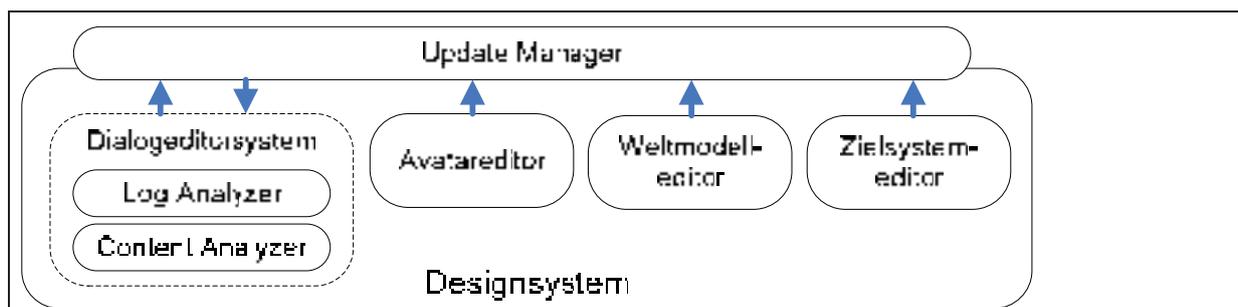


Abbildung 5-4: Designsystem der Referenzarchitektur für Avatarsysteme
Quelle: (Eigene Darstellung)

5.1.2.4. Reaktionsermittlungssystem

Das Reaktionsermittlungssystem als zentrale Komponente lässt sich zwischen Input und Output System lokalisieren. Es verarbeitet die konsolidierte Eingabe und gibt die ermittelte Reaktion an das Output System weiter. Federführend bei der Reaktionsermittlung ist der Dialogmanager, der bei der Verwendung eines Multiplexers im Input System gegebenenfalls mehrfach instanziiert wird. Das *reaction module* im Dialogmanager entscheidet, ob auf eine Eingabe eine reflexartige oder eine ordentliche Reaktion folgt und leitet die Eingabe

dementsprechend der ordentlichen Verarbeitung oder direkt dem Präsentationsmanager zu. Im Falle der ordentlichen Verarbeitung bestimmen das *task module* und der semantische Rahmen die Schrittfolge der Abarbeitung. Während der Verarbeitung werden die Dialoghistorie und der dynamische Teil der Umfeldinformationen durch den Dialogmanager aktualisiert.

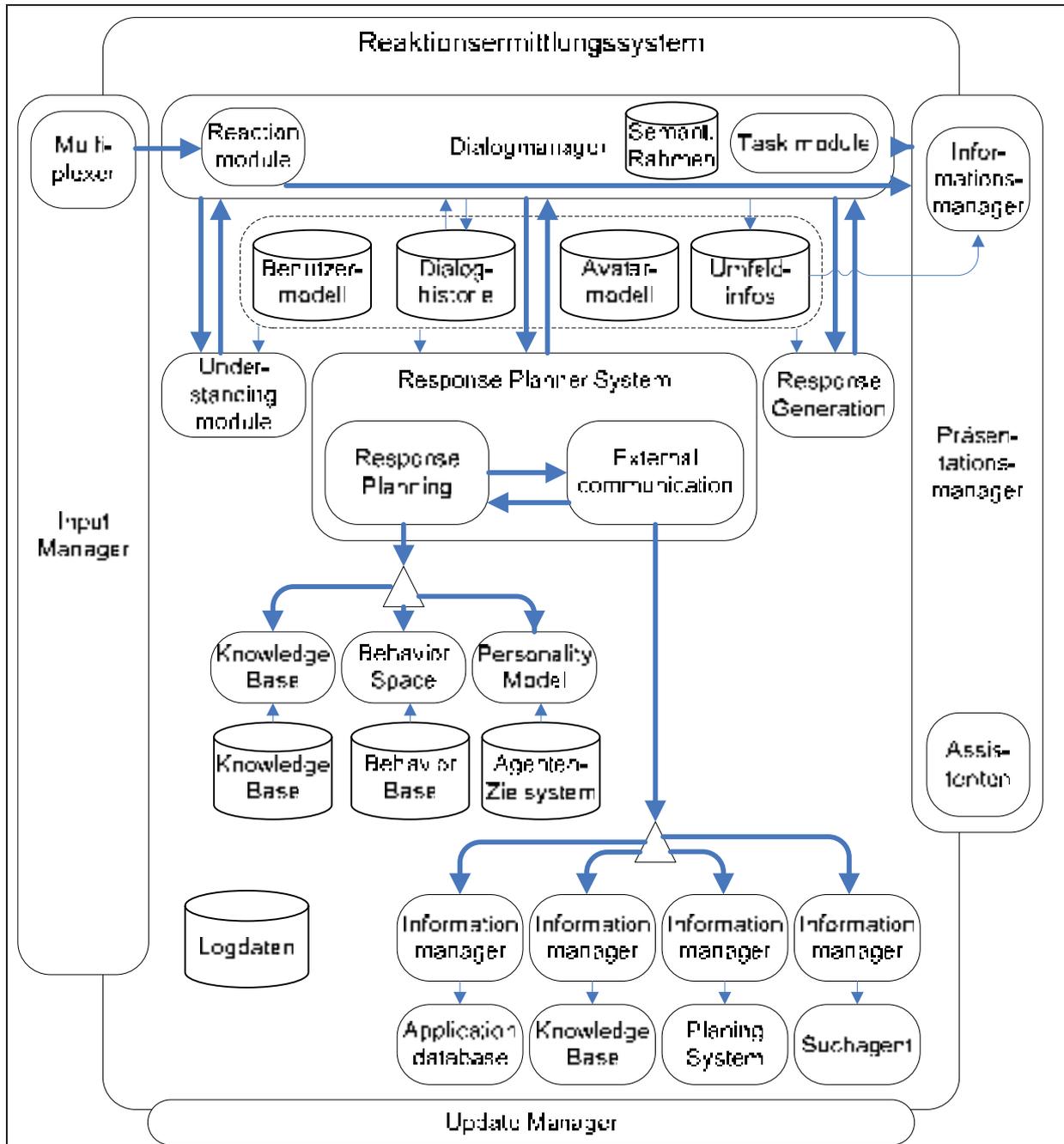


Abbildung 5-5: *Reaktionsermittlungssystem der Referenzarchitektur für Avatarsysteme*
 Quelle: (Eigene Darstellung)

Die ordentliche Verarbeitung erfolgt entlang der drei bereits bei der Komponentenübersicht identifizierten Bereiche des *understanding module*, des *response planner* und der *response generation*. Aufgrund der Zusammenstellung des *response planners* aus dem eigentlichen *response planning* und der *external communication* stellt dieser Bereich selbst wiederum ein

System dar. Die einzelnen Bereiche werden über den Dialogmanager in der entsprechenden Reihenfolge mit den notwendigen Daten versorgt.

Das *understanding module* verwendet, ebenso wie die beiden anderen Bereiche, das Benutzer- und gegebenenfalls das Avatarmodell, die Dialoghistorie sowie die Umfeldinformationen zur Auflösung der Benutzereingabe. Das *response planner* System beinhaltet die beiden Komponenten des *response planning* und der *external communication*. Das *response planning* kann auf Basis eines regelbasierten, eines vorkompilierten oder eines rein agenten-basierten Ansatzes erfolgen. Dementsprechend werden verschiedene Datencontainer verwendet, wobei die *behavior base* vom technischen Standpunkt einer *knowledge base* entspricht, die jedoch auf Inhalten eines Agentenzielsystems basiert. Die Vervollständigung der ermittelten Antwort erfolgt gegebenenfalls mithilfe der *external communication*, die vier Ausprägungen aufweisen kann. Im Gegensatz zum *response planning* können bei der *external communication* mehrere Datenquellen verwendet werden. Der Abschluss der ordentlichen Verarbeitung bildet die *response generation*.

Im Gegensatz zur Dialoghistorie werden die Logdaten von den meisten Komponenten dezentral generiert, so dass sie der Übersichtlichkeit halber nicht jeder einzelnen Komponente zugeordnet werden. Der Update Manager, der bei räumlicher Verteilung auch Teil des zentralen Systems ist, stellt die Verbindung zwischen den einzelnen Datencontainern und den entsprechenden Editoren her bzw. speichert Daten bis zur Synchronisation. Die beschriebenen Zusammenhänge sind in Abbildung 5-5 dargestellt. Da sowohl der Input Manager als auch der Präsentationsmanager Teil des Reaktionsermittlungssystems sein können, sind diese auch in der Darstellung enthalten. Die Datencontainer der ordentlichen Verarbeitung sind aus Gründen der grafischen Übersichtlichkeit zu einer logischen Einheit zusammengestellt worden. Dennoch können alle drei Bereiche der ordentlichen Verarbeitung auf alle vier zusammengefassten Datencontainer zugreifen.

5.1.2.5. Übersicht der Referenzarchitektur

Die gesamte Referenzarchitektur, die sich aus den vier genannten Systemen zusammensetzt, ist in Abbildung 5-6 dargestellt. Dabei sind die Datencontainer weitestgehend konsolidiert dargestellt, so dass gleiche Datenbestände redundanzfrei an einer Stelle vorgehalten werden. Ausnahmen dabei stellen die Sprachgrammatik im Input und Output System sowie die Umfeldinformationen im Input und Reaktionsermittlungssystem dar.

Die Verbindungen des Update Managers zu den einzelnen Datencontainern sind aus Gründen der Übersicht nicht dargestellt. Lediglich die Anbindung an die Animationsdatenbank ist visualisiert, da der Datencontainer nicht im Reaktionsermittlungs- sondern im Output System angesiedelt ist.

Die dargestellte Referenzarchitektur gibt somit die maximal möglichen Komponenten eines Avatarsystems an. Im Rahmen einer konkreten Umsetzung ist die Komplexität eines solchen Systems erwartungsgemäß deutlich geringer.

Software bedingt sichergestellt werden. Dies impliziert, dass keine weiteren Komponenten durch den Benutzer erweitert werden können.

- Ein Fahrzeug bietet aus Sicht des Herstellers die Möglichkeit, zusätzliche Hardware aufzunehmen und ist dadurch nicht so eingeschränkt wie ein mobiles Gerät, wie z. B. ein PDA (Personal Digital Assistant). Dadurch kann in die werksseitige Konfiguration des Avatarsystems prinzipiell eine breite Auswahl an Komponenten eingeplant werden.

Während sich die letzten beiden Punkte auf die Ausgestaltung der Architektur beziehen, eröffnet der erste Punkt ein neues Themenfeld: Die Verteilung der Komponenten bzw. Systeme eines Avatarsystems auf verschiedene Hardware. Im Folgenden soll deshalb nach einer generellen Betrachtung von Verteilungsmöglichkeiten, der Architektur und deren Verteilung für ein Avatarsystem im Fahrzeug festgelegt werden. Dies soll als Bauplan für die technische Implementierung des Systems in Kapitel 6.1 dienen.

5.1.3.1. Verteilung der Systeme und Komponenten

Einen systematischen Ansatz zur Analyse von möglichen Verteilungsformen zwischen verschiedenen Lokalitäten bzw. Hardwarestrukturen geben Krcmar und Strasburger (1993, 26). Im Rahmen der Verteilung von Client-Server-Architekturen stellen sie verschiedene Ausprägungen gegenüber und beschreiben ihre Eigenschaften und Anwendungsbereiche. Ebenso sollen an dieser Stelle auch die Systeme für Input, Output, Reaktionsermittlung und Design aus dem Blickwinkel der Mobilität auf die beiden Lokalitäten Client- und Server verteilt werden. Da sowohl In- als auch Output einen direkten Bezug zum Benutzer haben und somit analog Krcmar und Strasburger (1993) als Frontend bezeichnet werden können, sollen diese beiden Systeme auf der gleichen Ebene angesiedelt werden. Die systematische Anordnung des gesamten Avatarsystems auf der Grenze zwischen einem Server- und einem Client-Bereich ergibt die sieben in Abbildung 5-7 dargestellten Ausprägungen.

Die vollständige Umsetzung auf einem Server entspricht einem kontrollierten Laborsystem, wie es beispielsweise bei Cassell et al. (1998) vorgestellt wird. Eine Verschiebung des gesamten Avatarsystems zum Client-Bereich hin führt zur Verteilung der Interaktion. Diese Trennung erfolgt meist zwischen den Vorstufen und den eigentlichen Ein- und Ausgabekanälen. Ein Beispiel dafür ist die serverseitige Echtzeitgenerierung von Sprache, die auf Clientseite ausgegeben wird. In Fällen, in denen keine Vorverarbeitung der Ein- oder Ausgabeströme auf Seiten des Servers erforderlich ist, handelt es sich um eine rein clientseitige Abbildung der Interaktionssysteme, was bei den meisten internetbasierten oder mobilen Avatarsystemen der Fall ist (vgl. dazu Braun et al. 2003; Cavalluzzi et al. 2004).

Die Aufteilung der Reaktionsermittlung in der vierten Ausprägung führt zu einer kooperativen Verarbeitung. Dies kann der Fall sein, wenn integrale Bestandteile des zentralen Systems wie z.B. die Knowledge oder Behavior Base auf dem Server liegen, während die eigentliche Verarbeitung auf dem Client erfolgt. Die Umsetzung von Datencontainern auf dem Server kann aus Gründen der Aktualisierung sinnvoll sein.

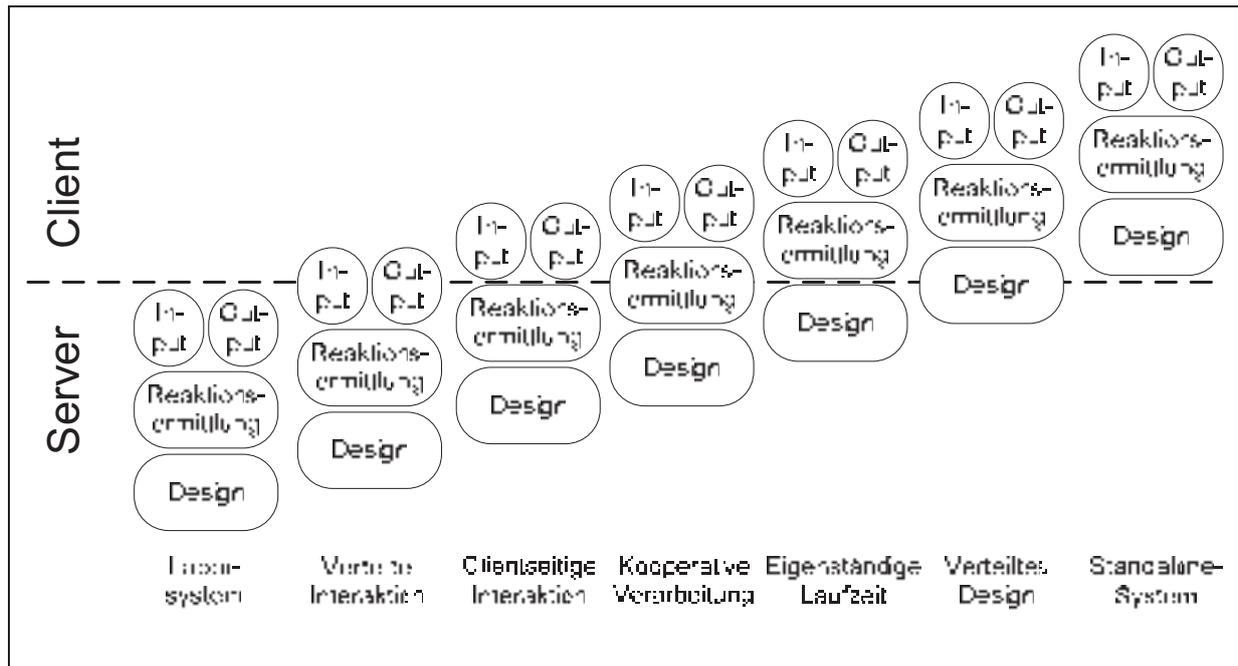


Abbildung 5-7: Verteilungsmöglichkeiten für Avatarsysteme

Quelle: (Eigene Darstellung, in Anlehnung an (Krcmar/Strasburger 1993, 26))

Die Ausprägungen 2 bis 4 setzen eine konstante Verbindung zwischen Client und Server zur Laufzeit voraus. Dies unterscheidet sie von den folgenden beiden Ausprägungen 5 und 6. Die Umsetzung der kompletten Laufzeitkomponenten auf Client-Seite erlaubt es, die Verbindung zwischen Client und Server nur sporadisch aufzubauen. Änderungen durch die Editoren der Designzeit werden im Upgrade Manager bis zum nächsten Verbindungsaufbau vorgehalten. Dieses Vorgehen gilt in Teilen auch für die sechste Ausprägung, in der das Design sowohl auf dem Client als auch auf dem Server erfolgen kann (vgl. Glass/Seneff 2003, 19f). Dabei kann von beiden Bereichen auf die gleichen Editoren und Daten zugegriffen werden oder es erfolgt eine Trennung nach Datenarten. Beispielsweise könnten die Animationsdaten vom Server gepflegt werden, während auf Clientseite neue Regeln erstellt werden. Die letzte Ausprägung schließlich, sieht die vollständige Umsetzung auf dem Client vor. Dabei spielt eine Verbindung zwischen den beiden Bereichen ebenso wie beim Laborsystem keine Rolle. Von Seiten des Servers kann kein Einfluss mehr auf das System genommen.

Von den vorgestellten Verteilungsmöglichkeiten kommt für den Einsatz im Auto derzeit lediglich die fünfte Ausprägung für eine eigenständige Laufzeit in Frage, wobei das Fahrzeug als Client und eine stationäre Hardware als Server angenommen wird. Die fehlende Gewährleistung einer konstanten Verbindung zwischen Fahrzeug und Server während der Fahrt und der gleichzeitige Anspruch, das Avatarsystem während der Fahrt bedienen zu können, erfordern eine Verteilung, welche die gesamte Laufzeit im Auto umsetzt und eine asynchronen Abgleich mit dem Update Manager ermöglicht. Diese Anforderungen werden nur durch die Ausprägungen 5 bis 7 erfüllt. Ein verteiltes Design macht beim Einsatz im Fahrzeug derzeit keinen Sinn, da in einem Fahrzeug nicht die notwendigen Interaktionsgeräte wie Maus, Tastatur und hochauflösender Bildschirm zur Verfügung stehen. Diese Geräte sind jedoch nötig, um Inhalte des Systems angemessen pflegen zu können. Die optimale Aufteilung der Komponenten sieht somit die Ansiedlung des Designsystems auf einer

zentralen Hardware vor, während alle anderen Komponenten im Fahrzeug vorgehalten werden.

5.1.3.2. Konkrete Systemarchitektur des Forschungsvorhabens

Unter Berücksichtigung derzeit verfügbarer Technologien und der erhöhten Zuverlässigkeit regelbasierter Ansätze im Rahmen des Response Planning ergibt sich für die Umsetzung eines Avatarsystems im Fahrzeug die in Abbildung 5-8 gezeigte Architektur. Das Input, Reaktionsermittlung- und Output-System sind im Client-Bereich, also im Fahrzeug, untergebracht während das Designsystem den Server-Bereich darstellt.

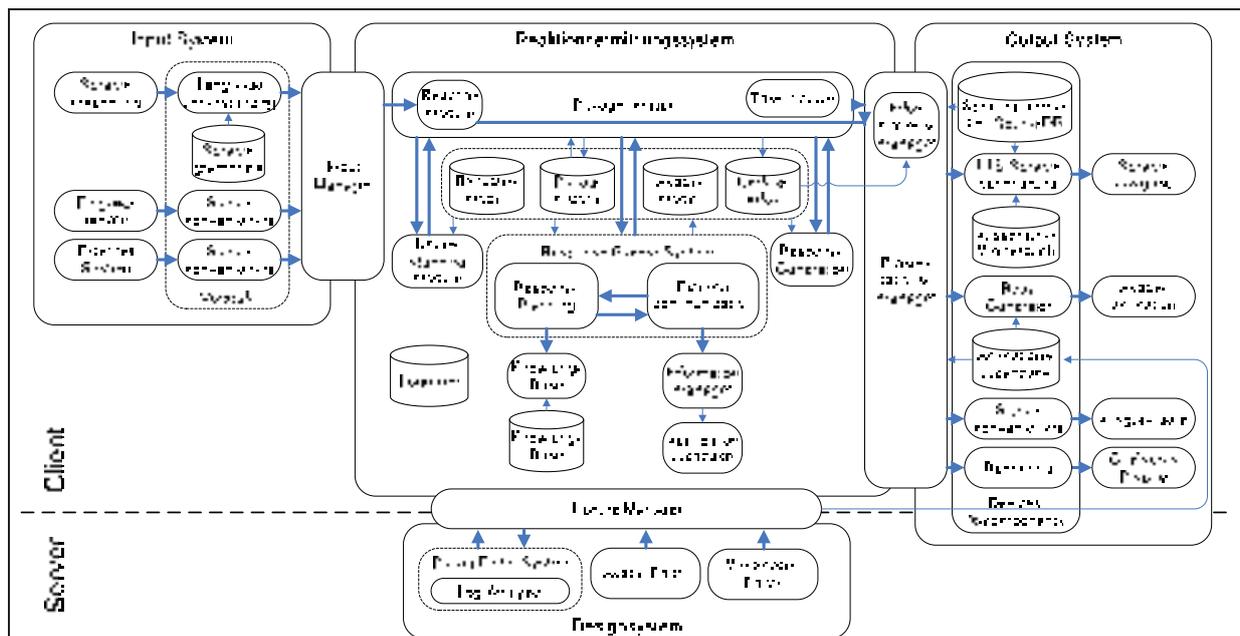


Abbildung 5-8: Architektur für Avatarsysteme im Fahrzeug
Quelle: (Eigene Darstellung)

Im Gegensatz zur Referenzarchitektur fallen im Input System zahlreiche Komponenten weg, da sie derzeit nicht in Fahrzeugen verfügbar sind oder wie im Falle von Mimik- und Gestenerkennung keinen Beitrag zur Bedienbarkeit des Systems im Fahrzeug leisten. Somit bleiben die Spracherkennung, die Betätigung der bestehenden haptischen Elemente im Fahrzeug und das Auslösen von Ereignissen durch Bordsysteme. Da keine verteilte Verarbeitung der Eingabe erfolgt, fällt auch die Komponente des Multiplexers weg.

Das Reaktionsermittlungssystem beinhaltet alle Komponenten, die zur regelbasierten Verarbeitung geeignet sind. Dadurch fallen der semantische Rahmen und entsprechende Ansätze des *response planning*s weg. Aufgrund der sporadischen Verbindung zu externen Systemen ist die Anbindung eines externen Systems, das auf dem Server liegt, kaum möglich. Um jedoch Regeln durch aktuelle Daten des Fahrzeugs zu vervollständigen, kann die Anbindung zu Bordinformationssystemen hergestellt werden. Durch den Wegfall eines agenten-basierten Ansatzes erübrigt sich auch die Pflege des Agentenzielsystems im Designsystem. Im Bereich der Ausgabe lassen sich unter Nutzung eingebauter Displays für Navigations- oder Bordbediensysteme alle möglichen Modalitäten umsetzen. Der Einsatz von

Vibration und Rückstellkraft, der besonders im Fahrzeug eine geeignete Erweiterung zu sein scheint (vgl. Akyol/Libuda/Kraiss 2001), ist mit serienmäßig eingebauten Geräten nicht umsetzbar. Durch die Ansteuerung von Ausgabegeräten lassen sich prinzipiell auch einfache Funktionen des tertiären Bereichs durch das Avatarsystem steuern (vgl. dazu Geiser 1985; Tönnis/Broy/Klinker 2006, 128f).

Die auf diese Weise abgesteckte Architektur für ein Avatarsystem bleibt hinter den Forderungen von Akyol, Libuda und Kraiss (2001) für ein Kfz-Informationssystem zurück. Angesichts des Umstandes, dass bisher noch keinerlei praktische Erfahrungen bezüglich der Umsetzung von Avatarsystemen im Fahrzeug veröffentlicht wurden, wird diese Architektur zunächst die Basis für eine genauere Untersuchung darstellen.

5.2. Inhaltliche Konzeption

Der technische Teil eines Avatarsystems stellt lediglich die Grundlage für die Erstellung und Präsentation von Inhalten dar und ist somit alleine nicht verwendbar. Churchill et al. (2000, 70ff) unterscheiden neben der technischen Umgebung, die sie in das *digital environment* einordnen, die drei gestalterischen Bereiche der Kommunikation bzw. des Domänenwissens (communication and domain expertise), der Persönlichkeit (personality) und des Aussehens (appearance) eines Avatarsystems (vgl. Abbildung 5-9). Während die Kommunikation und das Domänenwissen in diesem Kapitel betrachtet werden, erfolgt das Design des Aussehens in Kapitel 5.3 im Rahmen der Benutzerschnittstelle. Der Aspekt der Persönlichkeit, der keine kritische Komponente für den Betrieb eines Avatarsystems darstellt, wird im Rahmen des Kapitels 6.2.3.6 betrachtet.

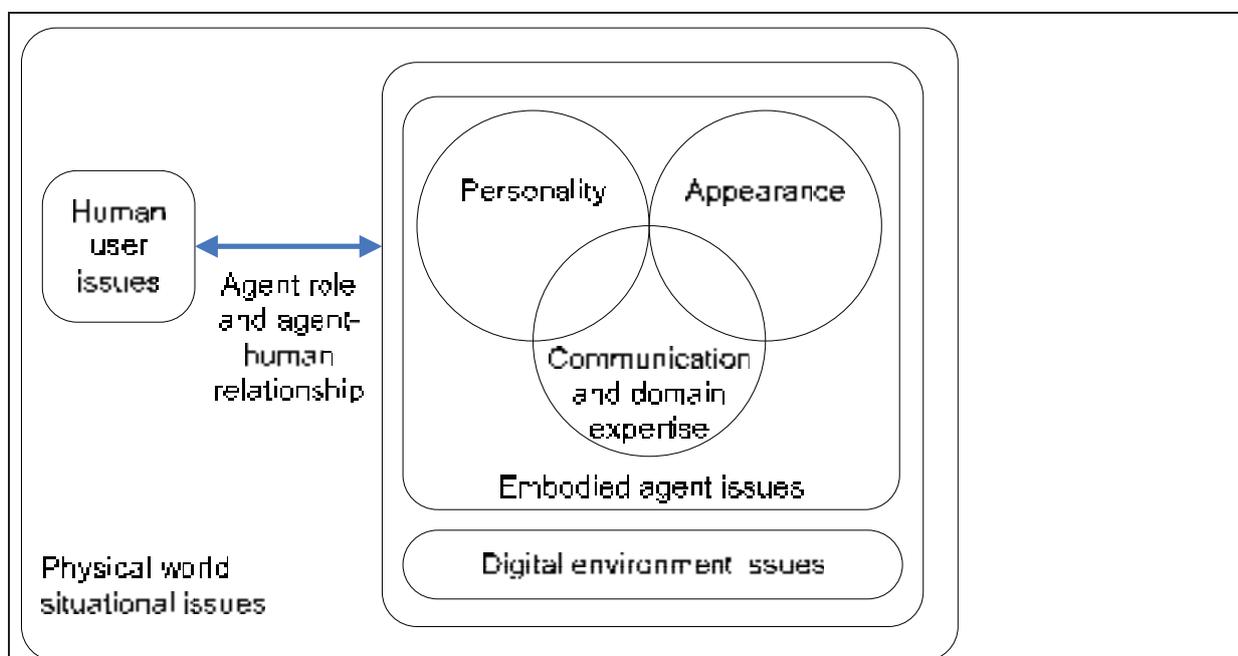


Abbildung 5-9: Designherausforderungen eines Embodied Conversational Agents (ECA)
Quelle: (Churchill et al. 2000, 70)

Die inhaltliche Gestaltung des Domänenwissens und der Kommunikation lässt sich entlang ihres Lebenszyklus in drei relevante Phasen unterteilen: Erstellung, Betrieb, Änderung. Während der Erstellung werden Inhalte zusammengetragen und miteinander verbunden. Während des Betriebs bestimmt insbesondere die Struktur der Inhalte die Möglichkeiten der Kommunikation. Änderungen führen zu Anpassungen und Erweiterungen des bestehenden Inhalts. Da alle Phasen von der Struktur und den damit vorgegebenen Möglichkeiten beeinflusst werden, soll zunächst der Aufbau und daraus abgeleitet die Funktionen konzipiert werden. Daran schließt sich die Betrachtung der Erstellung und Pflege an.

5.2.1. Gestaltung der Kommunikation und des Domänenwissens

Die Struktur der Dialoginhalte, im Falle eines regelbasierten Ansatzes also der Regeln, lässt sich aus drei verschiedenen Bereichen betrachten. Zum einen beschreiben der strukturelle Aufbau und die Organisation, wie Regeln gespeichert, aufgebaut und untereinander angeordnet sind. Die funktionelle Sicht untersucht die in den Regeln enthaltenen Funktionen zur Steuerung verschiedener Ausgabemodalitäten oder zur Abfrage fehlender Informationen im Rahmen einer externen Kommunikation. Der dritte Blickwinkel beleuchtet schließlich die inhaltlichen Bereiche, die durch die Regeln beim Einsatz für ein Avatarsystem im Fahrzeug abzudecken sind.

5.2.1.1. Struktureller Aufbau und Organisation der Inhalte

Die technische Organisation der verwendeten Regeln kann je nach Ausstattung der entsprechenden Komponente in Form einer Datenbank oder wie in Kapitel 3.3.3 beschrieben durch XML-Schematas erfolgen. Unabhängig von der Art der Speicherung bestehen Regeln aus einem Reaktionsteil und einem Eingabe- oder Identifikationsteil. Der Eingabeteil ermöglicht das Auffinden einer relevanten Regel, wenn eine bestimmte Eingabe erfolgt. Dabei kann es sinnvoll sein, Regeln nur für Eingaben eines bestimmten Modus zu aktivieren. So sollen bestimmte Regeln beispielsweise nur durch Ereignisse des angebundenen Bordsystems und nicht durch die Spracherkennung ausgelöst werden können. Dabei kann die Spracherkennung einfach von Eingaben anderer Modalitäten separiert werden, wenn eine gesonderte Sprachgrammatik verwendet wird. Eingaben, die nur über andere Eingabekanäle erfolgen können, werden dementsprechend nicht in die Sprachgrammatik aufgenommen. Im Gegenzug müssen alle möglichen Eingaben, die über die Spracherkennung erkannt werden sollen, in die Sprachgrammatik aufgenommen werden.

Ebenso wie auf der Eingabeseite muss auch auf der Ausgabeseite den verschiedenen Ausgangskanälen Rechnung getragen werden. Auch wenn die finale Aufarbeitung in der Komponente der *response generation* erfolgt, befinden sich im Reaktionsteil einer Regel bereits die meisten Steuerinformationen für die Ausgabe. So können neben anzuzeigenden oder zu sprechendem Text auch die Steuerung des Avatars oder von Ausgabegeräten enthalten sein (vgl. dazu auch Bengler 1995). Eine systematische Betrachtung der möglichen Funktionen erfolgt im folgenden Kapitel 5.2.1.2.

Die Auswahl passender Regeln aufgrund einer Eingabe erfolgt auch unter Verwendung der Dialoghistorie. Dies muss zum einen bei der Erstellung der Regeln bereits berücksichtigt

werden und zum anderen müssen verschiedene Regelarten unterschieden werden. So gibt es Regeln, die Inhalte transportieren oder zur Annäherung an diese dienen und es gibt rein steuernde Regeln, die beispielsweise ins Spiel kommen, wenn ein Eingabefehler durch Nachfrage korrigiert werden muss. Im Rahmen der Regelverarbeitung muss somit zwischen verschiedenen Gesprächssträngen unterschieden werden.

Der Reaktionsteil einer Regel muss bei Verwendung eines Benutzermodells diesem Rechnung tragen und Eigenschaften oder explizit gewählte Optionen eines Benutzers berücksichtigen. Dies kann insbesondere notwendig werden, wenn der Benutzer generell ein bestimmtes Niveau auszugebender Informationen angeben oder konkret eine bestimmte Art von Hinweisen von der zukünftigen Anzeige ausgeschlossen hat. Auch kann dadurch automatisch vermerkt werden, welche Inhalte dem Fahrer bereits vorgetragen wurden. Zudem können Informationen über die Vorkenntnisse des Benutzers in anderen Domänen genutzt werden, um den Detailreichtum einer Reaktion zu steuern (vgl. dazu auch Salmen 2002, 48). Ebenso müssen sowohl die statischen als auch dynamischen Umfeldinformationen bei der Gestaltung der Ausgabe berücksichtigt werden. Als statisches Merkmal kann beispielsweise die Ausstattung eines Fahrzeugs angesehen werden, wodurch die Erklärung nicht vorhandener Funktionen obsolet wird.

Die inhaltliche Gestaltung des Reaktionsteils selbst muss sich gemäß den Anforderungen zwei Maximen unterwerfen: Der Kürze und der Verständlichkeit. Die Kürze einer Reaktion ergibt sich zum einen daraus, dass die Kapazität der Informationschunks, die sich ein Mensch bei textueller Anzeige merken kann bei 7 ± 2 liegt. Dabei lässt sich ein Informationschunk nicht genau definieren: Er kann ein oder mehrere Wörter bzw. Bilder beinhalten (Miller 1956, 92f). Im Falle von Informationen, die nicht per Text, sondern akustisch vermittelt werden, umfasst die Anzahl an Chunks nur 5 ± 2 (Salmen 2002, 37). Die Maxime der Kürze erstreckt sich jedoch nicht nur auf Informationen, die als Text oder Sprache ausgegeben werden, sondern auch auf visuelle Darstellungen. Insbesondere während des Fahrens eines Autos sollte eine Anzeige dermaßen gestaltet sein, dass sie die maximal akzeptable Blickspanne von 1 bis 2 Sekunden nicht übersteigt bis der Benutzer ihren Sinn erschlossen hat (Kimura/Marunaka/Sugira 1997). Die durchschnittliche Blickspanne bei der Verwendung von fahrzeuginternen Systemen liegt zwischen 0,6 und 1,7 Sekunden (Wierwille et al. 1991; Wikman/Nieminen/Summala 1998). Bei der Umsetzung dieser Anforderung muss auch berücksichtigt werden, dass je nach Kenntnisstand oder Einstellungen des Benutzer bzw. des Umfeldes die Länge der Ausgabe variieren kann.

Die Verständlichkeit bezieht sich insbesondere auf den Sprachgebrauch des Benutzers (Heidmann/Biesterfeldt 2004, 63). So müssen Begriffe gebraucht werden, die im Alltag vorkommen und nicht technischen Fachausdrücken entsprechen (vgl. Salmen 2002, 38). Anstelle des Begriffs „Sicherungsstift“ kann von einem Türknopf gesprochen werden bzw. wenn eine „Wählhebelstellung“ erwähnt wird, ist ein Gang gemeint. Da dennoch verschiedene Benutzer verschiedene Wortschätze verwenden, muss versucht werden, alternative Eingabeformulierungen für eine möglichst breite Wortwahl abzudecken. Die Variation der Wortwahl kann nicht nur auf der Eingabe-, sondern auch auf Ausgabeseite erfolgen, um monotone Antworten zu vermeiden. Die Erwartungen an den Benutzer in Folge einer bestimmten Avatarreaktion können zudem kommuniziert werden, wenn die Reaktionen

als Frage oder Anleitung gekennzeichnet sind. Darüber hinaus können intern auch Eingaben nach verschiedenen Typen eingeteilt werden (vgl. Eckert 1996, 105).

Die beschriebenen Ansprüche an den Aufbau der Inhalte richten sich dabei nach dem informationslogistischen Prinzip aus, nach dem die richtige Information zum richtigen Zeitpunkt in der richtigen Menge am richtigen Ort und in der erforderlichen Qualität vorliegen soll (Augustin 1990). Lindner (2003, 15ff) formuliert speziell für Avatarsysteme drei Maximen zur Gestaltung des Inhalts:

1. Die Maxime der Quantität besagt, dass ein Redebeitrag gehaltvoll, der Situation angemessen sein und nicht mehr Informationen als gewünscht enthalten soll.
2. Die Maxime der Qualität beinhaltet die Forderung danach, nicht zu vermitteln, was nicht zutreffend ist, sondern was für eine zutreffende Information gehalten wird. Darüber hinaus soll nur zu Sachverhalten geantwortet werden, über die genug Informationen vorliegen.
3. Die Maxime der Form soll zur Vermeidung von unklaren Formulierungen und Doppeldeutigkeiten auffordern.

5.2.1.2. Bereitstellung von Funktionen zum Dialogablauf

Zur optimalen Gestaltung der Dialogabläufe sollen alle zur Laufzeit sinnvollen Funktionen in den Regeln verwendbar sein. Diese umfassen neben inhaltlichen Funktionen zur korrekten Verarbeitung von Eingaben, der Ermittlung von Reaktionen und der Pflege des Benutzermodells auch Aktivitäten, die das Umfeld beeinflussen. Zunächst sollen alle zur Verarbeitung und Ausgabe notwendigen Funktionen aufgeführt werden, bevor die umfeldsteuernden Tätigkeiten betrachtet werden.

Insbesondere die korrekte Zuordnung und Verarbeitung einer Eingabe erfordert den impliziten oder expliziten Zugriff auf die Datencontainer des Systems. Dabei sind vor allem das Benutzermodell, die Dialoghistorie sowie die Umfeldinformationen von Bedeutung (McTear 2002, 114ff). Im Rahmen der Regeln müssen somit gezielt Werte ausgelesen werden und im späteren Verlauf der Verarbeitung auch geschrieben werden können. Da dem Benutzer die Möglichkeit gegeben werden soll, jederzeit das Thema zu wechseln, muss zum einen sicher gestellt sein, dass das Avatarsystem das aktuelle Thema kennt und es auf der Grundlage einer entsprechenden Benutzereingabe ändern kann. Das Thema stellt neben der Dialoghistorie den zweiten Eckpfeiler der korrekten Interpretation einer Eingabe dar.

Da viele verschiedene Formulierungsmöglichkeiten zur gleichen Reaktion führen, müssen mehrere Eingaben derart zusammengefasst werden können, dass sie zur einfacheren Pflege zu einer einzigen Reaktion führen. Im Rahmen dieser Reaktion kann der auszugebende Teil wiederum variiert werden. Dafür muss entweder die Möglichkeit bestehen, die Eingabe wieder aufzugreifen oder per Zufall eine Variation zu bestimmen.

Zur Gewährleistung eines robusten Dialogs sind zwei Schritte und damit verbundene Funktionen notwendig: Zum einen das Erkennen eines Fehlers und zum anderen dessen Behandlung. Eckert (1996, 91) sieht vier mögliche Gründe für eine nicht korrekt verarbeitbare Eingabe im Rahmen eines sprachgestützten Dialogsystems:

1. Es wurden nicht alle gesprochenen Worte des Benutzers erkannt.
2. Dem Benutzer ist ein Fehler in der Wortwahl oder der Grammatik unterlaufen oder er hat sich versprochen.
3. Die gesprochene Phrase enthält nicht alle notwendigen Informationen, die für ihre Verarbeitung notwendig sind. Dies kann beispielsweise Folge des Zusammenziehens von Wörtern beim Sprechen sein, wie es in regionalen Dialekten häufig der Fall ist.
4. In der Regelbasis ist eine entsprechende Eingabe nicht vorgesehen.

Die Erkennung eines Fehlers kann dabei dermaßen erfolgen, dass zu einer Eingabe kein passender Regeleintrag besteht und aus diesem Grund im Rahmen einer Rückfallstrategie eine allgemeingültige Auffangregel aktiviert wird (vgl. Alexandersson/Maier/Reithinger 1994). Im Rahmen dieser Auffangregel kann versucht werden, evtl. fehlende Angaben aus Datencontainern des Systems aufzufüllen oder den Benutzer auf die nicht gegebene Verarbeitbarkeit hinzuweisen (McTear 2002, 114; Zue/Glass 2000, 1170). Bei der Verwendung einer Spracherkennung, die auf einer angepassten Sprachgrammatik basiert, verringert sich jedoch die Wahrscheinlichkeit, dass eine fehlerhafte Eingabe bis zum *understanding module* gelangt. Der Spracherkenner ermittelt den wahrscheinlichsten Treffer aus der gegebenen Sprachgrammatik, so dass auch bei kleineren Versprechern seitens des Benutzers dennoch die korrekte Eingabe erfolgen kann.

Die Verarbeitung von Daten im Rahmen des Regelwerks kann erst dann optimal genutzt werden, wenn auf Basis eines bestimmten Wertes eine Entscheidung getroffen werden kann. Diese Entscheidung kann beispielsweise im Rahmen einer Verzweigung im Regelwerk erfolgen.

Im Rahmen der Reaktion müssen zudem Funktionen zur Anzeige und Steuerung von Avatar und Umwelt realisierbar sein. Diese können die Steuerung der Sichtbarkeit, der Position und des Bewegungsablaufs beinhalten. Zudem kann auch die Anzeige von 2D-Objekten wie Bilder oder Videos notwendig werden. Neben der Steuerung der grafischen Komponente muss auch die Sprachausgabe individuell anpassbar sein. Dies beinhaltet neben der Aussprache von Wörtern auch die Möglichkeit, die prosodischen Aspekte einer Phrase zu beeinflussen. Die Aussprache kann dabei individuell in einzelnen Reaktionen oder durch die Verwendung eines Aussprachewörterbuchs optimiert werden.

Neben den Funktionen, die zur Verarbeitung und Ausgabe der Reaktion direkt verwendet werden, muss zudem die Möglichkeit bestehen, weitere Funktionen umzusetzen, die den Dialogfluss oder das Umfeld steuern. Eine solche Hilfsfunktion stellt beispielsweise die Reversibilität von Aktionen dar. Dabei müssen die Parameter der letzten Aktion vermerkt

werden, um diese Umkehren zu können. Darüber hinaus muss die Anzeige von Avatar und Umwelt insofern steuerbar sein, dass beispielsweise der Blickwinkel und die Position der einzelnen Anzeigeobjekte verändert werden kann. Zudem müssen sowohl der Avatar als auch sein Umfeld in ihrer Darstellungsart angepasst werden können. Dies kann z.B. derart erfolgen, dass die entsprechende Anzeige in einer bestimmten Situation ganz abgeschaltet wird. Alternativ dazu sind auch Formen denkbar, bei denen statt einer Animation nur unbewegte Bilder angezeigt werden, um den Benutzer nicht während der Fahrt abzulenken. Eine weitere sinnvolle Funktion ist die Verwendung einer Timeout-Regelung. Damit können nach einer vorgegebenen Zeit Benutzereingaben durch das System simuliert werden. Dies kann bei der Anzeige einer zusammengehörigen Informationssequenz dem Benutzer zusätzliche Eingaben ersparen. Die beschriebenen Funktionen sind als Übersicht in Tabelle 5-8 zu sehen.

Funktionsbereich	Funktion
Eingabe	Lesen und Schreiben von Werten aus Datencontainern des Systems
	Kenntnis über und Wechselmöglichkeit des Themas auf Benutzereingabe hin
	Zusammenfassung von Eingabemöglichkeit zu einer Reaktion
Verarbeitung	Variation der Reaktion auf Basis von Eingabe oder Zufallsmechanismus
	Umsetzung einer Auffangregel als Rückfallstrategie
	Rücksprache mit Benutzer/ Ergänzung fehlender Werte durch Zugriff auf Datencontainer in Auffangregel
	Umsetzung von Verzweigungen im Regelwerk
Reaktion	Steuerung von 3D-Objekten in der Anzeige
	Verwendung von 2D-Objekten
	Phonetische und prosodische Optimierung der Sprachausgabe
Hilfsfunktionen	Reversibilität einer Aktion
	Steuerung von Blickwinkel und Position von Objekten in der Anzeige
	Ein- und Ausblenden/ Variation der Darstellungsart von Anzeigeelementen
	Timeout-Funktion zur zeitgesteuerten Eingabe

Tabelle 5-8: *Funktionen zur Dialog- und Umfeldsteuerung*
Quelle: (Eigene Darstellung)

5.2.1.3. Abdeckung inhaltlicher Bereiche

Bühler (2003, 116ff) formuliert die Forderung an ein Avatarsystem, schön, schnell und schlau zu sein. Mit dem Aspekt der Schönheit verbindet er die Kommunikation einer klar abgesteckten Aufgabe, die auch durch Reaktionen auf Ereignisse erbracht werden soll. Darüber hinaus fordert er u.a. eine Abstimmung von Auftreten und Inhalt, sowie eine hohe Performance. Insbesondere die klare Abgrenzung erfordert die Identifikation möglicher Betätigungsfelder eines Avatarsystems im Auto und die Fokussierung auf besonders relevante Bereiche im Rahmen dieser Forschungsarbeit.

Die Abgrenzung der Inhalte und vor allem die Kommunikation dieser Grenzen steuern die Erwartungshaltung und damit in gewissem Rahmen auch die Akzeptanz des Fahrers. Als

grober Anhaltspunkt zur Einteilung von Inhalten kann dabei das gedruckte Handbuch eines Fahrzeugs dienen, wobei verschiedene interaktive Erweiterungen erst durch den Einsatz eines Avatarsystems möglich werden und somit die herkömmlichen Inhalte erweitern. Da das Avatarsystem im Fahrzeuginnenraum angesiedelt ist, eignen sich nur Funktionen und Bedienelemente im Fahrzeug für Erklärungen, die vom Innenraum aus erreicht werden können.

Die möglichen Bereiche lassen sich anhand vieler Kriterien auswählen und begründen. Einen Überblick über mögliche Inhaltsbereiche und deren Anordnung anhand der Dimensionen Tiefe und Breite lassen sich aus Abbildung 5-10 ablesen.

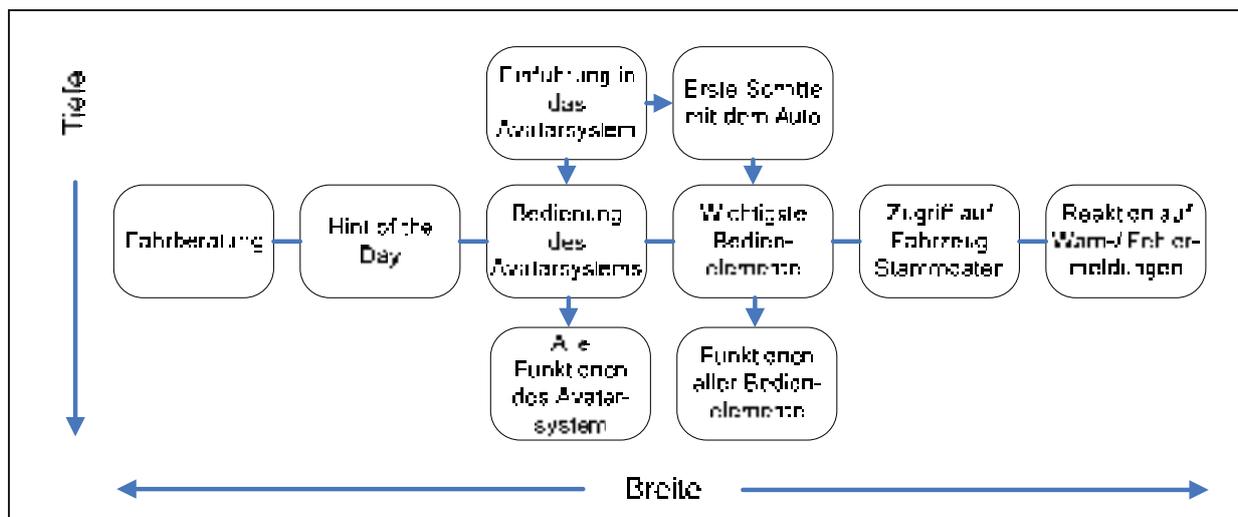


Abbildung 5-10: Tiefe und Breite der inhaltlichen Bereiche
Quelle: (Eigene Darstellung)

Da von den Benutzern keine Erfahrung im Umgang mit Avatarsystemen im Fahrzeug vorausgesetzt werden kann, stellt die Einführung in das Ziel und die Aufgaben des Systems den ersten Schritt dar. Ebenso wie die Beschreibung der ersten Schritte mit dem (neuen) Auto ist dieser Bereich als einführend einzustufen und befindet sich somit auf der obersten Stufe der Inhalte. Die mittlere Stufe der Inhalte hinsichtlich der Tiefe umfasst verschiedene Funktionen, welche die beiden Bereiche der obersten Stufe ergänzen können. Zum einen sind dies die Bedienung der grundlegenden Funktionen des Avatarsystems, sowie die Beschreibung der wichtigsten Bedienelemente. Aufbauend auf diesen Bereichen kann ein täglicher Hinweis zur Nutzung des Fahrzeugs auf Initiative des Avatarsystems hin als *Hint of the day* angeboten werden. Zudem kann der Benutzer Zugriff auf die Informationen zum spezifischen Fahrzeug nehmen und sich beispielsweise die Leistungsmerkmale erklären lassen. Die Stammdaten des Fahrzeugs sind dabei bereits Teil des Datencontainers mit statischen Umweltinformationen.

Während die bisher beschriebenen Bereiche mit Ausnahme des *Hint of the day* auf Initiative des Fahrers hin verwendet werden, stellen die beiden äußersten Bereiche Inhalte bereit, die durch äußere Ereignisse aufgerufen werden. Das an das Avatarsystem angeschlossene Bordinformationssystem des Fahrzeugs kann Warn- oder Fehlermeldungen weiterleiten, so

dass das Avatarsystem mit entsprechenden Hinweisen dem Fahrer zur Seite stehen kann. Diese Inhalte sind Teil des herkömmlichen Handbuchs, jedoch für die Beratung des Fahrers in kritischen Situationen gesondert aufgearbeitet.

Ein weiterer Inhaltsbereich, der nicht nur auf einem einzelnen Ereignis im Bordsystem reagiert, ist die Fahrberatung des Fahrers hinsichtlich eines bestimmten Kriteriums. Bei der Beobachtung eines gegenteiligen Verhaltens durch den Fahrer ist es dadurch möglich, beispielsweise Hinweise auf umweltschonendes, sicherheitsbewusstes oder komfortsteigerndes Fahren zu geben. Ebenso lässt sich diese Beratung auf Funktionen erweitern, die durch den Benutzer i.d.R. nicht verwendet werden, jedoch die Benutzung des Fahrzeugs erleichtern oder angenehmer machen.

Auf der dritten Ebene sind schließlich Bereiche angeordnet, die alle Details zu bestimmten Funktionen und Bedienelementen beinhalten. Während die bei der Erklärung der Detailfunktionen des Avatarsystems der gesamte, überschaubare Funktionsumfang die Menge des Inhalts bestimmt, erfolgt bei den Funktionen des Fahrzeugs eine Auswahl. Diese Auswahl im Rahmen des Prototyps ermöglicht die Fokussierung und tiefgehende Aufarbeitung bestimmter Inhalte. Bei der Auswahl dieser Inhalte bietet es sich an, Szenarien zu bestimmen, die im späteren Versuchsträger eine optimale Evaluation ermöglichen. Dies setzt zum einen voraus, dass die beschriebenen Funktionen als Ausstattungsmerkmale im späteren Versuchsfahrzeug enthalten sind. Zum anderen sollte sowohl ein Szenario bestimmt werden, das während der Fahrt funktioniert, als auch eines, das im Stehen des Fahrzeugs vorgestellt werden kann. Optimal ist in diesem Zusammenhang ein Szenario, das sowohl im Stehen als auch im Fahren gleich abläuft, um vergleichbare Ergebnisse zu erzielen.

Unabhängig von der Wahl der Szenarien ist insbesondere in diesen tiefgehenden Inhaltsbereichen die Umsetzung einer aktiven Gesprächsführung wichtig. Benutzer, die keine Kenntnisse über die zu erklärenden Funktionen des Fahrzeugs haben, benötigen dementsprechend Anhaltspunkte, welche Aspekte eine Funktion aufweist bzw. wie danach zu fragen ist (vgl. Chai et al. 2001, 286). Dies erfordert, dass das Avatarsystem den Benutzer im übertragenden Sinne an die Hand nimmt und durch die Erklärung der Funktionen hindurch führt. Dieses Konzept soll dabei nicht nur innerhalb eines Szenarios, sondern auch über alle Inhaltsbereiche hinweg umgesetzt werden, so dass mit der Führung des Avatarsystems alle Inhalte vorgetragen werden, die durch den Benutzer aktiv angewählt werden können. Sobald der Benutzer einmal weiß, dass es diese Inhalte überhaupt gibt, wird es ihm möglich sein, später gezielt danach zu fragen. Da diese Führung durch alle Inhalte eine umfangreiche Zeitspanne in Anspruch nehmen wird, muss dem Benutzer ermöglicht werden, die Führung jederzeit abubrechen und zu einem späteren Zeitpunkt fortzuführen.

Die dargestellten Bereiche stellen keinen Anspruch auf Vollständigkeit, sind jedoch die naheliegendsten Anwendungsgebiete für ein Avatarsystem im Fahrzeug und mit Fahrzeugbezug. Es sind weitere Inhalte denkbar, etwa ein interaktiver Fitness-Guide oder die Bedienung prototypischer Funktionen. Durch die Erweiterung der Inhalte und der entsprechenden statischen Datencontainer lassen sich beliebige Inhalte unter Verwendung des Avatarsystems darstellen und bedienen.

5.2.2. Inhaltserstellung und –pflege

Nachdem der Aufbau und die inhaltlichen Eckpunkte der Regelbasis betrachtet wurden, steht die Erstellung der konkreten Regeln und deren Pflege im Vordergrund. Dabei lassen sich die Frage nach dem Vorgehen zur Erstellung der Regelbasis sowie der Bedarf an Werkzeugen als Kernpunkte identifizieren.

5.2.2.1. Vorgehen zur Erstellung der Regelbasis

Der Prozess der Regelgenerierung, der zurzeit meist herstellergetrieben ist, unterliegt keinem standardisierten oder allgemein anerkanntem Vorgehen. Dies lässt sich einerseits mit den proprietären technischen Umgebungen, den verschiedenen Annotationen des Regelwerks oder mit verschiedenen Anforderungen erklären.

Ein grobes Vorgehensmodell speziell für die Entwicklung von Dialogen im Fahrzeug stellen Ahlers und Öl (2005) vor, wobei sich ihre Ausführungen nicht auf natürlichsprachliche Dialoge im Besonderen beziehen. Sie unterscheiden drei Schritte, die durch die Top Level Anforderungen an das Fahrzeug beeinflusst werden. Im ersten Schritt werden die resultierenden Anforderungen an die Dialoge abgeleitet. Darauf basierend werden im zweiten Schritt die Dialogabläufe entwickelt, was dem Aufbau einer Ablaufstruktur entspricht. Schließlich wird diese Ablaufstruktur im letzten Schritt, dem *wording*, mit konkreten Ein- und Ausgaben befüllt.

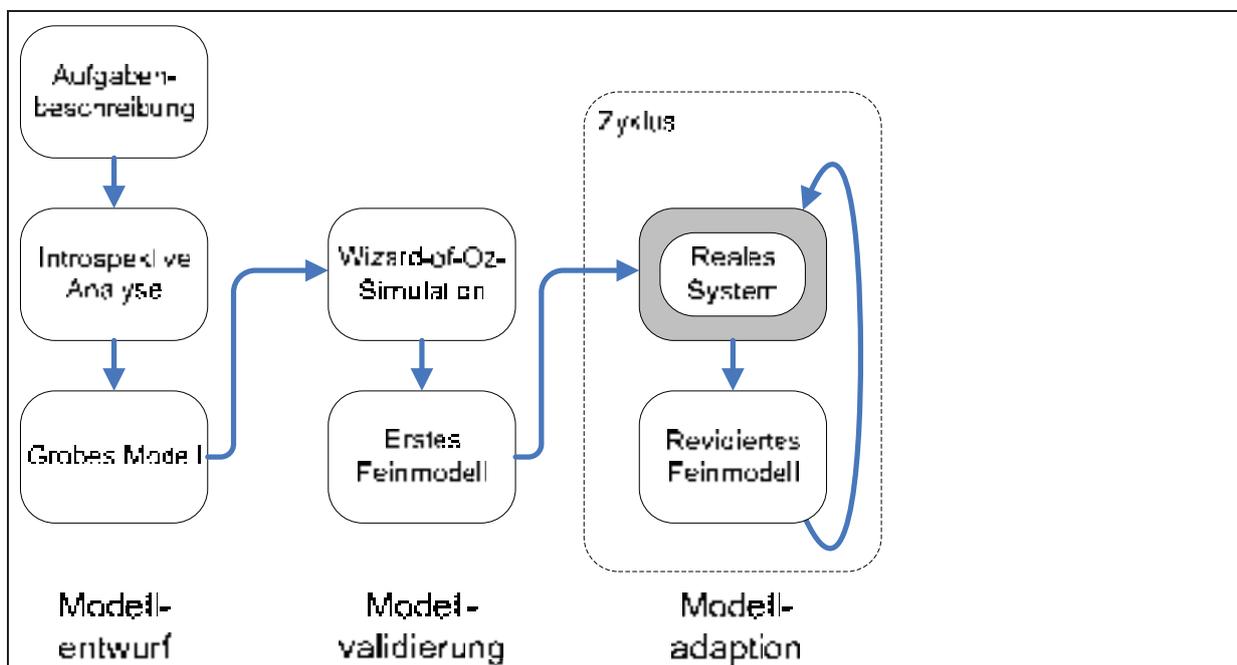


Abbildung 5-11: Verfeinerung des Inhaltsmodells
Quelle: (Eckert 1996, 93)

Eine differenziertere Betrachtung des Erstellungsprozesses nimmt Eckert (1996, 91ff) vor, der die oben beschriebenen Schritte teilweise aufgreift und in verschiedene Phasen einordnet (vgl. Abbildung 5-11). Dabei wird auch die Qualität der erstellten Regeln betrachtet. Der erste Schritt beginnt mit der Aufgabenstellung des neu zu entwickelnden Systems und stimmt

damit mit dem Modell von Ahlers und Öl überein. Im Rahmen einer introspektiven Analyse, d.h. einer Konzeption auf der Basis von Plausibilität unter Ausschluss von Benutzern, wird ein erstes grobes Modell erstellt. Dieses Modell enthält im Gegensatz zum ersten Vorgehensmodell auch bereits das *wording*, also die konkreten Ein- und Ausgaben. Die beschriebenen Schritte stellen die Phase des Modellentwurfs dar, die von der Modellvalidierung gefolgt wird.

Der erste Schritt der Modellvalidierung ist die Anwendung des groben Modells im Rahmen einer Wizard-of-Oz-Simulation. Diese Simulation basiert in Anlehnung an die Geschichte des Zauberers von Oz auf der PNAMBIC (Pay No Attention to that Man Behind the Curtain)-Idee, wie es Wallace (2002a) bezeichnet. Diese speziell im Rahmen von Dialogsystemen häufig verwendete Methode (vgl. Diaper/Shelton 1989, 102ff; Sanders/Scholtz 2000; Aimless 2003) versucht einer realen Testperson vorzugaukeln, dass sie sich nur mit einem Dialogsystem unterhält. Dabei ist bei diesem Test noch eine weitere reale Person beteiligt, die entweder die Antworten des Dialogsystems selbst bestimmt oder automatisch ermittelte Antworten vor ihrer Ausgabe kontrolliert. Dieses Verfahren ermöglicht es, unter Bedingungen, die für die Testperson der späteren Gebrauchssituation entsprechen, eine konzeptionelle Regelbasis zu testen. Die beteiligte Kontrollperson soll lediglich für einen kontinuierlichen Dialogfluss sorgen, wenn während des Tests Lücken oder Fehler erkannt werden. Das Ergebnis dieser Simulation und zugleich dieser Phase ist ein erstes Feinmodell, das wesentlich weniger Lücken, missverständliche Formulierungen oder Fehler enthält.

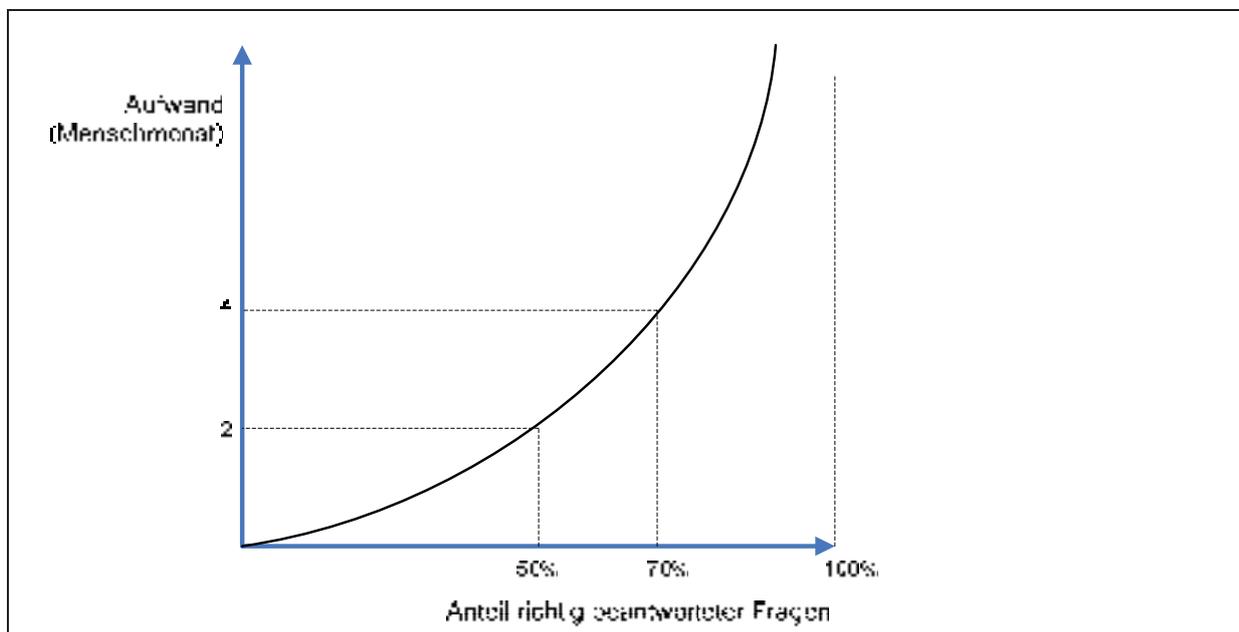


Abbildung 5-12: Aufwand zur Verbesserung der Beantwortungsrate
Quelle: (von Wendt 2003, 42)

Das Feinmodell geht im Rahmen der Modelladoptionsphase in den Betrieb auf dem realen Zielsystem über. Die dort erkannten Lücken in der Regelbasis führen fortlaufend zu revidierten Feinmodellen, die wiederum in das reale System eingehen. Somit wird im Laufe der Zeit die Quote der verarbeitbaren Eingaben verbessert. Dieses Vorgehen entspricht auch

der Praxis bei der Einführung von internetbasierten Avatarsystemen (Vetter 2003, 77; von Wendt 2003, 42). Von Wendt (2003, 42) berichtet davon, dass eine Abdeckung von 50% korrekt beantworteter Fragen nach einem Menschmonat erreicht werden kann, wobei sich die Rate exponentiell entwickelt (vgl. Abbildung 5-12). Vetter (2003, 77) geht davon aus, dass durch die kontinuierliche Pflege ein maximaler Wert von 90% korrekt beantworteter Fragen erreicht werden kann.

Die Beschreibung der Modelle von Ahlers und Öl (2005) sowie Eckert (1996, 91ff) zeigen, dass der Weg der Regelerstellung erforscht wird. Um einen praktischen Mehrwert für die praktische Arbeit mit Wissensbasen zu schaffen, müssen die einzelnen Schritte bis auf eine anwendbare Ebene detailliert werden. Die Beschreibungen von Wendts (2003) und Veters (2003) als Vertreter der kommerziellen Praxis lassen vermuten, dass in Unternehmen, die sich mit der Erstellung von internetbasierten Avatarsystemen beschäftigen, entsprechende Vorgehensmodelle bestehen. Da diese Modelle aus Gründen des Wettbewerbsvorteils nicht veröffentlicht wurden, ist es somit Teil dieses Forschungsvorhabens, im Rahmen der zweiten Forschungsfrage eine entsprechende Methode zumindest für den vorliegenden Fall zu erstellen.

5.2.2.2. *Werkzeuge zur Erstellung und Pflege*

Betrachtet man das Modell von Eckert (1996, 91ff), so lassen sich prinzipiell alle Schritte durch softwaregestützte Werkzeuge begleiten. Da die Aufgabenbeschreibung für die Umsetzung eines Avatarsystems als gegeben angesehen wird und im Bereich der softwaregestützten Anforderungserhebung und -verwaltung bereits eine Vielzahl an Forschungsarbeiten entstanden sind, soll dieser Schritt nicht aus diesem Aspekt betrachtet werden. Vielmehr beginnt die Werkzeugunterstützung mit der introspektiven Analyse, die aufgrund ihrer Situationsabhängigkeit und der meist unstrukturierten Daten als kreativer Prozess betrachtet werden muss. Dabei können neben der Analyse gegebenenfalls bestehender Dokumente auch Methoden wie das Brainstorming mit entsprechenden Tools angewandt werden.

Die Erstellung des Grob- und Feinmodells sowie dessen Revisionen erfolgen, je nach Ausprägung der Ausgabe mithilfe verschiedener Editoren, die kombiniert werden können. Auch wenn im Modell von Eckert (1996, 91ff) die verschiedenen Ausgabemodalitäten keine Beachtung finden, da er lediglich ein Sprachdialogsystem untersucht hat, so müssen die verschiedenen Outputs in die Betrachtung integriert werden.

Die grundlegende Komponente des Modells ist das Regelwerk, das primär die textuelle bzw. akustische Ausgabe steuert, aber auch Steuerbefehle für die weiteren Modalitäten enthalten kann. Die einfachste Umsetzung eines solchen Editors stellt ein rein textbasierter Ansatz dar, wie er beispielsweise beim GaitoBot Editor (http://www.gaito.de/content/produkte_gaitobot/editor.asp, zugegriffen am 13.04.2008), dem Lingubot Creator (von Wendt 2003, 42ff), dem Speech Builder (Glass/Seneff 2003) oder dem Carmeq Dialogstudio (Ahlers/Oel 2005, 38) Anwendung findet. Dabei werden Ein- und Ausgabe einer Regel in Textform und gegebenenfalls mit grafischer Hervorhebung und Anzeige von Zusammenhängen zwischen den Regeln visualisiert (vgl. Abbildung 5-13). Die grafische Visualisierung der textbasiert

erstellten Regeln, kann dabei auch stärker in den Vordergrund des Editors rücken, wie dies bei Laufer, Tatai und Nemeth (2005) der Fall ist. Meist werden die Editoren zudem mit integrierten Testfunktionen für die erstellte Regelbasis bereitgestellt.

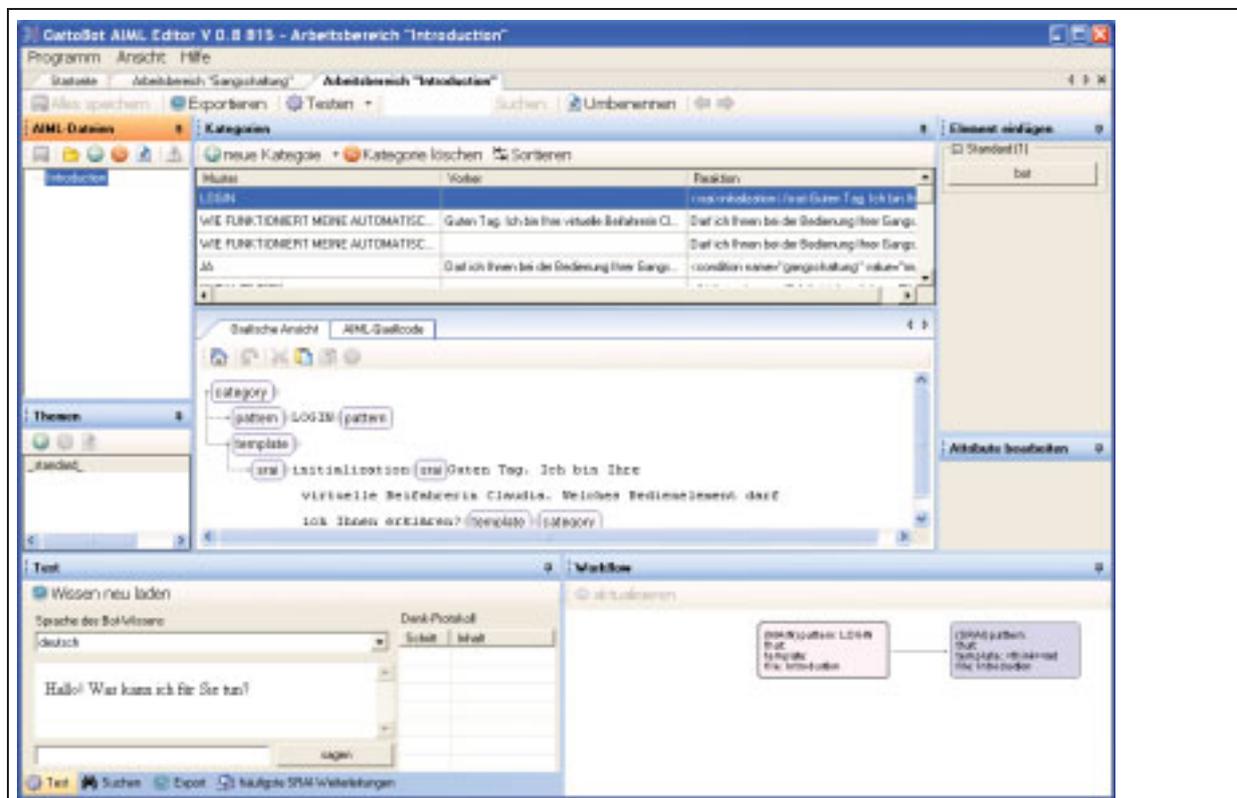


Abbildung 5-13: Screenshot des Regeleditors GaitoBot
Quelle: (Eigene Darstellung)

Da textbasierte Editoren recht nah an der Annotation der Regelbasis orientiert sind, erfordert der Umgang mit ihnen zumindest grundlegende Kenntnis der Annotation. Dies erschwert die Verwendung durch inhaltlich vertraute aber mit der Annotation unerfahrene Benutzer. Um diesem Problem zu begegnen, lässt sich die Gestaltung der Regelbasis auf grafischem Wege umsetzen, wie dies z.B. im SUEDE-Projekt erfolgte (Klemmer et al. 2000). Dabei ist jedoch zu bemerken, dass der SUEDE-Editor, der eine direkte grafische Modellierung zulässt, von seinen Funktionen her sehr eingeschränkt ist, kaum Transparenz über die genaue Umsetzung der grafisch modellierten Regeln bietet und nur für kleinere Modelle geeignet ist.

Neben der Erstellung der textbasierten Regelbasis erfordert die Modellgenerierung jedoch auch die Integration anderer Objekte wie beispielsweise Avatare, deren Animationen, 2D-Anzeigeelemente und Steuerdaten für die optimierte Sprachausgabe. Während zumindest für die grafischen Elemente allgemeine oder speziell für Avatarsysteme konzipierte Editoren bestehen (vgl. Balci 2005), lassen sich keine gesonderten Sprachoptimierungswerkzeuge identifizieren. Ein schwerwiegendes Problem aller Editoren ist die Trennung nach der Modalität, die durch sie gepflegt werden kann. Während auf Seiten des Outputs eine über alle Modalitäten hinweg synchronisierte Ausgabe des Avatarsystems erwartet wird, wird dies bisher von keinem integrativen Werkzeug unterstützt.

Der letzte Schritt aus dem Modell von Eckert (1996, 91ff), der hinsichtlich einer Werkzeugunterstützung untersucht werden muss, ist die Wizard-of-Oz-Simulation. Ein Werkzeug, das neben der Erstellung eines Modells auch die Durchführung des WoZ-Tests unterstützt ist SUEDE (Klemmer et al. 2000). Insgesamt ist bei der Durchführung solcher Simulationen im Bereich von Avatarsystemen anzumerken, dass ihre Durchführung nur unter Einschränkungen möglich und sinnvoll ist. So steigt die Komplexität der Vorbereitung und Durchführung mit zunehmender Zahl von Ausgabemodalitäten, da diese alle auf ihre Synchronität hin geprüft werden müssen. Bei einer großen Anzahl an entsprechenden Kanälen kann dies zu einer Verzögerung der Antwort aufgrund der manuellen Überprüfung führen. Darüber hinaus wird ein Teil der zur Ausgabe verwendeten Daten wie z.B. Animationsskripte, Aussprachewörterbücher oder prosodische Steuerbefehle bereits zur Designzeit erstellt. Dies macht es unmöglich, im Rahmen einer WoZ-Simulation solche Daten in angemessener Zeit zu generieren, wenn sie fehlen sollten.

Das größte Unterstützungspotenzial bei der Erstellung von Inhalten eines Avatarsystems lässt sich somit durch Editoren für die drei Schritte des Grob-, des Feinmodells und dessen Revisionen identifizieren. Dabei integriert ein optimaler Editor alle Aspekte der möglichen Ausgabemodalitäten und stellt dem Benutzer eine intuitive, grafische Modellierungsoberfläche zur Verfügung.

5.3. Konzeption der Benutzerschnittstelle

Die Benutzerschnittstelle eines Avatarsystems besteht zum einen aus der Darstellung eines Avatars selbst, einem Bereich in dem Inhalte vorgestellt werden können und gegebenenfalls einem Interaktions- bzw. Navigationsbereich, der den Umgang mit dem System ermöglicht oder vereinfacht. Aufgrund des hohen Stellenwerts des Avatars in diesem System soll seine Gestaltung zunächst betrachtet werden, bevor sein Umfeld und seine Integration darin beleuchtet werden.

5.3.1. Gestaltung des Avatars

Als zentrales (Inter-)Face spielt die Gestaltung des Avatars eine zentrale Rolle bei der Konzeption der Benutzeroberfläche. Neben Parametern, die das Aussehen oder die dahinterliegende Technik bestimmen, kann die Gestaltung auch charakterliche Wesenszüge umfassen (Schmidt 1998). Bei der Wahl der Merkmale müssen jedoch auch die Zusammenhänge mit anderen Komponenten des Avatarsystems berücksichtigt werden.

Zunächst muss bestimmt werden, welcher der drei Darstellungstypen nach Salem und Earle (2000, 94f) verwendet werden soll. Sie unterscheiden die natürliche bzw. reale, die abstrakte und die realistische Darstellungsform, welche sich an der Realität orientiert, aber künstlich produziert wird. Durch diese Wahl entscheiden sich zum einen die technischen Möglichkeiten der Darstellung und Animation sowie die Erwartungshaltung der Benutzer dem Avatarsystem gegenüber (Lindner 2003, 11; Nowak/Biocca 2003). Da der Avatar zur Erhaltung maximaler Flexibilität dynamisch animiert werden soll und eine zu hohe Erwartungshaltung einem prototypischen System gegenüber unangebracht ist, soll eine realistische Darstellung erfolgen.

Die Wahl des Geschlechts hängt ebenfalls wieder stark von der Erwartungshaltung des Benutzers ab. Während weibliche Wesen allgemein als sympathischer wahrgenommen werden, unterstellt man Männern eher technische Expertise. Weibliche Avatarsysteme im Internet sehen sich darüber hinaus häufiger Anzüglichkeiten ausgesetzt als männliche Exemplare (Lindner 2003, 11). Da die zu erklärenden Funktionen im Fahrzeuginnenraum überwiegend kein tiefgehendes technisches Wissen benötigen, soll der Sympathieeffekt eines weiblichen Avatars genutzt werden. Gegebenenfalls kann zu einem späteren Zeitpunkt ein männlicher Avatar für detaillierte technische Darstellungen ergänzt werden, wobei damit ein Mehraufwand einhergeht. Durch die Wahl des Avatargeschlechts ergibt sich die zwingende Synchronisierung der Sprachsynthese, d.h. zu einem weiblichen Avatar muss auch eine weibliche Stimme generiert werden (vgl. dazu Nass/Steuer/Tauber 1994; Heidmann/Biesterfeldt 2004, 63).

Eine weitere grundlegende Eigenschaft stellt die Ethnizität des Avatars dar. Lee und Nass (1998) konnten in Versuchen mit Avataren nachweisen, dass Avatare, welche die Volkszugehörigkeit des Benutzers haben, von den Benutzern als attraktiver und glaubwürdiger angesehen wurden. Darüber hinaus wurden die dargestellten Argumente als überzeugender bewertet. Angesichts der Anwendung des Avatarsystems im deutschen Sprachraum, ergibt sich somit eine mitteleuropäische Ethnizität des Avatars, um die dargestellten positiven Effekte nutzen zu können. Eine spätere Übertragung der Anwendung in eine andere Region der Erde, macht somit auch die Anpassung des Avatars selbst notwendig.

Ein weiteres Kriterium des Aussehens ist die Kleidung des Avatars. McBreen, Anderson und Jack (2001) haben gezeigt, dass ein Avatar seinem Umfeld angemessen gekleidet sein muss, um glaubwürdig zu erscheinen. Um ermitteln zu können, was angemessen ist, muss zunächst eine Referenzsituation gefunden werden, an der man sich orientieren kann. Dies kann beispielsweise die Auswahl des Fahrzeugs im Autohaus oder die Einweisung in die Benutzung des Fahrzeugs bei der Abholung sein. In beiden Fällen wird von den realen Personen Business Dress getragen, was somit auch für den Avatar gelten soll. Die genaue Farbgestaltung muss dabei in Abstimmung mit dem restlichen Anzeigebereich erfolgen. Eine Anpassung hinsichtlich der Corporate Identity von Audi kann in einem späteren Schritt erfolgen.

Angesichts der internationalen Ausrichtung des Automobilmarkts wurde auch das Thema der Lokalisierung des Avatarsystems betrachtet, was insbesondere Ethnizität und Sprache angeht. Neben der visuellen Gestaltung des Avatars hat die Anpassung für andere Kulturkreise auch schwerwiegende Auswirkungen auf die gesamte Regelbasis sowie gegebenenfalls die grafische Darstellung von Bedienelementen. Da der Fokus des Prototyps zunächst die erste Umsetzung und Erprobung eines Avatarsystems im Fahrzeug hat, soll der Avatar zunächst mitteleuropäischer Herkunft sein und lediglich die deutsche Sprache beherrschen. Die Planung und die konkrete Umsetzung sollen dabei jedoch nicht die Anpassung für andere Regionen verhindern.

Ein Aspekt der alle Merkmale des Aussehens betrifft ist die zu erwartende Größe des Displays und damit eng verbunden die Qualität der Darstellung des Avatars. Aufgrund der

aktuell geringen Auflösung von 480x240 Pixel im Farbdisplay des Audi MMI der zweite Generation soll auf zu viel Realitätsnähe verzichtet und stattdessen der Fokus auf eine klare und scharfe, anthropomorphe Darstellung gelegt werden. Die Anpassung der Detailtiefe kommt dabei auch dem beschränkten Speichervolumen für Daten im Fahrzeug entgegen. Abbildung 5-14 zeigt zwei Beispiele, die während der Konzeption des Avatareaussehens in der Modellierungssoftware Poser 7 entstanden sind.



Abbildung 5-14: Beispiele der konzeptionellen Avatargestaltung in Poser
Quelle: (Eigene Darstellung)

Neben der visuellen Konzeption eines Avatars spielt auch die emotionale und Persönlichkeitsmodellierung eines virtuellen Wesens eine Rolle (Picard 2000, 15). Auch wenn Bühler (2003, 116ff) eine ausgeprägte Persönlichkeit mit Ecken und Kanten als Erfolgsfaktor für Avatarsysteme im Internet ausmacht, ist unklar, ob dies auch für Avatarsysteme im Auto gilt. Während der Benutzer dem Avatarsystem im Internet jederzeit den Rücken kehren kann, sind Benutzer und Avatar im Fahrzeug praktisch auf kleinem Raum zusammengesperrt. Somit ist eine ausgeprägte Persönlichkeit des Avatars kein Teil der initialen Konzeption bis Erkenntnisse zur veränderten Situation vorliegen.

5.3.2. Gestaltung der Benutzeroberfläche

Auch wenn der Avatar die wichtigste wahrnehmbare Komponente des Systems ist, ist er in das Rahmenwerk einer Benutzerschnittstelle eingebunden. Dadurch können neben dem Avatar beispielsweise auch seine dreidimensionale Umwelt, zweidimensionale Bilder und Filme, Texte oder übergeordnete Navigations- und Informationsfunktionen angezeigt werden. Da das Lesen von Text einerseits einen Medienbruch und andererseits als Ablenkung während der Fahrt verstanden werden kann, ist die Nutzung von Text als Anzeigeobjekt zu vermeiden. Betrachtet man zudem die Memorisierung von Text im Vergleich zu Bildern, ist ein Bildüberlegenheitseffekt zu erkennen (Paivio 1979). Die angezeigten visuellen Bilder müssen dabei in der Lage sein, dem Benutzer ihren Informationsgehalt im Rahmen einer im Fahrzeug üblichen Blickspanne von 0,6 bis 2 Sekunden zu offenbaren (vgl. Wierwille et al. 1991;

Kimura/Marunaka/Sugira 1997; Williams/Helbig 2007, 302f). Trotz des dadurch vermittelten Informationsgehalts, darf die visuelle Darstellung nicht zu komplex sein, da sie sich sonst negativ auf die Fahrzeugsteuerung durch den Fahrer auswirkt (Bruckmayr/Reker 1994; Srinivasan/Jovanis 1997).

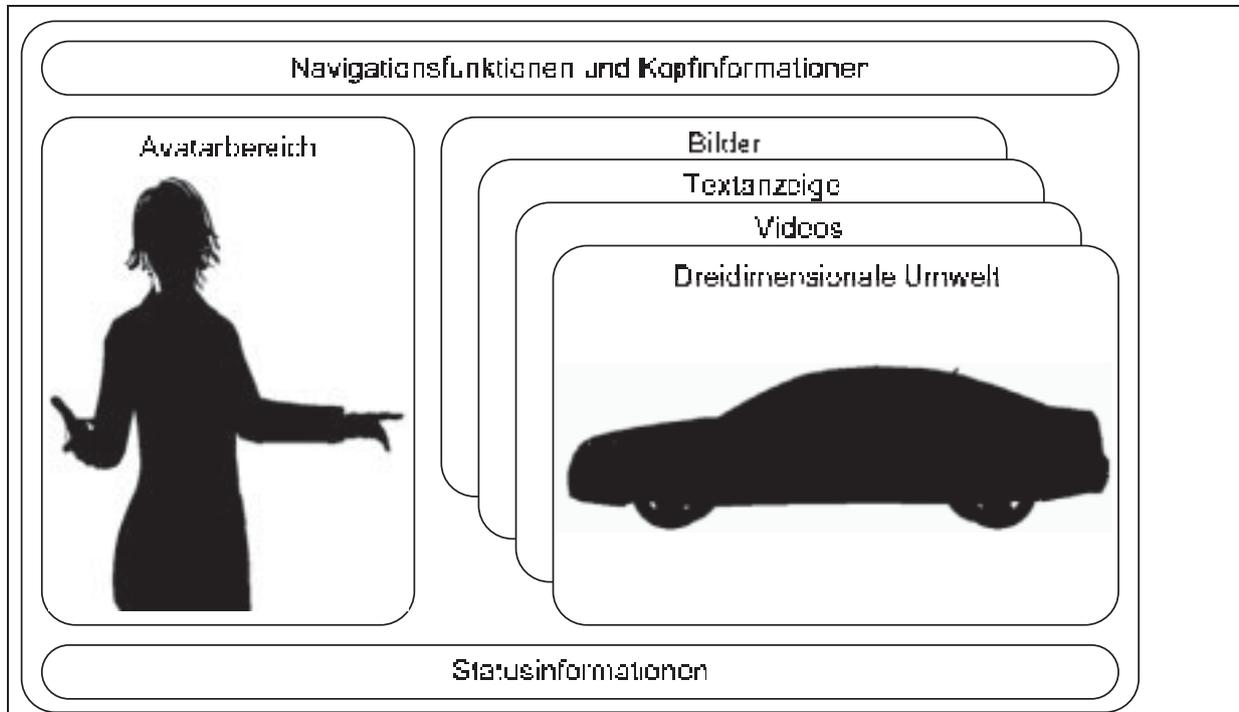


Abbildung 5-15: Konzeptionelle Anordnung der Teile der Benutzeroberfläche
Quelle: (Eigene Darstellung)

Die Abkehr von einer textorientierten Steuerung, wie wir sie von Computern her gewohnt sind und die Unerfahrenheit im Umgang mit avatarbasierten, interaktiven Sprachdialogsystemen lässt die Benutzung des Systems ungewohnt erscheinen. Um zumindest auf der Erfahrung der Benutzer aufbauen zu können, die gegebenenfalls bereits Computerprogramme bedient haben, sollte sich die neue Schnittstelle an Computerprogrammen orientieren (vgl. dazu auch Salmen 2002, 48). Daraus ergibt sich eine erste konzeptionelle Anordnung der Teile im Rahmenwerk der visuell wahrnehmbaren Benutzeroberfläche (vgl. Abbildung 5-15). Die Positionierung des Avatars innerhalb des Rahmenwerks kann dabei statt auf der linken auch auf der rechten Seite erfolgen. Diese Entscheidung hängt einerseits von der Position des Displays im Fahrzeug (vgl. Helal/Moore/Ramachandran 2001) sowie von der beabsichtigten Nähe zum Fahrer ab (vgl. Short/Williams/Christie 1976). Gegebenenfalls kann die Position und der Blickwinkel sowie die Entfernung des Avatars vom Benutzer angepasst werden.

Insbesondere die Kopf- und Statusinformationen des Frameworks stellen bekannte Elemente dar, während der Avatar und die angezeigten multimedialen Inhalte durch die exklusiv akustische Steuerung eine ungewohnte Komponente darstellen. Zudem repräsentiert die fehlende Anzeige möglicher Funktionen einen neuen Aspekt, der gemäß der Einteilung von Fischer (2001) die Menge der unbekannteren Funktionen vergrößert. Die Darstellung

verschiedener Symbole im Bereich der Statusinformationen soll deshalb dem Benutzer grundlegende Handlungsoptionen aufzeigen (vgl. Shneiderman/Plaisant 2005, 333).

Auch wenn die visuelle Komponente einen wichtigen Teil der Benutzerschnittstelle ausmacht, muss auch die akustische Interaktion betrachtet werden. Die Darstellung von Text oder Bildern spricht ebenso wie Informationen des Straßenverkehrs den visuellen Aufnahmekanal eines Menschen an. Je nach Nebentätigkeit nimmt dies 20% (Bedienung eines CD-Players im Auto) bis 50% (Telefonieren im Auto) des visuellen Kanals in Anspruch, was zu einer deutlichen Ablenkung führt (Wikman/Nieminen/Summala 1998; Airaksinen et al. 2004). Die akustische Ausgabe durch das Fahrzeug hingegen nutzt einen menschlichen Aufnahmekanal, der weniger ablenkt als der Blick auf ein Display (Parkes/Coleman 1990). Zudem wird der menschliche Ausgabekanal des Sprechens während der Fahrt nur bei Anwesenheit eines Beifahrers genutzt und ermöglicht somit eine bidirektionale Kommunikation (Mollenhauer et al. 1994).

Auch wenn Gstalter, Fastenmeier und Galsterer (1995) die Überlegenheit der akustischen Ausgabe im Gegensatz zu visuellen Informationen ermittelt haben, stellten Bengler (1995) sowie Srinivasan und Jovanis (1997) in Versuchen fest, dass die Kombination von akustischen und visuellen Informationen zu einer geringeren Belastung des Fahrers führt. Von besonderer Bedeutung ist dabei jedoch die inhaltliche und zeitliche Abstimmung der Modalitäten, da nur bei einer synchronisierten Ausgabe auch die Reaktionszeit des Fahrers bei Warn- oder Fehlermeldungen reduziert werden kann (Selcon/Taylor/McKenna 1995). Die verkürzte Reaktionszeit bei der Verwendung mehrerer Modalitäten lässt sich dadurch erklären, dass der Schwellenwert für die Aufmerksamkeit des Fahrers während der Fahrt leichter überschritten werden kann, wenn visuelle und akustische Informationen zusammen vermittelt werden (vgl. Hartmann 1984). Eine rein akustische Ausgabe hingegen begegnet nur unzureichend der Erwartungshaltung des Benutzers hinsichtlich einer visuellen Bestätigung einer Aussage (Airaksinen et al. 2004).

5.4. Zusammenfassung

Die Konzeption des Avatarsystems im Fahrzeug erstreckt sich über die drei Ebenen der Technik, des Inhalts und der Benutzeroberfläche. Insbesondere die technische Betrachtung erfordert eine umfangreiche Analyse bestehender Ausprägungen im Bereich von Avatar- und Sprachdialogsystemen. Auf Basis der darin identifizierten Bestandteile können Komponenten und Datencontainer extrahiert und nach einem Vergleich mit anderen Systemen vereinheitlicht beschrieben werden. Die bereinigten Komponenten ermöglichen es, sie neu zusammenzusetzen und daraus eine Referenzarchitektur für Avatarsysteme mit Sprachinteraktion zu gewinnen. Da die Verteilung der einzelnen Komponenten und Systeme der Architektur insbesondere beim Einsatz in mobilen Umgebungen eine bedeutende Rolle spielt, wird auch dieser Aspekt im Rahmen der Referenzarchitektur berücksichtigt und verschiedene Verteilungsmöglichkeiten dargestellt. Von dieser Referenzarchitektur kann schließlich eine konkrete Architektur für den Einsatz im Fahrzeug unter Berücksichtigung der spezifischen Gegebenheiten gewonnen werden. Die Erarbeitung einer Referenzarchitektur im Allgemeinen und für das Automobilumfeld im Besonderen stellt zugleich die Antwort auf die erste Forschungsfrage dar.

Im Rahmen der inhaltlichen Konzeption sind zunächst der Aufbau und die Organisation des Inhalts in Form der Regelbasis zu betrachten. Dabei spielt der Zusammenhang zwischen Eingabe und Reaktion ebenso eine Rolle wie die Möglichkeit, neben auszugebendem Text zusätzliche Funktionen in die Reaktion einzubinden. Die sinnvollen bzw. im konkreten Zusammenhang notwendigen Funktionen umfassen neben dem Zugriff auf Datencontainer und die Steuerung der Benutzerschnittstelle auch die Korrektur von Fehlern durch inhaltliche oder technische Störfaktoren bei der Benutzereingabe. Neben grundsätzlichen Punkten der Regelbasis sind auch die Abgrenzung und der Zusammenhang realisierbarer Inhalte in Form von Szenarien abzustecken. So können neben einer Einführung in das Avatarsystem eine Anleitung für die ersten Schritte mit dem Fahrzeug als grundlegende Szenarien betrachtet werden. Da die Umsetzung einzelner Szenarien eine umfangreiche Regelbasis zur Folge hat, steht der Aspekt der Erstellung und Pflege der einzelnen Regeln in das Gesamtszenario ebenfalls im Schlaglicht. Dabei lassen sich bisher überwiegend textbasierte und ansatzweise auch grafikbasierte Lösungen zur Bearbeitung der Regelbasis unterscheiden.

Die Benutzerschnittstelle wird vom Avatar dominiert, so dass dessen Gestaltung hinsichtlich wichtiger Merkmale wie Geschlecht, Ethnizität und Kleidung im Vordergrund stehen. Der Avatar ist eines von mehreren Bestandteilen, welche die gesamte Benutzerschnittstelle darstellen. So sind darüber hinaus noch ein Bereich zur Anzeige von Bildern, Texten, Videos oder dreidimensionalen Objekte sowie eine Kopf- und eine Fußleiste eingebunden. Neben der visuellen Komponente ist auch die Sprachinteraktion Teil der Benutzerschnittstelle und muss hinsichtlich verschiedener Kriterien mit der visuellen Ausgabe abgestimmt werden.

6. Implementierung eines interaktiven Hilfesystems im Fahrzeug

„Auch aus Steinen, die einem in den Weg gelegt werden, kann man Schönes bauen.“

Johann Wolfgang von Goethe (1749 – 1832), deutscher Dichter

Die Umsetzung der im vorangegangenen Kapitel dargestellten Konzeption umfasst im Wesentlichen zwei Bereiche. Zum einen ist dies die technische Umsetzung des Avatarsystems gemäß der erarbeiteten Systemarchitektur für den Einsatz im Fahrzeug. Dabei spielt zunächst die Wahl des konkreten Entwicklungsframeworks, der Technologien und schließlich die Umsetzung der einzelnen Komponenten eine Rolle. Zum anderen muss die technische Plattform mit Leben in Form von Regeln, Mediendaten, Animations- und Aussprachedaten erfüllt werden. Dabei spielt neben softwarebasierten Werkzeugen auch die Herangehensweise zur Erstellung von Inhalten eine wichtige Rolle. Dieser Aufteilung folgend wird in Kapitel 6.1 der technische Aspekt der Implementierung dargestellt während in Kapitel 6.2 ein Vorgehen für die Zusammenstellung und Aufarbeitung der Inhalte entwickelt wird (vgl. Abbildung 6-1).

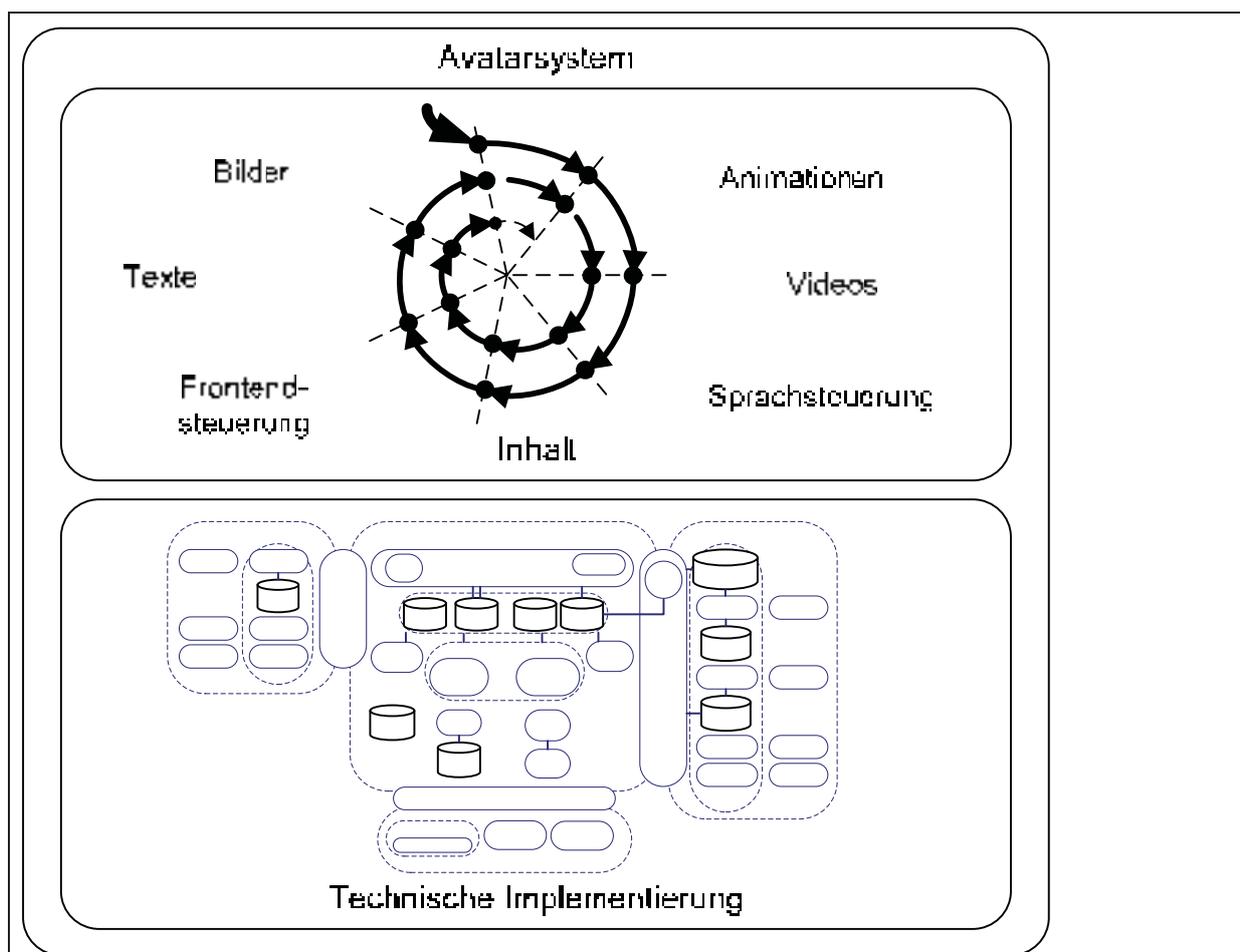


Abbildung 6-1: *Zweiteilung der Implementierung in Technik und Inhalt*

Quelle: (Eigene Darstellung)

6.1. Implementierung des technischen Systems

Wie beim Bau eines Hauses der Bauplan, so kann die Konzeption der Systemarchitektur als Rahmenwerk für die konkrete Umsetzung des Systems betrachtet werden. Die Frage mit welchen Hilfsmitteln und nach welchen Regeln gebaut wird, soll im Rahmen der Implementierung betrachtet werden. Dabei sind Standards hinsichtlich der Gestaltung und der Kommunikation zwischen technischen Komponenten von Bedeutung. Auf der Grundlage detaillierter Regeln von standardisierten Frameworks lassen sich die einzelnen Komponenten entwickeln. Zunächst soll betrachtet werden wie das Avatarsystem physisch in das Testfahrzeug eingebunden ist, um das notwendige Hintergrundwissen zu schaffen. Darauf aufbauend werden relevante Standards und Entwicklungsrahmen im Automobilumfeld beleuchtet. Im Rahmen der gewählten Entwicklungsumgebung werden die darin entwickelten Komponenten der Laufzeit genau analysiert und ihre Funktion beschrieben. Den Abschluss bildet die ganzheitliche Betrachtung des Systems und des Zusammenspiels der einzelnen Komponenten.

6.1.1. Technische Integration ins Fahrzeug

Aufgrund des prototypischen Charakters des Forschungsvorhabens wird die Laufzeit des Avatarsystems im Fahrzeug auf zusätzlicher Hardware ausgeführt und nicht auf dem bestehenden Bordsystem entwickelt. Die Grundlage des Avatarsystems stellt ein so genannter CarPC dar, der auf einem handelsüblichen PC des Formfaktors Mini-ITX basiert. Wegen des Einbaus auf engem Raum und der damit eingeschränkten Möglichkeiten zur Ableitung der Abluft, kommt der Rechner mit einer passiven Kühlung aus. Anstelle einer Festplatte befindet sich im Rechner eine SD-Flash-Karte mit 4 GB Speicher für den industriellen Einsatz, um auch bei starken Erschütterungen und Vibrationen Daten lesen und schreiben zu können. Die Speicherkarte ist standardmäßig im read-only Modus konfiguriert, so dass alle Daten bis auf Weiteres im 1 GB großen Hauptspeicher gehalten und nur mit einem gesonderten Befehl persistent gespeichert werden. Auf der Speicherkarte wird das Betriebssystem Windows XP betrieben, so dass Entwicklungen von herkömmlichen Rechnern mit dem gleichen Betriebssystem direkt übernommen werden können. Neben den Anschlüssen eines handelsüblichen PC's liegen im CarPC zudem zwei Anschlüsse für den CAN-Bus, eine GPS-sowie eine WLAN-Antenne, ein digitaler Input-/Outputkanal sowie mehrere Eingänge und Ausgänge für Videosignal vor (vgl. zum CAN-Bus auch Kapitel 6.1.3.2). Die Stromversorgung erfolgt ohne Transformation direkt über das Bordsystem, so dass der Rechner startet, sobald sich der Autoschlüssel in der Nähe des Fahrzeugs befindet¹⁹. Ebenso schaltet sich der CarPC ab, sobald das Bordsystem einige Minuten ohne Strom ist.

Die CAN-Bus-Anbindung erlaubt den gleichzeitigen Zugriff auf einen Low Speed und einen High Speed CAN-Bus, so dass die Bussysteme für den Antrieb und die Komfort-Funktionen parallel angesprochen werden können und Informationen über das Fahren des Fahrzeugs als auch die Aktivitäten des Fahrers im Innenraum vorliegen. Die Anzeige der visuellen Ausgabe des Avatarsystems erfolgt unter Nutzung des eingebauten MMI-Displays im Fahrzeug. Dazu

¹⁹ Audis Keyless-Go Technik ermöglicht die Erkennung des Fahrers, sobald der Fahrzeugschlüssel in der Nähe des Fahrzeugs ist.

wird zum einen das DVI-Ausgangssignal des CarPC in einem so genannten Display Switch Module (DSM) konvertiert, damit es dem internen Datenstrom für das Display entspricht. Das DSM schaltet das Videosignal auf dem MMI-Display vom ursprünglichen MMI-Signal und das Avatarsystem um, sobald im unteren Bereich der Mittelkonsole der Knopf „Info“ betätigt wird, der normalerweise zur Anzeige von Verkehrsmeldungen führt.

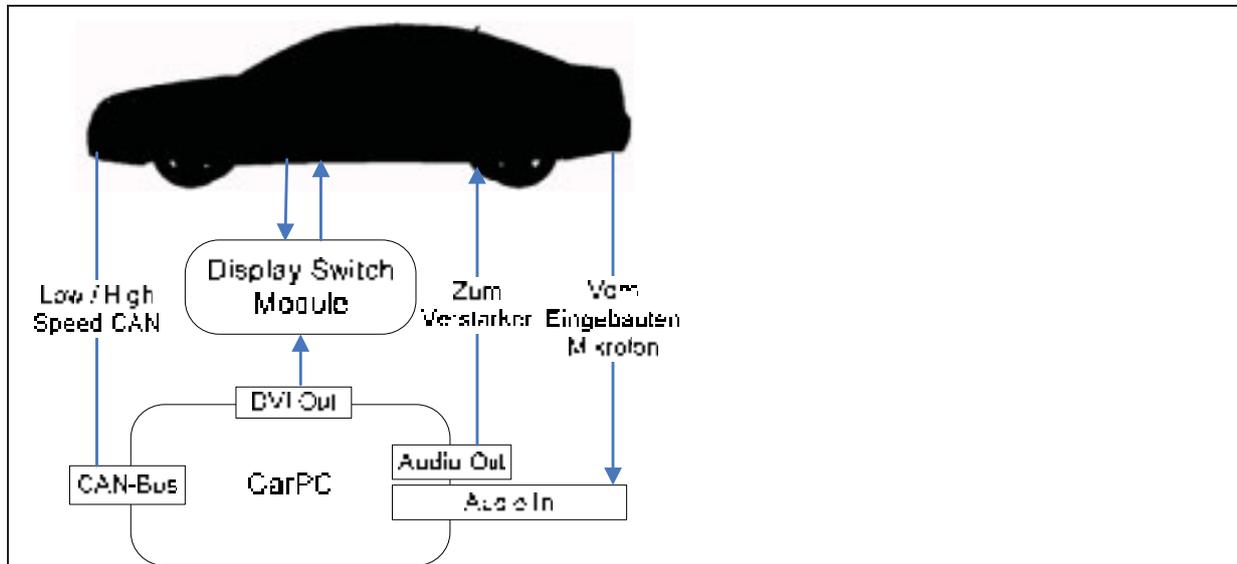


Abbildung 6-2: *Integration des CarPC's in das Fahrzeug*

Quelle: (Eigene Darstellung)

Die akustische Integration erfolgt auf Seiten des Audio-Ausgangs über die Anbindung an den zentralen Audio-Verstärker des Fahrzeugs, der dadurch akustische Signale des Avatarsystems wie Audiosignale eines Telefons interpretiert und alle anderen Audio-Ausgaben für die Dauer der Sprachsynthese unterdrückt²⁰. Der Audio-Eingang des CarPC ist mit einem gesondert verbauten Mikrofon verbunden, das im Dachhimmel im Bereich der Leseleuchte für den Fahrer angeordnet ist. Die Nutzung der serienmäßig eingebauten Mikrofone hätte ansonsten einen massiven Eingriff in die bestehende Bordelektronik notwendig gemacht. Eine schematische Darstellung der Integration des CarPC's in das Fahrzeug ist in Abbildung 6-2 zu sehen.

Der CarPC und das DSM wurden im Handschuhfach des Fahrzeugs verstaut, wohin auch die zusätzlichen Anschlüsse an den CAN-Bus, das MMI-Display sowie die Kabel für Mikrofon und Lautsprecher gelegt wurden. Abbildung 6-3 zeigt die Einbausituation, wobei der schwarze CarPC und das silberne DSM lose reingelegt wurden. Zeitweise wurde zur technischen Überwachung durch einen Laptop ein Netzkabel aus dem Handschuhfach raus gelegt.

²⁰ Aufgrund von technischen Problemen wurde während der Entwicklungs- und Evaluationszeit die Audio-Ausgabe überwiegend durch zusätzliche Lautsprecher umgesetzt, die sich jedoch für die Probanden unsichtbar im hinteren Teil des Fahrzeugs befanden.



Abbildung 6-3: *Unterbringung des CarPC und des DSM im Handschuhfach*
Quelle: (Eigene Darstellung)

6.1.2. Wahl eines Entwicklungsframeworks

Neben der Wahl einer Hardwareplattform oder der Programmiersprache trägt insbesondere die Verwendung eines geeigneten Entwicklungsframeworks zur standardisierten und modularen Entwicklung von Komponenten bei. Neben Richtlinien, welche Funktionen oder Schnittstellen einzelne Komponenten haben sollen, bieten Frameworks teilweise auch zentrale Dienste zur Verwaltung, zum Datenaustausch und zum Monitoring. Dadurch kann die Entwicklung beschleunigt werden während die Wiederverwendbarkeit und Austauschbarkeit einzelner Module gewährleistet wird.

Zunächst werden verschiedene Frameworks betrachtet, die im Automobilbereich verwendet werden, bevor das genutzte Rahmenwerk OSGi detailliert betrachtet wird. Schließlich werden unter Anwendung des verwendeten Frameworks zu implementierende Komponenten auf der Basis der konzipierten Systemarchitektur abgegrenzt.

6.1.2.1. Übersicht relevanter Frameworks

Im Rahmen der Softwareentwicklung im Automobilumfeld lassen sich verschiedene Standardisierungsbemühungen identifizieren, die meist zum Ziel haben, einheitliche Entwicklungsprozesse oder Schnittstellen zu etablieren. Insbesondere Automobilhersteller versuchen sich dadurch von zuliefererabhängigen Geräten und Schnittstellen loszusagen. Die ersten Bemühungen dieser Art erfolgten bereits 1995 mit dem Ansatz „Offene Systeme für die Elektronik im Kraftfahrzeug/ Vehicle Distributed Executive“ (OSEK/VDX) zur Standardisierung von Betriebssystemen im Fahrzeug (Zimmermann/Schmidgall 2007, 230ff).

Die Herstellerinitiative Software (HIS) stellt einen Zusammenschluss zwischen den Herstellern Audi, BMW, Daimler, Porsche und VW dar. Im Rahmen der Kooperation werden in fünf Bereichen gemeinsame Prozessmodelle und Architekturen mit dem Fokus auf der praktischen Anwendung entwickelt. Die bearbeiteten Bereiche sind dabei Standard-Software,

Prozess-Assessment, Simulation und Tools, Flashbase Software sowie Software-Test. Bezüglich der gemeinsamen Prozessangleichung entstand dabei die Anpassung des Reifegradmodells SPICE (Software Process Improvement and Capability Determination) zu Automotive SPICE. Die Konzeption und Umsetzung von Softwarearchitekturen erfolgt dabei in enger Zusammenarbeit mit einem weiteren Konsortium: AUTOSAR (AUTomotive Open System ARchitecture) (Chodura et al. 2004; Daginnus et al. 2007; Zimmermann/Schmidgall 2007, 254ff).

AUTOSAR bildete sich 2003 aus Bemühungen der Unternehmen BMW, Daimler, Bosch, Continental, Volkswagen und Siemens VDO, die sich zum Ziel machten, ein gemeinsames Rahmenwerk festzulegen, das jedoch individuell umgesetzt werden kann. Dabei stehen die Bereiche Antrieb, Chassis, Sicherheit, Multimedia, Komfort und Mensch-Maschine-Schnittstelle im Vordergrund (Heinecke et al. 2004). Eine umfangreiche Softwarearchitektur, die gemeinsam erarbeitet wurde, stellt dabei den Rahmen für die Umsetzung dar. Kern dieser Architektur ist der Virtual Functional Bus, der ein Zusammenspiel aller einzelnen Komponenten ermöglicht (Fennel et al. 2006). Eine konkrete Implementierung der AUTOSAR-Spezifikationen ist das Entwicklungssystem ASCET (Advanced Simulation and Control Engineering Tool) bzw. ASCET-SD der Bosch-Tochter ETAS, das die modellbasierte Entwicklung von eingebetteter Software im Automobil erlaubt (Stürmer/Weinberg/Conrad 2005; Zerfowski 2001; Zimmermann/Schmidgall 2007, 266ff).

Parallel zu den Kooperationen von HIS und AUTOSAR besteht eine weitere Kooperation von Standardisierungsgremien. Den Anstoß dazu gab die Initiative AMI-C (Automotive Multimedia Interface Collaboration), welche ebenfalls als Zusammenschluss von Automobilherstellern und Zulieferern verschiedene technische Prozesse und Implementierungsrahmen spezifizierte. Inzwischen sind die meisten gemeinsam erstellten und gepflegten Dokumente in die Obhut anderer Standardisierungsorganisationen übergegangen, unter ihnen die OSGi (Open Services Gateway initiative) Alliance mit der bereits zuvor eine enge Abstimmung bestand (vgl. <http://www.ami-c.org>, zugegriffen am 19.04.2008; Guglielmetti, 2003 #912).

Die OSGi Alliance, die 1999 gegründet wurde, stellt den ältesten formalisierten Zusammenschluss von Unternehmen zur Standardisierung in diesem Bereich dar. Dabei liegt der Fokus nicht nur auf dem Automobilbereich, sondern deckt auch die Bereiche Unternehmen, Heim, Mobil, Telematik und Consumer ab. Die Entwicklung einzelner Komponenten erfolgt dabei in Form so genannter Bundles. Das Framework bietet eine Laufzeitumgebung mit zentralen Diensten und unterstützt die Verwaltung der Bundles. Die lose Kopplung der Bundles sowie ein zentrales Service Repository und Lifecycle-Services führen nach der Unterscheidung von Baresi und Ghezzi (2004, 59ff) zur Einordnung der OSGi-Architektur als einen Service-orientierten Ansatz. Im Gegensatz zu den anderen genannten Standardisierungsgruppen ist das OSGi Framework nicht sprachen- bzw. plattformunabhängig, sondern von vornherein als Aufsatz auf Java konzipiert. Obwohl Java aufgrund der niedrigen Verarbeitungsgeschwindigkeit nicht die Sprache der Wahl für sicherheitskritische Anwendungen im Automobilbereich ist, findet das OSGi Framework sowohl im Prototyping als auch im Bereich des Infotainment in der Serie Anwendung (OSGi Alliance 2007a, 6; Saad 2003).

Bei den Dokumenten der OSGi Alliance handelt es sich, wie bei den anderen Frameworks auch, lediglich um Spezifikationen, die in konkreten Implementierungen umgesetzt werden. Neben kommerziellen Implementierungen existiert eine Vielzahl von Open Source Implementierungen wie Equinox, Oscar, Apache Felix, Knopflerfish oder Concierge. Die verschiedenen Implementierungen erlauben alle das Ausführen entwickelter Bundles, wobei Unterschiede in der Anzahl bereitgestellter Services und der Performance zu erkennen sind. Über die verschiedenen Laufzeitumgebungen hinaus, kann zur Entwicklung von Bundles ein Eclipse-Plug-in genutzt werden, das die Erstellung und Überprüfung des Quellcodes sowie zusätzlich benötigter Dokumente vereinfacht (OSGi Alliance 2007a, 6).

Von den vorgestellten Frameworks wurde OSGi für die Implementierung des technischen Teils des Avatarsystems gewählt. Die Wahl erfolgte aufgrund verschiedener Kriterien. So spielte zum einen die konkrete Bindung an eine Programmiersprache wie Java, die weite Verbreitung findet und dementsprechend eine Vielzahl von Dokumentation und eine große Community aufweist, eine Rolle. Zudem kann durch verschiedene Implementierungen die Gefahr reduziert werden, auf einen einzigen Hersteller angewiesen zu sein. Auch weist die Verwendung von OSGi im Prototyping auf die Möglichkeit einer schnellen Umsetzung hin. Hinsichtlich des Austauschs von Komponenten mit Audi-Fachabteilungen und dem INI.TUM-Projekt HIMEPP erweist sich OSGi ebenfalls als vorteilhaft, da sowohl im genannten Projekt als auch in den Abteilungen I/FP-34 und I/AEV-3 bereits Know-how und Bundles für OSGi bestehen.

Zur Laufzeit wurde die OSGi-Implementierung von Equinox (<http://www.eclipse.org/equinox>, zugegriffen am 19.04.2008) verwendet, die auch in der Entwicklungsumgebung Eclipse eingebunden ist. Neben dem Vorteil, Bundles editieren und direkt in Verbindung mit anderen Bundles laufen zu lassen, spielte auch die Integration von Bundles aus dem AEV-Umfeld eine Rolle. Insbesondere das Bundle zur technischen Anbindung des CAN-Bussystems, das von AEV auf Knopflerfish in der Version 3.3.6 und somit OSGi in der Version 3 entwickelt wurde führte zu dieser Entscheidung. Eine Migration dieses Bundles auf die neuere Version 4 von Knopflerfish bzw. OSGi ist ohne den Zugriff auf den Quellcode nicht möglich. Die Funktion des Eventings und damit der losen Kopplung von Bundles über Ereignisse ist jedoch erst ab OSGi in der Version 4 möglich. Nur durch die Verwendung von Equinox unter Zuhilfenahme von HIMEPP Bundles ist sowohl die Migration des CAN-Bundles ohne Quellcodeänderung auf eine neuere OSGi-Version als auch die Verwendung des OSGi-Eventings möglich gewesen.

6.1.2.2. Merkmale von OSGi

Die Wahl des OSGi Frameworks mit seinen Merkmalen bringt verschiedene Implikationen hinsichtlich der technischen Implementierung mit sich. Bevor auf das Mapping des OSGi-Frameworks auf die konzeptionelle Systemarchitektur eingegangen wird, sollen die Eigenschaften des Frameworks genauer betrachtet werden.

Die OSGi Plattform setzt auf einem Rechner auf, der neben der Bereitstellung der Hardware auch ein passendes Betriebssystem beinhaltet, auf dem das Java Runtime Environment lauffähig ist. Darauf baut das OSGi Framework auf, das aus dem Framework selbst, einer

Vielzahl von Services, einer Sicherheitskomponente sowie den selbstentwickelten Bundles besteht (OSGi Alliance 2007b, 1ff) (vgl. Abbildung 6-4). Durch den Aufsatz von OSGi auf die Java-Umgebung, profitieren OSGi-Anwendungen von der Plattformunabhängigkeit von Java.

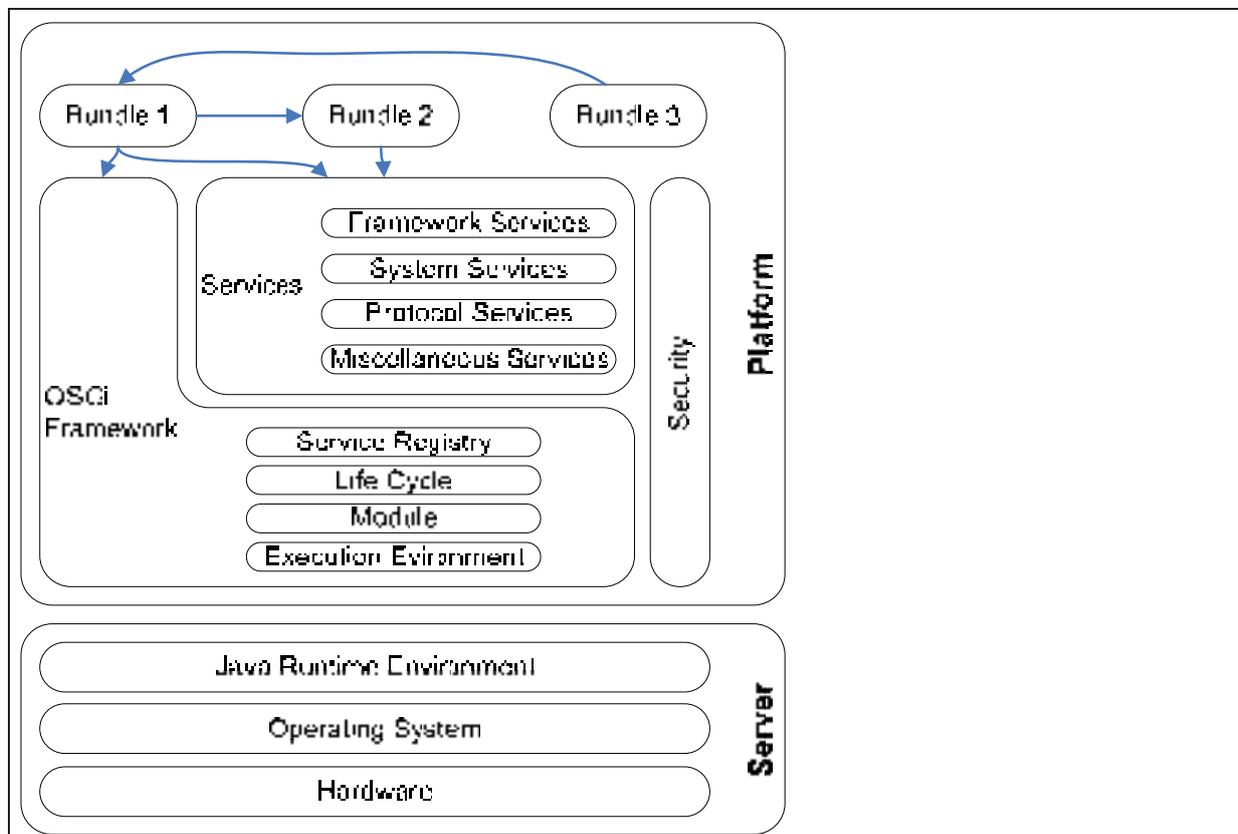


Abbildung 6-4: *Architektur des OSGi Frameworks*
Quelle: (In Anlehnung an (OSGi Alliance 2007b, 2))

Das Execution Environment des OSGi Frameworks enthält die minimale Menge an Teilen der Java Umgebung, um das OSGi Framework laufen lassen zu können. Insbesondere in mobilen Umgebungen werden nicht immer alle Bestandteile der Java Laufzeitumgebung installiert, so dass die bestehenden Bibliotheken überprüft werden müssen. Die Module-Schicht koordiniert den Ladevorgang von Java Archiven und deren Abhängigkeiten. Im Gegensatz zu herkömmlichen Java Anwendungen können OSGi-Bundles an vielen verschiedenen Orten liegen und müssen in der korrekten Reihenfolge ihrer Abhängigkeiten von diesen Ablageorten bezogen werden. Darüber hinaus können geschlossene Systeme im Rahmen der OSGi-Umgebung eingebunden werden. Die Life Cycle Ebene erlaubt das Installieren, Starten, Beenden, Deinstallieren und Aktualisieren von Bundles während der Laufzeit des OSGi Frameworks. Diese Dynamik macht einen flexiblen Koordinationsmechanismus zwischen den Bundles notwendig, der durch die Service Registry ermöglicht wird. Da in Bundles nicht davon ausgegangen werden kann, dass ein abhängiges Bundle immer installiert ist, kann dies über diese zentrale Stelle überprüft und das abhängige Bundle gegebenenfalls referenziert werden (OSGi Alliance 2008, 2007b, 1ff).

Im Rahmen der Framework-Services können Berechtigungen für den Zugriff auf Bundles oder die Startreihenfolge verwaltet werden. Darüber hinaus erfolgt durch diese Services die automatische Aufhebung von Verbindungen zwischen Bundles im Falle einer Aktualisierung zur Laufzeit. Die System-Services bieten neben Funktionen für das Logging und das Nachladen von Treiberklassen für angeschlossene Geräte auch eine Benutzerverwaltung. Der Event Admin-Service erlaubt zudem den Austausch von Informationen über Ereignisse zwischen Bundles, die nicht über Referenzen miteinander verbunden sein müssen. Die Protocol-Services ermöglichen den Zugriff auf verschiedene technische Protokolle wie http, um z.B. Bundles aus entfernten Quellen zu aktualisieren, sowie Universal Plug and Play und Device Management Trees, um dynamisch anschließbare Geräte wie z.B. MP3-Player oder Mobiltelefone ansprechen zu können. Unter den Miscellaneous-Services sind beispielsweise Dienste zum Verlinken von Bundles für einen gemeinsamen Ladevorgang beim Start des Frameworks oder ein XML Parser enthalten (OSGi Alliance 2008).

Die Sicherheitsebene baut im Wesentlichen auf den Sicherheitsmechanismen von Java auf und erlaubt einen vereinfachten Zugriff aus den Bundles heraus. Darüber hinaus sind insbesondere OSGi-spezifische Funktionen zur Signierung von Bundle-Dateien enthalten (OSGi Alliance 2007b, 11ff).

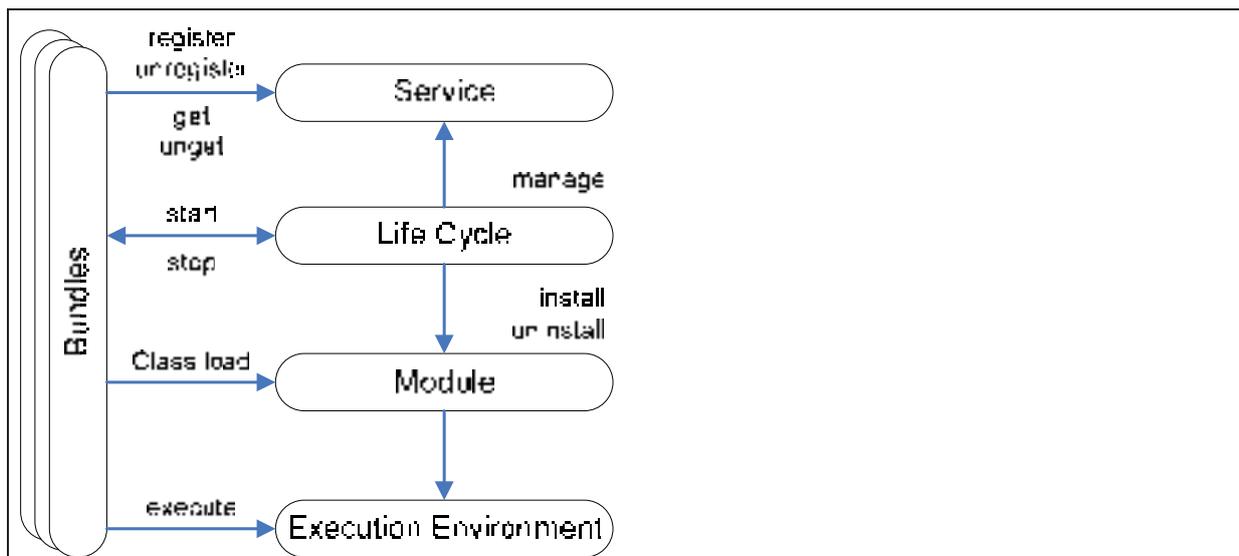


Abbildung 6-5: *Interaktion während des Bundle-Lebenszyklus*

Quelle: (OSGi Alliance 2007b, 3)

Die Bundles, welche die vorgestellten Bestandteile des OSGi Frameworks nutzen, haben eine Vielzahl von Interaktionen mit unterschiedlichen Komponenten und Services. Während viele Verbindungen von der jeweiligen Anwendung innerhalb eines Bundles abhängig sind, durchlaufen alle Bundles einige bestimmte Schritte während ihres Lebenszyklus im Framework. Sobald ein Bundle in das Framework aufgenommen und gestartet wird, erfolgt zunächst die Registrierung. Dadurch werden andere Bundles, die auf die Registrierung dieses Bundles warten, informiert und können das neu hinzugefügte Bundle referenzieren. Sobald der Startvorgang des neuen Bundles angestoßen wurde, wird das Archiv und die darin enthaltene Activator-Klasse des Bundles geladen und zur Ausführung an das Execution

Environment übergeben. Direkt nach der Registrierung eines neuen Bundles, erfolgt die Initialisierung des Bundles, was durch das Ausführen der Methode „start“ erfolgt. Dabei können durch das neue Bundle andere Bundles bzw. Services wie beispielsweise das Logging oder der Event Admin verwendet werden. Nach der Registrierung und Initialisierung kann das neue Bundle von anderen Bundles über entsprechende Ausführungsmethoden oder indirekt über Ereignisse angesprochen werden. Dies kann so lange erfolgen bis das Bundle wieder gestoppt wird. Dazu muss die Activator-Klasse jedes Bundle analog zur start-Methode eine stop-Methode enthalten, die referenzierte Bundles freigibt und gegebenenfalls verwendete Teile des Speichers wieder aufräumt. Nach der Beendigung der Bundle-Ausführung wird das Bundle deregistriert, so dass referenzierende Bundles darauf reagieren können (vgl. Abbildung 6-5).

6.1.2.3. Anwendung von OSGi auf die Systemarchitektur

Obwohl im Rahmen der konzeptionellen Systemarchitektur eine Vielzahl verschiedener Komponenten und Systeme identifiziert und in Zusammenhang gebracht wurden, ist es nicht sinnvoll, aus jeder Komponente ein eigenständiges OSGi-Bundle zu machen. Durch eine große Anzahl an Bundles nimmt die Menge an Interaktionen zwischen den Bundles zu und kann je nach Art und Umfang der ausgetauschten Informationen zu einem Leistungsabfall führen. Darüber hinaus ist zu berücksichtigen, dass Komponenten, die zur Laufzeit keine Rolle spielen, nicht zwangsläufig als OSGi-Bundles realisiert werden müssen. So können die meisten Komponenten der Designzeit mit beliebigen Rahmenwerken und Programmiersprachen erstellt werden, da ihre Schnittstelle zur Laufzeit primär im Format der Datencontainer besteht.

Von den Input Komponenten lässt sich der Spracheingabekanal mit der Spracherkennung und dem *language understanding* zu einem Spracherkennung-Bundle zusammenfassen. Es erkennt gesprochene Sprache und gibt den auf Grundlage einer Sprachgrammatik erkannten Text als Ergebnis zurück. Die Kanäle der externen Eingabegeräte und des externen Systems sind im Automobilumfeld ohne den Einbau weiterer Bedienelemente als gleichbedeutend anzusehen. Dies ist dadurch bedingt, dass die Benutzung von Bedienelementen wie beispielsweise das Aktivieren eines Schalters ebenso über das Bordsystem geleitet wird wie Warn- oder Fehlermeldungen. Dadurch können beide Eingabekanäle zum Bundle für die Bordsystemanbindung zusammengefasst werden. Durch die integrierte Signalkonvertierung kann auch von diesem Bundle als Ergebnis eine Nachricht in Textform ausgegeben werden. Da die Betätigung des Push-to-Talk-Buttons im Multifunktionslenkrad als Ereignis eines fahrzeuginternen Bussystems übermittelt wird, erfolgt auch das Anstoßen der Spracherkennung aus diesem Bundle heraus. Dadurch ist die Komponente des Input Managers zusammen mit der CAN-Bus Anbindung in einem Bundle zu finden.

Das Reaktionsermittlungssystem beinhaltet die Bundles für den Dialogmanager und den Chatbot. Zentrales Bundle der ordentlichen Verarbeitung ist das Chatbot-Bundle. Aufgrund des regelbasierten Ansatzes deckt es sowohl die Bereiche des *understanding modules*, des *response planner systems* als auch der *response generation* ab. Zustände der Umwelt d.h. des Fahrzeugs werden ebenfalls im Chatbot vorgehalten nachdem entsprechende Ereignisse vom Input Manager bei Statusänderungen versandt werden. Die Speicherung der

Statusinformationen im CAN-Bus-Bundle als Datenrekorder ist wenig sinnvoll, da der Austausch zwischen Chatbot und CAN-Bus-Bundle eine erhöhte interne Kommunikation erfordert und somit die Performance des Systems beeinflussen kann.

Im Output System lässt sich prinzipiell jeder Ausgabekanal als eigene Komponente erstellen, wobei die Vorstufe jeweils integriert ist. Aufgrund technischer Eigenschaften der verwendeten Avatar-Software ergibt sich der Bedarf, auch Komponenten im Output System einzubinden, die sich nicht als OSGi-Bundle umsetzen lassen. Der Präsentationsmanager, der als Bundle umgesetzt ist, nimmt stellvertretend für alle Ausgabekanäle die auszugebende Reaktion an und leitet sie an die Ausgabekanäle weiter. Dabei läuft die Abarbeitung durch die Ausgabekanäle sequentiell. Zunächst wird das GUI-Framework angesteuert, das seinerseits die Avatarkomponente ansteuert, welche schließlich synchronisiert die Sprachsynthese einbindet. Die drei Ausgabekanäle sind somit keine OSGi-Bundles, sondern Drittanbieter-Komponenten im Falle der Avatar- und Sprachsynthesoftware bzw. eine selbst erstellte HTML-Seite im Falle des GUI-Frameworks.

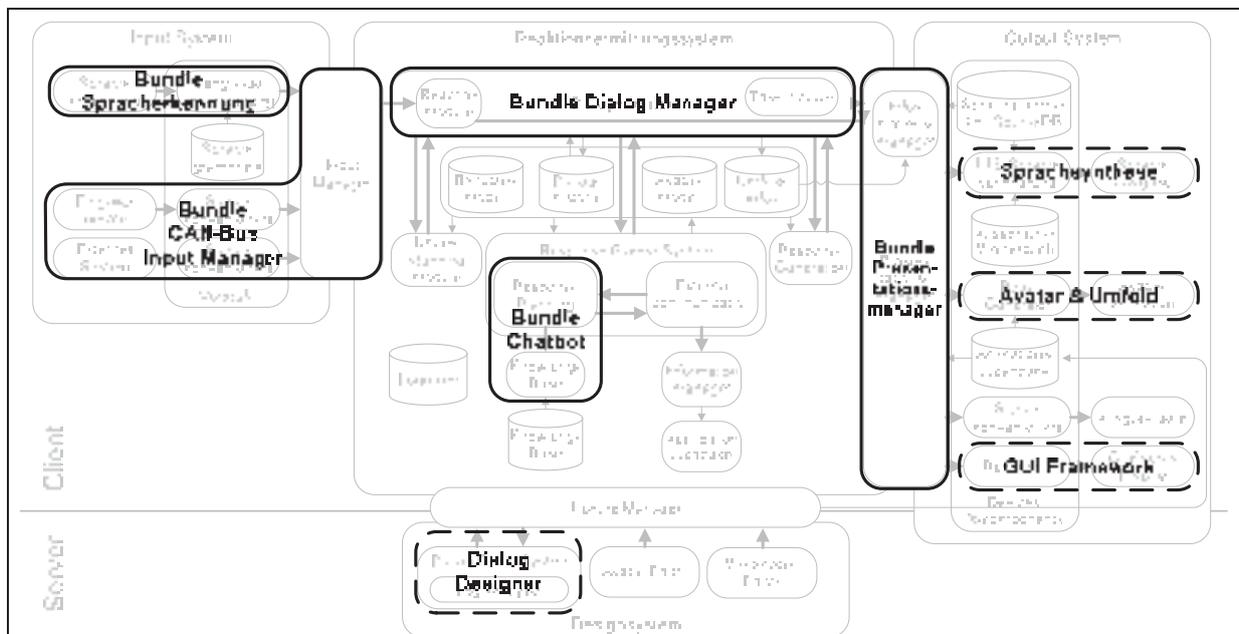


Abbildung 6-6: Aufteilung der Architektur in OSGi-Bundles und weitere Komponenten
Quelle: (Eigene Darstellung)

Der Dialogdesigner, der für die Erstellung der Knowledge Base zur Designzeit verwendet wird, stellt ebenfalls kein OSGi-Bundle dar, da er zur Laufzeit keine Rolle spielt. Ein Avatar Editor im eigentlichen Sinn besteht nicht, da die gewählte Lösung zur Avatardarstellung nur wenig Spielraum zur Gestaltung lässt (vgl. Kapitel 6.1.3.6). Einen Überblick über die zu entwickelnden Bundles bzw. weitere Elemente und ihre Abdeckung der Komponenten der konzeptionellen Systemarchitektur zeigt Abbildung 6-6.

6.1.3. Umsetzung der Komponenten

Die einzelnen Komponenten, die durch die Aufteilung der Systemarchitektur in OSGi-Bundles entstanden sind, sollen im Folgenden detailliert beschrieben werden. Dabei soll

sowohl auf die interne Arbeitsweise, die von außen ansprechbaren Funktionen und Schnittstellen und deren Integration in das Gesamtgefüge eingegangen werden. Der Vollständigkeit halber werden alle Softwarekomponenten der Laufzeit berücksichtigt, unabhängig davon, ob es sich um OSGi-Bundles oder damit verbundene Anwendungen handelt. Die Beschreibung erfolgt nach der Reihenfolge: Input, Reaktionsermittlung und Output. Die Designzeit-Komponente des Dialogeditors bzw. Dialogdesigners wird erst im Rahmen der Aufarbeitung von Inhalten in Kapitel 6.2.1 beschrieben, da er integraler Bestandteil der erarbeiteten Methode ist.

6.1.3.1. Spracherkennung

Da die Entwicklung einer eigenen Spracherkennungssoftware entweder den Umfang dieser Forschung gesprengt oder kaum angemessene Ergebnisse erzielt hätte, wurde ein bestehendes Softwarepaket verwendet. Dabei handelt es sich um VoCon 3200, das vom Hersteller Nuance (vormals ScanSoft) speziell für den Einsatz im Fahrzeug entwickelt wurde (Nuance 2008). Es wird bereits im Serieneinsatz bei Audi verwendet, so dass zum einen die Qualitätsansprüche erfüllt werden und zum anderen bereits Erfahrungen mit der Software vorliegen.

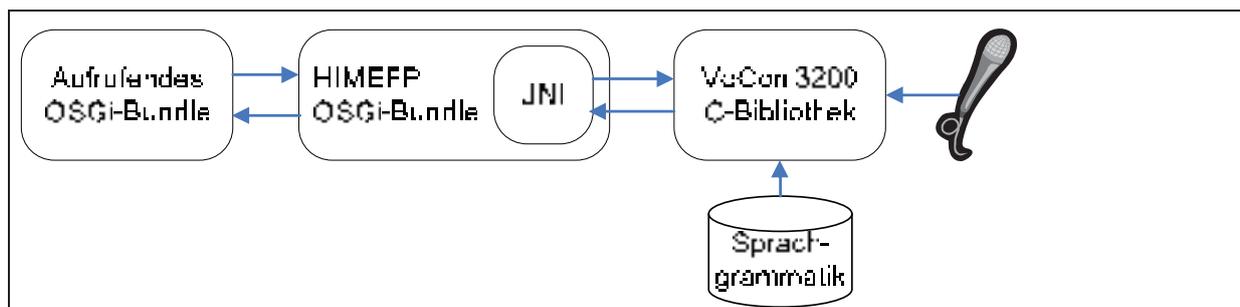


Abbildung 6-7: Verwendung der VoCon 3200 Bibliothek über JNI
Quelle: (Eigene Darstellung)

VoCon 3200 bietet eine umfangreiche Programmierschnittstelle, die es anderen Programmen ermöglicht, die Funktionen der Software zu nutzen (ScanSoft 2005). Die bereitgestellte Funktionsbibliothek liegt dabei jedoch als Funktionsbibliothek in der Programmiersprache C vor, so dass sie nicht ohne Weiteres aus Java heraus aufgerufen werden kann. Die Schnittstelle zwischen Java und der ursprünglichen C-Bibliothek erfolgt durch ein Java Native Interface (JNI), das ein Mapping zwischen C- und Java-Methoden beinhaltet. Dieses JNI wurde vom Volkswagen Electronic Research Lab (ERL) in Palo Alto entwickelt und im Rahmen eines Bestandteils der HIMEPP-Plattform in ein OSGi-Bundle integriert. Durch den Aufruf des HIMEPP-Bundles wird ein entsprechender Befehl an die VoCon-Bibliothek gesandt, der veranlasst, dass ab Eintreffen der Anweisung Sprachsignale zu erkennen versucht werden. Sobald eine Übereinstimmung in der Sprachgrammatik gefunden wird, wird diese über das JNI an das HIMEPP-Bundle und anschließend an das aufrufende Bundle zurückgeliefert (vgl. Abbildung 6-7). Wird kein Sprachsignal erkannt, weil der Benutzer nicht spricht oder Störgeräusche eine Erkennung unmöglich machen, wird die Eingabe „null“ erkannt. Die zur Erkennung verwendete Sprachgrammatik liegt im BNF+-Format vor (vgl. Kapitel 3.2.3) und wird beim Export aus dem Dialogeditor auf Basis der möglichen Benutzereingaben automatisch generiert.

Die Verwendung für eine natürlichsprachliche Kommunikation macht die Erkennung einer Vielzahl verschiedener und vor allem längerer Phrasen notwendig. Um die Erkennung dieser Eingabemöglichkeiten zuverlässig vorzunehmen, kann nur die Integration einer eigenen, eingeschränkten Sprachgrammatik als Ansatz verfolgt werden. Diese Sprachgrammatik enthält alle möglichen Formulierungen, die als Eingaben in der Knowledge Base des Chatbots vorhanden sind. Dadurch kann sichergestellt werden, dass nur Eingaben erkannt werden, die auch im späteren Verlauf verarbeitet werden können. Da im Rahmen des *language understanding*, das in VoCon 3200 integriert ist, die wahrscheinlichste, erkannte Eingabe aus der vorliegenden Sprachgrammatik zurückgemeldet wird, erfolgt dabei auch gleichzeitig in gewissem Rahmen eine Korrektur. Beispielsweise kann die gesprochene Phrase „Zeig mir mein Ziel an“ das Ergebnis „Zeig mir das Ziel an“ aus der Sprachgrammatik liefern, da in diesem Fall die Übereinstimmungswahrscheinlichkeit am höchsten war. Diese unbeabsichtigte Korrektur kann zum einen hilfreich sein und die umfangreiche Ausgestaltung von Sprachgrammatiken erleichtern. Auf der anderen Seite birgt sie jedoch die Gefahr, dass die Worte des Benutzers „verdreh“ werden. Insbesondere wenn kleine Details in einer Phrase den Sinn verändern, kann dies zu ungewollten Reaktionen führen. Somit muss die explizite Abdeckung möglichst vieler Formulierungen das primäre Ziel der Grammatikerstellung sein.

Die Aktivierung der Spracherkennung im Fahrzeug erfolgt durch die Betätigung des im Lenkrad integrierten Push-to-Talk-Buttons des Multifunktionslenkrads, das als Ausstattungsmerkmal für Audi-Fahrzeuge verfügbar ist. Das Signal des Knopfes wird über den CAN-Bus geleitet, der über ein weiteres OSGi-Bundle in das Avatarsystem eingebunden ist. Das CAN-Bus-Bundle, das auch als Input Manager fungiert, stößt somit die Spracherkennung an, sobald der Push-to-Talk-Button betätigt wurde und gibt den erkannten Inhalt an den Dialogmanager weiter. Während der Entwicklungsphase bestand für die Simulation ohne Fahrzeug auch die Möglichkeit, über eine provisorische Eingabemaske sowohl einen Push-to-Talk-Button zu simulieren als auch Text direkt einzugeben, der dem Dialogmanager übergeben wurde.

6.1.3.2. CAN-Bus und Input Manager

Das Bundle für die Anbindung des CAN-Bus im Fahrzeug ermöglicht zum einen das Einlesen von Ereignissen, die auf einem der CAN-Busse versendet werden und zum anderen prinzipiell auch das Versenden eigener Ereignisse. Durch die zentrale Bedeutung der CAN-Busse ist in diesem Bundle zudem die Funktion des Input Managers integriert, da die Steuerung der Spracheingabe ebenfalls durch dieses Bundle erfolgt. Die Umsetzung des Bundles beinhaltet weitere Bundles zum Zugriff auf CAN-Busse, die von der Fachabteilung AEV-3 erstellt wurden. Die Lauffähigkeit der für Knopflerfish in der Version 3 kompilierten Bundles unter OSGi 4 in der Equinox-Implementierung wird dabei durch die Überlagerung von Logging-Komponenten der AEV durch Elemente aus der HIMEPP-Plattform erreicht. Die Verarbeitung der eingehenden Nachrichten wurde für das Avatarsystem neu entwickelt.

Überblick wichtiger Bus-Systeme im Fahrzeug

In modernen Mittel- und Oberklassefahrzeugen wird eine Vielzahl verschiedener Bussysteme zur Kommunikation zwischen einzelnen Steuergeräten verwendet. Je nach

Anwendungsbereich und technischem Stand des Fahrzeugs kommen dabei vor allem vier Bussysteme zum Einsatz: CAN-, LIN-, MOST-Bus und Flexray.

Der CAN (Controller Area Network)-Bus wurde bereits in den 1980er Jahren von Bosch und Intel entwickelt, um digitale und analoge Punkt-zu-Punkt-Verbindungen zwischen Steuergeräten zu reduzieren und damit das Gewicht der benötigten Kabelbäume zu minimieren (Zimmermann/Schmidgall 2007, 37). Insbesondere bei europäischen, aber auch zunehmend bei amerikanischen Autoherstellern findet das System Anwendung. Es unterstützt dabei Übertragungsraten von bis 1 MBit/s, wobei diese von der Buslänge abhängen. Üblich sind 500 KBit/s für High Speed CAN und 125 KBit/s für Low Speed CAN, wobei die Datenraten nicht festgelegt sind und von Hersteller zu Hersteller variieren können (Zimmermann/Schmidgall 2007, 33ff). In so genannten CAN-Matrizen wird dabei in Tabellenform beschrieben, welches Steuergerät unter welchen Umständen, welche Daten mit welcher Priorität über welchen CAN-Bus versendet (Zimmermann/Schmidgall 2007, 37). Tabelle 6-1 zeigt die im verwendeten Testträger bestehenden CAN-Bussysteme und ihre wichtigsten Merkmale.

CAN-Bus	Geschwindigkeit	Beschreibung
Komfort	Low-Speed (100 Kbit)	Bussystem, das alle Status- und Ereignisinformation zu Bedienteilen des sekundären und tertiären Funktionsbereich im Fahrzeug übermittelt
Antrieb	High Speed (500 Kbit)	Übertragung kritischer Informationen zum Antrieb des Fahrzeugs (primärer Funktionsbereich)
Kombi	High Speed (500 Kbit)	Übertragung der Stati und Ereignisse des Kombiinstruments zur Anzeige von Geschwindigkeit, Motordrehzahl sowie von Meldungen im entsprechenden Display
Diagnose	High Speed (500 Kbit)	Gesonderter Bus, um Diagnosedaten aus den Fehlerspeichern des Fahrzeugs auszulesen

Tabelle 6-1: *CAN-Bussysteme im Audi A4 Testträger*
Quelle: (Eigene Darstellung)

Eine CAN-Nachricht besteht aus drei wichtigen Bestandteilen: einer ID, welche angibt, worauf sich die Nachricht bezieht; der Anzahl der genutzten Bits und die eigentlichen Daten in Form von bis zu 64 Bits. Jede Nachricht wird mit 64 Bits versendet, wobei jedoch nicht alle davon bei jedem Nachrichtentyp genutzt werden. Je nach Art der Nachricht, wird sie zyklisch im Abstand von einigen Millisekunden oder erst auf ein bestimmtes Ereignis wie das Betätigen eines Bedienelements hin versandt.

Da der CAN-Bus im gesamten Fahrzeug Anwendung findet und eine dementsprechend hohe Auslastung vorliegt, hat man mit dem Local Interconnect Network (LIN)-Bus versucht, eine kostengünstige Alternative zum Low Speed CAN umzusetzen. Sie findet beispielsweise für die Verkabelung innerhalb einer Tür oder eines Sitzes als eigenständiges System optimale Anwendung und entlastet damit den CAN-Bus (Zimmermann/Schmidgall 2007, 43). Ebenfalls stark am CAN-Bus orientiert sich auch Flexray, was versucht zum einen der geringeren Bandbreite zu begegnen und zum anderen die Steuerung der Übertragungsdauer zu ermöglichen. Mit einer Übertragungsrate von derzeit bis zu 10 MBit/s ist es hauptsächlich für

den Einsatz bei X-by-Wire-Systemen wie z.B. einer rein elektrisch gesteuerten Lenkung als Steer-by-Wire konzipiert, findet bisher jedoch noch kaum Anwendung (Zimmermann/Schmidgall 2007, 57ff).

Der Media Oriented Systems Transport (MOST)-Bus ist generell für Telematik- und Multimediadaten ausgelegt und erlaubt dadurch z.B. die Anbindung an das Autoradio, einen CD-Wechsler oder das Autotelefon. Da die übertragenen Daten sehr umfangreich sind, wird dabei eine größere Bandbreite mithilfe von Glasfaserleitungen auf Kosten der Übertragungssicherheit realisiert (Zimmermann/Schmidgall 2007, 69).

Funktionsweise des CAN-Bundles

Da die meisten Bedienelemente sowie kritische Informationen zu Warn- und Fehlermeldungen über den CAN-Bus übermittelt werden, stellt er das wichtigste Bussystem eines Fahrzeugs dar. Dies führt zur Fokussierung auf dieses Bussystem im Rahmen der Forschungsarbeit. Dem CAN-Bundle kommt dabei die Funktion eines Dispatchers zu: Es filtert zum einen die eingehenden Ereignisse und verarbeitet die gefilterten Ereignisse je nach ihren Eigenschaften auf drei verschiedene Weisen. Welche Verarbeitung bei welchem Signal erfolgt, wird dabei in einer zum Bundle gehörigen XML-Datei festgelegt, die in Abbildung 6-8 exemplarisch dargestellt ist. Für jeden der im Folgenden beschriebenen Verarbeitungstypen können verschiedene CAN-Nachrichten mit Ihren IDs und den entsprechenden Bitzuständen hinterlegt werden. Im Falle von Variablen und Nachrichten werden zudem weitere Angaben gemacht, die für das Mapping verwendet werden.

Einen besonderen Status haben Ereignisse, die den Push-to-Talk-Button betreffen. Durch solch ein Ereignis wird einerseits ein Ereignis im OSGi-Framework ausgelöst, das zur Anzeige des Mikrofon-Symbols im GUI führt und andererseits die Spracherkennung aktiviert. Obwohl nur ein Push-to-Talk-Button im Lenkrad enthalten ist, kann dieser je nach Dauer der Betätigung verschiedene CAN-Nachrichten versenden, so dass auch für diese Verarbeitungskategorie die Angabe mehrerer CAN-Nachrichten notwendig sein kann. Die von der Spracherkennung erkannte Phrase wird an das CAN-Bundle zurückgegeben und schließlich über das OSGi-Ereignis `de/avicos/SEND_MESSAGE_TO_MANAGER` an den Dialogmanager weitergeleitet.

Im Rahmen der Verwendung des Push-to-Talk-Buttons als Eingabetrigger ergaben sich im Laufe der Umsetzung zwei Herausforderungen. Zum einen stieß das Betätigen des Knopfes das werksseitig eingebaute Spracherkennungssystem des Fahrzeugs an, was zu ungewollten Reaktionen führte, die den Benutzer des Avatarsystems verwirrt hätten. Zum anderen führte das längere Drücken des Knopfes zur mehrfachen Versendung des relevanten CAN-Ereignisses. Das erste Problem konnte durch die Deaktivierung des bestehenden Spracherkennungssystems mithilfe einer Diagnosestation, wie sie auch in Vertragswerkstätten verwendet wird, behoben werden. Das wiederholte Versenden des Ereignisses wurde durch einen Timeout-Mechanismus im CAN-Bundle behoben. Dies führt dazu, dass nach dem ersten Empfangen des Ereignisses das gleiche Ereignis für den Push-to-Talk-Button erst nach einer ca. einer Sekunde wieder verarbeitet werden kann.

```
<ptt>
  <message>
    <text>PTT kurz druecken</text>
    <canid>66A</canid>
    <startbit>8</startbit>
    <mask>1</mask>
  </message>
</ptt>
<messages>
  <message>
    <text>CANReaction Sitzmemory Set</text>
    <canid>49a</canid>
    <startbit>38</startbit>
    <mask>1</mask>
  </message>
</messages>
<datarecorder>
  <variable>
    <text>CAN_Klima_Umluft</text>
    <value>aus</value>
    <CanID>66E</CanID>
    <startbit>3</startbit>
    <mask>0</mask>
  </variable>
</datarecorder>
```

Abbildung 6-8: *Aufbau der XML-Datei zur Steuerung der CAN-Signalverarbeitung*
Quelle: (Eigene Darstellung)

Die zweite Verarbeitungskategorie dient der Verarbeitung von Ereignissen durch den Chatbot. Die unter `<messages>` angegebenen Nachrichten führen dazu, dass der entsprechende Text über das OSGi-Ereignis `de/avicos/SEND_MESSAGE_TO_MANAGER` an den Dialogmanager gesendet wird und somit eine Reaktion im Chatbot hervorruft. Auch wenn keine technische Notwendigkeit dazu besteht, ist es sinnvoll den zu übergebenden Text mit einem Schlüsselwort zu versehen oder zu verschlüsseln, damit der Text und damit die entsprechende Reaktion nicht fälschlicherweise durch die Spracheingabe des Fahrers ausgelöst werden kann. Dies kann zudem dadurch sichergestellt werden, dass der entsprechende Eingabeschritt im Dialogdesigner, als stumm gekennzeichnet wird, d. h. der Text der Eingabe wird nicht in die Sprachgrammatik aufgenommen. Diese Art der CAN-Ereignisse wird insbesondere bei tutoriellen Szenarien verwendet, in denen mit dem Benutzer Schritt für Schritt bestimmte Tätigkeiten durchgegangen werden. Durch die Reaktion auf ein entsprechendes CAN-Ereignis kann in solchen Szenarien direktes Feedback auf das Verhalten des Fahrers gegeben werden. Reaktionen auf solche CAN-Ereignisse werden dabei jedoch nur im Zusammenhang von Tutorien, d.h. nicht als Einstiegspunkte in der Wissensbasis abgebildet (vgl. Kapitel 6.2.1.2), da sonst bei jedem entsprechenden CAN-Ereignis die Reaktion des Avatarsystems ausgelöst werden würde. Dennoch ist die Reaktion auf kritische

Warn- und Fehlermeldungen als Einstiegspunkt zu designen, damit in jedem Fall eine Hilfestellung dazu angeboten werden kann. Die Auslösekontrolle von CAN-Ereignissen wird dabei über die Darstellung des Hand-Symbols im GUI gesteuert. Nur wenn das Hand-Symbol angezeigt wird, wird überhaupt auf CAN-Ereignisse außer dem Push-to-Talk-Button reagiert.

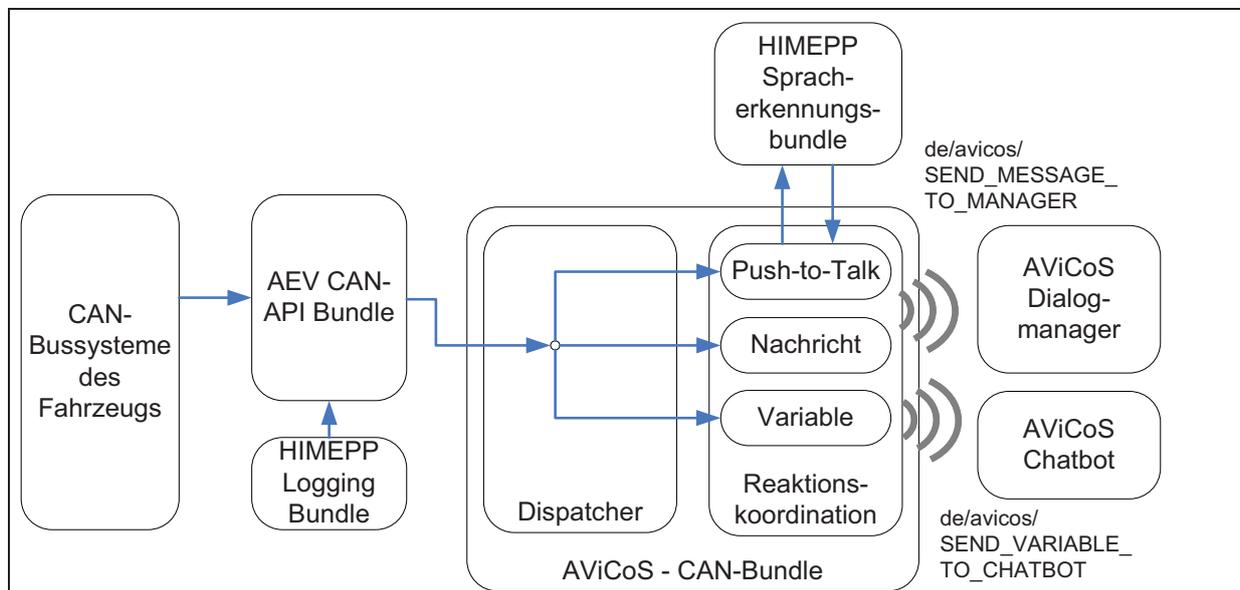


Abbildung 6-9: Verarbeitungsschema des CAN-Bundles

Quelle: (Eigene Darstellung)

Die dritte Art der Verarbeitung ist die Konvertierung von CAN-Ereignissen zu Werten von Variablen im Chatbot. Dabei wird bei Eingang einer definierten CAN-Nachricht das OSGi-Ereignis `de/avicos/SEND_VARIABLE_TO_CHATBOT` unter Angabe des Variablennamens und des neuen Wertes direkt an den Chatbot versendet. Im Chatbot wird die entsprechende Variable mit dem neuen Wert belegt. Handelt es sich dabei um eine Variable, die bei der Initialisierung des Chatbots als persistent angegeben wurde, so wird der neue Wert der Variable ebenfalls persistiert (vgl. dazu Kapitel 6.1.3.4). Alle anderen Variablen werden zur Laufzeit durch das entsprechende Ereignis erst angelegt und nicht persistent gespeichert. Die Unterscheidung in persistierte und volatile Variablen erlaubt die Abbildung der erwarteten Änderungshäufigkeit. So kann es Sinn machen, zu speichern, ob z.B. Audi Lane Assist aktiviert ist, während die aktuelle Fahrzeuggeschwindigkeit nicht gespeichert werden sollte, sondern nur zur Laufzeit verwendet wird. Das Verarbeitungsschema des CAN-Bundles mit den drei genannten Verarbeitungsvarianten ist in Abbildung 6-9 zu sehen.

Designoptionen

Die ursprüngliche Konzeption der Variablenverarbeitung sah vor, dass Variablen und ihre Werte im CAN-Bundle selbst zur Laufzeit gespeichert werden und vor der Reaktionsermittlung des Chatbots abgefragt werden. Aufgrund der ereignisbasierten Kommunikation zwischen den Bundles hätte dies zu einer Vielzahl von Ereignissen geführt, die zu diesem Zweck ausgetauscht worden wären. Da die Variablen und ihre Werte ohnehin im Chatbot bereits vorgehalten werden müssen, verringert die oben beschriebene Lösung die Kommunikationsmenge zwischen den Bundles und führt somit nicht zu

Performancebeeinträchtigungen. Sollten im Rahmen einer Erweiterung des Avatarsystems weitere Bundles auf solche Variablen zugreifen müssen, so kann die Einrichtung eines zentralen Datenrekorders im CAN-Bundle durchaus Sinn ergeben.

Die Anbindung an die CAN-Busse des Fahrzeugs kann neben der Verwendung als Eingabekanal auch als Ausgabekanal genutzt werden, um Funktionen des Fahrzeugs zu steuern. Dabei trifft man bei prototypischen Entwicklungen mithilfe des CAN-Busses auf den Umstand, dass das CAN-Protokoll nicht korrekt mit Nachrichten umgehen kann, die nicht bereits in der ursprünglichen Konzeption des Fahrzeugs in der CAN-Matrix vorgesehen waren. Das als „babbling idiot“-Verhalten bekannte Problem kann sogar dazu führen, dass das gesamte Bussystem lahm gelegt wird, wenn unkoordiniert Nachrichten abgesetzt werden (Anderl 2005, 22). Aufgrund des hohen damit verbundenen Sicherheitsrisikos wird die CAN-Bus-Anbindung im Avatarsystem nicht als Ausgabekanal genutzt.

6.1.3.3. Dialogmanager

Die Aufgaben eines Dialogmanagers innerhalb eines Avatarsystems hängen stark vom Einsatzgebiet ab. So schreiben Cassell et al. (1998) dieser Komponente die Organisation eines Dialogs zu, wobei Dialoge durch das System initiiert werden, in Ruhephasen das Wort ergriffen und auf alle Eingaben des Benutzers ein Feedback gegeben werden soll bzw. Rückfragen erfolgen. Diese Ansprüche beruhen auf der Beobachtung von Mensch-Mensch-Interaktion (Zue/Glass 2000, 1170). Die Zuweisung dieser Aufgaben setzt jedoch zum einen voraus, dass das Avatarsystem jederzeit von sich aus einen Dialog beginnen darf und dass der Benutzer den Dialog mit dem System als seine Hauptaufgabe in der entsprechenden Situation sieht. Beim Einsatz in einem Fahrzeug soll ein Avatarsystem jedoch nicht intrusiv sein, d.h. den Fahrer ohne schwerwiegenden Grund von seiner Hauptaufgabe, dem Fahren, ablenken. Beim Einsatz in einem Fahrzeug erfolgt eine Ausgabe des Avatarsystems somit nur bei einer entsprechenden Eingabe durch den Benutzer, die durch das Betätigen des Push-to-Talk-Buttons angekündigt wird, oder ein schwerwichtiges Ereignis in einem der Bussysteme. Durch diesen Umstand und die Verwendung eines Chatbots, der implizit auch die Aufgaben des *understanding modules* und der *response generation* übernimmt, gestaltet sich der Ablauf innerhalb des Dialogmanagers überschaubar.

Die Verarbeitung innerhalb des Dialogmanagers beginnt mit dem Eingang des Ereignisses „de/avicos/SEND_MESSAGE_TO_MANAGER“, das zugleich die Eingangsnachricht enthält. Im *reaction module* wird auf Basis der Steuerinformationen des *task modules* und des Nachrichteninhalts entschieden, wie die Eingabe zu verarbeiten ist. Im Falle einer priorisierten Verarbeitung wird direkt eine entsprechende Ausgangsnachricht generiert und über das Ereignis „de/avicos/SEND_MESSAGE_FROM_MANAGER“ an alle für dieses Ereignis registrierten Komponenten versendet. Diese priorisierte Verarbeitung findet beispielsweise statt, sobald der Push-to-Talk-Button betätigt wurde, um dem Benutzer ein Feedback in Form eines entsprechenden Symbols im Frontend bzw. einer Animation des Avatars zu geben. Neben der Anzeige des Mikrophon-Symbols wird zudem ein gegebenenfalls laufender Timeout-Thread abgebrochen. Enthält die eingehende Nachricht ein CAN-Ereignis, das mit dem Schlüsselwort „CANReaction“ gekennzeichnet ist, so wird überprüft, ob durch die letzte gegebene Antwort des Chatbots die Anzeige des Hand-Symbols bewirkt wurde. Nur

wenn das Hand-Symbol angezeigt wird, wird das eingehende CAN-Ereignis zur Verarbeitung an den Chatbot weitergegeben.

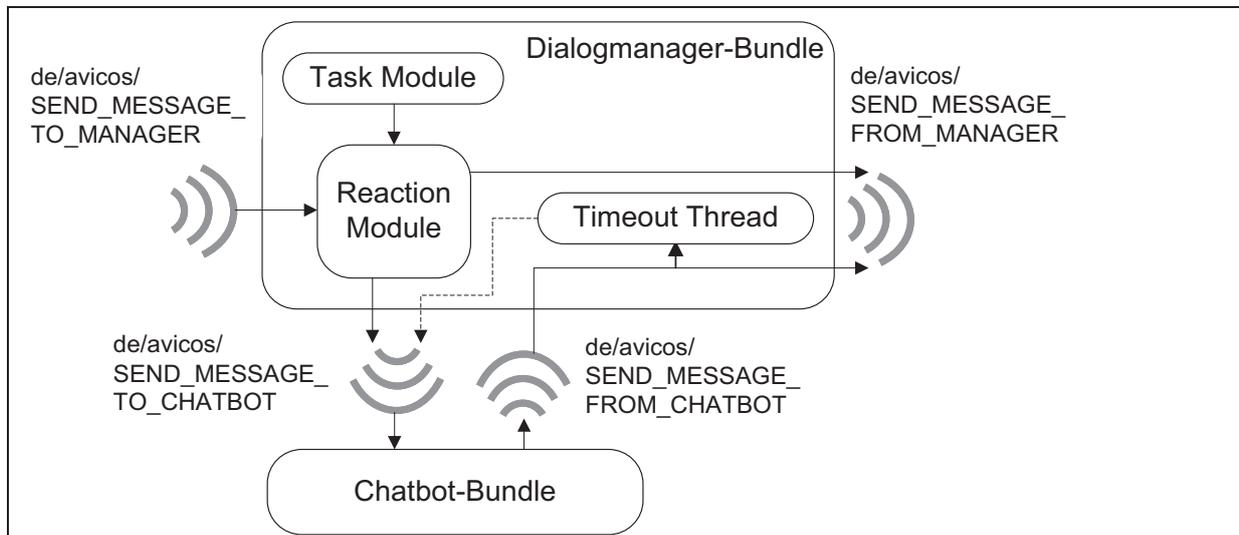


Abbildung 6-10: Steuerung der Antwortermittlung im Dialogmanager
Quelle: (Eigene Darstellung)

Die ordentliche Verarbeitung einer Eingangsnachricht wie beispielsweise eine akustische Eingabe des Benutzers oder das Ereignis eines Bussystems werden an den Chatbot mithilfe des Ereignisses „de/avicos/SEND_MESSAGE_TO_CHATBOT“ gesendet. Die ermittelte und gegebenenfalls konvertierte Antwort wird über ein entsprechendes Antwortereignis zurückgegeben an das Dialogmanager-Bundle. Enthält die Antwort des Chatbots eine Timeout-Funktion, so wird ein gesonderter Timeout-Thread im Dialogmanager eröffnet, der nach der vorgegebenen Timeout-Zeit die entsprechende Eingabe an den Chatbot sendet. Dadurch wird eine erneute Antwortermittlung im Chatbot angestoßen, die ihrerseits eine Antwort zurückgibt. Die Timeout-Funktion ist vor allen Dingen dann sinnvoll, wenn ein umfangreicher Teil des Handbuchs erklärt werden soll und er dazu in kleine Teile zerlegt wurde. Die Aufteilung dient dabei der besseren Wahrnehmung und Verarbeitung durch den Benutzer, da akustischen Informationen nur wenige Informationchunks entnommen werden können (Salmen 2002, 37). Die Steuerung der einzelnen Teile kann durch den Benutzer übernommen werden oder erfolgt automatisch nach einer gewissen Zeit. Dabei kann der Benutzer die Zeit als Option genau bestimmen oder die Timeout-Funktion gegebenenfalls ganz abschalten. Die beschriebenen Abläufe innerhalb des Dialogmanagers sind in Abbildung 6-10 dargestellt.

6.1.3.4. Chatbot

Die Betrachtung bestehender Chatbot-Software vermittelt ein zweigeteiltes Bild, bei dem kommerzielle, eigenständige Produkte auf der einen und Open Source Implementierungen auf der anderen Seite zu erkennen sind. Während eine Vielzahl kommerzieller Anbieter ausgemacht werden kann (z.B. von Wendt 2003; Samuelson 2003), wird die Open Source Entwicklung von der Software AliceBot (<http://www.alicebot.org>, zugegriffen am 02.05.2008) dominiert, die auch in verschiedenen wissenschaftlichen Anwendungen zum

Einsatz kam (vgl. Möbus et al. 2006; Braun et al. 2003). Neben der Offenheit von AliceBot stellt insbesondere auch die Definition des AIML-Standards für die Abbildung der Wissensbasis einen Vorteil dar. Im Vergleich zu proprietären, kommerziellen Formaten stellt dieser Standard eine einfach zu gestaltende und nachvollziehbare Form der Speicherung dar (vgl. als Gegenbeispiel von Wendt 2003, 43). Trotz der Einfachheit lassen sich mithilfe von Themengebieten (*topic-Tag*) und der Dialoghistorie (*that-Tag*) Informationen in die Antwortermittlung einbinden, die auch die Organisation großer Regelbasen ermöglicht. Zudem lassen sich in eingeschränktem Maße themenbezogene Regelbasen aus dem Internet beziehen und Fragen rund um die Regelerstellung durch zahlreiche Dokumente bzw. in einer Community klären.

Die Wahl von AliceBot bietet sich auch hinsichtlich der technischen Implementierung an, da die Funktionsweise von AliceBot, die auf den ersten Chatbot Eliza zurückgeht (Weizenbaum 1966), in einer Vielzahl von AIML-Interpretern umgesetzt wurde, die auf verschiedenen Programmiersprachen und Paradigmen basieren. Neben Python, Perl, Ruby, C/C++, PHP, .Net und anderen stehen auch verschiedene Umsetzungen in Java zur Verfügung (vgl. <http://www.alicebot.org/downloads/programs.html>, zugegriffen am 02.05.2008). Neben dem ersten noch verfügbaren Java-basierten AIML-Interpreter Programm D, das für den Webeinsatz mit einem integrierten Webserver oder als Servlet vorliegt, stellt das so genannte ChatterBean eine als Java Bean gekapselte Umsetzung dar. Dadurch lässt sie sich als Bestandteil einer Java-Anwendung einbinden, wie dies beispielsweise bei einem OSGi-Bundle der Fall ist. Die einfache Integration und die vollständige Kontrolle über die API führen dazu, dass das ChatterBean als AIML-Interpreter in die entsprechende OSGi-Komponente eingebunden wird. Bevor auf die Integration des ChatterBeans in das OSGi-Bundle eingegangen werden soll, bietet sich die Betrachtung der Eigenheiten des ChatterBeans insbesondere im Vergleich zur Referenzimplementierung durch Programm D an.

AliceBot ChatterBean

Durch die Kapselung der Interpreterfunktion ist der Funktionsumfang auf die Verarbeitung von AIML-Regeln fokussiert. Dies beinhaltet das Laden und Bereitstellen einer konkreten Bot-Instanz, die individuelle Merkmale hat. So werden für jede Bot-Instanz so genannte *bot predicates*, Konstanten der Instanz, so genannte *splitter* zur Identifizierung eines Satzendes bei Eingaben mit mehreren Sätzen und *substitutions* für jede Bot-Instanz eingelesen. Die *substitutions* ermöglichen insbesondere bei textbasierten Eingaben die Konvertierung von Rechtschreibfehlern, die Normalisierung von Zahlen oder die Umwandlung von Umlauten. Zu diesen instanzspezifischen Daten gehört auch die Regelbasis, die aus beliebig vielen AIML-Dateien bestehen kann. Auf der Grundlage dieser eingelesenen Regeln kann auf eine bestimmte Eingabe hin die passende Ausgabe ermittelt werden, wobei die *bot predicates*, die automatisch gespeicherte Dialoghistorie sowie das Themengebiet herangezogen werden können. Die Eingaben, die an das ChatterBean übergeben werden, sowie die Antworten, die ermittelt werden, werden automatisch für jede Instanzierung des ChatterBeans gesondert in einer Log-Datei gespeichert. Prinzipiell ist es dadurch möglich, mehrere Instanzen eines Bots parallel zu betreiben.

Über die durch das ChatterBean bereitgestellte API kann ein aufrufendes Programm zugreifen und die genannten Daten auslesen. Eine Änderung an diesen Werten ist während der Laufzeit nicht mehr möglich. Neben den *bot predicates* können jedoch auch *user predicates* verwendet werden, die sowohl durch Regeln als auch durch das aufrufende Programm gelesen und verändert werden können. Während in Programm D diese *user predicates* in Sets jeweils für verschiedene Benutzer gespeichert und bei einem Benutzerwechsel geändert wurden, steht diese Funktion im ChatterBean nicht zur Verfügung. Eng damit verbunden ist auch die Tatsache, dass eine Registrierung bzw. Anmeldung sowie eine automatische Identifizierung verschiedener Benutzer aufgrund der IP-Adresse ebenfalls nicht genutzt werden können. Dies ist der Hauptunterschied zwischen Programm D und dem ChatterBean, der sich durch die verschiedenen Einsatzbereiche begründen lässt. Während Programm D für den Mehrbenutzereinsatz über das Internet konzipiert wurde und dementsprechend auch über einen integrierten Webserver verfügt, ist das ChatterBean als universell integrierbarer Bestandteil konzipiert.

Das verwendete ChatterBeans in der Version 00.008 kann trotz des frühen Stadiums bereits alle wichtigen AIML-Tags identifizieren und verarbeiten. Dabei werden jedoch einige wenige Tags der AIML-Spezifikation 1.0.1 wie `<gender>` und `<person2>` noch nicht unterstützt. Das `gender`-Tag ermöglicht es, in einem Text geschlechtsbezogene Nennungen wie z.B. „sie“ in „er“ umzuwandeln. Dabei wird lediglich eine entsprechende Liste mit Konvertierungspaaren durchlaufen. Das gleiche Prinzip nutzt auch das `person2`-Tag, welche Verbenflexionen des ersten Falls Singular in den zweiten Fall Singular konvertiert, damit der Chatbot die Eingabe des Benutzers aufgreifen und zurückgeben kann. Ein Beispiel dafür ist die Eingabe „Ich fühle mich heute gut“ und die entsprechende Antwort des Chatbots „Wieso fühlst du dich heute gut?“. An diesem Beispiel wird auch die Problematik dieses Tags bei Verwendung mit der deutschen Sprache klar. Während bei Verwendung der englischen Sprache lediglich „i“ in „you“ umgewandelt werden muss, ändert sich im Deutschen auch die Form des Verbs. Das Fehlen dieser beiden Tags stellt im Deutschen somit keine ernst zu nehmende Einschränkung dar.

Herausforderungen beim Einsatz von AliceBot im Avatarsystem

Neben den beiden fehlenden Tags lässt sich jedoch ein Fehlverhalten im Umgang mit dem Tag `<that>` erkennen. Wird in einer Regel als Antwort mehr als ein Satz ausgegeben und sind die beiden Sätze durch einen Punkt als Trennzeichen getrennt, so kann diese Antwort nicht in nachfolgenden Regeln als Bezug genutzt werden. Abbildung 6-11 zeigt zwei solche Regeln: Unabhängig von der ersten Eingabe wird immer die Antwort „Erster Satz. Zweiter Satz.“ zurückgegeben. Auf jede weitere Antwort sollte als Antwort „Zweite Antwort“ zurückgegeben werden, was jedoch nicht erfolgt. Dieser Fehler kann einerseits dadurch umgangen werden, indem alle Punkte beispielsweise in Kommata umgewandelt werden, da bei einem Komma dieses Verhalten nicht zu beobachten ist. Der Austausch der Punkte führt jedoch dazu, dass die automatische Steuerung der Prosodie im Rahmen der Sprachsynthese nicht korrekt arbeiten kann. So wird im Falle eines Kommas die Stimme nicht soweit gesenkt, wie dies bei einem Punkt und somit einem Satzende erfolgt. Neben einer Änderung des Quellcodes des ChatterBeans, die sich aus Gründen der späteren Aktualisierung mit neuen Versionen als ungünstig erweisen kann, hat sich folgende Lösung ergeben, die auch

umgesetzt wurde. Durch die Löschung aller Splitter in der entsprechenden XML-Datei des ChatterBeans konnte das Problem behoben werden. Splitter werden bei der textbasierten Eingabe dazu genutzt, um mehrere Sätze aufzutrennen und getrennt zu verarbeiten. Da bei einer Spracheingabe zum einen keine Satzzeichen erkannt werden können und gesprochene Eingaben i.d.R. kürzer sind als Geschriebene, stellt die Löschung der Splitter keinen Nachteil dar. Der Zusammenhang zwischen Splitter und that-Tag ist jedoch nicht logisch nachvollziehbar, zumal im vorgestellten Beispiel auch ein Eintrag wie `<that>ZWEITER SATZ</that>` zu keinem korrekten Verhalten führte.

```

<category>
  <pattern>*</pattern>
  <template>
    Erster Satz. Zweiter Satz.
  </template>
</category>

<category>
  <pattern>*</pattern>
  <that>ERSTER SATZ ZWEITER SATZ</that>
  <template>
    Zweite Antwort.
  </template>
</category>

```

Abbildung 6-11: AIML Regeln zur Veranschaulichung des That-Problems
Quelle: (Eigene Darstellung)

Ein letztes Problem, das jedoch unabhängig vom verwendeten Interpreter auftritt, ist die Integration weiterer Befehle in den Antwortteil einer AIML-Regel. Diese Befehle werden beispielsweise genutzt, um eine bestimmte Avataranimation abzuspielen oder das Frontend zu steuern. Die Verwendung von XML-Tags, die nicht in den AIML-Spezifikationen enthalten sind, führen zu einer entsprechenden Fehlermeldung, da vom ChatterBean versucht wird, die Tags als AIML-Tags zu verarbeiten. Ein ähnliches Verhalten kann auch erkannt werden, wenn beispielsweise Animationsbefehle in ihrer ursprünglichen Syntax eingefügt werden, z.B. in der Form `$(DoMotion, greet/greet04)`. Kommt solch ein Befehl nur einmal in einer Antwort vor, wird er verarbeitet; bei einem zweiten Vorkommen führt dies jedoch zu einem Fehler. Diesem Verhalten kann zum einen dadurch begegnet werden, dass das ChatterBean um entsprechende Java Klassen für die neuen Tags erweitert wird. Da dies zu einer Änderung des AIML-Standards führt, ist dies nicht anzustreben. Vielmehr ist zum anderen eine gesondert für dieses Avatarsystem zu verwendende Syntax für zusätzliche Befehle zu bevorzugen. So werden für zusätzliche Befehle gesonderte Markierungen eingesetzt: `%%AV` für den Beginn eines Befehls und `AVE%%` für das Ende eines Befehls. Innerhalb dieser Markierungen gibt ein weiteres Schlüsselwort an, um welche Art von Befehl es sich handelt und ein Attribut gibt weitere Informationen an. Die Syntax und eine Kurzbeschreibung der zusätzlichen Befehle ist in Tabelle 6-2 zu sehen.

Beschreibung	Syntax	Beispiel	Konvertierte Syntax	Verarbeitungs-komponente
Avataranimation: spielt ein Animationskript ab, das im Frontendverzeichnis „character/motions“ liegt	%%AV ANI NAME=“...“ AVE%%	%%AV ANI NAME=“demoltouch03“ AVE%%	\$(DoMotion, <Dateiname>)	Avatar-Komponente
PlayMedia: Zeigt ein Bild, eine Bildsequenz oder ein Video an. Beim Fehlen einer Dateiendung wird im Chatbot Bundle nach voreingestellten Geschwindigkeitsschwellwerten und verfügbaren Medien selbst entschieden, ob ein Bild, eine Sequenz oder ein Video angezeigt wird. Die Pfadangabe beziehen sich auf das Frontendverzeichnis „images“ bzw. „video“	%%AV MEDIA NAME=“...“ AVE%%	%%AV MEDIA NAME=“video3.avi“ AVE%%	Unverändert	GUI-Komponente
Timeout: nach durch den Benutzer eingestelltem Timeout wird automatisch ein Befehl an Chatbot versandt	%%AV TIME ACTION=“...“ AVE%%	%%AV TIME ACTION=“WEITER“ AVE%%	Unverändert	Dialogmanager
Frontendsteuerung: Einblenden von Status- und Kopfinformationen. Die meisten Symbole werden automatisch eingebunden.	%%AV GUI ACTION=“...“ AVE%%	%%AV GUI ACTION=“WEITER“ AVE%%	Unverändert	Frontend-Komponente
3D-Objekt: Ruft eine CharaScript-Funktion zur Anzeige und Steuerung eines 3D-Objekts auf.	%%AV 3DO ACTION=“...“ AVE%%	%%AV 3DO ACTION=“showAudi“ AVE%%	\$(<CharaScript- Funktion>)	Avatar-Komponente

Tabelle 6-2: *Syntax für zusätzliche Befehle des Avatarsystems innerhalb von AIML*
Quelle: (Eigene Darstellung)

Die zusätzlichen Befehle werden nicht vom Chatbot selbst sondern von verschiedenen anderen Komponenten verarbeitet. Im Rahmen des Dialogmanagers, an den die Antwort des Chatbots übermittelt wird, wird je nach Art des Befehls die Syntax in die native Syntax der Komponente umgewandelt, die den Befehl verarbeitet. Die einzige Ausnahme dabei stellt der Timeout Befehl dar, der im Dialogmanager selbst dazu führt, dass er nach einer bestimmten Zeit eine Eingabe an den Chatbot sendet. Auf die genaue Verwendung und Arbeitsweise der einzelnen Befehle wird im Rahmen von Kapitel 6.2.1 näher eingegangen.

Da die Problematik der spitzen Klammern auch die Integration von Tags zur Steuerung von Phonetik und Prosodie bei der Sprachsynthese erschwert, erfordern auch diese Befehle eine Konvertierung. Dabei wird bei der Erstellung der Inhalte im Dialogeditor die ursprüngliche Syntax der Sprachsynthese verwendet, beim Speichern in AIML konvertiert und nach Rückgabe der entsprechenden Reaktion durch den Chatbot wieder in die ursprüngliche Syntax rücküberführt. Der Benutzer des Dialogeditors wird somit nicht mit der Problematik der Konvertierung konfrontiert. Diese Konvertierung erfolgt dabei jedoch nur für Befehle, die im Rahmen des Avatarsystems eine wichtige Rolle übernehmen (vgl. Tabelle 6-3). Eine

Übersicht aller Phonetik- und Prosodie-Befehle der verwendeten Sprachsynthese erfolgt in Kapitel 6.1.3.7.

Beschreibung	Syntax	Beispiel	Konvertierte Syntax
Phonetik: Steuerung der Aussprache von Wörtern	%%AV PHON ACTION="..." AVE%%	Hello, Mister %%AV PHON ACTION=" braUn" AVE%% Braun	Hello, Mister <phon sampa="braUn"/> Braun.
Prosodie - Pause: Steuerung der einer Pause in einer Phrase	%%AV BREAK TIME="..." AVE%%	Sentence one. %%AV BREAK TIME="2s" AVE%% Sentence two.	Sentence one. <break time="2s"/> Sentence two.
Prosodie - Tonhöhe: Steuerung der Tonhöhe in einer Phrase	%%AV PITCH LEVEL="..." AVE%%	Hello, %%AV PITCH LEVEL="140" AVE%% Miss Jones %%AV PITCH LEVEL="100" AVE%% arrived	Hello, <pitch level="140"> Miss Jones </pitch> arrived.
Prosodie - Geschwindigkeit: Steuerung der Geschwindigkeit in einer Phrase	%%AV SPEED LEVEL="..." AVE%%	Hello, %%AV SPEED LEVEL="300" AVE%% Miss Jones %%AV SPEED LEVEL="100" AVE%% arrived	Hello, <speed level="300"> Miss Jones </speed> arrived.
Prosodie - Lautstärke: Steuerung der Lautstärke in einer Phrase	%%AV VOL LEVEL="..." AVE%%	Hello, Miss %%AV VOL LEVEL="50" AVE%% Jones %%AV VOL LEVEL="100" AVE%% arrived	Hello, Miss <volume level="50"> Jones </volume> arrived.

Tabelle 6-3: Konvertierung von Phonetik- und Prosodieanweisungen im Avatarsystem
Quelle: (Eigene Darstellung)

Umsetzung des Chatbot-Bundles

Die Einbindung des ChatterBeans in das Chatbot-Bundle erübrigt die Entwicklung eines eigenen Algorithmus zur Antwortfindung. Dennoch muss die Verwaltung der *user predicates* ergänzt werden, um die Werte der Predicates über mehrere Sessions persistent zu halten. Beispielsweise sollen so Einstellungen, die vom Benutzer hinsichtlich der Anpassung des Avatarsystems vorgenommen wurden, auch über mehrere Gespräche hinweg erhalten bleiben.

Neben diesen *user predicates*, die aus den Regeln heraus mit Daten befüllt werden, stellen die aktuellen Zustände des Fahrzeugs eine Herausforderung dar. Diese Zustände können im Rahmen der Regeln für die Entscheidungsfindung verwendet werden. So könnte beispielsweise eine Reaktion nur ab oder bis zu einer gewissen Geschwindigkeit erfolgen oder die Antwort einer Regel enthält eine Rückmeldung über den aktuellen Ölstand. Zur Integration dieser Werte, die von den Bussystemen des Fahrzeugs stammen, sind drei Vorgehen denkbar:

1. Die Einbindung von Fahrzeugdaten in AIML-Regeln kann über die Implementierung eines gesonderten Tags erfolgen. Die Ausführung dieses Tags könnte dabei den aktuellen Zustand eines Bedienteils bei der Verarbeitung der Antwort ermitteln und einfügen. Dabei wird jedoch die standardisierte AIML-Spezifikation verletzt und verhindert die Verwendung der Regeln in anderen AIML-Interpretern.

2. Der zweite Lösungsansatz sieht eine zweistufige Antwortgenerierung vor. So werden anstelle der notwendigen Fahrzeugdaten entsprechende Platzhalter eingebunden, die außerhalb des ChatterBeans durch die realen Fahrzeugdaten ersetzt werden. Dieses Vorgehen gestaltet sich für den Fall, einen Wert lediglich ausgeben zu wollen, einfach. Werden Fahrzeugdaten jedoch für die Entscheidungsfindung im Rahmen einer Regeln verwendet, muss eine provisorische Antwort vom ChatterBean zurückgeliefert werden, die dazu führt, dass eine veränderte Eingabe generiert und wieder an das ChatterBean zurückgegeben wird, die den konkreten Wert des Platzhalters enthält. Aufgrund dieser zweiten Eingabe, erfolgt dann schließlich die finale Antwortermittlung. Dieser Weg führt zu einer Vermehrung der Regelanzahl und erhöht die Komplexität des Regelwerkes dermaßen, dass es nur noch schwer nachvollzogen werden kann.
3. Der dritte Ansatz bedarf weder neuer Tags noch weiterer Regeln. Er basiert darauf, dass beim Aufruf des Chatbot-Bundles die aktuellen relevanten Fahrzeugdaten aus einem Datenrekorder ausgelesen und dem ChatterBean übermittelt werden. Dadurch liegen die Fahrzeugdaten als gewöhnliche *user predicates* vor, wenn die Benutzereingabe im ChatterBean eingeht. Die Entscheidungsfindung bzw. die Einbindung der Fahrzeugdaten erfolgt dadurch wie bei gewöhnlichen *user predicates* auch. Wie bereits beim CAN-Bundle beschrieben wäre die Implementierung eines gesonderten Datenrekorders aus Sicht der Systemarchitektur das korrekteste Vorgehen. Da durch die ereignisbasierte Kommunikation zwischen den Bundles eine hohe Last entstehen würde, kann die Speicherung der aktuellen Wertzustände auch direkt im Chatbot erfolgen, wo sie ohnehin zur Verwendung in den Antworten vorgehalten werden müssen.

Durch die Wahl der dritten Möglichkeit ergeben sich die in Abbildung 6-12 dargestellten Ablaufsequenzen. Dabei ist zu beachten, dass das Chatbot-Objekt zusammen mit der Predicate Persistenz und dem ChatterBean zusammen ein OSGi-Bundle bilden. Im dargestellten Diagramm sind zum einen die Initialisierung und zum anderen die Antwortermittlung dargestellt. Bei der Initialisierung des Chatbot-Bundles erfolgt auch die Initialisierung der Predicate Persistenz. Dabei wird zunächst die entsprechende Datei mit den bereits gespeicherten *user predicates* und ihren Werten ausgelesen und gegebenenfalls um neu konzipierte *user predicates* ergänzt. Diese neuen *predicates* befinden sich in einer Referenzdatei, die alle persistenten *user predicates* enthält und auf der Grundlage der Modellierung im Dialogdesigner erstellt wird. Auf diese Weise hinzugefügte *predicates* haben den Defaultwert „undefiniert“. Die Liste aus *predicates* mit ihren Werten wird an die Chatbot-Komponente zurückgegeben und dort dazu verwendet, die *predicates* im ChatterBean entsprechend anzupassen bzw. zu erzeugen.

Die Sequenz zur Antwortermittlung wird durch ein Ereignis angestoßen, dass durch den Dialogmanager ausgelöst wird und dabei die Benutzereingabe an die Chatbot-Komponente weiterleitet. Im ChatterBean wird die korrekte Antwort ermittelt, wobei zur Entscheidungsfindung die aktuellen Werte von *user predicates* vorliegen. Die Antwort wird an die Chatbot-Komponente zurückgegeben. Dabei wird gegebenenfalls die Syntax der darin befindlichen zusätzlichen Befehle in native Formate der verarbeitenden Komponenten

konvertiert. Bei dieser Konvertierung wird teilweise auf *user predicates* zugegriffen, weshalb diese Konvertierung im Chatbot erfolgen muss, da außerhalb des Chatbots der Zugriff nicht möglich ist. So wird beispielsweise bei einer Timeout-Funktion überprüft, ob der Benutzer diese Funktion nicht generell abgeschaltet hat oder einen anderen Wert gewählt hat. Hat das *user predicate* „user_option_timeout“ einen Wert ungleich 0, so wird die Timeout-Funktion, die in der Antwort einer Regel vorgesehen ist, nicht herausgefiltert. In diesem Fall wird die Funktion um den vom Benutzer gewählten Timeout in Sekunden ergänzt, so dass der Aufruf die Form %%AV TIME ACTION=“WEITER“ TIME=“5“ AVE%% hat.

Bei der Anpassung der rückzumeldenden Antwort können auch aktuelle Werte aus den CAN-Bussen verwendet werden. Bei Änderungen von Werten, werden diese vom CAN-Bundle direkt an den Chatbot übermittelt. So kann beispielsweise abhängig von der Fahrzeuggeschwindigkeit die Anzeige bei Verwendung des Befehls „PlayMedia“ gesteuert werden. Als letzter Schritt der eigentlichen Antwortermittlung erfolgt die Rückgabe der Antwort zurück an den Dialogmanager. Da *user predicates* im Rahmen der Antwortgenerierung verändert werden können, werden im Anschluss die Werte der gepflegten *user predicates* aus dem ChatterBean abgefragt und an die Predicate Persistenz zur Speicherung weitergeleitet. Dieser Schritt erfolgt nach der Rückgabe der Antwort an den Dialogmanager, um den Prozess der Antwortermittlung nicht unnötig zu verlängern.

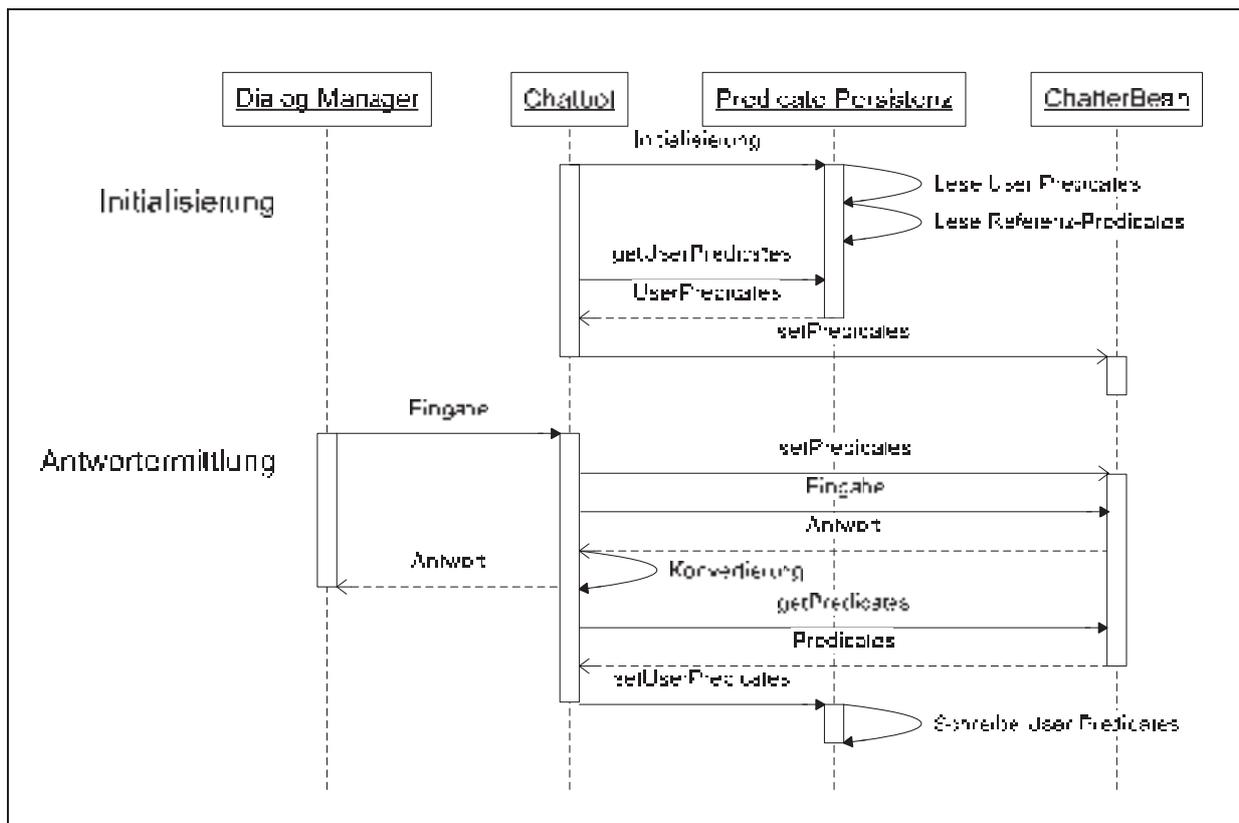


Abbildung 6-12: *Initialisierung und der Antwortermittlung der Chatbot-Komponente*
Quelle: (Eigene Darstellung)

Die Umsetzung des Chatbot-Bundles unter Verwendung des ChatterBeans erlaubt somit prinzipiell die Erweiterung der Nutzung hinsichtlich zweier Aspekte. Zum einen erlaubt die

Instanziierung des ChatterBeans die Umsetzung mehrerer paralleler Chatbots, die evtl. mit verschiedenen Eigenschaften und Regelbasen arbeiten können. Zum anderen erlaubt die Kapselung der Predicate Persistenz insbesondere hinsichtlich der *user predicates* die Nutzung des Avatarsystems durch verschiedene Benutzer, für die jeweils ein Set an *predicates* vorgehalten wird. So kann beispielsweise bei der Nutzung eines Fahrzeugs durch mehrere Familienmitglieder oder Mitarbeiter jeder Benutzer ein individuelles Profil haben, das bei Fahrtantritt in das Avatarsystem geladen wird.

6.1.3.5. GUI-Framework mit Präsentationsmanager

Die im Rahmen der Konzeption postulierte Parallelität des Zugriffs auf die Output-Komponenten setzt voraus, dass jede entsprechende Komponente eine abgetrennte Einheit zur Laufzeit bildet, die über eine wohldefinierte API ansprechbar ist. Obwohl sowohl die Sprachausgabe als auch die Avatarkomponente jeweils eine C/C++-API bereitstellen, kann diese nicht ohne ein JNI aus Java heraus angesprochen werden. Diese Parallelität wird zum einen durch die direkte Steuerung der Sprachausgabe durch die Avatarkomponente zur Realisierung einer lippensynchronen Ausgabe verhindert. Zum anderen verhindert die Verwendung der JavaScript-Schnittstelle bei der Ansteuerung der Avatarkomponente die parallele Steuerung der Output-Kanäle (vgl. 6.1.3.6). Durch diese Restriktionen gestaltet sich die Steuerung der Output-Kanäle sequentiell in der Form, dass zunächst das GUI-Framework Steuerbefehle für die Anzeige von Medien und Symbolen extrahiert, bevor Animationsanweisungen für den Avatar und die 3D-Umgebung in der Avatarkomponente verarbeitet werden. Schließlich wird der zu synthetisierende Text mit den entsprechenden phonetischen und prosodischen Befehlen in der Sprachausgabe abgearbeitet (vgl. Abbildung 6-13).

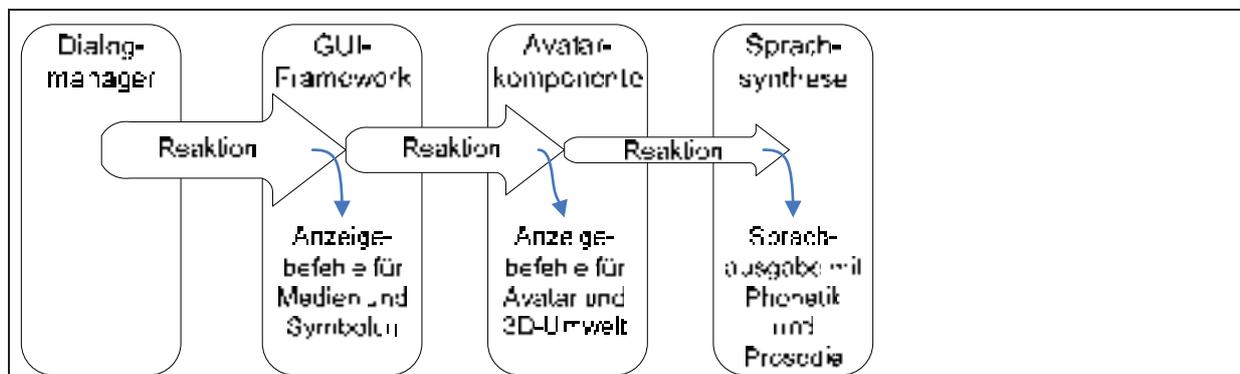


Abbildung 6-13: Befehlsverarbeitung im Output-System
Quelle: (Eigene Darstellung)

Da die Sprachausgabe derzeit nur Eingaben mit einer Länge von maximal 255 Zeichen fehlerfrei verarbeiten kann, wird bereits im Dialogdesigner darauf geachtet, dass Antworten diese Länge nicht überschreiten. Da diese Länge aufgrund diverser Steuerinformationen in manchen Fällen dennoch überschritten werden kann, erfolgt in der aktuellen Version des GUI Frameworks zur Gewährleistung der Funktion ein harter Schnitt nach 255 Zeichen bei der Weitergabe an die Avatarkomponente. Trotz der sequentiellen Abarbeitung kann die in einer Reaktion vorgegebene Reihenfolge von Sprachausgabe, Avataranimation und der Anzeige oder Veränderung der umgebenden Elemente eingehalten werden. Dadurch ist die

abgestimmte Präsentation einer Reaktion gemäß dem in einer Reaktion vorgesehenen „Drehbuch“ möglich, um einen möglichst menschlichen Eindruck des Avatars zu erwecken.

Das GUI-Framework ist aufgrund der Steuerung der Avatarkomponente über JavaScript eine HTML-Seite, die den Avatar als Plug-in integriert. Dadurch können zum einen jegliche Arten von Medien wie Bilder, Videos oder weitere Plug-ins unter Nutzung der Möglichkeiten des entsprechenden Web Browsers zur Anzeige verwendet werden ohne zusätzliche Programmierung durchzuführen. Der Avatar selbst und die dreidimensionale Umwelt werden durch das Avatar-Plug-in CharActor dargestellt und animiert.

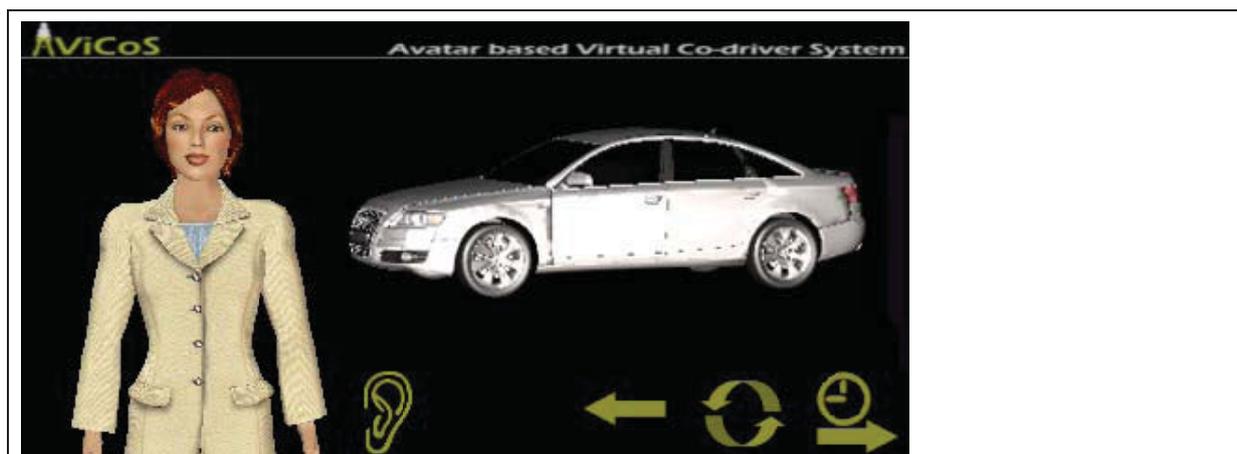


Abbildung 6-14: Screenshot des GUI Frameworks

Quelle: (Eigene Darstellung, enthält Avatar der Charamel GmbH)

Hinsichtlich der Abstimmung von Avatar und der Darstellung von Videos zeigte sich jedoch ein Problem bei der Überlagerung von Videos durch den Avatar. Versucht der Avatar aufgrund einer vorgegebenen Animation auf einen Bereich in einem Video zu zeigen, so verschwindet die Hand hinter dem Video. Dieses Verhalten kann dabei nur durch die Verwendung von Flash-Animationen behoben werden, die jedoch aufwändiger in der Erstellung sind (Charamel 2004c, 13f). Unterhalb der Darstellung von Medien oder 3D-Objekten deuten verschiedene Symbole auf allgemeine Interaktionsmöglichkeiten des Benutzers hin (vgl. Abbildung 6-14).

Eine direkte Ansprache des GUI-Frameworks aus OSGi heraus ist nicht sinnvoll, obwohl der Aufruf einer lokal aufgerufenen Webseite durch Java möglich ist. Die Reaktion könnte als Parameter der URL beim Aufrufen der entsprechenden Webseite übergeben werden, was jedoch insbesondere durch die Einbindung des Avatar-Plug-ins zu einem störenden Neuaufbau der Seite bei jeder Benutzereingabe führen würde. Aus diesem Grund nimmt ein OutputWrapper-Bundle die Reaktion des Dialogmanagers entgegen und speichert sie in eine Datei. Diese Datei wird mithilfe eines JavaScripts in der Framework-Webseite in regelmäßigen Zeitabständen von 100 Millisekunden eingelesen und verarbeitet. Dadurch muss die Webseite nur ein einziges Mal beim Starten der Plattform geladen werden und weist eine ruhige und konstante Darstellung auf.

Symbol	Position	Bedeutung
	4	Wiederholen: die aktuell ausgegebene Antwort kann nach Aufforderung durch den Benutzer wiederholt werden, wenn er zum Zeitpunkt der Ausgabe nicht zuhören oder zuschauen konnte. Diese Option besteht bei jeder Antwort des Avatarsystems.
	1	Mikrofon: Direkt nach dem Betätigen des Push-to-Talk-Buttons zeigt dieses Symbol an, dass das Avatarsystem der Eingabe des Benutzers lauscht; es verschwindet nach der Verarbeitung der Spracheingabe.
	5	Weiter: Die Erklärung des Avatarsystems ist nur ein Teil einer Erklärungssequenz. Der Benutzer kann durch Aufforderung den folgenden Teil sehen bzw. hören.
	3	Zurück: Der Benutzer kann sich die vorherige Reaktion des Avatarsystems erneut ausgeben lassen.
	2	Fragezeichen: Aufgrund der aktuellen Reaktion erwartet das Avatarsystem vom Benutzer eine Antwort oder Eingabe.
	2	Hinweis: Zur aktuellen Antwort besteht die Möglichkeit, sich weitere Hinweise wiedergeben zu lassen, was durch die Frage nach Hinweisen erfolgen kann.
	2	3D-Objekt: Das angezeigte grafische Objekt liegt dreidimensional vor so dass die Kameraperspektive auf das Objekt angepasst werden kann.
	5	Timeout: Die aktuelle Reaktion enthält einen Timeout-Befehl, der nach einer durch den Benutzer eingestellten Zeit aufgerufen wird. Ein Timeout-Befehl kommt meist in Verbindung mit einer Erklärungssequenz vor und vereinfacht das wiederholte Fortfahren.
	1	Hand: Erscheint die Hand, so kann der Benutzer durch Betätigung des eben erklärten Bedienelements weitere Informationen oder eine Bestätigung erhalten
	1	Ohr: Das Ohr signalisiert dem Benutzer die Wiedergabe einer Information aus dem Handbuch, bei der der Benutzer lediglich zuhören soll ohne ein Bedienelement zu betätigen.

Tabelle 6-4: **Symbole des GUI-Frameworks**
Quelle: (Eigene Darstellung)

Durch das zeitlich gesteuerte Einlesen der Datei wird der Inhalt nur verarbeitet, wenn er sich vom Inhalt des vorherigen Einlesens unterscheidet. Dies ist notwendig, da ansonsten der gleiche Inhalt im entsprechenden Intervall immer wieder verarbeitet und ausgegeben wird. Da trotzdem auf eine erneute Benutzereingabe die gleiche Antwort gegeben werden kann, wird der Dateiinhalt mit einem Zeitstempel versehen, der auch die wiederholte Ausgabe der gleichen Antwort ermöglicht, wenn sich der Zeitstempel unterscheidet. Der eingelesene Inhalt wird geparkt, wobei nur Befehle zur Anzeige von Medien und Symbolen von Belang für das GUI-Framework sind. Während Medien entsprechend der Reihenfolge in der Datei in die Ausgabe eingebunden werden, kommen Symbole erst nach Abschluss der akustischen und visuellen Ausgabe zur Anzeige, um die Aufmerksamkeit des Benutzers nicht vom Kerninhalt abzulenken. Eine Ausnahme bei der Darstellung der Symbole stellt das Mikrofonsymbol dar, das sofort angezeigt wird, damit der Benutzer ein direktes Feedback auf das Drücken des Push-to-Talk-Buttons erhält. Innerhalb des GUI-Frameworks ist eine Symbolmatrix abgelegt,

die zur Verarbeitung von GUI-Symbolen verwendet wird und angibt, an welcher der fünf verfügbaren Anzeigepositionen ein Symbol angezeigt wird und wie der Dateiname des Symbols lautet. Eine Übersicht der Symbole und ihrer Bedeutung ist in Tabelle 6-4 zu sehen. Die Bedeutung der Symbole wird dem Benutzer in einem einführenden Tutorial erklärt. Der Teil der Reaktion, der nicht im GUI-Framework verarbeitet wird, wird zur weiteren Verarbeitung über die JavaScript-Schnittstelle an das Avatar-Plug-in und somit indirekt auch an die Sprachausgabe weitergeleitet.

6.1.3.6. Animationskomponente

Die Animationskomponente umfasst zum einen die Darstellung des dreidimensionalen Avatars und dessen Bewegungen in Echtzeit als auch die Integration von dreidimensionalen Objekten in der Umwelt des Avatars. Da neben dem Anspruch, die Steuerung und das Rendering der Darstellung in Echtzeit darstellen zu können, auch die Notwendigkeit der Einbindung der Rendering-Engine in das Avatarsystem bestand, spielte bei der Auswahl der Technologie auch die Form der API eine bedeutende Rolle. Neben den Aspekten der Laufzeit ist zudem der Aufwand bei der Erstellung von Avatar und 3D-Objekten zu betrachten. Tümmler (2007, 34ff) stellt eine umfangreiche Auswahl relevanter Modellierungs- und Animationsprogramme hinsichtlich der Eignung für ein Avatarsystem vor. Dabei erweisen sich alle vorgestellten Komponenten als unangemessen wegen einem der folgenden Gründe:

- Aufwändige Modellierung von Avataren und Bewegungsabläufen
- Keine Komponente zur Echtzeitanimation des erstellten Avatars bzw. der 3D-Objekte
- Keine Möglichkeit zur vollständigen Integration in das Avatarsystem bzw. eine unzureichende API

Neben diesen allgemeinen Hindernissen für die Umsetzung eines in Echtzeit animierten Avatars, spielte darüber hinaus auch die Besonderheit des Einsatzes im Fahrzeug eine Rolle. So kann der verwendete CarPC aufgrund der Wärmeentwicklung und des Formfaktors nicht die gleiche Leistung bereitstellen, wie sie von einem Desktop-PC zur Animation erwartet wird. Zudem muss die Laufzeitkomponente des Animationssystems zusammen mit allen anderen Komponenten Platz auf der SD-Card finden, die im CarPC als Festplattenersatz verwendet wird.

Charamel CharActor

Nachdem die Anwendung der vorgestellten Lösungen als ungeeignet bewertet wurden, wurde ein Lösungsansatz gewählt, der bereits mehrfach für die Darstellung von animierten Avataren im Internet angewandt wurde: Charamel CharActor. Die Eignung für die Anwendung im Fahrzeug wird dabei auch dadurch belegt, dass in den beiden einzigen bisher veröffentlichten Projekten in diesem Bereich bei VW und Mercedes (Heise Autos 2007; Mercedes Scene 2007) die gleiche Software zum Einsatz kam. Der Charamel CharActor ist eine Laufzeitumgebung zur Darstellung und Animation von Avatar und Umwelt in Echtzeit, die von vornherein als integrierbare Komponente für Avatarsysteme entwickelt wurde (Charamel 2004c). Eine Besonderheit der Laufzeitumgebung ist zum einen der Motion Synthesizer, der

Bewegungsabläufe automatisch und übergangslos ineinander umrechnet, um fließende Bewegungen zu erzeugen. Zum anderen wird durch die direkte Ansteuerung der Sprachsynthese eine lippen-synchrone Sprachausgabe ermöglicht (Charamel 2004c, 6). Dabei können beliebige Sprachsyntheseprogramme über die von Microsoft spezifizierte Speech API (SAPI) eingebunden werden, so dass eine Koordination der Sprachsynthese durch den Outputmanager entfällt. Im Falle der verwendeten Sprachsynthese SVOX, die keine SAPI bereitstellt, wurde von Charamel eine gesonderte Ansteuerung speziell für dieses Forschungsvorhaben entwickelt.

Hinsichtlich der vorgestellten K.O.-Kriterien für Animationslösungen im Fahrzeug erleichtert die Lösung von Charamel die Modellierung zum einen durch die Bereitstellung von kundenspezifischen Avataren sowie einer umfangreichen Bibliothek mit Animationsskripten, die durch ihre Erstellung mit MOCAP realitätsnah wirken²¹. Für die Umsetzung des Forschungsvorhabens wurde von Charamel der bestehende Avatar namens Helen kostenfrei zur Verfügung gestellt (vgl. Abbildung 6-15). Darüber hinaus kann mit dem 3D Max Studio Exporter Tool der Export von 3D-Objekten aus der Modellierungsanwendung 3D Studio Max erreicht werden. Für dieses Forschungsvorhaben wurden 3D-Daten von Audi verwendet, die in das 3D Max Studio-Format 3DS konvertiert und mit dem Export Tool exportiert wurden. Die so bereitgestellten 3D-Objekte können in Echtzeit über die eigene Skriptsprache CharaScript gesteuert werden. Aufgrund der geringen Leistung des CarPC hinsichtlich der Echtzeit-3D-Berechnung konnten auf der Plattform im Auto jedoch keine 3D-Objekte zur Anzeige gebracht werden.

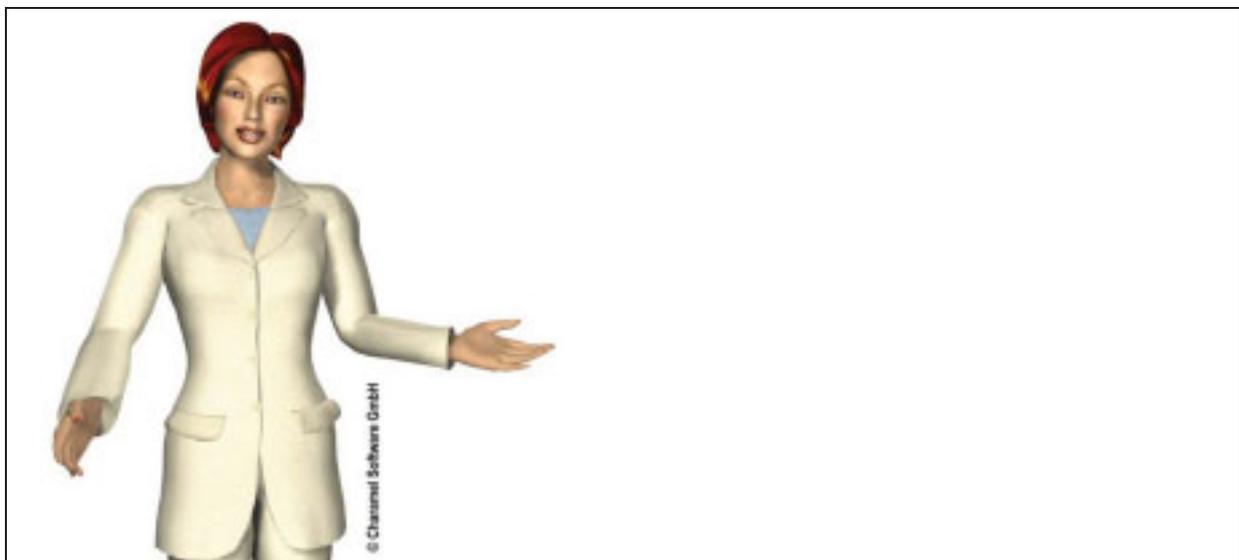


Abbildung 6-15: Avatar „Helen“ von Charamel
Quelle: (Charamel GmbH)

²¹ Die Anwendung Charamel CharactAriser ermöglicht zudem die Zusammenstellung von Animationsskripten und Sound-Dateien zu Szenen, speziell für den Einsatz im Internet (Charamel 2004a). Diese Software spielt in diesem Forschungsvorhaben wegen der Steuerung über die API in Echtzeit jedoch keine weitere Rolle.

Die Systemanforderungen von CharActor sind aufgrund der ursprünglichen Verwendung über das Internet skalierbar. Unabhängig davon, welches der drei unterstützten Renderingverfahren eingesetzt wird, kann der Avatar auf jedem aktuellen PC mit einem Microsoft Windows Betriebssystem dargestellt werden, wobei geringe Unterschiede in der Darstellungsqualität auftreten können. Es werden derzeit ein Software Renderer, ein Direct3D Renderer zur Verwendung von 3D Hardwarebeschleunigern sowie ein OpenGL Renderer unterstützt (Charamel 2004c, 6). Aufgrund der fehlenden Hardwareunterstützung zur 3D-Berechnung im CarPC wurde der Software Renderer verwendet. Die Ansteuerung von CharActor kann technisch auf vier verschiedene Arten erfolgen²²:

1. Das JavaScript-Interface wird für den Einsatz des CharActor Plug-ins in Webbrowsern verwendet und ist somit aus allen web-basierten Programmiersprachen wie z.B. (X)HTML, PHP, Flash und Java (Applets) ansprechbar. Das Plug-in, welches nur in Microsofts Internet Explorer ausgeführt werden kann, lässt sich dabei im Webbrowser (Browser Overlay Mode) oder auf dem gesamten Bildschirm (Desktop Overlay Mode) darstellen.
2. Eine C/C++ API C/C++ API in Form eines SDKs erlaubt, die direkte Steuerung aus C/C++ heraus.
3. Eine ActiveX-Variante der Software ermöglicht die Integration in Sprachen der Entwicklungsumgebung von .NET.
4. Die Socket-Schnittstelle stellt den allgemeinsten Ansatz dar und kann prinzipiell aus allen Programmier- und Skriptsprachen heraus verwendet werden, die Socket-Kommunikation anbieten.

Trotz der Vielzahl an bereitgestellten Schnittstellen lassen sich aus Java Anwendungen heraus, die nicht als Applet in Webseiten eingebunden sind, nur die Socket-Schnittstelle verwenden. Obwohl die Socket-Schnittstelle die einzig direkte Kommunikation auf Java-Basis erlaubt, stellt eine solche Integration der Avatarkomponente einen großen Aufwand dar, der für die prototypische Umsetzung des Avatarsystems nicht angemessen ist. Somit erfolgt die Steuerung des Avatars unter Verwendung des JavaScript-Interface aus dem GUI-Framework heraus. Im Rahmen der HTML-Seiten des Frameworks wird eine Datei mit dem jeweils aktuellen Steuerbefehl ausgelesen und verarbeitet. Die Datei mit dem Steuerbefehl wird dabei durch ein entsprechendes OSGi-Bundle erzeugt, das als Wrapper für die Ausgabeseite des Avatarsystems dient (vgl. dazu auch Kapitel 6.1.3.5).

Funktionsumfang und Steuerung

Die direkte Steuerung der Anzeige erfolgt über Funktionen, die mithilfe von CharaScript programmiert wurden. Sie steuern z.B. welches zuvor festgelegte Bewegungsskript wann und

²² Neben den vier vorgestellten Schnittstellen wurde zum Ende des Forschungsprojekts die Entwicklung eines gesonderten Java-SDK für CharActor angekündigt, der jedoch aufgrund der fortgeschrittenen Entwicklung keinen Einsatz mehr finden konnte.

in welcher Geschwindigkeit abgearbeitet wird und zu welchem Zeitpunkt oder in welches Morphtarget ein Objekt umgerechnet wird. Unter einem Morphtarget ist dabei eine bestimmte Deformation eines 3D-Objekts zu verstehen. So sind verschiedene Stellungen eines Mundes nur Morphtargets eines einzigen Objekts (vgl. Perlin/Goldberg 1996, 206ff). Standardmäßig liefert Charamel die in Tabelle 6-5 dargestellten, avatarspezifischen Funktionen aus, die durch eigene Skripte erweitert werden können. So müssen beispielsweise für 3D-Objekte, die eingebunden werden, ebenfalls Skripte erzeugt werden, um sie z.B. um ein bestimmte Achse rotieren zu lassen oder Animationen, die bereits während der Modellierung erstellt wurden, abzuspielen. Unabhängig davon, ob es sich um den Avatar oder ein Objekt handelt, bestehen zudem auch Funktionen, um die 3D-Objekte ein- und auszublenden. Neben diesen wichtigsten Funktionen, bietet CharaScript eine Vielzahl weiterer Möglichkeiten zur Steuerung von Objekten bzw. der Kamera und des Lichts im virtuellen Umfeld (Charamel 2004b).

Funktion	Beschreibung
SpeakTTS	Lippensynchrones Ausgeben von Text mithilfe der angebundenen Sprachsynthesesoftware
StopSpeaking	Beenden der Sprachausgabe über Sprachsynthese
DoMotion	Verarbeitung eines Bewegungsskripts des Avatars
DoEmotion	Steuerung des Gesichtsausdrucks des Avatars
SpeakSound	Lippensynchrone Ausgabe einer Sound-Datei
SetCamera	Setzen eines festgelegten Kombination von Kamerawinkel und Entfernung
ChangeToDress	Ändern der Texturen der aktuellen Kleidung (nicht der Kleidung selbst)
Show<Objektname>	Blendet entsprechendes Objekt an einer zuvor festgelegten Position ein
Hide<Objektname>	Blendet entsprechendes Objekt aus

Tabelle 6-5: *Werksfunktionen zur Steuerung des Avatars und seiner Umwelt*
Quelle: (Eigene Darstellung)

Eine Besonderheit stellt die Funktion SpeakTTS zur lippensynchronen Sprachausgabe dar. Dabei wird der auszugebende Text analysiert und in Viseme zerlegt. Diese Viseme sind das visuelle Gegenstück zu Phonemen und stellen den Zustand des Mundes bzw. des Gesichts bei der Aussprache eines bestimmten Phonems dar (vgl. Aschenberner/Weiss 2005; Benoît et al. 1992). Grafisch werden sie als Morphtargets, d.h. Deformationen des Gesichts, abgebildet. Beim Sprechen des Textes werden passend zu den Phonemen die entsprechenden Viseme nacheinander dargestellt, wobei durch den Motion Synthesizer fließende Mundbewegungen entstehen (vgl. Ezzat/Poggio 1998).

Während die Bibliothek für die Gesichtsregungen des Avatars überschaubar ist und eine Teilmenge der Bewegungsbibliothek darstellt, beinhaltet die Bibliothek für die Bewegung des Avatars selbst eine große Anzahl an Bewegungsabläufen. Sie umfassen neben dem Betreten und Verlassen der Szene und damit verbunden dem Grüßen und Verabschieden insbesondere Abläufe, die zum Gestikulieren während des Sprechens, zum Zeigen und betrachten bestimmter Bereiche der Szene verwendet werden können. Darüber hinaus bestehen auch

Skripte für den Fall, dass der Avatar nichts zu tun hat und unauffällig, aber lebendig wirken soll.

Die Verarbeitung einer Eingabe erfolgt standardmäßig nach der Reihenfolge der enthaltenen Funktionen, wobei das Sprechen eines Textes als Defaultfunktion festgelegt ist, wodurch der Befehl „SpeakTTS“ nicht explizit angegeben werden muss. Beispielsweise führt die Eingabe

„Herzlich Willkommen. \$(DoMotion, greet/greet04) Ich freue mich, Sie hier begrüßen zu können. \$(DoEmotion, smile)“

dazu, dass die Eingabe in einzelne Befehle zerlegt und nacheinander abgearbeitet wird. Zunächst wird der erste Teil des Textes lippensynchron ausgegeben. Danach wird eine entsprechende Animation zur Begrüßung abgespielt während der zweite Teil des Textes vorgetragen wird. Schließlich behält der Avatar ein Lächeln im Gesicht zurück. Neben dieser sequenziellen Verarbeitung lassen sich einzelne Abläufe auch explizit parallelisieren.

6.1.3.7. Sprachsynthese

Da der Markt für deutschsprachige Sprachsyntheseprogramme sehr ausgeprägt ist und die Ergebnisse nicht mehr an die ersten metallisch klingenden Versuche erinnern (Frötschl 2005), bietet sich die Verwendung eines bestehenden Systems anstelle einer Neuentwicklung an. Da aus einem vorangegangenen Forschungsprojekt (vgl. Hoffmann/Leimeister/Krcmar 2007) Erfahrungen vorliegen und die Ergebnisse der Sprachsynthese verhältnismäßig menschlich waren, wird an dieser Stelle die Software SVOX verwendet. Sie beinhaltet einen sprachenunabhängigen Kern, der den auszugebenden Text analysiert und darauf aufbauend die Sprache synthetisiert. Für beide Schritte wird eine sprachenabhängige so genannte Lingware verwendet. Sie beinhaltet zum einen grammatische Informationen über die Sprache und zum anderen eine entsprechende Stimme (SVOX 2007). Dies impliziert, dass für jede Kombination von Stimme und Sprache eine gesonderte Lingware vorliegt.

Zur SVOX-Anwendung wird eine Schnittstelle in Form einer C/C++-API mitgeliefert, über welche die Synthese angestoßen und gesteuert werden kann. Diese Schnittstelle wurde im Rahmen des INI.TUM-Projekts HIMEPP zunächst in ein Java Native Interface (JNI) eingebunden, das die Ansprache aus Java heraus erlaubt. Anschließend wurde diese Schnittstelle in ein OSGi-Bundle eingebunden. Aufgrund der Wahl der Avatar-Komponente CharActor von Charamel, konnte dieses Bundle jedoch nicht sinnvoll genutzt werden. Die CharActor Software spricht zur Realisierung einer lippensynchronen Sprachausgabe die angebundene Sprachsynthese direkt an, um Latenzzeiten und asynchrone Ausgaben zu vermeiden. In der Regel erfolgt die Steuerung von Sprachsyntheseprogrammen über die von Microsoft spezifizierte Speech API (SAPI), wie sie beispielsweise auch beim Avatarsystem REA genutzt wurde (Cassell et al. 1998, 27). Da SVOX diese API in der aktuellen Version nicht anbietet, wurde von Charamel eine direkte Ansprache umgesetzt. Dadurch ist die Einbindung des entsprechenden OSGi-Bundles hinfällig geworden.

Die Verwendung von SVOX lässt sowohl eine phonetische als auch eine prosodische Optimierung zu. Die phonetische Optimierung wird dabei vor allem mit SAMPA umgesetzt

und kann allgemein oder textspezifisch erfolgen. Die Möglichkeit der allgemeinen Optimierung umfasst neben einer Konvertierungsliste auch ein Aussprachewörterbuch. Die Konvertierungsliste wird dazu verwendet, um Abkürzungen wie „d.h.“ in „das heißt“ oder „PCM“ in „P, C, M“ umzuwandeln und dadurch eine korrekte Aussprache zu erreichen (SVOX 2006, 14). Das Aussprachewörterbuch ermöglicht vor allem die korrekte Aussprache von Eigennamen unter Verwendung von SAMPA. Dabei wird neben der Schriftform und der SAMPA-Notation auch angegeben, um welche Wortart es sich bei dem jeweiligen Begriff handelt, insbesondere ob es sich dabei um einen Wortstamm oder eine konjugierte bzw. deklinierte Form handelt. Dadurch können z.B. die meisten Wortformen eines Wortes nur durch Angabe der Aussprache eines Wortstamms abgedeckt werden (SVOX 2006, 15f). Die textspezifische, phonetische Optimierung erfolgt durch Verwendung gesonderter Tags, die in den zu synthetisierenden Text eingebunden werden und die SAMPA-Notation für ein Wort enthalten. So führt beispielsweise die Eingabe

„Mit einem Konzert der Staatlichen Koreanischen Philharmonie <phon sampa="sOul">Seoul</phon> wird heute der „Internationale Musiksommer 2007“ eröffnet.“

zur Ausgabe des entsprechenden Textes unter Verwendung der SAMPA-Notation für den Städtenamen „Seoul“. Zur Bestimmung der Aussprache von Fremdwörtern kann zudem auch ein landesabhängiges Alphabet für das <phon>-Tag verwendet werden (SVOX 2006, 19f). In der neuesten Version von SVOX lässt sich anstelle von SAMPA auch die IPA-Notation zur phonetischen Steuerung verwenden.

Die prosodische Optimierung erfolgt überwiegend automatisch, kann jedoch durch entsprechende Tags in einem Text gezielt gesteuert werden. Die automatische Optimierung erfolgt dabei vor allem auf Basis von Satzzeichen, die dazu genutzt werden, die Stimmhöhe anzupassen. So wird bei einem Komma die Tonhöhe leicht und bei einem Punkt stark gesenkt. Ein Fragezeichen hingegen führt zum Ansteigen der Tonhöhe. Die gezielte Steuerung erfolgt mit der SVOX Markup Language, die an STML bzw. SSML angelehnt ist (vgl. Kapitel 3.1.4.2). Dabei können neben der expliziten Angabe von Absatz- und Satzabgrenzungen auch das Tempo, die Lautstärke, die Tonhöhe und Unterbrechungen bestimmt werden (SVOX 2006, 17ff). Eine Übersicht der Tags für phonetische und prosodische Optimierung sowie weitere Steuerbefehle im Rahmen der SVOX Markup Language sind in Tabelle 6-6 dargestellt. Die Verwendung der wichtigsten Tags erfolgt im Dialogeditor, wobei intern eine Konvertierung der Syntax zur Verwendung in AIML vorgenommen werden muss (vgl. 6.1.3.4).

Die Anwendung der Tags im phonetischen Bereich spielt insbesondere bei der Aussprache von Eigennamen eine Rolle. Da im Inhalt eines Bedienerhandbuchs erwartungsgemäß keine Eigennamen vorkommen, ist die phonetische Optimierung nur bei langen Bindewörtern und Fachausdrücken sinnvoll. Die prosodische Optimierung kann zum einen dazu genutzt werden, bestimmte Passagen einer Erklärung zu betonen oder um gegebenenfalls Emotionen des Avatars zu untermalen.

Bereich	Befehl	Beschreibung	Beispiel
Phonetik	<phon sampa="..." />	Festlegung der Aussprache für nächstes Wortes mit SAMPA	Hello, Mister <phon sampa="braUn" /> Braun.
	<phoneme alphabet="..." ph="..." />	Festlegung der Aussprache des nächsten Wortes unter Verwendung eines bestimmten Alphabets	Hello, Mister <phoneme alphabet="x-SVOX-sampa_de-DE" ph="braUn" /> Braun.
	<preproccontext name="..."> ...</preproccontext>	Bestimmung des Preprocessing Contexts (Konvertierungsliste)	<preproccontext name="CID"> c312432186 </preproccontext>
Prosodie	<p> ... </p>	Explizites Abgrenzen eines Absatzes	<p> This is a paragraph. </p>
	<s> ... </s>	Explizites Abgrenzen eines Satzes	<s> This is a sentence. </s>
	<break time="..." />	Gibt die Dauer einer Pause zwischen Textteilen an	This is the first sentence. <break time="2s" /> This is the second sentence.
	<pitch level="..."> ... </pitch>	Setzt die Tonhöhe (50 bis 200) für den Text in Tags. Der Defaultwert ist 100, der nach den Tags automatisch wieder gesetzt wird	Hello, <pitch level="140"> Miss Jones </pitch> arrived.
	<speed level="..."> ... </speed>	Setzt die Geschwindigkeit (20 bis 500) für den Text in Tags. Der Defaultwert 100 gilt automatisch außerhalb Tags	Hello, <speed level="300"> Miss Jones </speed> arrived.
	<volume level="..."> ... </volume>	Setzt die Lautstärke (0 bis 20.000) des Textes in Tags. Der Defaultwert 100 gilt automatisch außerhalb Tags. Über 100 kann es zu Verzerrungen kommen.	Hello, Miss <volume level="50"> Jones </volume> arrived.
Sonstiges	<ignore> ... </ignore>	Ignoriert den Text in Tags bei der Synthese	Hello <ignore> any text </ignore> Mister Smith.
	<voice name="..."> ... </voice>	Gibt die Stimme für den Text in Tags an	<voice name="rt0co0a16">This is the english voice.</voice>
	<usesig file="..."> ... </usesig>	Ersetzt Textteile mit zuvor aufgenommenen Sounddateien	<usesig file="intro.wav"> Hello, Mister </usesig> Smith
	<play file="..." />	Spielt eine Sounddatei ab	Hello, Miss <play file="miller.wav" />.

Tabelle 6-6: *Tags der SVOX Markup Language*
Quelle: (Eigene Darstellung)

6.1.4. Entwicklung und Zusammenspiel der Komponenten

Die Vielzahl der vorgestellten Komponenten und der eingebundenen Systeme, die als Teile der Komponenten agieren, verhinderte eine parallele Entwicklung aller Komponenten. Vielmehr wurden zunächst einzelne Laborprototypen entwickelt, um die Machbarkeit und die Funktion einzelner Elemente zu untersuchen, bevor die Kombination in der Laufzeitumgebung und der Zielplattform vorgenommen werden konnte. Die Entwicklung und Abstimmung der Interaktion zwischen den Einzelteilen lässt sich demnach grob in drei Phasen unterteilen, die nachfolgend kurz beschrieben werden.

In der ersten Phase wurde die Kernkomponente des Chatbots in Form des ChatterBeans in ein Standalone Java-Programm eingebunden, um das Verhalten des AIML-Interpreters zu testen. Innerhalb dieses ersten Testprogramms wurde der Kern des späteren Präsentationsmanagers entwickelt, der zunächst nur die Antwort des Chatbots in eine Datei geschrieben hat. Auf Seiten der Ausgabe wurde ein HTML-basierter Demonstrator der Charamel GmbH mithilfe von JavaScript derart erweitert, dass er diese Datei einlas und den Text vom Avatar und der darüber angesprochenen Sprachsynthese vorlesen ließ. Durch die Kombination von Chatbot, Präsentationsmanager, Avatar und Sprachsynthese konnten bereits erste Funktionen getestet und die Wirkung des Avatars beobachtet werden.

Die zweite Phase beinhaltet insbesondere die Entwicklung von OSGi-Bundles mit Knopflerfish und damit auch OSGi in der Version 4, um die Kommunikation zwischen separaten Komponenten zu ermöglichen. Nach der Umsetzung von Chatbot und Präsentationsmanager als getrennte Bundles wurde die Entwicklung des Dialogmanagers notwendig, der zum einen Eingaben aus einem provisorischen Eingabefeld an den Chatbot leitete und dessen Antwort wiederum an den Präsentationsmanager weitergab. Die lose Kopplung der einzelnen Bundles wurde dabei mithilfe des OSGi-Eventings erreicht. Zur Integration der Spracherkennung wurde ein gesondertes Bundle entwickelt, das neben der Möglichkeit der Texteingabe auch einen virtuellen Push-to-Talk-Button bereitstellte und während des ganzen Projektes verwendet wurde, wenn keine Anbindung an den CAN-Bus des Fahrzeugs bestand. Diese Eingabeoberfläche aktivierte bei Betätigen des Push-to-Talk-Buttons die eingebundene HIMEPP Spracherkennung und leitete die erkannte Phrase an den Dialogmanager weiter. Neben der Erweiterung auf Eingabeseite wurde die Komponente des GUI-Frameworks entwickelt, durch die neue Anzeigemöglichkeiten, ein erweiterter Befehlssatz und ein entsprechender Parser hinzukamen. Sowohl das GUI-Framework als auch der Chatbot beeinflussten dabei die Gestaltung des Dialogdesigners, der parallel zur zweiten Phase entwickelt wurde. Ebenfalls parallel wurde das CAN-Bundle entwickelt, das zunächst jedoch separat auf Knopflerfish 3 und damit auch OSGi 3 lief. Die Tests des Bundles erfolgten dabei zum einen an einem Brettaufbau mit allen wichtigen Steuergeräten eines Fahrzeugs und zum anderen später im Testfahrzeug selbst.

Die letzte Phase der technischen Entwicklung war besonders durch die Migration von Knopflerfish nach Equinox gekennzeichnet. Diese war notwendig, um das Avatarsystem mit dem CAN-Bundle in einer einzigen Laufzeitumgebung unter Nutzung des OSGi-Eventings nutzen zu können. Der Betrieb des CAN-Bundles und verschiedener darin referenzierter Bundles der Fachabteilung AEV-3 auf OSGi 4, machte die Überlagerung verschiedener Logging-Klassen mit mithilfe eines HIMEPP-Bundles notwendig. Durch die Migration wurde zudem die Verwendung des Bundles HIMEPP Props erforderlich, um die Systemeigenschaften der Java Virtual Machine einzustellen, was unter Knopflerfish vereinfacht mit so genannten xarg-Dateien möglich war. Durch die Integration des CAN-Bundles konnte der Eingabeteil des Systems vervollständigt werden, so dass beim Einsatz im Fahrzeug der Push-to-Talk-Button des Multifunktionslenkrads zur Aktivierung der Spracherkennung genutzt werden konnte. Mit der Nutzung im Fahrzeug ging auch die Migration des gesamten Softwaresystems auf den CarPC einher. Einen Überblick aller neu entwickelten Komponenten, darin referenzierten Elemente und die Kommunikationswege zwischen ihnen zeigt Abbildung 6-16.

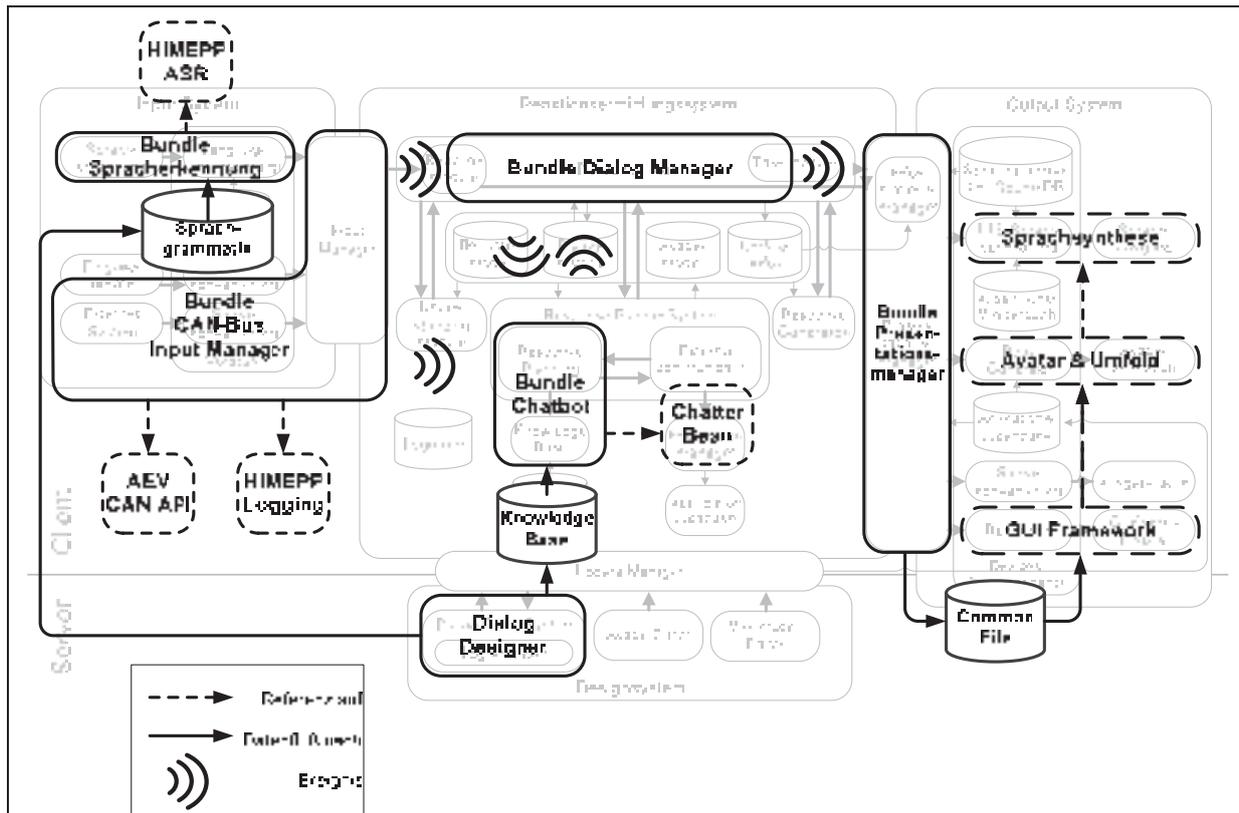


Abbildung 6-16: Zusammenspiel der Komponenten des Avatarsystems
Quelle: (Eigene Darstellung)

6.2. Erstellung von Inhalten

Die technische Plattform stellt alle notwendigen Funktionen zur Ausführung entsprechend gestalteter Inhalte bereit. Dementsprechend kann sie ohne Inhalte nicht funktionieren. Zudem entscheidet die Qualität der Inhalte des Avatarsystems über den Eindruck und die Menschlichkeit des Avatars. So kommt der Gestaltung des Inhalts in Form der Regelbasis und der darin referenzierten Medien eine bedeutende Rolle zu, die im Folgenden aus verschiedenen Blickwinkeln betrachtet werden soll. Zunächst lässt sich ein weiteres Werkzeug identifizieren, das die Erstellung des Inhalts unter Berücksichtigung aller beschriebenen Funktionen gewährleistet: Der Dialogeditor bzw. Dialogdesigner. Nachdem diese Anwendung der Designzeit vorgestellt wurde, werden auf Basis der in Kapitel 5.2.1.3 abgesteckten inhaltlichen Bereiche konkrete Szenarien zur Umsetzung systematisch ausgewählt. Auf der Grundlage zweier Einsatzbereiche werden zwei Methoden zur Erstellung von Inhalten für Avatarsysteme gewonnen: Eine Methode für bereits strukturierte Daten wie einem Handbuch und deren Erweiterung für unstrukturierte Informationen.

6.2.1. Gestaltung des Dialogeditors

Wie bereits in Kapitel 5.2.2.2 dargestellt wurde, existiert eine Vielzahl system- oder formatabhängiger Editoren, die meist einen textbasierten Ansatz zur Pflege der Regeln verfolgen. Insbesondere bei der Verwendung komplexer Regelformate erfordert der textbasierte Ansatz vom Benutzer sehr tiefgehende Kenntnisse der Notation und der

verfügbaren Funktionen. Ein grafischer Ansatz, wie ihn beispielsweise Laufer, Tatai und Nemeth (2005) vorstellen, ermöglicht es hingegen, auch mit wenigen Kenntnissen der Notation einen Dialog zusammenzustellen. Trotz der Vereinfachung durch die grafische Repräsentation der einzelnen Dialogschritte erfordert dieses Vorgehen grundlegende Kenntnisse der entsprechenden Notation.

Ein Ansatz, der unabhängig von der konkreten Implementierung arbeitet und dessen Grundlage Modelle sind, ist das Konzept der Model Driven Architecture (MDA) (Fettke/Loos 2003; Kempa/Mann 2005). Wesentliche Merkmale der MDA sind zum einen die Repräsentation in Platform Independent (PIM) und Platform Specific Modells (PSM) wobei als Plattform generische, technologie- oder herstellerspezifische Plattformen zu verstehen sind. Ein weiteres Merkmal ist die Verwendung einer Modellierungssprache wie z.B. UML oder Petri-Netze. Der dritte Bestandteil ist eine Transformationsdefinition, die festlegt wie die Modellierung des plattformunabhängigen Modells in das plattformabhängige Modell übertragen wird (Fettke/Loos 2003, 555f; Kleppe/Warmer/Bast 2003, 15ff). Bei der Umsetzung dieser Merkmale werden insbesondere verschiedene Beschreibungs- und Modellierungsstandards eingesetzt. So werden UML und die Object Constraint Language (OCL) zur Repräsentation der Modelle verwendet, während die Meta Object Facility (MOF) eine Metamodellierungssprache zur Beschreibung der Modellierungssprache ist. Der Austausch von Modellen zwischen Anwendungen kann dabei über XML Metadata Interchange (XMI) erfolgen (Fettke/Loos 2003, 556).

In Anlehnung an eine MDA ist der Dialogeditor derart angelegt, dass er die grafische Modellierung im Rahmen eines plattformunabhängigen Modells ermöglicht. Die Gestaltung des PIM wurde stark von AIML beeinflusst, ist jedoch auch für andere PSM geeignet. Als plattformabhängiges Modell wird AIML verwendet, wobei auch VoiceXML und weitere Formate mit entsprechenden Transformationsdefinitionen nutzbar sind. Die plattformunabhängige Modellierung richtet sich zwar nach festgelegten Regeln, die sich jedoch nicht an MOF orientieren. Dadurch ist auch die Verwendung von XMI erschwert. Da plattformabhängige Modelle jedoch auch nach Änderungen außerhalb des Dialogeditors in ein plattformunabhängiges Modell rücküberführt werden können, kann auf diesem Weg auch der Austausch mit anderen Werkzeugen erfolgen. Die Umsetzung dieser plattformunabhängigen Modellierung von Sprachdialogen versucht damit, den von Polifroni und Chung (2002) formulierten Bedarf nach einer domänenunabhängigen und schnellen Entwicklung von Sprachdialogen entgegenzukommen.

6.2.1.1. *Aufbau des Dialogeditors*

Da der Dialogeditor nicht Teil der Laufzeitumgebung des Avatarsystems ist, ist er als eigenständige Java-Anwendung ohne Verwendung des OSGi-Frameworks umgesetzt. Zur Speicherung der plattformunabhängigen Modelle wird die frei verwendbare Datenbank PostgreSQL in der Version 8.3 verwendet, da sie für den Einsatz mit dem Object-Relational Mapping (ORM)-Framework von Hibernate abgestimmt ist. Hibernate in der Version 3.2.6 vereinfacht die persistente Datenhaltung in relationalen Datenbanken unter Verwendung objektorientierter Programmiersprachen wie Java oder .NET. Es erstellt aus Zeilen in Tabellen unter Verwendung von Joins und Auflösung von Mehrfachbeziehungen

entsprechende Objekte, die unter Verwendung von set- und get-Methoden verändert werden können. Zur Umsetzung eines Transaktionskonzepts lassen sich die vorgemerkten Veränderungen mit einer weiteren Methode aus dem Objekt wieder in die Datenbank zurückschreiben. Die Benutzeroberfläche verwendet weitestgehend Elemente von Java Swing, wobei die Arbeitsfläche zur grafischen Modellierung mithilfe der JGraph Funktionsbibliothek umgesetzt wurde, welche die grafische Anzeige und Manipulation von Objekten vereinfacht (vgl. <http://www.jgraph.com>, zugegriffen am 25.05.2008).

Der Aufbau der Benutzeroberfläche ist in drei Bereiche in Form von Spalten unterteilt: Den Szenarien- und Dialogbereich auf der linken Seite, den Modellierungsbereich in der Mitte und den Funktions- und Ausprägungsbereich auf der rechten Seite (vgl. Abbildung 6-17). Szenarien dienen als Verzeichnisstruktur zur hierarchischen Gliederung von Dialogen. Die Erstellung von plattformunabhängigen Modellen erfolgt innerhalb der Dialoge, die Teil eines Szenarios sind. Diese Aufteilung trägt den Ergebnissen von Cassell et al. (1998, 22) Rechnung, wonach sich Gespräche in verschiedenen lange, hierarchisch zusammenhängende Abschnitte unterteilen lassen, die im Dialogeditor durch Szenarien, Dialoge und den Regeln bzw. Dialogschritten abgebildet sind.

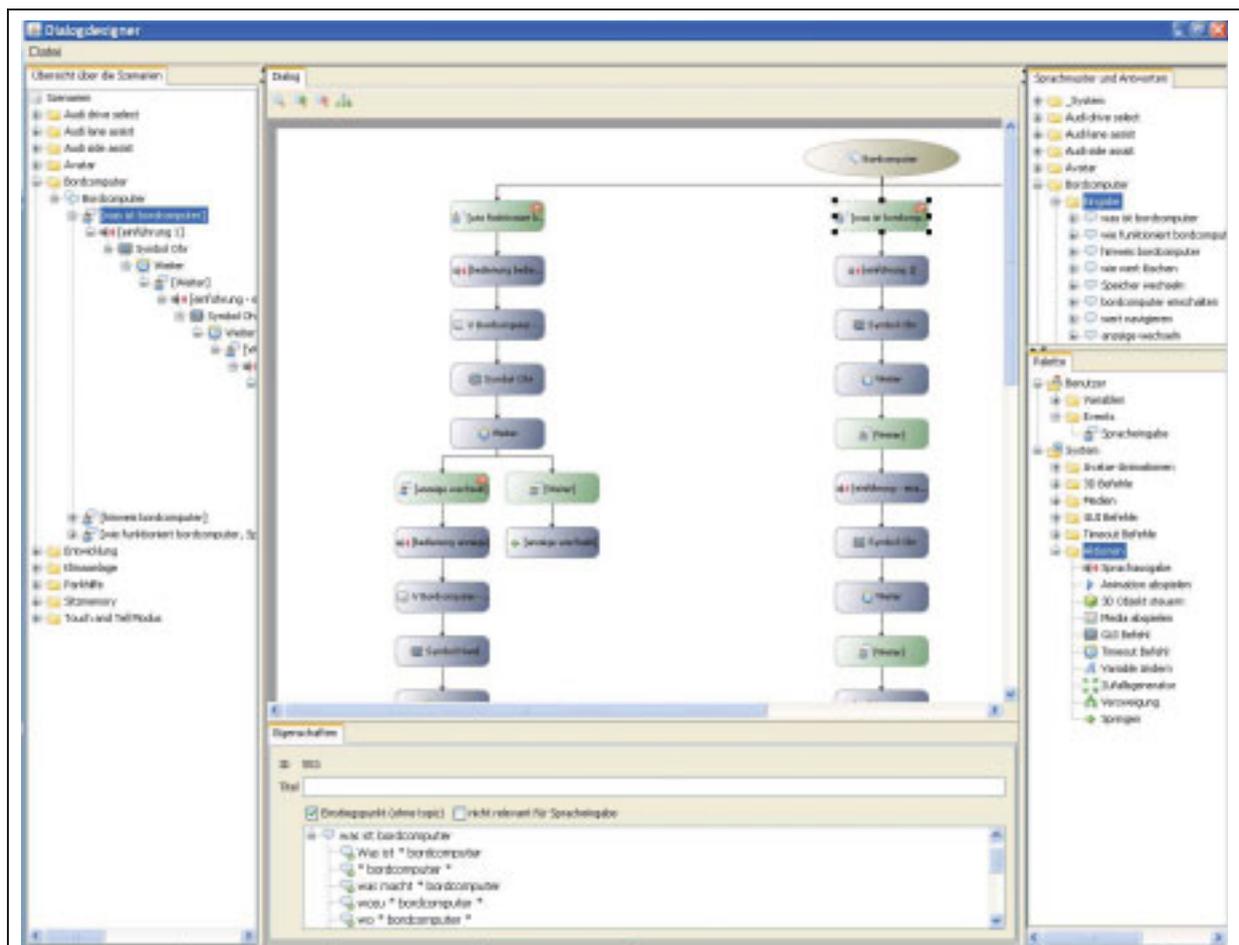


Abbildung 6-17: Aufbau des Dialogeditors
Quelle: (Eigene Darstellung)

„Hallo“, „Guten Tag“ und „Servus“ enthalten. Diese Äußerungen können nur den beiden Schritttypen der Sprachein- oder Sprachausgabe sowie einem Systemereignis zugewiesen werden. Ein Systemereignis ist eine technische Eingabe eines Bussystems, die keine Spracheingabe nutzt. Ein Systemereignis wird als Spracheingabeschritt mit einer bestimmten Einstellung abgebildet. Eine Übersicht aller Entitäten des Dialogeditors und ihrer Zusammenhänge gibt Abbildung 6-18, wobei die Attribute der Entitäten aus Gründen der Übersicht nicht verzeichnet sind.

6.2.1.2. Schritttypen

Die im Dialogeditor verwendeten Schritte weichen aufgrund der pragmatischen Herangehensweise unter Verwendung eines Chatbots von der Einteilung eines realen Gesprächs ab. Dennoch sollen die Grundlagen der Dialogeinteilung eines realen Dialogs kurz beschrieben werden, bevor die konkrete Umsetzung im Dialogeditor vorgestellt wird.

Grundlagen der Dialogeinteilung

Searle (1975) hat in seiner grundlegenden Betrachtung einer möglichen Verarbeitung von Dialogen durch Computer reale Dialoge untersucht. Dabei hat er weniger die sprachlichen Eigenschaften einer Aussage als mehr die Semantik d.h. den Inhalt bzw. die Absicht betrachtet. Im Rahmen dessen versteht er den *indirect speech act* als Sinn- bzw. Absichtsvermittlung. Er klassifiziert fünf verschiedene Arten solcher Sprechakte:

- Repräsentative (assertives): Der Sprecher verpflichtet sich auf die Wahrheit wobei der Grad der Verpflichtung variieren kann
- Direktive (directives): Der Sprecher versucht beim Zuhörer eine Handlung zu bewirken
- Kommissive (commissives): Der Sprecher verpflichtet sich auf zukünftige Handlungen
- Expressive (expressives): Der Sprecher drückt einen psychischen Zustand aus
- Deklaration (declaratives): Der Sprecher bewirkt die Änderung eines Zustands einer Entität (z.B. der Pfarrer traut ein Brautpaar)

Diese grundlegende Betrachtung soll als Ausgangsbasis für konkretere Anwendungen auf Chatbots und Avatarsysteme dienen. So versuchen Stede und Schlangen (2004) eine feingranularere Aufteilung von Sprechakten unter Berücksichtigung der sprechenden Entität, den dialoghistorischen Informationen sowie dem aktuellen Gesprächsthema. Sie identifizieren eine Vielzahl weiterer Sprechakte wobei sie nicht nur die Aussage eines Akts betrachten, sondern auch dessen Auswirkungen auf einen Gesprächsverlauf. Insbesondere die Initiierung und der Wechsel eines Gesprächsthemas sind dabei auch aus technischer Sicht von Bedeutung.

Eine technischere Sicht auf einen Sprechakt in Verbindung mit Avatarsystemen stellen Alexa et al. (2001, 156f) vor. Sie sehen neben dem Inhalt einer Aussage auch die damit verbundenen Steuerinformationen für die Ausgaben eines Avatarsystems. Sie verbinden die eigentliche Sprache mit einem Zustand im emotionalen Modell des Avatarsystems sowie einer bestimmten Blickrichtung des Avatars selbst. Dabei spielt das *turn taking*, also die Information darüber, wer derzeit reden soll bzw. darf, eine wichtige Rolle. Abschließend kann ein Sprechakt mit der Ausführung eines Skripts oder einer bestimmten Animation verbunden werden. Da diese pragmatische Betrachtung eines Sprechakts auch die technischen Herausforderungen der Avatarsystemsteuerung berücksichtigt bildet sie die Grundlage für die Schritttypen innerhalb des Dialogeditors.

Schritttypen des Dialogeditors

Die verschiedenen Schritttypen des Dialogeditors erlauben die Umsetzung von Funktionen, wobei jeder Typ individuelle Eigenschaften aufweist. Der Bereich der Eingabeschritte wird lediglich durch einen Typen repräsentiert: Der Spracheingabe. Dieser Schritttyp stellt den Beginn eines jeden Dialogs dar und kann eine beliebige Menge an Äußerungen enthalten. Es lassen sich mehrere Eingabeschritte als Beginn des Dialogs oder an einen Ausgabeschritt anhängen. Eine Sequenz von Eingabeschritten ist im Dialogeditor nicht zulässig. Ein Spracheingabeschritt kann als Einstiegspunkt markiert werden, d.h. unabhängig davon, über welches Thema eine Unterhaltung mit dem Avatarsystem gehalten wird und unabhängig von der Dialoghistorie kann ein Einstiegspunkt jederzeit aufgerufen werden. Durch diese Markierung werden Themenwechsel erst möglich, da ansonsten implizit davon ausgegangen wird, dass ein Eingabeschritt immer nur im entsprechenden Zusammenhang zum Tragen kommen kann. Einstiegspunkte werden visuell durch eine rote Flagge markiert. Die ersten Eingabeschritte direkt unter der Dialogmarkierung werden beim Export automatisch als Einstiegspunkte behandelt. Zudem kann ein Eingabeschritt als nicht relevant für die Spracherkennung markiert werden. Dadurch kann die damit verbundene Reaktion nicht durch den Benutzer, sondern nur durch eingehende technische Signale der Bussysteme ausgelöst werden. Regeln, die mit einem Eingabeschritt beginnen, der nicht für die Spracheingabe relevant ist, werden in einer gesonderten AIML-Datei exportiert und nicht in eine Sprachgrammatik konvertiert. Darüber hinaus werden solche Schritte mit einem zusätzlichen Symbol im Dialogeditor versehen.

Aus der Menge der Ausgabeschritte stellt die Sprachausgabe den wichtigsten Schritttyp dar, da sie den Text enthält, der als Reaktion auf eine Eingabe ausgegeben wird. Eine Sprachausgabe kann ebenso wie die Eingabeschritte eine beliebige Menge an Äußerungen mit den darin enthaltenen Formulierungen aufnehmen. Dabei wird nach einem AIML-Export zur Laufzeit per Zufall bestimmt, welche der Formulierungen in einem solchen Schritt zur Ausgabe kommt. Ein Sprachausgabeschritt beinhaltet zudem ein Attribut, das im Falle einer Verzweigung angibt, ob der mit diesem Schritt beginnende Zweig relevant ist. Geht einer Sprachausgabe ein Verzweigungsschritt voraus, in dem eine bestimmte Variable abgefragt wird, so steht jeder nachfolgende Sprachausgabeschritt für eine mögliche Reaktion, je nach dem Wert der Variable. Abbildung 6-19 zeigt ein Beispiel, in dem die Variable „a“ als Entscheidungsgrundlage einer Verzweigung verwendet wird. Beinhaltet die Variable den Wert „1“, so wird die Sprachausgabe mit dem Titel „Reaktion 1“ und alle darauf folgenden

Schritte durchgeführt. Ist in der nachfolgenden Reaktion keine Sprachausgabe vorgesehen, so kann ein Sprachausgabeschritt ohne Äußerungen verwendet werden. Wurde eine Variable zuvor nicht mit einem Wert belegt, so enthält sie den Inhalt „undefiniert“ entsprechend der Konfiguration des Chatbots. Der Wert einer Variablen kann explizit mit einem entsprechenden Schrittyp angepasst werden.



Abbildung 6-19: *Verzweigungsschritt in Verbindung mit Sprachausgabeschritten*
Quelle: (Eigene Darstellung)

Der Schrittyp der Animation spielt zur Laufzeit ein Animationskript für den Avatar ab, das in einer entsprechenden Datei abgelegt ist. Der Animationsname, unter dem der entsprechende Dateiname hinterlegt ist, stellt das einzige Attribut dieses Typs dar. Ebenso einfach ist der Schritt zur Integration einer 3D-Objektbefehls aufgebaut, er enthält lediglich den Namen des Befehls hinter dem der 3D-Befehl abgelegt ist. Durch Aufruf dieses 3D-Befehls zur Laufzeit wird eine entsprechende CharaScript-Funktion aufgerufen, die ein 3D-Objekt manipuliert. Die Integration eines Mediums erfolgt ebenfalls über den Namen des Mediums, hinter dem sich ein Dateiname verbirgt. Als Medium werden Bilder, Videos oder weitere Objekte bezeichnet, die vom Web Browser, in dem das Frontend zur Laufzeit betrieben wird, bereitgestellt werden. Zu jedem Video, das als Medium angelegt wird, muss eine gleichnamige Bilddatei existieren, die nicht zwangsläufig als Medium angelegt sein muss. Wird eine bestimmte Geschwindigkeit (derzeit 60 km/h) überschritten, wird anstelle der Videodatei die gleichnamige Bilddatei angezeigt, um den Fahrer nicht übermäßig abzulenken.

Ein Ausgabeschritt zur Einbindung eines GUI-Befehls kann zur gezielten Anzeige eines bestimmten Symbols oder sonstigen Elements der Benutzerschnittstelle verwendet werden. Obwohl die meisten Symbole automatisch auf Grundlage der Reaktion ermittelt und angezeigt werden, kann es beispielsweise im Falle eines Hinweises sinnvoll sein, ein Symbol explizit einzubinden.

Ein Timeout-Befehlsschritt führt zur automatischen Verarbeitung einer bestimmten Benutzereingabe nach einer gewissen Zeit. Der zeitliche Abstand bis zur Verarbeitung dieser Eingabe kann durch den Benutzer als eine Option des Avatarsystems eingestellt werden, wodurch auch die Möglichkeit zur vollständigen Abschaltung der Timeout-Befehle gegeben ist. Ist diese Funktion nicht abgeschaltet, so wird nach der entsprechenden Zeit eine Benutzereingabe durch das System simuliert, die zur Fortsetzung des Dialogs führt. Die Timeout-Eingabe ist wie bei den anderen Schritten dieser Art auch hinter einem Befehlsnamen verborgen.

Der Sprungschritt dient der Vermeidung von Redundanzen, indem der Inhalt anderer Antwortfolgen genutzt werden kann. Ein Sprungschritt enthält die ID eines Eingabeschrittes zu dem an der entsprechenden Stelle gesprungen werden soll. Abbildung 6-20 zeigt ein Beispiel, in dem dieser Schritt dazu genutzt wird, den Einstieg zu einer Antwortsequenz zum einen über einen jederzeit erreichbaren Einstiegspunkt („Erzähl mit etwas über ...“) und zum anderen über Fortführung einer begonnenen Dialogsequenz („Ja“ als Antwort auf eine zuvor gestellten Frage zur Überleitung zwischen zwei Themen) vorzunehmen. Dadurch kann der Benutzer sowohl mit der Eingabe „Ja“ als Fortführung des Dialogs oder jederzeit mit der Frage „Erzähl mir etwas über ...“ zur entsprechenden Antwort gelangen. Durch die Verwendung des Sprungbegriffs muss die Antwort jedoch nur einmal erstellt und gepflegt werden. Ein Sprungbefehl kann auch am Ende einer Reihe von Ausgabeschritten erfolgen. Dadurch werden beide Antwortsequenzen nacheinander abgearbeitet und ausgegeben. Als Sprungziele können dabei Spracheingabeschritte mit der Markierung als Einstiegspunkt gewählt werden.

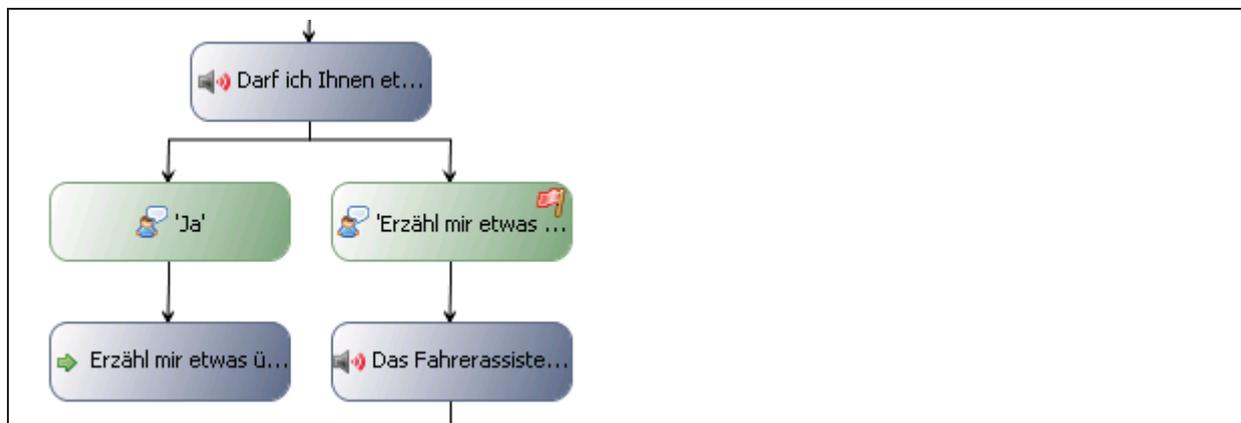


Abbildung 6-20: *Verwendung eines Sprungschritts*
Quelle: (Eigene Darstellung)

Da in manchen Fällen auch Schritte angesprungen werden müssen, die nicht als Einstiegspunkt markiert sind, ist es zudem möglich, beliebige Schritte mithilfe von AIML-Tags anzuspringen. Dazu kann ein Sprachausgabeschritt eingefügt werden, der den AIML-Befehl „<srai>123 WEITER</srai>“ beinhaltet. Der <srai>-Tag entspricht der direkten Umsetzung des Sprungschrittes in AIML. Der Wert innerhalb der Tags beinhaltet zum einen die ID des entsprechenden Eingabeschrittes und den Inhalt einer Formulierung dieses Schrittes. Die ID eines Sprachausgabeschrittes kann im unteren Bereich des mittleren Arbeitsplatzes abgelesen werden, wenn der entsprechende Schritt markiert ist. Die Funktionsweise basiert auf der Verwendung von Schattenschritten in AIML und wird im Kapitel 6.2.1.4 detailliert erklärt.

Eine Übersicht aller vorgestellten Schrittypen und ihrer Parameter ist in Tabelle 6-7 aufgeführt.

Bereich	Funktion	Attribute	Beschreibung
Eingabe	 Spracheingabe	Äußerungen , Einstiegspunkt, ASR- Relevanz	Spracheingabe des Benutzers oder technisches Signal eines Bus-Systems
Ausgabe	 Sprachausgabe	Äußerungen, Wert	Sprachausgabe; bei Verwendung als Zweig einer Verzweigung erfolgt die Ausgabe nur bei Übereinstimmung mit Wert der Variable der Verzweigung
	 Verzweigung	Variable	Entscheidet auf der Grundlage des Wertes eines Variable, welche Ausgabe danach erfolgt
	 Variable ändern	Variable, Wert	Verändert den Wert einer zugewiesenen Variable
	 Animation	Animationsname (Dateiname)	Spielt das in der Datei befindliche Animationskript des Avatars ab
	 3D-Objektbefehl	Befehlsname (3D- Befehl)	Ruft eine CharaScript-Funktion zur Manipulation eines 3D-Objekts zur Laufzeit auf
	 Medium	Medienname (Dateinamen)	Bindet ein Bild oder ein Video in die Anzeige des Avatarsystems ein
	 GUI-Befehl	Befehlsnamen (GUI- Befehl)	Zeigt ein bestimmtes Symbol im Frontend an.
	 Timeout-Befehl	Befehlsnamen (Timeout-Eingabe)	Lässt das Avatarsystem nach einer bestimmten Zeit eine Benutzereingabe simulieren
	 Zufallsgenerator	-	Wählt per Zufall einen der nachfolgenden Sprachausgabeschritte aus
	 Sprungbefehl	Schritt-ID	Erlaubt den Sprung zu einem beliebigen Einstiegspunkt

Tabelle 6-7: *Schritttypen des Dialogeditors und ihre Merkmale*
Quelle: (Eigene Darstellung)

6.2.1.3. Synonymgruppen

Im Laufe der Arbeit mit dem Dialogeditor wurde schnell klar, dass zur Abdeckung einer ausreichend großen Zahl an Eingabevariationen eine sehr große Menge von Formulierungen erfasst werden muss. Je nach Bedienelement, Fragestellung und Position innerhalb des Dialogs waren so circa zwanzig bis fünfzig Formulierungen von vorneherein notwendig, um Benutzereingaben hinreichend abzudecken. Dabei verursachte insbesondere die Verwendung von Synonymen eine Multiplikation der Anzahl an Formulierungen, die manuell erfasst werden mussten. Um diesen manuellen Aufwand zu reduzieren, wurden im Laufe der Szenarienentwicklung Synonymgruppen eingeführt, die als Stammdaten gepflegt werden und beim Export der Regeln automatisch berücksichtigt werden (vgl. von Wendt 2003, 43). Eine Synonymgruppe besteht aus einem Hauptsynonym und bis zu neun weiteren Synonymen (vgl. Abbildung 6-21).

Beim Export wird automatisch nach allen Hauptsynonymen in einer möglichen Benutzereingabe gesucht und zusätzliche Instanzen dieser Eingabe mit den weiteren Synonymen erzeugt. Dieser Vorgang wird für eine mögliche Eingabe mit jeder Synonymgruppe vorgenommen, so dass unbegrenzt viele Synonymgruppen in einer

möglichen Eingabe angewandt werden können. Durch die Anwendung der Synonymgruppen multipliziert sich die Anzahl möglicher Eingaben automatisch und es können mehr mögliche Benutzereingaben erkannt und zugeordnet werden.

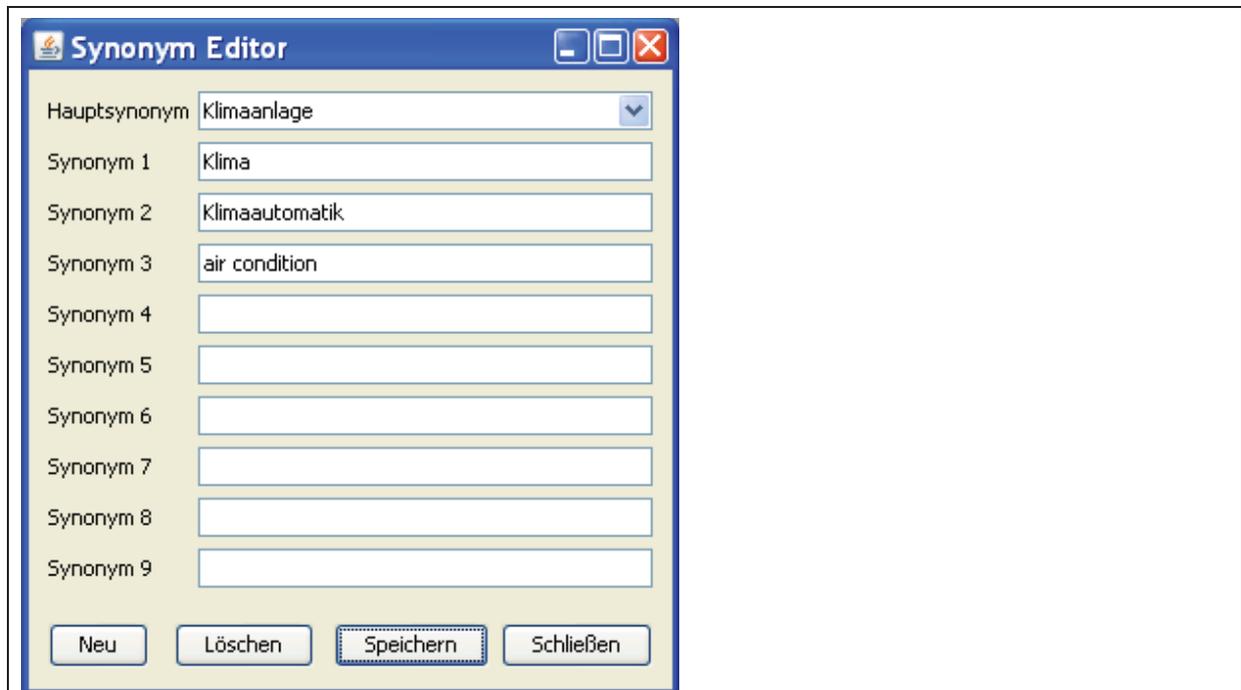


Abbildung 6-21: *Pflege einer Synonymgruppe im Synonymeditor*
Quelle: (Eigene Darstellung)

So werden aus der in Abbildung 6-22 dargestellten Formulierung „Wie kann ich die Klimaanlage einschalten“ durch Anwendung von zwei Synonymgruppen insgesamt sechzehn verschiedene Formulierungen generiert. Dabei muss darauf geachtet werden, in einer Formulierung das Hauptsynonym zu verwenden.

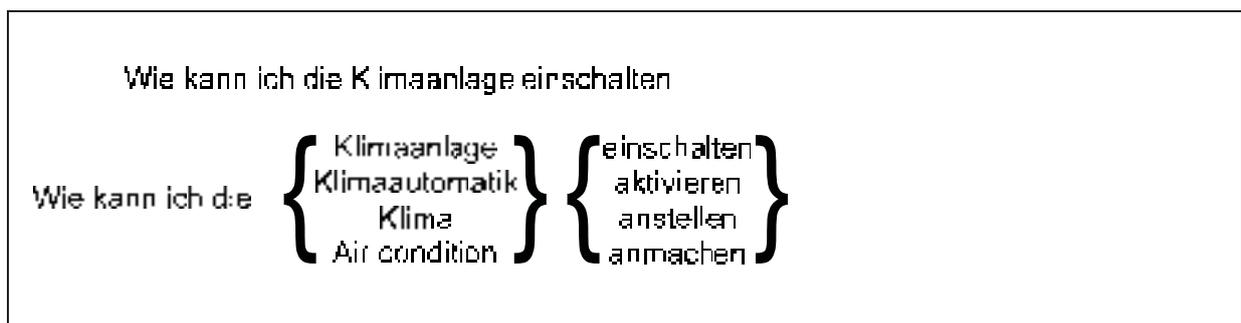


Abbildung 6-22: *Multiplikation von Eingabeformulierungen durch Synonymgruppen*
Quelle: (Eigene Darstellung)

Bei der Pflege der Synonyme ist lediglich darauf zu achten, dass das Hauptsynonym nicht Teil eines weiteren Synonyms in der Gruppe ist, da es sonst zu einer Endlosschleife beim Ersetzen kommen kann. So führt das Wort „Klima“ als Hauptsynonym mit dem Synonym „Klimaanlage“ in der gleichen Gruppe dazu, dass auch im Synonym „Klimaanlage“ das Hauptsynonym erkannt und ersetzt wird. Bei zu umfangreicher Gestaltung der

Synonymgruppen besteht zudem die Gefahr, semantisch sinnlose Varianten zu generieren, die nur die Spracheingabe unzuverlässiger machen. Ein Problem, das in der aktuellen Version des Avatarsystems noch nicht angegangen wurde, ist die Auftrennung von Verben aufgrund des Satzbaus. Erfolgt beispielsweise die Eingabe „Wie schalte ich die Klimaanlage ein?“ wird nicht erkannt, dass das Verb „einschalten“ verwendet wird, wodurch es nicht durch Synonyme ersetzt wird. Eine Möglichkeit, die zu beheben, wäre die Angabe von Synonymen mithilfe von Platzhaltern, d.h. beispielsweise in der Form „schalte * ein“ und „schalte * an“.

6.2.1.4. Austauschfunktionen des Dialogeditors

Im Rahmen des Transfers von Daten in und aus dem Dialogeditor können Imports und Exports verschiedener plattformabhängiger Formate vorgenommen werden.

Import von SGML

SGML-Tag	Beschreibung	Konvertierung
<H-KAP> <TITEL> ... </TITEL> </H-KAP>	Überschrift auf 1. Ebene, <Titel> enthält den Namen des Kapitels, gegebenenfalls mit Indizes (<IDX>)	Szenario
<KAP> <TITEL>... </TITEL> </KAP>	Überschrift zweiter Ordnung, <Titel> beinhaltet den Namen des Kapitels, gegebenenfalls mit Indizes (<IDX>)	Subszenario
<MODUL> <TITEL>... </TITEL> </MODUL>	Überschrift dritter Ordnung, <Titel> beinhaltet den Namen des Kapitels, gegebenenfalls mit Indizes (<IDX>)	Dialog
<VERWENDG> <EIGENSCHAFT>...</EIGENSCHAFT> <EIGENSCHAFT-TEXT>... </EIGENSCHAFT-TEXT> </VERWENDG>	Beinhaltet die vorausgesetzten Ausstattungsmerkmale für ein Modul, <Eigenschaft> und <Eigenschaft-Text> beziehen sich auf disjunkte Ausstattungsmerkmale	Variable und Verzweigungsschritt mit entsprechender Variable
<BESCHREIBG> <B-BLOCK> <ANLEITG> <ACHTUNG> <VORSICHT> <HINWEIS>	Verschiedene Tag-Paare umschließen einen Textblock im Handbuch, der als Beginn einer Erklärungssequenz umgesetzt wird	Spracheingabeschritt
<ABSATZ> <PUNKT> <HANDLUNG>	Verschiedene Tags markieren je nach umschließenden Textblock einzelne Absätze des Handbuchs	Sprachausgabeschritt(e) mit Timeout-Schritten zur Weiterleitung, Äußerungen
<ILLU> <BILD NAME="..."> <BILD-TITEL>...</BILD-TITEL> </ILLU>	Einbindung eines Bildes, <Name> ist der Alias für einen Dateinamen; <Bild-Titel> beinhaltet die Bildunterschrift	Media-Schritt und Mediaausprägung mit Dateinamen

Tabelle 6-8: Wichtige SGML-Tags des Audi Benutzerhandbuchs und ihre Konvertierung

Quelle: (Eigene Darstellung)

Die Standard Generalized Markup Language (SGML), zu deren Teilmengen beispielsweise HTML und XML gehören, wird bei der AUDI AG zur Speicherung der Daten des Benutzerhandbuchs verwendet. Aus dieser strukturierten Form werden sowohl das gedruckte

Handbuch als auch die Derivate des Online-Bordhandbuchs oder der Handbuch-CD für den Heim-PC generiert. Dabei erlaubt eine Vielzahl selbstdefinierter Tags eine detaillierte Strukturierung der SGML-Dateien. Tabelle 6-8 gibt dabei einen Überblick der wichtigsten Tags, die auch zum Import von SGML-Dateien im Dialogeditor herangezogen werden.

Da die Schriftform der Erklärung von Funktionen oft Sätze beinhaltet, die aufgrund ihrer Länge und der damit verbundenen, schwindenden Aufmerksamkeit des Hörers nicht für die akustische Wiedergabe geeignet sind, werden in solchen Fällen Absätze auf mehrere Sprachausgabeschritte verteilt. Benutzer können entweder proaktiv zum nächsten Sprachausgabeschritt navigieren oder warten die Ausführung des Timeout-Befehls ab.

Neben den dargestellten Tags werden noch weitere Anweisungen für Tabellen, die Einordnung eines Moduls in einen Bereich der Computer-CD, Querverweise, Synonyme und Hervorhebungen in der vorliegenden Version des Dialogeditors beim Import nicht verarbeitet. Während sich tabellarische Darstellungen für die Verarbeitung im Avatarsystem nicht eignen, könnten sich Querverweise prinzipiell als Hinweis des Avatars auf ähnliche Themen umsetzen lassen. Synonyme und Indizes könnten in automatisch generierte Regeln eingebunden werden, die z.B. eine Frage nach dem Muster „Was ist ...“ beantworten könnten. Hervorhebungen im Text könnten sich als prosodische Betonungen umsetzen lassen.

Export von AIML

Der Export des plattformunabhängigen Modells erfolgt aufgrund der späteren Verwendung im Chatbot des Avatarsystems nur in AIML. Die Formulierungen eines Eingabeschrittes werden zum Inhalt des AIML-Tags <pattern>. Liegen mehrere Formulierungen vor, so wird für jede mögliche Spracheingabe eine eigene Regel mit der gleichen Reaktion generiert. Verzweigungen werden mit den <condition>-Tags abgebildet, wobei die im Verzweigungsschritt eingebundene Variable mit den Werten der nachfolgenden Sprachausgabeschritte herangezogen werden. Die Formulierungen der Sprachausgabe werden unverändert übernommen, so dass sie AIML-erfahrenen Benutzern die Möglichkeit gibt, auch direkt AIML-Tags einzubinden. Liegen mehrere mögliche Formulierungen vor, so wird eine davon mit <random>-Tags per Zufall zur Laufzeit ausgegeben. Änderungen an Variablen werden mit <set>-Tags umgesetzt, wobei die umschließenden <think>-Tags die Ausgabe des Variablenwertes in der Sprachsynthese verhindern. Sprungschritte werden mit den <srai>-Tags umgesetzt, die das anspringen einer anderen Regel bewirken. Funktionen wie der Aufruf einer Animation oder die Steuerung eines 3D-Objekts, werden durch Einsetzen des entsprechenden Befehls in die interne Syntax des Avatarsystems konvertiert. Am Ende jeder Regel wird automatisch der Name des aktuellen Szenarios als Thema mit <topic>-Tags gesetzt. Dadurch wird die Menge der möglichen nachfolgenden Schritte eingeschränkt. Das Setzen eines neuen Themas mit diesem Tag wird zudem auch vor dem Absprung in einen Dialog in einem anderen Szenario mithilfe von <srai> durchgeführt.

Abbildung 6-23 zeigt ein Beispiel, dass die Transformation aller genannten Schrittypen mit Ausnahme des Medienschrittes demonstriert. Nach der Begrüßung des Benutzers, die durch einen Einstiegspunkt mit verschiedenen Formulierungen repräsentiert wird, erfolgt eine Verzweigung. Je nach dem, ob der Benutzer das Avatarsystem bereits verwendet hat, stellt

sich der Avatar zunächst vor und ändert die entsprechende Variable oder er beginnt, indem er den Benutzer seine Hilfe anbietet und zudem eine Begrüßungsanimation abspielt und im Hintergrund ein Audi-Fahrzeug einblendet, das sich dreht. Darüber hinaus wird ein entsprechendes 3D-Symbol angezeigt und ein Timeout-Befehl angestoßen.

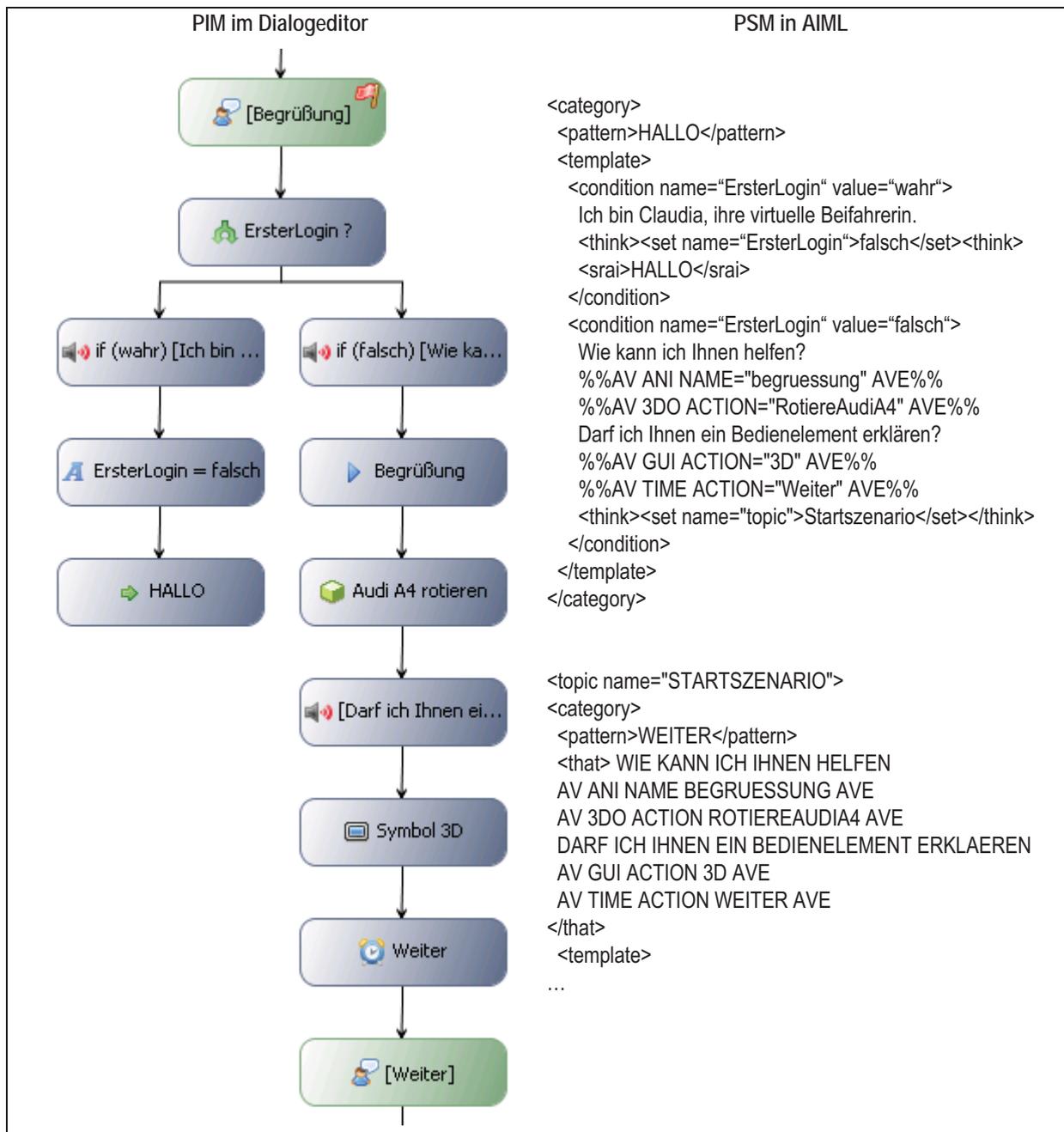


Abbildung 6-23: Umwandlung des PIM in ein AIML-PSM
Quelle: (Eigene Darstellung)

Im Beispiel wird zudem auch der Anfang der nachfolgenden Regel gezeigt, die kein Einstiegspunkt ist und deshalb dem aktuellen Thema und der Dialoghistorie Rechnung tragen muss. Dementsprechend schränkt das `<topic>`-Tag den Aufruf der Regel ein. Zudem beinhaltet das `<that>`-Tag den konvertierten Inhalt der letzten Antwort. Da die letzte Antwort

bei der Verwendung eines `<random>`-Tags verschieden ausfallen kann, muss für jede mögliche vorangegangene Antwort eine gesonderte Regel mit jeweils unterschiedlichem `<that>`-Inhalt generiert werden. Bei der Generierung der Inhalte von `<that>`-Tags werden während der Verarbeitung durch das ChatterBean Sonderzeichen entfernt oder in Leerzeichen umgewandelt. Da die Entscheidung, ob ein Sonderzeichen ersetzt oder gelöscht wird, nicht mit den Einstellungen in der Datei `substitution.xml` des ChatterBeans konsistent ist, erforderte dies eine gesonderte Abstimmung der Konvertierungsregeln für den Export im Dialogeditor.

Durch die Verwendung mehrerer Formulierungen in einer Äußerung aufgrund verschiedener möglicher vorhergehender Ausgaben steigt die Anzahl generierter Regel stark an, wodurch die Regelbasis bei der Pflege ohne einen solchen Dialogeditor schnell unüberschaubar werden würde.

Neben den eigentlichen Regeln wird zudem für jede Regel automatisch eine entsprechende Regel zum Wiederholen der letzten Antwort und zum Zurückspringen zur vorhergehenden Antwort generiert. Dabei beinhaltet jede dieser Schattenregeln als Eingabe die eindeutige ID des Eingabeschritts sowie eine der möglichen Formulierungen. Mit diesem Schema `<ID><Formulierung>` können somit auch Eingabeschritte angesprungen werden, die nicht explizit als Einstiegspunkte deklariert wurden. Die Funktionen „Wiederholen“ und „Zurück“ machen zudem die Speicherung der Historie aufgerufener ID's notwendig, was mithilfe mehrerer *user predicates* umgesetzt wird. Dabei reicht es aus, die jeweiligen IDs ohne die Eingabeformulierung zu speichern, da auch beim Aufruf `<srai>123 *</srai>` zu der Regel mit der entsprechenden ID gesprungen werden kann.

Das AIML-Modell der Spracheingabeschritte wird abschließend herangezogen, um die Sprachgrammatik zu generieren. Dabei werden die Inhalte der `<pattern>`-Tags für den Aufbau der Grammatik verwendet. Für Eingabeschritte des Systemereignistyps wird eine gesonderte AIML-Datei erzeugt, bei der die Umwandlung in eine Sprachgrammatik entfällt.

6.2.2. Auswahl von Szenarien

Im Rahmen der prototypischen Entwicklung des Avatarsystems können nicht alle Inhalte, die in der Konzeption vorgesehen sind, umgesetzt werden. Die Auswahl der Inhalte, welche im Dialogeditor zu Dialogen des Avatarsystems aufgearbeitet werden, hängt von verschiedenen Aspekten ab, die kurz beleuchtet werden.

Notwendigkeit und Bedeutung: Da das Avatarsystem selbst und dessen Bedienung nur wenige Parallelen zu softwaregestützten Systemen des Alltags haben, ist eine strukturierte Einführung notwendig. Dabei wird zum einen die sprachliche Interaktion, mögliche Optionen in Dialogen und die Bedeutung der Symbole in der Anzeige vorgestellt.

Eignung für den Prototyp: Der prototypische Charakter des Avatarsystems im Fahrzeug bedingt die Anpassung des Systems an die bereits serienmäßig im Fahrzeug befindlichen Systeme, die ihrerseits nicht für das Avatarsystem angepasst werden können. So werden bei der Wahl der Szenarien keine Funktionen des primären Funktionsbereichs und im sekundären Bereich nur unterstützende Funktionen berücksichtigt. Dies ist sinnvoll, um bis zur

Untersuchung des Ablenkungspotenzials eines Avatarsystems im Fahrzeug nicht bei kritischen Funktionen dem Fahrer reinzureden.

Eine weitere Einschränkung bei der Auswahl von Szenarien ergibt sich durch die **Verwendung des MMI-Displays** für die Darstellung der Benutzeroberfläche des Avatarsystems. Dadurch, dass die Anzeige des Avatarsystems durch Betätigen der Funktionstaste für Verkehrsinformationen eingeblendet wird (vgl. Kapitel 6.1.1), wird der Avatar wie der Modus „Verkehrsinformationen“ behandelt, wodurch keine anderen Anzeigen gleichzeitig im MMI-Display erfolgen können. Dieses Verhalten führt dazu, dass für die Szenarien des Avatarsystems keine Funktionen verwendet werden können, die das MMI-Display als primäre Anzeige nutzen. Funktionen wie z.B. die Klimaanlage, die das MMI-Display als zusätzliche Anzeige verwenden, werden bei der Funktionenauswahl jedoch berücksichtigt.

Einbindung als Tutorial: Für jedes Szenario des Avatarsystems soll sowohl eine informative als auch eine tutorielle Variante angeboten werden (vgl. Krause 1988). Ein Tutorial kann den Fahrer nur dann optimal unterstützen, wenn dem Avatarsystem Informationen über die Handlungen des Fahrers vorliegen. Solche Informationen werden bei Betätigung eines Bedienelements über den entsprechenden CAN-Bus versendet. Die meisten Aktionen des Fahrers und über Sensoren gemessene Änderungen des Umfelds werden so weitergegeben. Der Prototyp des Avatarsystems im Fahrzeug verfügt dabei nur über eine Anbindung an den Komfort CAN-Bus (vgl. Kapitel 6.1.3.2). Damit lässt sich die Bedienung der meisten Elemente im Fahrzeuginnenraum ermitteln, die keinen Einfluss auf primäre Fahrzeugfunktionen haben.

Verwendung für die Evaluation: Die ausgewählten Szenarien dienen neben der Entwicklung einer Methode zur Inhaltsaufarbeitung auch als Basis der Evaluation des Avatarsystems. Dadurch muss bereits bei der Auswahl der Szenarien die spätere Verwendung im Testträger unter Evaluationsbedingungen berücksichtigt werden. Somit können nur Funktionen in Betracht kommen, die im späteren Testfahrzeug auch vorhanden sind. Für die Vergleichbarkeit der Evaluation im Stand und während der Fahrt können zudem nur Funktionen gewählt werden, die in beiden Zuständen verwendet werden können. Dabei ist die Bedienung im Stand als größere Einschränkung zu betrachten, da die meisten Fahrzeugfunktionen für den Einsatz während der Fahrt konzipiert sind.

Um die Beeinflussung der Evaluationsergebnisse durch die gewählten Funktionen gering zu halten, müssen die Funktionen, die in den untersuchten Szenarien zum Einsatz kommen, verschieden sein. Als Unterscheidungskriterium bietet sich dabei die Komplexität und damit verbunden die Länge der Szenarien an. Diese wird maßgeblich durch die Anzahl an Bedienschritten innerhalb eines Szenarios beeinflusst. Unter einem Bedienschritt wird dabei eine Bedienungsart eines Bedienelements verstanden. So kann das Bedienelement „Scheibenwischerhebel“ in der Bedienart zur Aktivierung der Scheibenwischer oder der Bedienart zur Steuerung der Bordcomputer verwendet werden. Ein Szenario bzw. eine Aufgabenstellung für die Evaluation stellt somit eine Reihe von Bedienschritten dar, die einer alltäglichen Situation entsprechen. Zur Streuung der Szenarien werden diese in drei Kategorien anhand der enthaltenen Bedienschritte eingeteilt:

- Kategorie 1 umfasst einen Bedienschritt,
- Kategorie 2 umfasst zwei bis fünf Bedienschritte,
- Kategorie 3 umfasst sechs bis zehn Bedienschritte.

Eine vollständige Aufstellung möglicher Bedienschritte ohne die Nutzung des MMI-Displays und der wichtigsten Merkmale hinsichtlich des Einsatzes in einem Szenario des Avatarsystems ist in Anhang A zu finden. Tabelle 6-9 zeigt einen Auszug dieser Übersicht und stellt die ausgewählten Bedienschritte und deren Zusammenstellung zu Szenarien für die Evaluation und damit die Inhaltsaufarbeitung dar. Da es sich beim einführenden Avatarszenario nicht um fahrzeuggetriebenen Inhalt handelt, ist es nicht in der Liste aufgeführt.

Szenario	Funktion	Funktionsbereich	Bedienelement	Bedienart	CAN-Bus
Audi lane assist	Audi lane assist einschalten	Sekundär	Taste am Ende des Blinkerhebels	Taste drücken	Komfort
Audi side assist	Audi side assist einschalten	Sekundär	Audi side assist Taste, innen am Außenspiegel	Taste drücken	Komfort
Sitzmemory	Sitzmemory aktivieren	Tertiär	Taste „Off“ im Sitzmemory in Fahrertür	Taste drücken	Keiner
	Speicherung vorbereiten		Taste „Set“ im Sitzmemory in Fahrertür	Taste drücken	Komfort
	Sitzmemory Speicherung		Taste „1“ oder „2“ im Sitzmemory in Fahrertür	Taste drücken	Komfort
Bord-computer	Anzeige wechseln	Tertiär	„Reset“ Taste am Scheibenwischerhebel	Taste drücken	Komfort
	Angezeigten Wert wechseln		„Auf“- und „Ab“- Tasten am Scheibenwischerhebel	Taste drücken	Komfort
Klima-automatik plus	Aktivieren	Tertiär	Taste „Off“	Taste drücken	Komfort
	Temperatur einstellen		Drehregler	Drehregler drehen	Komfort
	Gebläse einstellen		Taste mit Piktogramm für Gebläse	Taste drücken	Komfort
			Drehregler	Drehregler drehen	Komfort
	Umluftbetrieb aktivieren		Taste mit Piktogramm für Umluftbetrieb	Taste drücken	Komfort
	Deaktivieren		Taste „Off“	Taste drücken	Komfort

Tabelle 6-9: *Bedienschritte und ihre Zusammenstellung zu Evaluationsszenarien*
Quelle: (Eigene Darstellung)

6.2.3. Konvertierung von Inhalten in Dialoge

Die Bereitstellung der Inhalte für ein Avatarsystem umfasst neben der Gestaltung der Regelbasis die Erstellung, Anpassung und Integration von multimedialem Material wie Bildern Videos oder 3D-Objekten. Ausschlaggebend ist dabei jedoch die Information, die durch Dialoge vermittelt werden soll. Dabei lassen sich neben strukturierten und bereits dokumentierten Informationen wie den Anweisungen im Benutzerhandbuch eines Fahrzeugs auch unstrukturierte Informationen unterscheiden (vgl. Pfaffmann 2000). Dieses unstrukturierte Wissen, das auch als *tazites Wissen* bezeichnet wird, wurde dabei noch nicht gesammelt, zueinander in Beziehung gesetzt und entsprechend festgehalten.

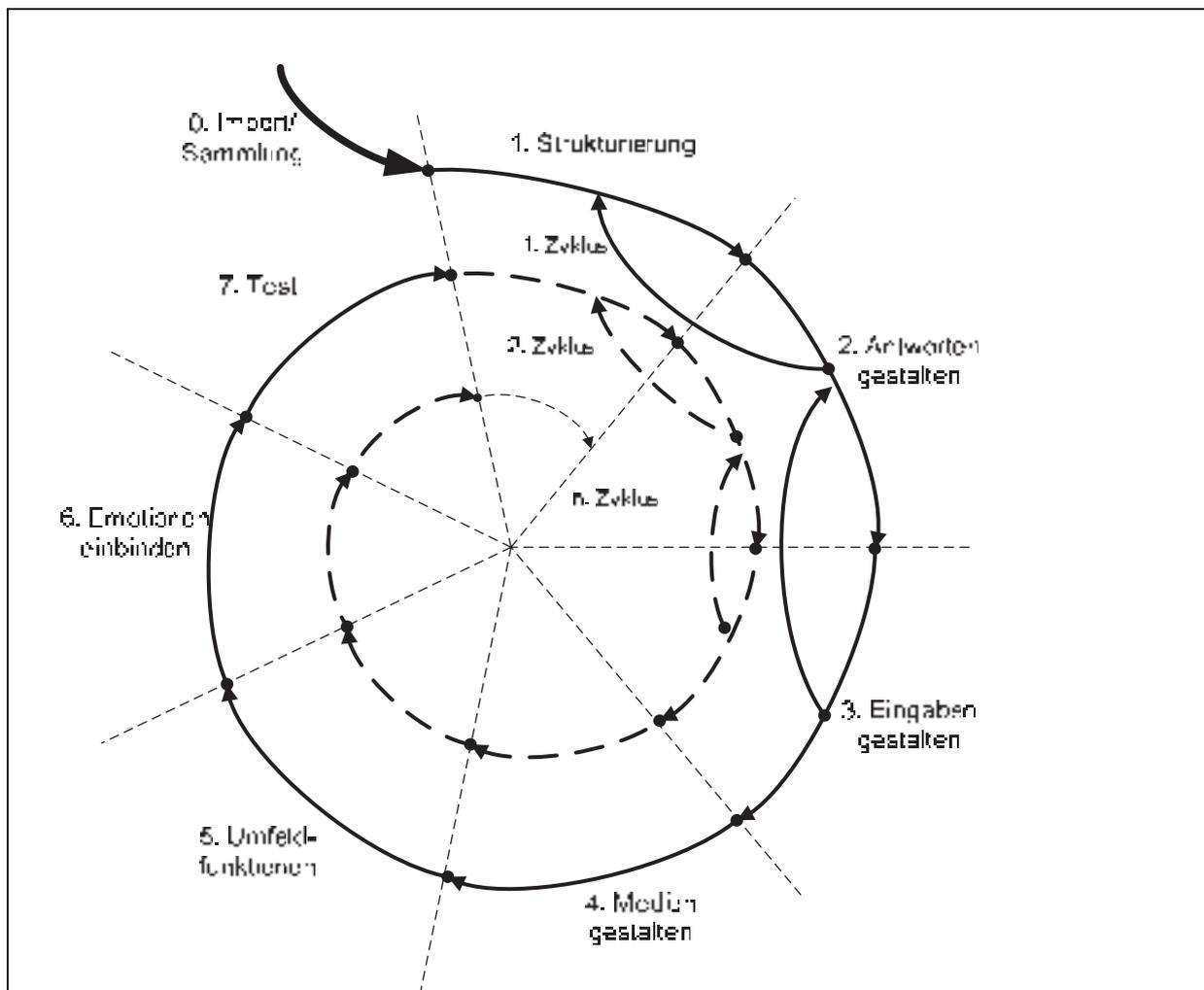


Abbildung 6-24: *Vorgehensmodells zur Konvertierung von Informationen in einen Dialog*
Quelle: (Eigene Darstellung)

Trotz dieser Unterschiede haben sich im Rahmen der Aufbereitung von Inhalten für das Avatarsystem ab einem bestimmten Zeitpunkt weitgehende Übereinstimmungen im Umgang mit den beiden Informationstypen ergeben. Die Anordnung der übereinstimmenden Schritte bei der Bearbeitung führte zu dem in Abbildung 6-24 dargestellten Vorgehensmodell in Form einer Spirale. Die Anordnung als Spirale versucht dabei der von Glass und Seneff (2003) beobachteten Problematik des Henne-Ei-Verhältnisses zwischen der Erstellung und der

Verbesserung von Inhalten für Dialogsysteme Rechnung zu tragen. Neben dem grundlegend verschiedenen Schritt des Imports bzw. der Sammlung von Informationen lassen sich sieben Phasen identifizieren, die in beiden Fällen erfolgen können. Die Spirale kann in mehreren Zyklen durchlaufen werden, wobei mit jedem Durchlauf der Anteil der korrekt beantworteten Fragen zunimmt (vgl. von Wendt 2003, 42) und somit ähnlich einem hermeneutischen Zirkel die Annäherung an das vollständige Verständnis erfolgt. Die Anzahl der Zyklen ist nicht begrenzt und hängt lediglich von den individuellen Zielvorgaben ab. Die nachfolgende Betrachtung der einzelnen Phasen sowie der darin erfolgenden Schritte basiert auf dem bereits vorgestellten Werkzeug des Dialogeditors.

Der Import von strukturierten Informationen wurde bereits in Kapitel 6.2.1.3 im Rahmen des Dialogeditors beschrieben. Die Sammlung und Zusammenstellung von unstrukturierten Informationen sowie weitere Besonderheiten bei diesem Typ werden später in Kapitel 6.2.3.8 dargestellt.

Die erste gemeinsame Bearbeitungsphase stellt die Strukturierung dar, welche die dialoggerechte Anordnung der verschiedenen Schritttypen im Dialogeditor beinhaltet. Die Gestaltung der Antworten der Regelbasis umfasst die angemessene Anpassung der Texte sowie gegebenenfalls die Einsatzplanung verschiedener Medien. Durch die Formulierung der Antworten können sich Änderungen auf die Struktur der Dialoge und Szenarien ergeben, so dass bereits innerhalb eines Zyklus die Notwendigkeit entstehen kann, erneut in die erste Phase zurückzukehren. Die Gestaltung der Antworten wird vor der Zusammenstellung möglicher Eingaben vorgenommen, da die Informationen, die in den Antworten enthalten sein sollen, meist bereits feststehen und eine endliche Menge darstellen, die meist nicht mehr erweitert wird. Anders hingegen verhält es sich mit der dritten Phase, die der Erfassung verschiedener Eingabemöglichkeiten des Benutzers dient. Die Vielzahl der möglichen Eingaben und Reaktionen des Benutzers machen in dieser Phase viele Anpassungen und Erweiterungen notwendig. Von dieser Phase können Rückschritte zur Gestaltung der Antworten notwendig werden, wenn die Aufteilung oder Differenzierung von Antworten notwendig wird. Die Gestaltung und Einbindung von Medien ist Gegenstand der vierten Phase. Unter Umfeldfunktionen werden alle nicht zwingend für einen fehlerfreien Dialog erforderlichen Regeln zusammengefasst. Die optionale Phase der Emotionen befasst sich mit der Abbildung eines emotionalen Modells des Avatars bzw. des Benutzers. Die letzte Phase beinhaltet Tests, die verschiedene Aspekte der Regelbasis abprüfen können.

Die klare Aufteilung der verschiedenen Arbeitsschritte in Phasen innerhalb des Spiralmodells erlaubt nicht nur einen sequentiellen Ablauf, sondern die Aufteilung auf verschiedene Personen bzw. Personengruppen (vgl. Tabelle 6-10). So kann die Strukturierung durch einen Redakteur in Zusammenarbeit mit einem Linguisten erfolgen. Die Gestaltung der Antworten wird maßgeblich durch einen Texter bestimmt, während die Sammlung möglicher Eingaben einen Linguisten oder Kommunikationswissenschaftler erforderlich macht, um verschiedene Formulierungen und Benutzerverhalten absehen zu können. Die Gestaltung und Einbindung von Medien erfolgt durch Grafiker und mithilfe eines Psychologen, um die nichtverbale Kommunikation des Avatars gezielt einzusetzen. Die möglichen Umfeldfunktionen können durch einen Linguisten erstellt werden, der mögliche Nach- oder Zwischenfragen absehen und integrieren kann. Die Ausgestaltung eines emotionalen Modells ist Aufgabe eines

Psychologen. Der Test am Ende eines Durchlaufs wird zum einen von Testpersonen und einem Testleiter bestimmt.

Phase	Personengruppen
Strukturierung	Redakteur, Linguist
Antworten gestalten	Texter, Redakteur
Eingaben gestalten	Linguist/ Kommunikationswissenschaftler
Medien gestalten	Grafiker, Psychologe/ Kommunikationswissenschaftler
Umfeldfunktionen	Linguist
Emotionen einbinden	Psychologe
Test	Testpersonen, Testleiter

Tabelle 6-10: *Mögliche Verteilung von Aufgaben an Personengruppen*
Quelle: (Eigene Darstellung)

Auf die einzelnen Phasen und deren Ausgestaltung wird im Folgenden detailliert eingegangen. Dabei werden in jeder Phase verschiedene Arbeitsschritte identifiziert, die durchlaufen werden müssen. Bei all diesen Schritten sollen zudem phasenabhängige Richtlinien als Leitfaden dienen.

6.2.3.1. Phase 1: Strukturierung von Informationen

André, Rist und Müller (1998b) unterscheiden in ihrem Avatarprojekt zwischen einem Ablaufmodell, das grob wichtige Punkte einer Gesprächs darstellt, und dem eigentlichen Interaktionsmodell, in dem die entsprechenden Eingaben und Ausgaben an der jeweiligen Stelle des Ablaufmodells festgehalten sind. Ein ähnliches Vorgehen legen auch Stede und Schlangen (2004, 5f) an den Tag, wobei sie in einen terminologischen und einen inhaltlichen Teil unterscheiden. Diesem Gedanken der grundsätzlichen Zweiteilung wird auch in der ersten Phase des Vorgehensmodells Rechnung getragen. Unabhängig von der konkreten Ausgestaltung der Inhalte, soll der korrekte Ablauf von Dialogen sichergestellt werden.

Direkt nach dem Import der Informationen aus dem Benutzerhandbuch liegen sie in einer Vielzahl von Dialogen vor, die entsprechend der Handbuchkapitel geordnet sind. Auch wenn die inhaltliche Nähe durch eine entsprechende Anordnung in Kapiteln gegeben ist, stehen sie unabhängig voneinander im Raum. Dadurch wird dem Benutzer der Zugang zu Informationen erschwert, deren Existenz ihm nicht bewusst ist. Erst durch die sinnvolle Verknüpfung der verschiedenen Inhalte bzw. Dialoge kann im Sinne einer aktiven Gesprächsführung auf verwandte Inhalte verwiesen werden. Diese Verknüpfung muss zum einen technisch erfolgen, indem ein entsprechender Sprungschritt eingefügt wird. Zum anderen muss verbal auf den verwandten Inhalt hingewiesen werden oder verdeutlicht werden, dass bereits in ein anderes Themengebiet abgesprungen wurde. Die verbale Überleitung erfolgt dabei bei der Pflege der Antworten in der zweiten Phase. So muss beispielsweise am Ende des Dialogs zur tiptronic-Gangschaltung auch auf die Möglichkeit verwiesen werden, die Steuerung der Schaltung über das Lenkrad zu erklären. Im Rahmen der Verknüpfung von Unterkapiteln des Handbuchs bzw. Teilen davon muss zudem geprüft werden, ob die Reihenfolge der Beschreibung im

Handbuch in einem Dialog noch angemessen ist. Während in einem gedruckten Buch eine starre Reihenfolge festgelegt werden muss, kann in einem Dialog eine Verzweigung zu verschiedenen Inhalten angeboten werden und damit ein dynamischer Ablauf der Informationen erfolgen. Die interaktive Bestimmung der vorgetragenen Inhalte im Avatarsystem macht zudem die redundante Verteilung von Informationen im Handbuch obsolet. So können gleiche Informationen zusammengeführt und von verschiedenen Dialogen auf sie verwiesen werden. Im Rahmen dieser Umstrukturierung kann es auch sinnvoll sein, automatisch generierte „Spaghetti-Dialoge“ mit unzähligen Schritten in mehrere Dialoge bzw. Stränge innerhalb eines Dialogs aufzuteilen.

Ein wichtiges Strukturelement sind Verzweigungen auf der Basis von Variablen wie beispielsweise Ausstattungsmerkmalen. Die Erstellung einer Variablen für die Abbildung eines Ausstattungsmerkmals erfolgt zwar während des Imports automatisch, jedoch sind dabei manuelle Nacharbeiten vorzunehmen. Zum einen muss die konsistente Verwendung von Bezeichnungen geprüft werden, da dies in der SGML-Version des Benutzerhandbuchs nicht immer gegeben ist. Zudem muss die Vollständigkeit der abgeprüften Ausstattungsmerkmale gewährleistet werden. An manchen Stellen im Handbuch werden nicht alle relevanten Ausstattungsmerkmale angegeben. Zudem wird die Relevanz von Ausstattungsmerkmalen für größere Textblöcke angegeben, was dazu führen kann, dass ein Textblock für zwei mögliche Merkmalsausprägungen als relevant gekennzeichnet ist während im Text zwischen den beiden Ausprägungen unterschieden wird. So sind Textblöcke im Handbuch beispielsweise für tiptronic- und multitronic-Schaltungen relevant, wobei im Textblock selbst wiederum zwischen den beiden Varianten unterschieden wird. Zudem ist es notwendig, für den Fall dass ein Fahrzeug nicht über das abgeprüfte Ausstattungsmerkmal verfügt, ein alternatives Informationsangebot vorzuschlagen. Bei der Verwendung von Variablen zur Verzweigung erweist sich der eingebundene Chatbot als starr, da nur diskrete Zustände, aber beispielsweise keine Wertebereich überprüft werden können.

Gemäß der Unterteilung von Krause (1988) in tutorielle Systeme, die eine interaktive Wissensvermittlung ermöglichen und informativen Auskunftssystemen, die als Nachschlagewerk dienen, muss bei den Antwortsträngen des Avatarsystems unterschieden werden. Während der Import des Benutzerhandbuchs bereits den informativen Teil bereitstellt, muss auf dessen Grundlage der interaktive Teil abgeleitet werden. Dabei muss das Avatarsystem dem Benutzer einerseits die entsprechende Auswahl anbieten und zum anderen neben verbalen Eingaben auch Systemereignisse wie das Betätigen eines bestimmten Bedienelements verarbeitet werden. So kann das Avatarsystem z.B. Schritt für Schritt mit dem Benutzer zusammen die Einstellung einer neuen Geschwindigkeit für das ACC vornehmen. Diese Unterscheidung in informative und tutorielle Szenarien macht jedoch nicht immer Sinn, da der Fahrer meist nach einer bestimmten Information fragt in der Absicht diese zu bedienen. Aus diesem Grund kann sich je nach Bedienelement auch eine Kombination von beiden Paradigmen als angemessen erweisen.

Bei der Umsetzung der genannten Schritte im Rahmen dieser Phase sind einige allgemeine Richtlinien zu befolgen. Grundsätzlich muss berücksichtigt werden, dass der Aufbau unabhängig von der konkreten Gestaltung von Ein- und Ausgaben erfolgt. Bei der Reorganisation von Dialogen und ggf. Szenarien muss großzügig geplant werden, da durch

folgende Schritte und Zyklen eine Vielzahl weiterer Schritte ergänzt werden. Die Überleitung von einem Dialog zum nächsten sollte im Idealfall den Durchlauf aller Handbuchinhalte ermöglichen. Bereits bei der Struktur der Dialoge sollte darauf geachtet werden, dass Ein- und Ausgaben einfach wiederverwendbar sind d.h. sie sollten sowohl für den tutoriellen und den informativen Einsatz genauso wie für gezielte Fragen zu Details verwendet werden können. Inhaltliche Bereiche, die nicht bereits durch den Import bereitgestellt werden, aber aufgrund der Strukturierung notwendig geworden sind, sollten mit leeren Schritten und entsprechenden Titeln angedeutet werden, um den Sinn der Strukturierung später in Phase zwei und drei nachvollziehen zu können. Tabelle 6-11 gibt einen Überblick der beschriebenen Schritte und Richtlinien der Strukturierungsphase.

Bereich	Bezeichnung
Schritte	Folge von Dialogen und Inhalten überprüfen
	Überleitungen zwischen Dialogen einbinden
	Konsolidierung und Prüfung der konsistenten Bezeichnung von Variablen für Ausstattungsmerkmale
	Erhöhung der Granularität der Merkmalsprüfungen
	Angebot von Informationsalternativen
	Unterscheidung zwischen tutoriellen und informativen Abläufen
Richtlinie	Struktur der Dialoge ist unabhängig von der Gestaltung von Ein- und Ausgaben
	Großzügiges Planen und Organisieren von Dialogen und Szenarien
	Verknüpfung aller Inhalte miteinander
	Wiederverwendbarkeit von Inhalten bei Strukturierung beachten
	Verwendung von Titeln für vorgesehene inhaltliche Bereiche

Tabelle 6-11: Schritte und Richtlinien für die Phase der Strukturierung

Quelle: (Eigene Darstellung)

6.2.3.2. Phase 2: Gestaltung von Antworten

Die Gestaltung der Antworten ist neben den Eingaben die wichtigste Phase in einem Zyklus, da sie zum einen direkt den Informationsgehalt und zum anderen dessen angemessene Präsentation bestimmt. Da diese Phase kritisch für den Erfolg und die Akzeptanz des Avatarsystems ist, gestalten sich sowohl die Schritte als auch die Richtlinien umfangreicher.

Schritte der Antwortphase

Neben der übergeordneten Struktur von Dialogen und Szenarien lässt sich auch in einer Antwort eine bestimmte Struktur, d.h. die Anordnung der einzelnen Schritttypen identifizieren. Zunächst ist im Rahmen der Struktur zu prüfen, ob die Reihenfolge der generierten Ausgabeschritte im Kontext des Avatarsystems sinnvoll ist. So sind beispielsweise häufig Bilder des gedruckten Handbuchs erst am Ende eines Textblocks oder an einer anderen für das Layout günstigen Stelle eingebunden. Im Rahmen der Antwort des Avatars muss die Darstellung eines Mediums jedoch an der Stelle der verbalen Erklärung erfolgen, weshalb das Verschieben von Mediensritten vorgenommen werden muss. Das gedruckte Handbuch unterliegt hinsichtlich der Medien auch der Einschränkung, nicht alle

Vorgänge bebildern zu können bzw. keine bewegten Bilder darstellen zu können. Dies macht die Ergänzung der importierten Medien um Neue notwendig.

In manchen Fällen kann ein Sachverhalt auch besser durch ein Video oder ein animiertes 3D-Objekt verdeutlicht werden. Dabei liegen viele Videos oder 3D-Objekte in der Phase der Antwortgestaltung noch nicht vor, so dass die entsprechenden Schritte zwar bereits eingebunden, aber erst in der vierten Phase, der Mediengestaltung, erstellt und in den Schritt eingefügt werden. Um eine optimale Abstimmung zwischen Antworttexter und Mediengrafiker sicherzustellen, bietet sich die gesonderte Pflege einer Medienliste an, in der Dateinamen, Inhalte und Timing von Animationen und deren Verwendung in Szenarien und Dialogen festgehalten wird.

Im Rahmen der Animation des Avatars und der Integration von Medien oder 3D-Objekten muss berücksichtigt werden, dass abhängig von der Fahrzeuggeschwindigkeit oder von Benutzeroptionen weder Avataranimationen noch bewegte Medien und 3D-Objekte angezeigt werden. Dadurch müssen Antworten darauf hin geprüft werden, ob sie den inhaltlichen Kern auch nur durch Sprachausgabe in Verbindung mit einem Bild vermitteln können. Ist dies nicht der Fall, muss eine entsprechende Abfrage der Geschwindigkeit vorgenommen werden und ein alternativer, detaillierterer Text ausgegeben werden. Weitere Antwortalternativen sind sinnvoll, wenn Tests in einem vorangegangenen Zyklus Verständnisschwierigkeiten bei bestimmten Antworten aufgedeckt haben. So können weiterführende Erklärungen, die nicht in der Schriftform des Handbuchs enthalten sind, zum Verständnis beitragen. Die Unterscheidung verschiedener Abstufungen einer Antwort ist auch bei der Erweiterung des Avatarsystems um ein Modell vom Benutzer und seinen Kenntnissen sinnvoll (vgl. Chin 1989; Paris 1989).

Beim Import des Handbuchinhalts werden Sprachausgabeschritte generiert, die bis zu 250 Zeichen an Text beinhalten können. Durch diese Aufteilung können ungleichmäßig große und unnötig viele Sprachausgabeschritte entstehen, die manuell konsolidiert werden müssen, um einen Dialog nicht unnötig lang zu gestalten. Bei der Umformulierung des Textes können Sprachausgabeschritte im Rahmen einer einzigen Ausgabe auch in mehrere Schritte aufgeteilt werden, um Avataranimation oder Mediendarstellungen mit dem Text synchronisiert darzustellen.

Der aus dem Benutzerhandbuch importierte Text liegt in der Schriftform vor, was sich zum einen im Satzbau, aber auch in der Wortwahl, widerspiegelt. Dies macht zum einen die sprachliche Kürzung der Sätze und zum anderen das Auflösen von reinen Schriftformen wie beispielsweise Klammern, Sonderzeichen, Verweisen oder die Verwendung umgangssprachlicher Ausdrücke notwendig. Neben dem Kürzen der Antworten aus Gründen des besseren Verständnisses und der Aufmerksamkeit, liegt derzeit auch in zweierlei Hinsicht eine technische Beschränkung vor. So kann zum einen die Datenbank des Dialogeditors keine Formulierungen mit mehr als 255 Zeichen speichern und zum anderen die Sprachsynthese keine Ausgaben mit mehr als 250 Zeichen verarbeiten.

Ein wichtiger Punkt ist speziell im Fall des Audi-Benutzerhandbuchs zudem die direkte Ansprache des Fahrers. In der Schriftform des Handbuchs wird häufig vom Fahrer in der

dritten Person gesprochen. In einem Dialog muss dies durch eine direkte Ansprache des Benutzers ersetzt werden. Bei der Formulierung einer Sprachausgabe muss die vermittelte Absicht berücksichtigt werden. Walker, Cahn und Whittaker (1997) zeigen, dass auch bei virtuellen Gesprächspartnern der Ton die Musik macht und durch verschiedene Formulierungen bestimmte Reaktionen beim Benutzer hervorgerufen werden können. Dennoch lässt die Vermittlung einer bestimmten Botschaft genug Spielraum, um sprachliche Variationen umzusetzen. Durch Variationen in der Sprachausgabe wirken insbesondere häufig ausgegebene Antworten weniger mechanisch und fördern den menschlichen Eindruck des Avatarsystems (Schatter/Schmolke/Zeller 2007). Im Rahmen der Erstellung neuer sprachlicher Antwortalternativen bietet es sich auch an, die zuvor in der ersten Phase erstellten strukturellen Verknüpfungen zwischen Dialogen mit überleitenden Texten zu versehen.

Nach der Überarbeitung und Variation der Formulierung kann dennoch der Gesamteindruck durch falsch ausgesprochene Wörter negativ beeinflusst werden. Somit muss insbesondere bei eingedeutschten Fremd- oder langen Bindewörtern manuell eine phonetische Optimierung vorgenommen werden. Eine prosodische Optimierung ist für die verständliche Vermittlung von Informationen nicht zwingend notwendig und ist deshalb Teil der optionalen sechsten Phase der Emotionsintegration.

Richtlinien der Antwortphase

Bei allen genannten Schritten sind einige Richtlinien zu berücksichtigen, welche die Gestaltung der Antworten vereinfachen und die Qualität verbessern können. Als oberstes Gebot kann die Kürze und Prägnanz der Antworten genannt werden, da sie dem Benutzer zum einen den klaren Auftrag des Avatarsystems vor Augen hält (Vetter 2003, 85f; vgl. auch Bergmann et al. 2005, 41) und zum anderen insbesondere in der Fahrsituation eine übermäßige Ablenkung vermeidet (Williams/Helbig 2007, 286f). Nur durch eine klare Antwort kann der Benutzer aktiv geführt und fehlerhafte Eingaben oder überhöhte Anforderungen an das System vermieden werden (Dybkjaer/Bernsen 2000). Dabei soll eine menschlich wirkende Formulierung gewählt werden, da diese vom Benutzer als angenehmer und natürlicher empfunden wird (Libuda/Kraiss 2004).

Hinsichtlich der Informationsvermittlung müssen die Mengen an präsentiertem Wissen dem Umfeld im Fahrzeug angemessen sein; d.h. sie dürfen den Fahrer in keiner Situation überfordern. Zudem müssen sie jedoch genug an Information vermitteln, dass sowohl der verbale als auch der grafische Teil der Antwort getrennt einen Mehrwert für den Benutzer darstellen. Insbesondere der verbale Teil, der in bestimmten Situationen nur mit einer eingeschränkten Visualisierung wiedergegeben wird, muss alle wichtigen Informationen enthalten. Somit ist der visuelle Teil als Unterstützung der verbalen Ausgabe anzusehen. Aufgrund der Gefahr der Ablenkung sollten nicht bei jeder Antwort zwangsläufig Medien eingesetzt werden. Auch wenn die Möglichkeit zur animierten Darstellung dazu verleitet, diese häufig einzusetzen, sollten statische Bilder bevorzugt werden, sofern dem Benutzer keine signifikanten Informationen durch die fehlende Animation vorenthalten werden.

Bei der Rechtschreibung und Punctuation muss berücksichtigt werden, dass Fehler bei der Sprachausgabe als wesentlich störender empfunden werden als es bei schriftlichen Medien der Fall ist. Somit ist auf eine korrekte Rechtschreibung und Zeichensetzung zu achten. Die Verwendung einer korrekten Punctuation hat zudem Auswirkungen auf den durch die Sprachsynthese automatisch generierten Teil der Prosodie. Somit sollte auch bei kurzen Antworten ein Abschluss durch ein Satzzeichen gesetzt werden, um dadurch ein Abnehmen der Tonhöhe zum Satzende hin zu bewirken. Die Verwendung von Kommata im Fall des konkreten Avatarsystems führte zu teilweise unnatürlich langen Pausen, so dass sie an manchen Stellen wieder entfernt wurden. Einen Überblick der Schritte und darin zu berücksichtigenden Richtlinien gibt Tabelle 6-12.

Bereich	Bezeichnung
Schritte - Struktur	Prüfung der Schrittfolge und –positionierung, besonders bei Medien
	Ergänzung neuer Medienschritte
	Umwandlung von Bildern in Videos oder 3D-Objekte
	Ggf. Bereitstellung einer Antwortalternative für reine Sprachausgabe
	Ggf. Bereitstellung weiterer Detailinformationen bei komplexen Inhalten
Schritte - Formulierung	Konsolidierung und ggf. Aufteilung von Sprachausgabeschritten
	Konvertierung der Texte von Schrift- in Sprachform
	Umsetzung einer direkten Ansprache des Benutzers
	Umsetzung von sprachlichen Variationen unter Berücksichtigung des Tonfalls
	Erstellung von verbalen Überleitungen
	Phonetische Optimierung
Richtlinie	Kürze und Prägnanz der Antworten
	Menschliche Formulierung und Satzbau
	Verwendung der Visualisierung zur Unterstützung der verbalen Information
	Bevorzugt Bilder statt Videos oder 3D-Objekte einsetzen
	Korrekte Rechtschreibung und Punctuation für angenehme Sprachsynthese

Tabelle 6-12: *Schritte und Richtlinien für die Phase der Antwortgestaltung*
Quelle: (Eigene Darstellung)

6.2.3.3. Phase 3: Gestaltung von Eingaben

Die dritte Phase umfasst die Erstellung und Optimierung der möglichen Eingaben, die einerseits durch den Benutzer oder durch Systemereignisse über den CAN-Bus ausgelöst werden können. Ein grundlegender Schritt ist die Überprüfung aller automatisch generierten Eingabeschritte daraufhin, ob sie ein Einstiegspunkt sein sollen, der ohne Kontext im Dialog aufgerufen werden kann. Einstiegspunkte aus dem Import sind insbesondere die Anfänge von Erklärungssequenzen, die zwar mehrere Eingabeschritte enthalten, aber meist jedoch nur Formulierungen wie „Weiter“ oder „Ja“ als Eingaben vorsehen. Der Festlegung dieser strukturellen Frage nach Einstiegspunkten folgt die Erstellung einer grundlegenden Formulierung für einen bestimmten Eingabeschritt wie z.B. „Wie funktioniert meine

Klimaanlage“. Dabei ist wichtig, dass die Absicht der Benutzereingabe aus dieser Grundformulierung klar hervorgeht, da auf dieser Absicht die Erstellung alternativer Eingabemöglichkeiten basiert.

Die Verwendung einer einzelnen Eingabealternative kommt der Verwendung eines einzelnen Kommandos gleich und entspricht nicht dem natürlichen menschlichen Verhalten, jeden Satz individuell zu formulieren. So wird die Vielfalt und Freiheit in der Formulierung von Spracheingaben gefordert (Rosset/Bennacef/Lamel 1999, 1535), wobei Zue und Glass (2000) auch auf die daraus resultierenden Probleme eingehen: Mit zunehmender Anzahl möglicher Eingabealternativen kann die Spracherkennung schwieriger ausfallen. Mit dem vorliegenden Avatarsystem und der darin verwendeten Spracherkennung konnten jedoch auch in Versuchen mit über 3000 verschiedenen Phrasen keine signifikanten Einflüsse auf die Spracherkennung erkannt werden. Auch wenn die Bandbreite möglicher Formulierungen der gleichen Absicht unfassbar groß erscheint, zeigt die Beobachtung von Zipf's Gesetz bei Chatbots auf, dass Benutzer bestimmte Formulierungen bevorzugen (Wallace 2002b). Diese bevorzugten Formulierungen liegen im ersten Zyklus jedoch noch nicht vor, so dass strukturiert nach Eingabealternativen gesucht werden muss. Zur Sammlung kann folgendes Vorgehen herangezogen werden, wobei sich die jeweils ergebenden Alternativen miteinander kombiniert werden müssen:

- Variation der Personen: Nach einer Information kann aus Sicht verschiedener Personen gefragt werden. So kann neben einer unpersönlichen Formulierung aus Sicht des Fahrers mit „ich“ oder „man“ gefragt werden. Darüber hinaus kann auch danach gefragt werden, wie das Avatarsystem („du“ oder „Sie“) bei einer bestimmten Aufgabe behilflich sein kann.
- Variation des Satzbaus: Neben einem grammatisch notwendigen oder sinnvollen Satzbau lassen sich in der gesprochenen Sprache viele weitere Variationen erkennen, die umgangssprachlich verwendet werden und nicht zwangsläufig eine korrekte Grammatik aufweisen.
- Variation der Verben: Für die gleiche Tätigkeit können verschiedene Verben verwendet werden, die u.U. Auswirkungen auf den Satzbau haben können. Zur Sammlung verschiedener Variationen lassen sich Synonym-Lexika heranziehen, die als Grundlage für die Pflege von Synonymgruppen für die infiniten Formen der Verben dienen können.
- Variation der Substantive: Insbesondere im technischen Bereich lassen sich bestimmte Teile sowohl mit Fachausdrücken als auch mit laienhaften Bezeichnungen benennen; so entspricht eine Wählhebelstellung umgangssprachlich einem eingelegten Gang. Sowohl auf fachlicher als auch auf laienhafter Seite lassen sich jedoch wiederum verschiedene gleichwertige Namen finden (z.B. Getriebe, Schaltung, Gangschaltung...). Ebenso wie für Verben lassen sich Synonym-Lexika verwenden, die zur Gestaltung von Synonymgruppen im Dialogeditor verwendet werden können. Da viele Substantive Teile des Fahrzeugs bezeichnen und Automobilhersteller eigene Begriffe für umgangssprachlich anders bezeichnete Teile verwenden, können viele Synonyme

erst in der Phase des Tests ermittelt werden. So haben sich beispielsweise für den Spurhalteassistenten erst im Test Begriffe wie „Spurlinienassistent“ oder „Spurhaltesystem“ ergeben, die zuvor nicht absehbar waren.

- Variation bzw. Auslassung weiterer Wortarten: Neben Verben und Substantiven lässt sich noch eine Vielzahl von Wortarten unterscheiden, die jeweils variiert werden können. Da die Berücksichtigung aller Optionen zu einem unüberschaubaren Aufwand führt, lassen sich diese Wortarten in den meisten Fällen ignorieren und durch Platzhalter ersetzen. Die Absicht einer Eingabe kann meist bereits durch die Verarbeitung von Substantiv und Verb korrekt erkannt werden. In bestimmten Fällen ist die Berücksichtigung von Personalpronomen sinnvoll, da evtl. nicht klar ist, wer eine bestimmte Handlung vornehmen soll. Bestimmte und unbestimmte Artikel stellen in der deutschen Sprache eine große Variationsmöglichkeit dar, die jedoch kaum zusätzliche Informationen bietet. So können Artikel meist vollständig durch Platzhalter ersetzt werden, um die Anzahl zu pflegender Formulierungen zu reduzieren.

Neben der expliziten Berücksichtigung aller möglichen Formulierungen kann insbesondere für den Einstieg in eine Erklärung eine vereinfachte Auffangregel wie z.B. „* Gangschaltung *“ erstellt werden, die auf alle Äußerungen mit dem Wort „Gangschaltung“ reagiert, sofern kein passenderer Treffer in der Regelbasis ermittelt wurde. Diese minimale Eingabe birgt dabei jedoch die Gefahr auch bei Eingabe aufgerufen zu werden, die nach einer bestimmten Information im Rahmen eines Bedienelements fragen und führt damit zu Frustration beim Benutzer. Der aus dieser Eingabe generierte Eintrag in der Sprachgrammatik kann zudem aufgrund seiner Kürze dazu führen, dass er erkannt wird, obwohl der Benutzer eine andere Eingabe vorgenommen hat.

Neben der Vielfalt einer Benutzereingabe kann ein Systemereignis jeweils nur eine Ausprägung haben. Insbesondere bei tutoriellen Dialogen, in denen zusammen mit dem Benutzer ein bestimmter Vorgang durchgearbeitet wird, ist die Berücksichtigung von Systemereignissen wie die Betätigung eines Bedienelements wichtig. Dabei ist jeweils darauf zu achten, dass diese Eingabeschritte zum einen neu erstellt werden müssen und zum anderen „stumm“ sind, d.h. nicht durch Spracheingabe ausgelöst werden können.

Das Ergebnis des Imports des Benutzerhandbuchs sind in Stücke zerlegte Erklärungssequenzen die in möglichst komprimierter Form einen Überblick über eine bestimmte Funktion geben, jedoch keine gezielten Fragen beantworten. So wird beispielsweise die Klimaanlage erklärt, jedoch nicht beschrieben, wofür die Beschriftung „AC“ auf einem der Knöpfe steht. Die Berücksichtigung solcher Detailfragen stellt den letzten Schritt dieser Phase dar. Sie macht neben der Formulierung neuer Eingabealternativen evtl. auch die Aufteilung oder den Auszug bestimmter Informationen einer Antwort in eine gesonderte Antwort notwendig. Dadurch kann insbesondere durch diesen Schritt die Notwendigkeit entstehen, im gleichen Zyklus zur Phase der Antwortgestaltung zurückzugehen und neue Detailantworten zu erstellen.

Bereich	Bezeichnung
Schritte	Überprüfung von Einstiegspunkten
	Erstellung einer Grundformulierung
	Strukturierte Suche nach Eingabealternativen
	Formulierung allgemeiner Auffangregeln mit Stichwörtern
	Integration von Systemereignissen
	Formulierung von Detailfragen
Richtlinie	Umgangssprachliche Formulierungen berücksichtigen
	Transformation von Fremdwörtern in die deutsche Sprache
	Dialogunabhängige Pflege häufig verwendeter Formulierungen

Tabelle 6-13: *Schritte und Richtlinien für die Phase der Eingabengestaltung*
Quelle: (Eigene Darstellung)

Da bereits viele Aspekte explizit in den Schritten dieser Phase eingebunden sind, bleiben nur wenige schrittunabhängige Richtlinien zu beschreiben. So lässt sich auch wie bei der Gestaltung der Antworten die Berücksichtigung der Umgangssprache nennen. Da in der modernen Sprache viele Fremdwörter eingedeutscht wurden und sich die Aussprache von der Schreibweise unterscheiden, muss auf die deutsche Transkription geachtet werden. So wird der Begriff „lane assist“ von der Spracherkennung nur in der Form „lein assist“ erkannt. Zudem werden einige Eingaben -durch die Darstellung in Antwortsequenzen bedingt- häufig verwendet. Äußerungen für Bedeutungen wie „Weiter“, „Wiederholen“, „Ja“, „Nein“ usw. sollten unabhängig von einem konkreten Dialog angelegt und gepflegt werden. Die gesamten Formulierungen für die Eingabe müssen nicht mit Satzzeichen versehen sein da sie ohnehin beim Export in AIML-Regeln und in die Sprachgrammatik herausgefiltert werden. Einen Überblick der genannten Schritte und Richtlinien dieser Phase gibt Tabelle 6-13.

6.2.3.4. Phase 4: Gestaltung von Medien

Nachdem textuelle Inhalte mit Eingaben und Antworten erstellt wurden, müssen die vorgesehenen Platzhalter für Medien mit passenden Bildern und Videos gefüllt werden. Da erst in dieser Phase auch die genaue Position von erklärten Bedienelementen in Medien und das Timing von Videos bestimmt werden, können Avataranimationen, die den Bezug zu einem Medium herstellen, auch erst in dieser Phase sinnvoll eingebunden bzw. angepasst werden.

Gestaltung und Verwaltung der Medien

Je nachdem wie sehr die Trennung der einzelnen Phasen mit einer personellen Trennung einhergeht, stellt die detaillierte Dokumentation der erforderlichen Medien einen wichtigen Brückenschlag zwischen der Phase der Antwortgestaltung und der Phase der Mediengestaltung dar. Wie bereits beschrieben, kann eine Medienliste mit den Namen der Medien im Dialogeditor bzw. dem Dateinamen und dem im Medium vorgesehenen Inhalt und zeitlichen Ablauf eine Arbeitshilfe darstellen. So kann darin beispielsweise spezifiziert werden, dass für das Szenario über den Bordcomputer das Video „bc_bedienen.avi“ mit den

Abmessungen 300x160 Pixel benötigt wird, dass den Scheibenwischerhebel zeigt, an dem nacheinander die Reset- sowie die Auf- und Ab-Taste betätigt werden.

Je nachdem welche Art von Medium die Medienliste vorsieht, kommen verschiedene Bezugsquellen in Frage. 3D-Objekte können nur auf Basis von Konstruktionselemente bzw. weiterverarbeiteten und texturierten Objekten konvertiert werden. Sind zu einem vorgesehenen 3D-Medium keine passenden Objekte vorhanden, so kann es neu konstruiert oder als 2D-Video umgesetzt werden. Zur Erstellung eines 2D-Videos können einerseits 3D-Objekte herangezogen werden, wenn eine Integration in das 3D-Umfeld keinen Mehrwert darstellt. Alternativ können 2D-Videos als Folge verschiedener Bilder erstellt werden. So bietet das virtuelle Cockpit, das auf der Handbuch-CD des Audi A4 erstmals enthalten ist, eine gute Quelle, um einzelne Screenshots zu erstellen und sie zu einem Video zusammenzustellen (vgl. Abbildung 6-25). Diese Quelle bietet sich neben den Abbildungen im digitalen Handbuch auch für die Gewinnung von 2D-Bildern an. Existieren zur Veranschaulichung einer Erklärung keine bestehenden digitalen Bilder, kann durch die Überarbeitung und Verfremdung eines Fotos ein entsprechendes Medium erzeugt werden.



Abbildung 6-25: *Virtuelles Cockpit der Audi Handbuch CD*
Quelle: (Audi 2007)

Die Gestaltung von Medien sollte sich unabhängig von der konkreten Ausprägung nach bestimmten Anforderungen für visuelle Darstellungen in Handbüchern richten. So darf eine Darstellung nur einen relevanten, überschaubaren und dennoch wiedererkennbaren Ausschnitt

zeigen (Zieten 1990, 105). Dies kann beim Einsatz in einem Fahrzeug derart erfolgen, dass z.B. nur der entsprechende Hebel dargestellt wird, auf dessen Bedienung eingegangen wird. Dadurch wird die Aufmerksamkeit des Fahrers gleich auf das relevante Bedienteil gerichtet. Die Darstellung des relevanten Ausschnitts soll dabei stark überzeichnet sein, um sich leichter orientieren zu können. Detailreiche Darstellungen und Fotos sind dazu ungeeignet (Zieten 1990, 126ff). Das erklärte Element eines Bedienteils wie beispielsweise ein bestimmter Knopf am Hebel soll dabei farblich hervorgehoben sein. Die farbliche Hervorhebung richtet zusammen mit der animierten Darstellung der beschriebenen Bedienungsart die Aufmerksamkeit voll auf das entsprechende Element und erleichtert es, die Funktion schnell zu begreifen. Videos, die einen Bewegungsablauf darstellen sollen dabei langsam ablaufen und müssen nicht die reale Benutzungsgeschwindigkeit repräsentieren. Im Falle von Videos muss zusätzlich zur Animation auch ein aussagekräftiges Standbild erstellt werden, das zum Einsatz kommt, sobald die entsprechende Geschwindigkeitsschwelle überschritten wird.

Integration von Avataranimationen

Neben der Darstellung der Medien können auch Avataranimation zur besseren Informationsvermittlung beitragen. Neben der Erklärung eines Bedienungsablaufs dienen die abgestimmten Bewegungen auch der nichtverbalen Kommunikation und somit zur Verdeutlichung der Kommunikationsabsichten (vgl. auch Churchill et al. 2000, 83f). Im Falle der vorliegenden Avatarplattform entfällt an dieser Stelle die Erstellung neuer Avataranimationen, da durch die Avatarkomponente von Charamel bereits eine umfangreiche Bibliothek an Animationen bereitgestellt wird. Darüber hinaus muss auch keine gesonderte Integration der vorgesehenen Avataranimationen in den aktuellen Bewegungsablauf des Avatars vorgenommen werden wie ihn z.B. Kurlander und Ling (1995) beschreiben. Die Transformation der aktuellen Avatarbewegung in die gewählte Animation wird durch die Avatarkomponente selbstständig vorgenommen.

Die Integration von Avataranimationen ist zwar auch schon in einer früheren Phase möglich, jedoch lässt sie sich erst in der vierten Phase mit den erstellten Medien abstimmen, um eine Präsentation aus einem Guss zu gewährleisten. Die Avataranimationen zur Verdeutlichung des Inhaltes nehmen Bezug auf den visuell dargestellten Inhalt und müssen sich sowohl hinsichtlich Positionierung und Timing an den bestehenden Medien orientieren. Da sowohl Medien als auch Avataranimationen in der gleichen Phase erstellt werden, kann eine abgestimmte Erstellung bzw. Integration der beiden Elemente erfolgen. Die Abstimmung von Ort und Zeit gewährleistet, dass der Avatar beispielsweise auf einen Bereich des Mediums verweist wenn dieser sowohl verbal beschrieben als auch durch eine Animation hervorgehoben wird.

Eine Einschränkung hinsichtlich der Abstimmung von Video und Avataranimation ergibt sich bei diesem Avatarsystem aufgrund der Verwendung in einem HTML-basierten GUI-Framework. Da der Avatar als Plug-in in die HTML-Seiten eingebunden ist, wird der Avatar durch Videos überlagert, so dass beispielsweise seine Hand nicht sichtbar ist, wenn sie auf einen Bereich des Videos zeigt. Dieses Verhalten lässt sich nur verhindern wenn anstelle von Videos Flash-Animationen verwendet werden, die jedoch einen erhöhten Erstellungsaufwand bedeuten. Eine weitere Herausforderung bei der Abstimmung zwischen Medien und Avatar

ergibt sich durch den dynamischen Austausch der Videos durch Standbilder ab einer entsprechenden Fahrgeschwindigkeit. Dabei muss somit sichergestellt werden, dass auch im Standbild auf das erklärte Element verwiesen werden kann.

Die Bedeutung der nonverbalen Kommunikation wird insbesondere bei Bente und Krämer (2000, 39f) deutlich, die verschiedene Ansätze zur Modellierung entsprechenden Verhaltens für anthropomorphe Schnittstellen gegenüberstellen. Sie unterscheiden den Top-Down- und den Bottom-Up-Ansatz zur Bereitstellung von Avataranimationen für eine Gesprächssituation. Beim Top-Down-Ansatz oder auch theory driven approach werden die Bewegungen des Avatars auf Basis von Algorithmen und verschiedener Umfeldinformationen wie z.B. der Dialoghistorie errechnet. Der Bottom-Up-Ansatz bzw. data driven approach verwendet gespeicherte Bewegungsmuster von realen Personen, um die Bewegungen des Avatars zu generieren und gewährleistet dadurch einen menschlicheren Bewegungsablauf. Die verwendete Avatarkomponente verwendet den data driven approach, indem sie gespeicherte Bewegungsabläufe einbezieht, die zuvor über Motion Capturing erfasst wurden.

André, Rist und Müller (1998b) identifizieren neben den manuell angewiesenen Animationen eines Avatars auch automatisch ablaufende Animationen. Dabei unterscheiden sie teilautonomes Verhalten hinsichtlich grundlegender Aktivitäten wie dem Bewegen über den Bildschirm, Leerlauf-Animationen, wenn das Avatarsystem keine Informationen wiedergibt oder bearbeitet, der Anzeige einer internen Verarbeitung als Ersatz einer Sanduhr und schließlich die direkte Reaktion auf eine Aktion des Benutzers. Während Leerlauf-Animationen bereits erfolgen und ein Ersatz für die Sanduhr aufgrund der schnellen Verarbeitungszeiten nicht notwendig ist, stellen die beiden verbleibenden Optionen Ansatzpunkte für die Verbesserung technischer Komponenten dar.

Eine Übersicht aller beschriebenen Schritte und Richtlinien innerhalb der Phase der Mediengestaltung ist in Tabelle 6-14 abgebildet.

Bereich	Bezeichnung
Schritte	Erstellung von Bildern bzw. Videos und Standbildern
	Erstellung von 3D-Objekten
	Integration von Avataranimationen abhängig vom Medium
Richtlinie	Darstellung von relevanten Bedienelementen
	Markierung von beschriebenen Bedienelementen
	Langsame Animationen
	Abstrakte Darstellung auf geringer Detailstufe

Tabelle 6-14: *Schritte und Richtlinien für die Phase der Mediengestaltung*
Quelle: (Eigene Darstellung)

6.2.3.5. Phase 5: Umsetzung von Umfeldfunktionen

Während die ersten vier Phasen zur Gestaltung des grundlegenden Inhalts innerhalb eines Dialogs beitragen, decken die Umfeldfunktionen alle Aktivitäten ab, welche die Bedienung

zusätzlicher Funktionen ermöglichen oder vereinfachen. Sie dienen somit nur indirekt dem Abrufen und Wiedergeben von Inhalten. Die Umfeldfunktionen lassen sich in zwei Bereiche unterteilen. Zum einen sind einige Umfeldfunktionen zur Handhabung des gesamten Systems notwendig und bereits im ersten Zyklus fast vollständig absehbar. Zum anderen dienen einige Umfeldfunktionen zur Verarbeitung unvollständiger Nachfragen innerhalb eines Szenarios und lassen sich ähnlich wie die Eingaben in der dritten Phase erst durch mehrere Iterationen vervollständigen. Grundlegende Umfeldfunktionen, die nach dem ersten Zyklus kaum noch verändert werden, dienen der Steuerung der Informationsabfolge und der Darstellung, der Fehlerbehandlung sowie der Anpassung des Systems.

Zur Steuerung des Ablaufs kann grundlegend das Benutzungsparadigma von Web-Browsern herangezogen werden, da es den meisten Benutzern bekannt ist und somit die Eingewöhnung an das Avatarsystem vereinfacht werden kann. Der erste Schritt dazu wird bereits bei der Gestaltung der Antworten gelegt, indem zwischen verschiedenen Teilen einer Erklärungssequenz mit dem Befehl „Weiter“ vorwärts navigiert werden kann. Um einem Web-Browser zu entsprechen, stehen damit noch die Funktionen des Zurückwechslens zum letzten Inhalt und das Aktualisieren des Inhalts aus. Das Aktualisieren des Inhalts gestaltet sich im Zusammenhang des Avatarsystems als Wiederholung des zuletzt vorgetragenen Inhalts. Beim Export der AIML-Regeln wird in jede Regel ein Mechanismus integriert, der die Historie der letzten Eingaben aktualisiert. Durch Zugriff auf diese Historie lässt es sich einfach bewerkstelligen, sowohl die aktuelle Antwort zu wiederholen als auch zu einer vorangegangenen Antwort zu springen. Der Rücksprung erfolgt dabei nicht durch den vom Dialogeditor angebotenen Sprungschritt, sondern durch eine direkte AIML-Anweisung (<SRAI>) unter Verwendung der ID der entsprechenden Eingabe, die aus der Historie bezogen wird.

Neben der Steuerung der ausgegebenen Antworten kann auch innerhalb von Antworten eingegriffen werden. Dies ist insbesondere der Fall, wenn Medien in die Antwort eingebunden sind. So kann im Fall eines 3D-Objekts das Objekt selbst gedreht oder die Betrachtungsperspektive geändert werden. Im Falle von Videos kann das Video erneut abgespielt werden. Selbst bei Bildern ist beispielsweise das Wegzoomen denkbar, dass die einfachere Einordnung des angezeigten Ausschnitts in den größeren Kontext des Fahrzeuginnenraums erlaubt.

Die Fehlerbehandlung deckt insbesondere die Reaktionen auf zwei bestimmte Fehlerfälle ab. So kann einerseits der Push-to-Talk-Button betätigt worden sein ohne dass Etwas gesagt wurde. Dadurch wird die leere Eingabe erfasst, auf die ein entsprechendes Feedback gegeben werden muss. Andererseits kommt es häufiger vor, dass der Benutzer eine Eingabe vornimmt, die zwar von der Spracheingabe erkannt und weitergegeben wird, inhaltlich jedoch als Eingabe auf die letzte Antwort nicht verarbeitet werden kann. Als Rückmeldung muss dem Benutzer entsprechend geantwortet werden, dass eine Eingabe erkannt wurde, sie aber nicht sinnvoll verarbeitet werden kann. Beide Fehlerfälle können zudem mit der „Zurück“-Funktion kombiniert werden, um dem Benutzer die letzte Antwort des Avatarsystems erneut vorzutragen.

Den dritten Bereich der grundlegenden Umfeldfunktionen stellen die Einstellungen des Benutzers dar. So können verschiedene Werte durch jeden Benutzer angepasst werden und in die Gestaltung der Reaktionen eingebunden werden. Beispielsweise lässt sich durch den Benutzer bestimmen, ob und wie lange eine Pause zwischen einzelnen Erklärungsschritten erfolgt. Ebenso können Videos generell nur als Bilder angezeigt oder der Avatar vollständig ausgeblendet werden.

Der Teil der Umfeldfunktionen, der in jedem Zyklus angepasst werden muss, gestaltet sich wesentlich unstrukturierter als die grundlegenden Funktionen und hängt weitestgehend von der Wortwahl und den Formulierungen der Benutzer ab. Diese Umfeldfunktionen beziehen sich auf Benutzereingaben unter Auslassung mindestens einer notwendigen Information. So kann es vorkommen, dass nach der Erklärung eines Bedienelements der Klimaanlage durch das Avatarsystem die Rückfrage kommt „Wo ist es?“. Dabei ist mit „es“ das Bedienelement der Klimaanlage gemeint; diese Information ist jedoch in der Eingabe nicht enthalten. Ebenso sind noch viele andere Fragestellungen denkbar, die sich auf das zuvor Gesagte beziehen, dies jedoch nur durch die Dialoghistorie ersichtlich ist (vgl. dazu auch Ball et al. 1997).

Diese Eingaben mit Auslassung lassen sich dabei auf verschiedenen kontextuellen Ebenen betrachten. So können sie auf Ebene des Bedienelements, auf Ebene des Bedienbereichs und auf globaler Ebene erfolgen. Am Beispiel der Klimaanlage sollen die drei Ebenen kurz erläutert werden. Mit einem Bedienelement ist beispielsweise die Einstellung des Gebläses an der Klimaanlage gemeint. Auf die Erklärung dieser Funktion hin kann die Eingabe erfolgen „Wo stelle ich das ein“. Da der Kontext bei dem entsprechenden Bedienelement liegt, kann direkt eine Antwort mit der Position der Gebläsetaste erfolgen. Auf der Ebene des Bedienbereichs, d.h. der Klimaanlage, kann solch eine direkte Antwort nicht erfolgen, obwohl der Kontext sich auf die Klimaanlage bezieht. Es muss gesondert nachgefragt werden, welches Bedienteil mit der Frage nach der Position oder ob die Klimaanlage als Ganzes gemeint ist. Auf globaler Ebene gestaltet sich die Bestimmung des betroffenen Bedienelements noch schwieriger, da der Kontext nicht den relevanten Bedienbereich enthält.

Bereich	Bezeichnung
Einmalige Schritte im ersten Zyklus	Erstellung von Funktionen zur Steuerung
	Erstellung von Funktionen zur Fehlerbehebung
	Erstellung von Funktionen zur Anpassung von Einstellungen
Wiederkehrende Schritte in allen Zyklen	Erstellung von Funktionen zur Beantwortung von Fragen unter Auslassung auf Ebene des Bedienelements
	Erstellung von Funktionen zur Beantwortung von Fragen unter Auslassung auf Ebene des Bedienbereichs
	Erstellung von Funktionen zur Beantwortung von Fragen unter Auslassung auf globaler Ebene

Tabelle 6-15: Schritte für die Phase der Umfeldfunktionen

Quelle: (Eigene Darstellung)

Somit sind entsprechende Umfeldfunktionen zur Beantwortung von Eingaben unter Auslassung auf allen Ebenen zu erstellen, um im Zweifelsfall das relevante Bedienelement zu ermitteln. Da die möglichen Inhalte der Frageformulierungen durch die Benutzer stark variieren und im ersten Zyklus nicht vollständig vorausgesehen werden können, ist eine iterative Pflege und Erweiterung dieser Funktionen notwendig. Eine Übersicht über die Schritte in der Phase zur Erstellung der Umfeldfunktionen gibt Tabelle 6-15.

6.2.3.6. Phase 6: Integration von Emotionen

Die optionale sechste Phase ermöglicht den Ausdruck von Emotionen durch den Avatar. Die Einbindung von Emotionsäußerungen dient nach Churchill et al. (2000, 64f) dazu, die Interaktion mit einem Avatarsystem mehr der Interaktion mit einem realen Menschen anzugleichen (vgl. auch Dryer 1997). Darüber hinaus kann das Zeigen von Emotionen durch den Avatar auch zur gezielten Beeinflussung der Emotionen des Benutzers eingesetzt werden (Picard 2000, 58). Die Voraussetzung für die Integration von Emotionen in ein Avatarsystem ist somit die Absicht, den Avatar sehr menschlich wirken zu lassen oder die Notwendigkeit, zu emotional beladenen Themen Stellung nehmen zu können. Da beide Gründe im Rahmen des konkreten Forschungsvorhabens nicht vorgelegen haben, soll nachfolgend eine konzeptionelle Betrachtung erfolgen.

Moffat (1997) unterscheidet anhand der Kriterien Dauer und Schwerpunkt die beiden Bereiche Persönlichkeit und Emotion, die das Verhalten eines künstlichen Charakters ausmachen. Dabei verändert sich die Persönlichkeit auf lange Dauer kaum, während Emotionen für eine kurze Zeit bestehen und vor allem durch Ereignisse getrieben sind und sich damit auch inkonsistent zur Persönlichkeit äußern können (Allport 1961). Somit muss für das Verhalten eines Avatars sowohl ein grundlegendes Persönlichkeitsmodell bestehen als auch der Einfluss von Ereignissen datentechnisch abgebildet werden.

Zur Abbildung der Persönlichkeit als auch der Emotionen eines Avatars bedarf es eines Modells, das den aktuellen Zustand abbildet. Bei der Modellierung der Persönlichkeit haben sich in der Psychologie eine Vielzahl verschiedener relevanter Kriterien herausgebildet, die sich durch mehrere, teilweise miteinander verbundene Skalen darstellen und somit auf künstliche Charakter übertragen lassen (Bates 1994; Cañamero 1997; Reilly 1996). Trotz zahlreicher Modelle auf diesem Gebiet lässt sich kein allgemein anerkanntes Modell erkennen. Die konkrete Ausprägung und Komplexität des Modells hängt zum einen von den technischen Möglichkeiten der Laufzeit als auch den Gestaltungsmöglichkeiten der Inhalte des Avatarsystems ab. So verwenden André et al. (2000, 225) beispielsweise die fünf Faktoren *offen*, *gewissenhaft*, *extrovertiert*, *angenehm* und *neurotisch* zur Abbildung der Persönlichkeit nach McCrae und John (1992). In der technischen Implementierung des Avatarsystems werden von diesen fünf Faktoren jedoch nur die Extravertiertheit und die Verträglichkeit des Charakters in diskreten Stufen umgesetzt (André et al. 2000, 230). Eine noch einfachere Implementierung eines emotionalen Modells setzen Arafa, Charlton und Mamdani (Arafa/Charlton/Mamdani 1999) um, die den emotionalen Zustand lediglich in die drei diskreten Stufen *negativ*, *neutral* und *positiv* unterteilen. Diese Beispiele zeigen deutlich, dass trotz umfangreicher Persönlichkeits- und Emotionsmodelle die technische Umsetzung

bisher nur ansatzweise erfolgt. Zudem ist der Versuch, Emotionen in diskrete Muster zu pressen bedenklich, da sie nicht trennscharf unterschieden werden können (Picard 2000, 169).

Neben der Modellierung der Persönlichkeit und der Emotionen des Avatars kann in einem Avatarsystem auch die Integration eines Zustandsmodells des Benutzers einen menschlichen Eindruck unterstützen. Bei der Abbildung eines Benutzers in einem Datenmodell lassen sich die gleichen Herausforderungen wie beim Avatar selbst identifizieren. Zudem kann der genaue Zustand des Benutzers nur dann erfasst werden, wenn dem Avatarsystem alle möglichen Eingabeformulierungen vorliegen und eine genaue Einteilung der darin enthaltenen emotionalen Nuancen vorgenommen wurde. Dabei muss zudem die Dialoghistorie berücksichtigt werden, um eine Eingabe hinsichtlich ihrer emotionalen Färbung richtig interpretieren zu können. Prinzipiell kann der aktuelle emotionale Zustand des Benutzers auch über die Analyse des eingehenden Sprachsignals erfolgen, da sich je nach Gemütszustand die Aussprache und Verschleifung von Silben ändern kann (Halberstadt/Niedenthal/Kushner 1995).

Die Abbildung von Persönlichkeit und emotionalen Zuständen ist lediglich der erste Schritt zu einem menschlich wirkenden Avatar. Obwohl intern eine bestimmte Emotion vorherrscht, muss diese nicht in jedem Fall nach außen getragen werden (Picard 2000, 48). Auch bei Menschen ist für Mitmenschen eine Emotion nicht immer von außen wahrnehmbar, so dass auch beim Avatar eine Filterung der Emotionen hinsichtlich ihrer Manifestation erfolgen muss. Arafa, Charlton und Mamdani (1999) differenzieren aus diesem Grund zwischen dem inneren Zustand des Avatarsystems und der Manifestation der entsprechenden Emotionen. Je nach konkreter Situation bzw. äußeren Einflüssen wird eine Emotion gar nicht oder verschieden abgestuft in das Verhalten des Avatars eingebunden. Dabei liegt das Wissen über die Beeinflussung des emotionalen Modells durch den entsprechenden Inhalt in der Wissensbasis selbst und wird vor der Laufzeit des Avatarsystems festgelegt. Eine harte Verdrahtung von Inhalten und Emotionen hingegen ist nicht sinnvoll (Picard 2000, 63). Abbildung 6-26 zeigt die relevanten Elemente bei der Bestimmung und Manifestation von Emotionen auf Seiten des Avatarsystems und des Benutzers.

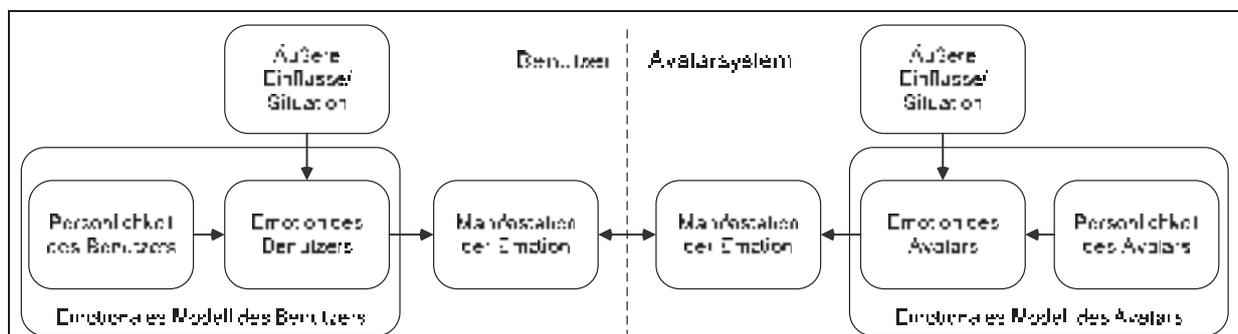


Abbildung 6-26: *Komponenten eines emotionalen Modells für Avatar und Benutzer*
Quelle: (Eigene Darstellung)

Die Manifestation von Emotionen bei einem Avatarsystem erfolgt einerseits auf verbaler als auch auf nonverbaler Ebene (Bente/Krämer 2000). Dies impliziert die Anpassung der gesprochenen Antwort, der Mimik und Gestik. Während die Mimik durch verschiedene

Gesichtsausdrücke direkt durch die Manifestation des emotionalen Zustands gesteuert werden kann, bedarf die verbale Äußerung einer Zwischenstufe. So kann entweder für jeden denkbaren Zustand eine eigene Antwort generiert werden oder es können verschiedene Formulierungen je nach Zustand ausgewechselt werden. Da das dynamische Austauschen von Phrasenteilen eine linguistische Herausforderung darstellt, werden bisher in Avatarsystemen verschiedene Antworten erstellt. Dieses aufwändige Vorgehen erklärt auch die Verwendung weniger diskreter Stufen im emotionalen Modell. Neben dem Inhalt der Sprachausgabe können auch phonetische und prosodische Parameter bei der Sprachsynthese Emotionen transportieren (Murray/Arnott 1993; Alm et al. 1993; Cahn 1990).

Tabelle 6-16 zeigt die notwendigen Schritte, die für die Integration eines emotionalen Modells für den Avatar als auch für den Benutzer notwendig sind.

Bereich	Bezeichnung
Schritte zur Integration eines emotionalen Modells für den Avatar	Identifizierung relevanter Faktoren und Erstellung eines emotionalen Modells für den Avatar
	Einbindung von Steuerschritten des Modells in Wissensbasis
	Einbindung von Avataranimation für Antworten abhängig vom emotionalen Modell
Schritte zur Integration eines emotionalen Modells für den Benutzer	Identifizierung relevanter Faktoren und Erstellung eines emotionalen Modells für den Benutzer
	Erweiterung möglicher Eingaben um emotional geprägte Formulierungen
	Einbindung von Steuerschritten des Modells in Wissensbasis abhängig von Eingaben
	Einbindung von Steuerschritten des Avatarmodells in Wissensbasis abhängig von Eingabe des Benutzers

Tabelle 6-16: *Schritte für die Phase der Emotionsintegration*
Quelle: (Eigene Darstellung)

6.2.3.7. Phase 7: Test der Dialoge

Das Testen der erstellten Regeln, Dialoge und Szenarien stellt den Abschluss der Inhaltserstellung für ein Avatarsystem dar. Ausführliche Tests der Inhalte werden nicht nur durch den wissenschaftlichen Ansatz der Design Science vorgesehen (Takeda et al. 1990, 12), sondern sind im Rahmen eines Avatarsystems dringend notwendig. Diese Notwendigkeit entsteht aus der Beobachtung, dass Avatarsysteme noch im Betrieb „reifen“, d.h. sie können zu Beginn ihrer Laufzeit nur einen Teil sinnvoller Eingaben verarbeiten, wodurch sich die Weiterentwicklung der Wissensbasis aufdrängt (von Wendt 2003; Vetter 2003). Während die Praxis, unfertige Systeme in Betrieb zu nehmen, bei internetbasierten Systemen üblich ist, birgt dies im Automobilbereich ein erhöhtes Risiko, wodurch ausführliche Tests und eine Erweiterung der Wissensbasis vor Inbetriebnahme unabdingbar sind. Ein weiterer Faktor ist zudem die spätere Aktualisierung der Inhalte während der Laufzeit. Während sich dies bei internetbasierten Systemen einfach bewerkstelligen lässt, kann dies im Auto nur schwer erfolgen (vgl. dazu auch Kapitel 5.1.2.3).

Technische Tests zur Funktion der Avatarplattform selbst sowie zur Gestaltung des Avatars bzw. der Benutzeroberfläche und deren Handhabung sind nicht Teil dieser Phase. Sie sollten

bereits vor der Erstellung und Erprobung von Inhalten erfolgt sein, so dass die Funktionstüchtigkeit des Systems vorausgesetzt werden kann. Im Rahmen der technischen Funktion wird auch die Abstimmung von Chatbot-Regeln mit der generierten Sprachgrammatik überprüft, so dass alle möglichen Eingaben der Chatbot-Regeln auch durch die Spracheingabe erkannt werden können. Ebenso wird auf Seiten der CAN-Busanbindung gewährleistet, dass alle in den Regeln verwendeten Eingaben auch korrekt durch CAN-Nachrichten ausgelöst werden.

Der Schwerpunkt der inhaltlichen Prüfung in dieser Phase liegt einerseits auf der Abdeckung aller möglichen Eingaben und auf dem erwartungskonformen Ablauf der Dialoge. Unter Verwendung des Dialogeditors und der darin verwendeten Synonymgruppen lassen sich bereits viele verschiedene Formulierungen abdecken. Da die manuelle Erstellung der möglichen Eingaben im Rahmen einer Regel jedoch dem subjektiven Empfinden des Bearbeiters unterliegt, können dadurch nicht alle Ausprägungen im Vorhinein bedacht werden. Einige Formulierungen ergeben sich erst aus der Verwendungssituation heraus oder sind durch den kulturellen und regionalen Hintergrund des Benutzers geprägt. Ein Aspekt der beim Einsatz im Fahrzeug zu beachten ist, ist neben der Spracheingabe auch die Überprüfung der Reaktionen, die auf Basis einer CAN-Nachricht erfolgen sollen.

Eine breite Abdeckung kann durch eine große Menge an Probanden erreicht werden oder indem die Probanden dazu gebracht werden, verschiedene Formulierungen zu probieren. Dies kann beispielweise dadurch erreicht werden, in dem den Probanden eine bestimmte Aufgabe gegeben wird, die nicht durch Spracheingabe gelöst werden kann und dies den Testpersonen nicht bewusst ist. Bei dem Versuch, die Aufgabe dennoch zu lösen, wird in den meisten Fällen eine Vielzahl verschiedener Formulierungen versucht, bevor die Probanden aufgeben. Dieses Vorgehen ist jedoch nicht zu empfehlen, wenn mehrere Aufgaben durch denselben Probanden zu absolvieren sind, da sich die fehlende Reaktion des Systems negativ auf die Motivation auswirkt. In jedem Fall sollte der Proband nach der entsprechenden Aufgabe darauf hingewiesen werden, dass die Aufgabe nicht lösbar war.

Neben möglichen Formulierungen müssen gegebenenfalls auch das Verhalten des Avatarsystems an die Erwartungen der Probanden angepasst werden. Unter dem Verhalten ist insbesondere die Auswahl und Präsentation von Inhalten zu verstehen. Dabei müssen die durch das Avatarsystem vorgetragenen Inhalte den mit der Eingabe verbundenen Erwartungen des Benutzers entsprechen. Zudem muss jedoch berücksichtigt werden, ob der Benutzer bereits genügend einführende Informationen zur Verwendung der Inhalte hat. Beispielsweise kann auf die Frage hin, wie eine Sitzposition im Auto gespeichert werden kann die Bedienfolge der Tasten erklärt werden. Jedoch fehlt dem Benutzer eventuell die Information, dass sich die Tasten dafür in der Fahrertür befinden. Bei der Auswahl der Inhalte ist somit eine Gratwanderung zwischen den Erwartungen des Benutzers und der sinnvollen Reihenfolge der Informationen notwendig.

Neben der Auswahl der zu präsentierenden Informationen ist auch die vorgetragene Antwort zu testen. So muss das gezeigte Medium zur gesprochenen Antwort zum einen inhaltlich und zum anderen von der zeitlichen Abstimmung der beiden Ausgabekanäle her synchron sein.

Diese Abstimmung muss zudem um die Perspektive der Avataramation und gegebenenfalls den zu einer Antwort an den Tag gelegten Emotionen erfolgen.

Der Ablauf der Tests hängt insbesondere vom Entwicklungsstand des Avatarsystems, der modalen Ausstattung und der Domäne ab. Vetter (2003, 91) identifiziert als groben Ablauf die Entwicklung eines Testkonzepts, danach die Erstellung eines Gesprächsprotokolls und schließlich dessen Auswertung. Eckert (1996, 93) schlägt für den Test der Inhalte einen Wizard of Oz –Test vor, bei dem der Proband mit dem Avatarsystem kommuniziert, jedoch von ihm unbemerkt eine reale Person in das Antwortverhalten des Avatarsystems eingreifen kann (vgl. auch Kapitel 5.2.2.1). Dieses Vorgehen bietet sich für Systeme an, die technisch noch nicht voll entwickelt sind oder nur wenige Ausgabemodalitäten beinhalten. Je mehr Ausgabemodalitäten d.h. Sprache, Medien, Avataramation oder Avataremotionen, desto schwieriger wird es für die eingreifende Person diese Kanäle zeitnah und synchronisiert zu manipulieren. Wenn der Entwicklungsstand des Systems es erlaubt oder die Anzahl an Modalitäten einen Wizard of Oz-Test verhindern, so kann eine formative Evaluation des Systems erfolgen.

Unabhängig von der gewählten Methode können dabei Beobachtungsmethoden wie Videoaufzeichnungen oder Logdatenanalysen einbezogen werden. Um den Probanden einen Leitfaden zu geben, können abhängig von den Aufgaben des Avatarsystems konkrete Aufgaben für den Test zusammengestellt werden oder nur ein grober Themenbereich vorgegeben werden. Sowohl bei der Verwendung von Aufgaben als auch beim Abstecken des Themenbereichs kann dabei meist der reale Anwendungsfall nur bedingt abgebildet werden. Beim Einsatz eines Avatarsystems im Fahrzeug z.B. kann erst durch Langzeituntersuchungen das Verhalten des Benutzers und die daraus resultierende Verwendung ermittelt werden. Tabelle 6-17 gibt eine Übersicht der beschriebenen Schritte und ihrer Reihenfolge in der Testphase.

Bereich	Bezeichnung
Schritte	Definition von Testfällen oder Themengebieten
	Test auf Vollständigkeit der Eingabeformulierungen aller Eingabemodalitäten
	Überprüfung der erwartungsgemäßen Zuordnung von Eingabe und Reaktion
	Inhaltliche und zeitliche Abstimmung von Medien und Avataramationen und –emotionen
	Gegebenenfalls Langzeituntersuchung ohne thematische Vorgaben

Tabelle 6-17: *Schritte für die Phase der Dialogtest*

Quelle: (Eigene Darstellung)

6.2.3.8. Besonderheiten bei der Konvertierung unstrukturierter Informationen

Im Gegensatz zur Überführung bereits strukturierter Informationen wie einem Handbuch, das optimalerweise digital vorliegt, müssen unstrukturierte Inhalte zunächst aufbereitet werden. Diese Aufbereitung besteht zunächst in der Sammlung und schriftlichen Aufnahme der Informationen, die später als Teil der Antworten im Avatarsystem dienen sollen. Die gesammelten Informationen müssen dabei bereits in eine inhaltlich sinnvolle Struktur

gebracht werden, die je nach weiterer Verwendung meist nicht so formal sein muss, wie für die Erstellung eines Handbuchs. Die Art der Strukturierung sollte sich dabei bereits an den Schritten und Richtlinien der vorgestellten ersten Phase orientieren. Bereits während der Sammlung muss der Bedarf an Medien festgestellt werden. Dabei muss die Verbindung von schriftlichen Inhalten und Medien sowie nach Möglichkeit bereits die genaue Ausgestaltung der Medien festgehalten werden.

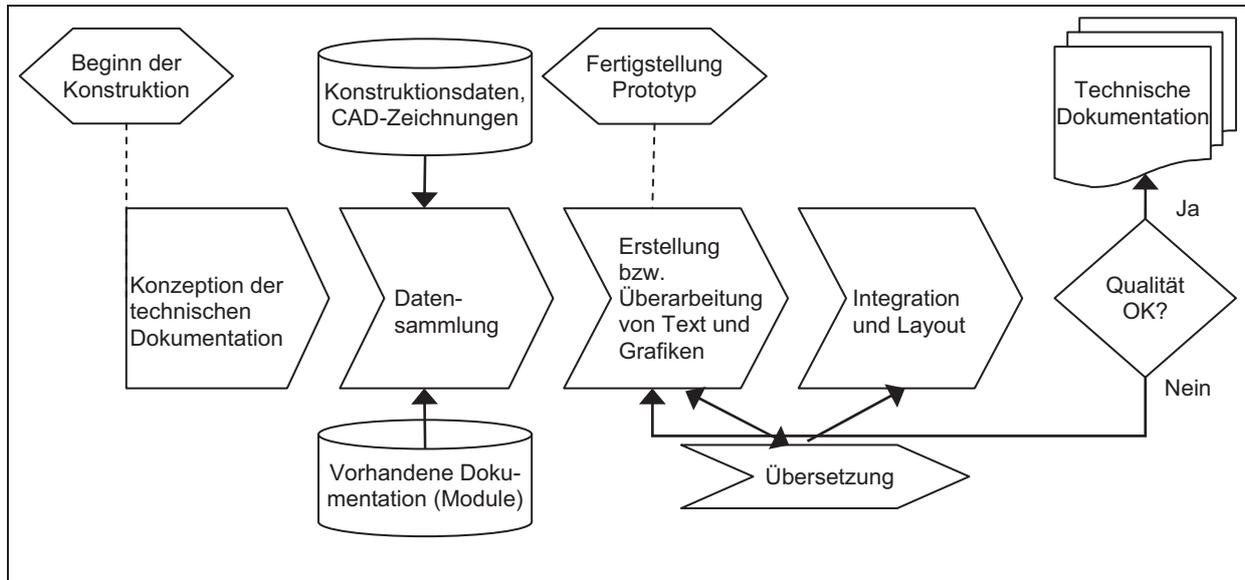


Abbildung 6-27: *Typischer Prozess zur Erstellung der technischen Dokumentation*
Quelle: (Hudetz/Friedewald 2001, 10)

Hudetz und Friedewald (2001) haben in einer Untersuchung bestehender Dokumentationsprozesse im Maschinen- und Anlagenbau einen typischen Prozessablauf für die Erstellung technischer Dokumentationen in diesen Bereichen erarbeitet (vgl. Abbildung 6-27). Sie identifizieren die Erstellung der Dokumentation als beiläufigen Prozess zur Konstruktion des technischen Artefakts und siedeln sie somit zeitlich sehr früh an. Ein weiterer Aspekt, der im vorgestellten Phasenmodell bisher keine Beachtung fand, ist die Übersetzung der Texte. Den Abschluss des Vorgehensmodells stellt die Qualitätsprüfung dar. Das Modell von Hudetz und Friedewald (2001, 10) kann somit als Leitfaden für die Sammlung, Koordination und Zusammenstellung bisher unstrukturierter Informationen angesehen werden und lässt sich anstelle des Imports strukturierter Daten in Phase 0 in das Phasenmodell einbinden.

6.3. Zusammenfassung

Die Implementierung des Avatarsystem lässt sich aus zwei Blickwinkeln betrachten: Der technischen Umsetzung und der Erstellung der Wissensbasis mit ihren Inhalten. Die technische Implementierung umfasst alle Komponenten und Schritte, um eine funktionsfähige Laufzeitumgebung für den Zugriff und die Wiedergabe der Inhalte in der Wissensbasis bereitzustellen. Sie beginnt beim physischen Umbau des Testfahrzeugs der zusätzlich mit einem CarPC und einem Display Switch Module zur Verwendung des verbauten Displays des MMI, einem zusätzlichen Mikrofons sowie der Anbindung an den CAN-Bus des Fahrzeugs

ausgestattet wurde. Aufbauend auf dieser Hardware-Plattform wird aus der Vielzahl für den Automobilbereich existierender Software-Entwicklungsframeworks das OSGi-Framework ausgewählt, das auch in anderen Bereichen Anwendung findet und zu dem bereits Know how im Entwicklungsumfeld besteht.

Unter Verwendung von OSGi werden die in der technischen Konzeption im vorangegangenen Kapitel identifizierten Komponenten entwickelt. Dabei werden in der Konzeption getrennt geplante Elemente technisch teilweise zusammengeführt und größtenteils auf OSGi basierend entwickelt. Die Komponenten für die Spracherkennung, die CAN-Bus-Anbindung und der Input Manager, der Dialogmanager und der Chatbot werden als OSGi-Bundles umgesetzt. Das gesamte GUI-Framework, die Avatar- und 3D-Komponente sowie die Sprachsynthese werden in einen HTML-basierten Rahmen unter Verwendung von JavaScript eingebunden. Bei der Implementierung der gesamten Komponenten wird sofern möglich auf bestehende Software zurückgegriffen, wobei der Best-of-Breed-Ansatz zum Tragen kommt.

Die Erstellung der Wissensbasis für das Avatarsystem gestaltet sich aufgrund der verschiedenen Ausgabemodalitäten als Herausforderung, die mit bestehenden Mitteln nur schwer bewältigt werden kann. Aus diesem Grund ist der erste Schritt der Inhaltsgenerierung die Entwicklung eines grafischen Dialogeditors, der auf einfache Weise alle möglichen Ein- und Ausgabemodalitäten verbindet, sowie die Verarbeitung von Informationen zur Antwortbestimmung erlaubt. Er verfügt über die Möglichkeit, das bestehende Handbuch des Testfahrzeugs im SGML-Format zu importieren und synchronisierte Sprachgrammatiken für die Spracherkennung und Regeln für den Chatbot zu generieren, die Steuerinformationen für alle Ausgabemodalitäten enthalten. Zur Abdeckung einer Vielzahl möglicher Eingabeformulierungen werden dabei Synonymgruppen verwendet, welche die automatische Variation von Wörtern bei der Regelgenerierung bewirken.

Die Entwicklung des Dialogeditors alleine erlaubt jedoch noch keine systematische Überführung der Inhalte des Handbuchs in eine für die Dialogführung geeignete Form. Diese Lücke schließt ein siebenstufiges Vorgehensmodell, das in mehreren Zyklen zu einer qualitätsgesicherten Wissensbasis führt. Dabei werden folgende Phasen unterschieden: Strukturierung, Gestaltung der Antworten, Gestaltung der Eingaben, Integration von Medien, Umfeldfunktionen, Integration von Emotionen und schließlich Test. Während in der Phase der Antwort- und der Eingabengestaltung Rücksprünge zur vorangegangenen Phase möglich sind, werden alle anderen Phasen sequentiell durchlaufen. Die Aufteilung des Zyklus in verschiedene Phasen erlaubt dabei auch die Aufteilung von Aufgaben auf verschiedene Personengruppen wie Redakteure, Grafiker oder Linguisten. Das Vorgehensmodell lässt sich mit verschiedenen Vorarbeiten auch für die Integration von unstrukturierten Inhalten in die Wissensbasis einsetzen.

Die Implementierung des Dialogeditors ist zusammen mit dem Vorgehensmodell eine Methode, die sich bei der Erstellung der Wissensbasis für das technische Avatarsystem im Rahmen dieses Forschungsprojekts bewährt hat und in kurzer Zeit die Umsetzung umfangreichen Inhalts erlaubt. Das Artefakt des Editors sowie das Vorgehensmodell lassen sich somit als Antwort auf die zweite Forschungsfrage nach der Erstellung von Inhalten für Avatarsysteme betrachten.

7. Evaluation des Hilfesystems

„Beginne eine Aufgabe mit dem Herzen, prüfe das Ergebnis mit dem Verstand.“

Otto Häsler (1880 – 1962), deutscher Architekt

Zur Bewertung und Verbesserung des entwickelten Avatarsystems und der damit vermittelten Inhalte ist die Evaluation des Systems mit Probanden im Fahrzeug unabdingbar, wie Christ (2003) ausführt. Aufgrund des Neuigkeitsgrades des Systems, der Vielzahl abgedeckter Aspekte sowie der Verwendung im Fahrzeug stellt die systematische Analyse des Avatarsystems eine Herausforderung dar, der mit dem bisher bekannten Methodenkanon begegnet werden soll. Im Folgenden werden zunächst die Überlegungen zur Konzeption der Evaluation sowie deren detaillierte Vorbereitung betrachtet. Die wichtigsten aus Befragung und Beobachtung gewonnenen Ergebnisse werden anschließend vorgestellt und eine Interpretation der beobachteten Sachverhalte versucht. Durch die Interpretation und die Abweichungen zwischen den verschiedenen Probandengruppen sollen darauf basierend Hypothesen formuliert werden, die es in zukünftigen Untersuchungen zu analysieren gilt.

Die Erkenntnisse aus der umfassenden Evaluation decken dabei das konkrete Avatarsystem als auch Avatarsysteme an sich aus vielerlei Aspekten ab. So werden neben technischen Problemen, die sich erst bei der groß angelegten Untersuchungen ergaben, auch Verbesserungen an der Gestaltung der Oberfläche, der Verknüpfung und Präsentation von Inhalten sowie der Bedienbarkeit des Systems im Fahrzeug betrachtet.

7.1. Planung und Durchführung der Evaluation

Die Planung der Evaluation gestaltet sich insofern als schwierig, als zum einen summativ der Unterschied zwischen der Verwendung des Avatarsystems und bisherigen Informationsquellen im Fahrzeug untersucht werden muss. Zum anderen hat eine summative Untersuchung einen hypothesenprüfenden Charakter (Bortz/Döring 2006, 111). Aufgrund fehlender Anhaltspunkte zur fundierten Hypothesenbildung für den Einsatz des vorgestellten Avatarsystems im Fahrzeug kann nur ein explorativer, hypothesengenerierender Ansatz gewählt werden, der zum einen durch den Vergleich mit der Benutzung des Handbuchs und zum anderen auf Basis von Beobachtungen der Verwendung in realen Situationen erste Tendenzen zur abschließenden Hypothesenformulierung liefert (vgl. dazu Bortz/Döring 2006, 480). Obwohl Untersuchungen mit Avataren (z. B. Sanders/Scholtz 2000; André/Rist/Müller 1998b; Qiu/Benbasat 2007) bereits verschiedene Ergebnisse aufgezeigt haben, variieren die Erkenntnisse jedoch mit jedem konkreten System und vor allem mit der Domäne des Einsatzes. Darüber hinaus ist das einzige dokumentierte und teilweise vergleichbare System REA (vgl. Cassell 2001) nur unter Laborbedingungen untersucht worden.

7.1.1. Konzeption der Evaluation

Da als Bezugsgröße der Evaluation in Ermangelung von etablierten Alternativen das derzeit vorherrschende Informationsmedium im Automotive Bereich das Benutzerhandbuch zu sehen

ist, muss die Verwendung des Handbuchs bzw. des Avatarsystems als Ersatz als ein mögliches Treatment der Evaluation angesehen werden. Zudem lässt sich unterscheiden, ob die Anwendung bei stehendem oder fahrendem Fahrzeug erfolgt. Auf Basis dieser beiden Merkmale lässt sich somit prinzipiell eine 2x2 Matrix zur Einteilung von Versuchsgruppen erstellen wie in ähnlich gelagerten Fällen zur Evaluation von Avatarsystemen (Qiu/Benbasat 2007) bzw. Untersuchungen im Fahrzeug erfolgt ist (Hu et al. 2007). Bei genauerer Betrachtung ergibt sich jedoch die Schwierigkeit bei der Kombination der Handbuchverwendung während der Fahrt. Da das Lesen eines gedruckten Textes während der Fahrt ein unannehmbar hohes Risiko birgt, kann diese Ausprägung nicht in einem Versuchsfahrzeug im realen Straßenverkehr untersucht werden. Um die verbleibenden drei Ausprägungen dennoch miteinander vergleichen zu können, ergeben sich daraus drei gleichwertige Gruppen mit folgenden Ausprägungen:

1. Verwendung des Handbuchs bei stehendem Fahrzeug mit erfolgter Zündung als Kontrollgruppe
2. Verwendung des Avatarsystems bei stehendem Fahrzeug mit erfolgter Zündung als Gruppe mit Treatment 1
3. Verwendung des Avatarsystems bei fahrendem Fahrzeug als Gruppe mit Treatment 2²³.

Dadurch werden die möglichen Ausprägungen berücksichtigt und die Vergleichbarkeit von jeweils mindestens zwei Gruppen gewährleistet. Jede der drei Gruppen hat dabei identische Aufgaben rund um die Bedienung des Fahrzeugs auszuführen. Es wurden nur Aufgaben zur Evaluation ausgewählt, die sowohl bei stehendem als auch bei fahrendem Fahrzeug durchgeführt werden können, um die Beobachtungen über alle Gruppen hinweg vergleichbar zu halten. Als mögliche Aufgaben kommen alle beobachtbaren Tätigkeiten rund um die Bedienelemente des Cockpits in Frage. Dabei wurde eine Vorauswahl anhand der Verwendung des MMI-Displays vorgenommen. Da zur Anzeige des prototypischen Avatarsystems die herkömmliche Anzeige im MMI-Display unterbrochen wird, können keine Aufgaben verwendet werden, die zu ihrer Bearbeitung das MMI-Display verwenden müssen. Diese Einschränkung ist insbesondere bedeutsam für die tutorielle Informationsvermittlung, bei der sowohl das Avatarsystem angezeigt werden muss als auch der Benutzer die entsprechende Aufgabe parallel durchführen können muss. Eine Übersicht aller relevanten Aufgaben und ihrer Eigenschaften ist in Anhang A zu finden.

Aus der dennoch umfangreichen Menge an Aufgaben wurden fünf verschiedene Aufgaben gewählt, die aufgrund der enthaltenen Bedienschritte und damit ihres Umfangs auf drei Kategorien verteilt sind. Der Tabellenausschnitt der gewählten Aufgaben und die Formulierung als Frage während der Evaluation sind in Tabelle 7-1 dargestellt (vgl. auch Kapitel 6.2.2). Ursprünglich bestand der fünfte Aufgabenkomplex zur Klimaautomatik aus

²³ Die aktivierte Zündung des Fahrzeugs ist in allen Fällen erforderlich da sich sonst die abgefragten Funktionen nicht betätigen lassen.

einer weiteren Aufgabe, die das Einschalten der Klimaautomatik zu Beginn vorsah. Aufgrund der hohen Temperaturen während des Evaluationszeitraums im September war die Klimaautomatik bereits zu Beginn jeder Beobachtung aktiviert, so dass diese Aufgabe entfiel.

Kategorie	Aufgabe	Teilaufgabe
1 Bedienschnitt	Spurhalteassistent	Aktivieren Sie den Spurhalteassistent bzw. Audi lane assist
	Spurwechselassistent	Aktivieren Sie den Spurwechselassistent bzw. Audi side assist
2 - 5 Bedienschnitte	Sitzmemory	Aktivieren Sie das Sitzmemory
		Speichern Sie die aktuellen Einstellungen unter Speicherplatz 2
		Deaktivieren Sie das Sitzmemory wieder
	Bordcomputer	Stellen Sie den Bordcomputer auf Speicher 1 bzw. den Einzelfahrtspeicher ein
		Navigieren Sie zum Wert der Reichweite im Bordcomputer
		Wechseln Sie zur Anzeige des Audi lane assist mit dem Bordcomputer
6 - 9 Bedienschnitte	Klimaautomatik	Reduzieren Sie die eingestellte Temperatur für den Fahrer um ein Grad
		Stellen Sie die Drehzahl des Gebläses für den Fahrer auf Stufe fünf ein
		Aktivieren Sie den Umluftbetrieb
		Schalten Sie Klimaanlage aus

Tabelle 7-1: *Einteilung und Wortlaut der Aufgaben während der Evaluation*
Quelle: (Eigene Darstellung)

7.1.1.1. Vorbereitung des Versuchsfahrzeugs

Für die technische Umsetzung als auch für die Evaluation wurde ein Audi A4 (interne Bezeichnung B8) der Modellreihe 2008 verwendet. Die starke Motorisierung (3,2l FSI mit Allradantrieb quattro) und die Vielzahl an zusätzlichen Ausstattungsmerkmalen wie Geschwindigkeitsregelanlage, Audi lane assist, Audi side assist, Klimaautomatik plus, MMI plus etc. machen das Fahrzeug von der Ausstattung her vergleichbar mit größeren Modellen des Herstellers.

Neben den Erweiterungen, die bereits für den Betrieb des Avatarsystems am Fahrzeug vorgenommen wurden, bedarf die Videoaufzeichnung der Evaluation weiterer Ergänzungen. So wurde für alle Versuchsgruppen eine Kamera aufgebaut, die sich im hinteren Bereich des Fahrzeugs zwischen den Sitzen befand. Abbildung 7-1 zeigt schematisch die Positionierung und den Blickwinkel der Kamera.

Die hohe Auflösung und Qualität der verwendeten Canon XL1 Kamera erlaubte die Aufnahme des gesamten Bedienbereichs des Fahrers, wobei trotz des großen Bereiches später alle Displays und Anzeigen in der Videoaufzeichnung abgelesen werden konnten. Dabei wurde sowohl das Sitzmemory in der Fahrertür, der schwer zugängliche Tempomat-Hebel auf der linken Seite als auch die gesamte Mittelkonsole mit dem MMI-Display erfasst. Je nach Sitzeinstellung wurde bei manchen Probanden der Blick auf das Sitzmemory verhindert, jedoch konnte die Bedienung über die eindeutigen akustischen Signale der Tasten

nachvollzogen werden. Die Kamera wurde dabei auf einem Dreibeinstativ aufgestellt, das während der Fahrt vom Betreuer festgehalten wurde (vgl. Abbildung 7-2). Die Kamera war für den Fahrer somit nur beim Blick in den Rückspiegel sichtbar und hat die Bedienung des Fahrzeugs nicht behindert.

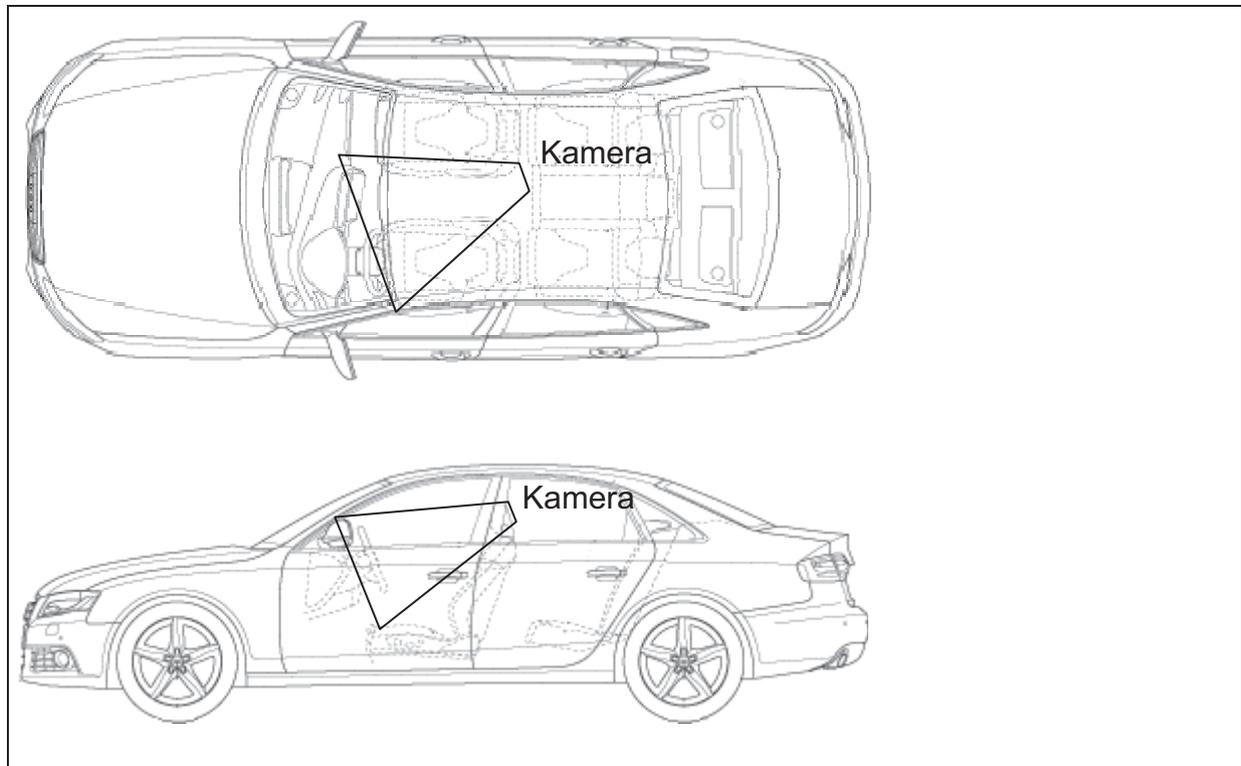


Abbildung 7-1: *Kameraperspektiven in Drauf- und Längssicht*

Quelle: (Eigene Darstellung, Quelle Fahrzeugschemata:

http://www.audi.de/audi/de/de2/neuwagen/a4/limousine/technische_daten/abmessungen.html, zugegriffen am 5.7.2008)

Auf weitere Kameras wie z.B. zur Aufnahme der Mimik und Gestik des Fahrers während der Fahrt wie es Hassenzahl und Seewald (2004) vorschlagen, wurde verzichtet. Diese Entscheidung beruht einerseits auf der Datenmenge, die durch mehrere Kameras produziert würde und nicht mehr von einer einzelnen Person während der Fahrt zu überwachen ist. Andererseits ist für die Auswertung von Mimik und Gestik nicht zwangsläufig als Teil einer Evaluation im Sinne der Design Science zu sehen.

Das Videosignal wurde mithilfe des verfügbaren FireWire-Anschlusses der Kamera direkt in einen zusätzlichen Rechner geleitet, wo das Video digital gespeichert wurde. Die Datenmenge, die bei der Beobachtung der insgesamt 67 Probanden in den drei Gruppen anfiel, belief sich am Ende der Evaluation auf ca. 450 GB. Die Stromversorgung der Kamera sowie der zusätzlichen Rechner (vgl. Kapitel 7.1.2) wurde aufgrund des Stromverbrauchs von ca. 150 Watt durch einen Wechselrichter bewerkstelligt, der die Bordspannung von 12 Volt auf 220 Volt Netzspannung konvertierte. Da sich nach dem „Abwürgen“ des Motors der Wechselrichter manchmal nicht mehr automatisch einschaltete, wurde er in den Fahrzeuginnenraum gelegt, um ihn gegebenenfalls manuell wieder aktivieren zu können.

Diese Notwendigkeit führte dazu, dass die Geräusche des Wechselrichterlüfters die Geräuschkulisse im Fahrzeug verstärkte.

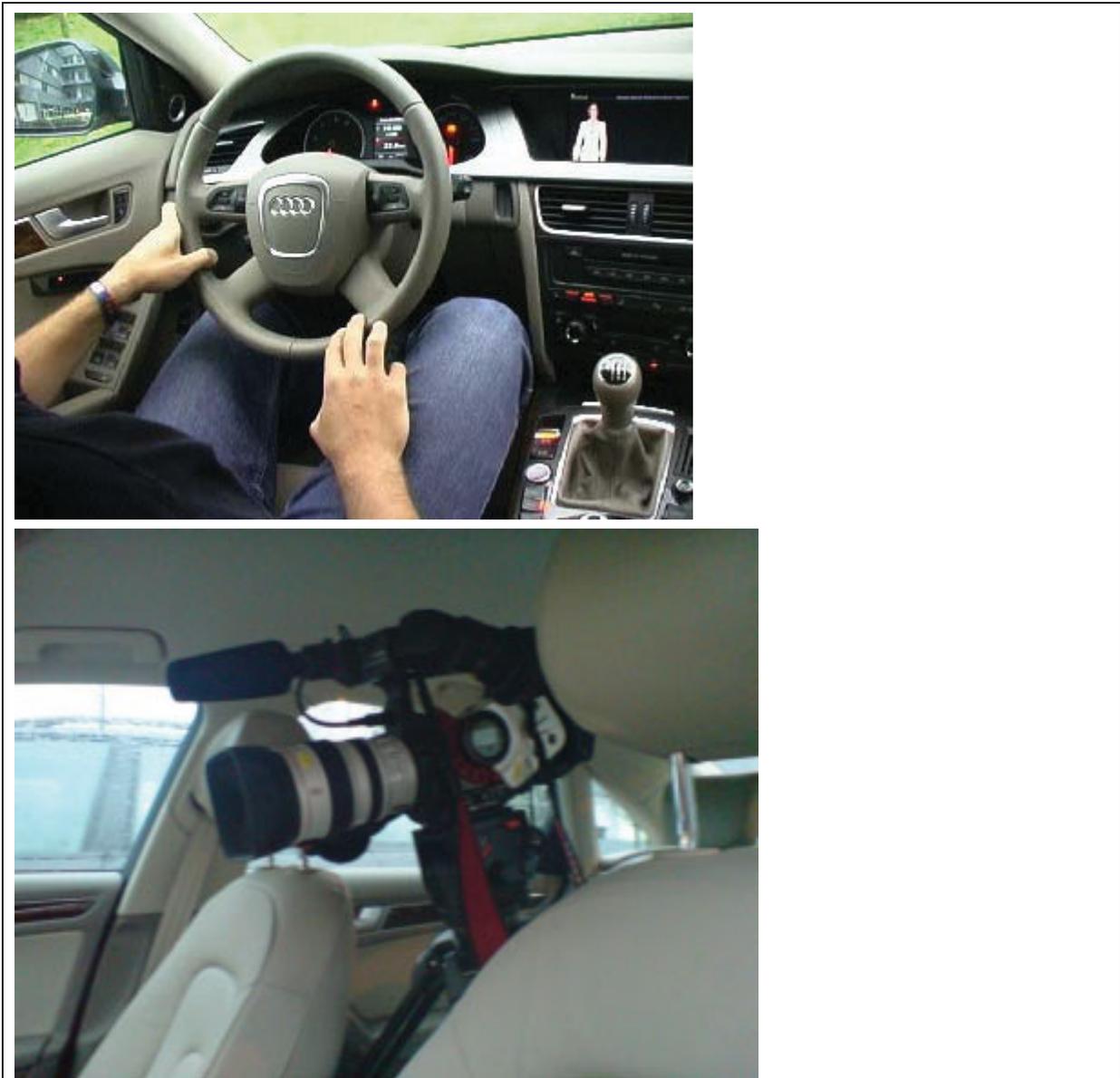


Abbildung 7-2: *Blinkwinkel und Aufbau der Kamera*
Quelle: (Eigene Darstellung)

7.1.1.2. Auswahl der Versuchspersonen

Bei der Auswahl von Versuchspersonen lassen sich im Zusammenhang dieser Untersuchung die Quantität in Form der Stichprobengröße sowie die Qualität in Form bestimmter individueller Voraussetzungen als wichtige Faktoren identifizieren. Der Stichprobenumfang ist dabei durch die Signifikanz des Versuchs als auch durch den erwarteten Effekt bestimmt. Die Ermittlung der Signifikanz beruht auf der Verwendung von Hypothese und Nullhypothese. Das Signifikanzniveau α beschreibt die Wahrscheinlichkeit mit der die Nullhypothese aufgrund der Ergebnisse einer Untersuchung fälschlich verworfen wurde. Das Signifikanzniveau, das auch als α -Fehler bezeichnet wird, ist dabei per Konvention auf 5%

oder 1% festgelegt. Liegt die Irrtumswahrscheinlichkeit bei 5% oder geringer spricht man von signifikanten Untersuchungsergebnissen. Bei einer Wahrscheinlichkeit von nur 1% handelt es sich sogar um sehr signifikante Ergebnisse (Bortz/Döring 2006, 494).

Analog zur Wahrscheinlichkeit, die Nullhypothese fälschlich zu verwerfen, besteht auch die Gefahr, die Nullhypothese fälschlicherweise nicht zu verwerfen. Diese Irrtumswahrscheinlichkeit β wird auch als β -Fehler bezeichnet. In der Regel wird eine Irrtumswahrscheinlichkeit in Höhe von 20%, 10% oder weniger als akzeptabel angenommen. Zwischen α - und β -Fehler vermutet Cohen (1988) eine Beziehung im Verhältnis 4:1, da der α -Fehler als viermal so gravierend angesehen wird. Dadurch ergibt sich bei einem α von 5% ein β von 20%. Aus β leitet sich die Teststärke (1- β) einer Untersuchung ab. Eine Teststärke von 80% besagt somit, dass mit einer Wahrscheinlichkeit von 80% korrekterweise die Hypothese akzeptiert wird (Bortz/Döring 2006, 501).

Die Effektgröße im Rahmen einer Untersuchung beschreibt den angenommenen Einfluss eines Treatments auf das Verhalten einer Gruppe und lässt sich in die Kategorien klein, mittel und groß einordnen. Die Effektgröße kann nach einer Untersuchung errechnet werden oder zu Zwecken der Bestimmung einer Stichprobengröße angenommen werden (Bortz/Döring 2006, 605). Da der Einsatz eines Avatarsystems als radikale Innovation angesehen werden kann, ist ein großer Effekt anzunehmen²⁴. Zusammen mit $\alpha=5\%$ und $\beta=20\%$ lässt sich ein minimaler Stichprobenumfang bestimmen. Da im Rahmen der Varianzanalyse neben der Kontrollgruppe zwei weitere Gruppen mit entsprechenden Treatments betrachtet werden, ergibt sich daraus eine minimal Stichprobengröße von 21 Personen pro Gruppe und somit insgesamt mindestens 63 Personen (Bortz/Döring 2006, 628).

Die Auswahl der individuellen Testpersonen ist u. a. an die Voraussetzung geknüpft, eine möglichst homogene Menge an Probanden zusammenzustellen, damit beobachtete Effekte nicht auf eine unterschiedliche Zusammensetzung der Gruppen zurückzuführen ist. Obwohl auf den ersten Blick die Zielgruppe der Audi-Fahrer für solch einen Versuch prädestiniert zu sein scheint, müssen folgende Dinge berücksichtigt werden: Zum einen handelt es sich bei dem Avatarsystem um eine Innovation, die in dieser Form frühestens in fünf bis zehn Jahren in ein Fahrzeug serienmäßig eingesetzt werden kann (vgl. Schäuuffele/Zurawka 2006, 20f). Zum anderen würden Personen, die bereits Erfahrungen mit dem Bedienkonzept eines Audi haben, das Ergebnis verfälschen, da sie nicht darauf angewiesen wären, mit dem Avatar zu interagieren. Darüber hinaus ergibt sich aufgrund der Umstände eine Einschränkung bei der Auswahl für die Testpersonen der dritten Gruppe, die das Avatarsystem während der Fahrt benutzt. Aus Versicherungsgründen können nur Mitarbeiter des Lehrstuhls für Wirtschaftsinformatik das Fahrzeug fahren und stellen somit komplett die Probanden für die Gruppe des zweiten Treatments (Avatarsystem während der Fahrt) dar.

²⁴ Im Falle einer Varianzanalyse, wie sie hier für den summativen Teil der Evaluation vorliegt, errechnet sich die Effektgröße E durch $E = \frac{\sigma_{\mu}}{\sigma}$, wobei σ_{μ} die Streuung der gemäß der Hypothese erwarteten Populationsmittelwerte und σ der Streuung des Merkmals in der Population entspricht (Bortz/Döring 2006, 614). Bei einer Varianzanalyse spricht man aber einer Effektgröße von 0,40 von einem großen Effekt (Bortz/Döring 2006, 606).

Ausgehend von dieser Vorgabe kann keine saubere Zufallsstichprobe gezogen werden, da somit die Zusammenstellung von Gruppe 3 vorgegeben ist und sich nicht an der Grundgesamtheit orientiert. Angesichts des explorativen und hypothesengenerierenden Charakters der beschriebenen Forschung liegt bei der untersuchten Menge an Probanden somit eine Pseudozufallsstichprobe vor (Bortz/Döring 2006, 480). Um dennoch homogene Gruppen zu erreichen, wurden die beiden verbleibenden Gruppen analog zur existierenden dritten Gruppe „manuell“ geschichtet. Die insgesamt 67 Probanden wurden im Umfeld der Universität rekrutiert. Als Anreiz zur Teilnahme wurde unter den Probanden ein aktuelles Apple iPod-Modell verlost. Die demographischen Daten der Probanden sind nach Gruppen aufgeteilt in Tabelle 7-2 zu sehen.

Ausgehend von Gruppe 3 wurden die anderen beiden Gruppe so gewählt, dass jeweils 3 weibliche Probanden unter- und oberhalb der 30-Jahre-Grenze liegen, während 12 männliche Probanden unter 30 und 4 männliche Probanden über 30 Jahren liegen. Die Alters- und Geschlechtsverhältnisse der vorgegebenen dritten Gruppe wurden somit weitestgehend auf die anderen beiden Gruppen übertragen.

		Alter				
Gruppe	Geschlecht	18-29	30-39	40-49	50+	
Gruppe 1	Weiblich	3	2	1		6
	Männlich	12	4			16
		15	6	1		22
Gruppe 2	Weiblich	3	2		1	6
	Männlich	11	5	1		17
		14	7	1	1	23
Gruppe 3	Weiblich	3	3			6
	Männlich	12	4			16
		14	6			22
Summe		44	20	2	1	67

Tabelle 7-2: Demographische Daten der Probanden

Quelle: (Eigene Darstellung)

7.1.1.3. Beschreibung der Versuchsstrecke

Die Experimente der Gruppen 1 und 2, die bei stehendem Fahrzeug durchgeführt wurden, erfolgten auf dem Forschungsgelände der TU München in Garching oder auf dem Stammgelände in der Innenstadt Münchens.

Die Evaluation aller drei Gruppen fand im September 2008 statt wodurch warmes und trockenes Wetter vorlag, was die Fahrten der dritten Gruppe somit nicht negativ beeinflusste. Lediglich zwei Testfahrten sind bei leichtem Nieselregen durchgeführt worden. Das Fahren eines unbekanntes Fahrzeugs unter der Verwendung eines unbekanntes Hilfesystems in einer Beobachtungssituation stellt bereits eine Grundablenkung dar. Aus diesem Grund wurde bei

der Auswahl der Versuchsstrecke eine Autobahn-Route gewählt, die das geringste Maß an Ablenkung durch die Fahrstrecke bedeutet. Nach Fastenmeier (1995) hängt die Ablenkung durch die Versuchsstrecke von der Menge an Kreuzungen, der Art der Knotenpunktregelung, sowie den Sicht- und Witterungsverhältnissen ab (vgl. auch von Benda 1977). Die geringsten Anforderungen an den Fahrer stellen somit Autobahnstrecken dar, da sie keine Kreuzungen mit Ampeln und Vorfahrtsregeln haben. Darüber hinaus ist der Gegenverkehr baulich getrennt und stehende Fahrzeuge und Passanten stellen keine Gefahr dar.

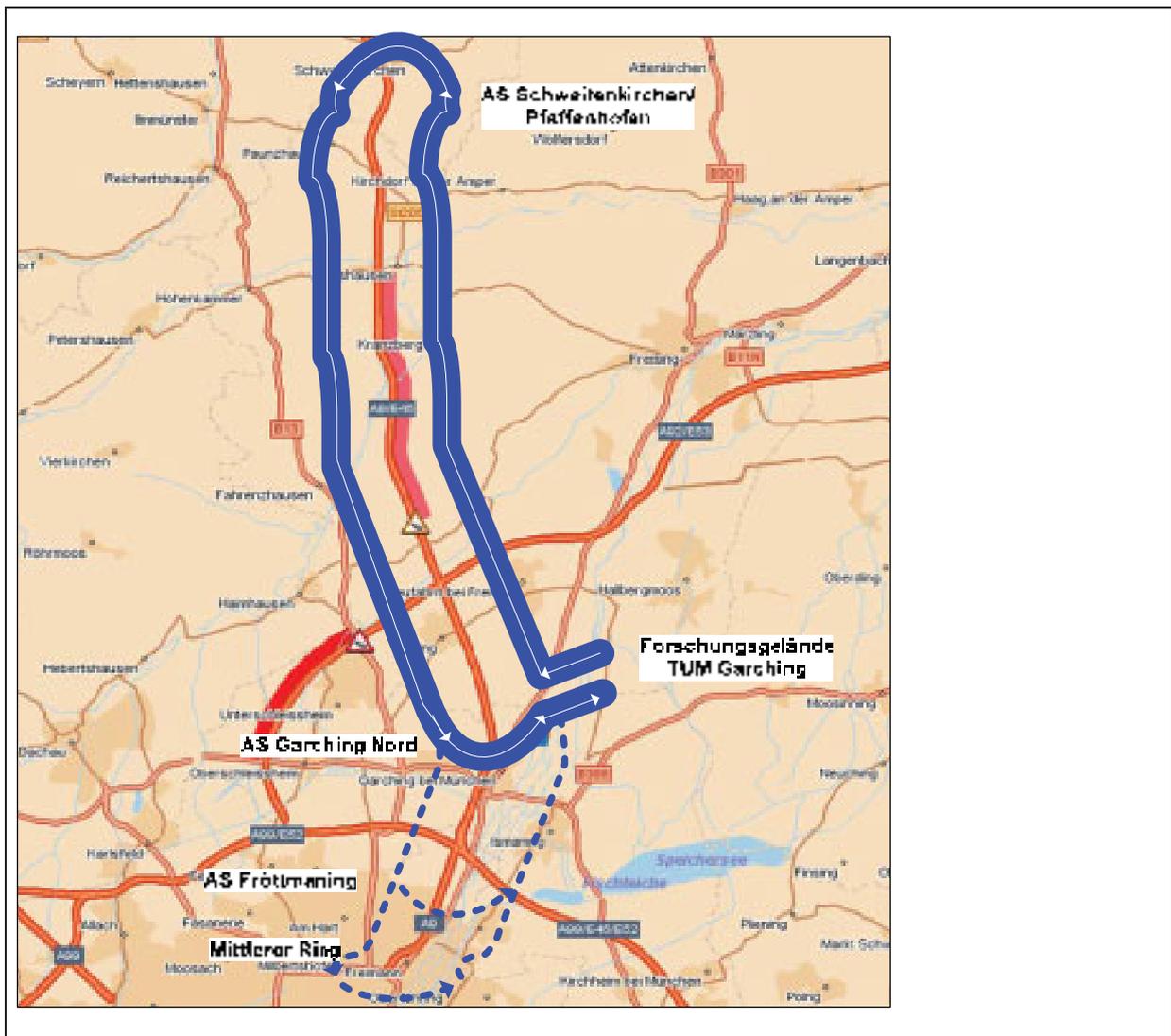


Abbildung 7-3: *Route der Versuchsstrecke und der Verlängerungsoptionen*
 Quelle: (Eigene Darstellung, Kartenmaterial von <http://www.map24.de>,
 zugegriffen am 16.10.2008)

Die gewählte Versuchsstrecke begann am Forschungsgelände der TU München in Garching und führte auf der A9 Richtung Nürnberg bis zur Ausfahrt Schweitenkirchen/ Pfaffenhofen. Dieses Teilstück diente der Gewöhnung der Probanden an das Fahrzeug, ohne eine zusätzliche Belastung durch Testaufgaben zu erfahren. Nach dem Wenden erfolgte die Bearbeitung der insgesamt fünf Themenkomplexe auf der Strecke von Schweitenkirchen bis zur Ausfahrt Garching Nord. Dauerte die Bearbeitung der Aufgaben länger an, so wurde die

Strecke verlängert, indem weitergefahren und später wieder gewendet wurde. Die Probanden wurden zu Beginn der Fahrt auf diese Verlängerungsoptionen hingewiesen, um keinen zeitlichen Stress aufkommen zu lassen. Die Kernstrecke sowie die Verlängerungsoptionen sind in Abbildung 7-3 dargestellt.

Die Fahrversuche wurden montags bis donnerstags zwischen 10:00 und 16:00 Uhr und freitags zwischen 10:00 und 14:00 Uhr durchgeführt, um nicht in die Verkehrsstoßzeiten zu kommen und dadurch zusätzliche Ablenkung zu erzeugen. Die Kernstrecke wies konstant viel Verkehr und mit einer Ausnahme keine Staus auf. Die zweite Verlängerungsoption wies zu der Zeit zwischen der Ausfahrt Freimann und dem mittleren Ring eine Baustelle mit Fahrbahnverengung in beide Fahrtrichtungen auf, was zu einer Unterbrechung der Aufgabenbearbeitung an der Stelle zur Folge hatte.

7.1.2. Durchführung der Evaluation

Jeder Proband durchlief gesondert eine dreiteilige Evaluation. Zunächst wurde den Teilnehmern ein präaktionaler Fragebogen zu den demographischen Eigenschaften und Vorkenntnissen hinsichtlich des Autofahrens sowie des Umgangs mit technischen Geräten und Anwendungen vorgelegt. Der Fragebogen war für alle drei Gruppen gleich und wurde meist im Fahrzeug ausgefüllt (vgl. Anhang B). Die Angaben im Fragebogen bestimmten, welche der ersten beiden Gruppen ein Proband zugeordnet wurde, sofern er nicht von vorneherein für Gruppe 3 vorgesehen war.

Nach der Beantwortung des ersten Fragebogens wurden die Probanden auf das weitere Vorgehen bei der Bearbeitung der Testaufgaben hingewiesen. Vor den Hinweisen zu den Testaufgaben wurden die Probanden um ihre Einwilligung gebeten, die Aufgaben auf Video aufzeichnen zu dürfen. Die Probanden wurden informiert, dass das Handbuch bzw. das Avatarsystem nur eine Hilfestellung sei und keine Pflicht zur Verwendung bestünde. Sie mussten den Abschluss jeder Teilaufgabe akustisch bemerkbar machen und konnten Teilaufgaben auslassen, wenn sie diese nicht bearbeiten konnten. Zudem wurde betont, dass die Aufgaben unabhängig von der Fahrerfahrung und den Kenntnissen über Audi-Fahrzeuge voraussichtlich nicht ohne Hilfe zu lösen sind, um bei den Probanden keinen Frust aufkommen zu lassen.

Teilnehmer in Gruppe 2 und 3 wurden zudem darüber informiert, dass das zu testende Hilfesystem vor allem vollständige Sätze als Fragen verstehen kann und kein Bedarf zur Verwendung von Kommandos oder Stichwörtern besteht. Zudem wurde der Einbauort des Mikrofons im Dachhimmel im Bereich der Leselampen erwähnt, damit die Probanden nicht die serienmäßig vorhandenen Mikrofone im Lenkrad verwendeten. Abschließend erhielten die Teilnehmer eine kurze Einweisung in die Benutzung des Push-to-Talk-Buttons im Lenkrad und mussten mit diesem Wissen selbst das Tutorial des Avatarsystem starten, indem Sie den Avatar fragen sollten, wie er funktioniert.

Teilnehmer der fahrenden Gruppe 3 wurden zudem über den Streckenverlauf mit den Verlängerungsoptionen informiert und darauf hingewiesen, dass sie jederzeit die Fahrt

abbrechen können, wenn ein entsprechender Grund vorläge. Ein Fahrabbruch kam jedoch nicht vor.

Alle Probandengruppen erhielten die gleichen Aufgaben, die jeweils durch einen Begleiter vorgelesen wurden. Dabei wurde zunächst der gesamte Aufgabenkomplex (z.B. Sitzmemory) mit allen Teilaufgaben vorgelesen und die Teilaufgaben später Schritt für Schritt zur Bearbeitung noch einmal vorgetragen. Zur Erhöhung der Sicherheit in der Testsituation wurden gleich zu Beginn in den ersten beiden Aufgaben der Spurhalte- und der Spurwechselassistent aktiviert. Der Begleiter saß entweder auf dem Beifahrersitz oder während der Fahrten hinter dem Beifahrersitz. Dort konnte er neben der Kamera und dem Computer zur Aufzeichnung im Falle von Gruppe 2 und 3 auch das Avatarsystem mit einem weiteren Rechner überwachen. Die technische Überwachung des Avatarsystems erwies sich in wenigen Fällen als notwendig, da sich das GUI-Framework manchmal „aufhängte“. Während der Fahrt kam dem Betreuer zudem die Aufgabe zuteil, das Kamerastativ gegen Umfallen abzusichern. Abbildung 7-4 zeigt den Arbeitsplatz des Betreuers auf der Rückbank des Fahrzeugs während der Fahrversuche mit den Computern zu Videoaufzeichnung und der technischen Überwachung des Avatarsystems sowie der Kamera.

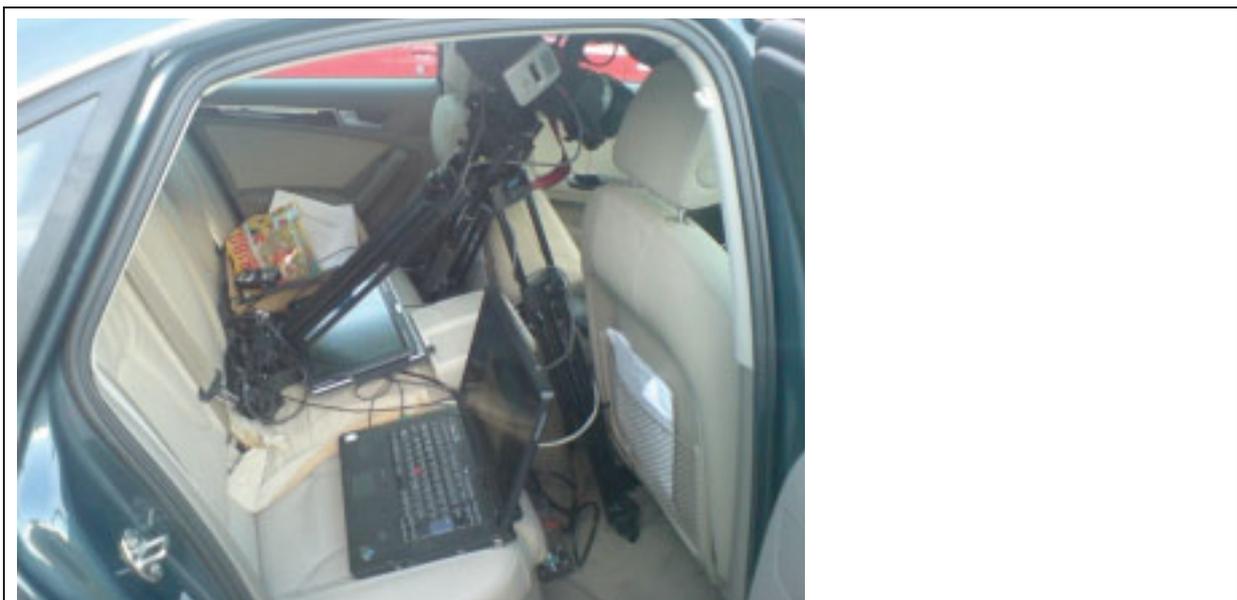


Abbildung 7-4: Arbeitsplatz des Betreuers auf der Rückbank während der Versuchsfahrten
Quelle: (Eigene Darstellung)

Nach Abschluss der Aufgabenbearbeitung bzw. der Versuchsfahrt wurde die Videoaufzeichnung beendet und der Proband erhielt einen postaktionalen Fragebogen (vgl. Anhang C). Dieser war je nach Verwendung des Avatarsystems oder des Handbuchs anders gestaltet. Gruppe 1 erhielt nur wenige Fragen zum Handbuch sowie einige Wissensaufgaben, während Gruppen 2 und 3 zusätzlich ausführlich zum Umgang mit dem Avatarsystem befragt wurden. Teilnehmer der Gruppen 2 und 3 erhielten zudem die Möglichkeit, in offenen Fragen frei ihre Eindrücke und Meinungen zu schildern. Zum Abschluss der Beobachtung erhielt jeder Proband eine Süßigkeit mit dem Hinweis auf die spätere Verlosung des iPod. Eine

schematische Übersicht der Evaluationsinstrumente und deren Unterscheidung nach Gruppen ist in Abbildung 7-5 dargestellt.

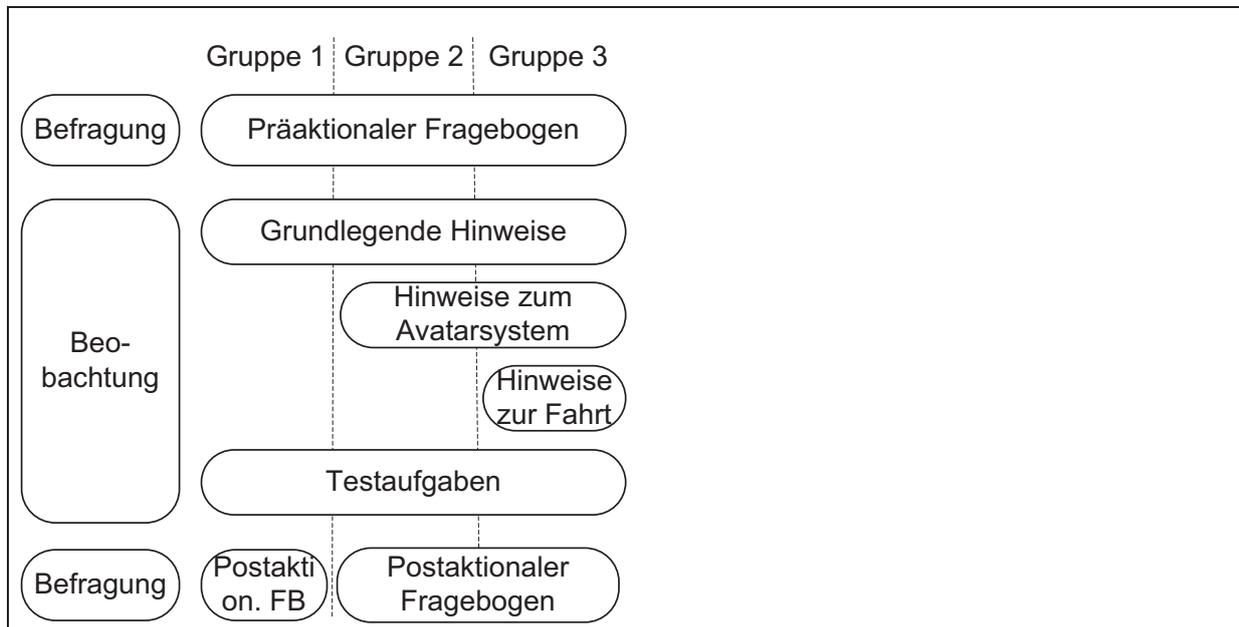


Abbildung 7-5: Schematischer Ablauf der Evaluation
Quelle: (Eigene Darstellung)

7.2. Ergebnisse

Die gedruckten, präaktionalen und postaktionalen Fragebögen wurden nach dem Durchlauf aller 67 Probanden digital erfasst. Die aufgenommenen Videos wurden nach der Evaluation analysiert, indem sie während des Betrachtens hinsichtlich verschiedener aufgabenspezifischer Werte erfasst wurden. So wurden für jede der zwölf relevanten Teilaufgaben folgende Werte mithilfe der Videos erfasst, da die Beobachtung durch die Betreuer zum Zeitpunkt der Evaluation deren besondere Relevanz vermuten ließen:

- Wurde das Handbuch bzw. das Avatarsystem verwendet und hat es zu der entsprechenden Teilaufgabe Informationen geliefert?
- Wurde die Teilaufgabe erfolgreich bearbeitet?
- Relevante Zeitpunkte:
 - Ende der Aufgabenbeschreibung durch den Betreuer
 - Erste Spracheingabe zur Teilaufgabe
 - Korrekte Bedienung des relevanten Bedienelements
 - Rückmeldung über Erfolg oder Misserfolg durch Benutzer

- Anzahl der Bedienung der richtigen bzw. falschen Bedienelemente
- Anzahl der Verwendung von Sprachbefehlen der Benutzeroberfläche (Zurück, Wiederholen, Weiter, Stopp, Hinweise)
- Anzahl falsch bzw. richtig und gar nicht erkannter Spracheingaben
- Anteil der Kommandos an Eingaben, die sich nicht auf die Benutzerführung beziehen

Die Werte der oben genannten Merkmale der Videoaufzeichnung wurden für jeden Probanden auf einen gesonderten Bogen erfasst, wobei das ablaufende Video jederzeit pausiert werden konnte, um Eintragungen vorzunehmen. Die Beobachtungsbögen wurden nach Abschluss aller Sichtungen digitalisiert. Insbesondere während Versuchsfahrten ist die Verbindung zwischen der Kamera und dem aufzeichnenden Computer ausgefallen, so dass 10 der insgesamt 804 Aufgabenbearbeitungen nicht optisch ausgewertet werden konnten.

Im Folgenden werden die wichtigsten Ergebnisse aus den Fragebögen sowie den Videoanalysen vorgestellt. Aus Gründen der Übersicht sind einige der insgesamt ca. 480 für jeden Probanden erfassten Werte ausgelassen. Zunächst werden die Vorkenntnisse der Probanden betrachtet, bevor die Ergebnisse der Aufgabenbearbeitung im Fahrzeug vorgestellt werden. Anschließend werden weitere Vergleichswerte herangezogen, die in jeder der drei Gruppen erfasst wurden. Die Evaluation der technischen Komponenten des Avatarsystems erfolgt anschließend nur für die Gruppen 2 und 3. Die Antworten auf die offenen Fragen der Teilnehmer in den Gruppen 2 und 3 werden danach kategorisiert und näher betrachtet. Aus den Ergebnissen der Evaluation werden abschließend fünf Hypothesen extrahiert. Signifikante Unterschiede, die sich auf der Basis von Varianzanalysen ergeben, werden im Folgenden explizit genannt. Erfolgt keine Nennung, so liegt kein statistisch signifikanter Unterschied vor.

7.2.1. Vorkenntnisse der Probanden

Die Vorkenntnisse der Probanden wurden hinsichtlich der allgemeinen Bildung, dem Umgang mit technischen Geräten und Anwendungen, der Fahrerfahrung und den Erfahrungen mit Fahrzeugen der Marke Audi untersucht.

7.2.1.1. Ausbildung der Probanden

Die Probanden wurden nach dem höchsten aktuellen Bildungsabschluss sowie der Fachrichtung befragt, wobei für die Fachrichtung eine Mehrfachnennung zugelassen war, um Mehrfachabschlüsse und interdisziplinäre Ausbildungen abbilden zu können. Eine Übersicht der erhobenen Daten ist in Tabelle 7-3 aufgeführt. Ein Chi-Quadrat-Test für jede einzelne

Gruppe ergab dabei Werte von 6,31 für Gruppe 1, 4,81 für Gruppe 2 und 2,13 für Gruppe 3, so dass eine Unabhängigkeit der Merkmale Bildungsabschluss und Fachrichtung vorliegt²⁵.

			Fachrichtung				
Höchster Bildungsabschluss	Anzahl	technisch	wirtschaftswiss.	naturwiss.	geisteswiss.	sozial	math.
Gruppe 1	Realschule/ Mittlere Reife	2		1	1		
	Abitur/ Fachabitur	11	7	3	6		
	Berufsausbildung	1	1		1		
	(Fach-) Hochschulabschluss	8	5	2	3	2	
	Summe		13	6	11	2	
Gruppe 2	Abitur/ Fachabitur	5	4	1	1		
	Berufsausbildung	2	1	1			
	(Fach-) Hochschulabschluss	15	9	8	5	2	
	Promotion	1		1			
	Summe Gruppe 2		14	11	6	2	
Gruppe 3	Berufsausbildung	1		1			
	(Fach-) Hochschulabschluss	19	12	12	4		1
	Promotion	2	2	2			
	Summe Gruppe 3		14	15	4		1
Gesamtergebnis			41	32	21	4	1

Tabelle 7-3: *Bildungsabschlüsse und Ausbildungsfachrichtungen der Probanden*
Quelle: (Eigene Darstellung)

Die Höhe des Bildungsabschlusses ist gruppenspezifisch verschieden verteilt. Während in der ersten Gruppe die meisten Probanden das Abitur und annähernd so viele Probanden einen Hochschulabschluss besitzen, liegt der Schwerpunkt der anderen beiden Gruppen wesentlich deutlicher auf dem Hochschulabschluss. Die Zusammenstellung der dritten Gruppe lässt sich damit erklären, dass sie nur aus Mitarbeitern des Lehrstuhls für Wirtschaftsinformatik zusammengestellt wurde, deren Mehrheit wissenschaftliche Mitarbeiter mit einer abgeschlossenen Hochschulausbildung sind.

Die Fachrichtung der jeweiligen Ausbildungen lassen über alle Gruppen hinweg einen Fokus auf dem technischen Aspekt erkennen. Wirtschaftswissenschaftliche Ausprägungen nehmen von Gruppe 1 zu Gruppe 3 hin zu, während der naturwissenschaftliche Aspekt von Gruppe 1 zu Gruppe 3 hin abnimmt. Geisteswissenschaftliche Ausrichtungen tauchen nur in den Gruppen 1 und 2 auf, während nur in Gruppe 3 soziale und mathematische Ausprägungen in geringer Zahl genannt wurden. Die weitere Option einer handwerklichen Ausbildung wurde von keinem Probanden angegeben.

²⁵ Bei einem angenommenen α -Fehler von 5% wird erst ab einem Wert von über 9,488 eine Abhängigkeit zweier Merkmale angenommen.

7.2.1.2. Technische Vorkenntnisse

Die technischen Vorkenntnisse der Probanden spiegeln die Erfahrungen mit verwandten oder von der Bedienung her ähnlichen Systemen wider. Darüber hinaus spielt auch der Einsatz solcher Systeme insbesondere im Umfeld eines Fahrzeugs eine Rolle und beeinflusst den Umgang mit dem neuartigen Avatarsystem im Fahrzeug.

Verwendungshäufigkeit		Technische Anwendung						
		Computer	Internet	Navigationssysteme im Auto	System mit Spracheingabe	Spracheingabe im Auto	Systeme mit Sprachausgabe	Avatarsysteme
Gruppe 1	Nie			8	7	20	10	9
	Wenige Male im Jahr			6	4	1	5	7
	Mindestens 1x/ Monat			5	3		4	4
	Mindestens 1x/ Woche	1		2			2	2
	Täglich	21	22	1	8	1	1	
Gruppe 2	Nie			7	6	18	8	6
	Wenige Male im Jahr			8	5	3	10	12
	Mindestens 1x/ Monat			4	2	1	4	5
	Mindestens 1x/ Woche			2	3			
	Täglich	23	23	2	7	1	1	
Gruppe 3	Nie			3	7	21	3	8
	Wenige Male im Jahr			6	4		6	12
	Mindestens 1x/ Monat			5	1	1	4	1
	Mindestens 1x/ Woche			6	1		8	1
	Täglich	22	22	2	9		1	

Tabelle 7-4: Technische Vorkenntnisse der Probanden
Quelle: (Eigene Darstellung)

Über alle Gruppen hinweg lässt sich bei den Versuchsteilnehmern ein homogenes Bild hinsichtlich der Benutzung von Computern und Internet erkennen, die von fast allen Befragten täglich genutzt werden. Die Verwendung eines festeingebauten Navigationssystems im Fahrzeug erfolgt in den Gruppen 1 und 2 überwiegend weniger als einmal in der Woche während die Verwendung in Gruppe 3 häufiger erfolgt. Die Verwendung von Systemen mit Spracheingabe weist über alle Gruppen hinweg eine U-Kurve auf. So verwenden die Teilnehmer solche Systeme überwiegend entweder nie oder täglich. Ein gleiches Bild über alle Gruppen hinweg ergibt sich bei der Verwendung von Spracheingabe im Fahrzeug: In jeder Gruppe haben über 78% nie eine Spracheingabe im Fahrzeug bedient. Systeme mit Sprachausgabe werden von über 45% von Gruppe 1 nie verwendet und die Verwendung zur täglichen Nutzung hin nimmt immer weiter ab. Ein deutlicheres Bild ergibt sich hingegen für

Gruppe 2 bei der über 78% solche Systeme nie oder wenige Male im Jahr einsetzen. Für Gruppe 3 ergibt sich eine Doppelspitze in der Verteilung bei der Verwendung weniger Male im Jahr und bei der wöchentlichen Benutzung.

Zur Verwendung von Avatarsystemen wurden die Versuchsteilnehmer in einer gesonderten Frage befragt, in der die Definition eines solchen System beschrieben und ein bekanntes Beispiel (der Bausparfuchs von Schwäbisch-Hall) genannt wurde. In Gruppe 1 wurde die Verwendung von Avatarsystemen derart beschrieben, dass der Schwerpunkt bei gar keiner Verwendung lag und bis zur wöchentlichen Benutzung hin abnahm. In den Gruppen 2 und 3 zeichnete sich ein Schwerpunkt bei der Benutzung weniger Male im Jahr. In allen Gruppen lag keine tägliche Verwendung vor und nur 4% aller Befragten habe eine wöchentliche Benutzung angegeben. Die erhobenen Daten zu allgemeinen technischen Vorkenntnissen sind in Tabelle 7-4 zu sehen. Ein Chi-Quadrat-Test über jede Gruppe deutet mit Werten von 145,08, 165,19 und 201,38 auf eine Abhängigkeit der beiden Merkmale hin.

7.2.1.3. *Vorkenntnisse im Bereich Autofahren und Erfahrungen mit der Marke Audi*

Die Vorkenntnisse beim Autofahren sowie die Erfahrung mit Fahrzeugen der Marke Audi beeinflussen das Wissen um bestimmte Funktionen und deren Bedienung. Die Fahrerfahrungen wurden ermittelt, indem die Probanden nach der Fahrhäufigkeit sowie nach der Dauer des Führerscheinbesitzes befragt wurden.

Die Fahrhäufigkeit ist in Gruppe 1 gleichmäßig verteilt, wobei der Großteil der Teilnehmer mehrere Male im Monat bzw. in der Woche Auto fahren. Diese Verteilung verschiebt sich über Gruppe 2 zu Gruppe 3 hin dahingehend, dass häufiger ein Fahrzeug geführt wird. Eine ähnliche Veränderung der Verteilung ist auch bei der Dauer des Führerscheinbesitzes zu beobachten. Während in Gruppe 1 die Dauer gleichmäßig verteilt ist, verschiebt sie sich zu Gruppe 3 hin zugunsten der Dauer über 5 Jahren. Diese Anhäufung erfahrener Fahrzeugführer in Gruppe 3 lässt sich teilweise dadurch erklären, dass den Mitarbeitern des Lehrstuhls freigestellt wurde, ob sie das Versuchsfahrzeug führen möchten oder nicht. Dadurch sind überwiegend erfahrenere Fahrer in Gruppe 3 verblieben, während die Teilnehmer, die das Auto nicht fahren wollten auf die Gruppen 1 oder 2 verteilt wurden. Die vollständigen Daten zur Fahrerfahrung der Probanden sind in Tabelle 7-5 aufgeführt.

Neben den allgemeinen Vorkenntnissen können auch die Erfahrungen mit der Marke Audi von Bedeutung sein. Die Probanden wurden befragt, ob sie regelmäßig ein Audi-Fahrzeug mit dem Baujahr nach 2000 fahren, da erst in etwa ab diesem Baujahr die neuartigen Assistenzsysteme in Fahrzeugen von Audi verbaut wurden. In den Gruppen 1 und 3 hatten jeweils 2 Probanden Erfahrungen in dieser Hinsicht, während es bei Gruppe 2 insgesamt 6 Probanden waren.

Die Probanden wurden zudem nach ihrer Einstellung zur Marke Audi befragt, wobei die Antwortmöglichkeiten auf einer Likert-5-Skala von „sehr schlecht“ über „neutral“ bis hin zu „sehr gut“ reichten. Die Mittelwerte bewegten sich bei allen Gruppen im Bereich von 4,05 bis 4,3 mit einer Standardabweichung σ von 0,65 bzw. 0,66 bei den Gruppen 1 und 3 sowie von 0,93 für Gruppe 2. Somit bewerteten alle drei Gruppen ihre Einstellung der Marke gegenüber

durchschnittlich als gut und besser. Die ermittelten Werte sind in Abbildung 7-6 grafisch als Balkendiagramme dargestellt, wobei die schwarzen Linien die Standardabweichungen vom Mittelwert ausgehend anzeigen.

		Dauer des Führerscheinbesitzes [Jahre]				
Fahrhäufigkeit		0-2	2-5	5-10	10+	Summe
Gruppe 1	1-11x/ Jahr	3		1		4
	1-3x/ Monat	1	3	1	2	7
	1-5x/ Woche	1	2	2	3	8
	Jeden Tag				3	3
	Summe	5	5	4	8	22
Gruppe 2	1-11x/ Jahr	1	1	2	2	6
	1-3x/ Monat			1	1	2
	1-5x/ Woche		3	1	3	7
	Jeden Tag			2	6	8
	Summe	1	4	6	12	23
Gruppe 3	1-11x/ Jahr			1	1	2
	1-3x/ Monat			2	2	4
	1-5x/ Woche			2	2	4
	Jeden Tag			4	8	12
	Summe			9	13	22
Gesamtergebnis		6	9	19	33	67

Tabelle 7-5: *Fahrerfahrung der Probanden*
Quelle: (Eigene Darstellung)

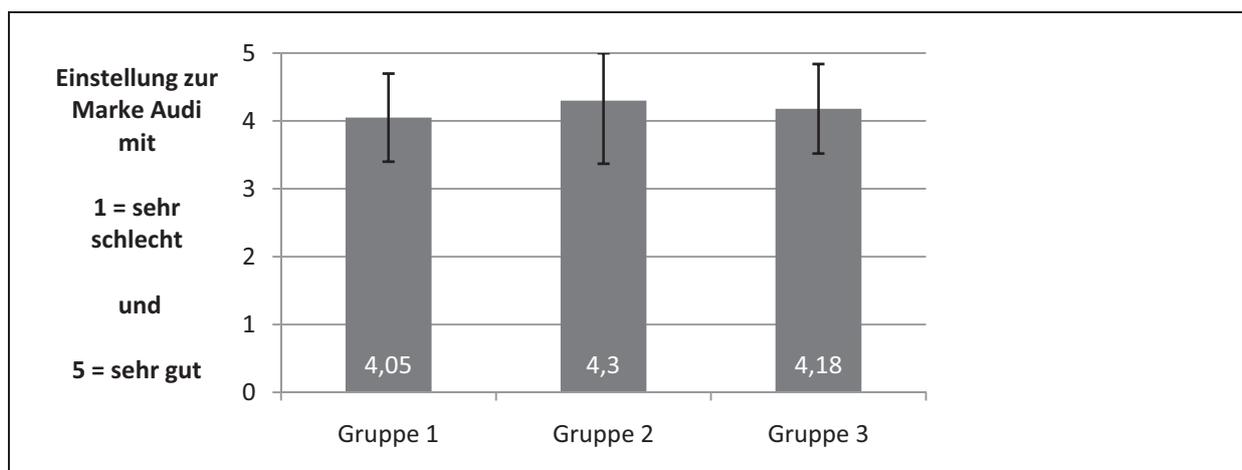


Abbildung 7-6: *Einstellung der Probanden gegenüber der Marke Audi*
Quelle: (Eigene Darstellung)

7.2.2. Bearbeitung der Testaufgaben

Im Rahmen der Aufgabenbearbeitung wurden aus den aufgenommenen Videos die Erfolge bei der Bearbeitung der verschiedenen Aufgaben, die Zeiten zur Bearbeitung und die Zielsicherheit bei der Benutzung der entsprechenden Bedienelemente extrahiert. Als Folge der Aufgabenbearbeitung kann auch das im postaktionalen Fragebogen geprüfte Wissen über die Funktionsweise der Bedienelemente aus den Testaufgaben angesehen werden.

7.2.2.1. Bearbeitungserfolge

Bei der Ermittlung des Bearbeitungserfolgs wurden nur Aufgaben als erfolgreich gewertet, die vom Probanden als erfolgreich rückgemeldet wurden und wirklich erfolgreich bearbeitet wurden, d.h. das entsprechende Bedienelement wurde korrekt bedient. Demzufolge wurden Aufgaben, die unbewusst erfolgreich gelöst wurden oder falsch gelöste Aufgaben mit positiver Rückmeldung als erfolglos gewertet. Die insgesamt zehn Aufgabenbearbeitungen, zu denen aufgrund eines technischen Problems keine Videoaufnahme bestand, wurden bei dieser Auswertung ausgelassen. Bei der Analyse des Bearbeitungserfolgs wurde zudem unterschieden, ob der Proband das Handbuch bzw. das Avatarsystem überhaupt verwendet hat. Im Falle des Avatarsystems galt es als verwendet, wenn der Proband eine zur Aufgabenstellung passende Antwort im Laufe der Aufgabenbearbeitung erhalten hat. Insgesamt konnten maximal zwölf Aufgaben von jedem Teilnehmer gelöst werden. Die im Folgenden analysierten Werte sind in Abbildung 7-7 zu sehen, wobei die schwarzen Linien die jeweilige Standardabweichung zeigen.

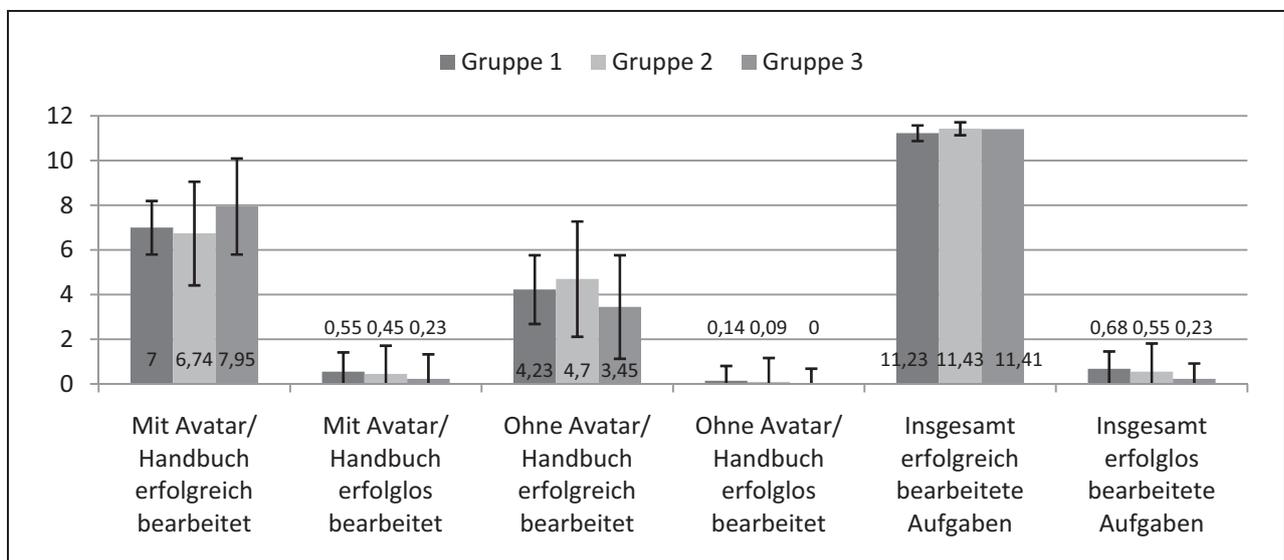


Abbildung 7-7: *Verwendung von Handbuch bzw. Avatarsystem und Bearbeitungserfolg*
Quelle: (Eigene Darstellung)

7.2.2.2. Bearbeitungszeiten erfolgreich bearbeiteter Testaufgaben

Bei den mit Handbuch bzw. Avatarsystem erfolgreich gelösten Aufgaben zeigt sich, dass Gruppe 2 knapp weniger Aufgaben lösen konnte als Gruppe 1, während die Probanden in Gruppe 3 durchschnittlich fast eine ganze Aufgabe mehr lösen konnte. Auch die Anzahl erfolglos gelöster Aufgaben nimmt von Gruppe 1 zu Gruppe 3 hin ab. Bei der Bearbeitung

ohne Handbuch bzw. Avatarsystem wurden in Gruppe 2 durchschnittlich am meisten Aufgaben gelöst während Gruppe 1 etwa eine halbe Aufgabe weniger lösen konnte und bei Gruppe 3 nochmals durchschnittlich fast eine ganze Aufgabe weniger gelöst werden konnte. Insgesamt kann bei der Bearbeitung der Aufgaben eine leicht verbesserte Erfolgsquote bei der Bearbeitung mit Avatarsystem im Gegensatz zum Handbuch beobachtet werden. Parallel dazu nimmt auch die Gesamtanzahl erfolglos bearbeiteter Aufgaben von Gruppe 1 zu Gruppe 3 hin ab.

Es lässt sich aufgrund dieser Beobachtungen vermuten, dass die Abhängigkeit vom Avatarsystem während der Fahrt größer war als bei der Verwendung des Avatarsystems im Stand und dies zur vermehrten Benutzung bei der Aufgabebearbeitung führte. Gleichzeitig konnten dadurch jedoch auch mehr Aufgaben erfolgreich gelöst werden.

Die Bearbeitungszeit der jeweiligen Testaufgaben ist ein Indiz für die Einfachheit des Umgangs des Benutzers mit dem Handbuch bzw. dem Avatarsystem. Es liegt der Schluss nahe, dass die Bearbeitung einer Aufgabe unter einem bestimmten Treatment umso einfacher ist, desto kürzer die entsprechende Zeit ist. Dabei muss jedoch auch die Aufgabenkategorie und damit der Umfang der Aufgabe betrachtet werden. Zur Ermittlung der Bearbeitungszeit wurde aus den Videoaufnahmen die Zeit vom Ende der Aufgabenbeschreibung durch den Betreuer bis zur Rückmeldung durch den Probanden aufgezeichnet. Erfolgte keine explizite Rückmeldung, so wurde die Zeit auf eine Sekunde vor Beginn der Beschreibung der nächsten Aufgabe geschätzt. Zudem wurde unterschieden, ob der Proband das Handbuch bzw. Avatarsystem überhaupt benutzt hat oder nicht. In Tabelle 7-6 sind alle ermittelten Durchschnittszeiten nach Gruppen dargestellt. Die Zahlen geben die Minuten und Sekunden an, die insgesamt in der Gruppe zur Bearbeitung der Aufgabe verwendet wurden, während die Angaben in Klammern die Werte für die Handbuch- bzw. Avatarsystembenutzer darstellen. Die Zahlen vor den Zeitangaben zeigen, wie viele Probanden innerhalb der Gruppe die Aufgabe erfolgreich gelöst haben und bei der Berechnung des Durchschnittswerts berücksichtigt wurden. Auf die Darstellung von Standardabweichungen wird aus Gründen der Übersichtlichkeit verzichtet.

Die Gesamtbearbeitungszeiten der Aufgaben der ersten Kategorie zeigen ein klares Bild: Die Gesamtbearbeitungszeit nimmt von Gruppe 1 zu Gruppe 3 hin um jeweils 20 bis 30 Sekunden zu. Betrachtet man jedoch die einzelnen Aufgaben, so zeigt sich, dass bei der ersten Aufgabe der Unterschied zwischen Gruppe 2 und 3 mit einer Differenz von fünf Sekunden verschwindend gering ist und bei der zweiten Aufgabe Gruppe 2 ein besseres Ergebnis mit dem Avatarsystem erzielt hat als Gruppe 1 mit dem Handbuch.

Bei den Aufgaben der Kategorie 2 ergibt sich ein differenzierteres Bild. Die Gesamtbearbeitungszeit von Gruppe 1 ist mit 6:50 Minuten deutlich besser als Gruppe 3 mit 9:29 Minuten. Jedoch weist Gruppe 3 damit ein besseres Ergebnis als Gruppe 2 mit 11:02 Minuten auf. Dieser Gesamteindruck zieht sich durch alle Teilaufgaben dieser Kategorie hindurch, so dass das Avatarsystem während der Fahrt schneller bedient wurde als im Stand und beide Avatarsystemgruppen langsamer waren als Gruppe 1 mit dem Handbuch. Während die Differenz zwischen den Avatarsystemgruppen und der Handbuchgruppe beim ersten

Aufgabenkomplex noch sehr deutlich ist, verringert sie sich jedoch bei der Aufgabe zum Bordcomputer.

Kategorie	Aufgabe	Teilaufgabe	Gruppe 1	Gruppe 2	Gruppe 3
1 Bedienschnitt	Spurhalteassistent		22: 1:32	23: 2:09 (23: 2:09)	20: 1:58 (19: 2:04)
	Spurwechselassistent		22: 1:08	23: 0:48 (22: 0:49)	21: 1:30
	Gesamt		2:39	2:56 (2:58)	3:30 (3:34)
2-5 Bedienschnitte	Sitzmemory	aktivieren	22: 0:51	23: 1:30 (21: 1:37)	20: 1:31
		Speichern	21: 0:32 (20: 0:33)	21: 1:07 (15: 1:20)	22: 0:59 (20: 1:04)
		Deaktivieren	22: 0:09 (8: 0:17)	23: 0:36 (12: 1:03)	22: 0:44 (15: 1:02)
		Gesamt	1:31 (1:40)	3:05 (4:00)	3:18 (3:37)
	Bordcomputer	Einzelfahrtspeicher	22: 2:32	23: 2:07 (22: 2:12)	22: 2:50
		Reichweite	22: 0:54 (17: 1:07)	23: 1:02 (9: 2:16)	21: 0:37 (7: 1:32)
		Audi Lane Assist	21: 1:12 (13: 1:35)	21: 1:32 (12: 2:33)	21: 1:04 (14: 1:31)
		Gesamt	4:36 (5:10)	4:42 (7:02)	4:30 (5:52)
	Gesamt		6:07 (6:50)	7:47 (11:02)	7:48 (9:29)
	6-9 Bedienschnitte	Klimaautomatik	Temperatur	22: 0:16 (4: 0:54)	23: 0:18 (7: 0:23)
Gebläse			22: 0:16 (4: 0:20)	23: 0:25 (6: 1:17)	22: 0:53 (12: 1:25)
Umluftbetrieb			22: 0:46 (7: 2:04)	23: 0:23 (11: 0:42)	21: 0:23 (13: 0:34)
deaktivieren			22: 0:19 (5: 1:06)	23: 0:09 (4: 0:31)	21: 0:10 (5: 0:28)
Gesamt			1:37 (4:24)	1:16 (2:53)	2:17 (4:04)
Gesamt			10:21 (13:59)	12:07 (16:52)	13:29 (16:56)

Tabelle 7-6: Bearbeitungszeiten erfolgreich bearbeiteter Testaufgaben
Quelle: (Eigene Darstellung)

Die schnellere Bearbeitung durch Gruppe 3 kann mit zwei verschiedenen Ansätzen erklärt werden. Zum einen versuchten die Probanden, die Aufgaben während der Fahrt so schnell wie möglich zu lösen, um sich wieder auf den Verkehr konzentrieren zu können. Zum anderen erhielten die Probanden dieser Gruppe bei verschiedenen Aufgaben als einzige Gruppe ein umfangreiches Feedback auf die Benutzung eines Bedienelements, was zu einer größeren

Gewissheit über den erfolgreichen Abschluss der Aufgabe führte. Beispielsweise reagiert die Anzeige des Audi lane assist nur ab einer Fahrgeschwindigkeit von ca. 65 Stundenkilometern mit einem Farbwechsel von gelb nach grün.

In der dritten Aufgabenkategorie verändert sich das Bild wiederum. Hier sind beide Avatarsystemgruppen teilweise wesentlich schneller in der Bearbeitung als die Handbuchgruppe. Nachforschungen zu diesem Sachverhalt ergaben, dass im Handbuch zur Beschreibung des Umluftbetriebs ein falsches Tastensymbol verwendet wurde, dass im Fahrzeug nicht existiert. Dadurch wurden die Probanden der Handbuchgruppe verwirrt und die Bearbeitung dieser Teilaufgabe zog sich in die Länge.

Die Gesamtbearbeitungszeit aller Aufgaben zeigt ein Bild, das sich teilweise bereits abzeichnet hat. Die Verwendung des Handbuchs hat mit 13:59 Minuten die schnellsten Erfolge herbeigeführt, während die Verwendung des Avatarsystems im Stand durchschnittlich 16:52 Minuten in Anspruch nahm. Eine ähnliche Zeit weist auch die Benutzung des Avatarsystems während der Fahrt mit 16:56 Minuten auf.

7.2.2.3. Zielsicherheit bei der Verwendung von Bedienelementen

Neben der Bearbeitungszeit einer Aufgabe gibt auch die Zielsicherheit, mit der die richtigen Bedienelemente gefunden und bedient werden, Aufschluss darüber, wie gut das Handbuch bzw. das Avatarsystem den Probanden die Verwendung einer bestimmten Funktion beschreiben können. Dazu wurde aus den Videoaufnahmen ermittelt, wie viele der Bedienungen am richtigen, für diese Funktion relevanten Bedienelement erfolgten.

Bei der Auswertung wurden nur erfolgreich gelöste Teilaufgaben analysiert, bei denen das Handbuch oder das Avatarsystem verwendet wurden. Teilaufgaben, die erfolgreich gelöst wurden, ohne ein Bedienelement zu bedienen, wurden ausgelassen. Dieser Zustand konnte erreicht werden, wenn beispielsweise in einer vorangegangenen Teilaufgabe bereits eine spätere Teilaufgabe gelöst wurde und dies dem Probanden bei der Bearbeitung der späteren Teilaufgabe auffiel. Tabelle 7-7 zeigt die prozentualen Anteile der Bedienung der korrekten Bedienelemente an der Gesamtanzahl von Betätigungen von Bedienelementen. Die Angaben vor den Prozentzahlen geben an, wie viele Probanden diese Aufgabe erfolgreich unter Verwendung eines Bedienelements und des Handbuchs bzw. des Avatarsystems gelöst haben. Die Werte in Klammern zeigen an, wie groß der Anteil über alle erfolgreichen Bedienungen auch ohne Handbuch oder Avatarsystem waren. Auch hier wurde aus Gründen der Übersichtlichkeit auf die Darstellung der Standardabweichungen verzichtet.

Die Aufgaben der ersten Kategorie wurden von Gruppe 1 und 2 in etwa gleich sicher bearbeitet, während die Zielsicherheit bei Gruppe 3 um ca. 10% über der der anderen Gruppen lag. Dies ist insbesondere auf die erste der beiden Aufgaben zurückzuführen, bei der in Gruppe 3 die Zielsicherheit 17% bis 19% über den anderen Gruppen lag.

In der zweiten Aufgabenkategorie liegen die drei Gruppen vom Gesamtwert her etwa gleich auf, wobei die Handbuchgruppe den besten Wert aufweisen kann. Während die Avatarsystemgruppen bei den meisten Teilaufgaben eine größere Zielsicherheit aufwiesen,

konnten die Teilnehmer der Gruppe 1 insbesondere bei der letzten Teilaufgabe zur Anzeige des Audi lane assist den Gesamtwert verbessern.

Kategorie	Aufgabe	Teilaufgabe	Gruppe 1	Gruppe 2	Gruppe 3		
1 Bedienschnitt	Spurhalteassistent		21: 73,89 % (72,43 %)	23: 75,02 %	18: 92,9 % (88,61 %)		
	Spurwechsellassistent		22: 97,73 %	22: 94,7 % (94,93 %)	20: 100 % (98,10 %)		
	Gesamt		85,81 % (85,08 %)	84,86 % (84,98 %)	96,45 % (93,36 %)		
2-5 Bedien-schritte	Sitzmemory	aktivieren	22: 95,96 %	20: 99,29 % (96,12 %)	21: 100 % (100 %)		
		Speichern	20: 100 % (100 %)	15: 100 % (100 %)	20: 100 % (100 %)		
		Deaktivieren	8: 100 % (100 %)	12: 100 % (100 %)	15: 100 % (100 %)		
		Gesamt	98,65 % (98,65 %)	99,76 % (98,71 %)	100 % (100 %)		
	Bordcomputer	Einzelfahrtspeicher	21: 85,69 % (85,43 %)	20: 88,07 % (89,63 %)	22: 84,24 %		
		Reichweite	15: 99,49 % (99,65 %)	6: 100 % (100 %)	5: 100 % (100 %)		
		Audi Lane Assist	11: 95,08 % (91,60 %)	9: 84,27 (77,92 %)	11: 87,71 % (88,24 %)		
		Gesamt	93,42 % (92,23 %)	90,78 % (89,18 %)	90,65 (90,83 %)		
	Gesamt		96,04 % (95,59 %)	95,27 % (93,95 %)	95,33 % (95,42 %)		
	6-9 Bedien-schritte	Klimaautomatik	Temperatur	4: 100 % (100 %)	7: 100 % (100 %)	10: 99,00 % (99,52 %)	
			Gebälse	4: 100 % (99,09 %)	6: 100 % (100 %)	12: 100 % (100 %)	
			Umluftbetrieb	2: 58,33 % (84,09 %)	10: 100 % (100 %)	12: 100 % (100 %)	
			deaktivieren	4: 100 % (100 %)	3: 100 % (100 %)	5: 100 % (100 %)	
Gesamt			89,58 % (95,80 %)	100 % (100 %)	99,75 % (99,88 %)		
Gesamt		90,48 % (92,16 %)	93,38 % (92,98 %)	97,18 (96,22 %)			

Tabelle 7-7: Anteil der Verwendung korrekter Bedienelemente
Quelle: (Eigene Darstellung)

In der dritten Aufgabenkategorie weisen die Avatarsystemgruppen eine um mehr als 10% höhere Zielsicherheit auf als die Benutzer des Handbuchs. Bei genauerer Betrachtung stellt sich jedoch auch hier wiederum die Aktivierung des Umluftbetriebs als Hauptursache heraus. Aufgrund des falschen Symbols im Handbuch haben die Probanden ein Bedienelement verwendet, das ein ähnliches Symbol aufwies, jedoch zur Bedienung des Audi drive select führte.

Insgesamt nimmt die Zielsicherheit von Gruppe 1 mit ca. 90 % über Gruppe 2 zu Gruppe 3 hin mit ca. 97 % mit gleich großen Abständen zu. Zudem liegt die Zielsicherheit der Benutzer von Handbuch und Avatarsystem über der durchschnittlichen Zielsicherheit aller Probanden. Die Überlegenheit der Avatarsystemgruppen bei dieser Beobachtung kann dadurch erklärt werden, dass im Avatarsystem entgegen zum Handbuch wesentlich mehr Abbildungen und vor allen Dingen auch Animationen gezeigt wurden, welche den Probanden die Benutzung der Bedienelemente gezeigt hat.

7.2.2.4. Wissen über die Bedienelemente

Das Wissen, das während der Aufgabenbearbeitung erworben wurde, wurde später in zehn Fragen des postaktionalen Fragebogens abgefragt. Dabei konnten einige Aufgaben nur korrekt gelöst werden, wenn über die gestellten Testaufgaben hinaus weitergelesen bzw. weitergefragt wurde. Das in Abbildung 7-8 visuell dargestellte Ergebnis zeigt, dass der Lerneffekt von Gruppe 1 über Gruppe 2 bis hin zu Gruppe 3 um insgesamt eine richtig beantwortete Frage zugenommen hat. Der Vergleich jeweils zweier Gruppen mithilfe von Varianzanalysen weist zwar keine signifikante Varianz auf, jedoch ist der F-Wert der Varianzanalyse von Gruppe 1 und Gruppe 3 mit einem F-Wert von 3,87 nahe dem Schwellenwert für eine Signifikanz mit einem F-Wert 4,3²⁶.

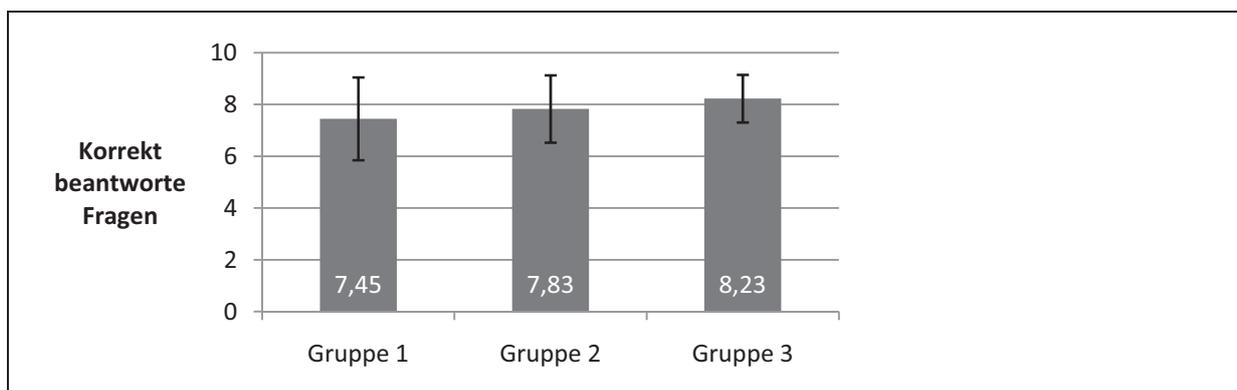


Abbildung 7-8: *Korrekt beantwortete Wissensfragen*

Quelle: (Eigene Darstellung)

Die besseren Werte bei den Avatarsystemgruppen lassen sich wiederum durch die Visualisierung der Verwendung von Bedienelementen erklären, was durch den

²⁶ Der Vergleichswert für den F-Wert liegt bei dem direkten Vergleich zweier Gruppen mit insgesamt über 40 Probanden bei 4,08. Liegt der ermittelte F-Wert darüber, handelt es sich um eine nicht durch auftretende Schwankungen erklärable Abweichung, so dass die Abweichung signifikant ist.

Bildüberlegenheitseffekt die Verinnerlichung von Sachverhalten vereinfacht (Paivio 1979). Dabei ist der Unterschied zwischen den beiden Avatarsystemgruppen durch das direktere und bessere Feedback der Betätigungen in Gruppe 3 während der Fahrt zu deuten.

7.2.3. Vergleich des Handbuchs mit dem Avatarsystem

Neben den Beobachtungen aus den Videoaufzeichnungen erlauben auch die Bewertungen im postaktionalen Fragebogen einen Vergleich der drei Gruppen, um die Eigenheiten der Verwendung des Handbuchs bzw. des Avatarsystems zu ermitteln. Zu diesem Zweck wurden die angegebenen Bewertungen der Aussagen im Fragebogen (vollkommene Ablehnung bis vollkommene Zustimmung), wie bei einer Likert-Skala üblich, in die Zahlen von eins bis fünf überführt, um Berechnungen durchführen zu können. Dabei steht eins für vollkommene Ablehnung und fünf für vollkommene Zustimmung, so dass ein hoher Durchschnittswert die Übereinstimmung mit einer Aussage darstellt. Auf diese Weise wurden Items in den Bereichen der Benutzerfreundlichkeit, des Nutzens und des Gesamteindrucks über die drei Gruppen miteinander verglichen.

7.2.3.1. Benutzerfreundlichkeit

Im Rahmen der Benutzerfreundlichkeit wurden die Probanden befragt, wie angenehm sie die Verwendung des Handbuchs bzw. des Avatarsystems empfunden haben. Dabei erhielten die Teilnehmer je nach Gruppenzuteilung eine Version des postaktionalen Fragebogens, in der jeweils nur nach dem Handbuch oder dem Avatarsystem gefragt wurde. Die grafische Darstellung der durchschnittlichen Antworten nach Gruppe sind in Abbildung 7-9 dargestellt. Die schwarzen Linien zeigen dabei die jeweilige Standardabweichung.

Ein wichtiges Indiz für die Benutzerfreundlichkeit ist, wie sehr es den Probanden Spaß gemacht hat, die entsprechende Hilfe zu verwenden. Dabei lässt sich eine deutliche Zunahme des Spaßfaktors von Gruppe 1 zu Gruppe 3 hin erkennen, wobei gleichzeitig auch die Standardabweichung abnimmt. So weist die Varianzanalyse der Werte von Gruppe 1 und 3 einen F-Wert von 8,67, was auf eine signifikante Abweichung hinweist, die nicht mehr mit gewöhnlichen Schwankungen erklärt werden kann. Eine mögliche Erklärung dieser Verteilung ist der erhöhte Spaß bei der Verwendung von Animationen zur Visualisierung von Inhalten im Allgemeinen (Harrison 1995; Payne/Chesworth/Hill 1992).

Bei der einfachen und intuitiven Benutzung der entsprechenden Hilfe wurde das Avatarsystem minimal besser bewertet. Dabei ist jedoch zu beachten, dass das Avatarsystem von seiner Bedienung her für die meisten Probanden eine fundamental neuartige Benutzerschnittstelle aufwies während die Verwendung eines Buches jedem Probanden bekannt war.

Die Akzeptanz des Avatarsystems, die sich in der ersten Frage abgezeichnet hat, lässt sich auch beim Aspekt der Wiederverwendung erkennen. Die Probanden der Gruppe 3 würden das Avatarsystem am liebsten wiederverwenden, während dieser Wunsch zu Gruppe 1 hin abnimmt. Auch hier ist zu beobachten, dass die Standardabweichung bei der dritten Gruppe deutlich geringer ist als bei den anderen beiden Gruppen.

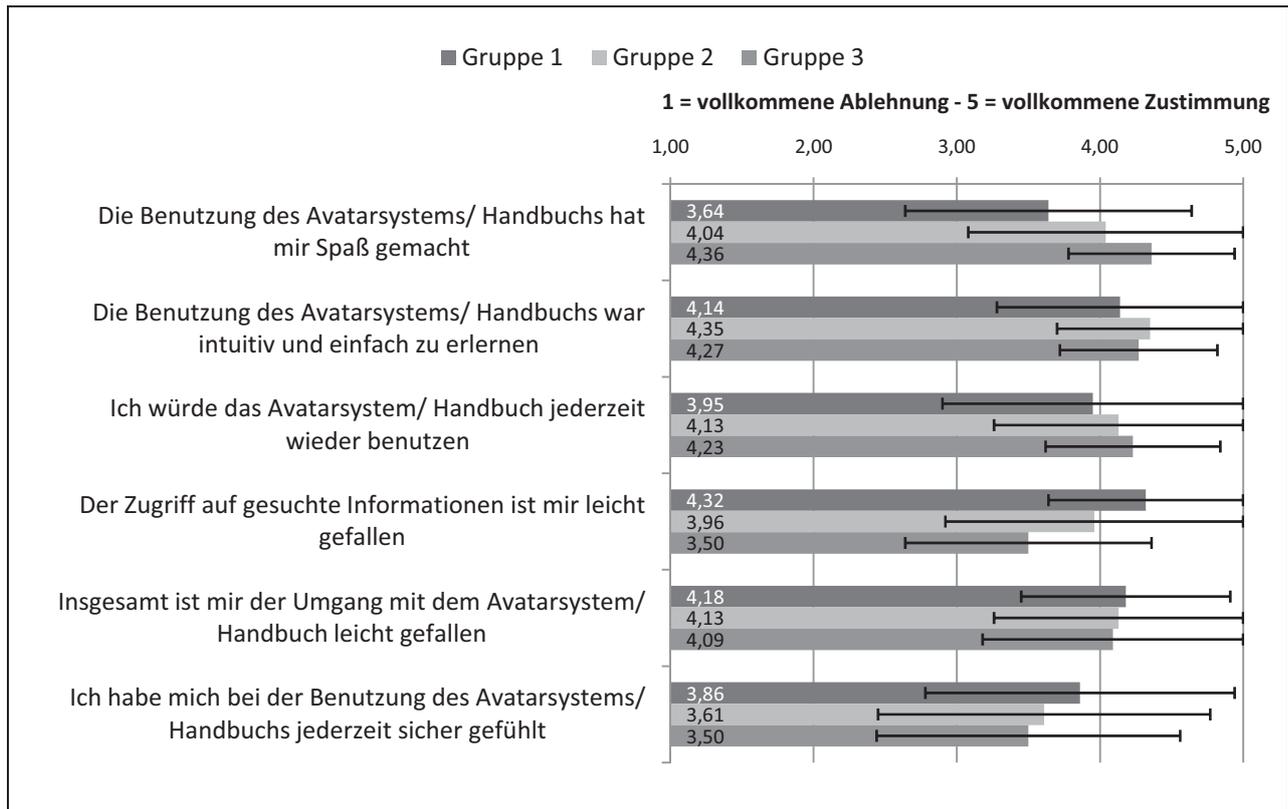


Abbildung 7-9: *Bewertung der Benutzerfreundlichkeit*

Quelle: (Eigene Darstellung)

Ein entgegengesetztes Bild ergibt sich bei der Frage des Informationszugriffs. Übereinstimmend mit der Analyse der Bearbeitungszeiten in Kapitel 7.2.2.2 wurde die Einfachheit des Zugriffs beim Handbuch am besten bewertet, wobei sie zu Gruppe 3 hin abnimmt. Somit zeigt sie deutlicher als die Bearbeitungszeiten die wahrgenommene Zugriffsgeschwindigkeit. Die Varianzanalyse zwischen Gruppe 1 und 3 weist mit einem F-Wert von 11,77 einen signifikanten Unterschied auf.

Eine ähnliche Verteilung ergibt sich auch bei den letzten beiden Fragen, in denen es um die Einfachheit und Sicherheit im Umgang mit der jeweiligen Hilfe ging. In beiden Bewertungen wurde das Handbuch als beste und das Avatarsystem während der Fahrt als schlechteste Hilfe angesehen. Dies kann einerseits darauf zurückzuführen sein, dass die Probanden nicht gewohnt sind mit einem rein sprachbasierten System umzugehen. Andererseits kann dies auch auf Verbesserungsbedarfe in der Benutzerführung hindeuten.

7.2.3.2. Nutzen

Während die Benutzerfreundlichkeit das rein subjektive Empfinden gegenüber dem Handbuch bzw. dem Avatarsystem zum Ausdruck bringt, kann der Nutzen einer Hilfe besser an den damit bearbeiteten Aufgaben gemessen werden. Die Antworten der Probanden auf die entsprechenden Fragen des postaktionalen Fragebogens wurden nach Gruppen getrennt in Abbildung 7-10 visualisiert. Dabei stellen die schwarzen Linien die Standardabweichungen der Durchschnittswerte dar.

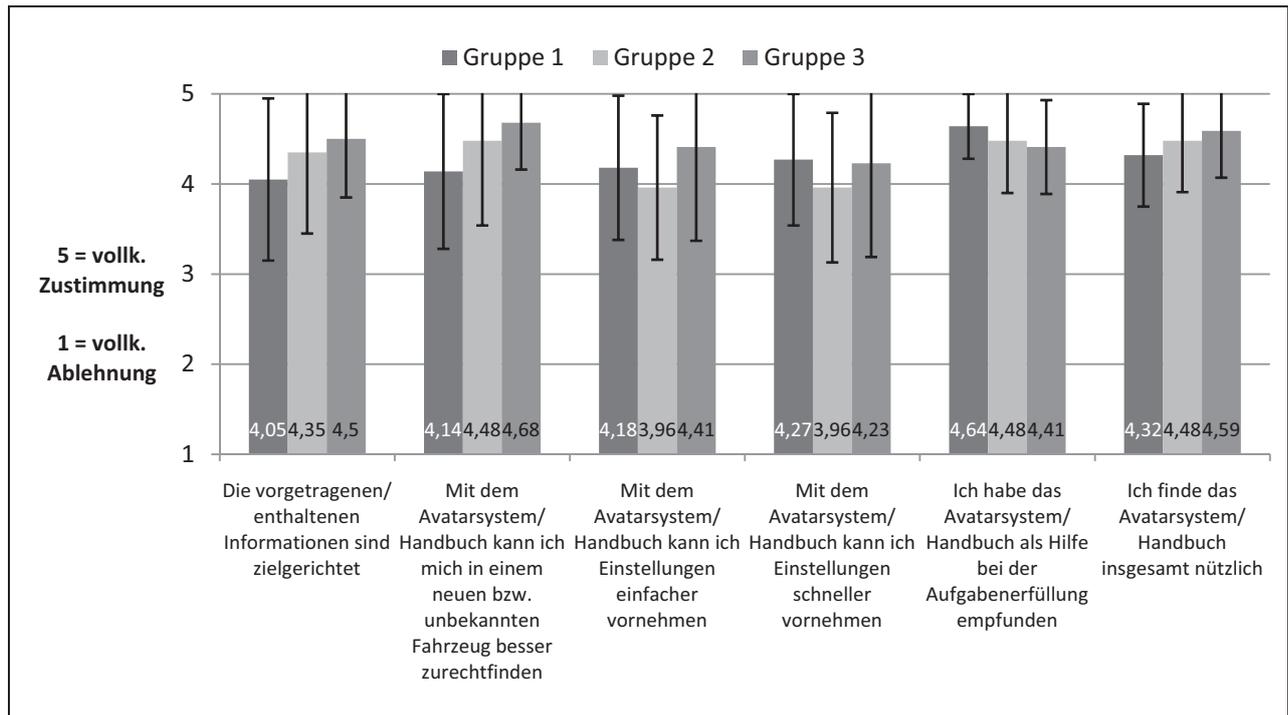


Abbildung 7-10: Bewertung des Nutzens
Quelle: (Eigene Darstellung)

Die vom Avatarsystem vorgetragenen Informationen wurden dabei als zielgerichteter empfunden als die enthaltenen Angaben im Handbuch. Die Bewertung erfolgt von Gruppe 1 zu Gruppe 3 hin zunehmend positiver. Eine ähnliche Verteilung weisen auch die Durchschnittswerte der Antworten zur Bewertung der Hilfen zur Verwendung eines neuen bzw. unbekanntem Fahrzeugs auf. Die Bewertung von Gruppe 1 zu Gruppe 3 hin wurde um einen halben Punkt besser und weist somit einen signifikanten Unterschied zwischen Gruppe 1 und 3 mit einem F-Wert von 5,42 auf.

Ein gemischtes Bild bei der Bewertung ergab sich bei der Frage, ob das Vornehmen bestimmter Einstellungen einfacher bzw. schneller erfolgen kann. Aufgrund der ähnlichen Verteilung bei diesen beiden Fragen, liegt der Schluss nahe, dass von Probanden kaum zwischen den beiden Dimensionen „einfach“ und „schnell“ unterschieden wurde. In beiden Fällen wurde das Avatarsystem von Gruppe 2 insgesamt am schlechtesten bewertet, gefolgt vom Handbuch. Am besten abgeschnitten hat in beiden Fragen das Avatarsystem während der Fahrt.

Widersprüchliche Angaben machten die Probanden bei den Fragen, ob sie das Handbuch bzw. das Avatarsystem als Hilfe empfunden haben und ob sie die entsprechende Hilfe als nützlich ansehen. Während das Handbuch von den meisten Probanden als Hilfe bewertet wurde, wurde es dennoch als am wenigsten nützlich angegeben. Das Gegenteil ist beim Avatarsystem während der Fahrt zu erkennen, das am wenigsten hilfreich, aber am nützlichsten angesehen wurde. Diese paradoxe Bewertung lässt sich gegebenenfalls dadurch erklären, dass die Probanden bei der Frage nach der Hilfestellung das Handbuch bzw. das Avatarsystem im konkreten Bezug zu den gestellten Aufgaben bewertet haben, während die Frage nach der

Nützlichkeit allgemein auf den Typ von Hilfe bezogen wurde. Dies bedeutet, dass das Avatarsystem während der Fahrt am wenigsten hilfreich bei der Aufgabenbearbeitung war, jedoch das Konzept des Systems als nützlich und ausbaufähig betrachtet wurde.

7.2.3.3. Gesamteindruck

Der letzte Vergleich aller drei Gruppen erfolgt bei der Gesamtbewertung der jeweiligen Hilfe, wobei die Probanden gebeten wurden, eine Schulnoten zu vergeben. Da alle Probanden zumindest bei einer Teilaufgabe das Handbuch oder das Avatarsystem als Hilfe herangezogen haben, wurden alle Probanden bei der Berechnung der Durchschnittsnote berücksichtigt.

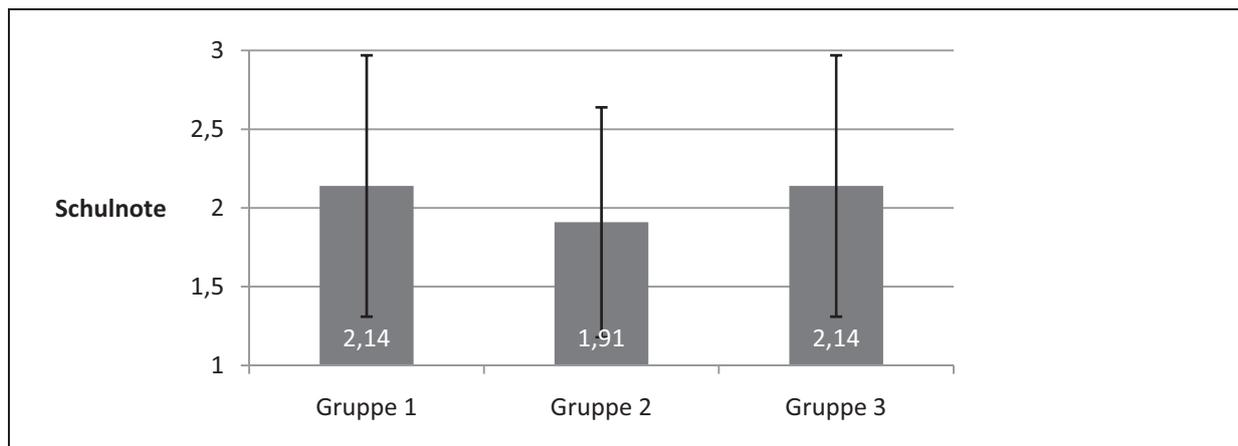


Abbildung 7-11: *Bewertung der verwendeten Hilfe mit Schulnoten*
Quelle: (Eigene Darstellung)

Die Benotungen durch Gruppe 1 und Gruppe 3 ergaben den exakt gleichen Wert von 2,14 während Gruppe 2 ihre Hilfe am besten mit 1,91 benotete (vgl. Abbildung 7-11). Dies kann derart interpretiert werden, dass das Avatarsystem mindestens als ebenbürtig zum Handbuch angesehen werden kann. Der Unterschied zwischen den beiden Avatarsystemgruppen kann insbesondere durch die Qualität der Spracherkennung im Stand bzw. während der Fahrt erklärt werden (vgl. dazu Kapitel 7.2.4.4). Die Anwendung der Varianzanalyse auf die Bewertung des verwendeten Hilfesystems durch Schulnoten ergab jedoch keine Signifikanz der Abweichungen.

7.2.4. Bewertung des Avatarsystems mit geschlossenen Fragen

Neben dem Vergleich der beiden Hilfen Handbuch und Avatarsystem wurde der postaktionale Fragebogen bei den Gruppen 2 und 3 auch dazu genutzt, formative Bewertungen einzelnen Komponenten des Avatarsystems zu erlangen. Da es sich bei der Bewertung um einen Black-Box-Test (vgl. Balzert 2001, 441f) handelt, konnten dabei nur Komponenten und Sachverhalte abgefragt werden, welche für die Probanden wahrnehmbar waren. Deshalb wurden Fragen zum Avatar selbst, zum Inhalt und dessen Visualisierung, zur Benutzeroberfläche, der Spracherkennung, der Sprachausgabe und zur Dialogsteuerung gestellt. Zu jedem Bereich wurden Aussagen formuliert, welche die Probanden ablehnen oder ihnen zustimmen konnten. Die dabei verwendete fünfstufige Likert-Skala wurde in die Zahlen von eins bis fünf überführt, so dass Berechnungen angestellt werden konnten. Neben den

technischen Komponenten wurde zudem der Gesamteindruck abgefragt. Die daraus gewonnen Ergebnisse werden im Folgenden dargestellt und können zur Gestaltung von Verbesserungen herangezogen werden.

7.2.4.1. Avatar

Der Avatar als Personifizierung des Fahrzeugs ist die augenscheinlichste Neuerung des Hilfesystems und hat somit besondere Aufmerksamkeit erhalten. Trotz der eingeschränkten Gestaltungsmöglichkeiten des Avatars in diesem Prototyp gilt es, Erkenntnisse für die weitere Gestaltung zu sammeln. Die ermittelten Durchschnittswerte und ihre Standardabweichungen sind in Abbildung 7-12 grafisch dargestellt. Im Gegensatz zur bisherigen vergleichenden Analyse zwischen den drei Gruppen spielen im Folgenden aufgrund des formativen Charakters der Ergebnisse auch die absoluten Werte eine Rolle, um die Stärke des Eindrucks der Probandengruppen zu bestimmen und die Dringlichkeit von Verbesserungen zu ermitteln.

Das Aussehen des Avatars wurde in beiden Gruppen mit 3,78 bzw. 4,27 als tendenziell angenehm angesehen, wobei die fahrende Gruppe das Aussehen in dieser Hinsicht mit einem F-Wert von 4,46 signifikant besser bewertet. Das Auftreten des Avatars wurde zudem mit 3,57 und 3,86 als tendenziell menschlich angesehen, wobei diese Eigenschaft somit nicht stark wahrgenommen wurde. Der Avatar wurde mit 2,39 und 1,86 insgesamt kaum als aufdringlich empfunden, wobei die fahrende Gruppe den Avatar (einen halben Punkt) weniger aufdringlich bewertete. Dies kann entweder damit zusammenhängen, dass die Fahrer zu sehr durch den Straßenverkehr abgelenkt waren, um alle Erklärungen des Avatarsystems wahrzunehmen oder sie haben die Erklärungen bzw. Bestätigungen des Avatars als hilfreicher in der jeweiligen Situation wahrgenommen.

Die grafische Darstellung wurde qualitativ von beiden Gruppen als gut bewertet, wobei die Bewertung in Gruppe 2 zwar besser ausfiel, jedoch mit einer Standardabweichung von 0,99 auch weiter streute. Dies kann daran liegen, dass Fahrer ihren Blick während der Fahrt nicht so lange auf den Avatar gerichtet haben wie die Probanden der Gruppe 2. Manche Probanden haben im Anschluss an die Evaluation verbal die Qualität der Darstellung bemängelt, die jedoch auf die geringe Auflösung des eingebauten Bildschirms zurückzuführen ist. Die Bewegungen des Avatars wurden von beiden Gruppen insgesamt als passend und flüssig bewertet.

Dementgegen wurde die Lippsynchronität von weniger Probanden wahrgenommen und dementsprechend schlechter bewertet. Diese Beobachtung war bei Gruppe 2 aus dem oben bereits dargestellten Grund stärker ausgeprägt. Tatsächlich konnte aufgrund der eingeschränkten Rechenleistung des verwendeten CarPC die Lippsynchronität nicht immer gewährleistet werden. Insbesondere bei den Antworten dieser Frage sind mit 1,28 und 0,99 große Standardabweichungen zu verzeichnen.

Die Kleidung des Avatars wurde von beiden Gruppen als gut bewertet, wobei die Bewertung der fahrenden Gruppe mit einem halben Punkt positiver ausfiel. Recht nah beieinander und mit 4,61 bzw. 4,77 deutlich positiv wurde der gezeigte Teil des Avatars (Oberkörper) wahrgenommen. Dies lässt vermuten, dass bei der gegebenen Bildschirmgröße die

Darstellung einer vollständigen anthropomorphen Gestalt nicht von Vorteil wäre. Insgesamt wurde das Gesamterscheinungsbild des Avatars als sehr positiv wahrgenommen, so dass sich bis auf die Verbesserung der Rechnerleistung zur Gewährleistung der Lippensynchronität und kleineren Anpassungen an der Zuordnung von Animationen zu Inhalten aus diesen Zahlen keine größeren Verbesserungsbedarfe zu ergeben scheinen.

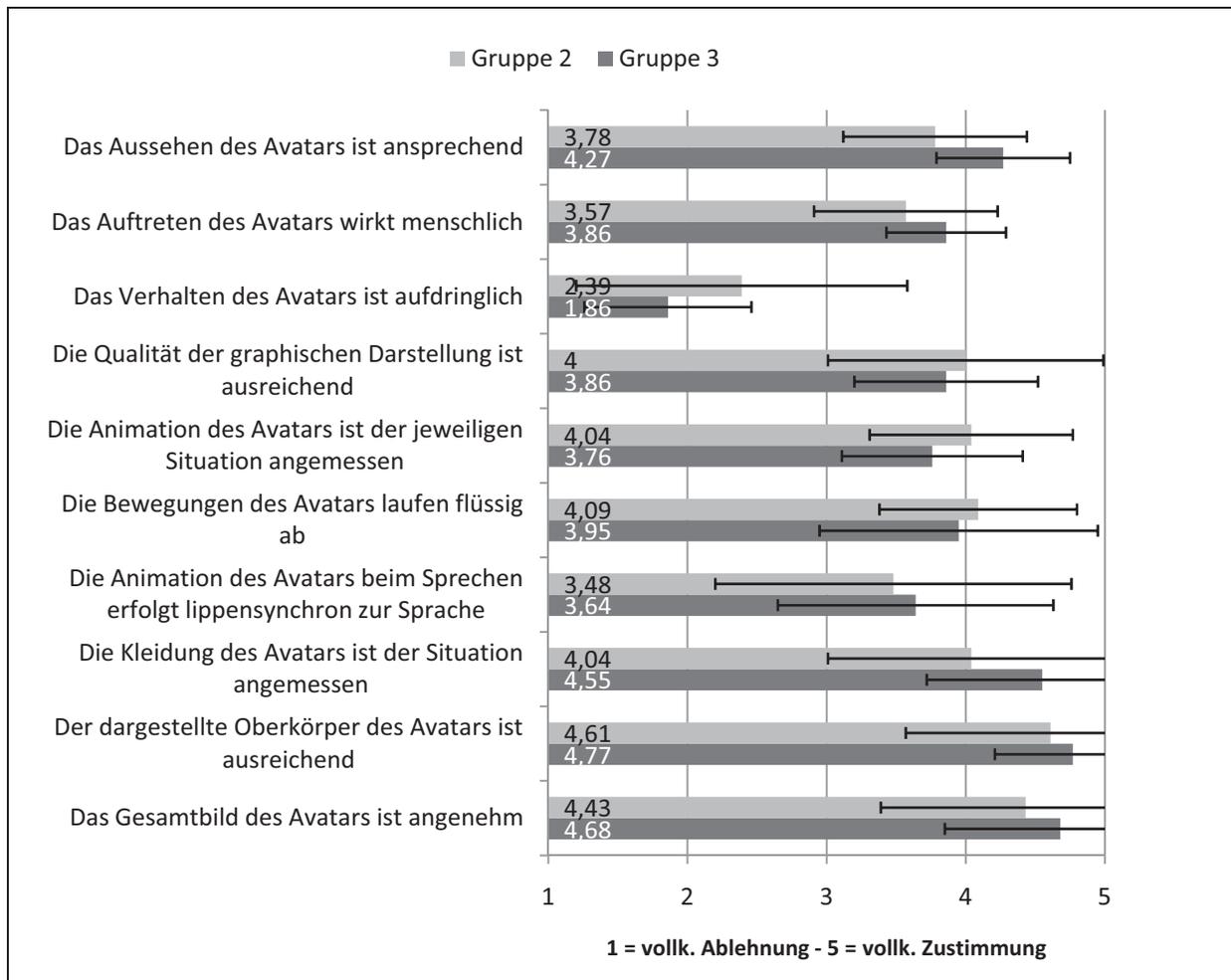


Abbildung 7-12: *Bewertung des Avatars*
Quelle: (Eigene Darstellung)

7.2.4.2. Inhalt und Visualisierung

Neben dem Avatar und der Benutzeroberfläche liegt das Augenmerk des Benutzers insbesondere auf dem Inhalt, der in Form von Bildern oder Animationen bzw. Videos dargestellt wird. Die meisten Antworten des Avatarsystems, die sich mit der Betätigung und der Funktion von Bedienelementen befassen, wurden durch ein entsprechendes Medium veranschaulicht. Die technisch mögliche Einbindung von 3D-Objekten war auf dem verwendeten CarPC aufgrund der suboptimalen Leistung im Grafikbereich während der Evaluation nicht möglich. Die Bilder und Videos wurden dabei entweder aus der digitalen Version des Benutzerhandbuchs oder dem virtuellen Cockpit übernommen und gegebenenfalls zu Animationen zusammengestellt. Die Fragen zum Inhalt und dessen

Visualisierung sowie die durchschnittlichen Bewertungen der Aussagen mit ihren Standardabweichungen sind in Abbildung 7-13 grafisch aufbereitet.

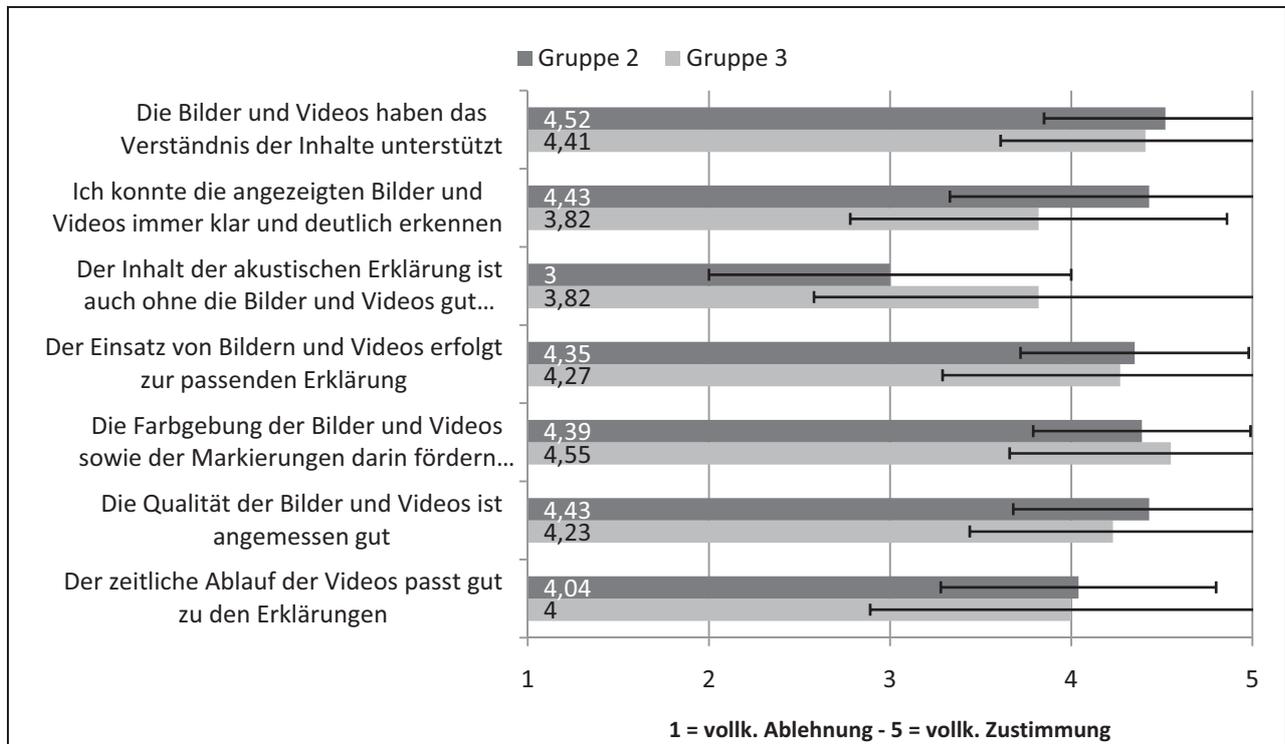


Abbildung 7-13: *Bewertung der Visualisierung*
Quelle: (Eigene Darstellung)

Die Probanden der beiden Gruppen waren sich darin einig, dass die Darstellung von Bildern und Videos das Verständnis der vermittelten Inhalte mit Mittelwerten von 4,52 und 4,41 gut bis sehr gut unterstützt hat. Durch die Darstellung von visuellen Informationen war es den Probanden u.a. besser möglich, die beschriebenen Bedienelemente zu finden. Insbesondere Teilnehmer der Gruppe 3 hatten Probleme, die visuellen Inhalte jederzeit klar und deutlich auszumachen, während die stehenden Probanden die Klarheit und Deutlichkeit besser bewerteten. Obwohl beide Gruppen diese Eigenschaften tendenziell positiv bewerteten, deutet dies auf Verbesserungspotenzial für die Mediendarstellung während der Fahrt hin.

Der Aussage, dass die akustischen Erklärungen auch ohne visuelle Darstellung gut verständlich seien, standen die Probanden der Gruppe 2 vollkommen neutral gegenüber, während Gruppe 3 dies tendenziell bejahte. Auch wenn diese unterschiedliche Bewertung mit einem F-Wert von 5,86 als signifikant einzustufen ist, können aussagekräftige Gestaltungsempfehlungen nur in ausdifferenzierten Experimenten mit und ohne akustische bzw. visuelle Informationen gewonnen werden.

Beide Gruppen waren der Meinung, dass die Bilder und Videos passend eingesetzt wurden und vom zeitlichen Ablauf her gut auf die Erklärungen abgestimmt waren. Aufgrund der großen Streuung beim Letztgenannten lässt sich Verbesserungspotenzial erkennen bei der Abstimmung von akustischer und visueller Erklärung, insbesondere Videos.

Die Farbgebung innerhalb der Bilder wurde von beiden Gruppen positiv bewertet, wobei Gruppe 3 eine minimal bessere Bewertung vornahm. Die Qualität der visuellen Medien wurde ebenfalls von beiden Gruppen mit 4,23 bzw. 4,43 als tendenziell sehr gut betrachtet. Wie auch später in Kapitel 7.2.5 noch ausgeführt wird, wurde die Visualisierung als eine der besten Komponenten des Avatarsystems bewertet und weist somit nur wenige Verbesserungsbedarfe auf.

7.2.4.3. Benutzeroberfläche

Die Benutzeroberfläche repräsentiert zum einen die technische Komponente des GUI-Frameworks, welche die Darstellung von Avatar und Medien zusammenführt, und zum anderen die Informationen über die Möglichkeiten zur Dialogsteuerung für den Benutzer. Die Ergebnisse der Bewertung bezüglich der Benutzeroberfläche des Avatarsystems sowie die daraus errechneten Standardabweichungen sind in Abbildung 7-14 visuell dargestellt.

Die Größe des Displays wurde von beiden Gruppen als ausreichend bzw. vollkommen ausreichend bewertet, wobei die fahrende Gruppe 3 die Größe als weniger angemessen sah und eine weite Streuung bei der Antwort aufwies. Dies kann damit zusammenhängen, dass während der Fahrt nur kurze Blicke auf den Bildschirm möglich waren und bei der geringen Größe gegebenenfalls Details nicht optimal wahrgenommen werden konnten.

Die Anzeige von Symbolen in der Benutzeroberfläche wurde positiv aufgenommen, wobei Gruppe 3 diese Einrichtung besser bewertete. Ähnlich gut bewertet, jedoch mit einer geringeren Streuung, wurde die Abstimmung zwischen Avatar und den gezeigten Medien angesehen. Die Farbgebung der Benutzeroberfläche wurde ebenfalls von beiden Gruppen als angenehm bis sehr angenehm betrachtet.

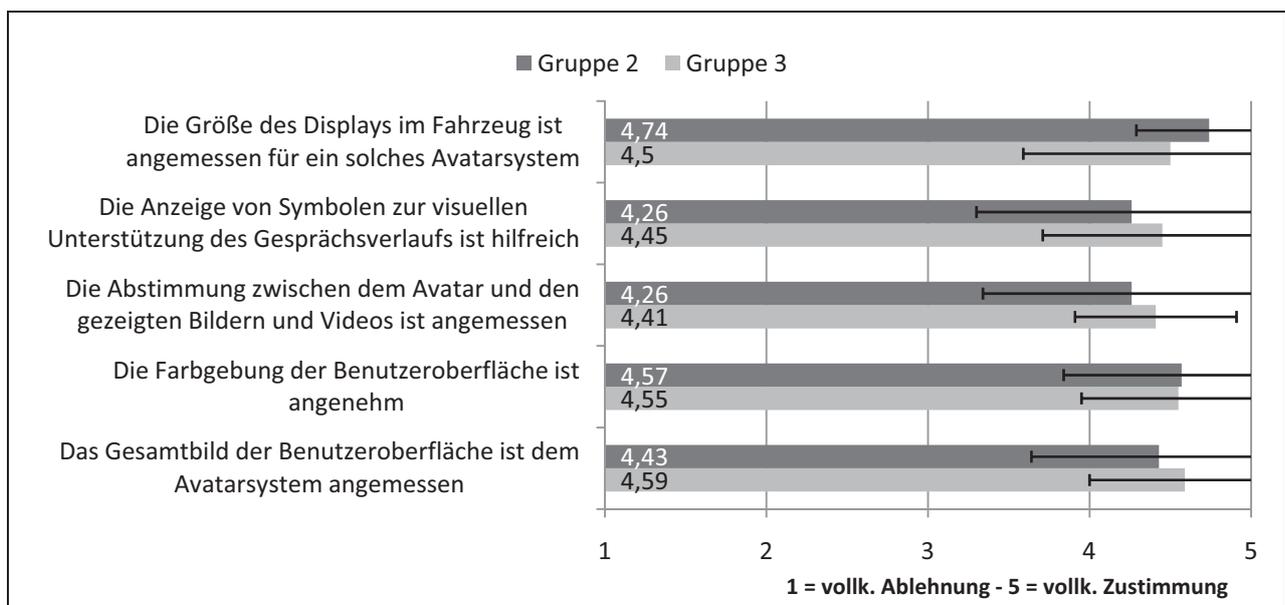


Abbildung 7-14: Bewertung der Benutzeroberfläche
Quelle: (Eigene Darstellung)

Der Gesamteindruck, den die Benutzeroberfläche gemacht hat, wurde als dem Avatarsystem angemessen bewertet, wobei Gruppe 3 eine minimal bessere Bewertung abgab. Als Verbesserungspotenzial sind somit die am schlechtesten bewerteten Punkte der Symbole und der Abstimmung zwischen Avatar und Medien anzusehen.

Die Auswertung des Videos ergab eine sehr rege Nutzung der durch die angezeigten Symbole signalisierten Befehlsoptionen der Benutzeroberfläche (vgl. Abbildung 7-15). So wurden insgesamt die Befehle „Weiter“ und „Stopp“ am häufigsten und der Aufruf von Hinweisen gar nicht verwendet, obwohl auf alle Symbole und die dazugehörigen Befehle im Tutorial des Avatarsystems hingewiesen wurde. Hinsichtlich der Verwendung in den beiden Gruppen ergaben sich deutliche Unterschiede. Während in der fahrenden Gruppe der Weiter-Befehl häufiger verwendet wurde, erfolgte die Benutzung des Stopp-Befehls in der stehenden Gruppe 2 deutlich häufiger als in der fahrenden Gruppe 3.

Die gehäufte Verwendung des Weiter-Befehls während der Fahrt kann darauf zurückgehen, dass Probanden weniger Zeit für die Erklärungen des Avatarsystems verwenden konnten und deshalb schneller zu einem bestimmten Teilaspekt einer Erklärung voranschreiten wollten. Damit verbunden sind die fahrenden Teilnehmer schneller an das Ende einer Erklärungssequenz gesprungen als es die Probanden in der stehenden Gruppe taten. Dadurch entfiel in vielen Fällen die Stopp-Anweisung, da kein automatisches Weiterlaufen der Erklärungsschritte am Ende einer Erklärungssequenz erfolgt.

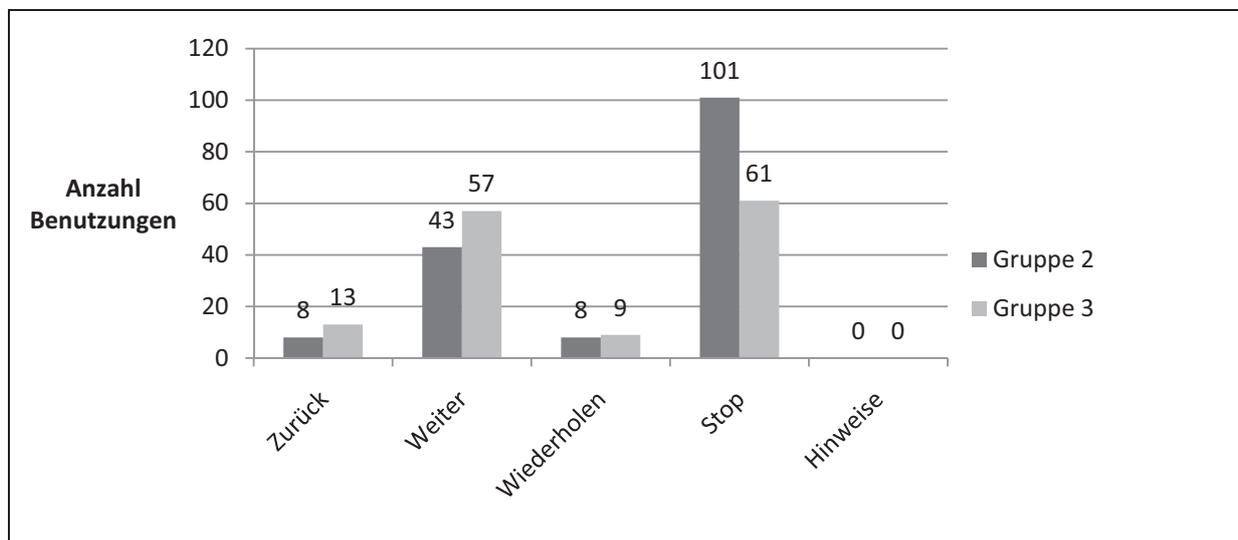


Abbildung 7-15: *Verwendete Befehle der Benutzeroberfläche*
Quelle: (Eigene Darstellung)

7.2.4.4. Spracherkennung

Der Spracherkennung des Avatarsystems kommt eine zentrale Rolle zu, da sie der einzige Eingabekanal des Systems ist, der immer aktiv ist. Der Benutzer ist somit darauf angewiesen, dass die Spracherkennung korrekt funktioniert und seine Eingaben erkennt, damit er mit dem Avatarsystem interagieren kann. Die Bewertung der Spracherkennung erfolgt einerseits auf

Basis der Angaben der Probanden in den postaktionalen Fragebögen und andererseits mithilfe der Analyse der aufgenommenen Videos.

Bewertung der Spracherkennung auf Basis der postaktionalen Fragebögen

Die Durchschnittsbewertungen aus den Fragebögen sind in Abbildung 7-16 grafisch aufgearbeitet. Die Teilnehmer beider Avatarsystemgruppen hatten tendenziell den Eindruck, dass die Spracherkennung sie nicht immer versteht. Während der Mittelwert der stehenden Gruppe im neutralen Bereich angesiedelt ist, bewerten die Fahrer diesen Aspekt negativer.

Hinsichtlich der Fehlerrobustheit hatten stehende Probanden eher den Eindruck, dass die Spracherkennung Pausen verzeiht, während Probanden der Gruppe 3 eher den Eindruck hatten, dass falsch ausgesprochene Wörter korrigiert werden. Diese Bewertungen können auf die Selbsteinschätzung der Probanden zurückgehen. Stehende Teilnehmer konnten sich tendenziell mehr Zeit nehmen, eine Spracheingabe zu formulieren, wodurch es unweigerlich zu Pausen im Sprechfluss kam. Probanden, die während des Fahrens eine Eingabe vornahmen, mussten sich auf den Verkehr konzentrieren und konnten somit nicht so viel Aufwand in die Formulierung von Spracheingaben stecken, wodurch es zu falsch ausgesprochenen Wörtern kam. Darüber hinaus sind die absoluten Werte der Antworten beider Fragen sehr niedrig, was zum einen damit erklärt werden kann, dass die Fehlerrobustheit insgesamt als schlecht angesehen wurde. Andererseits kann es auch ein Indiz dafür sein, dass die Probanden der Ansicht waren, nicht viele Pausen gemacht oder Wörter falsch ausgesprochen zu haben.

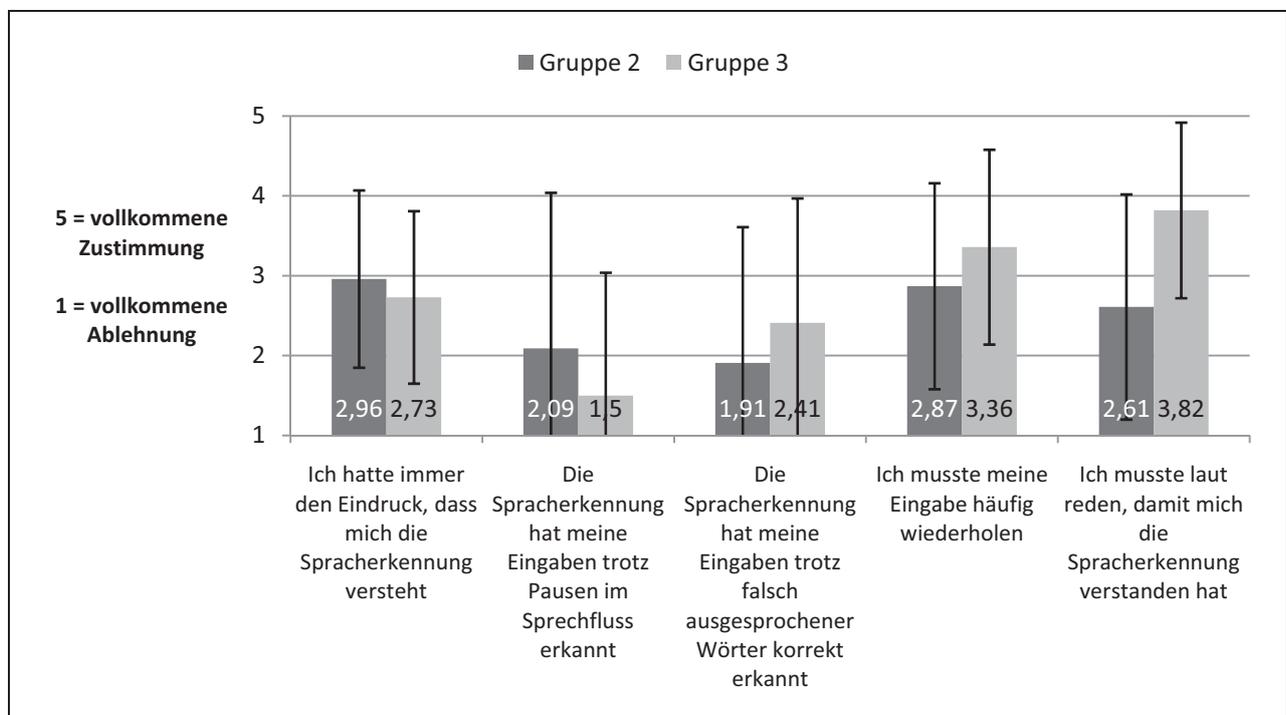


Abbildung 7-16: *Bewertung der Spracherkennung auf Basis der postaktionalen Fragebögen*
Quelle: (Eigene Darstellung)

Die Anpassung des Probandenverhaltens an Probleme mit der Spracherkennung erfolgt insbesondere auf zwei Arten: Der Wiederholung von Eingaben und der Erhöhung der Sprechlautstärke. So gaben insbesondere Teilnehmer der Gruppe 3 an, dass sie tendenziell häufig die Eingaben wiederholen mussten, während Probanden im stehenden Fahrzeug dies tendenziell verneinten. Ein ähnliches Bild ergab sich auch bei der Anpassung der Lautstärke, die während der Fahrt häufiger notwendig war, während sie im Stand tendenziell nicht vorgenommen wurde. Die Bewertung der Lautstärkenanpassung wurde dabei signifikant unterschiedlich bewertet, was der F-Wert von 10,29 deutlich zeigt. Die Anpassung der eigenen Spracheingaben ist insbesondere während der Fahrt notwendig, da die Fahrgeräusche eine Geräuschkulisse erzeugen, welche die Spracherkennung erschwert (vgl. Lach 2005). Die eindringenden Umfeldgeräusche im Versuchsfahrzeug waren aufgrund einer guten Schalldämmung jedoch eher gering. Den Vorschlag für eine Analysemethode zur Bestimmung der Spracherkennungsqualität abhängig von der Geschwindigkeit, dem Umfeld und dem Straßenbelag machen Westphal und Waibel (1999). Die Methode ist dabei aufgrund des einheitlichen Umfelds und Straßenbelags der verwendeten Teststrecke nicht zum Einsatz gekommen.

Bewertung der Spracherkennung auf Basis der Videoanalyse

Neben der subjektiven Bewertung der Spracherkennung durch die Probanden können aus dem aufgezeichneten Videomaterial objektive Daten extrahiert werden, die in Tabelle 7-8 aufgeführt sind. Dabei wurden zum einen die durchschnittliche Anzahl an Spracheingaben bei der Bearbeitung einer Aufgabe, die Quote vom Avatarsystem korrekt erkannter Spracheingaben und der Anteil an Kommando-Eingaben an den Spracheingaben erfasst. Unter einer Spracheingabe wird dabei jede beabsichtigte, verbale Äußerung in Richtung des Avatarsystems verstanden. Bei der Erkennung von Spracheingaben werden korrekt und falsch erkannte Eingaben unterschieden. Eine korrekt erkannte Spracheingabe ist eine Eingabe, auf die eine der beabsichtigten Antworten folgt. Eingaben, auf die per Zufall und nicht auf Basis der tatsächlich gesprochenen Wörter eine passende Antwort ausgegeben wird, werden als falsch erkannt gewertet.

Unter einem Kommando wird hier eine kommandoartige Eingabe verstanden wie z.B. „Klimaanlage ausschalten“ anstelle von „Wie kann ich die Klimaanlage ausschalten“. Nicht als Kommandos, aber dennoch als Spracheingaben gewertet, werden die Eingaben zur Steuerung des Dialogflusses wie „Weiter“ oder „Stopp“. Bei der Darstellung in der Tabelle wurden aus Gründen der Übersicht keine Teilaufgaben, sondern nur Aufgaben aufgeführt. Die Zahlen vor dem Mittelwert von Eingaben geben im Falle von Teilaufgaben die Anzahl der Probanden an, die das Avatarsystem zur Aufgabenbearbeitung erfolgreich genutzt haben und dementsprechend bei der Auswertung berücksichtigt wurden. Bei Summen oder aggregierten Betrachtungen über eine ganze Aufgabe entfällt die Angabe der Zahlen, da sie nicht einer Teilaufgabe eindeutig zugeordnet werden können.

Die Quote korrekt erkannter Spracheingaben ist bei der Gruppe im stehenden Fahrzeug in allen beobachteten (Teil-)Aufgaben höher ausgefallen als im fahrenden Fahrzeug, was mit den Fahrgeräuschen erklärt werden kann (vgl. dazu auch Akyol/Libuda/Kraiss 2001, 148ff). Der Verlauf der Quoten zeigt ein klares Muster an, zumal es in beiden Gruppen von der

Aufgabenkategorie 1 zu 2 abnimmt, wobei die Abnahme bei Gruppe 3 minimal ist. Die Erkennungsquote erhöht sich schließlich wieder im dritten Aufgabenkomplex auf ca. 76% bzw. ca. 71%. Die Gesamtrate weist bei Gruppe 2 mit knapp 76% eine fast 10% höhere Erkennungsrate als bei Gruppe 3 im fahrenden Fahrzeug auf.

Kategorie	Aufgabe	Wert	Gruppe 2	Gruppe 3
1 Bedienschritt	Spurhalteassistent	Eingaben	23: 4,10	19: 4,32
		Fehlerfrei	70,68 %	66,72 %
		Kommando	8,39 %	19,16 %
	Spurwechselassistent	Eingaben	22: 2,18	19: 2,26
		Fehlerfrei	92,80 %	61,80 %
		Kommando	24,13 %	27,51 %
	Gesamt	Eingaben	3,14	3,80
		Fehlerfrei	81,74 %	64,26 %
		Kommando	16,26 %	23,34 %
2-5 Bedienschritte	Sitzmemory	Eingaben	3,27	3,38
		Fehlerfrei	71,10 %	61,77 %
		Kommando	16,85 %	16,73 %
	Bordcomputer	Eingaben	4,79	3,72
		Fehlerfrei	66,34 %	65,72 %
		Kommando	21,45 %	28,96 %
	Gesamt	Eingaben	4,03	3,55
		Fehlerfrei	68,72 %	63,75 %
		Kommando	19,82 %	22,85 %
6-9 Bedienschritte	Klimaautomatik	Eingaben	2,07	2,56
		Fehlerfrei	76,21 %	70,75 %
		Kommando	22,56 %	43,04 %
Gesamt	Eingaben	3,08	3,3	
	Fehlerfrei	75,56 %	66,25 %	
	Kommando	19,55 %	29,74 %	

Tabelle 7-8: *Anzahl an Spracheingaben, Erkennungsraten und Kommandoanteile*
Quelle: (Eigene Darstellung)

Betrachtet man die Entwicklung der durchschnittlichen Spracheingaben, so lässt sich in beiden Gruppen eine kontinuierliche Abnahme der Anzahl der Eingaben mit zunehmender Anzahl an Aufgaben verzeichnen. Parallel dazu nimmt der Anteil an Kommandos mit zunehmender Anzahl der Aufgaben zu. Dies lässt sich zum einen als Indiz dafür ansehen, dass die Probanden mit zunehmender Erfahrung im Umgang mit Avatarsystem erkannt haben, auf welche Art von Spracheingaben das System besonders gut anspricht und sie damit die Anzahl an Eingaben reduzieren können. Zum anderen ist dieser Erklärungsversuch gepaart mit der Zunahme an Kommandos auch ein Indiz dafür, dass sich die Probanden bei der Verwendung von Sprachkommandos bessere Erfolgchancen ausrechnen, korrekt verstanden zu werden. Dies lässt den natürlichsprachlichen Ansatz des Avatarsystems als suboptimal erscheinen.

7.2.4.5. Sprachausgabe

Die Sprachausgabe ist insbesondere für die Probanden der fahrenden Gruppe 3 das zentrale Ausgabemedium gewesen, das auch in jeder Fahrsituation aufgenommen werden konnte. Schröter (2005) schlägt zur Bewertung der Sprachausgabe die drei Dimensionen der Genauigkeit bei der Synthese von problematischen Wörtern wie Fremdwörtern und Eigennamen, der Verständlichkeit und der Natürlichkeit der ausgegebenen Sprache vor. Diese drei Dimensionen sind in die Gestaltung der Fragen zu dieser Komponente eingeflossen. Die durchschnittliche Bewertung der verschiedenen Aspekte der Sprachausgabe mit ihren Standardabweichungen ist in Abbildung 7-17 dargestellt.

Die Geschwindigkeit der Sprachausgabe wurde von beiden Gruppen als angemessen bewertet, wobei die fahrenden Probanden das Tempo besser bewerteten. Eine insgesamt bessere Bewertung erhielt die Lautstärke, wobei diese während der Fahrt weniger gut angesehen wurde. Die Unterschiede in beiden Fragen sind dabei aufgrund von F-Werten von 4,68 bzw. 6,34 als signifikant einzustufen. Die unterschiedliche Bewertung kann damit zusammenhängen, dass die Lautstärke der Sprachausgabe unabhängig von der Geschwindigkeit gleich blieb, während die Fahrgeräusche mit zunehmender Geschwindigkeit zunahmen.

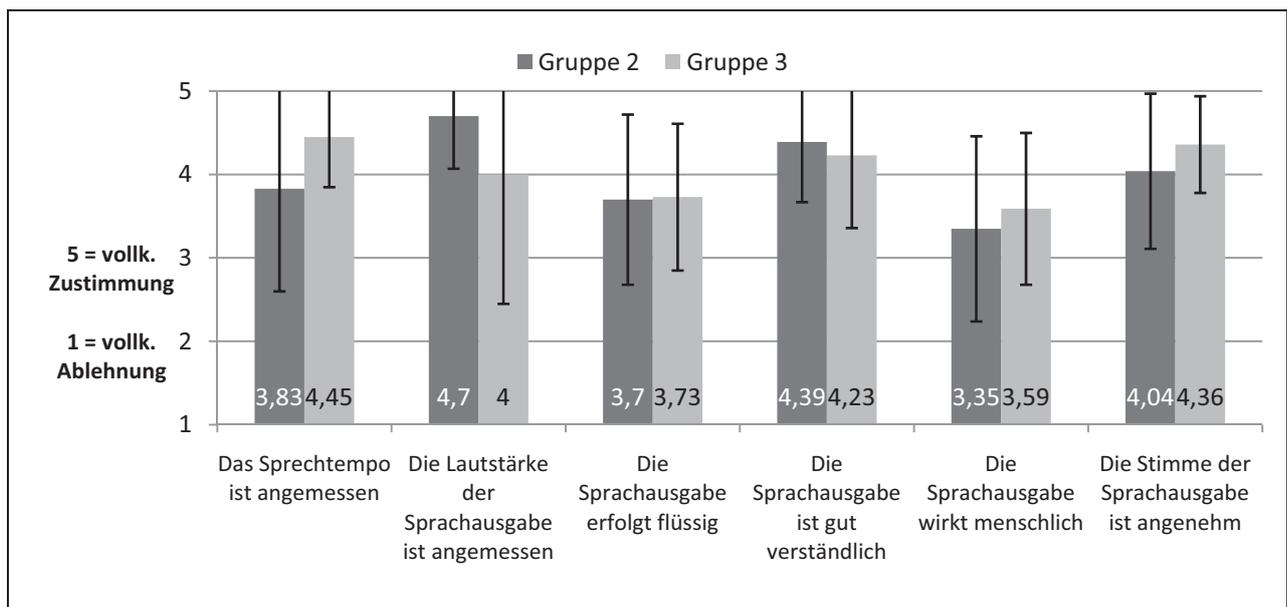


Abbildung 7-17: Bewertung der Sprachausgabe
Quelle: (Eigene Darstellung)

Der Aussage, dass die Sprachausgabe flüssig erfolgt ist, haben zwar beide Gruppen tendenziell zugestimmt, jedoch deuten die Werte zusammen mit der Bewertung der Menschlichkeit der Stimme auf eine abgehackte Sprachsynthese hin, die eine menschliche Satzmelodie vermissen lässt. Dies kann damit erklärt werden, dass aufgrund der Menge möglicher Antworten keine manuelle Sprachoptimierung in phonetischer oder prosodischer Richtung erfolgt ist und dadurch die reine Sprachsynthese zu hören war. Trotz der verhaltenen Kritik an diesen beiden Aspekten wurde die Verständlichkeit der Sprachausgabe dennoch deutlich als gut bewertet.

Die in der Sprachsynthese verwendete Stimme wurde von beiden Gruppen ebenfalls deutlich als angenehm bewertet, wobei Gruppe 3 dieser Aussage stärker zustimmte. Dies kann daran liegen, dass der monotone Vortrag der Informationen durch die Sprachausgabe zumindest neutral, vom Fahrer aufgenommen wurde im Gegensatz zu einer realen Stimme, die mit zunehmender Lautstärke während der Fahrt auch den Klang verändert hätte.

7.2.4.6. Dialogverlauf bzw. Dialogmanager

Der Dialogverlauf und damit die Arbeit des Dialogmanagers sind für die Probanden nur in wenigen Punkten ersichtlich, die nur indirekt die korrekte Funktion abbilden können. Die Durchschnittsbewertungen der Aussagen zum Dialogverlauf aus dem postaktionalen Fragebogen sowie die Standardabweichungen sind in Abbildung 7-18 visuell aufgearbeitet.

Das Verhalten des Avatarsystem wurde explizit nichtintrusiv gestaltet, um den Fahrer während der Fahrt nicht abzulenken. Schroyer (2000) konnte zeigen, dass eine systeminitiierte und somit intrusive Hilfe in einer Softwareanwendung kaum Aussichten auf Erfolg beim Benutzer hat. Dieser konzeptionelle Aspekt wurde von den Probanden in beiden Gruppen bestätigt, da sie es ablehnen, vom Avatarsystem häufiger angesprochen zu werden.

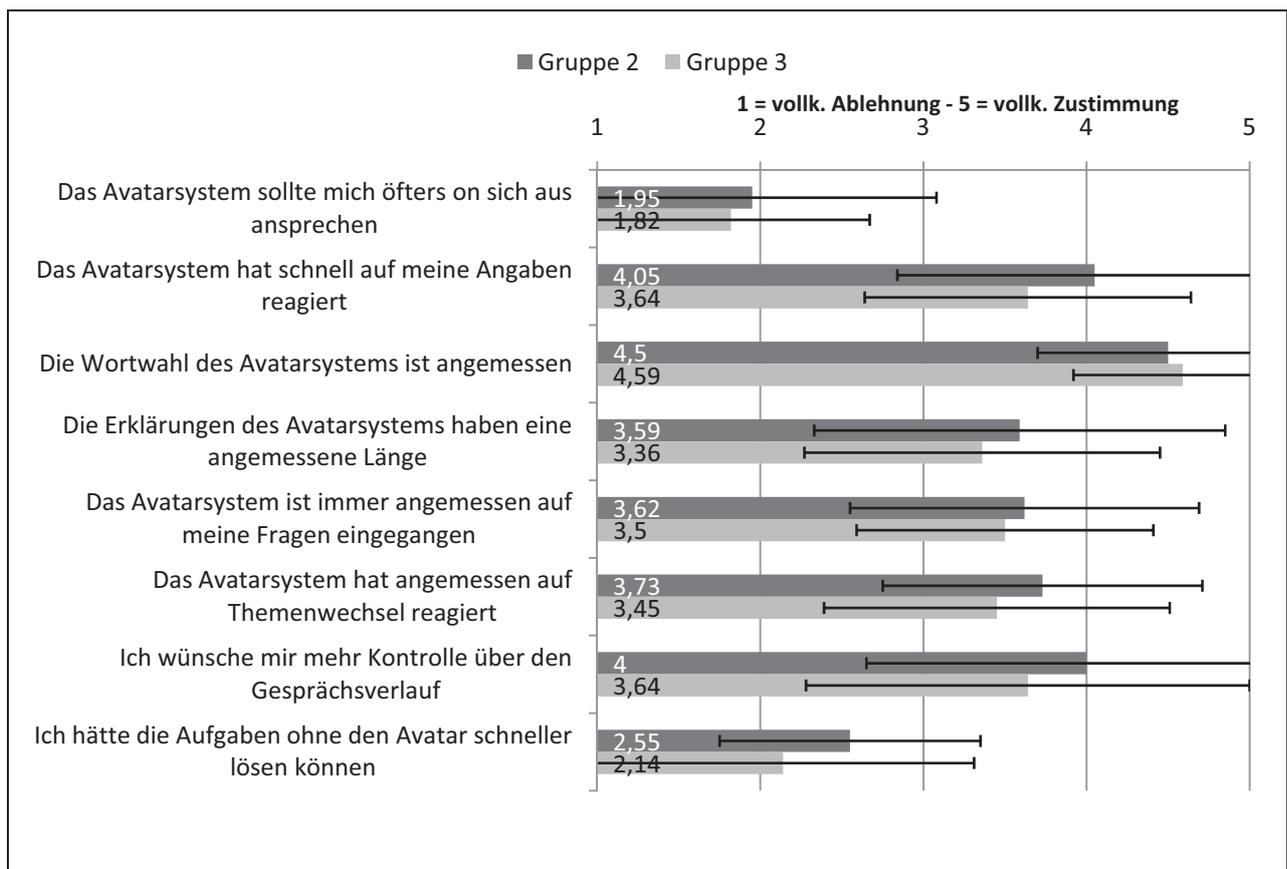


Abbildung 7-18: Bewertung des Dialogverlaufs
Quelle: (Eigene Darstellung)

Die Reaktion des Avatarsystems wurde von beiden Gruppen als schnell beurteilt, wobei die Probanden der fahrenden Gruppen dem weniger zustimmten. Dies kann damit zu tun haben,

dass die durch die Verkehrswahrnehmung erhöhte Aufmerksamkeit ein detaillierteres Zeitgefühl verleiht, bei dem manche Abläufe als langsamer wahrgenommen werden.

Die Wortwahl wurde einstimmig von beiden Gruppen als sehr angemessen angesehen. Tendenziell wurde auch die Länge der gegebenen Antworten als angemessen betrachtet, wobei die absoluten Werte im Vergleich zu den anderen Bewertungen dieses Bereichs möglicherweise auf zu lange oder zu kurze Antworten hinweisen.

Hinsichtlich der Interaktion zwischen Benutzer und Avatarsystems haben die Probanden zwar tendenziell den Eindruck gehabt, dass das Avatarsystem angemessen auf Fragen eingegangen ist und auf Themenwechsel reagiert hat, jedoch ist dies aufgrund der absoluten Werte im Vergleich zu den anderen Bewertungen ein möglicher Verbesserungsbedarf. Dies wird auch deutlich in dem ausgeprägten Wunsch, mehr Kontrolle über den Dialog ausüben zu können. Dieser Wunsch weist jedoch eine starke Streuung auf und ist in der fahrenden Gruppe insgesamt weniger stark ausgeprägt. Dies kann damit erklärt werden, dass das Ausüben von mehr Kontrolle auf den Gesprächsfluss von den Fahrern als zusätzliches Ablenkungspotenzial angesehen wird.

Beide Gruppen waren sich zudem einig, dass sie die gestellten Aufgaben ohne das Avatarsystem tendenziell nicht schneller hätten lösen können, wobei diese Meinung bei den Probanden der Gruppe 3 stärker ausgeprägt war.

7.2.4.7. Gesamteindruck

Aussagen, die sich nicht einer für den Probanden wahrnehmbaren Komponente zuordnen ließen, wurden im Gesamteindruck abgefragt, dessen Durchschnittsbewertung mit Standardabweichungen in Abbildung 7-19 dargestellt ist.

Die Verwendung des Avatarsystems wurde von beiden Gruppen als angenehm empfunden, wobei die fahrende Gruppe dies minimal positiver bewertet. Ebenso hat das Avatarsystem die Erwartungen an ein solches System erfüllt, wobei Gruppe 3 dies minimal weniger bejahte.

Die technische Reife wird von den beiden Gruppen unterschiedlich bewertet. Die fahrenden Teilnehmer der Gruppe 3 stehen der Aussage neutral gegenüber, während Gruppe 2 das System als signifikant ausgereifter einstuft. Die Varianzanalyse für die Antworten dieser Frage zeigt mit einem F-Wert von 4,57 eine deutliche Abweichung an. Unabhängig von der Bewertung der technischen Reife sind die Probanden in beiden Gruppen sehr davon überzeugt, dass Avatarsysteme in Zukunft häufig in Autos anzutreffen sein werden.

Ein Aspekt, der aus soziologischer Sicht interessant ist, ist die Frage danach, ob die Probanden den Avatar als soziales Wesen im Sinne der Theorie sozialer Präsenz wahrgenommen haben (vgl. Biocca/Harms/Borggon 2003; Short/Williams/Christie 1976; Lombard/Ditton 1997). Beide Gruppen stehen dieser Aussage neutral gegenüber, wobei Gruppe 2 den Avatar tendenziell als soziales Wesen empfindet.

Die Frage nach der Schwierigkeit der Aufgabenlösung wurde von beiden Gruppen verneint, was jedoch auch mit dem Aspekt zu tun haben kann, sich selbst nicht die Blöße zu geben. Nichtsdestotrotz wurden die gestellten Testaufgaben von Gruppe 3 als etwas schwieriger eingestuft, was durch die Doppelbelastung des Fahrens als Hauptaufgabe und der Aufgabenbearbeitung als Nebenaufgabe bedingt sein kann (vgl. dazu Schweigert/Bubb 2003; Totzke 2001; Piechulla et al. 2002).

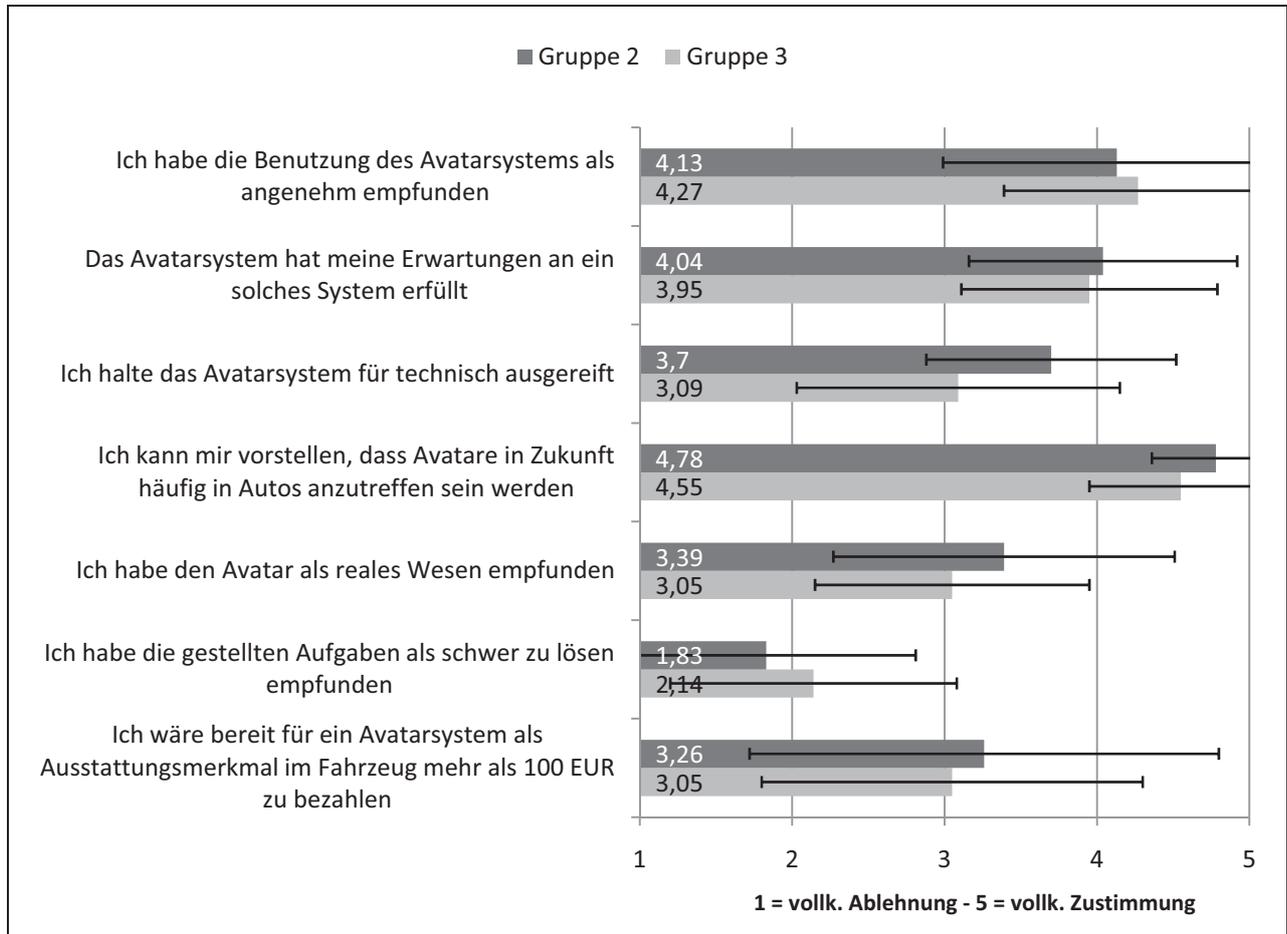


Abbildung 7-19: *Bewertung des Gesamteindrucks*

Quelle: (Eigene Darstellung)

Die letzte geschlossene Frage des postaktionalen Fragebogens hatte den Zweck, die prinzipielle Zahlungsbereitschaft der Probanden für solch ein System abzufragen. Dabei wurde ein fiktiver Preis von 100€ vorgegeben, um keine falschen Annahmen aufkommen zu lassen. Die Probanden der Gruppe 3 standen dem neutral gegenüber, während Gruppe 2 dies tendenziell bejahte. Die große Streuung in beiden Gruppen lässt jedoch vermuten, dass diese Einstellung der Teilnehmer von weiteren Faktoren abhängt, die gesondert geklärt werden müssen.

Auch wenn die einzelnen Komponenten des Avatarsystems aus verschiedenen Blickwinkeln betrachtet wurden, kann als Indiz für den Verbesserungsbedarf einzelner Komponenten der Mittelwert der Werte über beide Gruppen hinweg herangezogen werden. Dabei wurden nur Aussagen gewertet, die einen umso höheren absoluten Wert aufwiesen, desto zufriedener die

Probanden mit dem Aspekt der Komponente waren. Demnach ergibt sich das in Abbildung 7-20 gezeigte Bild, wonach insbesondere die Spracherkennung großen Verbesserungsbedarf erkennen lässt.

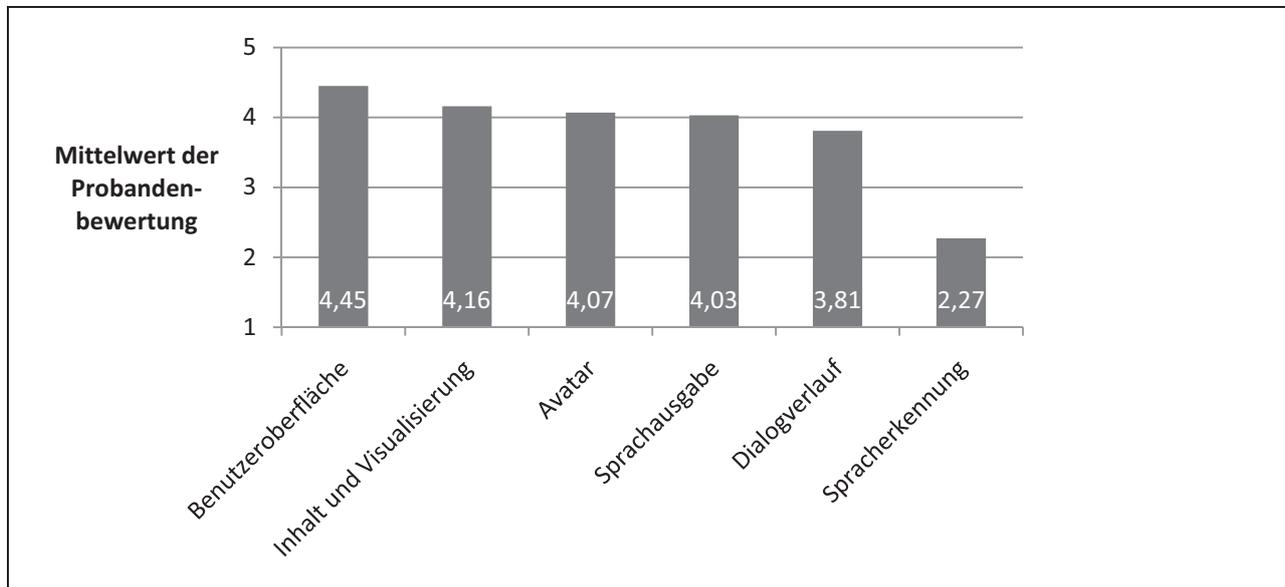


Abbildung 7-20: Gesamtbewertungsdurchschnitte als Indiz für den Verbesserungsbedarf
Quelle: (Eigene Darstellung)

7.2.5. Offene Bewertung des Avatarsystems

Um den ersten Eindruck über den Verbesserungsbedarf der Komponenten zu prüfen, der sich aus den Bewertungen der Probanden ergeben hat, sollen nun abschließend die Antworten auf offene Fragen am Ende der postaktionalen Fragebögen analysiert werden. Von diesen Fragen sind insbesondere diejenigen interessant, in denen Probanden beschrieben haben, welche Aspekte ihnen besonders positiv bzw. besonders negativ aufgefallen sind. Die genannten Verbesserungsvorschläge aus den Fragebögen sollen im Ausblick in Kapitel 8.3 aufgegriffen werden. Die von den Probanden genannten Aspekte wurden kategorisiert, wobei eine neue Kategorie ab drei ähnlichen Nennungen gebildet wurde. Dadurch lassen sich die Kategorien nicht immer eindeutig den technischen Komponenten des Avatarsystems zuordnen, jedoch lässt sich der Kreis der betroffenen Komponenten einschränken.

7.2.5.1. Positiv wahrgenommene Aspekte in den offenen Fragen

Die Anzahl der Nennungen positiver Aspekte in den jeweiligen Kategorien sind in Abbildung 7-21 nach Häufigkeit absteigend sortiert zu sehen. Insgesamt wurden 94 positive Äußerungen gemacht.

Am häufigsten wurde die Visualisierung der erklärten Sachverhalte als positiver Aspekt genannt. Dabei wurden meist die Existenz der Medien selbst sowie deren Qualität genannt. An zweiter Stelle wurde der schnelle Zugriff auf die gesuchten Informationen aufgeführt, wobei die einfache Verfügbarkeit sowie die schnelle Reaktion auf Spracheingaben wichtige Punkte waren. Die Gestaltung und Wiedergabe der Erklärungen wies die dritthäufigste

Nennung auf. Insbesondere die klar verständliche Präsentation der Informationen Schritt für Schritt sowie die Verbindung zu weiterführenden Themen wurden dabei erwähnt. Die bessere Vermittlung von Inhalten durch Animationen, die Schritt für Schritt ablaufen, konnten Palmiter und Elkerton (1991) nachweisen.

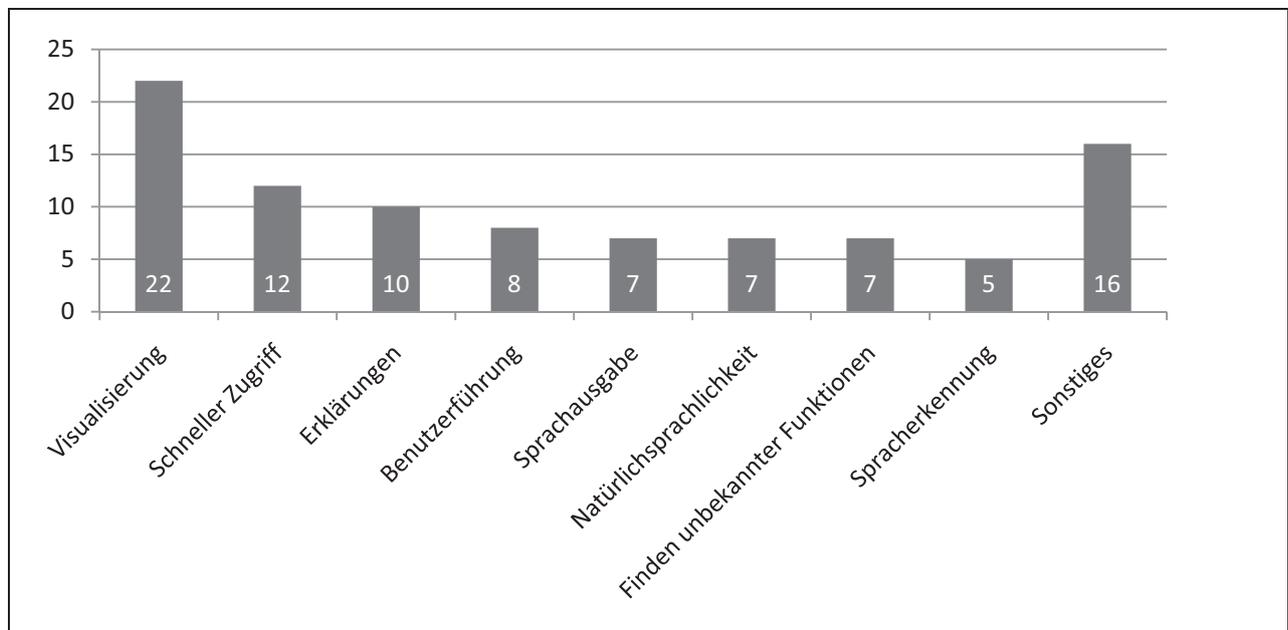


Abbildung 7-21: Anzahl der Nennungen positiver Aspekte nach Kategorien geordnet
Quelle: (Eigene Darstellung)

Die Benutzerführung kam mit acht Nennungen auf Platz vier, wobei die intuitive Bedienung und die Verwendung eines Push-to-Talk-Buttons der Grund dafür waren. An fünfter Stelle stehen die drei Bereiche Sprachausgabe, Natürlichsprachlichkeit und das Auffinden bzw. Kennenlernen neuer Funktionen. Im Rahmen der Sprachausgabe wurden die angenehme Stimme und die klare Aussprache angeführt. Die Verwendung offener Fragen und der breite, abgedeckte Wortschatz bei der Spracherkennung waren die meisten Nennungen im Bereich der Natürlichsprachlichkeit. Das Finden unbekannter Funktionen selbst sowie damit verbunden das Zurechtfinden in unbekanntem Fahrzeugen wurden beim letzten Aspekt auf diesem Platz angeführt. Die letzte explizit formulierte Kategorie ist die Spracherkennung, die in fünf Fällen als gut bewertet wurde.

Die Rubrik Sonstiges umfasst verschiedene Aspekte wie z.B. den Zugriff während der Fahrt, das Design und die Menschlichkeit des Avatars, die technische Reife, der nicht starr-sequentielle Gesprächsverlauf sowie das Feedback auf die Betätigung von Bedienelementen im Fahrzeug.

7.2.5.2. Negativ wahrgenommene Aspekte in den offenen Fragen

Die insgesamt 85 negativen Äußerungen wurden ebenfalls ab drei ähnlichen Nennungen zu Kategorien zusammengeführt, welche in Abbildung 7-22 grafisch aufgeführt sind.

Als größte negative Kategorie haben sich die Erklärungen hervorgetan, bei denen insbesondere der zu große oder zu kleine Umfang einer Erklärung, die Verständlichkeit sowie die Anpassung an den Wortschatz des jeweiligen Benutzers genannt wurden. An zweiter Stelle wurde die Spracherkennung angeführt, welche als nicht zuverlässig bezeichnet wurde. Zudem wurde die Notwendigkeit des lauten Redens zur Spracheingabe genannt. An dritter Stelle erschien die Sprachausgabe, die als zu langsam, zu schnell, zu leise oder unnatürlich bezeichnet wurde.

Im Rahmen der Benutzerführung wurde die fehlende Möglichkeit des Gesprächsabbruchs bzw. der Gesprächsunterbrechung zum Betätigen der Bedienelemente und des Überspringens von Inhalten moniert. Einigen Benutzer war darüber hinaus nicht immer klar, ob das Avatarsystem sie korrekt verstanden hat und an welcher Stelle sie sich inhaltlich befanden. Das wiederholte Vortragen gleicher Erklärungen sowie die fehlende Option, eine alternative Erklärung zu erhalten, waren ebenfalls Punkte dieser Kategorie. Die Ablenkung beim Autofahren durch das Avatarsystem sowie die eingeschränkte Erkennung von Frage- und Befehlsformulierungen und die damit verbundene Umformulierung wurden gleich oft genannt.

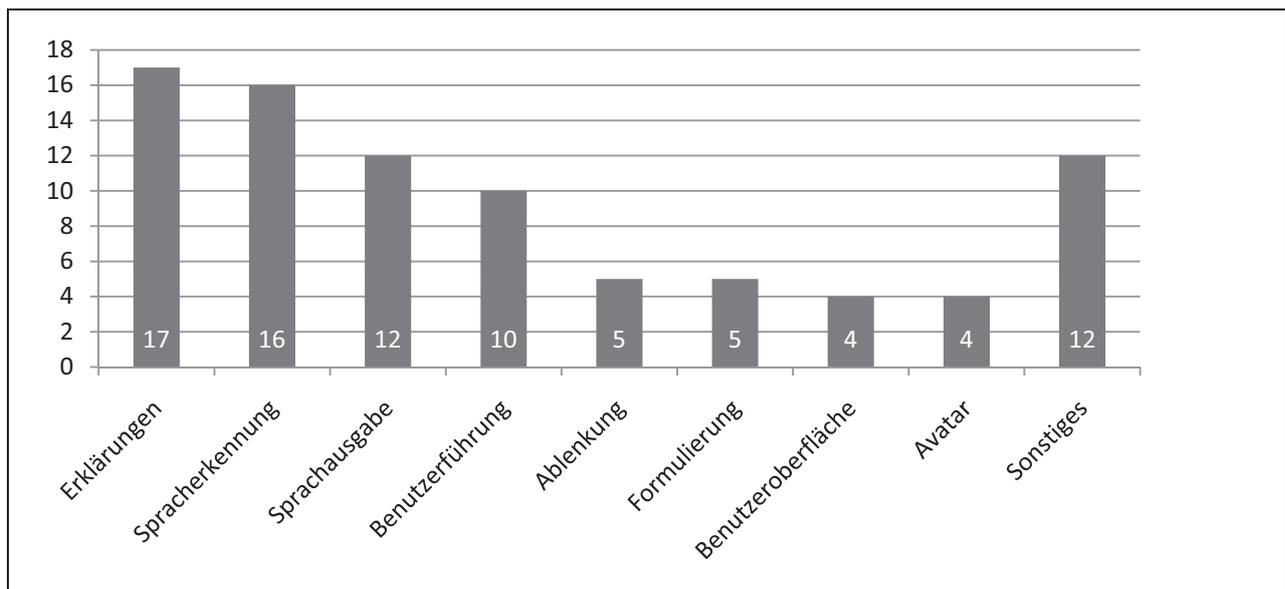


Abbildung 7-22: *Anzahl der Nennungen negativer Aspekte nach Kategorien geordnet*
Quelle: (Eigene Darstellung)

Auch wenn die Ablenkung nur von fünf Probanden genannt wurde, stellt dies einen Anteil von ca. 23% der Teilnehmer in Gruppe 3 dar und ist somit ein gewichtiger Aspekt, da durch den entstehenden Stress die Aufmerksamkeit auf den Verkehr reduziert wird und somit Unfälle erfolgen können (Matthews/Desmond 1995). Das Auftreten nicht verstandener Formulierung wird von Vetter (2003) auch im Fall von internetbasierten Avatarsystemen identifiziert. Dabei geht das mangelnde Verständnis aus seiner Sicht nicht auf die Formulierung des Benutzers, sondern auf die dafür fehlenden Regeln im Avatarsystem zurück. Da die Rückmeldung des Avatarsystems im Falle einer nicht verstandenen Formulierung sowie einer nicht verstandenen Spracheingabe die gleiche ist, ist zu vermuten,

dass ein Teil der negativen Äußerungen zur Spracherkennung auf fehlende oder mangelhafte Regeln zurückzuführen sind.

An der Benutzeroberfläche wurden die Symbole als nicht selbsterklärend genug und farblich suboptimal bezeichnet. Zudem wurde der Wunsch nach einer höheren Auflösung des Bildschirms und einer symbolischen Darstellung des Avatars anstelle der animierten Figur geäußert. Die Kritikpunkte am Avatar selbst bezogen sich auf die abgehackten und teilweise unpassenden Animationen, die altmodische Kleidung sowie den Namen „Claudia“.

Die Rubrik Sonstiges umfasst Äußerungen zu den langen Timeouts zwischen den Erklärungsschritten, der Positionierung des Mikrofons im Lenkrad anstelle des Dachhimmels, der Verzögerung zwischen Spracheingabe und der Antwort des Avatarsystems sowie der geringen Interaktion. Darüber hinaus wurde angemahnt, dass die Visualisierung dahingehend verbessert werden kann, dass bei jedem gezeigten Bedienelement verdeutlicht wird, wo es sich im Fahrzeug befindet. Als weiterer Kritikpunkt wurden die verwendete Fachterminologie bei der Eingabe sowie das fehlende Wissen über die Existenz von Funktionen im Fahrzeug beim Benutzer angeführt. Zudem wurden in einem Fall zwei Abstürze des Systems erwähnt, die während des Tests aufgetreten sind, sowie das ungeplante Auslösen einer Erklärung beim Überfahren einer Spurlinie mit eingeschalteten Audi lane assist.

7.3. Ableitung von Hypothesen

Die dargestellten Ergebnisse aus den Fragebögen und den Videoaufzeichnungen der Evaluation geben sowohl wichtige Anhaltspunkte zu den Anforderungen an die Gestaltung der technischen Komponenten, als auch Hinweise auf grundlegende Zusammenhänge in diesem Bereich. Die wichtigsten Ursache-Wirkung-Beziehungen, die sich aufgrund des Datenmaterials abzeichnen, werden im Folgenden in fünf Hypothesen formuliert, deren Wahrheitsgehalt in weiteren Untersuchungen detailliert überprüft werden muss.

Hypothese 1: Mit Avatarsystemen multimedial vermittelte Inhalte werden von Benutzern besser verinnerlicht als bei der Vermittlung auf andere Arten.

Die Ergebnisse der Wissensfragen im Nachgang der Evaluation in Kapitel 7.2.2.4 legen den Verdacht nahe, dass Inhalte, die mithilfe eines Avatarsystems unter Nutzung von Sprachausgabe und der Anzeige von Bildern oder Videos vorgetragen werden, besser durch den Benutzer verinnerlicht werden. André, Rist und Müller (1998b) konnten in ihren Untersuchungen jedoch keine signifikanten Auswirkungen der Präsentation von Informationen durch einen Avatar erkennen. Diese Diskrepanz kann in der Unterschiedlichkeit der Informationen und des Präsentationsumfelds begründet sein. Die Informationen zur Bedienung des Fahrzeugs konnten durch ihre Präsentation am Ort des Geschehens, d.h. direkt im Auto, sofort in die Tat umgesetzt werden, wodurch sich die Informationen besser einprägen konnten. Dieser Sachverhalt repräsentiert den Aspekt des richtigen Ortes des Informationszugriffs im Postulat des informationslogistischen Prinzips (vgl. Augustin 1990).

Hypothese 2: Mit dem Avatarsystem können gesuchte Informationen schneller gefunden und als relevant identifiziert werden als mit anderen Medien.

Weiterhin im Sinne des informationslogistischen Prinzips deuten die Ergebnisse der Evaluation darauf hin, dass das Avatarsystem insbesondere die Aspekte der Information zur richtigen Zeit und der Richtigkeit der Information unterstützt (vgl. Augustin 1990). Die positiven Äußerungen im offenen Teil des postaktionalen Fragebogens zeigen, dass der Informationszugriff als schnell wahrgenommen wurde. Diesem schnellen, initialen Zugriff steht jedoch die längere Bearbeitungszeit der Testaufgaben mithilfe des Avatarsystems gegenüber (vgl. Kapitel 7.2.2.2). Auch wenn die Gesamtbearbeitungszeit mit dem Avatarsystem höher ist, ist insbesondere bei den zu Beginn gestellten Aufgaben der Kategorie 1 ein deutlicher Zeitgewinn mithilfe des Avatarsystems zu verzeichnen. Die genauen Ursachen dieses Widerspruchs gehen nicht klar aus den erhobenen Daten hervor; sie können jedoch in der teilweise fehlenden Reife der technischen Komponenten, insbesondere der Spracherkennung oder in der ungewohnten Bedienung des neuartigen Avatarsystems begründet sein. Diese Lerneffekte, die sich zu Beginn der Bedienung eines neuartigen Systems einstellen, werden auch durch die Frage nach der Leichtigkeit der Bedienung bestätigt, wonach das Handbuch als am leichtesten zu bedienen angesehen wird (vgl. Kapitel 7.2.3.1).

Hypothese 3: Die Verwendung eines Avatarsystems führt zu einer zielsichereren Verwendung von Bedienelementen.

Nicht nur der Informationszugriff ist mit dem Avatarsystem schneller möglich, sondern auch die Anwendung der vorgetragenen Informationen. Die Auswertung der Zielsicherheit bei der Bedienung in Kapitel 7.2.2.3 deutet darauf hin, dass relevante Bedienelemente schneller gefunden bzw. sicherer bedient wurden. Dies liegt vermutlich zum einen an der visuellen Darstellung des Bedienelements im Fahrzeugkontext und zum anderen an der sprachlichen Beschreibung der Bedienungsart.

Hypothese 4: Die Verwendung eines Avatarsystems macht mehr Spaß als andere Hilfesysteme im Fahrzeug.

Die Benutzer des Avatarsystems in den Gruppen 2 und 3 haben im Rahmen der Bewertung der Benutzerfreundlichkeit (vgl. Kapitel 7.2.3.1) den Spaßfaktor bei der Lösung der Testaufgaben deutlich als höher eingestuft als die Benutzer des Handbuchs. Bestätigt wird diese Annahme zudem durch den deutlich geäußerten Wunsch der Wiederverwendung des Avatarsystems im Gegensatz zum Handbuch.

Hypothese 5: Die Akzeptanz des Avatarsystems ist insbesondere von der fehlerfreien Funktion der Spracherkennung abhängig.

Die Betrachtung der technischen Komponenten hat insbesondere die Spracherkennung als kritisch in diesem konkreten Avatarsystem identifiziert. Darüber hinaus wurde der Anteil an Sprachkommandos anstelle von natürlichsprachlichen Eingaben mit zunehmender Anzahl

bearbeiteter Aufgaben größer, was auf ein Umdenken in der Bedienung durch den Benutzer hinweist. Dies legt den Schluss nahe, dass diese Komponente somit auch einen deutlichen Einfluss auf die Akzeptanz des Systems hat. Die Korrelationskoeffizienten wichtiger Aspekte der Spracherkennung mit der Schulbenotung, dem Spaßfaktor und des Gesamteindrucks zeigen ein differenziertes Bild. Tabelle 7-9 zeigt die Korrelationskoeffizienten, wobei jeweils nach Gruppe 2 und Gruppe 3 getrennt wurde.

Demnach ergibt sich für die Frage danach, ob das Avatarsystem angenehm empfunden wurde und den Spaßfaktor ein klares Bild, das den vermuteten Zusammenhang begründet. Benutzer, die sich von der Spracherkennung immer verstanden gefühlt haben, haben das System als angenehmer und mit mehr Spaß behaftet empfunden. Die Notwendigkeit, eine Spracheingabe zu wiederholen oder besonders laut vorzunehmen, hat sich entsprechend negativ auf die Gesamtbewertung ausgewirkt. Die errechneten Korrelationskoeffizienten weisen dabei überwiegend auf signifikante Zusammenhänge hin (fett markiert). Erstaunlicherweise scheint die Benotung des Systems von der Bewertung der Spracherkennung im Fall von Gruppe 2 entgegengesetzt beeinflusst zu werden, während sich bei Gruppe 3 kein klares Bild erkennen lässt.

		Bewerten Sie das Avatarsystem mit Schulnoten	Ich habe die Benutzung des Avatarsystems als angenehm empfunden	Die Benutzung des Avatarsystems hat mir Spaß gemacht
Ich hatte immer den Eindruck, dass mich die Spracherkennung versteht	Gruppe 2	-0,62	0,62	0,59
	Gruppe 3	-0,17	0,43	0,24
Ich musste meine Eingabe häufig wiederholen	Gruppe 2	0,61	-0,54	-0,47
	Gruppe 3	0,14	-0,32	-0,20
Ich musste laut reden, damit mich die Spracherkennung verstanden hat	Gruppe 2	0,41	-0,25	-0,12
	Gruppe 3	-0,15	-0,49	-0,26

Tabelle 7-9: *Korrelationskoeffizienten der Spracheingabe und der Akzeptanz*
Quelle: (Eigene Darstellung)

7.4. Zusammenfassung

Zur Bewertung des entwickelten Avatarsystem sowie zur Gewinnung von Erkenntnissen zur Verbesserung des Systems, wurden im Rahmen einer Evaluation mit insgesamt 67 Probanden Informationen gesammelt. Die Evaluation bestand aus drei Teilen: Einem präaktionalen Fragebogen zur Erhebung demographischer Daten, dem Lösen von Testaufgaben in einem Versuchsfahrzeug und einem postaktionalen Fragebogen zur subjektiven Bewertung des verwendeten Hilfesystems. Die Probanden wurden dabei auf drei Gruppen verteilt. Die erste Gruppe diente als Kontrollgruppe und führte die Testaufgaben im stehenden Fahrzeug unter Zuhilfenahme des Bordhandbuchs durch. Die zweite Gruppe verwendete anstelle des Handbuchs das Avatarsystem bei stehendem Fahrzeug und die dritte Gruppe löste die

Testaufgaben mit dem Avatarsystem während der Fahrt. Neben den unterschiedlichen Testaufgaben unterschieden sich auch die postaktionalen Fragebögen zwischen Gruppe 1 und den avatarsystemverwendenden Gruppen 2 und 3. Die Bearbeitung der gleichen Testaufgaben wurde unabhängig von der Gruppenzugehörigkeit auf Video aufgezeichnet und später ausgewertet. Bei der Zusammensetzung der drei Gruppen wurde darauf geachtet, dass sie hinsichtlich Alter und Geschlechtsverteilung gleichverteilt sind.

Die Anzahl erfolgreich durch die drei Gruppen bearbeiteter Testaufgaben ergab kein eindeutiges Bild, jedoch zeigte die Gesamtbearbeitungsdauer einen Unterschied zwischen der Verwendung des Handbuchs und des Avatarsystem. Die Probanden, die ein Handbuch verwendeten, waren insgesamt schneller in der Aufgabenerledigung als die Teilnehmer der beiden anderen Gruppen. Ein anderes Bild ergab sich hingegen bei der Zielsicherheit der korrekt verwendeten Bedienelemente, bei der Gruppe 2 und 3 besser abschnitten. Ebenso war eine bessere Wissensvermittlung bzw. Verinnerlichung des neu Gelernten bei der Verwendung des Avatarsystems zu verzeichnen.

Im Bereich der Benutzerfreundlichkeit wurde das Avatarsystem als intuitiver in der Bedienung bewertet und vermittelte den Benutzern mehr Spaß bei der Verwendung. Insgesamt fiel der Umgang mit dem Handbuch jedoch leichter als mit dem Avatarsystem. Die Teilnehmer der Gruppen 2 und 3 bewerteten das Avatarsystem als die Nützlichere der beiden Alternativen und empfanden auch die vorgetragenen Informationen als zielgerichteter. Dennoch wurde es weniger als Hilfe bei der Aufgabenerfüllung angesehen als das Handbuch. Der mit Schulnoten widergegebene Gesamteindruck der Probanden zeigte das Handbuch mit dem Avatarsystem während der Fahrt gleich auf, während das Avatarsystem im Stand die beste Schulnote erhielt.

Neben den vergleichenden Fragen erhielten die Teilnehmer, die Erfahrungen mit dem Avatarsystem gemacht hatten, zudem gezielte Fragen zu den jeweiligen technischen Komponenten, um Schwachstellen aufzudecken. Dabei wurden die Benutzeroberfläche, die Inhalte und deren Visualisierung sowie der Avatar selbst am besten bewertet. Die deutlichste Schwachstelle zeigte sich bei der Komponente der Spracherkennung. Die in geschlossenen Fragen zu den Komponenten erhobenen Daten deckten sich zudem größtenteils mit der offenen Nennung von positiven und negativen Aspekten der Benutzung des Avatarsystems.

Die während der umfangreichen Evaluation gesammelten Daten geben bereits in dieser frühen Phase der Avatarsysteme im Fahrzeug verschiedene Indizien auf die Zusammenhänge von Faktoren und die Wirkungen eines solchen Systems auf die Benutzer. Auch wenn an zahlreichen Stellen Verbesserungspotenzial gesehen werden kann, lässt sich damit dennoch die dritte Forschungsfrage nach den Effekten eines Avatarsystems im Fahrzeug beantworten. Die auf Grundlage des Datenmaterials formulierten Hypothesen zeigen dabei verschiedene Richtungen, um die vermuteten Zusammenhänge detaillierter in gesonderten Untersuchungen zu bestätigen oder zu widerlegen.

8. Schlüsse aus der Umsetzung des Hilfesystems

„Der einzig wahre Realist ist der Visionär.“

Federico Fellini (1920 - 1993), italienischer Regisseur und Schriftsteller

Die vorliegende Forschungsarbeit deckt viele Aspekte der Gestaltung und des Einsatzes von Avatarsystemen im Fahrzeug ab. Während die bisherige Betrachtung überwiegend chronologisch erfolgte, werden in diesem Kapitel zunächst die wichtigsten Erkenntnisse aus Sichtweise der in Kapitel 1.3 formulierten Forschungsfragen beleuchtet. Im darauf folgenden Fazit der Arbeit werden sowohl positive als auch negative Erfahrungen aus dem Umgang mit der Materie wiedergegeben und die Konsequenzen für Forschung und Praxis kurz dargestellt. Abschließend werden Entwicklungsmöglichkeiten erläutert, die sich zum einen konkret auf das entwickelte Avatarsystem und zum anderen auf allgemeine Möglichkeiten der Weiterentwicklung aus dieser Arbeit beziehen.

8.1. Zusammenfassung der Ergebnisse

Zu Beginn dieser Forschungsarbeit wurden drei Forschungsfragen formuliert, die sich mit der Gestaltung einer angemessenen Architektur für Avatarsysteme im Allgemeinen und im Fahrzeug insbesondere, der Überführung von strukturierten Informationen in Dialogform für ein Avatarsystem und der Wirkung eines solchen Systems auf den Fahrer befassen. Aus der Perspektive der jeweiligen Frage sollen die bisherigen Ergebnisse im Folgenden resümiert werden.

8.1.1. Technische Entwicklung des Avatarsystems

Die erste Forschungsfrage zielte primär auf die Entwicklung einer Referenzarchitektur für Avatarsysteme im Allgemeinen sowie für den Einsatz in einem Fahrzeug im Besonderen ab. Die Fragestellung nach der Architektur alleine wird der Problemstellung jedoch kaum gerecht, da sich im Laufe der praktischen Anwendung der Architektur neue Aspekte eröffneten, die nur im implementierenden Kontext deutlich werden können. Um beiden Teilbereichen der Gestaltung des Avatarsystems gerecht zu werden, werden zunächst die Gewinnung der Architektur und anschließend die technische Umsetzung beleuchtet.

Die Konzeption eines Avatarsystems für den Einsatz im Fahrzeug hat bereits zu Beginn das Problem aufgeworfen, dass bisher nur wenige Avatarsysteme mit Sprachein- und -ausgabe sowie der Anbindung zahlreicher Bedienelemente umgesetzt worden sind und keines davon im Automotive-Bereich. Dies machte es notwendig, in benachbarten Forschungsbereichen zu recherchieren, wie z.B. den Sprachdialogsystemen, die auch eine Dialogstruktur jedoch ohne Avatar aufweisen. Aus der Forschung zu Avatarsystemen und Sprachdialogsystemen wurden markante Architekturkonzepte analysiert. Die Gegenüberstellung verschiedener Architekturen wurde durch die Zerlegung in ihre Komponenten ermöglicht. Es zeigte sich, dass trotz einer Vielzahl verschiedener Benennungen und Funktionsabgrenzungen, ein ähnlicher Stamm an Komponenten identifiziert werden konnte. Diese Komponenten wurden einem von fünf

Systemen zugeordnet: Der Eingabe, der Ausgabe, den Datenkomponenten, der Reaktionsermittlung oder dem Design.

Diese Modularisierung ermöglichte die Aufteilung des Aufwands, wodurch zunächst für jedes Teilsystem eine Referenzarchitektur entwickelt werden konnte. Insbesondere die Reaktionsermittlung erwies sich dabei als Herausforderung, da sie sowohl state-, frame- als auch agenten-basierte Ansätze darstellen können musste. Schließlich entstand aus der Verbindung der einzelnen Systeme eine Gesamtarchitektur, welche die aktuellen sowie zukünftigen Anforderungen an Avatarsysteme berücksichtigte.

Die Anwendung der Referenzarchitektur für den Einsatz im Fahrzeug zeigte, dass auch aufgrund der geplanten Verwendung eines frame-basierten Ansatzes der Reaktionsermittlung nur ein geringer Teil der Komponenten für die Reaktionsermittlung notwendig war. Die Mobilität des Avatarsystems durch den Einsatz im Fahrzeug machte es darüber hinaus notwendig, eine Systematik für die Verteilung von Komponenten auf das Fahrzeug bzw. einen unbeweglichen Teil zu konzipieren. So wurde im Sinne einer eigenständigen Laufzeit die gesamte Laufzeit für den Einsatz im Auto vorgesehen, während das Design der Inhalte außerhalb erfolgte.

Die konkrete technische Implementierung des Avatarsystems im Fahrzeug erforderte viele Entscheidungen, die zu Beginn zu treffen waren. So musste aus einer Vielzahl speziell für den Automotive-Bereich entwickelter Entwicklungskonzepte und -plattformen eine geeignete ausgewählt werden. Die Wahl fiel auf das OSGi-Konzept, welches die Gestaltung und das Zusammenspiel verschiedener Java-Bundles ermöglicht und durch verschiedene Implementierungen grundlegende Services bereitstellt. Als Implementierung wurde Eclipse Equinox ausgewählt, da es im Gegensatz zu anderen Lösungen alle bereits bestehenden und durch verschiedenen Fachabteilungen und Projekte entwickelten Bundles vereinen kann.

Bezüglich der Komponenten des Systems stand die klassische Make-or-Buy-Entscheidung an, die für den Fall der Spracheingabe und der -synthese, des Chatbots und des Avatars zugunsten von Standardkomponenten ausfiel. Deren Einbindung erzeugte dennoch einen hohen Aufwand, da sie bis auf den Chatbot keine Java-Schnittstellen enthielten. Die Umsetzung des Dialogmanager, des GUI-Frameworks sowie die Integration aller Komponenten mithilfe eines flexiblen Eventings erfolgte im Rahmen dieser Forschungsarbeit. Die Integration stellte im Rahmen der Implementierung und bei der Umsetzung des Inhaltes eine der größten Herausforderungen dar, der auf technischer Ebene mit der Entwicklung zahlreicher Bundles und auf inhaltlicher Ebene mit einer integrierten Authoring-Anwendung begegnet wurde. Neben der Bereitstellung von passenden Funktionen zur Steuerung des Dialogflusses erwies sich auch die Integration der multimodalen Eingabekanäle als Herausforderungen.

Neben der Softwareentwicklung war die Integration des Avatarsystems im Fahrzeug, d.h. im Feld, Teil des Vorhabens. Dabei musste den Leistungseinschränkungen eines CarPC's ebenso Rechnung getragen werden wie der Integration von Mikrofon und Lautsprecher im Fahrzeug. Insbesondere bei der Übertragung akustischer Signale spielten auch elektronische Belange wie die Entstörung von Bauteilen zur Begegnung von Interferenzen durch drehende

Elementen des Fahrzeugs eine Rolle. Darüber hinaus wurde das Handling des Avatarsystems den Gegebenheiten im Fahrzeug angepasst, ebenso wie die störungsfreie Integration der visuellen Anzeige des Avatarsystems bewerkstelligt wurde. Zudem erfolgte die Analyse des Aufbaus und des Verhaltens von Steuergeräten an den Bussystemen des Fahrzeugs, die bei der Gestaltung der entsprechenden Eingabemodalität berücksichtigt wurde.

8.1.2. Gestaltung der Inhalte

Die Laufzeit des Avatarsystems allein ist nicht lauffähig in dem Sinne, dass sie einen Mehrwert leisten könnte. Das Element, das ihr „Leben einhaucht“, ist der Inhalt mithilfe dessen Dialoge erst möglich werden. Im Falle des vorliegenden Avatarsystems basieren die Inhalte auf dem Benutzerhandbuch des Versuchsfahrzeugs, über dessen Bedienung Auskunft gegeben werden soll. Da bei diesem Prototyp nicht alle Inhalte umgesetzt und auf Verbesserungspotenzial bei der Überführung hin untersucht werden konnten, wurden nur ausgesuchte Szenarien betrachtet, die sich auch für die spätere Evaluation eigneten. Die Auswahlkriterien bezogen sich primär darauf, dass es sich um Inhalte über Bedienelemente handelte, die für den Fahrer im Stand und während der Fahrt bedienbar sind, die zu ihrer Bedienung keine Anzeige im MMI-Display benötigen und deren Zustand über den angeschlossenen Komfort-CAN-Bus abgefragt werden konnten.

Beim Handbuch handelt es sich um strukturierten Inhalt, der auch in digitaler Form vorliegt, so dass über eine (teil-)automatische Konvertierung des Handbuchs in Dialogform nachgedacht werden musste. Da weder für strukturierte noch für unstrukturierte Inhalte ein detailliertes Vorgehensmodell zur Überführung in eine Dialogform dokumentiert ist, musste ein eigenes Vorgehensmodell entwickelt werden. Bereits in der Konzeptionsphase dieses Modells zeichnete sich großes Automatisierungspotenzial ab, so dass die parallele Entwicklung eines Modells und eines unterstützenden Softwaretools geplant wurde.

Der Dialogdesigner bzw. Dialogeditor, der zur Unterstützung des Vorgehensmodells entwickelt wurde, erlaubt die grafische Entwicklung von Gesprächsabläufen, die in Szenarien und Dialogen organisiert werden können. Trotz der hierarchischen Organisation, müssen die entwickelten Inhalte nicht hierarchisch durchlaufen werden, sondern können durch den Benutzer beliebig gewechselt werden. Die Gesprächsabläufe enthalten Eingabe- und Ausgabeschritte, die verschiedene Eingaben bzw. Ausgaben enthalten können. Bei der Eingabe kann zwischen einer Spracheingabe und einer durch ein Bedienelement verursachten Auslösung unterschieden werden. Die Ausgabeschritte enthalten Sprachausgaben, die Anzeige von Bildern und Videos, die Manipulation der Benutzeroberfläche bzw. von 3D-Objekten, die zur Veranschaulichung angezeigt werden können. Es sind Verweise auf andere Dialoge und darin enthaltene Schritte ebenso möglich wie die Verwendung von Variablen oder Zufallsgeneratoren.

Die grafisch entwickelten Dialoge können exportiert werden, wobei zur Variation möglicher Eingabeformulierungen so genannte Synonymgruppen verwendet werden, die beispielsweise ein Stichwort wie „Klimaanlage“ auch Formulierungsmöglichkeiten um die Begriffe „air condition“ oder „Klimaautomatik“ ergänzen können. Der konkrete Export der als Metamodell

erstellten Dialoge erfolgt zum einen als Regeln für die Wissensbasis des Chatbots und zum anderen synchronisiert als Sprachgrammatik für die Spracheingabe.

Durch die Entwicklung dieses Dialogeditors wurde ein Vorgehensmodell ermöglicht, dass in sieben Phasen die Umwandlung von strukturiertem und auch unstrukturiertem Inhalt begleiten kann. Die einzelnen Phasen stellen Teile des Zyklus einer Spirale dar, die sich mit zunehmender Nähe zum Mittelpunkt einem vollständigen und ausgereiften Inhalt in Dialogform annähert. Die Aufteilung in Phasen erlaubt zudem die Verteilung der verschiedenen Aufgaben an Experten bzw. Verantwortliche.

Nach dem automatischen Import strukturierter Daten im Falle des Handbuchs oder der Sammlung von Informationen im Falle unstrukturierter Informationen, erfolgt in der ersten Phase die Strukturierung von Inhalten bei der redundante und unpassende Informationen entfernt und die verbleibenden Daten in eine sinnvolle Dialogstruktur gebracht werden. In der zweiten Phase werden die Antworten auf verschiedene Eingaben gestaltet, da sich die Antworten i.d.R. kaum noch verändern im Laufe des Entwicklungsprozesses. Häufige Änderungen bzw. Erweiterungen erfahren hingegen die möglichen Eingaben, die in der dritten Phase gepflegt werden. Die vierte Phase ist durch die Bereitstellung und Anpassung von Medien und Avataranimationen im Antwortteil einer Regel gekennzeichnet. In der fünften Phase werden so genannte Umfeldfunktionen erstellt, die in bestimmten Situationen zusätzliche Handlungsoptionen bzw. Eingabemöglichkeiten darstellen. So kann die Eingabe „Wie funktioniert das“ je nach Kontext verschieden verarbeitet werden. Die optionale sechste Phase spiegelt die Einbindung von Emotionen in den Gesprächsablauf wider. Dabei können verschieden detaillierte emotionale Modelle sowohl für den Avatar als auch für den Benutzer umgesetzt werden. Den Abschluss eines Zyklus stellt der Test der Inhalte dar, der Verbesserungsbedarfe aufdecken soll, die im nächsten Zyklus gedeckt werden sollen.

8.1.3. Wirkung auf Benutzer

Das Avatarsystem mit den erstellten Inhalten wurde in einer dreiteiligen Evaluation durch Benutzer getestet und bewertet. Alle Probanden erhielten einen präaktionalen Fragebogen, mussten unter Beobachtung fünf Aufgaben im Versuchsfahrzeug lösen und sollten danach einen postaktionalen Fragebogen ausfüllen. Die Probanden wurden in drei Gruppen aufgeteilt: Einer Kontrollgruppe, die das Benutzerhandbuch zur Lösung der Testaufgaben verwendete, sowie zwei Gruppe, die das Avatarsystem zum einen bei stehenden Fahrzeug und zum anderen bei fahrendem Fahrzeug zur Aufgabenlösung verwenden konnten. Neben dem Vergleich von Handbuch und Avatarsystem dienten die Daten aus den Fragebögen und den Videoaufzeichnungen der Aufgabenbearbeitung auch der formativen Evaluation des Avatarsystems.

Der präaktionale Fragebogen deckte eine weitgehend homogene Zusammensetzung der drei Testgruppen auf. Die Bearbeitung der Aufgaben selbst erfolgte durch die Benutzer des Avatarsystems langsamer, aber sicherer in dem Sinne, dass das korrekte Bedienelement schneller gefunden und bedient werden konnte. Das während der Aufgabenbearbeitung vermittelte Wissen über die Bedienung und die Funktion von Bedienelementen wurde im

Fälle der Avatarsystembenutzer besser verinnerlicht und im Wissenstest des postaktionalen Fragebogens entsprechend besser wiedergegeben.

Die Bedienung des Avatarsystems wurde intuitiver bewertet und machte den Probanden mehr Spaß, jedoch wurde es weniger als konkrete Hilfe bei der Aufgabenbearbeitung angesehen. Nichtsdestotrotz sahen die Versuchsteilnehmer das Avatarsystem nützlicher an als das Handbuch. Insgesamt wurden das Handbuch und das Avatarsystem in Schulnoten gleich bewertet, wobei das Avatarsystem im Stand die beste Note erhielt.

Die formative Bewertung des Avatarsystems diene auch der Identifizierung von Verbesserungspotenzialen bei den technischen Komponenten. Die Benutzer konnten aufgrund ihres fehlenden Einblicks in die technischen Hintergründe nur nach Symptomen befragt werden, wobei dies mithilfe geschlossener und offener Fragen erfolgte. Während die Probanden mit den wahrnehmbaren Teilen der Komponenten weitgehend zufrieden waren, stellten sich die Gestaltung der Erklärungen insbesondere hinsichtlich der Länge und die Zuverlässigkeit der Spracherkennung als verbesserungsbedürftig heraus.

Die Ergebnisse der Evaluation zeigten somit insgesamt, dass das Avatarsystem als Hilfesystem im Fahrzeug von den Probanden gut aufgenommen und als sinnvoller Ansatz angesehen wurde. Das teilweise schlechtere Abschneiden des Avatarsystems bei der Lösung der Testaufgaben deutet jedoch im konkreten Fall auf verschiedene Bereiche hin, die verbessert werden müssen, um eine gleich gute Unterstützung wie das Handbuch zu bieten. Unbestritten bleibt dabei in jedem Fall, dass das Avatarsystem ausreichende Hilfestellungen geben konnte, um die Testaufgaben zu lösen, was mit einem Handbuch während der Fahrt sonst nur mit einem realen Beifahrer erzielt worden wäre.

8.2. Fazit

Das ehrgeizige Projekt, einen virtuellen Beifahrer im Auto zu entwickeln und durch Benutzer testen zu lassen war in der Projektzeit von zwei Jahren nur durch die Verwendung von Standardsoftware und Standards möglich. Neben dem zeitlichen Vorteil stellt dies auch aus technischer Sicht einen wichtigen Faktor dar, da zum einen nach dem Best-of-Breed-Ansatz die besten Technologien in den betroffenen Bereich herangezogen werden konnte. Darüber hinaus können Updates des Komponentenherstellers eingespielt werden bzw. bei der Verwendung von Standards auch der Hersteller gewechselt werden wie dies z.B. bei der OSGi-Implementierung während der Entwicklung geschehen war. Diese Erkenntnisse sind im Praxisbereich nicht neu, jedoch werden die Vorteile auch im wissenschaftlichen Bereich deutlich. Durch die maßgebliche Verwendung von Standardkomponenten sind zumindest Teile der Ergebnisse reproduzierbar und durch andere Forscher nachvollziehbar.

Die Konzeption der Evaluation verliert durch die Vorgabe von Testaufgaben ein wenig vom Charakter einer Felduntersuchung, jedoch lassen sich über solch eine große Menge an Probanden ansonsten keine vergleichbare Daten zur Verwendung sammeln. Zur Untersuchung des realen Verhaltens der Probanden im Umgang mit dem Avatarsystem müssten Langzeituntersuchungen mit Personen durchgeführt werden, die bereits mit dem entsprechenden Fahrzeugtyp vertraut sind. Dies macht jedoch die Entwicklung eines

aufwändigen und robusten Messaufbaus notwendig und erfordert bereits von Beginn an die Abdeckung aller Inhalte des Handbuchs. Auch wenn diese Herangehensweise detailliertere Daten liefern kann, sind dafür die Ergebnisse dieser vorliegenden Forschungsarbeit als Grundlage hilfreich.

Diese Grundlagen für weitere Forschung bestehen zum einen aus der Anwendung des Dialogdesigners, der die integrierte Entwicklung von Dialogen mit Multimedia-Elementen und den synchronisierten Export von Regelbasis und Sprachgrammatik erlaubt. Zum anderen bestehen sie auch aus der Laufzeitumgebung des Avatarsystems, welche die Umsetzung der entwickelten Dialoge und ihrer Medien mithilfe einer sprachgesteuerten, avatarbasierten Benutzeroberfläche vornimmt.

Das gesamte System lässt sich sowohl technisch als auch inhaltlich flexibel erweitern. Die technische Erweiterbarkeit ist durch die lose Kopplung von OSGi-Bundles mithilfe von Ereignissen sichergestellt. Die inhaltliche Erweiterung kann durch den teilautomatisierten Import von Inhalten des elektronischen Benutzerhandbuchs sowie die automatische Generierung weiterer Eingabeformulierungen durch die Verwendung von Synonymgruppen einfach erzielt werden. Auch die Pflege weiterer Bedienelemente im Fahrzeug als Quelle für Eingaben des Avatarsystems kann durch Pflege der entsprechenden XML-Datei vorgenommen werden und fördert somit die Interaktion zwischen dem Avatarsystem und dem Fahrzeug. Das technische System und die genannten Erweiterungsoptionen machen das Avatarsystem zu einer Ausgangsbasis für Erweiterungen in technischer und inhaltlicher Sicht, wodurch neue Ideen und Konzepte direkt im Fahrzeug getestet werden können. Somit erleichtert das System im Sinne des Rapid Prototyping die Entscheidungsfindung im Bereich neuer Bedienkonzepte im Fahrzeug.

Die Evaluation hat neben der breiten Akzeptanz der Benutzer auch gezeigt, dass solch ein System selbst in einem so frühen Stadium bereits während der Fahrt eingesetzt werden kann und somit dem Handbuch in dieser Hinsicht überlegen ist. Die Bewertung durch die Probanden hat zudem Stärken und Schwächen des Systems aufgedeckt, so dass die Evaluationsergebnisse als Leitfaden für die weitere Entwicklung verwendet werden können. Insbesondere die Existenz eines natürlichsprachlichen Systems und damit auch dessen Bedienung sind für meisten Probanden jedoch noch ungewohnt, weshalb insbesondere über eine Vereinfachung der Bedienung bzw. die stärkere Anlehnung an bekannte Anwendungen nachgedacht werden sollte (vgl. auch Totzke/Meilinger/Krüger 2003).

Die vorliegende Arbeit zeigt die Richtung der weiteren Forschung im Bereich der Avatarsysteme auf. Auch wenn das vorgestellte System technisch noch nicht zuverlässig genug im Fahrzeug läuft und die Bedienung noch zu ungewohnt für manche Benutzer ist, zeigen die Ergebnisse klar auf, dass die Visualisierung von Informationen zur Bedienung des Fahrzeugs und der schnelle Zugriff darauf von Benutzern als wichtige Punkte angesehen werden. Somit sind zumindest diese Teilaspekte des Avatarsystems näher aus Sicht der Überführung in den praktischen Serieneinsatz zu betrachten. Konkrete Schritte, um einem verbesserten Stand des Systems herzustellen, werden zum einen konkret für das AViCoS-System und anschließend allgemein im folgenden Teilkapitel beleuchtet.

8.3. Ausblick

Bereits während der Entwicklung des Avatarsystems und seiner Designanwendung haben sich viele weiterführende Ideen ergeben, die im Folgenden nach ihrer Umsetzbarkeit geordnet aufgeführt werden. Zum einen werden Verbesserungsmöglichkeiten am AViCoS-System vorgestellt, die in naher Zukunft umgesetzt werden können. Zum anderen werden weiter reichende Neuerungen betrachtet, die teilweise im Rahmen dieser Forschungsarbeit bereits als Patentanmeldungen konzipiert und eingereicht wurden.

8.3.1. Entwicklungspotenzial beim AViCoS-Avatarsystem

Zunächst soll das vorliegende und ausführlich beschriebene Avatarsystem hinsichtlich technischer und inhaltlicher Erweiterungen analysiert werden.

8.3.1.1. Technische Weiterentwicklung

Ein technischer Schwachpunkt, der sich in der Evaluation gezeigt hat, und ein generelles Problem von Sprachdialogsystemen ist, stellt die Spracherkennung dar. Sie bestimmt darüber, wie flüssig und aus Sicht des Benutzers zufriedenstellend die Interaktion abläuft. Im Laufe der Arbeit haben sich drei Maßnahmen herauskristallisiert, bei denen die größten Verbesserungswirkungen vermutet werden.

Das für den Einsatz im Fahrzeug zusätzlich verbaute Mikrofon befand sich im Dachhimmel im Bereich der vorderen Innenbeleuchtung. Insbesondere während der Fahrt konnten die Probanden jedoch häufig dabei beobachtet werden, dass sie sich instinktiv zum Lenkrad vorbeugten, um ihre Spracheingaben vorzunehmen. Da im Lenkrad bereits ein Mikrofon serienmäßig verbaut ist, muss es das Ziel sein, dieses Mikrofon zu verwenden, um weniger Störgeräusche während der Fahrt einzufangen und damit die Qualität des eingehenden Sprachsignals zu verbessern.

Eine weitere Maßnahme ist die Berücksichtigung der Erkennungswahrscheinlichkeit einer bestimmten Eingabeformulierung. Bisher wurde lediglich der Eintrag aus der Sprachgrammatik zurückgemeldet, auf den am wahrscheinlichsten das Sprachsignal passt. Dabei wurde jedoch nicht berücksichtigt, wie hoch die absolute Wahrscheinlichkeit liegt. Nach der Festlegung eines bestimmten prozentualen Schwellenwerts könnte in Zukunft entschieden werden, ob die Eingabe als erkannt angesehen wird oder der Benutzer auf einen entsprechenden Fehler hingewiesen bzw. eine Rückfrage gestellt werden muss.

Die dritte Maßnahme zur Verbesserung der Spracherkennung ist die Generierung und Verwendung von kontextsensitiven Sprachgrammatiken. Dabei wird für jeden Eingabeschritt im Dialogdesigner eine gesonderte Sprachgrammatik generiert, die alle an dieser Stelle möglichen Eingaben enthält. Damit kann nicht mehr der Fall auftreten, dass fälschlicherweise eine Spracheingabe erkannt wurde, die zum einen nicht wirklich gesagt wurde und zum anderen an diesem Punkt im Dialog nicht sinnvoll ist.

Neben der Spracherkennung, die von Benutzern wahrgenommen werden kann, hat sich bei der Erstellung und Pflege von Inhalten die Funktionsweise des Chatbots AliceBot und der

verwendeten Notation AIML als suboptimal erwiesen. Auch wenn viele konzeptionelle Mängel von AIML durch die Exportfunktion des Dialogeditors abgefangen werden konnten, schränkt AIML die Funktionsvielfalt ein. So lassen sich beispielsweise keine arithmetischen Operationen in AIML abbilden und die Steuerung der Dialoghistorie gestaltet sich sehr aufwändig. Auch wenn der Chatbot keine Komponente ist, die in näherer Zukunft einfach ausgetauscht werden kann, muss dies mittelfristig ins Auge gefasst werden.

Weitere Veränderungen sind im Bereich der Bedienung des Systems angedacht. So führt derzeit beispielweise das Einblenden des Hand-Symbols als Zeichen für die Eingabemöglichkeit mit einem Bedienelement dazu, dass auf jedes bekannte Bedienelement reagiert wird, was zu Verwirrungen führen kann. Beispielsweise erklärt das Avatarsystem, wie ein Assistenzsystem aktiviert wird und zeigt die Hand an, um anzuzeigen, dass es auf die Betätigung des Bedienelements reagiert. Wenn der Benutzer nun die Klimaanlage bedient, beginnt das Avatarsystem, das entsprechende Bedienelement der Klimaanlage zu erklären. Um dieses Fehlverhalten abzustellen, muss während der Laufzeit berücksichtigt werden, welche Eingaben durch Bedienelemente an einem bestimmten Punkt des konzipierten Gesprächsverlaufs möglich sind und nur diese Ereignisse berücksichtigen.

Der Push-to-Talk-Button im Lenkrad kann derzeit nur verwendet werden, um eine Spracheingabe zu signalisieren. In vielen Fällen haben Probanden jedoch versucht, durch eine zweite Betätigung, den Erkennungsprozess abubrechen. Die Umsetzung einer solchen Funktionsweise verspricht somit eine bessere Bedienbarkeit. Von den Probanden wurde zudem explizit das Fehlen einer Pausenfunktion moniert, die es dem Benutzer ermöglicht, in der Erklärung zu pausieren, das Erklärte auszuprobieren und dann die Erklärung des Avatarsystems fortzusetzen. Ebenfalls mit der Steuerung des Gesprächsverlaufs ist die Beobachtung verbunden, dass Probanden, die Erfahrungen mit dem MMI von Audi hatten, instinktiv versucht haben, anstelle der Befehle „Zurück“, „Wiederholen“ und „Weiter“ Tasten des MMI zu verwenden.

Weitere Verbesserungen lassen sich beim Umgang mit falsch erkannten bzw. an der Stelle nicht sinnvollen Spracheingaben vornehmen. So muss der Benutzer bei dem Hinweis auf eine nicht verarbeitbare Eingabe manuell wieder in den letzten Kontext mit dem Befehl „Zurück“ wechseln. Dieser Kontextwechsel soll stattdessen automatisiert werden, so dass der Benutzer direkt seine korrigierte oder umformulierte Eingabe vornehmen kann.

Neben dem direkten Handling des Avatarsystems fehlen noch Möglichkeiten, Optionen durch den Benutzer einstellen zu lassen. So würde es dem Benutzer den Umgang erleichtern, persönliche Einstellung hinsichtlich Tempo und Lautstärke der Sprachausgabe, der Anzeige von Avatar und Medien oder dem Ein- und Ausblenden der Symbolleiste vorzunehmen. Das Vorgehen soll dabei analog zum derzeit bereits einstellbaren Wert für die Timeouts zwischen einzelnen Erklärungsschritten erfolgen.

Bevor die genannten technischen und inhaltlichen Erweiterungen vorgenommen werden können, müssen die Komponenten der Laufzeitumgebung ebenso wie der Dialogdesigner für die gewachsene Menge an Inhalten und Funktionen vorbereitet werden. Vor dem Hintergrund der im Vergleich zu einem herkömmlichen PC eingeschränkten Leistungsfähigkeit des

CarPC's ist dieser Schritt notwendig, um das Avatarsystem für ein breiteres Versuchsfeld vorzubereiten.

8.3.1.2. Inhaltliche Weiterentwicklung

Die Erweiterung des Inhalts lässt sich neben dem Umfang im Besonderen auch durch die verbesserte Verarbeitung von Informationen aus dem Gesprächsverlauf vornehmen.

Hinsichtlich des Umfangs der durch das Avatarsystem abgedeckten Informationen muss das Handbuch als vorläufiger Maßstab angesehen werden, so dass das übergeordnete Ziel der inhaltlichen Weiterentwicklung die Umsetzung des gesamten Handbuchs in Dialogform ist. Damit verbunden sollen Informationen nicht nur passiv auf eine Frage des Benutzers hin widergegeben werden, sondern in bestimmten Fällen proaktiv auf den Fahrer eingegangen werden. Wie die ursprüngliche Vision des Systems bereits vorsah, sollen Warn- und Fehlermeldungen vom Avatar selbstständig erklärt werden, so dass der Benutzer sich ein besseres Bild vom Grad der Beeinträchtigung machen kann.

Eine inhaltliche Dimension, die bisher nicht umgesetzt wurde, ist der Smalltalk, in dem bereits jetzt beispielsweise Informationen über die Tageszeit, die Witterungsverhältnisse, das Verhalten des Fahrers oder Fehler im Fahrzeug eingebunden werden könnten. Ebenso könnte eine Inhaltskategorie „Tip des Tages“ eingeführt werden aus der dem Fahrer bei Fahrtantritt ähnlich wie bei Computerprogrammen eine Information angeboten wird (vgl. Shneiderman/Plaisant 2005, 547).

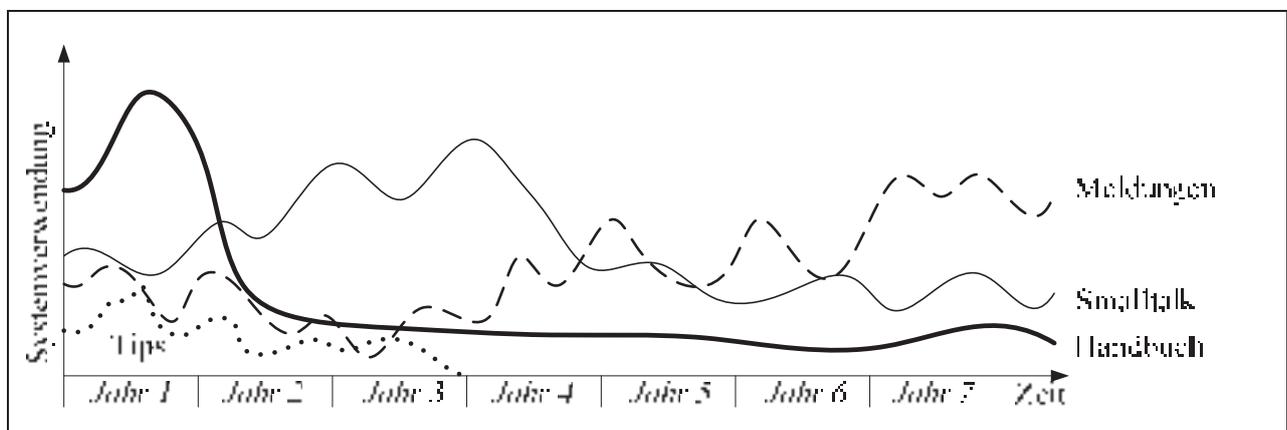


Abbildung 8-1: *Mögliches Nutzungsverhalten nach inhaltlichen Bereichen*
Quelle: (Eigene Darstellung)

Während die Nutzung der Informationen aus dem Handbuch mit der Zeit abnehmen wird, da der Fahrer alle Funktionen kennengelernt hat und nur noch selten nachfragen muss, kann nur durch die Erweiterung des Inhalts um weitere inhaltliche Bereiche eine hohe Nutzung und damit ein Mehrwert für den Benutzer gewährleistet werden. Die Tips des Tages werden dem Fahrer zu Beginn der Fahrzeugnutzung eher störend vorkommen und er wird sie erst nach einer gewissen Zeit bewusst wahrnehmen und nutzen. Nach der Nutzung dieser Inhaltskategorie wird er alle Tips gehört und gegebenenfalls auch verinnerlicht haben, so dass diese Rubrik keine Nutzung mehr erfahren wird. Der Smalltalk mit dem Avatar könnte nach

Bewältigung der elementaren Bedienfunktionen zunehmen und selbst nach längerer Zeit auf einem konstanten Level ablaufen. Die Nutzung dieser Kategorie hängt dabei jedoch von der Anbindung von Informationsquellen und der Gestaltung der entsprechenden Dialoge ab. Der einzige Bereich dessen Nutzung mit den Jahren voraussichtlich zunehmen wird, ist die Erklärung von Warn- und Fehlermeldungen. Während zu Beginn einige Meldungen auf das Fehlverhalten und Unwissen des Fahrers zurückgeführt werden können, können mit zunehmendem Alter des Fahrzeugs verstärkt Probleme auftreten. Ein möglicher Verlauf der bisher vorgestellten inhaltlichen Bereiche ist in Abbildung 8-1 zu sehen.

Durch Auswertung der Videoaufzeichnungen der Evaluation konnten häufig verwendete Frageformen und –formulierungen erkannt werden, die unabhängig vom jeweiligen Bedienelement verwendet wurden. So wurde beispielsweise häufig gefragt „Wo ist der Knopf von X“, „Wie funktioniert X“ oder „Was ist X“ (vgl. dazu auch Zipf's Gesetz in Wallace 2002b). Diese häufigen Fragen sollen systematisch zusammengestellt werden, so dass ein vorgegebener Katalog möglicher Eingaben vorliegt, der bei der Erstellung von Inhalten für ein Bedienelement abgearbeitet werden kann.

Aufgrund der eingeschränkten Rechenleistung der CarPC's konnten 3D-Objekte nicht zur Anzeige verwendet werden, jedoch ist zu erwarten, dass sie durch die Manipulationsmöglichkeiten bei der Anzeige einen Mehrwert für den Benutzer leisten können. Es sollte aus diesem Grund das Ziel sein, 3D-Objekte in das Avatarsystem im Fahrzeug oder zumindest in einer gesonderten Umgebung außerhalb des Autos in größerem Umfang zu integrieren und zu testen.

Die Reaktionsermittlung des Avatarsystems kann durch verschiedene Maßnahmen verbessert werden, indem beispielsweise zu jeder Erklärung eine kurze und eine ausführliche Version einer Antwort erstellt wird. Dabei kann die Länge einer wiedergegebenen Antwort entweder explizit angefordert oder als Option eingestellt werden. Der Vortrag der langen Version ist auch denkbar, wenn der Benutzer sich dahingehend äußert, dass er die kurze Erklärung nicht verstanden hat. Zum besseren Verständnis einer Funktion kann auch bei einer gezielten Frage überprüft werden, ob dem Fahrer bereits die Einleitung zum Bedienelement vorgetragen wurde und dies gegebenenfalls nachgeholt werden. Fragt der Benutzer z.B. danach, wie ein Assistenzsystem aktiviert wird und der Überblick zu diesem Thema wurde noch nicht vorgetragen, so gibt das Avatarsystem zunächst einen Überblick über das Assistenzsystem und erklärt anschließend dessen Aktivierung. Beim nächsten Mal, wenn nach der Aktivierung gefragt wird, wird diese direkt erklärt, da die Einführung bereits erfolgt ist.

Als Ergänzung zu den technischen Lösungen zur Verbesserung der Spracherkennung, kann auch eine kontextsensitive Reaktion auf nicht verarbeitbare Eingaben erfolgen. Wurde das Thema Klimaanlage erklärt und folgt danach eine nicht verarbeitbare Eingabe, so kann das Avatarsystem mögliche Themen im Bereich der Klimaanlage vorschlagen (vgl. dazu auch Hockey et al. 2003).

Die bisher verwendeten Antworten wurden hinsichtlich der Sprachausgabe von manchen Probanden als nicht menschlich bezeichnet, was an der fehlenden Sprachoptimierung lag. Diesem Umstand kann zum einen mit einer manuellen Sprachoptimierung begegnet werden,

die in jedem Fall zu einem gewissen Grad notwendig bleiben wird. Da die Inhalte des Avatarsystems bereits jetzt sehr umfangreich sind und eine Sprachoptimierung somit sehr aufwändig ist, sollte eine teilautomatische Optimierung angestrebt werden, was somit auch ein technisches Thema ist. Erste Vorarbeiten im Bereich der automatischen Prosodiesteuerung wurden dabei bereits vorgelegt (Raidt et al. 2004; Malfrère/Dutoit/Mertens 1998; Saul et al. 2002). Ebenso kann die Einbindung von Avataranimationen insofern teilautomatisiert werden, dass Animationen automatisch auf Basis von Schlüsselwörtern in der Antwort eingebunden werden (vgl. auch Qiu/Benbasat 2007, 344; Kopp/Tepper/Cassell 2004).

8.3.2. Weiterentwicklung von Avatarsystemen im Fahrzeug

Die vorgestellten Verbesserungsvorschläge des AViCoS-Systems können in naher Zukunft durchgeführt werden und stellen hauptsächlich eine weiterführende Umsetzung dar. Darüber hinaus lassen sich jedoch noch weitergehende Entwicklungsoptionen identifizieren, die prinzipiell als Erweiterung eines Avatarsystems im Fahrzeug verwendet werden könnten.

Während der Prototyp des Avatarsystems eine nachträgliche Erweiterung der bestehenden Systeme im Fahrzeug war, ermöglicht die frühzeitige Integration eines Avatarsystems in die Fahrzeugsysteme eine Vielzahl weiterer Optionen. So ermöglicht die Abstimmung der Anzeige von Avatar und anderen Funktionen des MMI die gleichzeitige Darstellung beider Systeme. Dadurch lassen sich die Erklärungen auch auf Bereiche des MMI erweitern und können durch eine funktionale Integration zu den Aktivitäten des Benutzers bessere Rückmeldungen geben. Alternativ ist auch die Anzeige in einem Head-up Display auf der Innenseite der Windschutzscheibe denkbar, wobei die Ablenkung durch einen animierten Avataren im direkten Sichtfeld vermutlich größer wäre. Das detaillierte Wissen über den Zustand anderer Geräte im Fahrzeug erlaubt auch das situationsabhängige, automatische Einblenden des Avatars, wenn erkannt wird, dass der Benutzer Probleme bei der Benutzung hat oder im Begriff ist, eine unvorteilhafte oder gefährliche Einstellung vorzunehmen.

Die enge Verbindung mit bestehenden Geräten im Fahrzeug erlaubt neben dem detaillierten Eingehen auf die aktuelle Bediensituation des Benutzers auch die Steuerung der Funktionen. So kann das Avatarsystem nicht nur Informationen über ein bestimmtes Bedienelement geben, sondern es nach Rückfrage beim Benutzer auch selbstständig bedienen. Nach Maes (1994) ändert dies damit das Assistenzparadigma des Avatarsystems von „tell me“ zu „do it“. Der Eingriff durch das Avatarsystems ist zwar technisch bereits im Prototyp möglich, wurde jedoch aus Sicherheitsgründen bislang nicht vorgenommen.

Ein weiterer Punkt, dessen technische Umsetzung derzeit noch im Forschungsstadium ist, stellt die Voice Activity Detection dar, bei der automatisch das Sprechen eines Benutzers erkannt werden kann. Während insgesamt bereits gute Ergebnisse erzielt werden (vgl. Singh/Boland 2007) ist die Anwendung im Fahrzeug noch problematisch, da Störgeräusche und weitere Insassen die Erkennung erschweren (Lavery 2005). Mögliche Lösungen könnten die Kombination mit einer Stimm und damit Sprechererkennung sein, um den Fahrer aus allen akustischen Signalen heraus erkennen zu können oder der Einsatz einer Kamera zur visuellen Spracherkennung anhand von Lippenbewegungen. Die Umsetzung eines solchen Mechanismus würde die Umsetzung von Dialogen mit verteilter Initiative ermöglichen, die

wesentlich menschlicher wirken (Ishizaki/Crocker/Mellish 1999; Glass/Seneff 2003). Der Benutzer müsste dadurch keinen Push-to-Talk-Button mehr drücken, um seine Sprechabsicht zu signalisieren. Eine kontinuierliche Spracherkennung erkennt, wenn er etwas sagen möchte und unterbricht so lange die Ausführungen des Avatarsystems. Je nach dem, was der Benutzer geäußert hat, wird die Ausführung fortgesetzt oder zu einem neuen Thema gesprungen.

Neben technischen Erweiterungen, die durch den Benutzer nur marginal wahrgenommen werden, stellt die Personalisierung des Avatarsystems einen wichtigen Punkt dar. Sowohl die Auswahl eines vorgegebenen Avatars als auch die Anpassung bzw. eigenständige Gestaltung des Avatars können die Akzeptanz durch den Benutzer steigern. Ein solches Bild ergab sich auch im Rahmen der Evaluation, bei der insbesondere die weiblichen Teilnehmer die fehlende Möglichkeit bemängelten, einen männlichen Avatar zu wählen. Die Personalisierung des Avatarsystems macht jedoch nicht nur beim Avatar halt; so sind auch die Auswahl einer Stimme für die Sprachausgabe, eine bestimmte Wortwahl oder das Design der Benutzeroberfläche mögliche Ansatzpunkte. Neben den visuellen und akustischen Anpassungen kann z.B. der Name des Benutzers angegeben werden, um eine persönliche Ansprache durch den Avatar bewerkstelligen zu können. Glass und Seneff (2003) schlagen dazu ein Vorgehen namens „speak and spell“ vor, bei dem der Benutzer neue Wörter Buchstabe für Buchstabe angeben kann. Die dynamische Sprachsynthese kann den Eigennamen danach sofort wiedergeben. Dieses Vorgehen kann dabei nicht nur für die Personalisierung des Avatarsystems, sondern beispielsweise für die Eingabe eines Straßennamens im Navigationsgerät oder die Benennung des Lieblingsradiosenders verwendet werden. Neben der benutzerinitiierten Anpassung von Einstellung ist auch die systeminitiierte Anpassung ein gangbarer Weg. So kann aufgrund von Eingaben und Verhalten des Benutzers ein Profil von ihm erstellt werden, dass zur Ausgabe angepasster Ausgaben und Hinweise führen kann (Stegmann 2006; Braun et al. 2003).

Mit der Absicht, das Avatarsystem im Fahrzeug mithilfe ausgedehnter Inhalte über Smalltalk zum persönlichen Begleiter des Fahrers zu machen, stellt die Mobilität des Avatars einen weiteren logischen Schritt dar. So können relevante Inhalte des Systems auf mobile Geräte übertragen werden, so dass jederzeit überall der gleiche Gesprächsstand besteht. Auch die spätere Übertragung auf einen Heimcomputer stellt dabei eine Option dar. Diese Idee mag wie eine technische Spielerei anmuten; sie hat jedoch einige bereits absehbare, positive Effekte. So kann das Avatarsystem mit seiner Visualisierung an Orte des Fahrzeugs mitgenommen werden, von denen aus das zentrale MMI-Display nicht eingesehen werden kann wie z.B. im Motorraum, im Kofferraum oder beim Reifenwechseln. Durch eine kabellose Verbindung mit dem zentralen Avatarsystem im Fahrzeug ermöglicht dies zudem die Fernabfrage von Informationen oder die Aktivierung bestimmter Funktionen wie z.B. der Standheizung.

Neben der bereits dargestellten Abdeckung vieler Inhalte mit Fahrzeugbezug stellt die Vernetzung mit anderen Informationsquellen ebenfalls einen Mehrwert für den Fahrer dar. So kann das Avatarsystem als Zugang zu Wetterdaten, location based services, den eigenen E-Mails oder dem Terminkalender eingesetzt werden (vgl. dazu auch Mercedes Scene 2007). Die Anbindung an die Bussysteme des Fahrzeugs erlaubt zudem die Bereitstellung einer Benutzerschnittstelle zu weiteren Funktionen, die keine Bedien- oder Anzeigeelemente

besitzen. So könnte der Zugriff auf bestehende, erweiterte Diagnoseinformationen z.B. die Hilfe bei einer Warn- oder Fehlermeldung maßgeblich verbessern.

Ein auch während der Evaluation genannter Aspekt ist die fehlende Visualisierung abrufbarer Inhalte im Avatarsystem. Während bewusst die Anzeige von Texten im Prototyp vermieden wurde, kann zukünftig beispielsweise durch die grafische Anzeige eines semantic webs mit dem aktuellen Gesprächskontext als Ausgangspunkt eine Hilfe bei der Erkundung von Inhalten erstellt werden. Diese Option würde jedoch die Pflege bzw. Generierung von Metadaten über die Regelbasis zur Erstellung des semantic web notwendig machen.

Abschließend sei die Herausforderung der Mehrsprachigkeit zu nennen, die im Rahmen des Prototyps nicht angegangen wurde. Aufgrund des starken Bezugs der entwickelten Inhalte zur Sprache, stellt die Mehrsprachigkeit bei einem Chatbot einen großen Aufwand dar (Churchill et al. 2000, 75). Neben fehlenden Erkenntnissen über die Gestaltung und Funktion von Chatbots in bestimmten Sprachen erschwert auch das Fehlen eines Konvertierungsmodells das Vorgehen. Darüber hinaus muss speziell im Automobilbereich berücksichtigt werden, dass sich Fahrzeugmodelle von Land zu Land unterscheiden können und damit auch die Erklärungen grundlegend verschieden aufgebaut sein müssen.

Literaturverzeichnis

- Aberg, J.; Shahmehri, N. (2000):** The role of human Web assistants in e-commerce: An analysis and a usability study. In: Internet Research: Electronic Networking Applications and Policy, Vol. 10 (2000) Nr. 2, S. 114-125.
- Aberg, J.; Shahmehri, N. (2001):** An empirical study of human Web assistants: Implications for user support in Web information systems. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, Seattle, WA 2001, S. 404-411.
- Aberg, J.; Shahmehri, N. (2003):** Live help systems. In: Human Factors and Web Development. Hrsg.: Ratner, J. Ed. Laurence Erlbaum Associates, Mahwah, NJ 2003, S. 287-309.
- Ahlers, J.; Oel, P. (2005):** Design von Sprachbediensystemen. In: automotive, (2005) Nr. 07-08, S. 36-38.
- Aimless, D. (2003):** A Tutorial for adding knowledge to your robot. In: <http://www.pandorabots.com/pandora/pics/tutorial/en/tutorial.htm>, zugegriffen am 6.8.2007.
- Airaksinen, T.; Aminoff, H.; Byström, E.; Eimar, G.; Mata, I.; Schmidt, D. (2004):** Automatic Parallel Parking Assistance System User Interface Design – Easier Said Than Done? In: <http://www.ida.liu.se/~HKGBB5/rapporter-04/grupp4.pdf>, zugegriffen am 10.05.2008.
- Akyol, S.; Libuda, L.; Kraiss, K.-F. (2001):** Multimodale Benutzung adaptiver Kfz-Bordsysteme. In: Kraftfahrzeugführung. Hrsg.: Jürgensohn, T.; Timpe, K.-P. Springer-Verlag, Berlin 2001, S. 137-154.
- Albert, H. (1968):** Traktat über kritische Vernunft, Verlag J.C.B. Mohr, Tübingen 1968.
- Albrecht, I.; Haber, J.; Seidel, H.-P. (2002):** Speech Synchronization for Physics-based Facial Animation. In: Proceedings of the 10th International Conferences in Central Europe on Computer Graphics, Visualization and Computer Vision (WSCG 2002), Plzen, Czech Republic 2002.
- Alexa, M.; Berner, U.; Hellenschmidt, M.; Rieger, T. (2001):** An animation system for user interface agents. In: 9th International Conference in Central Europe on Computer Graphics, Visualization and Computer Vision (WSCG 2001), Pilsen, Czech. Republic 2001, S. 154-160.
- Alexandersson, J.; Maier, E.; Reithinger, N. (1994):** The Dialogue Processing Component. Deutsches Forschungszentrum für Künstliche Intelligenz, 1994.
- Allbeck, J.M.; Badler, N.I. (1998):** Avatars á la Snow Crash. In: Proceedings of the Computer Animation 1998, S. 19-24.
- Allen, J.; Hunnicutt, M.S.; Klatt, D.H.; Armstrong, R.C.; Pisoni, D.B. (1987):** From Text To Speech: The MITTALK System, Cambridge University Press, New York 1987.
- Allen, J.F.; Schubert, L.K.; Ferguson, G.; Heeman, P.; Hwang, C.H.; Kato, T.; Light, M.; Martin, N.G.; Miller, B.W.; Poesio, M.; Traum, D.R. (1995):** The TRAINS Project: A case study in building a conversational planning agent. In: Journal of Experimental and Theoretical AI (JETAI), Vol. 7 (1995), S. 7-48.
- Allport, G.W. (1961):** Pattern and growth in personality, Rinehart & Winston, New York 1961.
- Alm, N.; Murray, I.R.; Arnott, J.L.; Newell, A.F. (1993):** Pragmatics and affect in a communication system for non-speakers. In: Journal of the American Voice I/O Society - Special Issue: People with Disabilities, Vol. 13 (1993), S. 1-15.

- Althoff, F.; McGlaun, G.; Spahn, G.; Lang, M.K. (2001):** Combining multiple input modalities for virtual reality navigation - a user study. In: Proceedings of HCI 2001: 9th International Conference on Human Computer Interaction, New Orleans, Louisiana 2001, S. 47-49.
- Anderl, T. (2005):** Kommunikation für Fahrantriebe - Entwicklungsmethoden am Beispiel eines Hybridantriebs. Dissertation, Technische Universität München 2005.
- Anderson, J.R. (1985):** Cognitive Psychology and its Implications. (2. Aufl.), W.H. Freeman and Company, New York 1985.
- Andersson, E.A.; Breitenbach, S.; Burd, T.; Chidambaram, N.; Houle, P.; Newsome, D.; Tang, X.; Zhu, X. (2001):** Early Adopter VoiceXML, Wrox Press, Birmingham 2001.
- André, E.; Rist, T. (1996):** Coping with temporal constraints in multimedia presentation planning. In: Proceedings of the Thirteenth National Conference on Artificial Intelligence, Portland, Oregon 1996, S. 142-147.
- André, E.; Rist, T.; Müller, J. (1998a):** Integrating Reactive and Scripted Behaviors in a Life-Like Presentation Agent. In: Proceedings of the Second International Conference on Autonomous Agents (Agents '98), Minneapolis, Minnesota, USA 1998a, S. 261-268.
- André, E.; Rist, T.; Müller, J. (1998b):** Webpersona: A life-like presentation agent for the world-wide web. In: Knowledge-Based Systems, Vol. 11 (1998b), S. 25-36.
- André, E.; Rist, T.; van Mulken, S.; Klesen, M.; Baldes, S. (2000):** The Automated Design of Believable Dialogues for Animated Presentation Teams. In: Embodied conversational agents. Hrsg.: Cassell, J.; Sullivan, J.; Prevost, S.; Churchill, E. MIT Press, Cambridge, MA 2000, S. 220-255.
- Arafa, Y.; Charlton, P.; Mamdani, A. (1999):** Engineering Personal Service Assistents with Lifelike Qualities. In: <http://ai.mit.edu/people/jvelas/ebaa99/arafa-ebaa99.ps>, zugegriffen am 14.09.2007.
- ARD (2007):** AusspracheDatenBank (ADB) der ARD. In: http://db.ard.de/abc/CONTENT.ergebnis?p_id=1406&p_typ=eg, zugegriffen am 8.12.2007.
- Aristoteles** Analytica posteriora - Buch I.
- Aschenberger, B.; Weiss, C. (2005):** Phoneme-Viseme Mapping for German Video-Realistic Audio-Visual-Speech-Synthesis. Universität Bonn, 2005.
- Augustin, S. (1990):** Information als Wettbewerbsfaktor: Informationslogistik - Herausforderung an das Management, Verlag TÜV Rheinland, Köln 1990.
- Aust, H. (1998):** Sprachverstehen und Dialogmodellierung in natürlichsprachlichen Informationssystemen,. Dissertation, RWTH Aachen 1998.
- Aust, H.; Oerder, M.; Seide, F.; Steinbiss, V. (1995):** The Philips automatic train timetable information system In: Speech Communication, Vol. 17 (1995) Nr. 3-4, S. 249-262.
- Badler, N. (1997):** Real Time Virtual Humans. *Proceedings of The Fifth Pacific Conference on Computer Graphics and Applications* (S. 4-14). Los Alamitos, CA: IEEE Computer Society Press.
- Badler, N.; Phillips, C.; Webber, B. (1993):** Simulating Humans: Computer Graphics, Animation, and Control, Oxford University Press 1993.
- Badler, N.I.; Bindiganavale, R.; Allbeck, J.; Schuler, W.; Zhao, L.; Palmer, M. (2000):** Parameterized Action Representation for Virtual Human Agents. In: Embodied conversational agents. Hrsg.: Cassell, J.; Sullivan, J.; Prevost, S.; Churchill, E. MIT Press, Cambridge, MA 2000, S. 256-286.
- Badzinski, D.M. (1991):** Children's cognitive representations of discourse: Effects of vocal cues on text comprehension. In: Communication Research, Vol. 18 (1991) Nr. 6, S. 715-736.

- Bailenson, J.N.; Blascovich, J. (2004):** Avatars. In: Encyclopedia of Human-Computer Interaction 2004, S. 64-68.
- Balci, K. (2005):** XfaceEd: Authoring Tool for Embodied Conversational Agents. In: 7th international conference on Multimodal interfaces, Toronto, Italy 2005, S. 208-213.
- Ball, G.; Ling, D.; Kurlander, D.; Miller, J.; Pugh, D.; Skelly, T.; Stankosky, A.; Thiel, D.; Van Dantzich, M.; Wax, T. (1997):** Lifelike Computer Characters: The Persona Project at Microsoft. In: Software Agents. Hrsg.: Bradshaw, J. AAAI/MIT Press, Menlo Park 1997, S. 191-222.
- Ball, J.E.; Ling, D.T.; Pugh, D.; Skelly, T.; Stankosky, A.; Thiel, D. (1994):** ReActor: a system for real-time, reactive animations. In: Conference companion on Human factors in computing systems, Boston, Massachusetts 1994, S. 39-40.
- Balzert, H. (2001):** Lehrbuch der Software-Technik; Software-Entwicklung. (2. Aufl.), Spektrum Akademischer Verlag, Heidelberg, Berlin 2001.
- Baresi, L.; Ghezzi, C. (2004):** Validation of Component and Service Federations in Automotive Software Applications. In: Automotive Software: Connected Services in Mobile Networks - First Automotive Software Workshop ASWSD 2004, San Diego, CA, USA 2004, S. 57-73.
- Bartmann, A. (1995):** Zur Erfassung von "Routine" beim Führen von Kraftfahrzeugen: Eine Feldstudie, Shaker Verlag, Aachen 1995.
- Baskerville, R.L. (1999):** Investigating Information Systems with Action Research. In: Communications of the Association for Information Systems, Vol. 2 (1999) Nr. 3.
- Bates, J. (1994):** The Role of Emotion in Believable Agents In: Communications of the ACM, Vol. 37 (1994) Nr. 7, S. 122-125.
- Bath, C. (2001):** Was können uns Turing-Tests von Avataren sagen? Performative Aspekte virtueller Verkörperungen im Zeitalter der Technoscience. In: Technik und Identität, Bielefeld 2001, S. 79-99.
- Bauer, J. (1988):** Konzepte und Prototypen interaktiver Hilfesysteme. Dissertation, Universität Stuttgart 1988.
- Becker, J.; Holten, R.; Knackstedt, R.; Niehaves, B. (2003):** Forschungsmethodische Positionierung in der Wirtschaftsinformatik - epistemologische, ontologische und linguistische Leitfragen (Arbeitsbericht Nr. 93). Westfälische Wilhelms-Universität Münster, Institut für Wirtschaftsinformatik, 2003.
- Becker, J.; Niehaves, B. (2006):** Epistemological Perspectives on IS Research: A Framework for Analyzing and Systematizing Epistemological Assumptions. In: Special Issue on Philosophy and Epistemology in Information Systems of the Information Systems Journal, (2006).
- Beckstein, C. (1988):** Zur Logik der Logik-Programmierung - Ein konstruktiver Ansatz, Springer-Verlag, Berlin 1988.
- Benbasat, I.; Goldstein, D.K.; Mead, M. (1987):** The Case Research Strategy in Studies of Information Systems. In: Management Information Systems Quarterly, Vol. 11 (1987) Nr. 3, S. 369-386.
- Bengler, K. (1995):** Gestaltung und experimentelle Untersuchung unterschiedlicher Präsentationsformen von Wegleitungsinformationen in Kraftfahrzeugen Dissertation, Universität Regensburg 1995.
- Benoît, C.; Lallouache, T.; Mohamadi, T.; Abry, T. (1992):** A set of Frenchvisemes for visual speech synthesis. In: Talking Machines: Theorie, Models and Designs. Hrsg.: Bailly, G.; Benoît, C.; Sawallis, T. Elsevier, Amsterdam 1992, S. 485-501.
- Bente, G.; Krämer, N.C. (2000):** Virtuelle Gesprächspartner: Psychologische Beiträge zur Entwicklung und Evaluation anthropomorpher Schnittstellen. In: Multimodale

- Interaktion im Bereich der Prozessführung: 42. Fachausschußsitzung Anthropotechnik 2000, S. 29-50.
- Bergmann, U.; Hartmann, H.; Kloke, S.; Weiß, C. (2005):** Betriebsanleitungen für den internationalen Markt, WEKA Media, Kissing 2005.
- Bernsen, N.O.; Buisine, S.; Martin, J.-C. (2005):** Children's Gesture and Speech in Conversation with 3D Characters. In: Proceedings of HCI International, Las Vegas 2005.
- Bernsen, N.O.; Dybkjær, H.; Dybkjær, L. (1998):** Designing Interactive Speech Systems: From First Ideas to User Testing, Springer Verlag, Berlin et al. 1998.
- Beskow, J.; McGlashan, S. (1997):** Olga - A Conversational Agent with Gestures. In: IJCAI'97 workshop on Animated Interface Agents - Making them Intelligent, Nagoya, Japan 1997.
- Beutnagel, M.; Conkie, A.; Schroeter, J.; Stylianou, Y.; Syrdal, A. (1999):** The AT&T next-gen TTS system. In: Proceedings of the Joint Meeting of ASA, EAA, and DAGA, Berlin, Germany 1999, S. 18-24.
- Bilvi, M.; Pelachaud, C. (2003):** Communicative and Statistical Eye Gaze Predictions. In: Proceedings of the Second International Joint Conference on Autonomous Agents (AAMAS 2003), Melbourne, Australia 2003.
- Biocca, F. (1997):** The cyborg's dilemma: Progressive embodiment in virtual environment. In: Journal of Computer-Mediated Communication, Vol. 3 (1997) Nr. 2.
- Biocca, F.; Harms, C.; Borggon, J.K. (2003):** Towards a more robust theory and measure of social presence: Review and suggested criteria. In: Presence: Teleoperators and Virtual Environments, Vol. 12 (2003) Nr. 5, S. 456-480.
- Blens, H.; Krämer, N.C.; Bente, G. (2003):** Virtuelle Verkäufer: Die Wirkung von anthropomorphen Interface Agenten in WWW und e-commerce. In: Mensch und Computer 2003: Interaktion in Bewegung. Hrsg.: Ziegler, J.; Szwillus, B.G. Teubner, Stuttgart 2003, S. 297-308.
- Block, N. (2004):** The Mind as the Software of the Brain. In: <http://web.comlab.ox.ac.uk/oucl/research/areas/ieg/e-library/sources/msb.pdf>, zugegriffen am 1.1.2008.
- Blumberg, B.; Galyean, T.A. (1995):** Multi-level direction of autonomous creatures for real-time virtual environments. In: 22nd annual conference on Computer graphics and interactive techniques 1995, S. 47-84.
- Böcker, M.; Holz auf der Heide, B. (2005):** Increasing Ease and Fun of Mobile Phone Use through Avatars. In: Proceedings of HCI International, Las Vegas 2005.
- Bolinger, D. (1958):** A theory of pitch accent in English. In: Word, (1958) Nr. 14, S. 109-149.
- Bolt, R.A. (1980):** "Put-That-There": Voice and Gesture at the Graphics Interface. In: ACM SIGGRAPH Computer Graphics, Vol. 14 (1980) Nr. 3, S. 262-270.
- Bolt, R.A. (1987):** The integrated multi-modal interface. In: The Transactions of the Institute of Electronic Information and Communication Engineers, Japan 1987, S. 2017-2025.
- Bortz, J.; Döring, N. (2006):** Forschungsmethoden und Evaluation: für Human- und Sozialwissenschaftler. (4., überarb. Aufl.), Springer, Heidelberg 2006.
- Braun, F. (2007):** BNF. In: <http://www-cgi.uni-regensburg.de/~brf09510/grammartypes.html>, zugegriffen am 21.12.2007.
- Braun, R.; Nicolescu, V.; Krcmar, H.; Schulz, W.F.; Kreeb, M. (2003):** Modern Communication Instruments for the Imparting of Sustainable Management - By the Example of an Avatar. In: The Information Society and Enlargement of the European Union - Proceedings of the 17th International Conference Informatics for Environmental Protection, Cottbus, Germany 2003, S. 94 -101.

- Brennan, S. (1990):** Conversation as direct manipulation: An Iconoclastic View. In: The Art of Human-Computer Interface Design. Hrsg.: Laurel, B. Addison-Wesley, Reading 1990, S. 393-404.
- Breuer, S.; Abresch, J.; Wagner, P.; Stöber, K. (2001):** BLF – Ein Labelformat für die maschinelle Sprachsynthese mit BOSS II. In: Tagungsband Elektronische Sprachsignalverarbeitung ESSV'2001, Bonn 2001, S. 100-106.
- Brouwer, L.E.J. (1907):** Over de grondslagen der wiskunde. Academisch Proefschrift, Maas & van Suchtelen, Amsterdam 1907.
- Brückmann, M. (2007):** CAD-Engineering-Software erlaubt schnelle Entwicklung von Formel-1-Bauteilen. In: <http://www.maschinenmarkt.vogel.de/themenkanaele/digitalefabrik/cadcam/articles/69174/>, zugegriffen am 29.07.2008.
- Bruckmayr, E.; Reker, K. (1994):** Neue Informationstechniken im Kraftfahrzeug. In: Zeitschrift für Verkehrssicherheit, Vol. 40 (1994), S. 12-23.
- Bubb, H. (2003):** Fahrerassistenz: primär ein Beitrag zum Komfort oder für die Sicherheit? In: VDI Berichte, (2003) Nr. 1768, S. 25-44.
- Bühler, K. (2003):** Schön - schnell - schlau: Online Marketing mit Avataren. In: Avatare - Digitale Sprecher für Business und Marketing. Hrsg.: Lindner, C. Springer-Verlag, Berlin et al. 2003, S. 111-120.
- Buld, S.; Krüger, H.-P.; Hoffmann, S.; Kaussner, A.; Tietze, H.; Totzke, I. (2002):** Wirkungen von Assistenz und Automation auf Fahrerzustand und Fahrsicherheit. Interdisziplinäres Zentrum für Verkehrswissenschaften an der Universität Würzburg (IZVW), 2002.
- Burnett, G.; Joyner, C. (1997):** An Assessment of Moving Map and Symbol-Basad Route Guidance Systems. In: Ergonomics and Safety of Intelligent Driver Inferfaces. Hrsg.: Noy, I.Y. Lawrence Erlbaum, Mahwah, NJ 1997, S. 115-137.
- Burrell, G.; Morgan, G. (1979):** Sociological Paradigms and Organizational Analysis, Heinemann, London 1979.
- Buschmann, M. (2003):** Strategien für Dialogführungssysteme - Automation der Kundenkommunikation im Kontaktkanal Internet. In: Avatare - Digitale Sprecher für Business und Marketing. Hrsg.: Lindner, C. Springer-Verlag, Berlin et al. 2003, S. 95-108.
- Bush, N. (2006):** Getting Started with Program D. In: http://www.aitools.org/Getting_Started_with_Program_D, zugegriffen am 25.02.2008.
- Büttemeyer, W. (1995):** Wissenschaftstheorie für Informatiker, Spektrum Akademischer Verlag, Heidelberg, Berlin, Oxford 1995.
- Cahn, J.E. (1990):** The generation of affect in synthesized speech. In: Journal of the American Voice I/O Society, Vol. 8 (1990), S. 1-19.
- Cañamero, D. (1997):** Modeling motivations and emotions as a basis for intelligent behavior. In: Proceedings of the 1st International Conference on Autonomous Agents, New York 1997, S. 148-155.
- Caporeal, L.R.; Heyes, C.M. (1996):** Why antropomorphize? Folk psychology and other stories. In: Anthropomorphism, Anecdotes, and Animals. Hrsg.: W.Mitchell, R.; Thompson, N.S.; Miles, H.L. University of New York Press, Albany 1996, S. 59-73.
- Carlson, R.; Hunnicutt, S. (1996):** Generic and domain-specific aspects of the Waxholm NLP and dialog modules. In: Proceedings of the Fourth International Conference on Spoken Language, Philadelphia, PA 1996, S. 677-680.
- Carroll, J.; Kellogg, W. (1989):** Artifact as Theory Nexus: Hermeneutics Meets Theory-Based Design. *Proceedings of the CHI '89*: ACM Press.

- Carstensen, K.-U.; Ebert, C.; Endriss, C.; Jekat, S.; Klabunde, R.; Langer, H. (Hrsg.) (2004):** Computerlinguistik und Sprachtechnologie – Eine Einführung (2. überarbeitete und erweiterte Aufl.). spektrum-Verlag, Heidelberg 2004.
- Cassell, J. (2000a):** Embodied conversational interface agents. In: Communications of the ACM, Vol. 43 (2000a) Nr. 4, S. 70-78.
- Cassell, J. (2000b):** Nudge Nudge Wink Wink: Elements of Face-to-Face Conversation for Embodied Conversational Agents. In: Embodied Conversational Agents. Hrsg.: Cassell, J.; Sullivan, J.; Prevost, S.; Churchill, E. The MIT Press, Cambridge 2000b, S. 1-28.
- Cassell, J. (2001):** Embodied Conversational Agents: Representation and Intelligence in User Interface. In: AI magazine, Vol. 22 (2001) Nr. 3, S. 67-83.
- Cassell, J.; Bickmore, T. (2000):** External manifestations of trustworthiness in the interface. In: Communications of the ACM, Vol. 43 (2000) Nr. 12, S. 50-56.
- Cassell, J.; Bickmore, T.; Billinghurst, M.; Campbell, L.; Chang, K.; Vilhjálmsón, H.; Yan, H. (1998):** An Architecture for Embodied Conversational Characters. In: First Workshop on Embodied Conversational Characters, Tahoe City, California 1998, S. 21-30.
- Cassell, J.; Bickmore, T.; Campbell, L.; Vilhjálmsón, H.; Yan, H. (2001):** The conversational Humanoid. In: <http://www.media.mit.edu/gnl/projects/humanoid/>, zugegriffen am 10.07. Access.
- Cassell, J.; Stone, M. (1999):** Living Hand to Mouth: Psychological Theories about Speech and Gesture in Interactive Dialogue Systems. In: AAAI 1999 Fall Symposium on Psychological Models of Communication in Collaborative Systems, North Falmouth, MA 1999, S. 34-42.
- Cassell, J.; Thórisson, K.R. (1999a):** The Power of a Nod and a Glance: Envelope vs. Emotional Feedback in Animated Conversational Agents. In: International Journal of Applied Artificial Intelligence, Vol. 13 (1999a) Nr. 4-5, S. 519-538.
- Cassell, J.; Thórisson, K.R. (1999b):** The power of a nod and a glance: Envelope vs. emotional feedback in animated conversational agents. In: Journal of Applied Artificial Intelligence, Vol. 13 (1999b) Nr. 4-5, S. 519-538.
- Cassell, J.; Torres, O.; Prevost, S. (1999):** Turn Taking vs. Discourse Structure: How Best to Model Multimodal Conversation. In: Machine Conversations. Hrsg.: Wilks, Y. Kluwer, The Hague 1999, S. 143-154.
- Cassell, J.; Vilhjálmsón, H.H.; Bickmore, T. (2001):** BEAT: the Behavior Expression Animation Toolkit. In: Proceedings of SIGGRAPH 2001, Los Angeles, CA, USA 2001, S. 477-486.
- Cavalluzzi, A.; De Carolis, B.; Pizzutilo, S.; Cozzolongo, G. (2004):** Interacting with embodied agents in public environments. In: Working conference on Advanced visual interface, Gallipoli, Italy 2004, S. 240-243.
- Chai, J.; Lin, J.; Zadrozny, W.; Ye, Y.; Stys-Budzikowska, M.; Horvath, V.; Kambhatla, N.; Wolf, C. (2001):** The Role of a Natural Language Conversational Interface in Online Sales: A Case Study. In: International Journal of Speech Technology, Vol. 4 (2001), S. 285-295.
- Chalmers, A.F. (2001):** Wege der Wissenschaft. (5 Aufl.), Springer, Berlin 2001.
- Charamel (2004a):** CharActor: CharactAriser Manual. Charamel GmbH, 2004a.
- Charamel (2004b):** CharaScript 3.5. Charamel GmbH, 2004b.
- Charamel (2004c):** Integration of Virtual Characters in Web-Pages with CharActor. Charamel GmbH, 2004c.

- Cheyner, A.; Julia, L. (1999):** InfoWiz: An Animated Voice Interactive Information System. In: Proceedings of the Workshop on Communicative Agents(Agents '99), Seattle 1999.
- Childers, D.; Cox, R.V.; DeMori, R.; Furi, S.; Juang, B.-H.; Mariani, J.J.; Price, P.; Sagayama, S.; Sondhi, M.M.; Weischedel, R. (1998):** The past, present, and future of speech processing. In: IEEE Signal Processing Magazine, Vol. 15 (1998) Nr. 3, S. 24-28.
- Chin, D. (1989):** KNOME: Modeling what the user knows in UC. In: User Models in Dialog Systems. Hrsg.: Kobsa, A.; Wahlster, W. Springer Verlag, London 1989, S. 74-107.
- Chmielewicz, K. (1994):** Forschungskonzeptionen der Wirtschaftswissenschaft. (3. Aufl.), Schäffer-Pöschel, Stuttgart 1994.
- Chodura, H.; Hofmann, P.-M.; Kalusche, B.; Knoblach, J.; Spohr, J.; Weber, T. (2004):** Standardisierung im Automotive-Umfeld. In: Elektronik Automotive, Vol. 4 (2004), S. 48-52.
- Christ, S.; Baur, A. (2003):** Die Rolle des Fahrers bei der Entwicklung neuer Infotainment- und Fahrerassistenz-systeme im Automobil. In: VDI Berichte, (2003) Nr. 1768, S. 257-269.
- Churchill, E.F.; Cook, L.; Hodgson, P.; Prevost, S.; Sullivan, J.W. (2000):** "May I Help You?": Designing Embodied Conversational Agent Allies. In: Embodied conversational agents. Hrsg.: Cassell, J.; Sullivan, J.; Prevost, S.; Churchill, E. MIT Press, Cambridge, MA 2000, S. 64-94.
- Cisco Systems Inc.; Comverse Inc.; Intel Corporation; Microsoft Corporation; Philips Electronics N.V.; SpeechWorks International Inc. (2002):** Speech Application Language Tags (SALT). In: <http://www.saltforum.org/saltforum/downloads/SALT1.0.pdf>, zugegriffen am 04.01.2008.
- Cohen, J. (1988):** Statistical power analysis for the behavioral sciences. (2. Aufl.), Lawrence Erlbaum Associates, Hillsdale 1988.
- Cohen, M.M.; Massaro, D.W. (1993):** Modeling Coarticulation in Synthetic Visual Speech. In: Models and Techniques in Computer Animation. Hrsg.: Thalmann, N.M.; Thalmann, D. Springer, Tokyo 1993, S. 141-155.
- Cohen, P.R. (1992):** The role of natural language in a multimodal interface. In: Proceedings of the 5th annual ACM symposium on User interface software and technology, Monterey, California, USA 1992, S. 143-149.
- Coker, C.H. (1985):** A Dictionary-Intensive Letter-to-Sound Program. In: Journal of the Acoustical Society of America Supplement 1, Vol. 78 (1985) Nr. 7.
- Comité Européen de Normalisation (1996):** Ergonomische Anforderungen für Bürotätigkeiten mit Bildschirmgeräten, Teil10: Grundsätze der Dialoggestaltung (DIN EN 9241-10) 1996.
- Crotty, M. (1998):** The Foundations of Social Research: Meaning and perspective in the research process, Sage Publications, London 1998.
- Daft, R.L.; Lengel, R.H. (1984):** Information richness: A new approach to managerial behavior and organization design. In: Research in Organizational Behavior, Vol. 6 (1984), S. 191-223.
- Daft, R.L.; Lengel, R.H.; Trevino, L.K. (1987):** Message equivocality, media selection, and manager performance: Implications for information systems. In: MIS Quarterly, Vol. 11 (1987) Nr. 3, S. 355-366.
- Daginnus, M.; Marx, D.; Belschner, R.; Barbehön, K.; Esch, S.; Spohr, J. (2007):** Herstellerinitiative Software. In: http://www.automotive-his.de/download/HIS_Praesentation_2007_v13.pdf, zugegriffen am 19.04.2008.

- Davis, K.H.; Buiddulph, R.; Balashek, S. (1952):** Automatic recognition of spoken digits. In: The Journal of the Acoustical Society of America, Vol. 24 (1952) Nr. 6, S. 637-642.
- de Angeli, A.; Gerbino, W.; Nodari, E.; Petrelli, D. (1999):** From tools to friends: Where is the borderline? In: Proceedings of the UM'99 Workshop on Attitude, Personality and Emotions in User-Adapted Interaction, Banff, Canada 1999, S. 1-10.
- de Angeli, A.; Johnson, G.I.; Coventry, L. (2001):** The unfriendly user: exploring social reactions to chatterbots. In: Proceedings of The International Conference on Affective Human Factors Design 2001.
- den Os, E.; Boves, L.; Lamel, L.; Baggia, P. (1999):** Overview of the ARISE project. In: Sixth European Conference on Speech Communication and Technology (Eurospeech '99), Budapest, Hungary 1999, S. 1527-1530.
- Deray, K. (2001):** Avatars: A Shifting Interaction. In: Pan-Sydney Area Workshop on Visual Information Processing (VIP2001) - Volume 11, Sydney, Australia 2001, S. 129-138.
- Dettweiler, H. (1981):** An approach to demisyllable speech synthesis of German words. In: Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP '81) 1981, S. 110-113.
- DFKI (2007):** MaryXML. In: <http://mary.dfki.de/documentation/maryxml>, zugegriffen am 9.12.2007.
- Diaper, D.; Shelton, T. (1989):** Dialogues with the Tin Man: A Natural Language Grammar for Expert System Naive Users. In: Recent developments and applications of natural language processing. Hrsg.: Peckham, J. Kogan Page, London 1989, S. 98-116.
- Dingler, H. (1931):** Der Zusammenbruch der Wissenschaft und der Primat der Philosophie. (2. Aufl.), Ernst Reinhardt, München 1931.
- Döring, N. (1999):** Sozialpsychologie des Internet: Die Bedeutung des Internet für Kommunikationsprozesse, Identitäten, soziale Beziehungen und Gruppen, Hogrefe, Göttingen 1999.
- Dowding, J.; Ganron, J.M.; Appelt, D.; Bear, J.; Cherny, L.; Moore, R.; Moran, D. (1993):** Gemini: a natural language system for spoken language understanding. In: Proceedings of the 31st Annual Meeting of the ACL, Columbus, OH 1993, S. 54-61.
- Draxler, C. (1995):** Introduction to the Verbmobil-PhonDat Database of Spoken German. In: Proceedings of the Practical Applications of Prolog Conference, Paris 1995.
- Drews, P.; Frees, W.; Schellhase, J. (2006):** Entwicklungstendenzen im Bereich der Fahrzeugelektronik - Generische Plattform für Automotive und Automation. In: Wirtschaftsinformatik, Vol. 48 (2006) Nr. 2, S. 128-132.
- Dryer, D.C. (1997):** Ghosts in the machine: Personalities for socially adroit software agents (Technical Report FS-97-02). AAAI Press, 1997.
- Dutoit, T. (1997):** High-Quality Text-to-Speech Synthesis : an Overview. In: Journal of Electrical & Electronics Engineering, Australia: Special Issue on Speech Recognition and Synthesis, Vol. 17 (1997) Nr. 1, S. 25-37.
- Dybkjaer, L.; Bernsen, N.O. (2000):** Usability issues in spoken dialogue systems. In: Natural Language Engineering, Vol. 6 (2000) Nr. 3-4, S. 243-271.
- Eckert, W. (1996):** Gesprochener Mensch-Maschine-Dialog, Shaker Verlag 1996.
- Elting, C.; Rapp, S.; Möhler, G.; Strube, M. (2003):** Architecture and Implementation of Multimodal Plug and Play. In: Proceedings of the 5th international conference on Multimodal interfaces, Vancouver 2003, S. 93-100.
- Endsley, M.R.; Kaber, D.B. (1999):** Level of automation effects on performance, situation awareness and workload in a dynamic control task. In: Ergonomics, (1999) Nr. 42, S. 462-492.

- Erickson, T. (1997):** Designing Agents as if People Mattered. In: Software Agents. Hrsg.: Bradshaw, J. AAAI/MIT Press, Menlo Park 1997, S. 79-96.
- Evans, G.; Belsey, J.; Blenkhom, P. (2000):** A Multi-Lingual Speech Synthesiser for Blind People. In: IEE Seminar on Speech and Language Processing for Disabled and Elderly People 2000, S. 3/1-3/4.
- Ezzat, T.; Poggio, T. (1998):** MikeTalk: A Talking Facial Display Based on Morphing Visemes. In: Computer Animation Conference, Philadelphia, Pennsylvania 1998, S. 96-102.
- Färber, B. (1987):** Geteilte Aufmerksamkeit: Grundlagen und Anwendung im motorisierten Straßenverkehr (Vol. 20), Verlag TÜV Rheinland, Köln 1987.
- Fastenmeier, W. (1995):** Die Verkehrssituation als Analyseeinheit im Verkehrssystem. In: Autofahrer und Verkehrssituation. Neue Wege zur Bewertung von Sicherheit und Zuverlässigkeit moderner Strassensysteme (Vol. 33). Hrsg.: Fastenmeier, W. Verlag TÜV Rheinland, Köln 1995, S. 27-28.
- Fenn, J.; Raskino, M.; Davies, J.; O'Donovan, P.; Fiering, L.; Schlegel, K.; Andrews, W.; Cramoysan, S.; Morrison, S.; Costello, R.; Jopling, E.; Williams, M.; Kolsky, E.; Kitagawa, M.; Tully, J.; Ball, R.J.G.; Bell, T.; De Azevedo Filho, W.A.; Prentice, S. (2007):** Hype Cycle for Human-Computer Interaction (G00148734). Gartner Inc., 2007.
- Fennel, H.; Bunzel, S.; Heinecke, H.; Bielefeld, J.; Fürst, S.; Schnelle, K.-P.; Grote, W.; Maldener, N.; Weber, T.; Wohlgemuth, F.; Ruh, J.; Lundh, L.; Sandén, T.; Heitkämper, P.; Rimkus, R.; Leflour, J.; Gilberg, A.; Virnich, U.; Voget, S.; Nishikawa, K.; Kajio, K.; Lange, K.; Scharnhorst, T.; Kunkel, B. (2006):** Achievements and exploitation of the AUTOSAR development partnership. In: http://www.autosar.org/download/AUTOSAR_Paper_Convergence_2006.pdf, zugegriffen am 10.09.2007.
- Fettke, P.; Loos, P. (2003):** Model Driven Architecture (MDA). In: Wirtschaftsinformatik, Vol. 45 (2003) Nr. 5, S. 555-559.
- Fischer, G. (2001):** User Modeling in Human-Computer Interaction. In: User Modeling and User-Adapted Interaction, Vol. 11 (2001) Nr. 1-2, S. 65-86.
- Fischer, S.; Kießling, W.; Holland, S.; Fleder, M. (2002):** The COSIMA prototype for multi-objective bargaining. In: Proceedings of the first international joint conference on Autonomous agents and multiagent systems: part 3. Hrsg. ACM Press, Bologna, Italy 2002, S. 1364-1371.
- Frank, U. (2003):** Einige Gründe für die Wiederbelebung der Wissenschaftstheorie. In: DBW, Vol. 63 (2003) Nr. 3, S. 278-292.
- Frankfurter Allgemeine Zeitung (2007):** Audi-Audio. In: Frankfurter Allgemeine Zeitung, (21.04.2007).
- Franz, E.A.; Zelaznik, H.N.; Smith, A. (1992):** Evidence of Common Timing Processes in the Control of Manual, Orofacial, and Speech Movements. In: Journal of Motor Behavior, Vol. 24 (1992) Nr. 3, S. 281-287.
- Frötschl, B. (2005):** Vergleich von Sprachsynthese-Systemen mit deutscher Sprachausgabe. In: <http://www.8hertz.com/tts/tts.html>, zugegriffen am 05.05.2008.
- Furui, S. (2001):** Digital Speech Processing, Synthesis, and Recognition. (2. überarbeitete und erweiterte Aufl.), Marcel Dekker Inc., New York, Basel 2001.
- Garcia-Mateo, C.; Zhou, Q.; Lee, C.-H.; Pargellis, A. (1998):** A Voice User Interface Demonstration System for Mexican Spanish. In: Proceedings of the 5th International Conference on Spoken Language Processing, Sydney, Australia 1998.
- Geiser, G. (1985):** Man Machine Interaction in Vehicles. In: ATZ, (1985) Nr. 87, S. 74-77.

- Gentner, D.; Nielson, J. (1996):** The Anti-Mac Interface. In: *Communications of the ACM*, Vol. 39 (1996) Nr. 8, S. 70-82.
- Gibbon, D. (1995):** SAMPA-D-VMlex. In: <http://coral.lili.uni-bielefeld.de/Documents/sampa-d-vmlex.html>, zugegriffen am 31.10.2008.
- Gibbon, D. (1998):** Sprachlaute: Phonologie (Wortphonologie). In: <http://coral.lili.uni-bielefeld.de/Classes/Summer98/Grundkurs98/Vorlesung/grundkursvorlesung/node7.html>, zugegriffen am 31.10.2008.
- Glass, J.; Seneff, S. (2003):** Flexible and Personalizable Mixed-Initiative Dialogue Systems. In: *Human Language Technology Conference, Workshop on Research Directions in Dialogue Processing (HLT-NAACL 2003)*, Edmonton, Canada 2003, S. 19-21.
- Görtz, M. (2006):** „Helen“ – Die persönliche Informationsagentin im Fahrzeug - Usability-Evaluierung eines anthropomorphen Interface-Agenten in einem multimodalen Fahrerinformationssystem der Blaupunkt GmbH. Magisterarbeit, Universität Hildesheim 2006.
- Gregg, D.G.; Kulkarni, U.R.; Vinzé, A.S. (2001):** Understanding the Philosophical Underpinnings of Software Engineering Research in Information Systems. In: *Information Systems Frontiers*, Vol. 3 (2001) Nr. 2, S. 169–183.
- Grice, M.; Baumann, S.; Benz Müller, R. (2007):** GToBI. In: <http://www.uni-koeln.de/phil-fak/phonetik/gtobi/index.html>, zugegriffen am 9.12.2007.
- Grünweg, T. (2007):** Bestseller Bedienungsanleitung: Didaktik im Handschuhfach. In: <http://www.spiegel.de/auto/aktuell/0,1518,479458,00.html>, zugegriffen am 31.10.2008.
- Gstalter, H.; Fastenmeier, W.; Galsterer, H. (1995):** Ein elektronisches Leitsystem im Vergleich mit anderen Formen der Navigation im Fahrzeug. In: *Autofahrer und Verkehrssituation – Neue Wege zur Bewertung von Sicherheit und Zuverlässigkeit moderner Straßenverkehrssysteme (Vol. 33)*. Hrsg.: Fastenmeier, W. Verlag TÜV Rheinland, Köln 1995, S. 123-139.
- Guba, E.G. (1990):** *The Paradigm Dialog*, Sage, Newbury Park 1990.
- Guba, E.G.; Lincoln, Y.S. (1994):** Competing Paradigms in Qualitative Research. In: *Handbook of Qualitative Research*. Hrsg.: Denzin, N.K.; Lincoln, Y.S. Sage Publications, Thousand Oaks, London et al. 1994, S. 105-117.
- Haas, J.; Gallwitz, F.; Kornwachs, J.; Schröder, M. (2004):** Spracherkennung und Sprachdialog: Stand der Technik, Einsatzbeispiele und zukünftige Trends. In: *Design&Elektronik Entwicklerforum*, München 2004.
- Hagen, E.; Said, T.; Eckert, J. (2004):** Spracheingabe im neuen BMW 6er. In: *ATZ*, (2004).
- Halberstadt, J.B.; Niedenthal, P.M.; Kushner, J. (1995):** Resolution of lexical ambiguity by emotional state. In: *Psychological Science*, Vol. 6 (1995) Nr. 5, S. 278-282.
- Haller, R. (2003):** The Display and Control Concept iDrive - Quick Access to All Driving and Comfort Functions. In: *ATZ/MTZ Extra (The New BMW 5-Series)*, (2003), S. 51-53.
- Hamberger, W.; Mauter, G. (2003):** Audi Multi Media Interface (MMI): Neue Spezifikationsmethoden zur interdisziplinären Bedienkonzeptentwicklung. In: *VDI-Berichte: Der Fahrer im 21. Jahrhundert*, (2003) Nr. 1768, S. 217-233.
- Hanrieder, G.; Hamerich, S.W. (2004):** Modeling Generic Dialog Applications for Embedded Systems. In: *Proceedings of the 8th International Conference on Spoken Language Processing (Interspeech 2004)*, Jeju Island, Korea 2004, S. 237-240.
- Harrison, S.M. (1995):** A comparison of still, animated, or non-illustrated on-line help with written or spoken instructions in a graphic user interface. In: *Proceeding of CHI '95 Conference: Human Factors in Computing Systems 1995*, S. 82-89.

- Hartmann, E. (1984):** Lichttechnik und Physiologie des Sehens. In: Der Dunkelheitsunfall. Hrsg.: Gramberg-Danielsen, B.; Hartmann, E.; Giehring, H. Enke-Verlag, Stuttgart 1984.
- Hassenzahl, M.; Seewald, F. (2004):** Vom kritischen Ereignis zum Nutzungsproblem: Die qualitative Analyse in diagnostischen Usability Tests. In: Usability Professionals 2004. Hrsg.: Hassenzahl, M.; Peissner, M. GC-UPA e.V., Stuttgart 2004, S. 142-146.
- Hayes-Roth, B.; van Gent, R. (1997):** Story-Making with Improvisational Puppets. In: Proceedings of the First International Conference on Autonomous Agents, New York 1997, S. 1-7.
- Heeman, P.A.; Allen, J.F. (1997):** Intonational boundaries, speech repairs, and discourse markers: modeling spoken dialog. In: Proceedings of the 35th Annual Meeting of the ACL and the 8th Conference of the European Chapter of the Association for Computational Linguistics, Madrid, Spain 1997, S. 254-261.
- Heidmann, F.; Biesterfeldt, J. (2004):** Akzeptanz und Usability von Sprachapplikationen in Deutschland. Fraunhofer Institut für Arbeitswirtschaft und Organisation, 2004.
- Heinecke, H.; Schnelle, K.-P.; Fennel, H.; Bortolazzi, J.; Lundh, L.; Leflour, J.; Maté, J.-L.; Nishikawa, K.; Scharnhorst, T. (2004):** AuTomotive Open System Architecture – An Industry-Wide Initiative to Manage the Complexity of Emerging Automotive E/E-Architectures. In: Proceedings of the 2004 International Congress on Transportation Electronics, Detroit 2004, S. 325-332.
- Heise Autos (2007):** VW: Avatar „Carla“ hilft bei Bedienung der Radionavigation. In: <http://www.heise.de/autos/artikel/s/4590>, zugegriffen am 15.11.2007.
- heise online (2006):** Frischer Wind für den Spiele-PC In: <http://www.heise.de/newsticker/meldung/print/77644>, zugegriffen am 1.4.2008.
- Heisterkamp, P. (2001):** Linguatronic: Product-Level Speech System for Mercedes-Benz Cars. In: 1st International Conference on Human Language Technology Research (HLT 01), San Diego, CA 2001, S. 1-2.
- Helal, A.; Moore, S.E.; Ramachandran, B. (2001):** Drishti: An Integrated Navigation System for Visually Impaired and Disabled. In: Proceedings of the Fifth International Symposium on Wearable Computers (ISWC'01) 2001, S. 149.
- Hess, W. (2006):** Systeme der akustischen Mensch-Maschine-Kommunikation - Spracheingabe. In: http://www.ikp.uni-bonn.de/dt/lehre/materialien/sammk/sam_2f.pdf, zugegriffen am 31.10.2008.
- Hevner, A.R.; March, S.T.; Park, J.; Ram, S. (2004):** Design Science in Information Systems Research. In: MIS Quarterly, Vol. 28 (2004) Nr. 1, S. 75-105.
- Heyting, A. (1930):** Die formalen Regeln der intuitionistischen Logik. In: Sitzungsberichte der preußischen Akademie der Wissenschaften, phys.-math. Klasse 1930, S. 42-65.
- Higuchi, N.; Hirai, T.; Sagisaka, Y. (1997):** Effect of speaking style on parameters of fundamental frequency contour. In: Progress in Speech Synthesis. Hrsg.: van Santen, J.; Sproat, R.W.; Olive, J.P.; Hirschberg, J. Springer-Verlag, 1997, S. 417-428.
- Hildebrand, A.; Sá, V. (2000):** EMBASSI: Electronic Multimedia and Service Assistance. In: Proceedings of the IMC2000 – Intelligent Interactive Assistance and Mobile Multimedia Computing, Rostock-Warnemünde, Germany 2000.
- Hipp, E.; Schaller, K.-V. (2003):** Fahrerassistenzsysteme für Nutzfahrzeuge: Stand und Ausblick. In: VDI Berichte, (2003) Nr. 1613, S. 49-64.
- Hirschberg, J. (1991):** Using text analysis to predict intonational boundaries. In: Proceedings of Eurospeech 91, Geneve 1991, S. 1275-1278.
- Hirschheim, R.; Klein, H.K. (1989):** Four Paradigms of Information Systems Development. In: Communications of the ACM, Vol. 32 (1989) Nr. 10, S. 1199-1216.

- Hirst, G. (1981):** Anaphora in Natural Language Understanding: A Survey (Vol. 119), Springer-Verlag, Berlin et al. 1981.
- Hockey, B.A.; Lemon, O.; Campana, E.; Hiatt, L.; Aist, G.; Hieronymus, J.; Gruenstein, A.; Dowding, J. (2003):** Targeted Help for Spoken Dialogue Systems: Intelligent Feedback Improves Naive User's Performance. In: Proceedings of the 10th Conference on European Chapter of the Association for Computational Linguistics 2003, S. 147-154.
- Hodges, A. (1992):** Alan Turing: The Enigma, Vintage, London 1992.
- Hof, A. (2007):** Entwicklung eines adaptiven Hilfesystems für multimodale Anzeige-Bedienkonzepte im Fahrzeug. Dissertation, Universität Regensburg 2007.
- Hoffmann, H.; Leimeister, J.M.; Krcmar, H. (2007):** Pilotierung mobiler Dienste im Automobilsektor. In: Mobile Dienste im Auto der Zukunft. Hrsg.: Reichwald, R.; Krcmar, H.; Reindl, S. Eul Verlag, Lohmar 2007, S. 125-205.
- Hohage, K. (2000):** A new way of learning. In: SAP Info, (2000) Nr. 75, S. 55.
- Horn, E.; Reinke, T. (2002):** Softwarearchitektur und Softwarebauelemente: Eine Einführung für Softwarearchitekten, Carl Hanser Verlag, München 2002.
- Hu, J.; Winterboer, A.; Nass, C.I.; Moore, J.D.; Illowsky, R. (2007):** Context & Usability Testing: User-Modeled Information Presentation in Easy and Difficult Driving Conditions. In: Proceedings of the SIGCHI conference on Human factors in computing systems (CHI 2007), San Jose, CA, USA 2007, S. 1343-1346.
- Huber, R. (2004):** Vom einfachen Bussystem zum anspruchsvollen Datennetz. In: Automotive, (2004) Nr. 05-06, S. 47-49.
- Hudetz, W.; Friedewald, M. (2001):** Technische Produktdokumentation im Maschinen- und Anlagenbau: Eine Bestandsaufnahme. Fraunhofer-Institut für Systemtechnik und Innovationsforschung (FhG-ISI), 2001.
- Hunicut, S. (1980):** Grapheme-to-Phoneme rules: a Review (QPSR 2-3). Speech Transmission Laboratory, Royal Institute of Technology, 1980.
- IBM (2007):** IBM Research Demonstrates Innovative 'Speech to Sign Language' Translation System. In: <http://www-03.ibm.com/press/us/en/pressrelease/22316.wss>, zugegriffen am 29.12.2007.
- Institut für Kommunikationsforschung und Phonetik (2001):** IKP Forschung: Phonetik BOMP In: <http://www.ikp.uni-bonn.de/dt/forsch/phonetik/bomp/BOMP.html>, zugegriffen am 12.8.2007.
- International Phonetic Association (2007):** IPA. In: <http://www2.arts.gla.ac.uk/IPA/index.html>, zugegriffen am 8.12.2007.
- Ishizaki, M.; Crocker, M.; Mellish, C. (1999):** Exploring Mixed-Initiative Dialogue Using Computer Dialogue Simulation. In: User Modeling and User-Adapted Interaction, (1999) Nr. 9, S. 79-91.
- Janich, P.; Kambartel, F.; Mittelstraß, J. (1974):** Wissenschaftstheorie als Wissenschaftskritik, Campus Verlag, Frankfurt a. M. 1974.
- Jensen, C.; Farnham, S.; Drucker, S.; Kollock, P. (2000):** The effect of communication modality on cooperation in online environments. In: Proceedings of the SIGCHI Conference on Human factors in computing systems (CHI'00), The Hague, The Netherlands 2000, S. 470-477.
- Johnson, M.P.; Wilson, A.; Blumberg, B.; Kline, C.; Bobick, A. (1999):** Sympathetic interfaces: using a plush toy to direct synthetic characters. In: Proceedings of the SIGCHI conference on Human factors in computing systems: the CHI is the limit, Pittsburgh, Pennsylvania, United States 1999, S. 152-158.
- Kamlah, W.; Lorenzen, P. (1996):** Logische Propädeutik - Vorschule des vernünftigen Redens (3. Aufl.), Verlag J. B. Metzler, Stuttgart 1996.

- Keller, A. (2004):** Dialogsysteme. In: Computerlinguistik und Sprachtechnologie. Hrsg.: Carstensen, K.-U.; Ebert, C.; Endriss, C.; Jekat, S.; Klabunde, R.; Langer, H. Spektrum Akademischer Verlag, Heidelberg 2004, S. 533-539.
- Kempa, M.; Mann, Z.A. (2005):** Model Driven Architecture. In: Informatik-Spektrum, Vol. 28 (2005) Nr. 4, S. 298-302.
- Kennaway, R. (2001):** Synthetic Animation of Deaf Signing Gestures. In: International Gesture Workshop 2001, London 2001, S. 146-157.
- Kiesler, S.; Sproull, L. (1997):** "Social" human-computer interaction. In: Human values and the design of computer technology. Hrsg.: Friedman, B. CSLI Publications/Cambridge University Press, Stanford 1997, S. 191-199.
- Kießling, W.; Fischer, S.; Holland, S.; Ehm, T. (2001):** Design and Implementation of COSIMA - A Smart and Speaking E-Sales Assistant (2001-1). Universität Augsburg, Institut für Informatik, 2001.
- Kimura, K.; Marunaka, K.; Sugira, S. (1997):** Human Factors Considerations for Automotive Navigation Systems: Legibility, comprehension, and Voice Guidance. In: ergonomics and Safety of Intelligent Driver Interfaces. Hrsg.: Ian Noy, Y. Lawrence Erlbaum Associates, Mahwah, N.J. 1997, S. 153-167.
- Kitahara, Y.; Tohkura, Y. (1989):** Prosodic components of speech in the expression of emotion. In: Journal of the Acoustical Society of America, Supplement 1, Vol. 84 (1989).
- Klatt, D.H. (1987):** Review of text-to-speech conversion for English. In: The Journal of the Acoustical Society of America, Vol. 82 (1987) Nr. 3, S. 737-793.
- Klemmer, S.R.; Sinha, A.K.; Chen, J.; Landay, J.A.; Aboobaker, N.; Wang, A. (2000):** SUEDE: A Wizard of Oz Prototyping Tool for Speech User Interfaces. In: Proceedings of the 13th Annual ACM Symposium on User Interface Software and Technology (UIST 2000) 2000, S. 1-10.
- Kleppe, A.; Warmer, J.; Bast, W. (2003):** MDA Explained - The Model Driven Architecture: Practice and Promise, Pearson Education, Boston 2003.
- Knuth, D.E. (1964):** Backus Normal Form versus Backus Naur Form. In: Communications of the ACM, Vol. 7 (1964) Nr. 12, S. 735-736.
- Koch, B. (2004):** Zuverlässige Software fürs Auto. In: Fraunhofer Magazin, (2004) Nr. 4, S. 24-25.
- Koda, T.; Maes, P. (1996):** Agents with Faces: The Effect of Personification. In: Proceedings of the 5th IEEE international workshop on robot and human communication, Tsukuba, Japan 1996, S. 189-194.
- Kopp, S.; Tepper, P.; Cassell, J. (2004):** Towards Integrated Microplanning of Language and Iconic Gesture for Multimodal Output. In: Proceedings of the 6th international conference on Multimodal interfaces, State College, PA, USA 2004, S. 97-104.
- Krahmer, E.; Landsbergen, J.; Pouteau, X. (1997):** How to Obey the 7 Commandments for Spoken Dialogue? In: Proceedings of the ACL/EACL Workshop on Interactive Spoken Dialog Systems, Madrid, Spain 1997, S. 82-89.
- Krämer, N.C.; Bente, G. (2002):** Virtuelle Helfer: Embodied Conversational Agents in der Mensch-Computer-Interaktion. In: Virtuelle Realitäten. Hrsg.: Bente, G.; Petersen, A.; Krämer, N.C. Hogrefe, Göttingen 2002, S. 203-225.
- Krämer, N.C.; Nitschke, J. (2002):** Ausgabemodalitäten im Vergleich: Verändern sie das Eingabeverhalten der Benutzer? In: Bedienen und Verstehen. Tagungsband 4. Berliner Werkstatt Mensch-Maschine-Systeme 2002, S. 231-248.
- Krause, J. (1988):** Adaptierbarkeit, Adaptivität, Intervenierbarkeit und Hilfesysteme (Forschungsbericht). Universität Regensburg, 1988.

- Krause, J. (1992):** Natürlichsprachliche Mensch-Computer-Interaktion als technisierte Kommunikation: Die Computer-Talk-Hypothese. In: Computer Talk. Hrsg.: Krause, J.; Hitzenberger, L. Georg Olms Verlag, Hildesheim 1992, S. 1-29.
- Krcmar, H.; Strasburger, H. (1993):** Informationsmanagement und Informationssystem-Architekturen - Vorteile und Risiken von Client-Server-Architekturen aus der Sicht des Informationsmanagements. In: Client-Server-Architekturen - Herausforderung an das Informationsmanagement. Hrsg.: Krcmar, H. AIT Angewandte Informationstechnik Verlags GmbH, Hallbergmoos 1993, S. 9-30.
- Kshirsagar, S.; Magnenat-Thalmann, N.; Guye-Vuillème, A.; Thalmann, D.; Kamyab, K.; Mamdani, E. (2002):** Avatar Markup Language. In: Workshop on Virtual environments 2002 Barcelona, Spain 2002, S. 169-177.
- Kurlander, D.; Ling, D.T. (1995):** Planning-based control of interface animation. In: Proceedings of the SIGCHI conference on Human factors in computing systems, Denver, Colorado, USA 1995, S. 472-479.
- Labiale, G. (1992):** Driver Characteristics and In-Car Map Display Memory Recall Performance. In: The 3rd International Conference on Vehicle Navigation & Information Systems, Oslo, Norway 1992, S. 227-232.
- Lach, A. (2005):** Sprachverstehen im Störlärm bei 60 und 80dB mit dem HSM-Satztest in der Computerversion bei den 30 besser hörenden Probanden einer Gruppe von 60 Normalhörenden um die 50 Jahre. Dissertation, Julius-Maximilians-Universität 2005.
- Lankes, J.; Matthes, F.; Wittenburg, A. (2005):** Architekturbeschreibung von Anwendungslandschaften: Softwarekartographie und IEEE Std. 1471-2000. In: Softwareengineering 2005, Essen 2005, S. 43-54.
- Laroche, L.; Stylianou, Y.; Moulines, E. (1993):** HNS: Speech modification based on a harmonic + noise model. In: Proceedings of the IEEE International Conference on Acoustics, Speech, Signal Processing (ICASSP), Minneapolis 1993, S. 550-553.
- Laufer, L.; Tatai, G.; Nemeth, B. (2005):** Dialogue Modeling in Embodied Communicational Agents. In: Proceedings of the Symposium on Dialogue Modelling and Generation, Dortmund, Germany 2005.
- Laurel, B. (1997):** Interface Agents: Metaphors with Character. In: Software agents. Hrsg.: Bradshaw, J.M. MIT Press, Cambridge, MA 1997, S. 207-219.
- Laverty, S.W. (2005):** Detection of Nonstationary Noise and Improved Voice Activity Detection in an Automotive Hands-free Environment. Master Thesis, Worcester Polytechnic Institute 2005.
- Lee, A.S. (1999):** Researching MIS. In: Rethinking Management Information Systems. Hrsg.: Currie, W.; Galliers, B. Oxford University Press, New York 1999, S. 7-27.
- Lee, E.-J.; Nass, C. (1998):** Does the Ethnicity of a Computer Agent Matter? An Experimental Comparison of Human-Computer Interaction and Computer-Mediated Communication. In: Proceedings of the WECC Conference, Lake Tahoe, CA 1998.
- Lee, K.; Nass, C. (2003):** Designing social presence of social actors in human computer interaction. In: Proceedings of the ACM SIGCHI Conference on Human Factors in Computing Systems (CHI'03), Ft. Lauderdale, FL 2003, S. 289-296.
- Lee, P.S.; Badler, B.J.; Badler, I.N. (2002):** Eyes Alive. In: ACM Transactions on Graphics, Vol. 21 (2002) Nr. 3, S. 637-644.
- Lefebvre, P.; Duncan, G.; Poirier, F. (1993):** Speaking with computers: a multimodal approach. In: Proceedings of the 3rd European Conference on Speech Communication and Technology, Berlin, Germany 1993, S. 1665-1668.
- Lester, J.; Voerman, J.; Towns, S.; Callaway, C. (1997a):** Cosmo: A life-like animated pedagogical agent with deictic believability. In: Working Notes of the IJCAI '97

- Workshop on Animated Interface Agents: Making Them Intelligent, Nagoya, Japan 1997a, S. 61-69.
- Lester, J.C.; Converse, S.A.; Kahler, S.E.; Barlow, S.T.; Stone, B.A.; Bhogal, R.S. (1997b):** The Persona Effect: Affective Impact of Animated Pedagogical Agents. In: Human factors in computing systems – Proceedings of the CHI '97, Atlanta, Georgia, USA 1997b, S. 359-366.
- Lester, J.C.; Towns, S.G.; Callaway, C.B.; Voennan, J.L.; FitzGerald, P.J. (2000):** Deictic and Emotive Communication in Animated Pedagogical Agents. In: Embodied conversational agents. Hrsg.: Cassell, J.; Sullivan, J.; Prevost, S.; Churchill, E. MIT Press, Cambridge, MA 2000, S. 123-154.
- Levinson, S.E.; Olive, J.P.; Tschirgi, J. (1993):** Speech Synthesis in Telecommunications. In: IEEE Communications Magazine, Vol. 31 (1993), S. 46-53.
- Lewis, J.P. (1991):** Automated Lip-Sync: Background and Techniques. In: Visualization and Computer Animation, (1991) Nr. 2.
- Lieberman, M.Y.; Church, K.W. (1992):** Text analysis and word pronunciation in text-to-speech synthesis. In: Advances in Speech Signal Processing. Hrsg.: Furuy, S.; Sondhi, M.M. Dekker, New York 1992, S. 791-831.
- Libuda, L.; Kraiss, K.-F. (2004):** Dialogassistenz im Kraftfahrzeug (DGLR-Bericht 2003-04). DGLR, 2004.
- Lilienthal, J. (2003):** Erweiterung der Mensch-Maschine-Schnittstelle zum verteilten Bediensystem. In: VDI Berichte, (2003) Nr. 1768, S. 147-170.
- Lindner, C. (2003):** Wer braucht wofür Avatare? Konzeption und Implementierung natürlichsprachlicher Systeme - Zur Einführung. In: Avatare - Digitale Sprecher für Business und Marketing. Hrsg.: Lindner, C. Springer-Verlag, Berlin et al. 2003, S. 5-24.
- Loftin, R.B.; Scerbo, M.W.; McKenzie, F.D.; Catanzaro, J.M. (2004):** Training in Peacekeeping Operations Using Virtual Environments. In: IEEE Computer Graphics and Applications, Vol. 24 (2004) Nr. 4, S. 18-21.
- Lombard, M.; Ditton, T. (1997):** At the heart of it all: The concept of presence. In: Journal of Computer-Mediated Communication, Vol. 3 (1997) Nr. 2.
- Lorenzen, P. (1987):** Lehrbuch der konstruktiven Wissenschaftstheorie, BI Wissenschaftsverlag, Mannheim, Wien, Zürich 1987.
- Lowerre, B.T. (1976):** The Harpy Speech Recognition System. Dissertation, Carnegie-Mellon University 1976.
- Luft, A.L. (1981):** Software-Engineering und konstruktive Wissenschaftstheorie - Ein Beitrag zur Methodologie des Software-Engineering. In: Angewandte Informatik, (1981) Nr. 3, S. 93-99.
- MacGregor, R.; Bates, R. (1987):** The LOOM knowledge representation language (TechReport ISI-87-188). University of Southern California, Information sciences Institute, 1987.
- Maekawa, K. (1999):** Phonetic and phonological characteristics of paralinguistic information in spoken Japanese. In: Proceedings of the 5th International Conference on Spoken Language Processing, Sydney, Australia 1999.
- Maes, P. (1994):** Agents that Reduce Work and Information Overload. In: Communications of the ACM, Vol. 37 (1994) Nr. 7, S. 30-40.
- Malfrère, F.; Dutoit, T.; Mertens, P. (1998):** Fully Automatic Prosody Generator for Text-to-Speech Synthesis. In: International Conference on Speech and Language Processing, Sidney, Australia 1998, S. 1395-1398.
- March, S.T.; Smith, G.F. (1995):** Design and Natural Science Research on Information Technology. In: Decision Support Systems, Vol. 15 (1995) Nr. 4, S. 251-266.

- Marcus, M. (1995):** New trends in natural language processing: statistical natural language processing. In: Voice Communication Between Humans and Machines. Hrsg.: Roe, D.; Wilpon, J. National Academy Press, Washington, DC 1995, S. 482–504.
- Marrenbach, J.; Kraiss, K.-F.; Libuda, L.; Bengler, K. (2001):** Development and Multimodal Operation of a Multimedia Car Instruction Manual. In: Analysis, Design and Evaluation of Human-Machine Systems: Proceedings of the 8th IFAC/IFIP/IFORS/IEA Symposium on Human-Machine Systems 2001.
- Marshall, S. (1998):** Speech Synthesis in Interactive Spoken Dialogue Systems. In: <http://citeseer.ist.psu.edu/cache/papers/cs/2876/http:zSzzSzwww.leafdigital.comzSzSamzSzspeech-essay.pdf/marshall98speech.pdf>, zugegriffen am 31.10.2008.
- Matthews, G.; Desmond, P.A. (1995):** Stress As a Factor in the Design of In-Car Driving Enhancement Systems. In: Travail Humain, Vol. 58 (1995) Nr. 2, S. 109-129.
- Matthews, S.W.; Miller, K.H. (2003):** An open architecture for defense virtual environment training systems. Master Thesis, Naval postgraduate school, Monterey 2003.
- McBreen, H.M.; Anderson, J.; Jack, M. (2001):** Evaluating 3D Embodied Conversational Agents in Contrasting VRML Retail Applications. In: Proceedings of the AAMAS01 Workshop on Representing, Annotating, and Evaluating Non-Verbal and Verbal Communicative Acts to Achieve Contextual Embodied Agents, Montreal, Canada 2001, S. 83-87.
- McCrae, R.R.; John, O.P. (1992):** An introduction to the five-factor model and its implications. In: Journal of Personality and Social Psychology, Vol. 60 (1992), S. 175-215.
- McGlashan, S.; E., B.; Fraser, N.; Gilbert, N.; Heisterkamp, P.; Youd, N. (1990):** Managing oral dialogues (Research Report). Social and Computer Sciences Research Group, University of Surrey, 1990.
- McGurk, H.; MacDonald, J. (1976):** Hearing lips and seeing voices. In: Nature, Vol. 264 (1976) Nr. 5588, S. 746-748.
- McTear, M.F. (2002):** Spoken dialogue technology: enabling the conversational user interface. In: ACM Computing Surveys (CSUR), Vol. 34 (2002) Nr. 1, S. 90-169.
- Medizinische Informationstechnologie (2007):** Einsatz von Diktiersystem und Spracherkennung in der Radiologie. In: <http://www.medizin.uni-tuebingen.de/mit-ima/diktiersystem060504.pdf>, zugegriffen am 22.12.2007.
- Meier-Arendt, G.; Abel, H.-B. (2003):** Zukünftige Fahrerinformationssysteme im Kraftfahrzeug - Ergonomische Optimierung des Human Machine Interfaces (HMI). In: VDI Berichte, (2003) Nr. 1768, S. 75-88.
- Meng, H.; Ching, P.C.; Chan, S.F.; Wong, Y.F.; Chan, C.C. (2004):** ISIS: An Adaptive, Trilingual Conversational System With Interleaving Interaction and Delegation Dialogs. In: ACM Transactions on Computer-Human Interaction (TOCHI), Vol. 11 (2004) Nr. 3, S. 268-299.
- Mercedes Scene (2007):** IAA: F 700. In: http://www.mercedes-scene.de/html/de/content/Websites/TOP_THEMA/IAA_F_700, zugegriffen am 15.11.2007.
- Mercer; Fraunhofer IPA; Fraunhofer IML (2004):** Future Automotive Industry Structure (FAST) 2015 - die neue Arbeitsteilung in der Automobilindustrie (32). Verband der Automobilindustrie e V, 2004.
- Miller, G.A. (1956):** The Magical Number Seven, Plus or Minus Two: Some Limits on Our Capacity for Processing Information. In: Psychological Review, Vol. 63 (1956) Nr. 2, S. 81-97.

- Möbus, C.; Eißner, A.; Feindt, J.; Janßen, C.; Krefeldt, J.; Sieverding, S.; Sölbrandt, S.; Stumpe, J.; de Vries, H.; Willer, S. (2006):** Web-Kommunikation mit OpenSource, Springer Verlag, Berlin, Heidelberg, New York 2006.
- Moffat, D. (1997):** Personality parameters and programs. In: Creating personalities for synthetic actors. Hrsg.: Trappl, R.; Petta, P. Springer-Verlag, New York 1997, S. 120-165.
- Mollenhauer, M.A.; Lee, J.; Cho, K.; Hulse, M.C.; Dingus, T.A. (1994):** The Effects of Sensory Modality and Information Priority on In-Vehicle Signing and Information Systems. In: Proceedings of the Human Factors and Ergonomics Society 38th Annual Meeting, Nashville, Tennessee 1994, S. 1072-1076.
- Moore, R.C. (1995):** Integration of speech with natural language understanding. In: Voice Communication Between Humans and Machines. Hrsg.: Roe, D.; Wilpon, J. National Academy Press, Washington, DC 1995, S. 254-271.
- Morningstar, C.; Farmer, F.R. (1991):** The lessons of Lucasfilm's habitat. In: Cyberspace: first steps. Hrsg.: Benedikt, M. MIT Press, Cambridge, MA, USA 1991, S. 273-302.
- Moulines, E.; Charpentier, F. (1990):** Pitch-synchronous waveform processing techniques for text-to-speech synthesis using diphones. In: Speech Communication, (1990) Nr. 9, S. 453-467.
- Muhr, R. (2007):** Österreichisches Aussprachewörterbuch, Österreichische Aussprachedatenbank - Kurzinformation.
- Murray, I.R.; Arnott, J.L. (1993):** Toward the simulation of emotion in synthetic speech: A review of the literature on human vocal emotion. In: Journal Acoustical Society of America, Vol. 93 (1993) Nr. 2, S. 1097-1108.
- Nagao, K.; Takeuchi, A. (1994):** Speech dialogue with facial displays: Multimodal human-computer conversation. In: Proceedings of the 32nd Annual Meeting of the Association for Computational Linguistics (ACL-94), Las Cruces, Mexico 1994, S. 102-109.
- Nass, C.; Lee, K. (2001):** Does computer-synthesized speech manifest personality? Experimental tests of recognition, similarity-attraction, and consistency-attraction. In: Journal of Experimental Psychology (Applied), Vol. 7 (2001), S. 171-181.
- Nass, C.; Moon, Y.; Fogg, B.J.; Reeves, B.; Dryer, D.C. (1995):** Can computer personalities be human personalities? In: International Journal of Human-Computer Studies, Vol. 43 (1995) Nr. 2, S. 223-239.
- Nass, C.; Steuer, J.; Tauber, E.R. (1994):** Computers are social actors. In: Proceedings of the Conference on Human Factors in Computing Systems, Boston, Massachusetts, USA 1994, S. 72-78.
- Newcomb, D. (2008):** IBM's Drumming Car Reads Your Lips. Seriously. In: <http://blog.wired.com/cars/2008/06/your-future-car.html>, zugegriffen am 26.07.2008.
- Ngugi, K.; Okelo-Odongo, W.; Wagacha, P.W. (2005):** Swahili text-to-speech system. In: African Journal of Science and Technology (AJST), Vol. 6 (2005) Nr. 1, S. 80-89.
- Noma, T.; Badler, N.I. (1997):** A Virtual Human Presenter. In: Proceedings of the IJCAI-97 Workshop on Animated Interface Agents: Making them Intelligent, Nagoya 1997, S. 45-51.
- Nowak, K.; Biocca, F. (2003):** The effect of the agency and body anthropomorphism of virtual humans on users sense of presence, copresence, and social presence. In: Presence, Vol. 12 (2003) Nr. 5, S. 481-494.
- Nuance (2001):** Grammar Developer's Guide.
- Nuance (2008):** VoCon 3200: Next-Generation Speech User Interface for Automotive Applications. In: <http://www.nuance.com/vocon/3200/>, zugegriffen am 22.04.2008.

- Nunamaker, J.F.; Chen, M.; Purdin, T.D.M. (1991):** Systems Development in Information Systems Research. In: Journal of Management Information Systems, Vol. 7 (1991) Nr. 3, S. 89-106.
- ORF (2007):** Aussprache-Wörterbuch des Österreichischen. In: http://www-gewi.uni-graz.at/deuph/meldungen/docs/ORFONScience_Aussprache-Woerterbuch.htm, zugegriffen am 8.12.2007.
- OSGi Alliance (2007a):** About the OSGi Service Platform: Technical Whitepaper. In: <http://www.osgi.org/documents/collateral/OSGiTechnicalWhitePaper.pdf>, zugegriffen am 11.01.2008.
- OSGi Alliance (2007b):** OSGi Service Platform Core Specification. OSGi Alliance, 2007b.
- OSGi Alliance (2008):** OSGi Technology. In: <http://www.osgi.org/About/Technology>, zugegriffen am 20.04.2008.
- Paivio, A. (1979):** Imagery and Verbal Processes, Lawrence Erlbaum Associates, Hillsdale, NJ 1979.
- Palmiter, S.; Elkerton, J. (1991):** An evaluation of animated demonstrations for learning computer-based tasks. In: Proceedings of CHI '91 Conference: Human Factors in Computing Systems, New Orleans, Louisiana 1991, S. 257-263.
- Pandzic, I.S. (1998):** Facial Communication in Networked Collaborative Virtual Environments. Dissertation, Université de Genève 1998.
- Paris, C.L. (1989):** The use of explicit user models in a generation system for tailoring answers to the user's level of expertise. In: User Models in Dialog Systems. Hrsg.: Kobsa, A.; Wahlster, W. Springer Verlag, London 1989, S. 200–232.
- Parkes, A.M.; Coleman, N. (1990):** Route Guidance Systems: A Comparison of Methods of Presenting Directional Information to the Driver. In: Contemporary Ergonomics. Hrsg.: Lovesey, E.J. Tylor & Francis, London 1990, S. 480-485.
- Payne, S.J.; Chesworth, L.; Hill, E. (1992):** Animated demonstrations for exploratory learners. In: Interacting with computers, Vol. 4 (1992) Nr. 1, S. 3-22.
- Peckham, J. (1991):** Speech Understanding and Dialogue over the Telephone: an Overview of the ESPRIT SUNDIAL Project. In: Proceedings of the Speech and Natural Language Workshop, Pacific Grove, CA 1991, S. 14–27.
- Peckham, J. (1993):** A new generation of spoken dialogue systems: Results and lessons from the SUNDIAL project. In: Proceedings of 3rd European Conference on Speech Communication and Technology (Eurospeech'93), Berlin, Germany 1993, S. 33-40.
- Peirce, C.S. (1976):** Schriften zum Pragmatismus und Pragmatizismus, Suhrkamp, Frankfurt am Main 1976.
- Peissner, M.; Sell, D.; Steimel, B. (2006):** Akzeptanz von Sprachapplikationen in Deutschland 2006. Initiative Voice Business, Fraunhofer -Institut für Arbeitswirtschaft und Organisation IAO, 2006.
- Perlin, K.; Goldberg, A. (1996):** Improv: A System for Scripting Interactive Actors in Virtual Worlds. In: Proceedings of the 23rd annual conference on Computer graphics and interactive techniques 1996, S. 205-216.
- Pfaffmann, E. (2000):** Knowledge Maturity of Products, Modularity, and the Vertical Boundaries of the Firm. In: Competence, Governance, and Entrepreneurship: Advances in Economic Strategy Research. Hrsg.: Foss, N.; Mahnke, V. Oxford University Press, Oxford, New York 2000, S. 250-275.
- Pfau, T.; Ruske, G. (1996):** Inkrementelle Erstellung von Worthypothesengraphen (Bericht 107). Verbmobil-Report, 1996.
- Picard, R. (2000):** Affective Computing, MIT Press, Cambridge 2000.
- Piechulla, W.; Mayser, C.; Gehrke, H.; König, W. (2002):** Echtzeit-Fahrerbeanspruchungsschätzung. In: Information und Mobilität. Proceedings des 8.

- Internationalen Symposiums für Informationswissenschaft, Konstanz 2002, S. 365-368.
- Poggi, I.; Pelachaud, C. (2000):** Performative Facial Expressions in Animated Faces. In: Embodied conversational agents. Hrsg.: Cassell, J.; Sullivan, J.; Prevost, S.; Churchill, E. MIT Press, Cambridge, MA 2000, S. 155-188.
- Polifroni, J.; Chung, G. (2002):** Promoting portability in dialogue management. In: Proceedings of the ICSLP, Denver, Co 2002, S. 2721-2724.
- Pompino-Marschall, B. (1992):** PHONDAT. Verbundvorhaben zum Aufbau einer Sprachsignaldatenbank für gesprochenes Deutsch (30). Institut für Phonetik und Sprachverarbeitung, 1992.
- Popper, K. (1994):** Logik der Forschung. (10., verbesserte und vermehrte Auflage Aufl.), Mohr Verlag, Tübingen 1994.
- Potamianos, G.; Neti, C.; Iyengar, G.; Helmuth, E. (2001):** Large-Vocabulary Audio-Visual Speech Recognition by Machines and Humans. In: Proceedings of the Seventh European Conference on Speech Communication and Technology (Eurospeech 2001), Aalborg, Denmark 2001, S. 1027-1030.
- Price, P. (1990):** Evaluation of Spoken Language Systems: the Atis Domain. In: Proceedings of the DARPA Speech and Natural Language Workshop, Hidden Valley, Pennsylvania, USA 1990, S. 91-95.
- Purao, S. (2002):** Design Research in the Technology of Information Systems: Truth or Dare, 2002.
- Qiu, L.; Benbasat, I. (2007):** An Investigation into the Effects of Text-to-Speech Voice and 3D Avatars on the Perception of Presence and Flow of Live Help in Electronic Commerce. In: ACM Transactions on Computer-Human Interaction (TOCHI), Vol. 12 (2007) Nr. 4, S. 329-355.
- Raidt, S.; Bailly, G.; Holm, B.; Mixdorff, H. (2004):** Automatic Generation of Prosody: Comparing Two Superpositional Systems. In: Speech Prosody 2004, Nara, Japan 2004.
- Reeves, B.; Nass, C. (1998):** The Media Equation - How People Treat computers, Television, and New media Like Real People and Places, CSLI Publications, Stanford, California 1998.
- Reilly, S.N. (1996):** Believable Social and Emotional Agents. Dissertation, Carnegie Mellon University 1996.
- Reitter, D.; Covaci, S.; Oltean, F.; Bacanu, C.; Serbanuta, T.F. (2001):** Hybrid Natural Language Processing in a Customer-Care Environment. In: Proceedings of the 11th Student Conference on Computational Linguistics (TaCoS '01), Heidelberg, Germany 2001.
- Reyelt, M.; Grice, M.; Benzmüller, R.; Mayer, J.; Batliner, A. (1996):** Prosodische Etikettierung des Deutschen mit ToBI. In: Natural Language Processing and Speech Technology. Results of the 3rd KONVENS Conference Proceedings, Berlin 1996, S. 144-155.
- Rice, R.E. (1992):** Task analyzability, use of new medium and effectiveness: A multi-site exploration of media richness. In: Organization Science, Vol. 3 (1992) Nr. 4.
- Rickel, J.; Johnson, W.L. (1997):** Integrating pedagogical capabilities in a virtual environment agent. In: Proceedings of the First International Conference on Autonomous Agents (Agents'97), Marina del Rey 1997, S. 30-38.
- Rickel, J.; Johnson, W.L.; Thiebaut, M. (2000):** Steve: A Virtual Human for Individual and Team Training. In: <http://citeseer.ist.psu.edu/490323.html>, zugegriffen am 31.10.2008.
- Rist, T.; Baldes, S.; Gebhard, P.; Kipp, M.; Klesen, M.; Rist, P.; Schmitt, M. (2002):** CrossTalk: An interactive installation with animated presentation agents. In:

- Proceedings of the 2nd Conference on Computational Semiotics for Games and New Media (COSIGN) 2002, S. 61-67.
- Rosset, S.; Bennacef, S.; Lamel, L. (1999):** Design Strategies For Spoken Language Dialog Systems. In: Proceedings of the Sixth European Conference on Speech Communication and Technology (Eurospeech'99), Budapest, Hungary 1999, S. 1535-1538.
- Rudolph, S.; Krcmar, H. (2004):** Stand digitaler Produktionen (Arbeitspapier Nr. 3). Lehrstuhl für Wirtschaftsinformatik, TU München, 2004.
- Rupp, C.; die SOPHISTen (2007):** Requirements-Engineering und -Management: Professionelle, iterative Anforderungsanalyse für die Praxis. (4. aktualisierte und erweiterte Aufl.), Carl Hanser Verlag, München 2007.
- Saad, A. (2003):** Prototyping bei der BMW Car IT GmbH. Java Spektrum, S. 49-53.
- Sacher, H.; Bubb, H. (2006):** Was bedient der Fahrer? - Feldversuche zu Erfassung der Fahrer-Fahrzeug-Interaktion. In: MMI-Interaktiv, (2006) Nr. 11, S. 14-25.
- Sadek, M.D.; de Mori, R. (1998):** Dialogue systems. In: Spoken Dialogues with Computers. Hrsg.: de Mori, R. Academic Press, London 1998, S. 523-561.
- Sagisaka, Y.; Campbell, N.; Higuch, N. (Hrsg.) (1997):** Computing Prosody – Computational Models for Processing Spontaneous Speech. Springer-Verlag, Berlin 1997.
- Sagisaka, Y.; Tokhkura, Y. (1984):** Phoneme duration control for speech synthesis by rule. In: Journal of the Institute of Electronics and Communication Engineers of Japan, Vol. J67-A (1984) Nr. 7, S. 629-636.
- Sahel, S. (2006):** Struktur der deutschen Sprache1: Das Wort. In: http://www.uni-bielefeld.de/lili/projekte/el-germling/veranstaltungen/struktur_dt_sprache1/StrukturI_ss06/phoneme.pdf, zugegriffen am 31.10.2007.
- Salem, B.; Earle, N. (2000):** Designing a nonverbal language for expressive avatars. In: Proceedings of the 3rd International Conference on Collaborative Virtual Environments (CVE'00), San Francisco, CA 2000, S. 93-101.
- Salmen, A. (2002):** Multimodale Menüausgabe im Fahrzeug, Herbert Utz Verlag, München 2002.
- Samuelson, P. (2003):** Die Bedeutung von natürlichsprachlichen Dialogsystemen im Internet-Business. In: Avatare - Digitale Sprecher für Business und Marketing. Hrsg.: Lindner, C. Springer-Verlag, Berlin et al. 2003, S. 27-35.
- Sanders, G.A.; Scholtz, J. (2000):** Measurement and Evaluation of Embodied Conventional Agents. In: Embodied conversational agents. Hrsg.: Cassell, J.; Sullivan, J.; Prevost, S.; Churchill, E. MIT Press, Cambridge, MA 2000, S. 346-373.
- Saul, L.K.; Lee, D.D.; Isbell, C.L.; LeCun, Y. (2002):** Real time voice processing with audiovisual feedback: toward autonomous agents with perfect pitch. In: Advances in Neural Information Processing Systems 15 (NIPS 2002) 2002, S. 1205-1212.
- ScanSoft (2004):** Vocon 3200 Development Formalisms.
- ScanSoft (2005):** ScanSoft VoCon 3200 Embedded Development System: Vocon 3200 Developer's Guid, 2005.
- Schatter, G.; Schmolke, S.; Zeller, B. (2007):** Elements of an Acoustic Anthropomorphisation of an User Adaptive Digital Radio DAB. In: 8. Workshop Digitaler Rundfunk, Ilmenau 2007.
- Schäuffele, J.; Zurawka, T. (2006):** Automotive Software Engineering. (3. Aufl.), Vieweg & Sohn Verlag, Wiesbaden 2006.
- Schierge, F. (2005):** Welche Fahrerassistenz wünschen sich die Fahrer?: Methoden und Ergebnisse zur Nutzerakzeptanz von FAS aus der

- Forschungsinitiative INVENT. In: VDI Berichte, (2005) Nr. 1919, S. 207-219.
- Schmidt, A.P. (1998):** E-CYAS, E_ndo-CY_bernetic A_rtficial S_tar im Internet. In: <http://www.heise.de/tp/r4/artikel/2/2367/1.html>, zugegriffen am 19.11.2007.
- Schröter, J. (2005):** Text to-Speech (TTS) Synthesis. In: http://www.research.att.com/~ttsweb/tts/papers/2005_EEHandbook/tts.pdf, zugegriffen am 31.10.2008.
- Schroyer, R. (2000):** Actual readers versus implied readers: Role conflicts in Office 97. In: Technical communications, Vol. 47 (2000) Nr. 2, S. 238-240.
- Schukat-Talamazzini, E.G. (1995):** Automatische Spracherkennung: Grundlagen, statistische Modelle und effiziente Algorithmen, Vieweg Verlag, Braunschweig, Wiesbaden 1995.
- Schulte, J. (2003):** IT in der Kommissionierung. In: Hebezeuge und Fördermittel, Vol. 43 (2003) Nr. 9, S. 400-403.
- Schütte, R. (1999):** Basispositionen in der wirtschaftsinformatil - ein gemäßigt-konstruktivistisches Programm. In: Wirtschaftsinformatik und Wissenschaftstheorie. Hrsg.: Becker, J.; König, W.; Schütte, R.; Wendt, O.; Zelewski, S. Verlag Dr. Th. Gabler, Wiesbaden 1999, S. 211-242.
- Schweigert, M.; Bubb, H. (2003):** Einfluss von Nebenaufgaben auf das Fahrerblickverhalten. In: VDI Berichte, (2003) Nr. 1768, S. 59-74.
- Scripture, E.W. (1901):** Phonetic Notation. In: Modern Language Notes, Vol. 16 (1901) Nr. 6, S. 161-164.
- Searle, J.R. (1975):** Indirect Speech Acts. In: Syntax and Semantics III: Speech Acts, (1975), S. 59-82.
- Seiffert, H. (1992):** Einführung in die Wissenschaftstheorie - Dritter Band: Handlungstheorie, Modellogik, Ethik, Systemtheorie (Vol. 1). (12 Aufl.), Beck, München 1992.
- Selcon, S.J.; Taylor, R.M.; McKenna, F.P. (1995):** Integrating Multiple Information Sources: Using Redundancy in the Design of Warnings. In: Ergonomics, Vol. 38 (1995) Nr. 11, S. 2362-2370.
- Seneff, S.; Lau, R.; Polifroni, J. (1999):** Organization, Communication, and Control in the GALAXY-II Conversational System. In: Proceedings of the Sixth European Conference on Speech Communication and Technology (Eurospeech '99), Budapest, Hungary 1999, S. 1271-1274.
- Sengers, P. (2000):** The avatar as interface in virtual environments (eRENA-D6.5.2). GMD, 2000.
- Shah, H.; Henry, O. (2005):** The Confederate Effect in Human-Machine Textual Interaction. In: <http://www.alicebot.org/articles/huma/Huma.pdf>, zugegriffen am 19.05.2008.
- Sheridan, T.B. (1988):** Task allocation and supervisory control. In: Handbook of Human-Computer Interaction. Hrsg.: Helander, M. Elsevier Science Publishers, Amsterdam 1988, S. 159-173.
- Shneiderman, B. (1997):** Direct Manipulation Versus Agents: Paths to Predictable, Controllable, and Comprehensible Interfaces. In: Software agents. Hrsg.: Bradshaw, J.M. AAAI Press, 1997, S. 97-106.
- Shneiderman, B. (2000):** The limits of speech recognition. In: Communications of the ACM, Vol. 43 (2000) Nr. 9, S. 63-65.
- Shneiderman, B.; Plaisant, C. (2005):** Designing the User Interface: Strategies for Effective Human-Computer Interaction. (4. Aufl.), Pearson Addison-Wesley, Boston et al. 2005.
- Short, J.; Williams, E.; Christie, B. (1976):** The Social Psychology of Telecommunication, John Wiley & Sons, London, UK 1976.
- Silverman, K.; Beckman, M.; Pierrehumbert, J.; Ostendorf, M.; Wightman, C.; Price, P.; Hirschberg, J. (1992):** ToBI: A Standard Scheme for Labeling Prosody. In:

- International Conference on Spoken Language Processing., Banff, Canada 1992, S. 867-879.
- Simon, H. (1996):** The Sciences of the Artificial. (3 Aufl.), The MIT Press, Cambridge et al. 1996.
- Singh, D.; Boland, F. (2007):** Voice Activity Detection In: ACM Crossroads, Vol. 13 (2007) Nr. 4.
- Sjölander, K.; Beskow, J.; Gustafson, J.; Lewin, E.; Carlson, R.; Granström, B. (1998):** Web-based educational tools for speech technology. In: Proceedings of the 5th International Conference on Spoken Language Processing (ICSLP'98), Sydney, Australia 1998, S. 3217-3220.
- Slater, M.; Steed, A. (2001):** Meeting People Virtually: Experiments in Shared Virtual Environments. In: The Social Life of Avatars: Presence and Interaction in Shared Virtual Environments. Hrsg.: Schroeder, R. Springer Verlag, London 2001, S. 146-171.
- Smith, R.; Hipp, D.R. (1994):** Spoken Natural Language Dialog Systems: A Practical Approach, Oxford University Press, New York 1994.
- Sölbrandt, S. (2006):** Der Avatar "EIBY". In: Web-Kommunikation mit OpenSource. Hrsg.: Möbus, C.; Eißner, A.; Feindt, J.; Janßen, C.; Krefeldt, J.; Sieverding, S.; Sölbrandt, S.; Stumpe, J.; de Vries, H.; Willer, S. Springer Verlag, Berlin, Heidelberg, New York 2006, S. 117-138.
- Speech Experts (2003):** Whitepaper. In: <http://www.speech-experts.com/de/whitepaper.pdf>, zugegriffen am 22.12.2007.
- Sproat, R.; Hunt, A.; Ostendorf, M.; Taylor, P.; Black, A.W.; Lenzo, K.; Edginton, M. (1998):** SABLE: A standard for TTS markup. In: International Conference on Spoken Language Processing, Sydney 1998, S. 1719-1724.
- Sproat, R.; Taylor, P.; Tannenblatt, M.; Isard, A. (1997):** A markup language for text-to-speech synthesis. In: Proceedings of the Fifth European Conference on Speech Communication and Technology (Eurospeech '97), Rhodes, Greece 1997, S. 1747-1750.
- Sproull, L.; Subramani, M.; Kiesler, S.; Walker, J.; Waters, K. (1997):** When the Interface is a Face. In: Human Values and the Design of Computer Technology. Hrsg.: Friedman, B. CSLI Publications/ Cambridge University Press, Stanford, CA 1997, S. 163-190.
- Srinivasan, R.; Jovanis, P.P. (1997):** Effect of In-Vehicle Route Guidance Systems on Driver Workload and Choice of Vehicle Speed: Findings from a Driving Simulator Experiment. In: Ergonomics and Safety of Intelligent Driver Interfaces. Hrsg.: Noy, I.Y. Lawrence Erlbaum Associates, Mahwah, NJ 1997, S. 97-114.
- Stallard, D.; Bobrow, R. (1992):** Fragment processing in the DELPHI system. In: Speech and Natural Language Workshop, Harriman, N.Y. 1992, S. 305-310.
- Stede, M.; Schlangen, D. (2004):** Information-Seeking Chat: Dialogue Management by Topic Structure. In: Proceedings of the 8th Workshop on Semantics and Pragmatics of Dialogue, CATALOG 04, Barcelona, Spain 2004.
- Stegmann, R. (2006):** LINGUINI: Acquiring Individual Interest Profiles by Means of Adaptive Natural Language Dialog. Dissertation, Technische Universität München 2006.
- Stenzel, R. (1999):** Natürlichsprachliche Eingabe für das Automobil. Diplomarbeit, Technische Universität München 1999.
- Stephenson, N. (1992):** Snowcrash, Bantam Books, New York 1992.
- Stiller, C. (2005):** Fahrerassistenzsysteme – von realisierten Funktionen zum vernetzt wahrnehmenden, selbstorganisierenden Verkehr. In: Fahrerassistenzsysteme mit

- maschineller Wahrnehmung. Hrsg.: Maurer, M.; Stiller, C. Springer, Berlin 2005, S. 1-20.
- Stone, B.A.; Lester, J.C. (1997):** Dynamically sequencing an animated pedagogical agent. In: Proceedings of AAAI-96, Portland, Oregon 1997, S. 424-431.
- Stürmer, I.; Weinberg, D.; Conrad, M. (2005):** Overview of Existing Safeguarding Techniques for Automatically Generated Code. In: Proceedings of the second international workshop on Software engineering for automotive systems St. Louis, Missouri 2005, S. 1-6.
- Suhendra, A. (2001):** The development of System for Talking Avatar. Dissertation, Universität Kassel 2001.
- Sun Microsystems (1999):** API Markup Language Specification. In: <http://java.sun.com/products/java-media/speech/forDevelopers/JSML/index.html>, zugegriffen am 9.12.2007.
- Sun Microsystems (2007):** Grammar Format Specification. In: <http://java.sun.com/products/java-media/speech/forDevelopers/JSGF/JSGF.html#11803>, zugegriffen am 22.12.2007.
- SVOX (2006):** SVOX Manual: Speech Output Engine SDK 4.0.0. SVOX AG, 2006.
- SVOX (2007):** SVOX Architecture. In: <http://www.svox.com/SVOX-Architecture.aspx>, zugegriffen am 05.05.2008.
- Takayama, L.; Nass, C. (2008):** Driver safety and information from afar: An experimental driving simulator study of wireless vs. in-car information services. In: International Journal of Human-Computer Studies, Vol. 66 (2008) Nr. 3, S. 173-184.
- Takeda, H.; Tomiyama, T.; Yoshikawam, H.; Veerkamp, P.J. (1990):** Modeling design processes (CS-R9059). Centre for Mathematics and Computer Science, 1990.
- Takeuchi, A.; Nagao, K. (1993):** Communicative facial displays as a new conversational modality. In: Proceedings of the INTERCHI '93 conference on Human factors in computing systems, Amsterdam, The Netherlands 1993, S. 187-193.
- Takeuchi, A.; Naito, T. (1995):** Situated facial displays: towards social interaction. In: Human factors in computing Systems: CHI'95 Conference Proceedings 1995, S. 450-455.
- Taylor, P.; Black, A.W. (1997):** Assigning Phrase Breaks from Part-of-Speech Sequences. In: Proceedings of the Fifth European Conference on Speech Communication and Technology (Eurospeech '97), Rhodes, Greece 1997, S. 995-998.
- Taylor, P.; Isard, A. (1997):** SSML: A Speech Synthesis Markup Language. In: Speech Communication, Vol. 21 (1997) Nr. 1, S. 123-133.
- Taylor, P.A.; Black, A.; Caley, R. (1998):** The architecture of the festival speech synthesis system. In: The Third ESCA Workshop in Speech Synthesis, Jenolan Caves, Australia 1998, S. 147-151.
- Teichgräber, U.K.M.; Ehrenstein, T.; Lemke, M.; Liebig, T.; Stobbe, H.; Hosten, N.; Keske, U.; Felix, R. (1999):** Digitale Spracherkennung bei der Erfassung computertomographischer Befundtexte. In: RöFo - Fortschritte auf dem Gebiet der Röntgenstrahlen und der bildgebenden Verfahren, Vol. 171 (1999) Nr. 2, S. 396-399.
- Tesch, R. (1990):** Qualitative Research - Analysis Types & Software Tools, The Falmer Press, New York, Philadelphia, London 1990.
- Theune, M. (2001):** ANGELICA: choice of output modality in an embodied agent. In: International Workshop on Information Presentation and Natural Multimodal Dialogue (IPNMD-2001), Verona, Italy 2001, S. 89-93.
- Thórisson, K.R. (1996):** Communicative Humanoids: A Computational Model of Psychosocial Dialogue Skills. Dissertation, Massachusetts Institute of Technology 1996.

- Thórisson, K.R. (1998):** Real-Time Decision Making in Multimodal Face-to-Face Communication. In: Proceedings of the Second international conference on Autonomous agents, Minneapolis, Minnesota, USA 1998, S. 16-23.
- Timpe, K.-P. (1998):** Unterstützungssysteme als interdisziplinäre Herausforderung. In: Wohin führen Unterstützungssysteme? Entscheidungshilfe und Assistenz in Mensch-Maschine-Systemen, 2. Berliner Werkstatt Mensch-Maschine-Systeme, Berlin 1998, S. 1-20.
- Tönnis, M.; Broy, V.; Klinker, G. (2006):** A Survey of Challenges Related to the Design of 3D User Interfaces for Car Drivers. In: IEEE Virtual Reality Conference (VR 2006), Alexandria, Virginia, USA 2006, S. 127-134
- Totzke, I. (2001):** Die Humanisierung multimodaler HMI - die Bedeutung der Modalität und Situation der Mensch-Maschine-Interaktion. In: 2. IIR-Fachkongress "E-Car Infotainmentplattform & Telematikdienste für das Multimedia Auto", Stuttgart 2001.
- Totzke, I.; Meilinger, T.; Krüger, H.-P. (2003):** Erlernbarkeit von Menüsystemen im Fahrzeug - mehr als „nur“ eine Lernkurve. In: VDI Berichte, (2003) Nr. 1768, S. 171-196.
- Traber, C. (1993):** Syntactic Processing and Prosody Control in the SVOX TTS System for German. In: Proceedings of the Second European Conference on Speech Communication and Technology (Eurospeech '93), Berlin, Germany 1993, S. 2099-2102.
- Tschamler, H. (1996):** Wissenschaftstheorie: eine Einführung für Pädagogen. (3. erweiterte und neubearbeitete Aufl.), Julius Klinkhardt, Heilbrunn 1996.
- Tümmler, J. (2007):** Avatare in Echtzeitsimulationen. Dissertation, Universität Kassel 2007.
- Turing, A. (1950):** Computing machinery and intelligence. In: <http://www.abelard.org/turpap/turpap.htm>, zugegriffen am 02.09.2006. Access.
- Turk, M.; Kölsch, M. (2003):** Perceptual Interfaces. In: Emerging Topics in Computer Vision. Hrsg.: Medioni, G.; Kang, S.B. Prentice Hall, 2003.
- Turk, M.; Robertson, G. (2000):** Perceptual user interfaces. In: Communications of the ACM, Vol. 43 (2000) Nr. 3, S. 32-34.
- Vaishnavi, V.; Kuechler, W. (2004):** Design Research in Information Systems
In: <http://www.isworld.org/Researchdesign/drisISworld.htm>, zugegriffen am 20.01.2007. Access.
- Van Compernelle, D. (1997):** Speech Recognition in the Car - From Phone Dialing to Car Navigation. In: Proceedings of the Fifth European Conference on Speech Communication and Technology (Eurospeech '97), Rhodes, Greece 1997, S. 2431-2334.
- van Dam, A. (1997):** Post-WIMP user interfaces. In: Communications of the ACM, Vol. 40 (1997) Nr. 2, S. 63-67.
- Verlinden, M.; Tijsseling, C.; Frowein, H. (2001):** A Signing Avatar on the WWW. In: Revised Papers from the International Gesture Workshop on Gesture and Sign Languages in Human-Computer Interaction London, UK 2001, S. 169-172.
- Vetter, M. (2003):** "Ich habe Ihre Eingabe leider nicht verstanden" - Qualitätskriterien für Online-Tests von Bots. In: Avatare - Digitale Sprecher für Business und Marketing. Hrsg.: Lindner, C. Springer-Verlag, Berlin et al. 2003, S. 73-92.
- VoiceXML Forum (2000):** Voice eXtensible Markup Language - VoiceXML. In: <http://www.voicexml.org/specs/VoiceXML-100.pdf>, zugegriffen am 9.12.2007.
- Vollrath, M.; Totzke, I. (2000):** In-vehicle communication and driving: an attempt to overcome their interference. In: http://www.psychologie.uni-wuerzburg.de/methoden/texte/2000_vollrath_totzke_In_vehicle_communication_and_driving.pdf, zugegriffen am 04.01.2008.

- Vollrath, M.; Totzke, I. (2003):** Möglichkeiten der Nutzung unterschiedlicher Ressourcen für die Fahrer-Fahrzeug-Interaktion. In: VDI Berichte, (2003) Nr. 1768, S. 47-58.
- von Benda, H. (1977):** Die Skalierung der Gefährlichkeit von Straßenverkehrssituationen. Teil 1: Ein Klassifikationssystem für Verkehrssituationen aus Fahrersicht im Auftrag der Bundesanstalt für Straßenwesen (FP 7320). Technische Universität München, 1977.
- von Glasersfeld, E. (1992):** Konstruktion der Wirklichkeit und der Begriff der Objektivität. In: Einführung in den Konstruktivismus. Hrsg.: Gumin, M. Piper, München 1992, S. 1-26.
- von Wendt, K.L. (2003):** Technische Grundlagen von natüersprachlichen Dialogsystemen. In: Avatare - Digitale Sprecher für Business und Marketing. Hrsg.: Lindner, C. Springer-Verlag, Berlin et al. 2003, S. 39-47.
- W3C (2007):** EMMA: Extensible MultiModal Annotation markup language. In: <http://www.w3.org/TR/emma/>, zugegriffen am 04.01.2008.
- Walker, J.; Sproull, L.; Subramani, R. (1994a):** Computers are Social Actors. In: Human Factors in Computing Systems: CHI '94 Conference Proceedings 1994a.
- Walker, J.H.; Sproull, L.; Subramani, R. (1994b):** Using a human face in an interface. In: Proceedings of the SIGCHI conference on Human factors in computing systems: celebrating interdependence, Boston, Massachusetts, United States 1994b, S. 85-91.
- Walker, M.A.; Cahn, J.E.; Whittaker, S.J. (1997):** Improvising Linguistic Style: Social and Affective Bases for Agent Personality. In: Proceedings of the First International Conference on Autonomous Agents 1997.
- Wallace, R.S. (2002a):** The Anatomy of A.L.I.C.E. In: <http://www.alicebot.org/documents/Anatomy.doc>, zugegriffen am 10.04.2002.
- Wallace, R.S. (2002b):** Zipf's Law. In: <http://www.alicebot.org/articles/wallace/zipf.html>, zugegriffen am 31.10.2008.
- Wäller, C.; Wu, Y.; Bohnenberger, T.; Lilienthal, J. (2006):** Schutzrecht 537 Volkswagen AG: Vorrichtung und Verfahren zur interaktiven Informationsausgabe und/oder Hilfestellung für den Benutzer eines Kraftfahrzeugs 2006
- Wandke, H.; Wetzenstein, E.; Polkehn, K. (2005):** Handlungsbezogene Elementarbausteine für Fahrerassistenzsysteme. In: VDI-Berichte, (2005) Nr. 1919, S. 41-62.
- Watchel, T. (1989):** Discourse Structure in LOQUI. In: Recent developments and applications of natural language processing. Hrsg.: Peckham, J. Kogan Page, London 1989, S. 215-243.
- Waters, K.; Levergood, T.M. (1993):** DECface: An Automatic Lip-Synchronization Algorithm for Synthetic Faces. Digital Equipment Corporation - Cambridge Research Lab, 1993.
- Wedekind, H.; Ortner, E. (1980):** Systematisches Konstruieren von Datenbankanwendungen - Zur Methodologie der angewandten Informatik, Hanser, München, Wien 1980.
- Weizenbaum, J. (1966):** Eliza - A computer program for the study of natural language communication between man and machine. In: Communications of the ACM, Vol. 9 (1966) Nr. 1, S. 36-45.
- Wells, J. (1997):** SAMPA computer readable phonetic alphabet. In: Handbook of Standards and Resources for Spoken Language Systems. Hrsg.: Gibbon, D.; Moore, R.; Winski, R. Mouton de Gruyter, Berlin, New York 1997, S. 684-732.
- Wells, J.C. (1995):** SAMPROSA (SAM Prosodic Transcription). In: <http://www.phon.ucl.ac.uk/home/sampa/samprosa.htm>, zugegriffen am 9.12.2007.
- Westphal, M.; Waibel, A. (1999):** Towards Spontaneous Speech Recognition for On-Board Car Navigation and Information Systems. In: Proceedings of the Sixth European

- Conference on Speech Communication and Technology (Eurospeech '99), Budapest, Hungary 1999, S. 1955-1958.
- Weyl, H. (1990):** Philosophie der Mathematik und der Naturwissenschaften. (6. Aufl.), Oldenbourg Verlag, München 1990.
- Whittaker, S.J.; Attwater, D.J. (1996):** The design of complex telephony applications using large vocabulary speech technology. In: Proceedings of the 4th International Conference on Spoken Language Processing (ICSLP-96), Philadelphia, PA 1996, S. 705-708.
- Wierwille, W.W.; Hulse, M.C.; Fischer, T.J.; Dingus, T.A. (1991):** Visual Adaption of The Driver to High-Demand Driving Situations While Navigating with In-Car Navigation Systems. In: Vision in Vehicles III. Hrsg.: Brown, I.D.; Haslegrave, C.M.; Moorhead, I.; Taylor, S.P.; Gale, A.G. North-Holland Publishing, Amsterdam 1991, S. 79-87.
- Wightman, C.W. (2002):** ToBI Or Not ToBI? In: Proceedings of Speech Prosody 2002, Aix-en-Provence, France 2002, S. 25-29.
- Wikman, A.S.; Nieminen, T.; Summala, H. (1998):** Driving Experience and Time-Sharing During In-Car Tasks on Roads of Different Width. In: Ergonomics, Vol. 41 (1998) Nr. 3, S. 358-372.
- Wilcox, S.K. (1998):** Web Developer.com – Guide to 3D Avatars, Wiley Computer Publishing, New York 1998.
- Williams, M.; Helbig, R. (2007):** Sicherheitsleitfaden für mobile Dienste im automobilen Nutzungsbereich. In: Mobile Dienste im Auto der Zukunft. Hrsg.: Reichwald, R.; Krcmar, H.; Reindl, S. Eul Verlag, Lohmar 2007, S. 273-318.
- Winograd, T.; Flores, F. (1989):** Erkenntnis, Maschinen, Verstehen. (1. Aufl.), Rotbuch-Verl. 1989.
- Winston, P.H. (1987):** Künstliche Intelligenz, Addison-Wesley Verlag, Bonn 1987.
- Witt, B.I.; Baker, F.T.; Meritt, E.W. (1994):** Software Architecture and Design: Principles, Models and Methods, Van Nostrand Reinhold, New York 1994.
- Wyard, P.J.; Simons, A.D.; Appelby, S.; Kaneen, E.; Williams, S.H.; Preston, K.R. (1996):** Spoken language systems - beyond prompt and response. In: BT Technology Journal, Vol. 14 (1996) Nr. 1, S. 187-205.
- Yin, R.K. (1994):** Case Study Research – Design and Methods. (3 Aufl.), Sage Publications, Thousands Oaks, London, New Delhi 1994.
- Zboril, D. (1997):** Einführung in die Sprachsynthese. In: http://www.phonetik.uni-muenchen.de/Lehre/Skripten/Seminare/HS_WS1997/Synthese.html, zugegriffen am 31.08.2007.
- Zelewski, S. (2007):** Kann Wissenschaftstheorie behilflich für die Publikationspraxis sein? In: Wissenschaftstheoretische Fundierung und wissenschaftliche Orientierung der Wirtschaftsinformatik. Hrsg.: Lehner, F.; Zelewski, S. Gito, Berlin 2007, S. 71-120.
- Zerfowski, D. (2001):** ASCET-SD im Praxis-Einsatz bei Knorr-Bremse. In: Automotive Electronics - Sonderausgabe von ATZ und MTZ und AUTOMOTIVE ENGINEERING PARTNERS, (2001), S. 10-14.
- Zieten, w. (1990):** Gebrauchs- und Betriebsanleitungen - direkt, wirksam, einfach und einleuchtend, Verlag Moderne Industrie, Landsberg am Lech 1990.
- Zimmermann, W.; Schmidgall, R. (2007):** Bussysteme in der Fahrzeugtechnik: Protokolle und Standards. (2. Aufl.), Friedrich Vieweg & Sohn Verlag, Wiesbaden 2007.
- Zue, V.W.; Glass, J.R. (2000):** Conversational Interfaces: Advances and Challenges. In: Proceedings of the IEEE 2000, S. 1166-1180.

Anhang

Anhang A: Mögliche Aufgaben für die Evaluation im Fahrzeug

Aufgabe	Sektor	Zeitpunkt	Bedienelemente	Funktionen d. Bedienelemente	Min. # Schritte	Max. # Schritte	Seiten im Handbuch
Vordersitze einstellen	Sekundär	Immer	3 Einstellungsschalter an der Sitzseite	Schalter ziehen/drücken	1	3	1
Sitzeinstellung mit Sitzmemory vornehmen	Tertiär	immer	Sitzmemorypanel (4 Tasten)	Tasten betätigen	1	2	0,5
Spurwechselassistenten ein-/ausschalten	Sekundär	Immer	Audi side assist Knopf	Taste betätigen	1	1	6
Blinken	Sekundär	Fahren	Blinkerhebel	Hebel heben/senken	1	1	0,5
Parklicht ein-/ausschalten	Sekundär	Stand	Blinkerhebel	Hebel heben/senken	1	1	0,5
Fernlicht ein-/ausschalten	Sekundär	Immer	Blinkerhebel	Hebel drücken	1	1	0,5
Lichthupe betätigen	Sekundär	Immer	Blinkerhebel	Hebel ziehen	1	1	0,5
Audi lane assist ein-/ausschalten	Sekundär	Fahrt	Blinkerhebel mit einem seitlichen Knopf	Knopf drücken	1	1	1,5
Zentralverriegelung	Tertiär	Immer	Zentralverriegelungspanel (2 Tasten)	Tasten betätigen	1	1	0,5
Fahrer-/Beifahreraußenspiegel verstellen	Sekundär	Immer	Außenspiegelpanel(4 Knöpfe und ein Drehknopf)	Drehknopf drehen und mit dem Stick justieren	1	2	0,5
Spiegelbeheizung ein-/ausschalten	Sekundär	Immer	Außenspiegelpanel(4 Knöpfe und ein Drehknopf)	Drehknopf drehen	1	1	0,5
Beide Außenspiegel anklappen	Tertiär	Stand	Außenspiegelpanel(4 Knöpfe und ein Drehknopf)	Drehknopf drehen	1	1	0,5
Gepäckraumklappen öffnen	Tertiär	Stand	Entriegelungstaste	Taste ziehen	1	1	0,5
Fenster (automatisch) schließen/öffnen	Tertiär	Immer	Festerheberpanel (5 Tasten)	Schalter drücken/ziehen	1	1	0,5
Kindersicherung ein-/ausschalten	Tertiär	Immer	Festerheberpanel (5 Tasten)	Knopf drücken	1	1	0,5
Motorraumklappe entriegeln	Tertiär	Stand	Motorraumentriegelungsschalter	Hebel ziehen	1	1	0,5
Fahrlicht ausschalten	Sekundär	Immer	Lichtschalterpanel (2 Tasten, 2 Drehräder)	Drehtaste drehen	1	1	0,5
Automatische Lichtfahrtsteuerung einschalten	Sekundär	Immer	Lichtschalterpanel (2 Tasten, 2 Drehräder)	Drehtaste drehen	1	1	0,5

Tabelle A-1: Mögliche Aufgaben für die Evaluation im Fahrzeug (Teil 1)

Aufgabe	Sektor	Zeitpunkt	Bedienelemente	Funktionen d. Bedienelemente	Min. # Schritte	Max. # Schritte	Seiten im Handbuch
Standlicht einschalten	Se-kundär	Stand	Lichtschalterpanel (2 Tasten, 2 Drehräder)	Drehtaste drehen	1	1	0,5
Abblendlicht einschalten	Se-kundär	Immer	Lichtschalterpanel (2 Tasten, 2 Drehräder)	Drehtaste drehen	1	1	0,5
Standlicht ein-/ ausschalten	Se-kundär	Immer	Lichtschalterpanel (2 Tasten, 2 Drehräder)	Drehtaste drehen/Knopf drücken	1	2	0,5
Nebelschlussleuchte ein-/ ausschalten	Se-kundär	Immer	Lichtschalterpanel (2 Tasten, 2 Drehräder)	Drehtaste drehen/Knopf drücken	1	2	0,5
Instrumentenbeleuchtung regulieren	Tertiär	Immer	Lichtschalterpanel (2 Tasten, 2 Drehräder)	Drehtaste ein-/ausrasten und drehen	1	3	0,5
Leuchtweite regulieren	Se-kundär	Immer	Leuchtweitenregulierungsdrehrad	Drehtaste drehen	1	1	0,5
Checkdurchlauf von Hand starten	Tertiär	Bei eingeschalteter Zündung	SET-Taste	Taste drücken	1	2	12,5
Warnschwelle 1 einstellen	Se-kundär	Fahrt	SET-Taste	Taste drücken	1	1	0,5
Warnschwelle 2 einstellen	Se-kundär	Immer	MMI	MMI bedienen	?	?	0,5
Uhrzeit, Datum, Temperaturanzeige und Kilometerzähler anzeigen	Tertiär	Stand	SET-Taste	Taste drücken	1	1	0,5
Tageskilometerzähler zurücksetzen	Tertiär	Immer	0.0-Taste	Taste drücken	1	1	0,5
Lenkradposition einstellen	Se-kundär	Immer	Lenksäulenverstellungshebel	Hebel ziehen/Lenkradposition verstellen/Hebel drücken	3	3	0,5
Warnblickanlage ein-/ausschalten	Se-kundär	Immer	Warnblinkanlagenknopf	Knopf drücken	1	1	0,5
Fahrmodus auswählen	Se-kundär	Immer	Audi drive select (Anzeigetafel und 2 Knöpfe)	Mit zwei Knöpfen zwischen den verschiedenen Modi wechseln	1	1	2
Einparkhilfe einschalten	Se-kundär	Stehen/Langsame s Fahren	P-Schalter	Entweder dem Rückwärtsgang einlegen oder den P-Schalter betätigen	1	1	3,5

Tabelle A-1: Mögliche Aufgaben für die Evaluation im Fahrzeug (Teil 2)

Aufgabe	Sektor	Zeitpunkt	Bedienelemente	Funktionen d. Bedienelemente	Min. # Schritte	Max. # Schritte	Seiten im Handbuch
Einparkhilfe ausschalten	Se-kundär	Stehen/Langsam es Fahren	P-Schalter	Schneller als 10 km/h fahren oder den P-Schalter betätigen oder die Zündung ausschalten	1	1	1
Audi-hold assist ein-/ausschalten	Se-kundär	Stand	Audi hold assist-Knopf	Taste drücken	1	1	1
Parkbremse lösen/schließen	Primär	Stand	Parkbremsenschalter	Schalter drücken/ziehen	1	1	0,5
Motor starten	Primär	Stand	Kupplungshebel/Schalthebel/Startknopf	Kupplungspedal durchtreten/Schalthebel in Leerlaufstellung bringen/Startknopf drücken	3	3	0,5
Motor abstellen	Primär	Stand	Start-/Stop-Taste	Taste drücken	1	1	0,5
Scheibenwischer ausschalten	Se-kundär	Immer	Scheibenwischerhebel mit 3 Knöpfen	Hebel heben/senken	1	1	0,5
Scheibenwischer einmal drüber wischen lassen	Se-kundär	Immer	Scheibenwischerhebel mit 3 Knöpfen	Hebel kurz nach unten ziehen	1	1	0,5
Scheibenwischer auf Langsam/Schnell einstellen	Se-kundär	Immer	Scheibenwischerhebel mit 3 Knöpfen	Hebel heben/senken	1	1	0,5
Scheibenwischer auf Intervallbetrieb schalten	Se-kundär	Immer	Scheibenwischerhebel mit 3 Knöpfen	Hebel heben/senken & Rädchen am Hebel drehen	1	2	0,5
Windschutzscheibe reinigen	Se-kundär	Immer	Scheibenwischerhebel mit 3 Knöpfen	Hebel ziehen	1	1	0,5
Gewünschte Geschwindigkeit in Geschwindigkeitsregelanlage speichern	Se-kundär	Stand	Geschwindigkeitsregelhebel mit Knopf und Regelrädchen	Mit gewünschter Geschwindigkeit fahren und Knopf betätigen	1	2	0,5
Gespeicherte Geschwindigkeit am in Geschwindigkeitsregelanlage ändern	Se-kundär	Fahrt	Geschwindigkeitsregelhebel mit Knopf und Regelrädchen	Hebel kurz nach oben bzw. Nach unten ziehen	1	1	0,5
Geschwindigkeit in Geschwindigkeitsregelanlage vorwählen	Se-kundär	Stand	Geschwindigkeitsregelhebel mit Knopf und Regelrädchen	Hebel zu sich ziehen und nach oben bzw. nach unten drücken	2	2	0,5

Tabelle A-1: Mögliche Aufgaben für die Evaluation im Fahrzeug (Teil 3)

Aufgabe	Sektor	Zeitpunkt	Bedienelemente	Funktionen d. Bedienelemente	Min. # Schritte	Max. # Schritte	Seiten im Handbuch
Geschwindigkeitsregelanlage (ACC) vorübergehend ausschalten	Primär	Fahrt	Geschwindigkeitsregelhebel mit Knopf und Regelrädchen/Kupplungspedal/Bremspedal/*	Bremspedal treten oder Kupplungspedal länger treten oder Hebel drücken oder länger als 5 min schneller als 10 km/h über gespeicherten Wert fahren	1	1	0,5
Zündung mit dem Schlüssel ein-/ausschalten	Sekundär	Stand	Zündschloss	Zündschlüssel reindrücken	1	1	0,5
Innenraumüberwachung ausschalten	Tertiär	Immer	Taster für Innenraum- und Abschleppschutzüberwachung	Taste betätigen	1	1	0,5
Abschleppschutzüberwachung ausschalten	Tertiär	Immer	Taster für Innenraum- und Abschleppschutzüberwachung	Taste betätigen	1	1	0,5
Vordere Innenbeleuchtung ein-/ausschalten	Tertiär	Immer	Innenbeleuchtungspanel für vorne mit 3 Knöpfen	Taste betätigen	1	1	0,5
Spiegel abblenden	Sekundär	Immer	Taste am Rückspiegel	Taste betätigen	1	1	0,5
Vordere Kopfstützen nach oben verstellen	Tertiär	Stand	Keine	Kopfstütze nach oben ziehen	1	1	0,5
Vordere Kopfstützen nach unten verstellen	Tertiär	Stand	Knopf an der Kopfstütze	Knopf drücken und gleichzeitig die Kopfstütze nach unten ziehen	2	2	0,5
ESP ausschalten	Sekundär	immer	ESP-OFF Schalter	Taste betätigen	1	1	0,5
Fahrerinformationen über den Bordcomputer nacheinander anzeigen lassen	Tertiär	immer	Scheibenwischerhebel mit 3 Knöpfen	Den Schalter nach oben/nach unten drücken	1	1	0,5
Fahrerinformationsanzeige zwischen den verschiedenen Bordcomputern wechseln	Tertiär	immer	Scheibenwischerhebel mit 3 Knöpfen	Den Schalter kurz drücken	1	1	0,5
Einen der Werte des Bordcomputers löschen	Tertiär	immer	Scheibenwischerhebel mit 3 Knöpfen	Navigationstaste nach oben/unten drücken und die Löschtaste mindestens eine Sekunde lang gedrückt halten	1	2	0,5

Tabelle A-1: Mögliche Aufgaben für die Evaluation im Fahrzeug (Teil 4)

Anhang B: Präaktionaler Fragebogen der Evaluation

Gestaltung avatarbasierter natürlichsprachlicher Hilfesysteme für den Einsatz in Fahrzeugen

Präaktionale Befragung

Bei meiner Doktorarbeit geht es um die Untersuchung eines Avatarsystems für den Einsatz im Auto als interaktives Hilfesystem. Dabei werden verschiedene Aufgaben eines Fahrers auf die mögliche Unterstützung durch ein intelligentes Softwaresystem mit Sprachein- und ausgabe untersucht.

Mit Ihrer Teilnahme an den Versuchsreihen leisten Sie einen wichtigen Beitrag zur wissenschaftlichen Beantwortung der Frage, ob künstliche Intelligenz die Benutzung eines Autos erleichtern kann.

Die erhobenen Daten sind anonym und werden nicht außerhalb des oben genannten Dissertationsprojekts verwendet.

Vielen Dank für Ihre Teilnahme,
Valentin Nicolescu



Bitte diesen Abschnitt nicht ausfüllen!

Betreuer:

Gruppe:

Fragebogennr.:

Version: 1.0

Uhrzeit:

Datum:

Teil A: Fragen zu Ihrer Person

1. Sie sind ...

- Weiblich Männlich

2. Wie alt sind Sie?

- 18-29 30-39 40-49 50+

3. Was ist Ihr höchster Bildungsabschluss?

- | | |
|--|--|
| <input type="checkbox"/> Volksschul-/Hauptschulabschluss | <input type="checkbox"/> Realschule/Mittlere Reife |
| <input type="checkbox"/> Abitur/Fachabitur | <input type="checkbox"/> Abgeschlossene Berufsausbildung |
| <input type="checkbox"/> Hochschul-/Fachhochschulabschluss | <input type="checkbox"/> Promotion |

4. Welche Fachrichtungen hatte Ihre Ausbildung/Ihr Studium?

(Eine Mehrfachnennung ist möglich.)

- | | |
|--|---|
| <input type="checkbox"/> technisch | <input type="checkbox"/> betriebswirtschaftlich |
| <input type="checkbox"/> geisteswissenschaftlich | <input type="checkbox"/> naturwissenschaftlich |
| <input type="checkbox"/> sozial | <input type="checkbox"/> handwerklich |
| <input type="checkbox"/> andere: _____ | |

Teil B: Fragen zu Ihrem Fahrverhalten

5. Wie lange haben Sie Ihren Führerschein?

- 0-2 Jahre 2-5 Jahre 5-10 Jahre 10+ Jahre

6. Wie oft in etwa fahren Sie Auto?

- 1-11x / Jahr 1-3x/ Monat 1-5x/ Woche jeden Tag

7. Welche Fahrzeugklasse fahren Sie aktuell?

(Wählen Sie bei mehreren Fahrzeugen bitte nur die Klasse Ihres größten Fahrzeugs aus.)

- Bis Klasse A (z.B. Audi A3, 1er BMW, Mercedes A-Klasse und kleinere Motorisierung)
- Klasse B (z.B. Audi A4, 3er BMW, Mercedes C-Klasse)
- Klasse C (z.B. Audi A6, 5er BMW, Mercedes E-Klasse)
- Klasse D (z.B. Audi A8, 7er BMW, S-Klasse und größere Motorisierung)
- Momentan kein Auto

8. Fahren Sie regelmäßig oder haben Sie regelmäßig ein Fahrzeug von Audi mit einem Baujahr ab 2000 gefahren?

- ja nein

9. Wie schätzen Sie Ihre persönliche Grundhaltung gegenüber der Automarke Audi ein?

- | | | | | |
|--------------------------|--------------------------|--------------------------|--------------------------|--------------------------|
| Sehr schlecht | Schlecht | Neutral | Gut | Sehr gut |
| <input type="checkbox"/> |

Teil C: Fragen zu Ihrem Umgang mit technischen Geräten

10. Wie häufig nutzen Sie die folgenden technischen Geräte oder Anwendungen?

(Bitte wählen Sie in jeder Zeile nur einen Zeitraum aus.)

	Nie	Wenige Male im Jahr	Mindestens 1x / Monat	Mindestens 1x / Woche	täglich
Computer/Laptop	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Internet	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Navigationssystem im Auto	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Systeme mit Spracheingabe (z.B. Handy)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Spracheingabe im Auto	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Systeme mit Sprachausgabe (z.B. mobiles Navigationssystem)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

11. Wie schätzen Sie Ihre persönliche Einstellung gegenüber neuen Technologien, Geräten und Anwendungen ein?

(Bitte wählen Sie nur eine Option aus.)

- Wenn ich von einer neuen Technologie erfahre, suche ich eine Möglichkeit, sie auszuprobieren.
- In meinem Freundeskreis bin ich normalerweise der Erste, der neue Technologien ausprobiert.
- Ich experimentiere gerne mit neuen Technologien.
- Beim Ausprobieren von neuen Technologien bin ich eher zögerlich.

12. Unter einem Avatarsystem versteht man eine Softwareanwendung, die mithilfe einer virtuellen Abbildung (Avatar) mit dem Benutzer kommuniziert (vgl. Abbildung). Solche Systeme werden vor allem auf Websites verwendet, um Besuchern bei der Suche nach Informationen zu helfen (z.B. Bausparfuchs auf www.schwaebisch-hall.de, Anne auf www.ikea.de oder Eve auf www.yellostrom.de).

Wie häufig interagieren Sie mit einem solchen System?

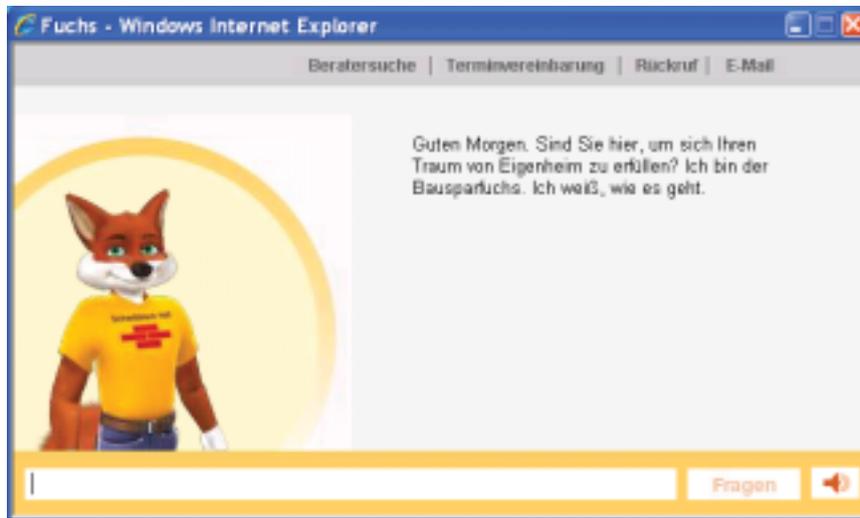


Abbildung 1: Beispiel eines Avatarsystems im Internet

- | | | | | |
|--------------------------|--------------------------|--------------------------|--------------------------|--------------------------|
| Nie | Wenige Male im Jahr | Mindestens 1x / Monat | Mindestens 1x / Woche | täglich |
| <input type="checkbox"/> |

Teil D: Fragen zu Ihrem Umgang mit gedruckten Bedienungsanleitungen aller Art

13. Ich nehme eine gedruckte Bedienungsanleitung zur Hand, wenn...

(Bitte wählen Sie in jeder Zeile nur eine Aussage aus.)

	Trifft überhaupt nicht zu	--	-	0	+	++	Trifft voll und ganz zu
...ich gezielt bestimmte Einstellungen vornehmen möchte (ohne es zuvor intuitiv zu versuchen).	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
...wenn ich intuitiv nicht mehr weiter komme.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
...ein Problem während der Bedienung des Systems auftritt (Fehlermeldung o.ä.)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
...ich mir einen Überblick über die Funktionen verschaffen möchte.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
...ich zum ersten Mal ein neues Gerät bediene.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
...ich mich vorab über die Eigenschaften eines neuen Gerätes informieren möchte.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	

Anderer Grund, und zwar: _____

14. Welche Kriterien sind für Sie wichtig, damit Sie eine gedruckte Bedienungsanleitung gerne benutzen?

(Bitte wählen Sie in jeder Zeile nur eine Aussage aus.)

	Gar nicht wichtig	--	-	0	+	++	Sehr wichtig
Das Auffinden der gesuchten Informationen ist schnell und einfach.	<input type="checkbox"/>						
Die Bedienungsanleitung ist kurz und prägnant.	<input type="checkbox"/>						
Die Handlungsanleitungen sind Schritt für Schritt erklärt.	<input type="checkbox"/>						
Die Handlungsanleitungen sind ausführlich bebildert.	<input type="checkbox"/>						
Auf technisches Hintergrundwissen wird ausführlich eingegangen.	<input type="checkbox"/>						
Das Design und die Aufmachung des Handbuchs sprechen mich an.	<input type="checkbox"/>						

15. Wann haben Sie das letzte Mal eine gedruckte Bedienungsanleitung zu Rate gezogen und für welches Gerät bzw. für welche Anwendung haben Sie dies getan?

Noch nie	Im letzten Jahr	In den letzten 3 Monaten	Letzten Monat	Letzte Woche
<input type="checkbox"/>				

Für das Gerät / die Anwendung: _____

Teil E: Fragen zu Ihrem Umgang mit elektronischen Bedienungsanleitungen im Allgemeinen

16. Welche der folgenden elektronischen Bedienungsanleitungen haben Sie bisher benutzt?
(Mehrfachnennungen möglich)

- Elektronische Dokumente (z.B. PDF- und Word-Dateien ...)
- Online-Hilfen in Anwendungen, Webseiten und Frequently asked Questions (FAQ)
- Elektronische Tutorials und multimediale Anleitungen mit Bildern, Videos und Ton,
- E-Learning Plattformen mit Lernbetreuung und Erfolgskontrolle
- andere: _____

17. Bitte teilen Sie uns mit, wie Sie persönlich elektronische Bedienungsanleitungen beurteilen.
(Bitte wählen Sie in jeder Zeile nur eine Aussage aus.)

	Ich lehne vollkommen ab	--	-	0	+	++	Ich stimme vollkommen zu
Elektronische Bedienungsanleitungen finde ich übersichtlich.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
Elektronische Bedienungsanleitungen bieten mir schnellen Zugriff auf gewünschte Informationen.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
Elektronische Bedienungsanleitungen liefern mir mehr Informationen als gedruckte Handbücher.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
Das Lernen mit elektronischen Bedienungsanleitungen fällt mir leicht.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
Elektronische Bedienungsanleitungen mit Film und Ton vereinfachen mir das Merken von Informationen.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	

18. Wann haben Sie das letzte Mal eine elektronische Bedienungsanleitung zu Rate gezogen und für welches Gerät bzw. für welche Anwendung haben Sie dies getan?

- Noch
nie Im letzten Jahr In den letzten 3 Monaten Letzten Monat Letzte Woche
-

Für das Gerät / die Anwendung: _____

Anhang C: Postaktionaler Fragebogen der Evaluation

Gestaltung avatarbasierter natürlichsprachlicher Hilfesysteme für den Einsatz in Fahrzeugen

Postaktionale Befragung

Sie haben nun einige Aufgaben im Fahrzeug bewältigt und sich einen Eindruck von den Funktionen und ihrer Bedienung machen können.

Bitte teilen Sie im Folgenden Ihre Eindrücke und Erfahrungen mit.



Bitte diesen Abschnitt nicht ausfüllen!

Betreuer:

Gruppe:

Fragebogennr.:

Version: 1.1

Uhrzeit:

Datum:

Teil D: Fragen zum vermittelten Inhalt**9. In welcher Reihenfolge bedienen Sie die Taster des Sitzmemory, um die aktuellen Einstellungen des Sitzes zu speichern?**

(Nur eine der möglichen Antworten ist richtig)

- Zunächst den Speichermodus aktivieren, dann den Set-Knopf drücken und schließlich die Taste mit der Nummer des Speichers drücken
- Zunächst den Set-Knopf drücken, dann den Speichermodus aktivieren und schließlich die Taste mit der Nummer des Speichers drücken.
- Zunächst die Taste mit der Nummer des Speichers drücken, dann den Speichermodus aktivieren und schließlich den Set-Knopf betätigen

10. Mit welchem Knopf schalten Sie die Klimaanlage aus?

(Nur eine der möglichen Antworten ist richtig)

- Mit dem AC-Schalter.
- Mit dem Drehregler der Klimaanlage.
- Mit dem Off-Schalter.

11. Dieses Symbol  markiert in der Klimaanlage die Taste für ...

(Nur eine der möglichen Antworten ist richtig)

- ...die Einstellung der Luftverteilung.
- ...die Aktivierung des Umluftbetriebs.
- ...die Einstellung des Gebläses.

12. Wie können Sie einen Wert im Bordcomputer zurücksetzen?

(Nur eine der möglichen Antworten ist richtig)

- Den „Cursor hoch“-Knopf des Bordcomputers eine Sekunde lang drücken.
- Die Reset-Taste eine Sekunde lang drücken.
- Den Knopf am Ende des Scheibenwischerhebel eine Sekunde lang drücken.

13. Wo befindet sich die Taste für den Spurwechselassistent (side assist)?

(Nur eine der möglichen Antworten ist richtig)

- Am Ende des Blinkerhebels.
- Am Ende des Scheibenwischerhebels.
- An der Innenseite des Fahrerspiegels.

14. Wie hilft Ihnen der Spurhalteassistent (lane assist)?

(Nur eine der möglichen Antworten ist richtig)

- Er gibt visuell und akustisch an, wenn eine Begrenzungslinie überfahre.
- Er hält automatisch die Spur.
- Er weist mich darauf hin, dass ich die Spur gewechselt habe.

15. Dieses Symbol  bei der Luftverteilung bedeutet, dass ...

(Nur eine der möglichen Antworten ist richtig)

- Die Luft wird nur zu Fahrer und Beifahrer geblasen.
- Die Luft wird nur zu den Scheiben geblasen.
- Die Luft wird zu den Scheiben, zum Fahrer und dem Beifahrer geblasen.

16. Mit dem Drehregler der Klimaanlage kann man Folgendes einstellen:

(Nur eine der möglichen Antworten ist richtig)

- Man kann mit dem Drehregler nur die Temperatur einstellen.
- Man kann mit dem Drehregler die Temperatur, die Sitzheizung, das Gebläse und die Luftverteilung einstellen.
- Man kann mit dem Drehregler die Temperatur, das Gebläse und die Luftverteilung einstellen.

17. Wie viele Bordcomputer bzw. Speicher für Bordcomputer hat das Fahrzeug?

(Nur eine der möglichen Antworten ist richtig)

- Das Fahrzeug hat einen Speicher.
- Das Fahrzeug hat zwei Speicher.
- Das Fahrzeug hat drei Speicher.

18. Zum Laden einer Einstellung des Sitzmemory müssen Sie was tun?

(Nur eine der möglichen Antworten ist richtig)

- Die Nummer des Speicherplatzes so lange drücken bis die Endposition des Sitzes erreicht ist.
- Die Nummer des Speicherplatzes kurz drücken.
- Die Set-Taste des Sitzmemory kurz drücken.

Teil E: Fragen zum Gesamteindruck

19. Wie ist ihr Gesamteindruck vom Avatarsystem?

(Bitte wählen Sie in jeder Zeile nur eine Aussage aus.)

	Ich lehne vollkommen ab	--	-	0	+	++	Ich stimme vollkommen zu
Ich habe die Benutzung des Avatarsystems als angenehm empfunden.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
Ich halte das Avatarsystem für technisch ausgereift.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
Ich kann mir vorstellen, dass Avatarsysteme in Zukunft häufig in Autos anzutreffen sein werden.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
Ich habe den Avatar als sozialen Akteur empfunden.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
Der Avatar hat in mir positive Emotionen ausgelöst.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
Das Avatarsystem hat meine Erwartungen an ein solches System erfüllt.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
Ich habe die gestellten Aufgaben als schwer zu lösen empfunden.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
Ich wäre bereit für ein Avatarsystem als Ausstattungsmerkmal im Fahrzeug mehr als 100€ zu bezahlen.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	

20. Bewerten Sie das Avatarsystem mit Schulnoten.

Sehr gut	Gut	Befriedigend	Ausreichend	Mangelhaft
<input type="checkbox"/>				

21. Was hat Ihnen am Avatarsystem gefallen?

(Verwenden Sie auch die Rückseite, wenn der Platz nicht ausreicht)

22. Was hat Ihnen am Avatarsystem NICHT gefallen?

(Verwenden Sie auch die Rückseite, wenn der Platz nicht ausreicht)

23. Welchen Eindruck hat das Avatarsystem auf Sie gemacht?

(Verwenden Sie auch die Rückseite, wenn der Platz nicht ausreicht)

24. Welche Funktionen bzw. welches Verhalten wären im Avatarsystem Ihrer persönlichen Meinung nach sinnvoll und wünschenswert?

(Verwenden Sie auch die Rückseite, wenn der Platz nicht ausreicht)
