

# **Evaluation multimodaler Computer-Systeme in frühen Entwicklungsphasen**

Ein empirischer Ansatz zur Ableitung von Gestaltungshinweisen  
für multimodale Computer-Systeme

von Diplom-Psychologin  
Katharina Seifert

von der Fakultät V – Verkehrs- und Maschinensysteme der  
Technischen Universität Berlin  
zur Erlangung des akademischen Grades

Doktorin der Ingenieurwissenschaften  
- Dr.- Ing. –

genehmigte Dissertation

Promotionsausschuss:

Vorsitzender: Prof. Dr.-Ing. H. Mertens

Gutachter: Prof. Dr. K.-P. Timpe

Gutachter: Prof. Dr.-Ing. L. Blessing

Tag der wissenschaftlichen Aussprache: 2. September 2002

Berlin 2002  
D 83

## Danksagung

Ich möchte den Abschluss meiner Dissertation nutzen, um mich bei den Menschen zu bedanken, die meinen bisherigen beruflichen Werdegang angeleitet, unterstützt und engagiert mit getragen haben.

Herrn Professor Timpe möchte ich ganz besonders für seine großzügige Haltung allen ungewöhnlichen organisatorischen Entscheidungen gegenüber, ohne die diese Arbeit jetzt noch nicht vorliegen würde, und für seine sorgfältige fachliche Betreuung danken. Er förderte und unterstützte meine wissenschaftliche Tätigkeit in hohe Maße und war mir stets ein besonnener Ratgeber, auf dessen Kompetenz ich mich verlassen konnte. Frau Professor Blessing möchte ich sowohl für die Übernahme des Zweitgutachtens herzlich danken, als auch für ihre Bereitschaft, sich mit dem Thema multimodale Mensch-Computer-Interaktion zu beschäftigen. Herrn Professor Mertens danke ich als Vorsitzenden des Promotionsausschusses für eine angenehme Prüfungsatmosphäre. In seiner Funktion als Dekan danke ich ihm für die Unterstützung meiner Promotion in den Ingenieurwissenschaften.

Die empirischen Arbeiten zur Dissertation konnte ich im Rahmen des Projektes „mUltimo3D“ am Heinrich-Hertz-Institut für Nachrichtentechnik Berlin GmbH durchführen. Dafür bin ich Herrn Dr.-Ing. Pastoor, der mich in das von ihm geleitete Projekt aufnahm und mir die nötigen Arbeitsvoraussetzungen schuf, außerordentlich dankbar. Frau Dr.-Ing. Liu, Herrn Dipl.-Inf. Stachel und Herrn Dipl.-Inf. Renault danke ich für die umfangreiche technische Unterstützung, besonders für die Umsetzung der Experimentalaufbauten, und für ihre Offenheit, Erkenntnisse aus der Evaluation in die weitere Systementwicklung einfließen zu lassen. Auch allen studentischen Mitarbeitern sei für ihren Einsatz bei der Programmierung gedankt. Es gibt aber noch Personen namentlich zu benennen, ohne deren engagierte Mitarbeit, die weit über das für studentische Mitarbeiter zu erwartende Maß an Identifikation mit ihrer Aufgabe und an Kompetenz hinausging, die anstrengende empirische Arbeit nicht hätte gelingen können. Jörn Hurtienne unternahm mit mir die ersten Schritte zur benutzungsorientierten Evaluation im Projekt „mUltimo3D“ und verhalf mit seiner Flexibilität und seinem Organisationstalent zum zügigen Start der empirischen Arbeiten. Thorb Baumgarten bereicherte bald unser kleines Team mit seinen fachlichen und menschlichen Qualitäten, die für die Informationsübertragung von Datenformat zu Datenformat genauso wie von Psychologen zu Informatikern ausgesprochen hilfreich waren. Thomas Vöhringer-Kuhnt übernahm bereits nach kürzester Einarbeitungszeit verlässlich und routiniert verantwortungsvolle Aufgaben in der Vorbereitung, Durchführung und Auswertung von experimentellen Erhebungen. Nicole Kelling ist für ihr großes Engagement bei der Organisation und Koordination der Untersuchungsteilnehmer und in der inhaltlichen Vorbereitung und Auswertung der empirischen Untersuchung zu danken. Petra Tschoppe unterstützte dankenswerterweise die Testdurchführung.

Herrn Professor Eyferth verdanke ich den Zugang zur Thematik Mensch-Maschine-Interaktion und zur Methodik der Blickbewegungsregistrierung. Er hat mir damit mein fachliches „Zuhause“ gewiesen. Meinen Kolleginnen und Kollegen im Zentrum Mensch-Maschine-Systeme und im Fachgebiet Mensch-Maschine-Systeme bin ich für ein andauernd gutes und anregendes Arbeitsklima dankbar. Elke Fadel und Mario Lasch gilt mein besonderer Dank. Sie halfen mir immer in freundschaftlicher Weise, wenn die Zeit zu knapp war für lange Erklärungen oder für das Lesen von Rundbriefen und Handbüchern. Matthias Rötting danke ich dafür, dass er mich bei den ersten beruflichen Schritten ermuntert und mir über Jahre beharrlich zu jeder Tages- und Nachtzeit Diskussionspartner, Informationsagent, Berater und Freund war.

Die Abend- und Wochenendarbeit der vergangenen Jahre wäre nicht zu bewältigen gewesen ohne meine verständnisvolle Familie. Ich danke meinem Sohn Jacob für seine Liebe und Geduld mit einer häufig nicht ganz aufmerksamen Mutter und für die Selbstständigkeit, mit der er seine schulischen Aufgaben erfüllte. Steffen Seifert danke ich für sein selbstverständliches Sorgen um viele alltäglichen, häuslichen Dinge. Meinen Eltern und meiner Bruderfamilie bin ich für alle emotionale Unterstützung und für das in mich gesetzte Vertrauen dankbar, dass ich es trotz vieler Umwege „schaffen werde“.

# Inhalt

1	Zielstellung der Arbeit.....	1
2	Multimodale Mensch-Computer-Interaktion.....	3
2.1	Definition der multimodalen Mensch-Computer-Interaktion.....	3
2.1.1	Definition von Mensch-Computer-Interaktion.....	3
2.1.2	Definition von Multimodalität.....	4
2.1.3	Ziele der Gestaltung multimodaler Mensch-Computer-Interaktion.....	6
2.2	Bereiche multimodaler Mensch-Computer-Interaktion.....	7
2.2.1	Sprachinteraktion.....	8
2.2.1.1	Technische Aspekte der Sprachinteraktion.....	9
2.2.1.2	Psychologische Aspekte der Sprachinteraktion.....	12
2.2.1.3	Einsatzgebiete der Sprachinteraktion.....	14
2.2.2	Blickinteraktion.....	16
2.2.2.1	Technische Aspekte der Blickinteraktion.....	16
2.2.2.2	Psychologische Aspekte der Blickinteraktion.....	18
2.2.2.3	Einsatzgebiete der Blickinteraktion.....	19
2.2.3	Gestikinteraktion.....	21
2.2.3.1	Technische Aspekte der Gestikinteraktion.....	21
2.2.3.2	Psychologische Aspekte der Gestikinteraktion.....	25
2.2.3.3	Einsatzgebiete der Gestikinteraktion.....	27
2.2.4	Weitere Interaktionsformen.....	29
2.2.5	Integration mehrerer Modalitäten.....	31
3	Systementwicklung und Evaluation.....	34
3.1	Modelle der Systementwicklung.....	34
3.2	Evaluationsmodelle.....	38
3.3	Entwicklung eines Evaluationsmodells für die Entwicklung multimodaler Systeme bei nicht spezifiziertem Anwendungskontext.....	42
3.3.1	Multimodale Systementwicklung als Gegenstand der Forschung.....	42
3.3.2	Evaluation der multimodalen Interaktion an prototypischen Aufgaben.....	43
3.3.2.1	Beschreibung von Aufgaben und Aufgabenmerkmalen.....	45
3.3.2.2	Verknüpfung von Systemeigenschaften und Aufgaben.....	50
3.4	Empirische Evaluationsmethoden.....	54
3.4.1	Explorative Methoden.....	55
3.4.2	Methoden zur Erfassung von Urteilen.....	57
3.4.3	Experimentelle Methoden.....	59

4	Gegenstand der Systementwicklung: mUltimo3D .....	60
4.1	Entwicklungsziele von mUltimo3D.....	60
4.1.1	Generelle Entwicklungsziele .....	60
4.1.2	Spezifische Gestaltungsbereiche .....	60
4.2	Anwendung des Evaluationsmodells auf die Entwicklung von mUltimo3D.....	63
5	Empirische Untersuchung zur Evaluation von mUltimo3D .....	67
5.1	Ableitung von Aufgaben für die Untersuchung.....	67
5.2	Explorative Experten-Evaluation.....	68
5.2.1	Problemstellung der Experten-Evaluation.....	68
5.2.2	Durchführung der Experten-Evaluation .....	68
5.2.2.1	Beschreibung der Testanwendungen.....	68
5.2.2.2	Auswahl der Evaluatoren und Durchführung der Evaluation .....	70
5.2.3	Ergebnisse der Experten-Evaluation .....	71
5.2.4	Interpretation der Ergebnisse und Konsequenzen für die Systementwicklung.....	74
5.3	Modalitätsspezifische Untersuchungen der Blickinteraktion .....	75
5.3.1	Problemstellung und Hypothesen der Untersuchung der Blickinteraktion .....	75
5.3.2	Versuchsplan und Aufgabe für die Untersuchung der Blickinteraktion .....	78
5.3.3	Durchführung der Untersuchung der Blickinteraktion.....	80
5.3.3.1	Versuchsteilnehmer .....	80
5.3.3.2	Versuchsablauf.....	80
5.3.4	Ergebnisse der Untersuchung der Blickinteraktion.....	82
5.3.5	Interpretation der Ergebnisse der Untersuchung der Blickinteraktion.....	84
5.4	Modalitätsspezifische Untersuchungen der Sprachinteraktion.....	84
5.4.1	Problemstellung und Hypothesen der Untersuchung der Sprachinteraktion.....	84
5.4.2	Versuchsplan und Aufgabe für die Untersuchung der Sprachinteraktion.....	86
5.4.3	Durchführung der Untersuchung der Sprachinteraktion .....	87
5.4.3.1	Versuchsteilnehmer .....	87
5.4.3.2	Versuchsablauf.....	88
5.4.4	Ergebnisse der Untersuchung der Sprachinteraktion .....	88
5.4.5	Interpretation der Ergebnisse.....	90
5.5	Untersuchung prototypischer Aufgaben .....	90
5.5.1	Problemstellung und Hypothesen der Untersuchung prototypischer Aufgaben .....	90

5.5.2	Versuchsplan und Aufgabe der Untersuchung prototypischer Aufgaben .....	93
5.5.2.1	Unabhängige Variablen.....	93
5.5.2.2	Testapplikation zur Konstruktion.....	94
5.5.2.3	Testapplikation zur Informationssuche .....	96
5.5.2.4	Abhängige Variablen.....	97
5.5.3	Durchführung der Untersuchung prototypischer Aufgaben .....	100
5.5.3.1	Versuchsablauf.....	100
5.5.3.2	Versuchsteilnehmer .....	101
5.5.4	Ergebnisse der Untersuchung prototypischer Aufgaben .....	102
5.5.5	Interpretation der Ergebnisse der Untersuchung prototypischer Aufgaben .....	112
6	Integration der Ergebnisse und Diskussion .....	116
7	Verallgemeinerbarkeit des Evaluationsmodells und Ausblick.....	122
8	Literatur .....	124
9	Verzeichnisse .....	139
9.1	Abbildungsverzeichnis.....	139
9.2	Tabellenverzeichnis .....	141
10	Anhänge.....	142
	Anhang 1: Validierung der SEA-Skala am NASA-TLX.....	142
	Anhang 2: Prioritätsliste für die Gestaltung der Anwendungen.....	144
	Anhang 3: SEA-Skala.....	147
	Anhang 4: Semantisches Differential (Instruktion).....	148
	Anhang 5: NASA-TLX deutsche Fassung .....	149

## Zusammenfassung

Es gibt wohl keine technische Erfindung, die in einer ähnlich rasanten Weise Verbreitung gefunden hat wie der Computer im letzten Jahrzehnt. Seit dem Bau der ersten automatischen Rechenmaschine durch Konrad Zuse 1935 wandelte sich die Computertechnologie in beeindruckender Weise von einem Rechenautomaten zu einer interaktiven, „intelligenten“ Maschine. Dieses früher nur von ausgebildeten Experten zu bedienende Werkzeug ist heute ein Massenprodukt, das von jedem Menschen für unterschiedlichste Zwecke verstanden und benutzt werden soll. Damit dieser Anspruch erfüllt werden kann, wurden ab den 80-er Jahren verstärkt Konzepte entwickelt, die auf die Vereinfachung der Mensch-Rechner-Interaktion ausgerichtet waren. Neue Entwicklungen fokussieren die Anpassung der Computertechnik auf unterschiedlichste Anwendungssituationen und -umgebungen. Ein bedeutsamer Trend liegt dabei in der universellen Vorfindbarkeit von Computern, die auch neue Interaktionsformen zwischen Benutzer und Maschine zu entwickeln erfordert. Mit diesen neuartigen Mensch-Maschine-Schnittstellen - in Verbindung mit „intelligenten“ Software-Architekturen wie Multi-Agentensystemen - leistungsfähige und benutzungsfreundliche Computertechnik zu entwickeln, erweist sich als ein komplexes Gestaltungsproblem. Nicht allein die schnell zunehmende Menge von Hard- und Softwareentwicklungen, sondern auch die Besonderheiten der Einsatzdomäne und der Benutzer machen die Gestaltung von Systemen mit neuen Interaktionsformen und deren Bewertung zum interdisziplinären Forschungsgegenstand.

Die vorliegende Arbeit stellt eine Vorgehensweise vor, die es bereits in sehr frühen Phasen der Systementwicklung solcher neuartigen Computersysteme möglich macht, Designentscheidungen zu treffen und Empfehlungen für die Interaktionsgestaltung abzuleiten sowie erste Evidenzen für zukünftige Anwendungsdomänen zu gewinnen. Damit die Bedeutung einer Systementwicklung für unterschiedliche prototypische Aufgaben empirisch geprüft werden kann, sind Entwicklungsarbeiten zu einem System-Prototypen zu leisten. Dabei können Gestaltungsentscheidungen im Rahmen der Prototypenentwicklung bereits durch empirische Studien unterstützt werden. Dafür werden in dieser Arbeit drei Schritte der Evaluation unterschieden: (1) die explorative Expertenbefragung zum ersten lauffähigen Prototypen, (2) die experimentelle Untersuchung von Gestaltungsalternativen der einzelnen interaktiven Komponenten und (3) die experimentelle Untersuchung des integrierten Systems bezüglich der Unterstützung prototypischer Aufgaben.

Am Beispiel einer Prototypenentwicklung mit multimodalen Eingaben und autostereoskopischem Display, die im Projekt mUltimo3D<sup>1</sup> am Heinrich-Hertz-Institut für Nachrichtentechnik Berlin GmbH durchgeführt wurde, wird die empirische Begleitung der Systemgestaltung demonstriert. Neben der Darstellung des neuartigen Konzepts der multimodalen Interaktion sowie unterschiedlicher Modelle der Systementwicklung und -bewertung wird ein benut-

---

<sup>1</sup> Finanziert vom Bundesministerium für Bildung und Forschung (Förderkennzeichen 01BK410 und 01BK802) und Mitteln des Landes Berlin und durchgeführt unter Leitung von Dr.-Ing. Siegmund Pastoor.

zungsorientierter Ansatz für die informationstechnologische Grundlagenforschung im Bereich der Mensch-Rechner-Interaktion erläutert, der den Rahmen dieser Arbeit bestimmt. Die Vorstellung des Prototypen mUltimo3D schließt sich an und leitet die Darstellung der empirischen Arbeiten ein. Es werden vier Untersuchungen vorgestellt, die parallel zum Systemgestaltungsprozess durchgeführt wurden und die bei der Auswahl von Gestaltungsalternativen und der Bewertung der Nutzungseigenschaften des Prototypen hilfreich waren. Die erprobte Vorgehenssystematik lässt sich auf ähnliche Forschungs- und Entwicklungsprojekte übertragen und kann zur Erweiterung der empirischen Basis über innovative interaktive Systemkomponenten genutzt werden.



# 1 Zielstellung der Arbeit

*„Wir haben so viele technologische Lösungen,  
dafür kennen wir die Probleme noch gar nicht“*

*H.-J. Bullinger*

*(Abendvorlesung „Reale Arbeit in einer virtuellen Welt“  
auf dem 42. Kongress der Deutschen Gesellschaft für Psychologie)*

Gekoppelt an die schnell voranschreitende Leistungsfähigkeit von Hardwarekomponenten rücken auch neue Paradigmen der Mensch-Computer-Interaktion in den Bereich des Machbaren. Eines dieser gegenwärtig noch in den Kinderschuhen steckenden Paradigmen für die Entwicklung interaktiver Software heißt „multimodale Mensch-Computer-Interaktion“. Multimodale Mensch-Computer-Interaktion zielt auf die Erweiterung der Interaktion zwischen Mensch und Maschine durch neue Interaktionskanäle, die in der natürlichen Kommunikation von Menschen genutzt werden und die nun auch vom Computer erkannt und interpretiert werden können. Die Umsetzung dieses Konzeptes verknüpft Hardwarekomponenten, die vom Benutzer erzeugte Informationen aufnehmen, mit Software, die diese Informationen interpretiert und ein adäquates Systemverhalten anstößt. Deshalb wird im Rahmen dieser Arbeit von Systementwicklung i.S. der Erstellung „einer Gesamtheit geordneter Elemente, z.B. Funktionen oder technische Gebilde, die aufgrund ihrer Eigenschaften durch Relationen verknüpft und durch eine Systemgrenze umgeben sind“ (Pahl & Beitz 1997) und nicht von Softwareentwicklung gesprochen.

Neben den vordergründigen technischen Problemen, die für erste Systemrealisierungen zu überwinden sind, besitzen für interaktive Systeme Fragen der Interaktionsgestaltung zwischen den zukünftigen Nutzern und dem Computer für diese neuen Kanäle zentrale Bedeutung. Ob eine neue Technologie die Bedürfnisse ihrer potentiellen Nutzer befriedigt, trägt letztendlich wesentlich zu ihrem Erfolg und zu ihrer Verbreitung bei. Allerdings treffen die Systementwicklerinnen<sup>2</sup> im Bereich der Gestaltung der Mensch-Rechner-Kommunikation für neuartige Systeme auf viele ungeklärte Probleme, z.B.:

- Eingeführte Gestaltungsrichtlinien oder Normen können häufig die im erweiterten Gestaltungsraum auftretenden Fragen nicht beantworten helfen. Ein Grund dafür besteht darin, dass nicht alle technischen und paradigmatischen Fortentwicklungen in solchen Gestaltungshilfen antizipierbar sind.
- Am Beginn der Prototypenentwicklung kann über den zukünftigen spezifischen Anwendungskontext noch Unklarheit bestehen.

---

<sup>2</sup> Im weiteren Verlauf der Arbeit werden sowohl die männliche als auch die weibliche Form zur Kennzeichnung von Personen benutzt, dabei sind immer auch die Personen des anderen Geschlechts mit gemeint.

- Erste Systemkonzepte werden von Informationstechnik-Spezialisten, die keine Experten in den antizipierten Anwendungsdomänen sind, in Forschungsinstituten und in Forschungsabteilungen von Unternehmen umgesetzt. Die vordefinierte Ausrichtung der Prototypenentwicklung auf einen spezifischen Anwendungskontext erfordert die Generierung von Wissen über die Benutzung des Systems, z.B. situative Einschränkungen der Interaktion bei der Ausführung der Tätigkeit, wesentliche Tätigkeitsprozesse, die das System unterstützen muss oder dessen Integration in organisationale Abläufe. Dafür werden zusätzliche Ressourcen im Projekt benötigt.
- In dem Maße, in dem universellere Aussagen zur Gestaltung innovativer Systemkomponenten im Forschungsprozess gewonnen werden sollen, kann eine zu starke Ausrichtung auf einen spezifischen Anwendungsfall hinderlich sein.

Gegenwärtig besteht daher für die multimodale Systementwicklung noch intensiver Forschungsbedarf, um zu Richtlinien und Standards für die Systemauslegung in verschiedenen Anwendungskontexten zu gelangen. Für die empirische Untersuchung von Konzepten der multimodalen Mensch-Computer-Interaktion im Rahmen des gegenwärtigen Stands der Technik ist es notwendig, über einen lauffähigen Prototypen zu verfügen. In der Entwicklung des Prototypen sind Phasen zu unterscheiden, in denen Konzepte implementiert und Gestaltungsfragen mittels empirischer Methoden geklärt werden. Diese Phasen verlaufen teilweise sequentiell und teilweise parallel. Die Produkte der beiden Entwicklungsstränge Implementierung und empirische Analyse, ein weiterentwickelter Prototyp und hinzugewonnenes gestalterisches Wissen, müssen im Projektrahmen für eine erfolgreiche Entwicklung stets wieder zusammengeführt werden.

Wie aber kann für interaktive Systeme, die vorerst als Prototypen im Rahmen informationstechnischer Grundlagenforschung realisiert wurden, bereits Wissen über eine angemessene Interaktionsgestaltung gesammelt werden?

In der vorliegenden Arbeit wird zur Beantwortung dieser Frage ein Vorgehensmodell vorgeschlagen, welches aus der human- und ingenieurwissenschaftlichen Forschung entlehnte Methoden verknüpft. Im Rahmen formativer Evaluation wird die Prototypenentwicklung durch empirisch fundierte Designentscheidungen unterstützt. Am Ende des Designzyklus wird dann die Wirkung des prototypischen Systems mittels prototypischer Aufgaben summativ bewertet und die Systemmerkmale werden mit einem antizipierten Nutzungskontext verknüpft. Mit Hilfe eines solchen zweistufigen Evaluationsprozesses mit formativen und summativen Zyklen kann der Designprozess auf zunehmend höherem Informationsniveau vorangebracht werden. Die empirische Begleitung der Prototypenentwicklung bietet schließlich die Möglichkeit, die Zielerreichung des Entwicklungsprojektes zu bewerten und damit zum Erkenntnisgewinn für das Forschungsfeld beizutragen.

## **2 Multimodale Mensch-Computer-Interaktion**

### **2.1 Definition der multimodalen Mensch-Computer-Interaktion**

#### **2.1.1 Definition von Mensch-Computer-Interaktion**

Mensch-Computer-Interaktion bezeichnet eine ganze Disziplin, die sich mit dem Design, der Evaluation und der Implementation interaktiver Computersysteme für den Gebrauch durch Menschen und den damit verbundenen Phänomenen beschäftigt (ACM SIGCHI 1992).

Spezifischer steht der Begriff für den wechselseitigen Informationsaustausch zwischen Mensch und Computer und damit auch für eine wechselseitige, aufeinanderbezogene Beeinflussung des Verhaltens. Habermas (1981) versteht Interaktion unter inhaltlichen Aspekten als eine besondere Form der Kommunikation. Während für die Kommunikation allgemein kennzeichnend ist, dass wechselseitig Informationen ausgetauscht und rückgekoppelt werden mit verschiedenen Intentionen, ist mit dem Begriff der Interaktion der Informationsaustausch zum Zwecke der Handlungskoordination im speziellen gemeint (Bahrtdt 1984). Für zwischen Interaktionspartnern koordinierte Handlungen sind Intentionen abzustimmen und eine Übereinkunft über das Handlungsziel zu treffen. Pateau (1990) verweist darauf, dass unter dem formalen Aspekt die Interaktion gegenüber der Kommunikation der allgemeinere Begriff ist, dessen Bestimmungsstück die Bezogenheit des Handelns von zwei oder mehreren Personen ist. Die für diesen Wechselwirkungsprozess notwendigen Abstimmungen müssen nicht ausdrücklich wie in der Kommunikation geschehen. Bedeutsam ist allein, die Intention des anderen zu erfassen, um das eigene Handeln zustimmend oder ablehnend daran auszurichten. Die Übertragung des Interaktionsbegriffs auf das Zusammenwirken von Mensch und Computer impliziert, dass über eine kommandobasierte Kommunikation (siehe auch Geiser 1990, S. 141 ff. über verschiedene Kommunikationsformen) hinaus eine auf der Erkennung von Intentionen basierende Reaktion des Systems möglich ist. Waren solche Anforderungen an Computersysteme bisher nicht realisierbar, sehen neuere Ansätze der Gestaltung des Informationsflusses zwischen Mensch und Computer eine solche Erkennung von Benutzerintentionen vor. Der Computer wird zunehmend als ein „intelligentes Werkzeug“ konzipiert, das beispielsweise aus dem Benutzerverhalten Rückschlüsse auf dessen Expertisegrad zieht, um sich daran ohne ausdrückliche Mitwirkung des Benutzers zu adaptieren. Da die vorliegende Arbeit auf die Entwicklungsunterstützung für ein solches innovatives Interaktionskonzept, die multimodale Mensch-Computer-Interaktion, ausgerichtet ist, liegt die Verwendung des Terminus Mensch-Computer-Interaktion für die Beschreibung des wechselseitigen Informationsaustausches zwischen Mensch und Computer nahe.

## 2.1.2 Definition von Multimodalität

formationsaustausch parallel über mehrere Kommunikationskanäle stattfindet, wie dies in der zwischenmenschlichen Interaktion typischerweise geschieht. Der aus der Physiologie stammende Begriff der Modalität bezieht sich darauf, dass jedes Sinnesorgan eine spezifische Reaktionsweise auf physikalische Umgebungsmerkmale zeigt und für die Aufnahme einer bestimmten Art von Informationen, sogenannten adäquaten Reizen, spezialisiert ist, beispielsweise das gustatorische Sinnesorgan für Geschmacksreize (David 1984). In der multimodalen Mensch-Computer-Interaktion bedeutet dies, dass Verhaltensäußerungen des Benutzers beispielsweise in Form von Sprache, Gestik oder Mimik durch entsprechende rechnerseitige Sensoren aufgenommen und erkannt werden. Ein multimodales System ist zudem in der Lage, den Inhalt der Information der verschiedenen Input-Modalitäten automatisch auf einem höheren signaltechnischen Niveau (d.h. ein vom Benutzer intendiertes Kommando) zu modellieren und nicht allein ein Signal, etwa in Form einer Audio-Datei, zu übertragen (Nigay & Coutaz 1993). Der Vorteil derartiger multimodaler Interaktionen gegenüber der heute noch überwiegenden unimodalen, gestisch basierter Informationseingabe mit Tastatur und Maus wird in einer verbesserten Anpassung der Systemsteuerung an die menschlichen Leistungsvoraussetzungen gesehen. Mit der multimodalen Mensch-Computer-Interaktion ist der Begriff des *perceptual user interface* (Pentland 2000) eng verknüpft, der die notwendige computerseitige Voraussetzung für einen multimodalen Mensch-Maschine-Dialog in den perzeptiven Fähigkeiten des Interfaces fokussiert.

In Abgrenzung zu diesen Konzepten wird bei multimedialen Benutzungsoberflächen besonderes Augenmerk auf die Darbietung von Informationen gelegt, die entsprechend der perzeptiven und kognitiven Fähigkeiten von Benutzergruppen aufbereitet werden (Turk & Robertson 2000). So können die Erkenntnisse der Multimedia-Forschung für die Informationsausgabe in multimodalen Systemen genutzt werden, wobei der Aspekt der Interaktivität für multimodale Systeme definierend ist. Eine weitere Unterscheidung ist zwischen multimodaler Mensch-Computer-Interaktion und „Virtueller Realität“ (VR) zu treffen. Letztere ist darauf ausgerichtet, durch eine immersive, multisensorische Erfahrung die Illusion zu schaffen, in einer synthetischen Umgebung integriert zu sein (Schomaker et al. 1995).

Die Abbildung 1 veranschaulicht die multimodale Mensch-Rechner-Interaktion, bei der der Computer Sensoren für unterschiedliche Verhaltensäußerungen des Benutzers enthält. Die Verhaltensäußerungen können erkannt und auf der Basis von Regeln interpretiert sowie durch die regelbasierte Verknüpfung verschiedener Eingabeinformationen disambiguiert werden, um sichere Schlüsse über die Intentionen bzw. den Zustand des Benutzers zu ziehen. Das registrierte Benutzerverhalten führt dann zur rechnerseitigen Informationsausgabe, die im engeren Sinne interaktionsrelevant (Rückmeldung des erkannten Sprachbefehles) oder aufgabenrelevant (Ausführung des Befehls oder Angebot einer Änderungsoption) ist. Auch für die Informationsausgabe stehen in modernen Computersystemen unterschiedliche Medien zur

Verfügung, die spezifische Sinnesmodalitäten des Menschen ansprechen, beispielsweise akustische Ausgabemedien oder Krafrückmeldungen für die taktile oder haptische Wahrnehmung.

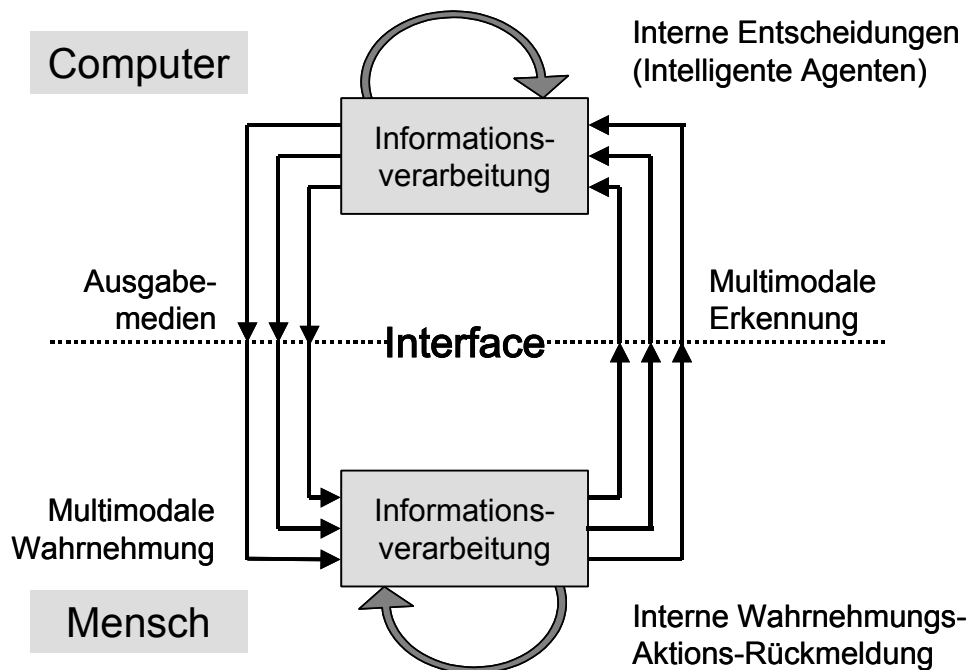


Abb. 1: Modell der multimodalen Mensch-Computer-Interaktion

Neben dem zuverlässigen technischen Funktionieren eines Interaktionskanals ist die Frage bedeutsam, welche physikalisch-technischen, physiologischen und kommunikationstheoretischen Eigenschaften einen Interaktionskanal zur Übertragung spezifischer Informationen geeignet machen. Für die Beantwortung dieser Frage sind gegenwärtig noch viele verschiedene Aspekte der multimodalen Mensch-Computer-Interaktion zu erforschen, sowohl auf der Ebene der einzelnen Modalitäten als auch auf der ihrer Integration. Da Erkenntnisse aus der natürlichen menschlichen Interaktion nicht ohne Prüfung auf den Mensch-Maschine-Dialog übertragen werden können, gilt es die Besonderheiten der multimodalen Mensch-Rechner-Kommunikation zu erfassen. Eine für den Benutzer effiziente und zufriedenstellende multimodale Interaktion erfordert es, nicht nur die angebotenen Kanäle für den Informationsaustausch günstig zu gestalten, sondern auch intelligente Verarbeitungsprozesse zu entwickeln, die nutzer- und situationsadaptive Reaktionen des Systems erzeugen.

### **2.1.3 Ziele der Gestaltung multimodaler Mensch-Computer-Interaktion**

Die Ziele, die mit der Entwicklung multimodaler Systeme verbunden werden, lassen sich nach Hedicke 2000, Oviatt 1998, Oviatt und andere 2000, Oviatt & Wahlster 1997, Vo & Waibel 1993, Vo und andere 1995, Waibel und andere 1997 zusammenfassen unter den in Abbildung 2 genannten vier Begriffen:

1. Natürlichkeit,
2. Sicherheit,
3. Flexibilität und
4. Leistungsfähigkeit der Interaktion.

Mit der Natürlichkeit der multimodalen Mensch-Computer-Interaktion ist gemeint, dass der Benutzer seine Erfahrungen aus ähnlichen Situationen der natürlichen Interaktion auf den Dialog mit dem Computer übertragen kann, dass er also wenig spezielles computerbezogenes Interaktionswissen braucht.

Unter der Interaktionssicherheit wird die Zuverlässigkeit verstanden, mit der Computer die Intentionen des Benutzers zu interpretieren vermag. Dies soll durch die Analyse verschiedener Eingabeinformationen und deren Integration zu komplexeren Interpretationsmustern erreicht werden, z.B. für die Disambiguierung einer Sprachäußerung durch die Interpretation des Blickortes.

Die Flexibilität der Interaktion bezieht sich auf örtliche, zeitliche und situative Parameter sowie personenbezogene Besonderheiten der Interaktion, z.B. für die behindertengerechte Interaktion.

Die Leistungsfähigkeit multimodaler Systeme begründet sich darin, dass die Qualität übertragbarer Information zunimmt, beispielsweise durch zusätzliche taktile Rückmeldungen, dass mehr Informationen vom Benutzer im System verarbeitet werden und dass verschiedene Kanäle der Informationsübertragung je nach Erfordernis verfügbar sind.

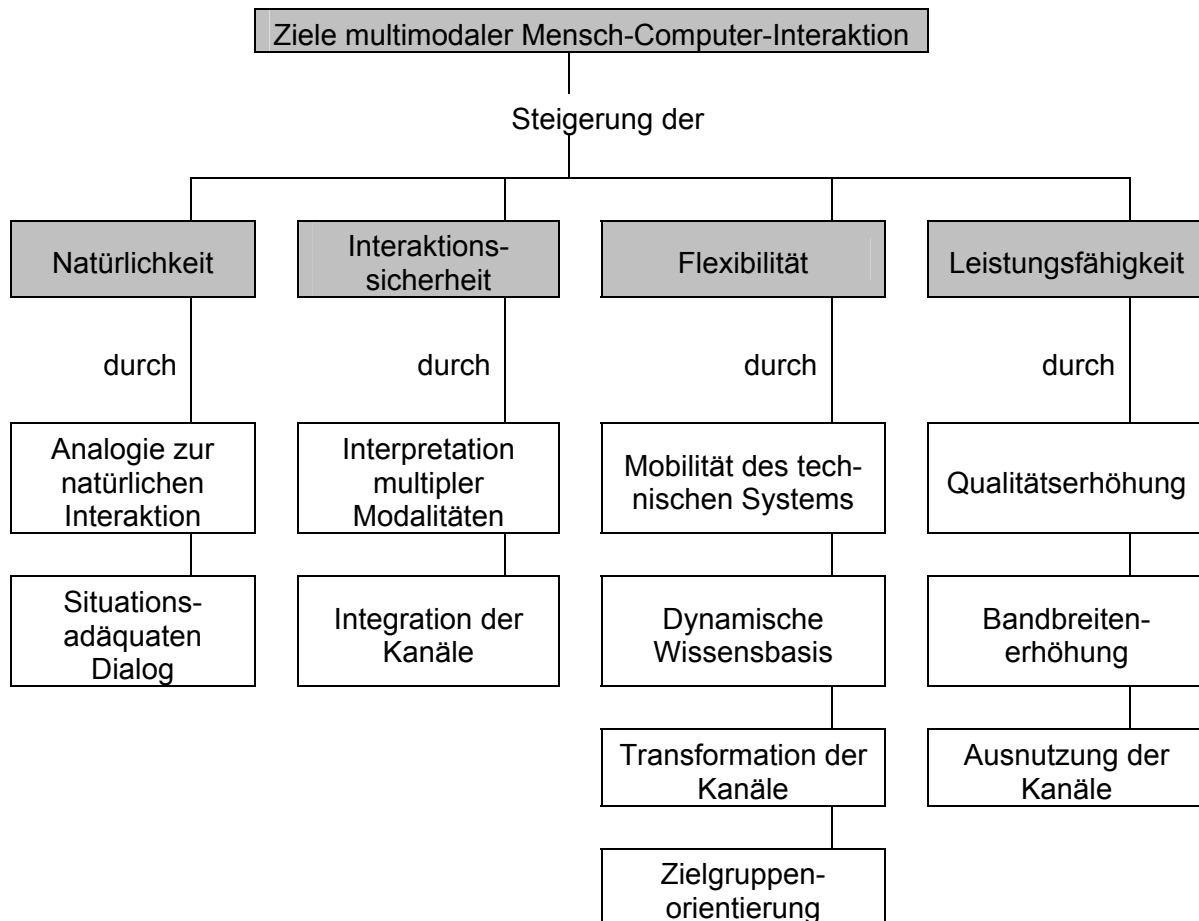


Abb. 2: Ziele der Entwicklung multimodaler Mensch-Computer-Interaktion

## 2.2 Bereiche multimodaler Mensch-Computer-Interaktion

In einer natürlichen Kommunikationssituation nehmen Menschen parallel verschiedene Verhaltensaüßerungen ihres Kommunikationspartners mit den verschiedenen verfügbaren Sinnen wahr. Diese aufgenommenen Informationen werden üblicherweise automatisch integriert und ganzheitlich interpretiert. Die verschiedenen von Benutzern erzeugten Informationen, die heute mindestens prototypisch über multimodale Schnittstellen registrierbar sind, können unter zwei interaktionsbestimmenden Gesichtspunkten charakterisiert werden:

1. dem Grad, in dem die registrierten Verhaltensaüßerungen interpretiert werden, und
2. dem Grad der Bewusstheit, mit dem die verschiedenen Verhaltensaüßerungen erzeugt werden.

Die im Kommunikationsprozess mittels verschiedener Modalitäten übermittelten Informationen können vom Rezipienten als übereinstimmend oder divergierend wahrgenommen werden. Dies wirkt sich auf die Interpretation, die Bedeutungszuweisung von wahrgenommenen Verhaltensaüßerungen, aus. Genannt seien beispielsweise die Einstufung der Dringlichkeit einer Aufforderung oder das Erkennen von Ironie. Für die multimodale Mensch-Computer-Interaktion müssen einzelne aufgenommene Informationen zu einer „Bedeutungseinheit“ in-

tegriert werden, die eine spezifische Systemantwort auslöst. Dieser Interpretationsprozess stellt eine äußerst schwierige und ressourcenfordernde Aufgabe dar. Die Problematik besteht vorrangig in der potentiellen Vielfalt kommunizierter Bedeutungen, des weiteren in der inter- und intraindividuellen Varianz, spezifische Bedeutungen zu kommunizieren und in der Notwendigkeit, eindeutige und umfassende Interpretationsregeln zu generieren (vgl. auch Roy & Pentland 1997). Lösungsansätze für diese Problematik bestehen in einer starken Beschränkung des situativen Kontexts, etwa der Bearbeitung weniger Aufgaben innerhalb der multimodalen Interaktion (z.B. Cohen et al. 1997), der Berücksichtigung sehr weniger Verhaltensäußerungen, etwa einer Zeigegeste und eines sprachlichen Kommandos (z.B. Bolt 1980) oder der gezielten Suche nach Verhaltensindikatoren für nur wenige „Bedeutungen“, z.B. Verhaltensindikatoren, die als Überraschung interpretiert werden (z.B. Ark, Dryer & Lu 1999).

Verhaltensäußerungen des Benutzers werden mehr oder minder bewusst erzeugt bzw. kontrolliert. Dies muss bei deren Analyse und Interpretation berücksichtigt werden. Verhaltensäußerungen, die auch unwillkürlich auftreten können, beispielsweise Blickbewegungen, werden u.U. durch Umgebungsbedingungen hervorgerufen, die das multimodale Computersystem gar nicht registriert. Werden solche Verhaltensäußerungen ohne weitere Prüfung als interaktionsrelevanter Ausdruck von Intentionen interpretiert, können unangemessene Systemantworten daraus folgen. Diese lassen dem Benutzer das System intransparent erscheinen. Da auch die hochgradig unter bewusster Kontrolle stehenden Verhaltensäußerungen, wie verbale Äußerungen, häufig uneindeutig und schwierig zu interpretieren sind, ist es grundsätzlich sinnvoll, auch weniger kontrolliert erzeugtes Verhalten zur Auflösung von Mehrdeutigkeit mit zu berücksichtigen. Die parallele Analyse verschiedener vom Benutzer erzeugter Informationen stellt nach wie vor eine Herausforderung für die Gestaltung multimodaler Mensch-Computer-Interaktion dar. Dabei spielen die Probleme der zeitlichen Koordination von registrierten Verhaltensäußerungen und die Zusammenfassung der verschiedenen Informationen zu einer eindeutigen Interpretation eine hervorragende Rolle (Lenzmann 1998) (siehe auch Abschnitt 2.2.5). In dem Maße, in dem es gelingt, die vielfältigen Verhaltensäußerungen von Benutzerinnen zuverlässig zu interpretieren, kann die multimodale Mensch-Computer-Interaktion zu einer „intentionsbasierten“ Kommunikation ausgebaut werden.

### **2.2.1 Sprachinteraktion**

Der Sprachinteraktion wird im Rahmen multimodaler Systementwicklung eine zentrale Rolle beigemessen, da sie aufgrund ihrer Vertrautheit aus der zwischenmenschlichen Kommunikation als ein intuitiver Zugang zur Technik betrachtet wird. Schukat-Talamazzini (1995) verweist darauf, dass über gesprochene Kommunikation zwischen 120 - 250 Wörter/Minute ohne Übung übertragbar sind, während geschultes Personal über die Tastatur nur 100 - 150 Wörter/Minute einzugeben vermag. Allerdings stehen dem breiten Einsatz der sprachbasierten Mensch-Computer-Interaktion noch verschiedene Probleme entgegen, die im Folgenden bei der Skizzierung der technischen Ansätze und einiger in diesem Rahmen bedeutsam erscheinenden psychologischen Gesichtspunkte der Sprachinteraktion kurz angerissen werden. Eine



umfassende und aktuelle Einführung in die technischen und physiologischen Zusammenhänge der Sprachkommunikation zwischen Mensch und Maschine findet sich bei O'Shaughnessy (2000).

### **2.2.1.1 Technische Aspekte der Sprachinteraktion**

Als Komponenten der verbalen Mensch-Computer-Interaktion lassen sich die automatische Spracherkennung und die Sprachsynthese unterscheiden.

Die automatische Spracherkennung für die multimodale Mensch-Rechner-Interaktion dient dazu, die sprachliche Äußerung eines Benutzers zu analysieren und in eine adäquate rechnerinterne Darstellung zu überführen, um eine angemessene bzw. vom Benutzer intendierte Systemreaktion zu veranlassen.

Schukat-Talamazzini (1995) gruppiert die Probleme, die bei der automatischen Spracherkennung zu überwinden sind, in nachstehende Klassen:

1. **Kontinuität:** die Grenzen zwischen möglichen Analyseeinheiten, z.B. Wörtern oder Silben, sind nicht eindeutig aus dem Signal bestimmbar.
2. **Variabilität:** eine spezifische Spracheinheit wird akustisch stets unterschiedlich realisiert, z.B. aufgrund akustischer Störquellen, der Sprechweise oder kontextueller Aussprachevariationen durch die jeweilige lautsprachliche Umgebung. Dies erschwert einerseits deren automatische Zuordnung zu einem prototypischen Referenzmuster und andererseits die Abgrenzung von anderen Mustern.
3. **Komplexität:** die Anforderungen an die Rechenleistung und Speicherkapazität von Computern zur Spracherkennung sind sehr hoch durch die hohe Abtastfrequenz, die große Menge von Erkennungseinheiten, die als Referenzen gespeichert und deren Abgleich mit eingehenden Signalen berechnet werden muss, und die vielfältigen Kombinationsmöglichkeiten bei der Satzbildung.
4. **Ambiguität:** Mehrdeutigkeiten zwischen Spracheinheiten und ihren Realisierungen können auf der phonologischen, syntaktischen und semantischen Ebene sowie bei Komposita oder Derivationen auftreten.

Mit den unterschiedlichen Anforderungen, die an die Funktionsweise eines automatischen Spracherkennungssystems gestellt werden bzgl. der zu analysierenden Einheiten (fließend gesprochene Sprache oder Einzelworterkennung) oder des Umfangs des zu erkennenden Vokabulars sowie des Trainingsaufwandes (sprecherabhängig oder sprecherunabhängig), sind diese generellen Schwierigkeiten in leicht variierendem Maße verknüpft.

Für die automatische Spracherkennung werden üblicherweise folgende Prozesse durchlaufen (Ainsworths 1999):

1. Die Vorverarbeitung, die das über das Mikrofon eingehende analoge Signal digitalisiert.
2. Die Merkmalsextraktion, mittels derer solche Signaleigenschaften hervorgehoben werden, die der Identifikation der Äußerung dienen und solche ausblenden, welche auf Eigenheiten der Sprecherin, auf Umgebungseinflüssen und elektrischen Übertragungseigenschaften beruhen.
3. Die Mustererkennung, bei der die aus dem eingehenden Signal berechneten Merkmale mit gespeicherten Referenzmustern von Sprachsignalen, die sich aus unterschiedlichen lautsprachlichen Einheiten konstituieren können, verglichen werden.

Die Merkmalsextraktion basiert zumeist auf einer Frequenzanalyse, welche zu einer Repräsentation des Sprachlautes als Frequenz-Intensitäts-Zeit-Struktur führt. Dabei werden innerhalb kurzer Zeitfenster Merkmale berechnet, wobei mit verschiedenen Transformationsverfahren, beispielsweise der Fouriertransformation (vgl. z.B. Press et al. 1989), das Signal von Störgrößen bereinigt wird. Die Merkmale setzen sich aus einer Sequenz von Vektoren zusammen, die die spektrale Energieverteilung des zugehörigen Signalausschnitts abbilden (Ainsworths 1999). Für die Mustererkennung stehen verschiedene Ansätze zur Verfügung, die die aus dem Signal berechneten Merkmale mit gespeicherten Mustern vergleichen, um sie zu identifizieren. Neben den Klassifikationsverfahren der Dynamischen Programmierung und der künstlichen neuronalen Netze werden aktuell vorrangig Hidden-Markov-Modelle (HMM) (siehe auch Rabiner & Juang 1993, Ruske 1994) zur Mustererkennung eingesetzt (vgl. Kamm & Helander 1997). Der Einbezug von linguistischen Kategorien, beispielsweise der Syntaktik oder des Sprechrhythmus, in die automatische Spracherkennung können die Zuverlässigkeit der Erkennung weiter erhöhen. Dieses in Expertensystemen abgelegte Wissen kann in zukünftigen Spracherkennungssystemen dazu eingesetzt werden, die durch die statistischen Methoden gewonnenen Ergebnisse zu evaluieren (O'Shaughnessy 2000). Bei der multimodalen Mensch-Computer-Interaktion kann die *word spotting* Technik (Kamm & Helander 1997) gewinnbringend eingesetzt werden, bei welcher Zielwörter oder -phrasen aus dem kontinuierlichen Sprachfluss erkannt werden. Die Technik ermöglicht es dem Benutzer, normal zu sprechen, ist jedoch nur für die Erkennung eines kleineren Vokabulars geeignet.

Im Folgenden soll kurz auf die technischen Aspekte der Sprachsynthese eingegangen werden. Für die Sprachausgabe werden von Burkhardt (2000) drei Varianten der künstlichen Sprach-erzeugung angeführt:

- Die Sprachwiedergabe, die zuvor aufgenommene Sprachsegmente von mindestens Wortgröße, u.U. in veränderter Reihenfolge, wiedergibt.
- Die Sprachsynthese, die eine lautsprachliche Repräsentation schriftsprachlicher Einheiten generiert.
- Die Resynthesverfahren, die nur in der Forschung verwendet werden, um ausge- wählte Merkmale von Sprachsignalen zu manipulieren.

Der Prozess der Sprachsynthese ist abstrahiert in Abbildung 3 nach O'Shaughnessy (2000) dargestellt.

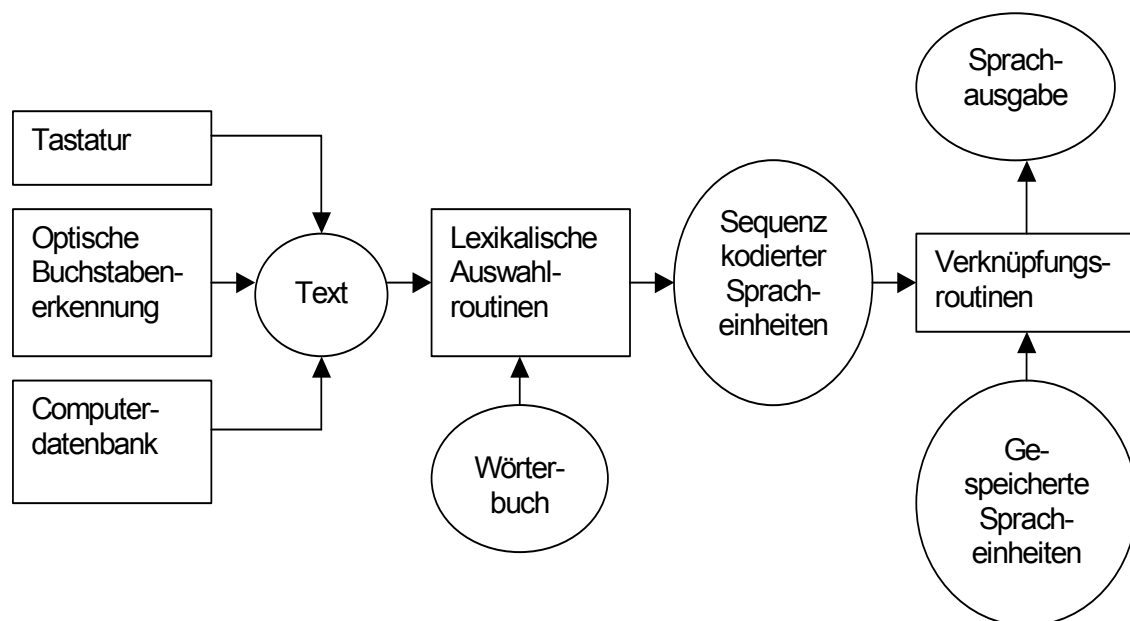


Abb. 3: Blockdiagramm der Sprachsynthese (nach O'Shaughnessy 2000)

Bei der Sprachsynthese muss zwischen der Effizienz der Kodierungsmethoden und der Quali-tät der erzeugten Sprache abgewogen werden: die Speicherung größerer Spracheinheiten (Phrasen oder Sätze) steigert die Sprachqualität, erfordert jedoch sehr viel Speicherplatz.

Ein Ansatz für umschriebene Kontexte, zum Beispiel bei der Fahrplanauskunft, besteht darin, dass größere Einheiten für wiederkehrende Äußerungen verwendet werden, die Freistellen aufweisen, in welche aus kleineren Spracheinheiten generierte flexible Inhalte eingefügt werden (Spiegel & Streeter 1997). Bei uneingeschränktem Text, der synthetisiert werden soll, wird die Sprache aus kleineren Einheiten, beispielsweise Silben oder Halbsilben, erzeugt.

Die Überführung von Texten in gesprochene Sprache stößt bereits bei der Textanalyse auf Schwierigkeiten: so sind Abkürzungen zu interpretieren sowie Akronyme und Eigennamen phonetisch richtig umzusetzen (Burkhardt 2000). Weitere Probleme, die insbesondere die Natürlichkeit und Verständlichkeit der Sprachausgabe beeinflussen, sind die Diskontinuitäten, die zwischen den verknüpften Sprachsegmenten entstehen, und die gering variierenden Prosodie der ausgegebenen Sprache, d.h. der Sprechmelodie und des -rhythmus.

Für die Gestaltung von Sprachausgabe muss neben den linguistischen und informatorischen Aspekten von Sprache auch darauf geachtet werden, welche Eindrücke durch die Wahl der Stimme, der Sprechgeschwindigkeit und Sprechfrequenz bei der Hörerin hervorgerufen werden (Spiegel & Streeter 1997). Die damit assoziierten vokalen Stereotypen können sich auf die Akzeptanz von Sprachausgabe sowohl positiv als auch negativ auswirken.

### **2.2.1.2 Psychologische Aspekte der Sprachinteraktion**

Die Sprache bildet das mächtigste und differenzierteste Interaktionssystem des Menschen, dem sich die psychologische Forschung aus verschiedenen Perspektiven nähert. Sprachliche Kommunikation lässt sich in die Sprachproduktion und die Sprachrezeption untergliedern. Nach Herrmann (1992) umfasst die Sprachproduktion die Erzeugung geordneter Folgen von Sprachlauten und die psychischen Prozesse, die diesem Vorgang planerisch vorausgehen und regulierend beigeordnet sind. Zur Sprachproduktion wird auch die Erzeugung nicht-lautlicher, geschriebener Sprache gezählt.

Die Sprachrezeption wird definiert als die Wahrnehmung partnerseitigen Sprechens mit den psychischen Prozessen, in denen das Gehörte kognitiv und emotional weiterverarbeitet wird, sowie die handlungsbezogene Nutzung der verstandenen Äußerungen des Partners. Auch auf Seiten der Sprachrezeption ist das Lesen von Geschriebenem oder Gedrucktem eingeschlossen.

Winograd (1983) beschreibt sprachliche Kommunikation als einen informationsverarbeitenden Prozess, der im Sprecher und im Hörer abläuft, und bei dem Wissen über die Sprache, die Welt und die Situation einfließt. Verstehen, bei dem es um die Rekonstruktion von Bedeutungen geht, setzt voraus, dass wenigstens teilweise gemeinsames Wissen vorhanden ist.

Die sprachliche Kommunikation wird in der Psychologie häufig durch Sprecher-Hörer- bzw. Sender-Empfänger-Modelle veranschaulicht. Abbildung 4 stellt ein solches allgemeines Kommunikationsmodell dar. Der Diskurs zwischen Sprechern und Hörern wird hier als ein mehrstufiger Prozess der Generierung und der Dekodierung einer Nachricht aufgefasst. Die Stufen dieses Prozesses können teilweise parallel verlaufen (Herrmann & Hoppe-Graff 1988). Es werden dabei aus einem vorhandenen „Bedeutungsvorrat“ von subjektiven Repräsentationen von Dingen, Sachverhalten usw., eine bestimmte Menge von Inhalten ausgewählt und in eine Sequenz gebracht, die dann durch ausgewählte Zeichen, z.B. Wörter, repräsentiert werden. Die entstandene Zeichensequenz wird dem Empfänger übermittelt. Der Empfänger kann

bei ausreichender Übereinstimmung zwischen seinem Zeichenvorrat und der eingehenden Zeichensequenz die Nachricht dekodieren und ihr somit aus seinem Bedeutungsvorrat Inhalte zuordnen. Das Verstehen einer Nachricht besteht demnach in der Rekonstruktion ihres Bedeutungsgehaltes. Probleme, die das Gelingen der sprachlichen Kommunikation beeinträchtigen, lassen sich an diesem Modell in verschiedenen Bereichen ableiten. Dem Sender kann es misslingen, adäquate Inhalte auszuwählen und in eine Reihenfolge zu bringen oder den Bedeutungen passende Zeichen zuzuordnen. Der Zeichenvorrat kann zu begrenzt sein, um die ausgewählten Inhalte zu vermitteln oder es werden unpassende Zeichensequenzen produziert. Die Nachrichtenübertragung kann gestört werden, z.B. durch Umgebungsgeräusche. Dem Empfänger kann es am probaten Zeichenvorrat mangeln oder den Zeichen werden falsche Bedeutungen zugeordnet.

Generell ist für sprachliche Äußerungen zu beachten, dass sie nicht überdauernd sind, so dass sie leicht verloren gehen oder vergessen werden können. Diese Probleme beeinflussen auch die sprachbasierte Mensch-Computer-Interaktion, so dass es notwendig ist, Strategien zu entwickeln, die in Anlehnung an zwischenmenschliche Strategien dazu beitragen, das Verstehen zwischen Gesprächspartnern zu unterstützen, z.B. Integration von Domänenwissen oder Diskurswissen.

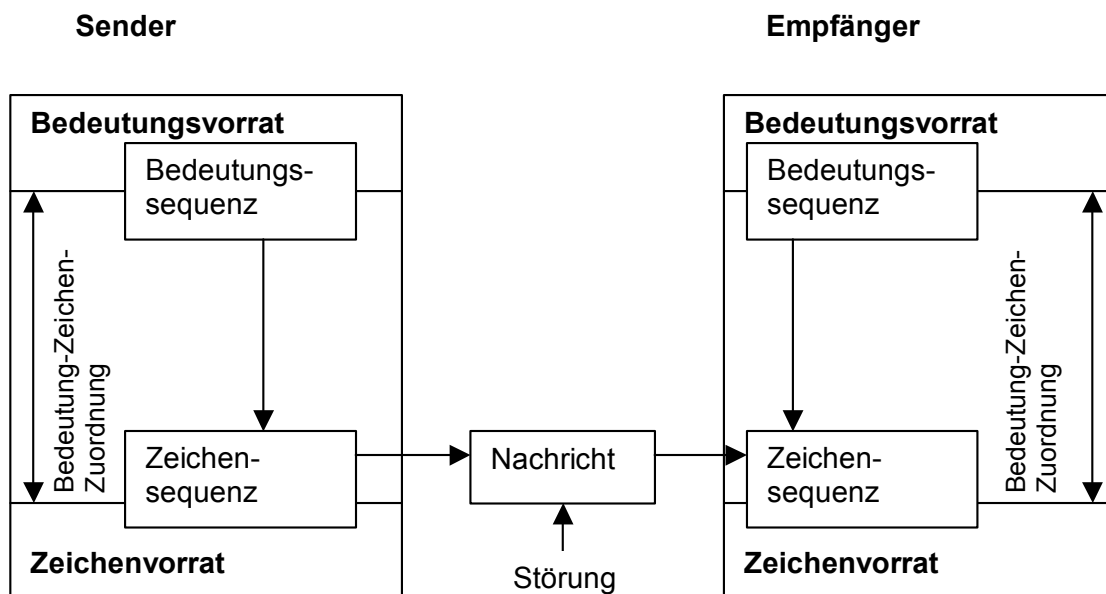


Abb. 4: Allgemeines psychologisches Kommunikationsmodell (nach Herrmann 1995)

Herrmann (1995) kritisiert an den dichotomen Sender-Empfänger-Modellen, dass sie weder den dynamischen Aspekt noch die Situations- und Kontextabhängigkeit von sprachlicher Kommunikation abbilden.

Direkte sprachliche Kommunikationssituationen sind demnach dadurch geprägt, dass:

- die Rollen von Sprecher und Hörer gewechselt werden,
- der Austausch von Nachrichten zielgerichtet und partnerbezogen ist,
- der aktuelle Sprecher sich in der Regel auf die vorangehende Äußerung bezieht,
- sowohl die Produktion als auch die Rezeption bzw. Interpretation einer Nachricht durch situative Merkmale beeinflusst wird,
- die Gesprächspartner eine Vorstellung über die jeweilige Sprecherabsicht haben,
- sich im Verlaufe der Kommunikation die interne Repräsentation verändert (Wissen wird aufgebaut, modifiziert etc.),
- für sie soziale Konventionen existieren,
- non-verbale Informationen die Rezeption des Gesprochenen beeinflussen.

Für die funktionale Perspektive auf sprachliche Kommunikation bietet die Sprechakttheorie (Austin 1962, Searle 1969) einen eingeführten Rahmen. In der Sprechakttheorie werden die sprachlichen Äußerungen als Handlungen klassifiziert (Hickmann 2000), die direkt auf das Sprechen (lokutionärer Akt), auf den Inhalt der Äußerung (illokutionärer Akt) oder auf den Effekt der Äußerung (perlokutionärer Akt) bezogen sind. Eine Äußerung wird in einer bestimmten Weise verwendet, etwa als Warnung, und kann auf einen Hörer über den Gesprächsrahmen hinaus verhaltensmodifizierend wirken. Dieser perlokutionäre Akt wird nicht mehr durch Sprachkonventionen nach Regeln vollzogen, sondern vielmehr durch die Verknüpfung der Äußerung mit dem prozeduralen, d.h. dem anwendungsbezogenen (vgl. Opwis 1988) Wissen des Hörers. Die Erkenntnisse aus der Analyse des funktionalen Zusammenhangs von Gesprochenem und der hervorgerufenen oder intendierten Wirkung beim Rezipienten kann dazu verwendet werden, sprachbasierte Mensch-Computer-Interaktion zu gestalten. Hierfür ist es notwendig, den kontextspezifischen Rahmen für die handlungsgenerierende Interpretation einer Äußerung abzuleiten.

### **2.2.1.3 Einsatzgebiete der Sprachinteraktion**

Die Sprachinteraktion als integraler Bestandteil multimodaler Mensch-Computer-Interaktion kann je nach dem Anwendungsfokus eines Systems auf unterschiedlichen Performanzstufen realisiert sein. Je nach dem funktionalen Umfang, der für die Aufgabenbearbeitung notwendig ist, ist für eine adäquate Sprachinteraktion ein mehr oder minder großer Wortschatz erforderlich, der die verschiedenen inhaltlichen Bereiche der Interaktion hinreichend abdeckt. Für multimodale Systeme, die für öffentliche Bereiche oder für viele Benutzer Unterstützung bieten sollen, ist eine sprecherunabhängige gegenüber einer sprecherabhängigen Spracherken-

nung vorzuziehen. Sprecherabhängige Systeme, deren Erkennungsleistung für umfangreiche Wortschätze trainingsbedingt höher ist, sind dort akzeptabel, wo ein oder sehr wenige Benutzer ein großes domänenspezifisches Vokabular benötigen und eine längerfristige und häufige Nutzung des Systems wahrscheinlich ist, z.B. in der Histopathologie beim Mikroskopieren, wo Hände und Augen beschäftigt sind und somit die gleichzeitige Steuerung einer herkömmlichen Computeranwendung unmöglich ist.

Grasso, Ebert und Finin (1998) verweisen in einem Vergleich zwischen direkt-manipulativen und sprachbasierten Benutzungsoberflächen darauf, dass Sprachinteraktion:

- Augen- und Handfreiheit gewährt,
- kompliziertere und verknüpfte Aktionen ermöglicht,
- nicht notwendigerweise sichtbare Bezüge benötigt und
- unterschiedliche Möglichkeiten bietet, auf die selben Entitäten zu verweisen.

Sprachbasierte Systeme, die fließend gesprochene Sprache mit hinreichender Zuverlässigkeit erkennen und einen Dialog mit dem Benutzer führen oder gewünschte Aufträge erfüllen können, befinden sich trotz langjähriger Forschung noch in einem prototypischen Status. Diese prototypischen Systeme werden häufig dazu genutzt, Prinzipien der sprachlichen Mensch-Computer-Interaktion in verschiedenen Anwendungen zu erforschen. Ein Beispiel dafür ist „SpeechActs“ für E-Mail-, Kalender-, Börsennachrichten- und Wettervorhersage-Anwendungen (Yankelovich, Levow & Marx 1995). Jörding und Wachsmuth (2001) zeigen an ihrem sprachlich zu steuernden anthropomorphen Agenten „Hamilton“ die Schwierigkeiten auf, die selbst bei der Interpretation einfacher sprachlicher Äußerungen über räumliche Bezüge auftauchen. Eine Äußerung wie „Gehe nach rechts“ kann einmal aus der Perspektive des Benutzers oder aus der des Agenten interpretiert werden. Welchen Referenzrahmen der Benutzer bei seiner Anweisung gewählt hat, kann nicht ohne weiteres erschlossen werden. Häufig wird für die Klärung mehrdeutiger oder unterspezifizierter Äußerungen ein expliziter oder impliziter Aushandlungsprozess benötigt, für den Diskursmodelle unter Einsatz verschiedener Modellierungsmethoden entwickelt wurden (vgl. z.B. Alexandersson et al. 1998, Cheyer & Julia 1999a, Denecke 1997, Heeman et al. 1998, Stolcke et al. 1998, Zue 1997). Die automatische Spracherkennung für die Mensch-Computer-Interaktion wird seit einigen Jahren auch durch Benutzermodellierungsansätze aus der KI unterlegt. Es besteht die Hoffnung, dass Benutzermodelle, die Wissen über die Präferenzen, den Expertisegrad u.ä. Merkmale des Benutzers zur Verfügung stellen, dazu beitragen können, auf den Benutzer abgestimmte und relevante Systemantworten bezogen auf einen verbalen Dialog oder eine Aufgabe zu generieren (Zukerman & Litman 2001).

Die Bedeutung, die der Spracherkennung und -synthese für die Mensch-Computer-Interaktion beigemessen wird, zeigt sich auch darin, dass fast alle gegenwärtig vom Bundesministerium für Bildung und Forschung (BMBF) geförderten Leitprojekte „Mensch-Technik-Interaktion“

sprachbasierte Schnittstellen für die unterschiedlichen Anwendungen explizit in ihrem Programm aufführen (vgl. DLR 2001). In den verschiedenen Leitprojekten wird sprachbasierte Interaktion für technische Geräte im Haushalt, im Kraftfahrzeug, in öffentlichen Terminals, für die Steuerung teilautonomer Roboter in Industrie und Haushalt, für Arbeitsplatzrechner oder für mobile Endgeräte prototypisch realisiert. Teilweise stehen diese Anwendungsbereiche schon längere Zeit im Fokus der Sprachtechnologie, speziell zur Bedienung von Systemen im Kraftfahrzeug (vgl. z.B. Aldridge & Lansdown 1999, Bengler et al. 2000) oder in der Robotersteuerung (vgl. z.B. Lueth et al. 1994). Die simultane Übersetzung in der Telephonie (Wahlster 2000), die Prozesskontrolle in der Flugsicherung (Bierwagen & Vielhauer 2000) oder die verbale Navigation im Internet (Morley et al. 1999) sind weitere zukunftssträchtige Anwendungen für die automatische Spracherkennung und die Sprachsynthese. Im Projekt MASEA (Luczak 1997) wurden umfangreiche arbeitspsychologische Studien über den Einsatz der Spracheingabe in verschiedenen Bereichen der Fertigung durchgeführt. Generell wurde die Spracheingabe u.a. für solche Bereiche als besonders dienlich beschrieben, die geprägt sind durch extreme Umgebungsbedingungen, in denen Schreiben unmöglich ist, für die parallele Protokollierung oder Steuerung neben anderen manuellen Tätigkeiten erforderlich ist oder für mobile Tätigkeiten.

## **2.2.2 Blickinteraktion**

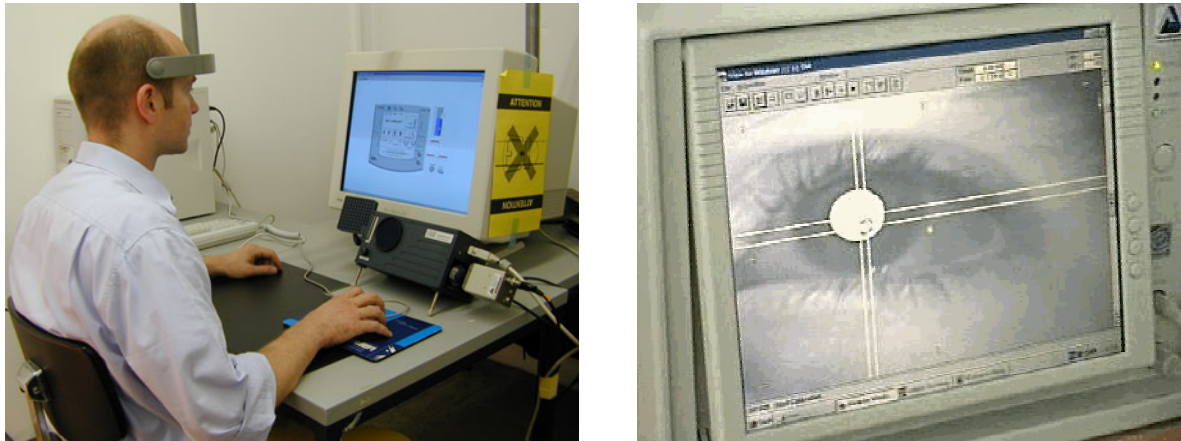
Die Blickinteraktion wird häufig in virtuellen oder so genannten angereicherten (*augmented*) Umgebungen eingesetzt, weniger in multimodalen Systemen. Die räumliche Auflösung der heutigen Messtechnik, die zudem weitgehend beeinträchtigungsfrei sein muss, genügt nicht, um auf üblichen Computerdisplays Schaltflächen zu differenzieren. In virtuellen Umgebungen und auf größeren Displays kann blickgestützte Interaktion mit sinnvoller Funktionalität hinterlegt werden, wie die unter 2.2.2.3 aufgeführten Beispiele belegen.

### **2.2.2.1 Technische Aspekte der Blickinteraktion**

Für die Messung der Blickposition von Benutzerinnen stehen verschiedene Methoden zur Verfügung, die bereits als Messsysteme realisiert sind. Einen Überblick über die verschiedenen Methoden zur Messung der Blickposition gibt Rötting (1999). Ergänzend finden sich Hinweise auf aktuelle kommerzielle Messsysteme im Internet bei ZMMS Eyes Tea (2002). Diese Systeme sind jedoch für den Einsatz im Bereich der multimodalen Mensch-Computer-Interaktion weniger geeignet, da sie üblicherweise weder online Datenanalysen noch die Kommunikation mit anderen Systemen unterstützen. Da für die Mensch-Computer-Interaktion kontaktfreie Methoden der Blickpositionserfassung wegen des Komforts und der Einfachheit der Benutzung wünschenswert sind, erscheinen Verfahren wie das Elektro-Okulogramm (EOG), kontaktlinsenbasierte Verfahren und solche Messsysteme, die eine auf dem Kopf zu tragende Messeinheit benötigen, als ungeeignet. Kontaktfrei messende Systeme für die Blickpositionsbestimmung basieren üblicherweise auf der automatischen Analyse definierter Merkmale im Videobild des Auges, die sich systematisch bei Blickpositionsverände-



rungen modifizieren, z.B. die Verschiebung der als dunkelsten Bereich gut detektierbaren Pupille gegenüber dem mit einer punktförmigen Infrarotlichtquelle erzeugten Cornea-Reflex (siehe Abbildung 5).



*Abb. 5: Blickrichtungsbestimmung aus dem Verhältnis von Pupillenmittelpunkt und cornealem Reflex (rechts: Fadenkreuze durch Pupillenschwerpunkt und Cornea-reflex) bei einem kontaktfreien Messsystem (links: iview-System der Firma SMI)*

Damit die Blickrichtung zuverlässig und ohne größere Bewegungseinschränkungen des Benutzers berechnet werden kann, ist es notwendig, dass die Aufnahmekamera bei Kopfbewegungen automatisch nachführt und mit den Kopfbewegungen verbundene Fehler kompensiert werden können. Dazu kann es günstig sein, die Kopfposition mit einem separaten Messsystem zu erfassen. Dieser Ansatz wurde beispielsweise bei dem unter 4.1.2 näher erläuterten System mUltimo3D gewählt. Für viele kontaktfreie Blickbewegungsregistriersysteme galt bislang, dass Bewegungen in der z-Achse, d.h. auf die Kamera zu, schwer ausgeglichen werden konnten. Mit neuen Autofokuskameras kann dieses Problem jedoch ausgeräumt werden, so dass der Bewegungsfreiraum der Benutzerinnen erhöht wird.

Zur interaktionsbezogenen Interpretation des Blickes von Benutzern können grundsätzlich unverrechnete Messdaten der Registriereinheit eingesetzt werden oder parametrisierte Daten, insbesondere Fixationen, bei denen das Auge auf ein Objekt ausgerichtet ist und in welchen weitestgehend die visuelle Informationsaufnahme stattfindet (Matin 1974). Für die systemseitige Interaktionsführung können noch die Zeitcharakteristika oder die räumliche Verteilung der Fixationen analysiert werden, um daraus Schlussfolgerungen über die Aufgabenbewältigung und eventuellen Unterstützungsbedarf zu ziehen (vgl. Baumgarten 2002).

Für eine zufriedenstellende blickbasierte Mensch-Computer-Interaktion muss eine hinreichend hohe Messfrequenz (ca. 50 Hz) und eine adäquate Geschwindigkeit der online Analyse der Messwerte für die Parametrisierung vorausgesetzt werden, um die Interaktionsgeschwindigkeit nicht zu beeinträchtigen. Die örtliche Genauigkeit muss gut genug sein, um Blicke

zuverlässig zu blickinteraktionsrelevanten Arealen in der grafischen Benutzungsoberfläche zuordnen zu können.

Da es zwischen Personen Unterschiede darin gibt, wie sich die für die Registrierung genutzten Merkmale des Auges bei Blicksprüngen ändern, bedarf es für eine genaue Messung einer Kalibrierung. Die Kalibrierung des Systems auf den jeweiligen Benutzer sollte möglichst unaufwendig sein und über eine Rückmeldung der Kalibriergüte verfügen. Eine solche vereinfachte Ein-Punkt-Kalibrierung mit grafischer Rückmeldung der Kalibriergüte wurde beispielsweise im Projekt mUltimo3D am Heinrich-Hertz-Institut für Nachrichtentechnik Berlin entwickelt. Um bei der multimodalen Interaktion an einem autostereoskopischen Bildschirm die Blickposition einfach integrieren zu können, werden die Kalibrierdaten eines passenden Referenznutzers ausgewählt.

Die Messgenauigkeit der bekannten Messsysteme nimmt üblicherweise über längere Messzeiträume ab. Die dann nötigen Rekalibrierungen können im Interaktionsverlauf „versteckt“ werden, wozu sich aufleuchtende oder bewegende Reize im peripheren Sehbereich eignen, da sie üblicherweise eine unwillkürliche Blickbewegung zu ihnen hin auslösen.

### **2.2.2.2 Psychologische Aspekte der Blickinteraktion**

In der natürlichen Kommunikation spielen verbale und nonverbale Äußerungen zusammen. Die mimischen Ausdrucksformen werden hauptsächlich von der Augen- und Mundpartie gesteuert. In interkulturellen Studien konnte belegt werden, dass die grundlegenden menschlichen Emotionen über den mimischen Ausdruck zuverlässig über Kulturkreise hinweg erkannt werden (Ekman 1972). Kulturell geformte Regeln der nonverbalen Kommunikation umfassen in starkem Maße auch die Art und Weise des Blickkontaktes. Der mimische Ausdruck, den die Art des Blickens wesentlich beeinflusst, trägt bedeutsam zur nonverbal vermittelten Wahrnehmung der emotionalen Befindlichkeit durch andere Personen bei. In einem natürlichen Dialog signalisiert der Blick auf den Gesprächspartner bereits vor dessen Ansprache, dass eine Äußerung an ihn gerichtet ist. Maglio und andere (2000) konnten zeigen, dass dies auch für technische „Gesprächspartner“ gilt. Ein Problem der sprachbasierten Mensch-Computer-Interaktion, bei dem der Blickort Mehrdeutigkeiten auszuräumen verhilft, besteht darin, den gemeinten Gesprächspartner zu ermitteln. Ob eine Äußerung des Benutzers an den Computer oder eine andere Person im Raum bzw. weitere technische Einrichtungen mit Spracherkennungskomponenten gerichtet ist, lässt sich durch die zusätzliche Analyse des Blickes des Sprechers identifizieren.

Die Ausrichtung des Blickes kann einem menschlichen oder auch einem „maschinellen“ Kommunikationspartner den Fokus der gegenwärtigen Aufmerksamkeit vermitteln (Driver et al. 1999, Friesen & Kingstone 1998, Langton, Roger & Bruce 2000). Das sogenannte „Visuelle Zeigen“ wurde von Gullberg und Holmqvist (1999) beim Erzählen gefunden, wobei der Hörer seinen Blick auf reale Objekte, die vom Erzähler angeschaut werden, ausrichtet.

Bedingt durch die Charakteristika der Anatomie und Physiologie des menschlichen Auges ist keine hohe örtliche Genauigkeit zu erwarten, es kann vielmehr nur auf die bestimmten Regionen zugeteilte visuelle Aufmerksamkeit geschlossen werden. Eine Problematik bei der Interpretation von Blicken besteht in dem weithin bekannten Phänomen des „Ins-Leere-Sehens“, während die Aufmerksamkeit auf personeninterne Vorgänge ausgerichtet ist. Für eine zielgerichtete Interaktion über einen umschriebenen Zeitraum ist aber anzunehmen, dass ein solches „Abdriften“ der visuellen Aufmerksamkeit sehr selten auftritt sowie ein enger Zusammenhang zwischen dem Blick und dem Inhalt der kognitiven Prozesse besteht. Die Ausrichtung des Blickes auf der Ebene der einzelnen Fixationen ist ein nicht kontinuierlich unter bewusster Kontrolle stehendes Verhalten, d.h. einzelne Blicksprünge (Sakkaden) können automatisch initiiert werden. Um die Intention einer Person aus ihrem Blick schließen zu können, müssen demnach sowohl Umgebungs- als auch Aufgabenbedingungen berücksichtigt werden.

### **2.2.2.3 Einsatzgebiete der Blickinteraktion**

Ein Anwendungsfall der Blickinteraktion im Rahmen multimodaler Mensch-Computer-Interaktion kann in einer „Vorauswahl“ liegen. Der Computer kann ggf. langwierige Operationen, beispielsweise das Laden zusätzlicher Programmbibliotheken, bereits starten sobald der Benutzer dem entsprechenden Menüeintrag seinen Blick zuwendet und bevor ein entsprechender Befehl auf herkömmliche Art initiiert wird. Ebenso bietet es sich an, den Cursor an dem Rande des augenblicklich fixierten Bereiches zu positionieren, um motorische Aktionen zu verkürzen. Zwei Varianten eines solchen Ansatzes wurden von Zhai, Morimoto und Ihde (1999) untersucht und von den Versuchspersonen positiv aufgenommen.

Die Interpretation des Blickes als Aufmerksamkeitsbereich kann auch zur verbesserten Darstellung der betrachteten Objekte genutzt werden. Denkbar sind hier automatische Zoom- oder *scroll*-Operationen (vgl. Goldberg & Schryver 1995). Diese sollten, wie alle automatischen Aktionen, durch den Benutzer einfach zu bestätigen bzw. abubrechen sein. Wird ein länger angeblicktes Objekt herangezoomt, sollte dies mittels Sprache, Tastatur o.ä. bestätigt oder abgebrochen werden können. Die Interpretation des Blickes als Aufmerksamkeitsfokus wurde von O’Sullivan et al. (2000) zur Optimierung von Computergrafikprozessen eingesetzt. Dabei wurde die Auflösung und die Qualität des aufgebauten Bildes oder der Animation blickortabhängig verfeinert (siehe Abbildung 6). So sollen neue Techniken entwickelt werden, die das Detaillierungsniveau des Echtzeit-Rendering komplexer Grafiken entsprechend der Charakteristika der visuellen Wahrnehmung adaptieren (vgl. auch O’Sullivan & Dingliana 2001). Ein *head-mounted Display*, welches ebenfalls blickortabhängig hochaufgelöste Grafik darzustellen in der Lage ist, wurde von Yoshida, Rolland und Reif bereits 1995 vorgestellt und prototypisch realisiert.

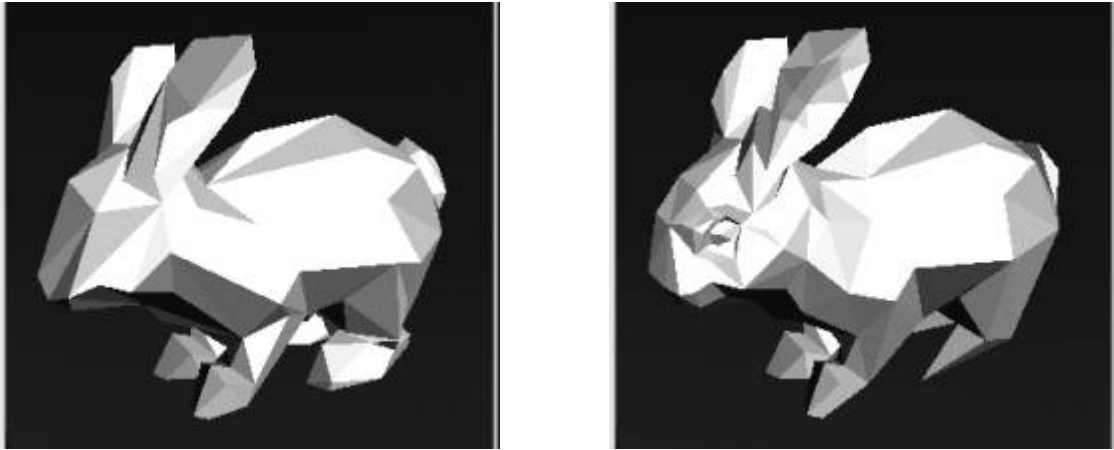


Abb. 6: *Blickrichtungsabhängiges Objekt-Rendering zur Optimierung von Berechnungszeiten für die Computergrafik; links: 358 Polygone gleichmäßig über das gesamte Objekt verteilt, rechts: 358 Polygone detaillierter am Kopf und niedrigauflösend an den Hinterbeinen verteilt (O'Sullivan et al. 2000).*

Salvucci und Anderson (2000) erprobten ein System, in dem neben der Maus und der Tastatur die Blickrichtung für verschiedene Aktionen in der grafischen Benutzungsschnittstelle eingesetzt werden konnte. Durch einen zusätzlichen Algorithmus, der auf der Basis von typischen Interaktionsverläufen im Interface Wahrscheinlichkeiten für die Auswahl einer spezifischen Schaltfläche abhängig von vorangegangenen Auswahlen lieferte, konnte die blickbasierte Auswahl zuverlässiger berechnet werden. Evaluationsuntersuchungen zeigten, dass die Auswahl eines Interface-Elementes per Blick effizient und akzeptiert war, wohingegen Bewegungsoperationen per Blick (*drag-and-drop*) den Benutzern schwierig und unkomfortabel erschienen.

Die Erfassung und Interpretation des Blickortes des Benutzers kann mit anderen Eingabemodalitäten nutzbringend kombiniert werden (Wachsmuth et al. 2001). Einerseits können so komplexere Interaktionsmuster erfasst und zum Aufbau eines Benutzermodells herangezogen werden. Ein solches Benutzermodell erlaubt die Vorhersage wahrscheinlicher Handlungsabfolgen, kann das System an die aktuellen Anforderungen adaptieren und ggf. spezifische Benutzerunterstützung bereitstellen. Andererseits disambiguieren kombinierte Eingabemodalitäten mehrdeutige Eingaben einer singulären Modalität. Als Beispiel ist folgende Spracheingabe denkbar: „Drehe *das* da!“, die durch die Berücksichtigung des Blickortes auf ein spezifisches Objekt interpretiert werden kann.

Der Einsatz der Blickinteraktion für die Steuerung von Computern durch behinderte Personen weicht von den hier vorgestellten Anwendungen ab. Die Frage nach der Natürlichkeit und Effizienz der Interaktionsgestaltung tritt hier zurück hinter dem Abwägen der verbleibenden Alternativen zur Steuerung hilfreicher Anwendungen für die spezifische Benutzergruppe.

### 2.2.3 Gestikinteraktion

Unter dem Begriff der Gestenerkennung oder Gestikinteraktion werden eine Vielzahl unterschiedlicher Interaktionsformen subsummiert, die sowohl auf der Ebene technischer Realisierungen als auch auf der Ebene menschlichen Verhaltens differenziert werden können. So fallen Bildverarbeitungsverfahren ebenso wie Geräte, die Gesten erkennen, um eine Krafrückmeldung zu generieren, oder berührungssensitive Displays für die stiftbasierte Eingabe oder Datenhandschuhe unter die technischen Ansätze zur Gestikinteraktion. Bei den analysierten und interpretierten Verhaltensäußerungen handelt es sich beispielsweise um freie Handgesten, um Kopfbewegungen, um die Handschrift oder die Handposition in Bezug auf ein virtuelles Objekt, welche u.a. für die Modellierung einer Krafrückmeldung wichtig ist.

#### 2.2.3.1 Technische Aspekte der Gestikinteraktion

Gestikinteraktion im Rahmen der Mensch-Computer-Interaktion geht über das übliche, auf die natürliche Umgebung bezogene Verständnis von Gestik hinaus. Hand- und Armbewegungen, die in der physikalischen Welt zur direkten Manipulation von Gegenständen oder zu deren taktiler bzw. haptischer Exploration verwendet werden, würden eher als Bewegungen zur Ausführung einer Handlung und nicht als symboltragende Gesten aufgefasst. In der Interaktion mit virtuellen Objekten jedoch, die typisch für die Mensch-Rechner-Interaktion ist, stellen die Handbewegungen oder über Eingabegeräte vermittelte Bewegungen nur Verweise auf die Manipulationsart dar und sind insofern Gesten. Diese müssen erst durch das Computersystem interpretiert werden und können erst dann in eine Veränderung an den Artefakten oder, für den Fall der Exploration, in eine Krafrückmeldung überführt werden.

Für die gestische Interaktion ist es erforderlich, zwei generelle Probleme zu lösen:

- die zuverlässige Erkennung einer Geste trotz erheblicher inter- und intraindividuel-  
ler Varianz bei der Ausführung ein und derselben Geste und deren eindeutige Zu-  
ordnung zu einer Bedeutung sowie
- die zuverlässige Segmentierung des Anfangs und des Endes einer Geste aus einer  
Folge von gestischen Äußerungen.

Zur Behandlung dieser Probleme werden verschiedene Ansätze gewählt. So kann die Anzahl von zu erkennenden Gesten kontextabhängig eingeschränkt werden oder es werden spezifische, determinierte Gesten von den Benutzerinnen gelernt, z.B. die Zeichensprache für Gehörlose oder spezielle Schriftzeichen bei der Handschriftenerkennung. Die Segmentierung von Gesten kann durch den Einbezug typischer dynamischer Muster unterstützt werden (Pavlović, Sharma & Huang 1997). Für die Zuordnung von Gesten zu gespeicherten Modellen werden vorrangig Klassifikationsmethoden eingesetzt, die auch bei der Spracherkennung Verwendung finden: Hidden Markov Modelle, künstliche neuronale Netze und Dynamische Programmierung (*dynamic time warping*).

Die eingeführten Interaktionstechniken mit Maus und Tastatur zählen ebenfalls zu den gestisch vermittelten Eingabeverfahren. Die eingeführten Eingabegeräte sollen hier jedoch nicht erörtert werden. Eine Darstellung über die Entwicklung und Funktionsweise der Maus ist bei Soojung-Kim Pang (2002) zu finden. Mit der Entwicklung von 3D-Display-Techniken wird es wichtig, die Funktionalität der Maus für eine adäquate Interaktion im dreidimensionalen Raum zu erweitern. Zhai (1998) gibt einen Überblick über verschiedene Eingabegeräte mit sechs Freiheitsgraden (drei für Translationen in die Koordinatenrichtungen x, y, z und drei für die entsprechenden Rotationen). Diese Eingabegeräte besitzen unterschiedliche Funktionalität. Isometrische Geräte, etwa der Joystick, übersetzen die Eingabevariable (Kraft oder Weg) ausgehend von einer Nullposition in eine entsprechende Cursorbewegung, d.h. sie unterstützen die Bewegungsübersetzung einer Geste bezogen auf ein grafisches Element im *graphical user interface* (GUI). Bei so genannten isotonischen Eingabegeräten, beispielsweise die Maus, die keine Nullposition haben, wird die Eingabevariable Positionsverschiebung direkt auf einen Ort in der Benutzungsschnittstelle skaliert. Diese Eingabegeräte unterstützen die Positionskontrolle, z.B. beim Zeigen auf virtuelle Objekte.

Stiftbasierte Eingaben auf einem druckempfindlichen Display werden für das Zeigen bzw. Auswählen von angezeigten Informationen genutzt und die handschriftliche Eingabe alphanumerischer oder grafischer Informationen. Solche Eingabeformen werden bereits in handelsüblichen mobilen elektronischen Planern eingesetzt. Einen Überblick über Funktionsprinzipien und Gestaltungsanforderungen für stiftbasierte und Touchscreen-Eingabeformen ist bei Greenstein (1997) zu finden. Für die Handschriftenerkennung werden zur Verbesserung der Erkennungsleistung spezifische Zeichenkonventionen genutzt. Andere Ansätze bestehen darin, aufwendigere Berechnungen vorzunehmen, die neben dem erzeugten Muster selbst auch die dynamischen Aspekte der Zeichengenerierung mit analysieren und zur Interpretation heranziehen (z.B. Waibel et al. 1995).

Für die Erfassung von Gesten und für die Krafrückmeldung stehen unterschiedliche Erfassungsmethoden bzw. Eingabegeräte zur Verfügung. So können für die reine Gestenerfassung, ohne Krafrückmeldung, Kameras und automatische Bildverarbeitungsverfahren, aber ebenso Datenhandschuhe und ähnliche am Körper zu tragende Sensoren eingesetzt werden. Ein Beispiel für die Gestenerkennung mittels automatischer Bildverarbeitung ist in Abbildung 7 dargestellt.

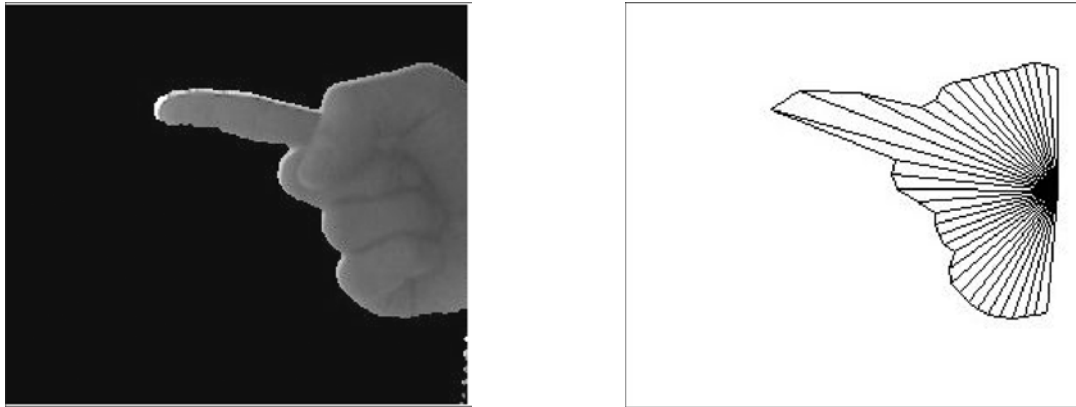


Abb. 7: Segmentierte Hand (links) und ihre Handvektoren (rechts) zur Erkennung der Handgesten ©HHI

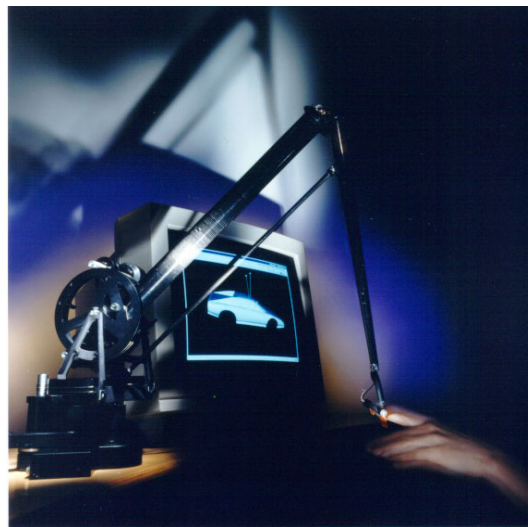
Während die Gestenerkennung mit automatischer Bildverarbeitung den Vorteil der Kontaktfreiheit hat, sind jedoch für eine robuste Erfassung verschiedene Probleme zu lösen (vgl. Mulder 1996), beispielsweise:

- die sichere Erkennung der Hand oder anderer einbezogener Körperteile unter wechselnden Beleuchtungsverhältnissen,
- die Entdeckung von Bewegungen der verfolgten Körperteile,
- die Positionierung der Kameras, um Verdeckungen auszuschließen und die Position einzelner Finger detailliert zu erfassen,
- die Abdeckung eines angemessenen Interaktionsraumes.

Datenhandschuhe oder ähnliche Gestenerkennungssysteme arbeiten mit unterschiedlichen Verfahren, um die Beugung und Bewegung der Hand und der Finger zu detektieren. So werden beispielsweise Lichtleiter benutzt, die netzartig an den Fingern entlang laufen, und je nach Krümmungswinkel der Finger unterschiedliche Lichtintensitäten durchleiten. Weitere Verfahren basieren auf der Veränderung des elektrischen Widerstandes von Materialien bei deren Stauchung oder Streckung. Zur Positionserkennung von Gliedmaßen oder anderen Körperteilen sowie der Bestimmung ihrer Lage im Raum werden auch elektromagnetische und ultraschallbasierte Verfahren eingesetzt. Eine umfangreiche Linksammlung zu diesen *Virtual-Reality*-Eingabetechnologien ist auf der DLR-Internetseite zu finden (DLR 2002).

Nachteilig ist an diesen Eingabesystemen, dass sie nicht kontaktfrei arbeiten. Dies beeinträchtigt nicht allein den Komfort der Benutzung, sondern schränkt auch die Effizienz der Nutzung dadurch ein, dass es aufwendiger ist, diese Eingabengeräte temporär in den Interaktionsverlauf zu integrieren.

Andererseits erlauben es diese Eingabegeräte, die Interaktion durch die Rückmeldung von Kräften bzw. Druck bei der Objektmanipulation anzureichen. Dafür wurden jedoch noch weitere Eingabegeräte entwickelt. Das PHANTOM (Massie & Sallisbury 1994) mit seinem an einem frei beweglichen Arm angebrachten Fingerhut, über den Kräfte an die Fingerspitze vermittelt werden, hat hohen Bekanntheitsgrad erreicht (siehe Abbildung 8). Dieses System kann zwar unaufwendig befristet in den Interaktionsverlauf integriert werden, verändert jedoch die natürliche Art, Objekte taktil zu explorieren, da nur eine Fingerspitze taktil stimuliert wird (Jansson 1999).



*Abb. 8: Das Phantom zur haptischen Interaktion. Der Finger, der die Krafterückmeldung empfindet, steckt in einem beweglichen Fingerhut am Ende des Auslegers, der in den drei Achsen bewegt werden kann (SensAble Technologies 2001).*

Die Krafterückmeldung verhilft dem Benutzer, die Berührung eines virtuellen Objektes zu spüren oder beim Manipulieren eines Objektes auf physikalische Eigenschaften wie das Gewicht, die Steifigkeit oder die Geschwindigkeit rückzuschließen. Da für eine kontinuierliche haptische Wahrnehmung eine Auffrischungsrate von ca. 1KHz (Bordegoni et al. 2000) notwendig ist, müssen Kräfte sehr schnell zurückgemeldet werden, um realitätsnah wahrgenommen zu werden. Floyd (1999) schlägt vor, in Teleoperationssystemen Objektinformationen in das System einzuspeisen, welches für die Modellierung der Krafterückmeldung zuständig ist, um Systemantworten bereits vorzubereiten. Potts (2000) verweist auf die Probleme, die Berührung eines virtuellen Objektes zu erkennen und eine adäquate Krafterückmeldung zu modellieren und stellt einen Lösungsalgorithmus für das PHANTOM vor. Das PHANTOM arbeitet mit nur drei Freiheitsgraden in einem Bereich von der Größe eines Mauspads. Dadurch ist es aber möglich, präziser als bei Datenhandschuhen Kollisionen zu detektieren und die Krafterückmeldung zu generieren.

Gegenwärtig besteht für gleichermaßen robuste wie ergonomische gestische und haptische Eingabetechniken noch weiterer Forschungs- und Entwicklungsbedarf. So stellten Sowa und



Wachsmuth (2000) ein auf empirischen Untersuchungen aufbauendes Modell für die Erkennung von Gesten bei der Manipulation virtueller Objekte vor. Ein Beispiel, wie durch den Einsatz neuer Materialien mit spezifischen Eigenschaften innovative Interaktionstechniken erprobt werden können, bietet die Entwicklung eines drucksensitiven Eingabegerätes für die 3D-Manipulation virtueller Objekte aus polymeren Materialien, das gegenwärtig vom Fraunhofer-Institut für Rechnerarchitektur und Softwaretechnik FIRST Berlin (Becker, Hausen & Rüsseler 2001) entwickelt wird.

### 2.2.3.2 Psychologische Aspekte der Gestikinteraktion

Gestik umfasst im weitesten Sinne alle Bewegungen des Körpers, die eine Bedeutung ausdrücken sollen. Gesten haben eine typische Dynamik (McNeill & Levy 1982), deren Analyse zur Segmentierung bei der automatischen Gestenerkennung benutzt wird. Gesten haben eine Vorbereitungsphase, eine Ausführungsphase und eine Rückführungsphase, in der die Körper- oder spezifischer die Hand-Arm-Haltung zu einer Ruheposition zurückkehrt.

In der Mensch-Computer-Interaktion werden verschiedene Taxonomien verwendet, um Gesten, die durch Bewegungen von Arm und Hand ausgeführt werden, funktional zu klassifizieren. Ein Überblick über verschiedene Klassifizierungsansätze für Gesten wird bei Mulder (1996) gegeben. Pavlović, Sharma & Huang (1997) schlagen die in Abbildung 9 veranschaulichte Gliederung vor.

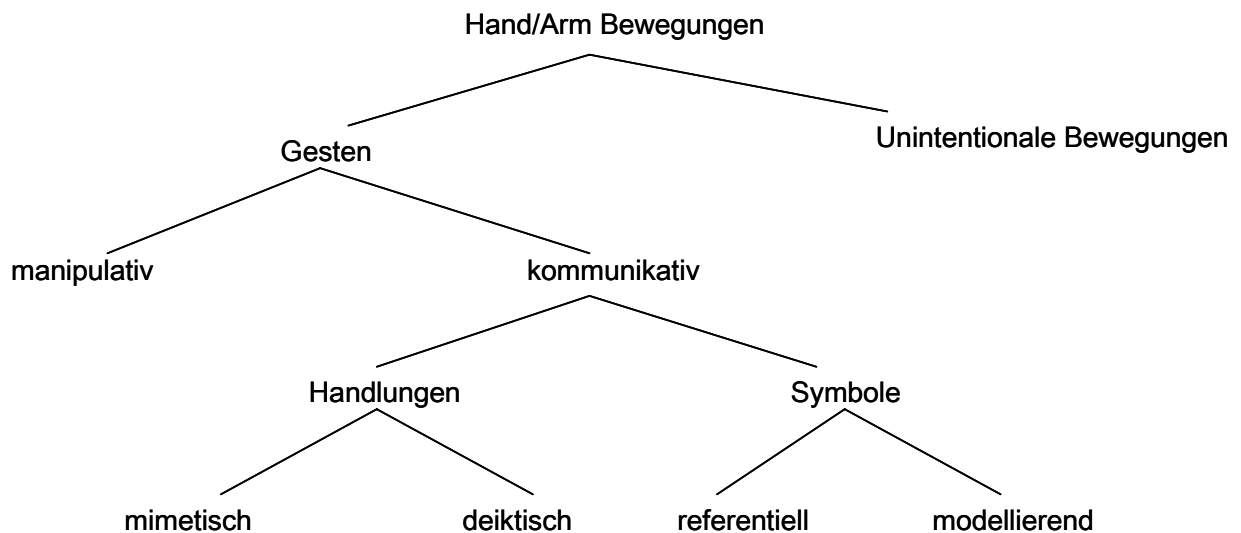


Abb. 9: Gestentaxonomie für die Mensch-Computer-Interaktion (nach Pavlović, Sharma & Huang 1997)

Die Unterscheidung zwischen manipulativen und kommunikativen Gesten verweist darauf, dass Gesten für eine Handlung an einem Objekt benutzt werden können oder kommunikativen Zwecken dienen; in natürlichen Situationen üblicherweise in Verbindung mit Sprache. Symbolische Gesten, die eine sprachliche Funktion besitzen, beinhalten eine referenzierende Handlung. Beispielsweise wird ein Rad durch die kreisende Bewegung des Zeigefingers sym-

bolisiert. Sie können aber in ihrer modellierenden Funktion auch sprachliche Äußerungen erweitern oder spezifizieren. Bewegungen, die direkt interpretiert werden können, werden als gestische Handlungen bezeichnet. Mimetische Gesten imitieren eine Handlung, während deiktische Gesten auf etwas zeigen.

Die Taxonomie verweist darauf, dass nicht alle Hand- und Armbewegungen als Gesten interpretiert werden können. Gesten werden von den Autoren per se als intendierte Bewegungen definiert, was jedoch zumindest für kommunikative Gesten kontrovers diskutiert wird. Die offene Frage, ob besonders koverbale Gesten intendiert erzeugt werden (z.B. McNeill 1992) oder eher spontan und unwillkürlich auftreten bei der Sprachproduktion (z.B. Kopp & Wachsmuth 2000, Krauss & Hadar 1999), sollte bei der Gestaltung gestischer Mensch-Computer-Interaktion berücksichtigt werden. Die Interpretation gestischer Informationen sollte durch weitere Informationen aus anderen Modalitäten abgesichert werden, um falschen Systemreaktionen entgegenzuwirken. Steininger, Schiel und Louka (2001) zeigten in einer Studie zu multimodaler Mensch-Computer-Interaktion, dass in einer reibungslosen Interaktion vorrangig mimetische und deiktische Gesten auftreten, während diffuse Hand- und Armbewegungen häufig eine Unterbrechung der laufenden Interaktion anzeigen. Zuletzt genannte Bewegungen können u.U. in multimodalen Systemen genutzt werden, um die Qualität des Interaktionsverlaufs zu prüfen und das System auf neue oder korrigierte Eingaben vorzubereiten.

Gestische Informationen in die Mensch-Computer-Interaktion einzubeziehen erscheint plausibel, um das Ziel einer natürlichen und intuitiven Interaktion zu unterstützen. Einschränkend ist jedoch anzumerken, dass gestische Interaktion auch in der natürlichen Kommunikation mit Interpretationsschwierigkeiten verbunden ist. Diese bestehen zum Beispiel darin, dass gestischer Ausdruck kulturell und sozial determiniert ist, dass es, mit Ausnahme der Zeichensprachen für Gehörlose, für nur wenige der Gesten klare Interpretationsregeln gibt und dass stets neue Gesten kreiert oder bekannte in abgewandelter Form erzeugt werden können. Des Weiteren ist für einen spezifischen Anwendungsbereich multimodaler Interaktion zu prüfen, welche Strategien von Benutzern situationsangemessen entwickelt und als natürlich empfunden werden. Smith und andere (2000) fanden beispielsweise, dass Benutzer andere Strategien der Hand-Auge-Koordination für das Zeigen auf dem Bildschirm wählten, wenn sie Eingabegeräte nutzten, als beim Zeigen mit der Hand auf physikalische Ziele. Steininger (2000) verwendete einen empirischen Ansatz, um ein Beschreibungssystem für Gesten und andere multimodale Eingaben sowie eine entsprechende Datensammlung in einem speziellen Anwendungsszenario zu generieren.

Eine weitere menschliche Sinnesmodalität, die eine bedeutsame Rolle in der multimodalen Mensch-Computer-Interaktion spielt, ist die haptische Wahrnehmung. Die haptische Wahrnehmung wird auch als Tastsinn bezeichnet und ist nicht durch ein spezifisches Sinnesorgan, sondern durch über den Körper verteilte Rezeptoren für verschiedene physikalische Qualitäten repräsentiert. Sie konstituiert sich aus taktilen und kinästhetischen Sinneseindrücken. Während Druck, Vibration, Temperatur und Schmerz taktil empfunden werden, liefert die

kinästhetische Komponente Informationen über die Position und Bewegung von Körperteilen sowie Informationen der auf sie einwirkenden Kräfte (vgl. Burdea 1996). Für die taktile Wahrnehmung von Kanten, statischen Kräften, Geschwindigkeit und Vibration sind verschiedene sogenannte Mechanorezeptoren in der Haut verantwortlich, die für Reize unterschiedlicher Frequenzen sensitiv sind. Für die kinästhetische Wahrnehmung sorgen sowohl Rezeptoren in Gelenken, Muskeln und Sehnen als auch das Vestibularorgan, der Gleichgewichtssinn. Haptische Wahrnehmung ist mit der aktiven, tastenden Exploration der Umgebung verbunden und ist daher die einzige Sinnesmodalität, bei der im Wahrnehmungsprozess Energie auch vom Körper an die Umwelt vermittelt wird. Menschliche Erfahrungsbildung ist von den ersten Lebenstagen an wesentlich mit diesem Sinn verbunden, da er einen direkten Bezug zwischen eigenen Aktivitäten und der Umwelt ermöglicht. Definitorische Klassifikationen für diese Sinnesmodalität stimmen in der Literatur nicht immer überein, was der Komplexität dieser Sinnesmodalität und dem Stand der Forschung geschuldet sein mag. Für die Mensch-Maschine-Interaktion bietet jedoch die Feinheit haptischer Wahrnehmung und die Geschwindigkeit der Informationsverarbeitung dieser Modalität, die teilweise reflexhafte, sehr schnelle Reaktionen hervorruft, interessante Anwendungsaspekte.

### **2.2.3.3 Einsatzgebiete der Gestikinteraktion**

Aufgrund der unter 2.2.3.1 bereits angedeuteten Vielfältigkeit von Eingabetechnologien und den damit übertragenen Informationen, die unter dem Begriff der Gestikinteraktion subsumiert werden, können nachfolgend nur einige Anwendungsbeispiele aufgeführt werden, die gegenwärtig im Zentrum der Forschung stehen.

In den derzeit vom BMBF geförderten Leitprojekten zur Mensch-Technik-Interaktion wird die gestische Interaktion in unterschiedlichen Anwendungsszenarien und unter Verwendung verschiedener technologischer Ansätze erforscht und entwickelt. So werden neben der gestischen Interaktion mit Haushaltsgeräten (EMBASSI 2001), *e-commerce*-Anwendungen (INVITE 2001) oder mit mobilen Robotern (morpha 2001) auch die stiftbasierte Interaktion mit mobilen Geräten und neuen Arbeitsplatzcomputern (MAP 2001, SmartKom 2001) in die multimodalen Benutzungsschnittstellen integriert. Stiftbasierte gestische Interaktion, bei der zum Teil auch kursive Schriftzeichen erkannt werden, eignet sich besonders gut für die Eingabe in mobile Endgeräte mit berührungssensitiven Bereichen im Display, z.B. der PVA mit Stift- und Spracheingabe (Bers, Miller & Makhoul 1998). Solche Computer können ähnlich wie der althergebrachte Notizblock mit Stift benutzt werden, z.B. der Palm (Palm Inc. 2002).

Gestikinteraktion wird zumeist in Kombination mit Sprachinteraktion eingesetzt. Gesten können in diesem Rahmen genutzt werden, um auf Objekte oder Orte zu referenzieren, wie dies bereits in Bolts „*Put-that-there*“-System (Bolt 1980) realisiert wurde. Bimler (1999) gibt ein Beispiel für die Manipulation von Objekten auf einem horizontalen Display, einem „virtuellen Tisch“ mittels Gesten. Besonders geeignet ist gestische Interaktion für solche Anwendungen, wo keine sehr präzise orts- oder formbezogene Veränderung an virtuellen Objekten erforderlich ist, sondern eine unaufwendige Positions- oder Formänderung im Vordergrund

steht, z.B. in kreativen Prozessen. Jung und andere (2000) setzten Gestik in Kombination mit Sprache zum virtuellen Konstruieren ein. Dabei wurde ein Set aus deiktischen, mimetischen und referentielle Gesten zur Eigenschaftsbeschreibung von Objekten verwendet. Alle Gesten, die mittels Datenhandschuh erfasst wurden, waren von sprachlichen Äußerungen begleitet. Die Interpretation der multimodalen Eingaben wurde durch eine domänenspezifische Wissensbasis unteretzt, um uneindeutige Benutereingaben und Ungenauigkeiten der VR-Eingabegeräte auszugleichen. Das System ZYKLOP (Kohler 2001) bietet ebenfalls eine gestische Interaktion mit einer einfachen CAD-Anwendung. Die Gestenerkennung basiert auf der automatischen Bilderkennung, die sowohl statische als auch dynamische Merkmale auswertet. Ein solches System erlaubt dem Benutzer ohne Aufwand zwischen freier gestischer Interaktion und anderen Eingabegeräten, wie Maus und Tastatur, zu wechseln. Damit ist dem Benutzer freigestellt, für jeden Bearbeitungsschritt die angemessenen Eingabemöglichkeit je nach erforderlicher Präzision oder Eignung auszuwählen. Andere Anwendungsbereiche werden für dieses System für die Steuerung von Präsentationen oder für die Auswahl von Monitoren in einer Multi-Display-Umgebung, zum Beispiel in Leitständen, sowie unter sterilen Bedingungen im Operationssaal für einfache Steuerungsvorgänge von Geräten, etwa der Lichtquelle oder für die Steuerung von Haushaltsgeräten für Behinderte gesehen. Keates und Robinson (1999) kamen in einer Studie mit motorisch behinderten Menschen zu dem Ergebnis, dass Gesteninteraktion auch zu erhöhten Anforderungen an die Benutzer führen kann, wenn komplexere Gesten, etwa simultane Kopf- und Handbewegungen, erforderlich sind, um beispielsweise die Erkennungszuverlässigkeit des Systems zu erhöhen. Dies verdeutlicht, dass eine benutzergruppenspezifische Evaluation solcher plausibel erscheinender Unterstützungssysteme wichtig ist, um Gestaltungsmängel, die die Akzeptanz des Systems gefährden, frühzeitig beheben zu können.

Eine weitere Anwendungsdomäne für die Gesteninteraktion wird in der visuell vermittelten Kommunikation gesehen, die sich auf rechnerunterstützte Kommunikation zwischen Menschen beziehen kann, etwa in Telekonferenzen (z.B. Sherrah et al. 2000), oder auf die Interaktion mit dem Computer selbst. So stellten Cutler und Turk (1998) ein videobasiertes Verfahren zu Erkennung von Gesten vor, das kleinen Kindern ermöglicht, mit dem Computer gestisch zu interagieren und verschiedene Spiele zu spielen. Die eingehenden gestischen Informationen, die in so genannten *motion blobs* segmentiert werden, werden über deren dynamische Charakteristik regelbasiert erkannt.

Auch bei der Steuerung von Komfortfunktionen im Kraftfahrzeug wird die gestische Interaktion als eine Möglichkeit gesehen, ohne Blickabwendungen von der Fahrbahn Einstellungen vorzunehmen. Eine prototypische Anwendung unter Einsatz der automatischen Bilderkennung für statische Gesten wurde dafür von Akyol und anderen (2000) vorgestellt. Die Autoren verweisen darauf, dass aufgrund zu geringer Erkennungszuverlässigkeit der Gestenerkennung eine Anwendung für sicherheitskritische Funktionen unmöglich ist.

Die gestische Steuerung mobiler Roboter auf der Basis automatischer Bildverarbeitung bietet ein weiteres Einsatzfeld, vorwiegend um generelle Richtungsanweisungen zu geben oder Be-

wegungen zu initiieren und zu stoppen. Ohne zusätzliche Eingabegeräte können Benutzer mit einem solchen System über Distanzen interagieren (vgl. z.B. Boehme et al. 1998, Ehrenmann, Lütticke & Dillmann 2001). Im Leitprojekt des BMBF zur Mensch-Technik-Interaktion „morpha“ wird über die Bewegungssteuerung des mobilen Roboters hinaus die gestische Interaktion genutzt, um dem Roboter neue Bewegungsabläufe durch Nachahmung beizubringen (morpha 2001). Bei komplexen und präzisen Steuerungserfordernissen von Robotern in Teleoperationsanwendungen werden zumeist spezifische Eingabegeräte verwendet (vgl. DLR 2002), die zudem auch Informationen für die haptische Wahrnehmung bereitstellen sollten. Ein Sonderforschungsbereich der DFG „Wirklichkeitsnahe Telepräsenz und Teleaktion“ an der TU München (TU München 2002) beschäftigt sich mit der Entwicklung dieser anspruchsvollen Technologie für Anwendungsbereiche wie der minimalinvasiven Chirurgie, der Produktionstechnik und der Mikromontage. Die Unterstützung haptischer Wahrnehmung ist auch im Bereich der Trainingssimulationen sehr wichtig, wobei hier diese Form multimodaler Interaktion besonders dafür notwendig ist, einen realitätsnahen Eindruck beim Trainierten zu erzeugen. Diese aufwendigen Trainingssimulationen werden vorwiegend eingesetzt, um die notwendigen Fertigkeiten für die Benutzung und Steuerung komplizierter, sicherheitskritischer Technik zu erlernen, z.B. Trainingssysteme im Bereich der endoskopischen Chirurgie (Çakmak & Kühnapfel 2000) oder Flugsimulatoren für die Aus- und Weiterbildung von Piloten (z.B. Caro 1988). Haptische bzw. taktile Interaktion kann auch für Sehbehinderte in unterschiedlichen Anwendungszusammenhängen den Zugang zu technischen Systemen erleichtern und räumliche Strukturen leichter erfahrbar machen, als es über eine verbale Beschreibung möglich ist (vgl. z.B. Kurze 1999).

## **2.2.4 Weitere Interaktionsformen**

An dieser Stelle soll auf einige weitere Interaktionsformen verwiesen werden, auf die aber im empirischen Teil dieser Arbeit nicht weiter eingegangen wird. Der gegenwärtige Trend in der Mensch-Computer-Interaktion, Emotionen zu erkennen und die Interaktionszuverlässigkeit zu erhöhen, sorgt jedoch für ein zunehmendes Forschungsinteresse an den hier vorgestellten Modalitäten. Nicht betrachtet werden bislang sehr spärlich untersuchte Interaktionsformen, wie gustatorische oder olfaktorische Benutzungsschnittstellen. Da sich die multimodale Mensch-Computer-Interaktion darauf konzentriert, natürliche Interaktionsformen zu integrieren, werden hier auch solche Interaktionsansätze ausgenommen, die nur technisch vermittelt vorkommen, wie die auf der Ableitung von Gehirnströmen beruhende Interaktion.

Das automatische Lippenlesen, die Erkennung typischer Mundbewegungen beim Sprechen verschiedener Laute oder Worte, ist eine weitere Modalität, die mittels automatischen Bildverarbeitung und Mustererkennung in die Mensch-Computer-Interaktion einbezogen werden kann. Das Ziel dieser Forschung besteht in erster Linie darin, die gesprochene Sprache durch die zusätzliche optische Analyse der dabei auftretenden Mundmotorik auch bei Umgebungslärm und anderen Störvariablen zuverlässiger zu erkennen (vgl. z.B. Goldschen, Garcia & Petajan 1995, Luettin, Thacker & Beet 1996). Für diesen Ansatz spricht, dass die Lippenbe-

wegungen und die gleichzeitig produzierte Sprache in der natürlichen Kommunikation integriert wahrgenommen werden. Dies belegen auch experimentelle Untersuchungen über die Veränderung der akustischen Wahrnehmung, die durch asynchrone Lippenbewegungen erzeugt wird, dem so genannten McGurk-Effekt (MacDonald & McGurk 1978). Einen Überblick über Ansätze des automatischen Lippenlesens und über Systemrealisationen ist bei Goldschen (1997) sowie bei Chen und Rao (1998) zu finden.

Die Erkennung von Emotionen aus dem mimischen Ausdruck und aus physiologischen Parametern ist aktuell interessant für die Mensch-Computer-Interaktion. Dient der Gesichtsausdruck dazu, Emotionen zu erkennen, werden für die Analyse spontaner Mimik automatische Bildverarbeitungs- und Mustererkennungsverfahren eingesetzt. Ekman und Friesen (1978) klassifizierten emotionale mimische Ausdrücke nach den in einem Emotionsausdruck typischerweise aktiven Muskelgruppen. Die automatische Erkennung nutzt häufig diese detaillierte Taxonomie. Allerdings ist die automatische Analyse spontaner Mimik mit verschiedenen Problemen konfrontiert. Sie bestehen beispielsweise darin, dass die Kopfposition nicht immer frontal zur Kamera ausgerichtet ist, die zur Erkennung herangezogenen Vergleichsmuster jedoch üblicherweise Frontalansichten darstellen oder das Teile des Gesichts verdeckt sind durch Brillen etc. oder auch, dass Gesichtsausdrücke sehr schnell wechseln (Cohn et al 2001). Ekman (1993) verweist auf einige psychologische Probleme, die gleichfalls die automatische Emotionserkennung aus dem mimischen Ausdruck erschweren: die Intensität mimischen Ausdrucks unterliegt breiten inter- und intraindividuellen Schwankungen, soziale Normen beeinflussen den Ausdruck von Emotionen und Emotionen müssen sich nicht notwendigerweise im mimischen Ausdruck wiederfinden.

Für die Erkennung von Emotionen werden auch physiologische Werte analysiert, die an speziell entwickelten Eingabegeräten abgeleitet werden können, z.B. der „*Emotion Mouse*“ (Ark, Dryer & Lu 1999). Speziell Hautleitwiderstand, Blutvolumen, Elektrokardiogramm (EKG) und Elektromyogramm (EMG), Atmung und Temperatur werden als Indikatoren von Emotionen in der automatischen Emotionserkennung diskutiert (Picard 1997). Die Forschung befasst sich damit, Merkmale bzw. Merkmalskombinationen aus den physiologischen Maßen zu extrahieren, die die Emotionen hinlänglich differenzieren, um daraus Vergleichsparameter für die automatische Emotionserkennung zu generieren. Als Zielstellung dieser Entwicklungen wird argumentiert, dass eine an den emotionalen Zustand des Benutzers angepasste systemseitige Interaktion produktivitätssteigernd sei, wie dies in der Kooperation zwischen Menschen beobachtet wurde (Dryer & Horowitz 1997). Verschiedene Probleme lassen die zuverlässige Erkennung bzw. Diskriminierung von Emotionen über physiologische Maße fraglich erscheinen. Einerseits besitzen physiologische Maße keine ausgeprägte Spezifität, d.h. sie sind durch verschiedene kognitive und physiologische Prozesse beeinflussbar. Andererseits sind die auftretenden Parameterveränderungen, die eine Veränderung des emotionalen Zustandes anzeigen, häufig sehr klein und messtechnisch daher häufig nur im Labor zuverlässig zu erfassen.

Als Beispiel kann die elektrodermale Aktivität gelten, die im Mikrosiemens-Bereich analysiert wird und die durch geringe motorische Aktivitäten am Ableitungsort gestört wird. Des Weiteren können viele physiologische Maße nur invasiv, d.h. durch am Körper angebrachte Sensoren, zuverlässig abgeleitet werden. In der Mensch-Computer-Interaktion dürften sich stets emotionale Zustände mit kognitiven Anforderungen und physiologischen Schwankungen überlagern, so dass die Interpretation emotionaler Zustände allein aus physiologischen Parameter störanfällig sein dürfte. Inwiefern Benutzer die kontinuierliche Erfassung ihres emotionalen Zustandes durch den Computer speziell während der Arbeit akzeptieren werden, scheint eine weitere offene und stärker aus ethischer Perspektive zu reflektierende Frage zu sein.

Die Erkenntnisse aus der Forschung zu gestischem und mimischen Ausdruck von Menschen in der Kommunikation findet jedoch gegenwärtig auch breiten Eingang in die Animation von anthropomorph visualisierten Software-Agenten, so genannten Avataren. Diese sollen möglichst natürlich mit dem Benutzer kommunizieren und als menschenähnliche Dialogpartner erscheinen (vgl. z.B. Massaro 1998).

### **2.2.5 Integration mehrerer Modalitäten**

Eine der zentralen Herausforderungen multimodaler Mensch-Computer-Interaktion besteht darin, die über die einzelnen Modalitäten aufgenommenen Informationen vom Benutzer zu integrieren. Dabei werden zwei Typen multimodaler Architekturen unterschieden: solche, die die Eingangssignale auf dem Merkmalsniveau zusammenführen (*early fusion*) und solche, die die Informationen aus den Eingangssignalen auf der semantischen Ebene integrieren (*late fusion*) (vgl. auch Oviatt et al. 2000). Die verschiedenen Modalitäten, die in einer multimodalen Systemarchitektur verbunden werden können, weisen jedoch unterschiedliche zeitliche Charakteristika auf (vgl. Abbildung 10), die einerseits aus dem natürlichen Kommunikationsverlauf und andererseits auf die technischen Parameter der automatischen Erkennungssysteme zurückzuführen sind.

Für die technisch bedingten zeitlichen Verzögerungen bietet die weitere Optimierung von Erkennungsprozessen eine Lösung. Für den Umgang mit natürlicherweise entstehenden Zeitdifferenzen zwischen semantisch gekoppelten Informationen wird von Lenzmann (1998) und Wachsmuth (2000) eine zeit-ereignisabhängige Segmentierung der kontinuierlichen Benutzereingaben im Rahmen einer multimodalen Eingabeagentur vorgeschlagen. Gegenwärtig werden für die Analyse und Interpretation von multimodalen Benutzereingaben üblicherweise Agentensysteme benutzt, in denen Programme, so genannte Software-Agenten, in einer vernetzten Umgebung selbstständig Aufgaben ausführen und untereinander kommunizieren können (vgl. auch Brenner, Zarnekow & Wittig 1998). Diese Agenturen unterstützen die asynchrone Verarbeitung und Weiterleitung von Informationen, die Verteilung von Verarbeitungsprozessen und die gezielte Abfrage von und Suche nach Informationen bei anderen Agenten.

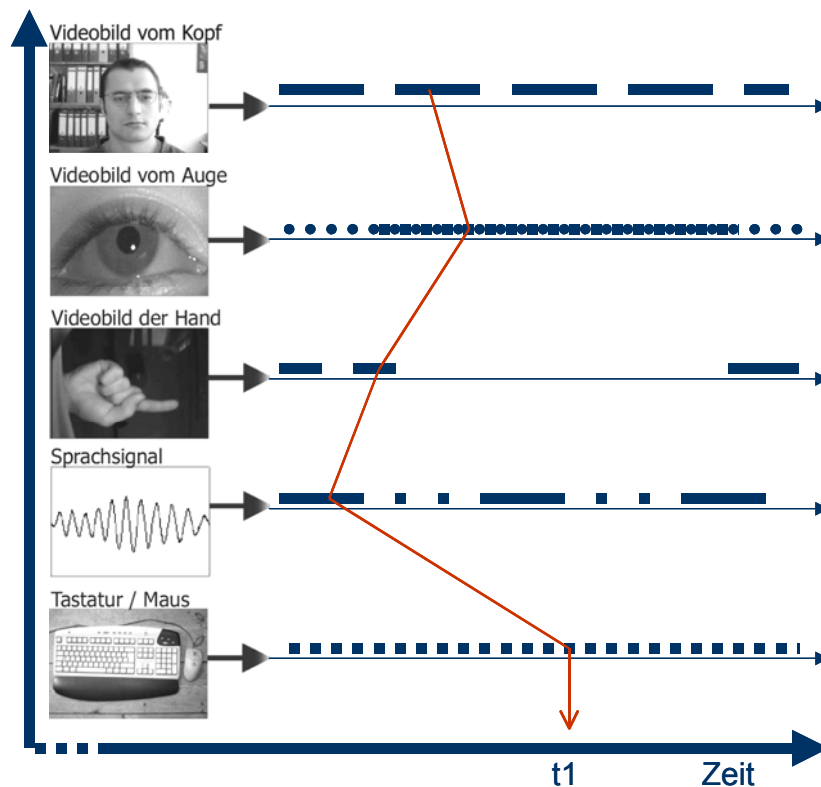


Abb. 10: Integration der Eingangssignale verschiedener Modalitäten mit unterschiedlichen Signalcharakteristiken

Ein weiterer Aspekt, der die Verarbeitung in multimodalen Systemen bestimmt, bezieht sich auf die gegenseitige Disambiguierung der Informationen aus den einzelnen Modalitäten. Die Bewertung der Informationen kann auf der empirisch ermittelten Erwartung über die Erkennungsrate für eine Modalität beruhen und gewichtend mit einbeziehen, wie kontrolliert eine Verhaltensäußerung in einer Modalität erzeugt wird. Weiterhin können auch redundante oder spezifizierende Informationen in anderen Eingabemodalitäten für die Interpretation der Benutzerintention herangezogen werden.

Lippenbewegungen und gesprochene Sprache bieten hochgradig redundante Informationen. Andere Verhaltensäußerungen weisen eine stärkere Komplementarität auf, beispielsweise Sprache und Gesten (Oviatt, DeAngeli & Kuhn 1997), und können zur gegenseitigen Spezifikation verwendet werden. Verfügbare äquivalente Modalitäten, die, je nach situativen Kontext, wahlweise zur Interaktion genutzt werden können, erhöhen nicht nur die Flexibilität, sondern machen das System auch weniger störanfällig gegen sinkende Erkennungsraten einer gleichwertigen Modalität. Als äquivalente Modalitäten können in funktionsbezogenen Dialogen beispielsweise gesprochene und geschriebene Sprache angesehen werden.



Oviatt und andere (2000) beschreiben folgende Systemanforderungen speziell für multimodale stift- und sprachbasierte Eingaben verarbeitende Architekturen, die jedoch auch für Systeme mit anderen multimodalen Eingabekombinationen gelten können:

- parallele Erkennungskomponenten, die eine Anzahl zeitlich kodierter bedeutungstragender Informationen für jede Modalität liefern,
- eine Systemkomponente, in der die bedeutungstragenden Informationen aus den einzelnen Modalitäten gemeinsam repräsentiert sind,
- ein Gruppierungsprozess, der zeitbezogen zusammengehörige Bedeutungsfragmente bestimmt,
- Prozeduren, die semantisch und temporär passende Bedeutungsfragmente zusammenfügen,
- eine datengetriebene statistische Absicherung der getroffenen Gesamtinterpretation der multimodalen Eingaben,
- eine flexible, asynchrone Architektur für multiple Verarbeitungsprozesse und für eine mit den Benutzereingaben schritthaltende Verarbeitung, u.U. auch für mehrere Benutzer,
- eine Benutzungsschnittstelle, die die sich ergänzenden Modalitäten synergetisch kombiniert.

Hinzuzufügen sind ergonomische Anforderungen an die Eingabegeräte und Messsysteme, die vom Benutzer generierte Signale aufnehmen. So sollten die Eingabegeräte leicht zu handhaben sein und Messsysteme für die Gestik- oder Blickpositionserfassung u.a. möglichst kontaktfrei und ohne Einschränkung des natürlichen Bewegungsfreiraumes des Benutzers funktionieren. Die Berücksichtigung der ergonomischen Systemanforderungen trägt ebenfalls zu dem Ziel bei, durch Multimodalität eine natürlichere Mensch-Computer-Interaktion zu schaffen.

## **3 Systementwicklung und Evaluation**

### **3.1 Modelle der Systementwicklung**

Für die Entwicklung von Mensch-Maschine-Systemen wurden in den letzten 40 Jahren unterschiedliche Prozess- bzw. Phasenmodelle erstellt, die dazu beitragen sollen, die mit der möglichen funktionalen Komplexität von Systemen einhergehende Problematik der Systemgestaltung zu reduzieren. Dabei werden die einzelnen Arbeitsschritte in der Systementwicklung mit den durch sie erzeugten Ergebnissen systematisch über den Zeitverlauf gegliedert. Die systematische Betrachtung erstreckt sich in den verschiedenen Modellen auf die Entwicklungsschritte bis zur Inbetriebnahme oder darüber hinaus auf den gesamten Lebenszyklus des Systems. Die eingeführten Prozessmodelle optimieren den Systementwicklungsprozess auf verschiedene Ziele hin, z.B. die Produktqualität. Im Folgenden werden die einflussreichsten Prozessmodelle kurz charakterisiert.

Das Wasserfallmodell (Benington 1956, Boehm 1981, Royce 1970) leitet zur schrittweisen Abarbeitung und Dokumentation der Entwicklungsschritte von der Anforderungsdefinition bis zur Wartung während des Betriebes an. Es ist mit geringem Managementaufwand verbunden und übersichtlich, da Rückkopplungen nur auf angrenzende Stufen begrenzt werden (Balzert 1998).

Im V-Modell (siehe Abbildung 11a) sind weitere bewertende Schritte im Systementwicklungszyklus eingefügt. Der Prozess der Verifikation testet eine Systemkomponente auf Übereinstimmung mit ihrer Spezifikation, wohingegen sich die Validation auf die Zweckerfüllung des Gesamtsystems bezieht. Das V-Modell sichert durch die Testschritte auf Subsystem und Komponentenebene die Qualität bei der Entwicklung komplexer Systeme. Allerdings werden dadurch höhere Anforderungen an die Verwaltung der verschiedenen Aktivitäten gestellt (Blanchard & Fabrycky 1998).

Das Spiralmodell (Boehm 1986) sieht iterative Schritte der Systementwicklung unter besonderer Beachtung der Risikominimierung vor (siehe Abbildung 11b). Innerhalb der einzelnen Schritte werden Prototypen generiert und getestet zur Abschätzung der mit der Entwicklung verbundenen Risiken. Das Modell zeichnet sich durch Flexibilität aus, da basierend auf lokalen, zyklenspezifischen Entscheidungen risikoarme Strategien für die Systementwicklung ausgewählt werden. Probleme bei diesem Prozessmodell bestehen in der umfassenden Analyse von risikobeeinflussenden Faktoren und in den hohen Anforderungen an das Projektmanagement.

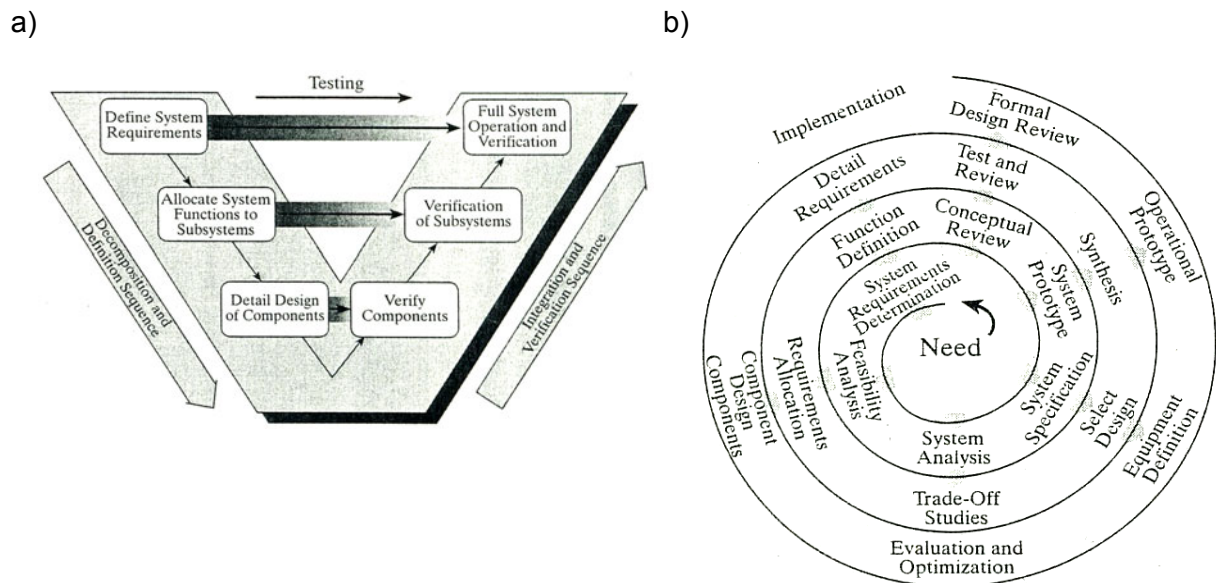


Abb. 11: Das V-Modell (links) und das Spiralmodell (rechts) für die in Phasen gegliederte Systementwicklung (Blanchard & Fabrycky 1998)

Das als *concurrent engineering*, *simultaneous engineering* oder *parallel engineering* bekannte Vorgehensmodell aus der Fertigungsindustrie wird sehr häufig im Zusammenhang mit der Verkürzung von Produktentwicklungszeiten diskutiert (Eiff 1991). Das auch als nebenläufiges Prozessmodell bezeichnete Vorgehen vereint alle an der Entwicklung beteiligten Abteilungen in einem Team, um möglichst viele Entscheidungsaspekte von Anfang an zu berücksichtigen und früher sequentiell durchgeführte Schritte zu parallelisieren. Neben dem hohen Personalaufwand ist zu berücksichtigen, dass sich spätere Revisionen in einem Segment negativ auf weitere parallelisierte Arbeitsfolgen auswirken.

Die vorgestellten Prozessmodelle fokussieren bei der Systemgestaltung die technischen Komponenten von Mensch-Maschine-Systemen. Die Leistungsfähigkeit des gesamten Mensch-Maschine-Systems wird jedoch nicht allein durch den Einsatz spezifikationsgerecht gestalteter Technik gesichert, sondern insbesondere dadurch, dass Mensch und Maschine effizient zusammenwirken können. Dafür ist es notwendig, im technischen Entwurf die Auswirkungen auf die Tätigkeit des Menschen mit seinen individuellen Leistungsvoraussetzungen und seiner organisationalen Einbindung zu berücksichtigen.

Das parallel-iterative Vorgehen bei der Gestaltung eines Mensch-Maschine-Systems versucht durch zyklische Prozesse der Spezifikation und Prüfung die menschlichen Ressourcen für die Leistung des gesamten Systems in hohem Maße nutzbar zu machen. In die parallel-iterative Systemgestaltung fließen auch Aspekte der eingeführten Modelle ein. Die systemtechnische Perspektive erweitert den eng auf das technische System bezogenen Blick um Aspekte der System-Organisation-Interaktion sowie der System-Umwelt-Interaktion (Timpe & Kolrep 2000). Für die Entwicklung von Mensch-Maschine-Systemen, deren Domäne und zukünftige

Benutzer bereits eindeutig definiert sind, kann dieses Phasenmodell erfolgreich für die Gestaltung angewandt werden (vgl. Müller, Giesa & Anders 1999).

Für die Entwicklung von informationstechnischen Systemen sind noch weitere Modelle bedeutsam.

Mit der Verbreitung objektorientierter Programmiersprachen gewann der Aspekt der Wiederverwendbarkeit von Software stärkere Bedeutung, z.B. in Form von Klassenbibliotheken (Gamma et al. 1993). Die Auswahl und die Archivierung wiederverwendbarer Komponenten für den eigenen Entwicklungszweck erfordern neue Aktivitäten im Entwicklungsprozess.

Auf der Optimierung des Entwicklungszeitraumes liegt das Hauptaugenmerk, wenn evolutionäre oder inkrementelle Modelle zur Softwareentwicklung eingesetzt werden (Floyd 1984). Bei evolutionären Modellen werden sowohl die Anforderungen als auch das vollständige Produkt, dessen Kernversion dann bereits in Betrieb genommen wurde, sukzessive entwickelt. Im Unterschied dazu geht das inkrementelle Modell von vollständig definierten Systemanforderungen aus, so dass spätere Entwicklungsschritte nicht zu einer kompletten Revision des Systems führen.

Aus dem Kontext des *usability engineering* stammt das Vorgehensmodell des *usability life cycle*, das besonders die Integration der Gebrauchstauglichkeitskriterien in die objektorientierte Softwareentwicklung bezweckt (Mayhew 1999). Hierbei werden drei Phasen der Systementwicklung mit iterativer Lösungsgenerierung abgegrenzt, die durch verschiedene Methoden der Informationsgewinnung und Bewertung flankiert werden. Die drei Phasen umfassen die Anforderungsanalyse, die Design-Test-Entwicklungsphase und die Installationsphase. In der Anforderungsanalyse werden die Ziele der Systementwicklung bzgl. der Gebrauchstauglichkeit auf der Basis des Benutzerprofils (Qualifikationsvoraussetzungen), der Aufgabenanalyse (Tätigkeits- bzw. Anwendungskontext), der Plattformvoraussetzungen (Definition technischer Merkmale für die Integration in bestehende Netzwerke) und allgemeiner Designprinzipien analysiert und beschrieben. In der daran anschließenden Phase wird die Tätigkeit nach den Möglichkeiten, die die Automatisierung bietet, reorganisiert und das konzeptuelle Modell der Systemfunktionalität abgeleitet. Mit Hilfe von *mockups* und Designprototypen werden die Designalternativen auf jeder Stufe der Modellierung mit Benutzerbeteiligung evaluiert. Erst bei erfolgreicher Testung kann zum nächsten Schritt im Entwicklungsprozess übergegangen werden. Auch in der letzten Phase, der Installation des Systems, wird noch einmal eine Benutzerrückmeldung erhoben, die zu weiteren Anpassungen führt, wenn nicht alle Auflagen erfüllt wurden. Für einfache Anwendungen sind explizit Abkürzungen im Entwicklungsprozess vorgesehen.

Das Prototypenmodell ermöglicht es in der Softwaretechnik, einerseits Risiken durch frühzeitige Erprobung von Teillösungen des Gesamtsystems zu minimieren und andererseits die denkbaren Lösungen interaktiv mit den Auftraggebern schrittweise über den Entwicklungsprozess zu diskutieren (Balzert 1998). Bei diesem Modell können verschiedene Arten von

Prototypen nach ihrer Nähe zum zukünftigen Gesamtsystem und nach ihrem Zweck unterschieden werden (vgl. Balzert 1998). Eine weitere Unterscheidung von Prototypen bezieht sich auf die Ebenen, die vom zukünftigen Softwareprodukt repräsentiert werden. Floyd (1984) unterscheidet horizontale und vertikale Prototypen (siehe Abbildung 12). Während erstere nur spezifische Ebenen der Software, allerdings annähernd vollständig, realisieren, z.B. die Benutzungsoberfläche ohne dahinterliegende Funktion, werden bei letzteren ausgewählte Teile des zu entwickelnden Produkts durch alle Ebenen der Software hindurch implementiert.

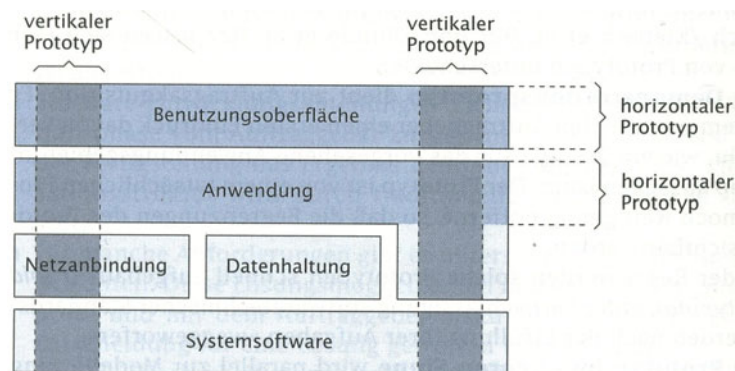


Abb. 12: Realisierungsebenen von vertikalen gegenüber horizontalen Prototypen (Balzert 1998)

Software-Prototypen, die das Zielprodukt in wesentlichen Dimensionen repräsentieren, werden genutzt, um bedeutsame Anforderungen und Vorgehensweisen der Entwicklung zu spezifizieren. Sie können in vielfältigen Ausprägungen entstehen, um über die günstigsten Varianten zu entscheiden und sie können zu Forschungszwecken benutzt werden. Das Prototypenmodell ist geeignet, in informationstechnischen Forschungsprojekten eingesetzt zu werden, beispielsweise um neue Mensch-Computer-Interaktionsformen zu untersuchen. Dabei stehen anfangs besonders Fragen der Realisierbarkeit von Konzepten im Vordergrund und die Prüfung des Systemverhaltens, weniger die für die Produktentwicklung wesentliche Anpassung an Kundenanforderungen, z.B. die Plattformkompatibilität oder den Funktionsumfang betreffend. Durch diese fokussierte Vorgehensweise können wiederholte *redesigns* kosten- und zeitgünstig durchgeführt werden, um unterschiedliche Lösungsansätze zu testen. Bereits in diesen frühen Phasen der Systementwicklung können basierend auf dem Prototypenmodell über die technischen Aspekte hinaus bereits Erfahrungen zur Nutzung des innovativen Systems gewonnen werden. In Verbindung mit psychologischen Methoden, bei denen bestimmte Systemeigenschaften simuliert werden (z.B. *Wizard-of-Oz*-Methode), können mit Prototypen Erkenntnisse darüber gewonnen werden, wie Benutzer mit einem System interagieren wollen, das gegenwärtig noch über den Stand der Technik hinausgeht (Cheyer & Julia 1999b, Dahlbäck, Jönsson & Ahrenberg 1993). Mit Hilfe von Evaluationsuntersuchungen, wie sie im *usability life cycle*-Ansatz vorgeschlagen werden, lassen sich die verschiedenen Systemeigen-

schaften nutzerbezogen optimieren, Auswahlentscheidungen geeigneter Systemkomponenten können unterstützt und avisierte Einsatzzwecke können erstmals erprobt werden.

## 3.2 Evaluationsmodelle

Als Evaluation wird sowohl der Prozess als auch das Ergebnis einer nach Nutzenkriterien vollzogenen Bewertung von Maßnahmen, Programmen oder Produkten bezeichnet. Stufflebeam (1972) unterscheidet nach dem Zweck der Evaluation:

- die Steuerungs- und Optimierungsfunktion über einen Prozess hinweg,
- die Bewertungs- und Beurteilungsfunktion am Ende einer durchlaufenen Maßnahme oder eines Prozesses,
- die Bewertung von Rahmenbedingungen und Zielen von Maßnahmen sowie
- die Bewertung des Ressourceneinsatzes für die Zielerreichung.

Die Evaluation mit Steuerungs- und Optimierungsfunktion wird auch als formative Evaluation und die mit Bewertungs- und Beurteilungsfunktion für ein Produkt eines Prozesses als summative Evaluation bezeichnet.

Bisher gibt es keine anerkannte Definition oder ein allgemeines Modell der Evaluation (Majonica 1996). Dies kann darauf zurückgeführt werden, dass Evaluationen im sozialen, administrativen und wirtschaftlichen Bereich mit unterschiedlichen Zielsetzungen und unter Einsatz verschiedenster methodischer Vorgehensweisen und Verfahren durchgeführt werden. Einige wesentliche Übereinstimmungen liegen nach Wottawa und Thierau (1990) darin, dass Evaluationen als Planungs- und Entscheidungshilfen und zur Bewertung von Handlungsalternativen dienen. Die systematisch durch die Evaluation gewonnenen Daten und Befunde werden regelgeleitet an festgesetzten Maßstäben bewertet und münden in einer Stellungnahme. Ziele von Evaluationen bestehen darin, Maßnahmen und Vorgehensweisen zu verbessern, zu legitimieren oder über sie zu entscheiden. Der Fokus der Evaluation ist auf Maßnahmen und Prozesse, nicht auf die individuelle Leistungsbeurteilung gerichtet.

Bewertende Studien (empirische Evaluationsstudien) wurden zuerst bei reformerischen Programmen Ende des 19. Jahrhunderts und Anfang des 20. Jahrhunderts bei den *Experimental Stations and Extension Programs* der Land Grant Colleges in den USA durchgeführt. In der Mitte des 20. Jahrhunderts wurden von der UNESCO beauftragte Evaluationen von Entwicklungshilfeprojekten durchgeführt. Die staatlichen Reformprogramme in den 60er und 70er Jahren des letzten Jahrhunderts in den USA und Deutschland, beispielsweise im Bildungs- und Gesundheitswesen sowie zur Humanisierung der Arbeitswelt, waren bedeutungsvolle Gegenstände der Evaluation und stellten neue methodische Herausforderungen an die empirische Evaluationsforschung (vgl. House 1993, Koch & Wittmann 1990).

Die Bewertung von Geschäfts- und Produktionsprozessen ist in vielen Wirtschaftszweigen unter dem Begriff des Qualitätsmanagements verbreitet. Ein in den letzten Jahren an Bedeutung gewinnender wirtschaftlicher Bereich für die Evaluation ist die Informations- und Kommunikationstechnik. Dieses Interesse an Evaluation wird mit der hohen strategischen Bedeutung von Software für viele Unternehmen begründet und den Wettbewerbsvorteilen, die sich die Unternehmen der Software-Industrie davon erhoffen. Besonders für die Softwareerstellung wurden verschiedene Evaluationsansätze entwickelt, die zur Standardisierung von Softwareentwicklungsprozessen führen sollen, z.B. durch die Methode SPICE (*Software Process Improvement and Capability dEtermination*) (Thaller 1998). Die Ansätze gehen übereinstimmend davon aus, dass systematische Entwicklungsprozesse zu verbesserten Produkten führen. Einschränkend sei jedoch darauf verwiesen, dass für Innovationsprojekte Technologie-Risiken und Know-how-Risiken eine bedeutsamere Rolle für den Projekterfolg spielen, als die Prozessrisiken (Rezagholi 2000).

Auch die Produkte der informationstechnischen Systementwicklung unterliegen der qualitätssichernden Bewertung, wobei interaktive Systeme, die eine Benutzungsschnittstelle für den Menschen haben, nicht allein technischen Funktionskriterien genügen müssen. Die Evaluation von interaktiven Computersystemen dient dazu, die Gebrauchstauglichkeit der Schnittstelle, üblicherweise unter den Kriterien der Effektivität, Effizienz und Akzeptanz (DIN ISO 9241-11) zu testen.

Werden in frühen Phasen der Systementwicklung im *parallel design* gleichzeitig mehrerer Alternativen für ein und dieselbe Teilaufgabe entworfen, besteht das Ziel der Evaluation darin, einen Entwurf auszuwählen, der dann weiter verfolgt und ausgearbeitet wird (Preim 1999). Die Bewertung des Produktes kann in unterschiedlichen Realisierungsgraden erfolgen. Bereits in der Konzeptphase der Entwicklung interaktiver Systeme können mittels Papierskizzen Benutzungsprobleme identifiziert werden (Virzi, Sokolov & Karis 1996).

Bei der Evaluation interaktiver Systeme sind drei Ansätze zu unterscheiden:

- die Evaluation auf der Basis formaler Benutzermodelle,
- die Expertenevaluation und
- die Evaluation mit Benutzern.

Die formalen Benutzermodelle beschreiben die Mensch-Rechner-Interaktion als Abfolge von Systemzuständen in Zusammenhang mit den Aktionen eines „gedachten“ Benutzers. Geiser (1990) führt verschiedene Methoden für die Interaktionsmodellierung auf, z.B. Zustandsübergangsnetze, Petri-Netze und kontextfreie Grammatiken. Zustandsübergangsnetze beschreiben die Aktionsfolge und damit erzeugte Systemzustände als Zustandsübergangsdiagramme, die Rekursionen sowie zulässige und unzulässige Übergänge enthalten können. Die so darstellbaren Aktionsfehler lassen sich jedoch nur als Verletzung von systemseitigen Interaktionsregeln abbilden. Sie geben keinen Hinweis, welche kognitiven Vorgänge den Fehler konstituieren.

Petri-Netze sind geeignet, um diskrete parallele und konkurrierende Ereignisse bei der Informationsübertragung darzustellen. Schlick (2000) benutzte Petri-Netze zur Modellierung der Mensch-Maschine-Interaktion in autonomen Produktionszellen. Kontextfreie Grammatiken beschreiben eine Sprache in Produktionsregeln, z.B. in Backus-Naur-Form, einer formalen Notation der Syntax einer Sprache (Naur 1960). In diese Produktionsregeln können sowohl kognitive Aktivitäten einbezogen, als auch die Interaktion hinsichtlich der strukturellen Konsistenz und Komplexität beurteilt werden. Die Darstellung der Interaktionen von Benutzern mit dem Computersystem in Backus-Naur-Form erschwert es aber, Interaktionsdauern abzubilden.

Ein äußerst einflussreiches konzeptuelles Modell der formalen Benutzermodellierung ist das GOMS-Modell von Card, Moran und Newell (1983), welches mit verschiedenen Methoden modelliert wurde. GOMS steht für *goals*, *operators*, *methods* und *selection rules*. Mittels dieser vier Begriffe lässt sich die Benutzung eines interaktiven Systems charakterisieren. Ziele (*goals*) werden als vom Benutzer angestrebte Zustände definiert, die in Unterziele heruntergebrochen werden können. Für die Erstellung eines Briefes muss beispielsweise ein Schreibprogramm geöffnet und eine Schriftart ausgewählt werden. Operatoren (*operators*) sind die im System verfügbaren elementaren Interaktionstechniken, z.B. Mausbewegungen und Tastaturelemente. Methoden (*methods*) werden durch Operatoren sequentiell ausgelöst, um ein Ziel zu erreichen. Selektionsregeln (*selection rules*) werden dann angewandt, wenn es mehrere Methoden zur Erreichung eines Zieles gibt. Mit dem GOMS-Modell kann die erwartete Effizienz geübter Benutzer analysiert werden, indem den Modellkomponenten Zeiten zugeordnet werden. Die Interaktionseffizienz für weniger geübte Benutzer abzuschätzen ist mit Hilfe des GOMS-Modells nicht möglich. Die Erstellung des Modells kann für funktional umfangreiche Systeme sehr aufwendig und im Rahmen verkürzter Produktentwicklungszeiten nicht mehr einsetzbar sein. Marrenbach (2000) stellt ein Werkzeug vor, das die Erstellung von GOMS-Modellen unterstützt und über die NGOMSL-Analyse (Kieras 1988) eine Lernzeitabschätzung ermöglicht. Problematisch an diesem Ansatz bleibt jedoch, dass Gestaltungsziele, wie Intuitivität der Interaktionsmetaphern sowie ähnliche auf der pragmatischen und semantischen Gestaltungsebene angesiedelte Probleme mit GOMS-Modellen schwerlich beurteilt werden können. Ein weiterer bedeutsamer Kritikpunkt an GOMS-Modellen besteht darin, dass die Entscheidung über die Körnungsgröße der Analyseeinheiten willkürlich festgelegt werden kann. Der Detaillierungsgrad der Modellierung beeinflusst jedoch das Evaluationsergebnis, da für ein feiner aufgelöstes Modell mit vielen Einheiten, denen allen eine Zeitausdehnung zugeordnet wird, die Wahrscheinlichkeit steigt, dass die Interaktion als zeitaufwendig eingestuft wird.

Expertenevaluationen werden häufig auch als heuristische Evaluationen bezeichnet, da die Bewertung des Systems auf sogenannten Faustregeln und erlernten Zusammenhängen beruht. Die Expertenevaluationen können in unterschiedlichen Phasen der Systementwicklung durchgeführt werden. Wird die Beurteilung nur auf der Basis der schriftlichen Spezifikation vorgenommen, wird auch von einer „kalten“ Einschätzung gesprochen. Eine „warme“ Einschät-



zung kann mit Hilfe eines Prototypen mit Teilfunktionalität und eine „heiße“ Einschätzung am fast vollständigen Produkt durchgeführt werden. Bei Expertenevaluationen werden verschiedene Verfahren eingesetzt, z.B. *cognitive walkthroughs*, bei denen sich Experten in die Rolle eines wenig erfahrenen Benutzers versetzen (vgl. Lewis & Wharton 1997) oder die Heuristische Evaluation nach Nielsen, die unter 3.4.1 näher beschrieben wird. Expertenevaluationen eignen sich besonders, um Probleme anhand von Richtlinien zu ermitteln und die Benutzungsschnittstelle von groben Fehlern zu befreien (vgl. Kantner 1997). Sie sind deshalb günstig vor der Durchführung von Tests mit Benutzern anzuwenden. Die meistens zahlreich diagnostizierten Schwachstellen erfordern jedoch eine systematische Gewichtung, da sie unterschiedlich eng mit benutzerorientierten Akzeptanz- und Leistungsmaßen zusammenhängen und die im Projekt verfügbare Zeit für die gleichrangige Abarbeitung aller Hinweise üblicherweise nicht ausreicht.

Die Evaluation mit potentiellen oder realen Benutzern kann mit Hilfe unterschiedlicher Methoden vorgenommen werden. Es kann in unterschiedlichen Entwicklungsphasen zu bewertenden, summativen oder konstruktiven, formativen Zwecken evaluiert werden. Für die Beurteilung einer Benutzungsschnittstelle nach ISO-Norm-Kriterien (vgl. DIN ISO EN 9241 Teil 10) stehen beispielsweise Fragebögen zur Verfügung, wie der IsoMetrics (Gediga & Hamburg 1999). *Focus groups*, in denen mit Benutzern Konzepte und Eigenschaften eines Systems diskutiert werden, Kreativitätssitzungen, in denen vorzugsweise geeignete Interface-Metaphern gesucht werden, oder die Auswertung von Rückmeldungen von Benutzern an die Servicebereiche geben einen Eindruck, wie vielfältig Daten für die Systemgestaltung gewonnen werden können (vgl. Wixon & Wilson 1997). Ein aufgrund des gewählten testtheoretischen Vorgehens ökonomischer Ansatz zur Beurteilung der Gebrauchstauglichkeit nach den ISO-Norm-Kriterien wurde mit dem Ergo-Norm-Prüfverfahren (Dzida et al. 2001) vorgestellt. Das Ergo-Norm-Prüfverfahren prüft ein Software-Produkt darauf, inwiefern es von der ISO-Norm abweicht. Ausgehend von der Annahme der Konformität wird versucht, an objektiv und valide aus dem Einsatzkontext ableitbaren Anforderungen an das Produkt die Konformitätsannahme zu falsifizieren.

In Usability-Laboren ist die Durchführung von Tests, die experimenteller Arbeitsweise folgen, verbreitet. Für die Beurteilung der Benutzungsschnittstelle werden dabei unterschiedliche Daten ausgewertet: automatische Interaktionsprotokolle (*logfiles*), Videoaufzeichnungen, Beobachtungen, Protokolle über Benutzeräußerungen, psychophysische Messwerte, Leistungsmaße, Interviews und andere mehr (vgl. Baggen & Hemmerling 2000). Empfehlenswert scheint eine Kombination verschiedener Methoden für die Systemevaluation, da so die jeweils spezifischen Schwächen einer Methode kompensiert und unterschiedliche Typen von Problemen identifiziert werden können (vgl. Landauer 1988). Tests mit Benutzern setzen voraus, dass für die zukünftige Nutzung kritische oder besonders wichtige Systemfunktionalitäten eingegrenzt werden können, da eine quantitative und inferenzstatistische Absicherung der gewonnenen Daten für eine große Anzahl von experimentellen Variationen enorm aufwendig ist.

Sofern für die Datenanalyse Software-Routinen zur Verfügung stehen, lassen sich Effizienz-nachteile von Benutzungstests gegenüber Expertenevaluationen kompensieren. Die auf wesentliche Systemmerkmale fokussierten Tests können bei entsprechend schneller Ergebnisaufbereitung wichtige Informationen bieten, die mit moderatem Zeitaufwand in die Systementwicklung umgesetzt werden können. Eine Problematik von Tests mit Benutzern ist es, dass für Systeme, die sehr weit verbreiteten Einsatz finden, der „durchschnittliche“ Benutzer schwer zu bestimmen und für die Teilnahme an Tests nicht leicht zu gewinnen ist. Bei Systementwicklungen für eine spezifische Anwendungsdomäne sind die zukünftigen Benutzer und ihre Voraussetzungen, wie Ausbildung etc., zwar weitestgehend bekannt, aber sie müssen für die Teilnahme an Tests gewonnen werden. Dies beeinflussen zumeist Faktoren, die eher im organisationalen betrieblichen Umfeld angesiedelt sind, z.B. die Unterstützung der Benutzerbeteiligung in der Systementwicklung durch den Auftraggeber.

### **3.3 Entwicklung eines Evaluationsmodells für die Entwicklung multimodaler Systeme bei nicht spezifiziertem Anwendungskontext**

#### **3.3.1 Multimodale Systementwicklung als Gegenstand der Forschung**

Für neue Konzepte im Bereich der Mensch-Computer-Interaktion besteht die Notwendigkeit, Prototypen auf der Basis innovativer Technologien zu realisieren. Einerseits, um die technische Umsetzbarkeit der Konzeption im Sinne von Machbarkeitsstudien zu prüfen und andererseits, um Erfahrungen über die Interaktionsgestaltung zu gewinnen.

In diesem frühen, forschungsbezogenen Stadium der Systementwicklung kann der Anwendungskontext einer neuen Technologie oftmals erst grob abgeschätzt werden, da es auf der technischen Seite vorerst gilt, Probleme der Genauigkeit, der Zuverlässigkeit und der Geschwindigkeit von Informationsverarbeitungsprozessen des Systems zu lösen. Ein weiteres Problem solcher neuen interaktiven Systeme besteht in einem Mangel an Gestaltungsstandards. Deshalb ist es notwendig, wenn die entsprechenden soft- und hardwaretechnischen Voraussetzungen geschaffen sind, an dem Prototypen Erfahrungen über die Gestaltung der Interaktion mit Hilfe von Testapplikationen zu gewinnen.

In einem iterativen Entwicklungsprozess werden die Phasen der Implementierung von Soft- und Hardware von Schritten zur Bewertung von Gestaltungsentscheidungen begleitet. Die formative Systembewertung fokussiert die Fragen zur Gestaltung im jeweiligen Implementierungsstatus. Die benutzungsorientierten Untersuchungen während der Systemerstellung dienen dazu, Probleme in der entsprechenden Gestaltungsphase aufzudecken sowie Gestaltungsalternativen zu vergleichen und auszuwählen, um schließlich einen Prototypen nach dem aktuellen Stand der Erkenntnis vorliegen zu haben. Die summative Bewertung zielt auf eine abschließende Beurteilung dieses in mehreren iterativen Schritten entwickelten prototypi-

schen Systems auf der Basis der damit verbundenen Zielstellungen. Durch diese Evaluations-schritte können sowohl Informationen für die Gestaltung des spezifischen Systems gewonnen als auch verallgemeinerbare Gestaltungshinweise für die neuen Interaktionsformen abgeleitet werden.

Die Ausbildung von Standards und Gestaltungsrichtlinien für neue interaktive Technologien kann jedoch nicht auf der Basis nur eines prototypischen Systems gelingen, da konzeptionell sehr unterschiedliche Systemrealisationen denkbar sind. Diese Unterschiedlichkeit begründet sich einerseits aus der Vielzahl von technischen Einzelkomponenten, die ein System konstituieren können, andererseits aber auch aus dem antizipierten Anwendungszusammenhang. Vor diesem Hintergrund spannt sich ein breites Forschungsfeld auf, das umfangreiches Wissen über die Gestaltung und den potentiellen Einsatz innovativer Technologien im Bereich der multimodalen Mensch-Computer-Interaktion generiert. Die während dieser Phase gewonnenen Erfahrungen können insbesondere dann leicht in späteren produktorientierten Systementwicklungsphasen genutzt werden, wenn es gelingt, sie in systematisch abgelegte Informationen zu transponieren.

### **3.3.2 Evaluation der multimodalen Interaktion an prototypischen Aufgaben**

Eine Vorgehensweise, die sich für die strukturierte Informationsdarstellung und -gewinnung in frühen Phasen der Konstruktion in den Ingenieurdisziplinen als nützlich erwiesen hat, ist die Darstellung in einer n-dimensionalen Matrix, dem „Morphologischen Kasten“ (Pahl & Beitz 1997), der von Zwicky (1966) entwickelt wurde. Im morphologischen Kasten werden systematisch Informationen kombiniert, z.B. konstruktive Lösungen für spezifische Funktionen, deren Verknüpfung zu einer Gesamtlösung führen. Diese Methodik kann für die Informationsdarstellung über Systemeigenschaften von multimodalen Computersystemen und deren Wirkung auf spezifische Aufgaben abgewandelt werden. Der in diesem Falle 2-dimensionale Morphologische Kasten stellt den möglichen Lösungsraum für die Systementwicklung vor dem Hintergrund geforderter Funktionalität dar. Zu prüfen ist dann für ein spezifisches zu gestaltendes System, welche Lösungsvarianten bzw. Systemeigenschaften widerspruchsfrei miteinander kombinierbar sind. Je mehr Informationen Eingang finden in die Matrix, um so effektiver kann sie für die zukünftige, produktorientierte multimodale Systemgestaltung genutzt werden.

Wenn die in der Matrix enthaltene Information durch weitere informationstechnische und benutzungsorientierte Forschung vervollständigt wird, kann sie zunehmend mehr Orientierungshilfe für die Systemgestaltung anbieten. Neben den Erkenntnissen darüber, wie einzelne Modalitäten sinnvoll zu gestalten und zu kombinieren sind, lassen sich auch Hinweise auf einen vorteilhaften Einsatz in unterschiedlichen Domänen finden. Die Dokumentation der Funktionalität eines Systems für unterschiedliche, definierte Testaufgaben in der Matrix kann für zukünftige domänenspezifische Entwicklungen genutzt werden. Werden für eine Produktentwicklung Aufgaben analysiert, die durch das neu zu entwickelnde System unterstützt wer-

den sollen, so können diese mit den in der Matrix aufgeführten Testaufgaben verglichen werden. Übereinstimmende Aufgabenbereiche helfen dann, die korrespondierenden Systemeigenschaften zu ermitteln, die sich in sehr frühen Phasen der multimodalen Systementwicklung als zweckmäßig erwiesen haben. Im Rahmen der in dieser Arbeit vorgestellten empirischen Arbeit wird dazu ein erster Beitrag geleistet.

Der vergleichenden Bewertung von multimodalen Systemen stehen gegenwärtig noch verschiedene Probleme entgegen. Aufgrund des prototypischen Status' der Entwicklung multimodaler Systeme und der stark divergierenden potentiellen Anwendungsbereiche sind sie bisher nicht miteinander vergleichbar. Es liegen noch wenig Informationen aus benutzungsorientierten Evaluationsstudien für die existierenden Prototypen vor, die aufgrund unterschiedlicher methodischer Vorgehensweise zudem schwer zu integrieren sind. Eine Gegenüberstellung eines solchen Systems mit herkömmlichen Anwendungen ist schwierig, da multimodale Systementwicklung auf einem anderen Interface-Konzept beruht, welches auch Veränderungen in der Systemarchitektur bzw. Plattform nach sich zieht. Die hierfür entwickelten Testanwendungen bieten üblicherweise nicht den Funktionsumfang kommerzieller Anwendungen oder unterstützen völlig veränderte Tätigkeitsabläufe, z.B. das *Field Medic Information System* (siehe Abbildung 13).

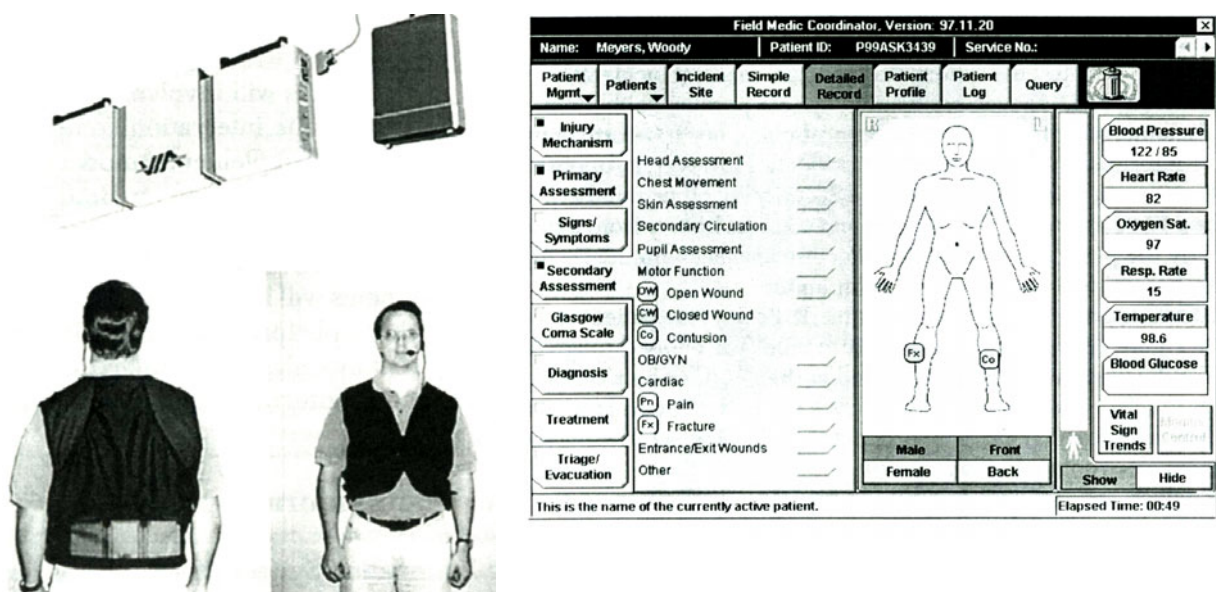


Abb. 13: *Field Medic Information System* von NCR Corporation (aus Oviatt et al. 2000): die *Field Medic Associate Hardware* mit Spracherkennung ist in die Weste des Notarztes integriert (links), das Interface des *Field Medic Coordinator* mit alternativer Sprach- oder Stifteingabe (rechts).

Dieses System ist für den Einsatz eines Notarztes im Außendienst entwickelt worden. Der Arzt kann den Status des aufgesuchten Patienten über Spracheingabe während der Untersuchung in das System eingeben. Außerdem steht ihm ein digitales Patientenprotokoll zur Verfügung, in das er mit einem Stift Eingaben vornehmen kann. Die Daten werden per Funknetz

an das aufnehmende Krankenhaus übermittelt, wo bei Ankunft des Patienten bereits Vorbereitungen zu dessen weiterer Behandlung getroffen sein können. Bisher bestand für den Notarzt keine Möglichkeit, die Daten zum Gesundheitszustand des behandelten Patienten sofort zu dokumentieren und das aufnehmende Krankenhaus verfügte bis zur Ankunft des Patienten nur über wenig detaillierte Informationen.

Eine Möglichkeit der Bewertung solcher innovativer Mensch-Rechner-Interaktionskonzepte besteht darin, ihre Wirkung im Zusammenhang mit den Aufgaben, die die prototypische Systemrealisation unterstützen soll, zu untersuchen. Ein Computersystem, als informationsverarbeitendes und informationsdarstellendes Werkzeug, wird zweckmäßigerweise daran gemessen, inwiefern es Personen dabei unterstützt, spezifische mit ihrer Aufgabe verbundene Anforderungen zu erfüllen. Für die Entwicklung von Testaufgaben bedarf es einer Beschreibungssystematik, die die Aufgaben mit den implizierten Anforderungen an den Menschen zu definieren hilft. Im Rahmen der Untersuchung prototypischer Systemrealisierungen ist eine Beschreibungssystematik nützlich, die es erlaubt, Aufgaben domänenunabhängig nach ihren wesentlichsten Anforderungen zu klassifizieren. Die benutzten Beschreibungsdimensionen sollten leicht verständlich und konsensfähig sein sowie mit geringem Aufwand auf den jeweiligen Untersuchungsgegenstand übertragen werden können. Werden die Evaluationsergebnisse für spezifische multimodale Systeme vor dem Hintergrund ein und derselben Aufgabensystematik diskutiert, können die Prototypen mit unterschiedlichen Systemeigenschaften und die Ergebnisse unterschiedlicher Evaluationsparadigmen besser verglichen und integriert werden. In der humanwissenschaftlichen Forschung haben sich Ansätze mit jeweils unterschiedlichen Zielstellungen herausgebildet, um Aufgaben und die mit ihnen verbundenen Anforderungen an den Menschen strukturiert abzubilden.

### **3.3.2.1 Beschreibung von Aufgaben und Aufgabenmerkmalen**

Die Mensch-Computer-Interaktion im Arbeitskontext steht üblicherweise im engen Zusammenhang mit der Unterstützung informationsverarbeitender Prozesse, die für die Erfüllung der Arbeitsaufgabe relevant sind. Als Arbeitsaufgabe wird nach Hacker (1986) die Übernahme und Redefinition des objektiven, rechtlich und betriebswirtschaftlich festgelegten Arbeitsauftrages durch die Arbeitsperson verstanden. Aufgaben sind geprägt durch die jeweiligen kognitiven und sensumotorischen Anforderungen, die die Person bei der Aufgabenbearbeitung zu bewältigen hat.

Abgesehen von der Problematik, Anforderungen intersubjektiv bestimmbar, d.h. unabhängig von den *wahrgenommenen* Anforderungen zu betrachten, werden Aufgabenanforderungen dennoch als objektivierbare Entitäten analysiert (vgl. Hacker 1995, Wieland-Eckelmann 1992). In einer sehr umfangreichen Systematik über Aufgabencharakteristika beschreiben Fleishman und Quaintance (1984), neben der Möglichkeit zur Klassifikation, auch die Nützlichkeit einer aufgabenanalytischen Perspektive für die Prädiktion von Leistungen.

Aufgabenanalytische Ansätze stoßen jedoch auf Hindernisse, die insbesondere bei der Computerarbeit bedeutsamen mentalen Prozesse für die Aufgabenerfüllung zuverlässig zu erfassen und zu beschreiben, da diese der direkten Beobachtung nicht zugänglich sind (Sheridan 1997). Auf die an der Aufgabenbearbeitung beteiligten kognitiven Prozesse kann nur indirekt geschlossen werden. Die Adäquatheit und Vollständigkeit der inferierten mentalen Ereignisse bei der Aufgabenbeschreibung sind daher per se schwierig zu prüfende Kriterien der aufgabenanalytischen Methode. Für die Beschreibung und Klassifizierung kognitiver Prozesse im Rahmen arbeitswissenschaftlicher und arbeitspsychologischer Analysen liegen eine Vielzahl von Beschreibungskategorien vor, die sich in ihrem Detaillierungsgrad und ihrer theoretischen Fundierung unterscheiden.

R.B. Miller veröffentlichte bereits ab der Mitte des letzten Jahrhunderts zahlreiche Artikel über Prinzipien und Vorgehensweise bei der Aufgabenbeschreibung und Analyse (z.B. Miller 1962, 1966), die viele weitere Forschungsarbeiten in der Domäne befruchteten. Sein pragmatischer Ansatz berücksichtigt Informationsverarbeitungsprozesse in engem Bezug zu Anforderungen bei der Überwachung von komplexen Systemen (Monk & Gilbert 1995). Für die Analyse von Tätigkeiten in komplexen Mensch-Maschine-Systemen liegen weitere, aktuelle Ansätze vor, z.B. von Rasmussen, Pejtersen und Goodstein (1994) und von Hollnagel (1998). Rasmussens Kategorisierung kognitiver Prozesse als fertigungs-, regel- und wissensbasiert kann nur im Hinblick auf den Wissensstand des Operateurs operationalisiert werden, da Lernprozesse die Zuordnung kognitiver Anforderungen zu den Ebenen beeinflussen. Hollnagels Beschreibungsdimensionen kognitiver Leistungen sind, dem Zweck der Methode entsprechend, stark auf die konkreten Anforderungen bei Überwachungstätigkeiten in dynamischen Systemen ausgerichtet. Vier kognitiven Funktionen - Beobachten, Interpretieren, Planen und Ausführen - werden kognitive Aktivitäten zugeordnet, dabei werden besonders feine sprachliche Unterscheidungen für das Beobachten und die Regulierung von Parametern getroffen. Eine Untersuchung von Giesa und Köhler (2001) zeigt, dass Experten mit der Zuordnung von kognitiven Aktivitäten zu den vier postulierten kognitiven Funktionen sowie den Definitionen mehrerer Aktivitäten nicht übereinstimmen. Für den infrage stehenden Nutzungszweck, innovative Interaktionsformen zwischen Mensch und Computer zu untersuchen, erscheint eine abstraktere Beschreibung der zentralen Anforderungsdimensionen der bearbeiteten Aufgaben angemessener.

Tätigkeitsanalyseverfahren, die auf der Handlungstheorie aufbauen, untersuchen und ordnen Tätigkeiten nach dem Grad von Planung und Kontrolle, die für ihre Ausführung nötig sind. Mit diesen Verfahren sollen insbesondere potentielle Einschränkungen in der Persönlichkeitsentwicklung durch unvollständige Tätigkeiten mit geringen Planungs- und Kontrollanteilen ausgemacht werden (Hacker 1995). Arbeitsgestalterische Maßnahmen können dann gezielt eingesetzt werden, um eine interessante und sinnvolle Tätigkeit am analysierten Arbeitsplatz zu implementieren. Aufgaben bzw. Handlungen werden einem spezifischen Niveau von Planung und Kontrolle zugeordnet, welches sehr abstrakt beschrieben ist, um auf möglichst viele Tätigkeiten anwendbar zu sein.

Arbeitswissenschaftliche Erhebungsverfahren basieren überwiegend auf dem Belastungs-Beanspruchungskonzept, welches gegenwärtig jedoch in ergonomischen, handlungsorientierten und stresstheoretischen Ansätzen unterschiedlich definiert ist (vgl. Schönplflug 1987). Der AET (Rohmert & Landau 1983) ist ein Beispiel für diese Erhebungsverfahren. Anforderungen an die Arbeitsperson, die über personeninterne Moderatorvariablen das Beanspruchungsmaß beeinflussen, werden bei dieser Analyse nicht auf die motorischen, perzeptiven und höheren kognitiven Anforderungen der konkreten Arbeitsaufgabe beschränkt, sondern umfassen auch belastende Umgebungsfaktoren. Ein sehr umfangreiches arbeitspsychologisches Verfahren ist der TAI (Frieling et al. 1993), der insgesamt 2000 Beurteilungssitems zur Arbeitssystem-, Aufgaben- und Tätigkeitsanforderungsanalyse umfasst. Allerdings werden kognitive Prozesse, die der direkten Beobachtung nicht zugänglich sind, nicht erfasst. Weitere arbeitsanalytische Verfahren wurden für spezifische Arbeitsplätze entwickelt, z.B. für die Arbeit an CNC-Werkzeugmaschinen (Weber et al. 1994). Alle Verfahren gehen von einer detaillierten Beschreibung der Tätigkeit aus, die dann der theoretischen Konzeption folgend, entsprechenden Kategorien bewertend zugeordnet wird, um arbeits- und organisationsgestalterische oder eignungs- und qualifikationsbezogene Erfordernisse abzuleiten. Dunckel (1999) betont, dass „die psychologische Arbeitsanalyse ohne Beteiligung der konkret arbeitenden Individuen nicht denkbar“ ist. Damit sind sie für den Zweck, personenunabhängig allgemeine Beschreibungsdimensionen für menschliche Leistungen zu liefern, weniger angebracht.

Eine von den aufgabenanalytischen Ansätzen verschiedene, wesentlich detailliertere Perspektive auf kognitive Prozesse findet sich bei der psychologischen Intelligenz- bzw. Eignungsdiagnostik. Hier wurden speziell durch den Einsatz der Faktorenanalyse eine Vielzahl kognitiver Leistungen voneinander abgegrenzt. Die Absicht der Untersuchungen stand jedoch nicht in Zusammenhang mit der Analyse von arbeitsbezogenen Aufgaben und den dafür erforderlichen mentalen Fähigkeiten, sondern vielmehr in der Entwicklung einer Theorie und Struktur von menschlichen Fähigkeiten (vgl. z.B. Conrad 1983, Ekstrom, French & Harman 1979, Funke & Vaterrodt-Plünnecke 1998). Für die benutzer- und domänenunabhängige Beschreibung prototypischer Aufgaben erweisen sich die betrachteten Ansätze als ungeeignet.

Berliner, Angell und Shearer (1964) entwickelten ein Schema zur Verhaltensklassifikation, welche die Leistungsmessung in einem möglichst breiten Bereich von Tätigkeiten unterstützen sollte (siehe Tabelle 1). Damit sollte sichergestellt werden, dass in Simulatortrainings erworbene Fähigkeiten quantitativ und differenziert bewertet werden können, da das externe Kriterium der Trainingszeit allein keine ausreichende Beurteilung des Lernfortschritts zulässt. Die Deskriptoren des Schemas sind hierarchisch angeordnet, wobei vier behaviorale Prozesse die höchste Ebene bilden, darunter ordnen sich sechs Aktivitäten, denen auf der unteren Ebene 47 spezifische verhaltensbeschreibende Verben zugeordnet werden. Die Verben zur Beschreibung spezifischen Verhaltens sollten weitestgehend verständlich und semantisch genügend von einander abgegrenzt sein. Die Deskriptoren der unteren Ebene wurden danach ausgewählt, ob sie zuverlässig zu identifizieren waren, ob sie für einfache Handlungen mit quantifizierbaren Eigenschaften standen und ob sie in verschiedenen Tätigkeiten anzutreffen wa-

ren. Die Systematik wurde vervollständigt, indem den Kategorien Messinstrumente hinzugefügt wurden, so dass eine Matrix mit den Koordinaten Verhalten, Messvariablen und Instrumente entstand (Rabideau 1964). Für das Design von Untersuchungen zur Systembewertung mit Hilfe von Leistungsdaten kann diese Matrix unterstützend genutzt werden.

Die Termini der Taxonomie wurden nicht näher definiert, so dass „private Definitionen“ in die Analyse eingehen können. Dies erschwert u.U. die übereinstimmende Kategorisierung von vorgefundenen Verhaltensabläufen, die bei der Erfüllung einer Aufgabe von verschiedenen Beurteilern beobachtet werden. Das Kategorisierungsproblem, besonders für nicht direkt beobachtbares Verhalten wie Interpolieren, besteht generell für alle Dekompositionsmodelle, so dass Praktikabilitätsabwägungen im jeweiligen Untersuchungskontext auswahlrelevant sind. Christensen und Mills (1967) verweisen bei der Anwendung der Taxonomie von Berliner, Angell und Shearer darauf, dass mit geringem Aufwand zur Verinnerlichung der Beschreibungsdimensionen eine Interraterübereinstimmung mit einer Median-Korrelation von +.97 erreicht wurde. Für den Zweck des Kreierens oder der Zuordnung von Testaufgaben dienen die Beschreibungsdimensionen in erster Linie als Anhaltspunkt, welche spezifischen Verhaltensausprägungen für eine prototypische Aufgabe kennzeichnend sind. Die „persönlichen“ Definitionen der Evaluatoren werden dadurch transparent, dass die konkrete Untersuchungsaufgabe in der Evaluationsstudie zu beschreiben ist. Da die Systematik entwickelt wurde, um Leistungen zu bewerten, ist sie für die Evaluation von Computersystemen insofern geeignet, als dass in diesem Bereich ebenfalls Effektivität und Effizienz bedeutsame Bewertungskriterien darstellen. Durch die experimenteltechnische Eliminierung oder Kontrolle von Personen- und Trainingsvariablen sind Leistungsunterschiede auf Systemmerkmale zurückführbar.



Tab. 1: Berliner Klassifikationsschema

Prozesse	Aktivitäten	Spezifisches Verhalten
Perzeptive Prozesse	Suche und Empfangen von Informationen	Entdecken
		Prüfen
		Beobachten
		Lesen
		Empfangen
		Absuchen
		Begutachten
	Objekte, Handlungen, Ereignisse identifizieren	Unterscheiden
		Identifizieren
		Lokalisieren
Vermittelnde Prozesse	Informationsverarbeitung	Kategorisieren
		Kalkulieren
		Kodieren
		Berechnen
		Interpolieren
		Spezifizieren
		Ordnen
		Übersetzen
	Problemlösen und Entscheiden	Analysieren
		Kalkulieren
		Auswählen
		Vergleichen
		Berechnen
		Schätzen
		Planen
		Kommunikative Prozesse
Antworten		
Kommunizieren		
Anweisen		
Hinweisen		
Informieren		
Unterweisen		
Anfragen		
Übermitteln		

Tab. 1: Berliner Klassifikationsschema (Fortsetzung)

Motorische Prozesse	komplex-kontinuierlich	Einstellen
		Justieren
		Regulieren
		Synchronisieren
		Verfolgen
	einfach-diskret	Auslösen
		Schließen
		Verbinden
		Trennen
		Zusammenfügen
		Bewegen
		Drücken
		Fixieren

### 3.3.2.2 Verknüpfung von Systemeigenschaften und Aufgaben

Als Ansatz zur systematischen Darstellung von interaktionsbezogenen Evaluationsergebnissen für multimodale Computersysteme kann in Anlehnung an das Prinzip des Morphologischen Kastens eine Matrix aufgebaut werden, in der die spezifische prototypische Systemrealisierung eine Koordinate der Matrix bildet, während die abstrakteren Beschreibungskategorien der Aufgaben nach dem Berliner Klassifikationsschema auf der anderen Koordinate der Matrix eingetragen werden. Die Matrix wird im Folgenden als System-Aufgabeneigenschaftsmatrix bezeichnet. In Tabelle 2 ist das Gerüst der Matrix dargestellt, die am Ende der Darstellung mit den im Rahmen dieser Arbeit erhobenen Evaluationsergebnisse gefüllt wird.

Unter den Systemeigenschaften wird interaktionsbezogen zwischen der Informationsübertragung vom Menschen zur Maschine und von der Maschine zum Menschen unterschieden. Innerhalb der Kategorie der Informationsübertragung vom Menschen zur Maschine sind mögliche Modalitäten mit den verschiedenen Varianten der systemseitigen Informationserkennung aufgeführt. Für die Informationsübertragung von der Maschine zum Menschen sind verschiedene vom Menschen wahrnehmbare Kanäle nutzbar, wobei auch die Informationsausgabe unterschiedlich gestaltet werden kann. Die Aufgabeneigenschaften definieren sich über die Leistungen, die für ihre Bearbeitung vom Menschen zu erbringen sind. Dabei wird entsprechend Berliners Klassifikationsschema zwischen perceptiven, vermittelnden (im engeren Sinne kognitiven) Prozessen, motorischen Prozessen und kommunikativen Prozessen unterschieden. Die gesonderte Kategorie „kommunikative Prozesse“ verweist auf die über die Maschine vermittelte Kommunikation mit anderen Personen, ein bedeutsames Einsatzgebiet von Computersystemen. Die übrigen Prozesse werden noch in jeweils zwei Verhaltenskategorien, die verschiedene Aspekte der Prozesse abbilden, unterteilt.

Für die Evaluation von Systemeigenschaften können mit Hilfe der Verben aus der unteren Ebene des Berliner Klassifikationsschemas Aufgaben zu den Prozessen und Verhaltenskategorien nach ihren schwerpunktmäßigen Anforderungen zugeordnet werden. Die Beschreibungsdimensionen der unteren Ebene des Berliner Klassifikationsschemas können aber auch genutzt werden, um Testaufgaben zu kreieren, die als relevant erachtete Leistungen für die Aufgabenbewältigung erfordern.

Die Matrix beinhaltet die aus der Evaluation gewonnen Informationen über die Wirkung der geprüften Gestaltungsvarianten der multimodalen Prototypen im Zusammenhang mit den jeweils fokussierten prototypischen Aufgaben (vgl. Tabelle 2). So können für das multimodale Interaktionskonzept, für das es aufgrund seiner Neuartigkeit vorerst an Gestaltungsrichtlinien und Standards mangelt, aus den Evaluationsergebnissen Gestaltungsempfehlungen abgeleitet werden.

Die systematische empirische Betrachtung von spezifischen Systemen unter prototypischen Aufgaben verbreitert die Wissensbasis für den Einsatz und die Gestaltung von multimodalen Systemen. Sie erleichtert das Auffinden von Invarianten, die als Richtlinien bzw. Heuristiken für die Gestaltung etabliert werden können. Da jedes informationstechnische System das Ergebnis eines Problemlöseprozesses mit kreativem Anteil darstellt, ist es angemessener, empirisch fundierte Heuristiken für die multimodale Systemgestaltung abzuleiten, als ein vollständiges Regelwerk abbilden zu wollen. Die Heuristiken ermöglichen es, den Schritt von der allein auf der Intuition der Entwickler beruhenden Auslegung eines multimodalen Systems hin zu einer stärker angeleiteten Gestaltung zu gehen. In einer weitgehend gefüllten Matrix, die die Evaluationsergebnisse für verschiedene Systemauslegungen an unterschiedlichen Aufgaben darstellt, können Systemkomponenten und Interaktionskonzepte nach ihrer grundsätzlichen Eignung für spezifische Aufgabenkontexte kombiniert werden. Die Matrix kann demnach als „Findehilfe“ für zukünftige Systementwicklungen ausgebaut werden.

Tab. 2: Matrix für die strukturierte Zuordnung von Systemeigenschaften zu Aufgabeneigenschaften

Mensch>Maschine		Systemeigenschaften			Aufgabeneigenschaften				
		Maschine>Mensch	Perzeptive Prozesse		Vermittelnde Prozesse	Kommunikative Prozesse	Motorische Prozesse		
Modalitäten	Eingabe			Suche nach und Empfangen von Informationen	Identifizieren von Objekten, Handlungen und Ereignissen				
	Rohdaten der Blickerfassung	wahrnehmbare Medien	Ausgabe			Informationsverarbeitung	Problemlösen und Entscheiden	Komplex-kontinuierlich	Einfach-diskret
		visuell							
Blick		auditiv							
		haptisch							
	Parametrisierte Rohdaten der Blickerfassung	visuell							
Sprache		auditiv							
		haptisch							
	Fließend gesprochene Sprache	visuell							
Sprache		auditiv							
		haptisch							
	Einzelworte, Wortgruppen	visuell							
		auditiv							
		haptisch							

Systemeigenschaften		Aufgabeneigenschaften					
Mensch>Maschine	Maschine>Mensch	Perzeptive Prozesse		Vermittelnde Prozesse		Kommunikative Prozesse	Motorische Prozesse
Modalitäten	Wahrnehmbare Medien	Suche nach und Empfangen von Informationen	Identifizieren von Objekten, Handlungen und Ereignissen	Informationsverarbeitung	Problemlösen und Entscheiden		Komplex-kontinuierlich
Eingabe	Ausgabe						Einfach-diskret
2D-kontinuierlich (z.B. Maus, Stift & Tablett)	visuell auditiv haptisch						
2D-diskret (z.B. Tastatur)	visuell auditiv haptisch						
3D-diskret/kontinuierlich (z.B. Bilderkennung, Phantom)	visuell auditiv haptisch						
Ggf. weitere Modalitäten							

Die Kriterien, an denen die Prototypen gemessen werden können, sind grundsätzlich beliebig wählbar. Für einen besseren Vergleich von Systemeigenschaften und deren Wirkung in einem Aufgabenrahmen ist es jedoch sinnvoll, die im Rahmen der Bewertung von ergonomischen Untersuchungen eingeführten Kriterien zu berücksichtigen. Neben der Leistung und der Beanspruchung, die in der Mensch-Maschine-Systemforschung von zentraler Bedeutung sind, sind die in der software-ergonomischen Evaluation häufig herangezogenen subjektiven Beurteilungen der Benutzungseigenschaften von Software allgemein anerkannte Bewertungskriterien. Die DIN ISO EN 9241 Teil 11 benennt ebenfalls die drei Bereiche der Bewertung, die Effektivität, die durch die Messung von Leistung erfolgen kann, die Effizienz, die den Aufwand berücksichtigt, der für die Erbringung der Leistung nötig ist und die Akzeptanz, die auf der Zufriedenheit und der Beeinträchtigungsfreiheit der Benutzer basiert. Für die Erfassung der Effizienz können unter standardisierten Umgebungsbedingungen und der Kontrolle wichtiger moderierender Personenvariablen, wie Motivation, kognitive Leistungsvoraussetzung und Vorerfahrung, Beanspruchungsmaße erhoben werden. Die Evaluation von Systemen im Labor bietet für die Standardisierung der Untersuchungsbedingungen und die Kontrolle von Personenvariablen in der Benutzergruppe günstige Bedingungen. Die Akzeptanz eines Systems kann mit verschiedenen Methoden der Befragung eingeschätzt werden. Im nächsten Abschnitt werden für die Systembewertung hilfreiche Methoden erläutert. Bei der Evaluation multimodaler Systeme sollten die Bewertungskriterien, die in den ISO-Norm-Leitsätzen formuliert sind, berücksichtigt werden, um die angestrebte Vergleichbarkeit zu erreichen. Darüber hinaus können mit Hilfe verschiedener humanwissenschaftlicher Methoden noch weitere wichtige Informationen über das spezifische System gesammelt werden.

### **3.4 Empirische Evaluationsmethoden**

Die Evaluation von prototypischen multimodalen Systemen zur Ableitung von Gestaltungsrichtlinien kann auf ein breites Inventar humanwissenschaftlicher Methoden aufbauen, das sich in Untersuchungen der Mensch-Computer-Interaktion teilweise bereits bewährt hat. Die Auswahl der Methoden sollte sich einerseits nach dem Wissensstand richten, der über die Wirkung von Gestaltungsvarianten vorhanden ist, und andererseits nach der Zielstellung der Evaluation. Bei geringem Wissensstand eignen sich solche Methoden, mit deren Hilfe effizient große Mengen an Information gewonnen werden können. Mit zunehmender Kenntnis können auch feinauflösendere, eng fokussierte Methoden eingesetzt werden, beispielsweise ein Entscheidungsexperiment über mehrere spezifische Gestaltungsvarianten. Die Zielstellung der Evaluation kann darauf gerichtet sein herauszufinden, welche Effizienzvorteile eine Gestaltungsvariante mit sich bringt oder wie sie potentiellen Benutzerinnen anmutet. Erstere Zielstellung erfordert, solche Methoden zu wählen, die quantifizierbare Aussagen über Leistungen und Aufwände erlauben. Die letztere Zielstellung verlangt Methoden zur Erhebung von verbalen oder auch physiologischen Daten, die affektive Reaktionen von Personen zu erfassen vermögen. Im folgenden werden die für die in dieser Arbeit vorgestellten Untersuchungen verwendeten Methoden erläutert. Physiologische Maße wurden in diesen Untersuchungen

nicht abgeleitet, da für die in Frage kommenden keine kontaktfreien Verfahren verfügbar sind, so dass der Bewegungsfreiraum der Probanden stark eingeschränkt würde.

### 3.4.1 Explorative Methoden

Mit Hilfe explorativer Verfahren können große Informationsmengen erhoben werden, sie eignen sich deshalb sehr gut für die erste empirische Annäherung an einen Forschungsgegenstand, für den noch keine spezifischen Hypothesen abgeleitet werden können. Ein Überblick über qualitative Verfahren, die in der Evaluationsforschung eingesetzt werden können, findet sich bei Bortz und Döring (1995). Die Evaluation im Rahmen der Systementwicklung stellt besondere Anforderungen an die eingesetzten Methoden. Sie sollten möglichst geringen Aufwand in Vorbereitung, Durchführung und Auswertung verursachen, damit die Ergebnisse zügig in die Entwicklungsarbeit einfließen können. Eine praktikable und angemessene Evaluation stellt einen Kompromiss von Güte und Aufwand dar. Hierfür eignet sich eine explorative Evaluation mit Experten. Die Zuverlässigkeit und Gültigkeit der ermittelten Benutzbarkeitsprobleme hängt von den Fähigkeiten der Evaluatoren ab. Die Fähigkeiten beziehen sich nach Nielsen (1993) auf drei Bereiche:

- Erfahrungen mit der Benutzung von Software,
- Erfahrungen mit der software-ergonomischen Beurteilung von Benutzungsoberflächen,
- Kenntnisse bezüglich der geplanten Einsatzumgebung der Software.

Die Heuristische Evaluation nach Nielsen ist ein an allgemeinen Gestaltungsregeln ausgerichtetes pragmatisches Verfahren der Systembewertung, welches auf Expertenurteilen über Gestaltungsmängel beruht. Softwareexperten beurteilen anhand von *Usability*-Heuristiken (Molich & Nielsen 1990; vgl. Tabelle 3) die Benutzbarkeit aus der antizipierten Sichtweise des zukünftigen Benutzers. Die Heuristiken bieten Anker für die Beurteilung des Systems und strukturieren so die Aussagen. Dies unterstützt die zügige Auswertung der erhobenen Daten und damit die Berücksichtigung bei der weiteren Systementwicklung. Nielsen (1993) konnte zeigen, dass bereits fünf voneinander unabhängige Experten in der Lage sind, 75% der insgesamt auffindbaren Benutzbarkeitsprobleme aufzudecken. Die Heuristische Evaluation einer ersten Systemrealisierung setzt voraus, dass bereits grobe Ziele der Systementwicklung spezifiziert und implementiert wurden. Sollen nicht allein visuelle Merkmale einer Benutzungsoberfläche getestet werden, ist eine rudimentäre Applikation nötig, die eine Interaktion mit dem System erlaubt.

Tab. 3: Usability-Heuristiken nach Nielsen (1994)

<b>Heuristik</b>	<b>Beschreibung</b>
<i>Einfacher und natürlicher Dialog</i>	Dialoge sollen keine irrelevanten oder selten benötigten Informationen enthalten. Alle Informationen sollen in einer natürlichen und logischen Ordnung erscheinen.
<i>Verwendung der Sprache des Benutzers</i>	Der Dialog soll in Worten, Phrasen und Konzepten ausgedrückt sein, die dem Benutzer vertraut sind und nicht in systemorientierter Terminologie.
<i>Minimierung der Gedächtnisbeanspruchung beim Benutzer</i>	Der Benutzer soll sich nicht bei der Bearbeitung der aktuellen Aufgabe an Details vergangener Schritte erinnern müssen. Bedienhinweise sollen jederzeit einfach erreichbar sein.
<i>Konsistenz</i>	Benutzer sollen nicht durch verschiedene Bezeichnungen, Situationen oder Aktionen, die den gleichen Gegenstand meinen, verwirrt werden.
<i>Rückmeldungen des Systems</i>	Das System soll dem Benutzer ständig zurückmelden, in welchem Zustand es sich befindet bzw. welche Prozesse im Moment verarbeitet werden.
<i>Klar markierte Ausstiegspunkte</i>	Falls Systemfunktionen fälschlicherweise ausgewählt wurden, wird ein klar markierten „Notausstieg“ benötigt.
<i>Shortcuts für erfahrene Benutzer</i>	Hat der Benutzer den Status des Anfängers verlassen, sollen ihm Möglichkeiten zur schnelleren Aufgabenbewältigung angeboten werden.
<i>Gute Fehlermeldungen</i>	Fehlermeldungen sollen durch verständlichen Klartext, nicht durch kryptische Fehlercodes angezeigt werden.
<i>Fehlervermeidung durch das System</i>	Besser als gute Fehlermeldungen im Fehlerfall sind Funktionen des Systems, die das Auftreten von Fehlern vermeiden.
<i>Hilfe und Dokumentation</i>	Alle Informationen über das System und seine Funktionen sollen leicht aufzusuchen und kontextsensitiv bezüglich der eben ausgeführten Aufgabe sein sowie konkrete, effiziente Ausführungsschritte enthalten.

Die Methode des Lauten Denkens (vgl. Dix et al. 1995) ist bei aller Problematik bzgl. der Beeinflussung des untersuchten Gegenstandes die einzige Methode, mit der aktuelle Bewusstseinsinhalte erfasst werden können. Sie macht es möglich, in Entscheidungs- und Problemlösungssituationen aktuelle kognitive, affektive und evaluative Prozesse zu erheben (Wirth, Plasa & Schubert 1999). Bei der Mensch-Rechner-Interaktion werden Interaktionsfolgen sequentiell abgearbeitet und Ziele werden erreicht, indem nacheinander kleinschrittige Aktionen ausgelöst werden, die strukturiert zu kommentieren sind. Deshalb wird die Methode des Lauten Denkens in diesem Bereich bevorzugt eingesetzt. Der Anwender wird gebeten, seine Gedanken, die er während der Interaktion mit dem System hat, zu äußern: sein gegenwärtiges Ziel, seine Vorstellung über die aktuell im System ablaufenden Prozesse, seine Verwunderung über unerwartete Systemreaktionen und ähnliches. Die Aktionen des Anwenders werden üblicherweise mit seinen Äußerungen auf Video aufgezeichnet. Dabei vermögen es Experten, häufig elaboriertere Aussagen zu treffen und sind weniger gehemmt sich zu äußern als ungeübtere Personen. Mit ihnen ist auch eine kooperative Bewertung (Monk et al. 1993) möglich, in der sie neben der Erläuterung ihrer eigenen Aktivitäten auch bewertende Kommentare abgeben. Das mit der Methode häufig diskutierte Problem der Beeinflussung des Verhaltens der



Testteilnehmer durch das Verbalisieren kann bei *Usability*-Experten als relativ gering eingestuft werden. Sie sind hochgeübt darin, neue Systeme zu explorieren und parallel dazu mindestens personenintern Bewertungen vornehmen. Sie bietet neben den durch die Heuristische Evaluation erfassten generellen Einschätzung weitere Informationen, wann im konkreten Interaktionsverlauf unerwartete oder problematische Zustände auftraten.

Eine für die Versuchsteilnehmer komfortable Methode, ihre Gedanken zum System zu äußern, ist das fokussierte Interview (Merton & Kendall 1979) im Anschluss an die Interaktion mit dem System. Anhand eines Leitfadens werden die Benutzer gebeten, ihre Interpretation zu wesentlichen, im Leitfaden definierten Aspekten des Systems mitzuteilen. Der Interviewer hat dabei vier Kriterien zu berücksichtigen: den Interviewten nicht zu beeinflussen, die Äußerungen mit spezifischen Situationen der vorangegangenen Ereignisse während der Systeminteraktion zu verknüpfen, einen möglichst breiten Bezugsrahmen zu erfassen und den Interviewten zu persönlichen, wertenden Aussagen anzuregen. Diese Einschätzungen können neben experimentell erhobenen Daten häufig zusätzliche spezifische Informationen über Probleme bei der Interaktion mit dem System hervorbringen, die zur Interpretation von quantitativen Daten hilfreich sein können. Die Möglichkeit sich im Anschluss an die experimentellen Untersuchungen zur Systemevaluation zu äußern, wird von Teilnehmern üblicherweise sehr positiv aufgenommen. Es unterstreicht die ihnen zugebilligte Kompetenz als potentielle Nutzer des getesteten System, wichtige Informationen zur weiteren Systemgestaltung beizusteuern.

Mit Hilfe dieser Vorgehensweise können keine direkten Festlegungen zur Beseitigung der Benutzungsprobleme getroffen werden. Die Zuordnung der aufgetretenen Probleme zu einem verletzten Dialogprinzip und zu geäußerten Benutzererwartungen helfen jedoch, das Design anhand von vorliegenden Gestaltungsregeln und Erfahrungen zu überarbeiten.

### **3.4.2 Methoden zur Erfassung von Urteilen**

Eingeführt sind Fragebögen im Rahmen experimenteller Untersuchungen zur Erhebung biographischer Daten und zu für den Untersuchungskontext bedeutsamen stichprobenbeschreibenden Informationen. Darüber hinaus machen es Fragebögen mit geschlossenen und offenen Items (vgl. Bortz & Döring 1995) auch möglich, zusätzliche wertende Informationen von den Untersuchungsteilnehmern zu erhalten. Geschlossene Items erlauben eine gezielte Abfrage von Informationen mit vorgegebenen Antwortoptionen, während offene Items den Untersuchungsteilnehmern Freiraum geben für die Äußerung eigener Erfahrungen oder Bewertungen im Zusammenhang mit der experimentellen Systemevaluation (vgl. Scheele & Groeben 1988). Der Umfang der Antworten auf offen formulierte Fragen kann jedoch sehr unterschiedlich sein, z.B. aufgrund des Ermüdungsgrades der Untersuchungsteilnehmer.

In der ergonomischen Forschung sind Daten über den Beanspruchungszustand (vgl. Hacker & Richter 1984, Ulich 1994) der Person bei der Aufgabenbearbeitung von großer Bedeutung. Dafür stehen, neben häufig eingesetzten physiologischen Messverfahren, validierte Beurtei-

lungsskalen zur Verfügung, wie der NASA-TLX (Hart & Staveland 1988) oder die SEA-Skala (Eilers, Nachreiner & Hänecke 1986). Da physiologische Messverfahren sehr störanfällig gegenüber unterschiedlichen Personen- und Umgebungsfaktoren sind, wurde für die Systemevaluation auf die subjektive Beurteilung mittels eingeführter Messinstrumente zurückgegriffen. Während der NASA Task Load Index die Beanspruchung mit sechs Subskalen vor dem Hintergrund eines mehrdimensionalen Beanspruchungskonzeptes erfasst, basiert die SEA-Skala (Skala zur Erfassung subjektiv erlebter Anstrengung) auf der eindimensionalen Einschätzung der mentalen Beanspruchung. Dabei wird nach Sanders (1983) die Anstrengung (*effort*) als wesentliches Kriterium des Beanspruchungserlebens beurteilt. In einer methodenvergleichenden Untersuchung im Rahmen der Systemevaluation der in Abschnitt 5.3 dargestellten Untersuchung ergaben sich für die SEA-Skala und den NASA-TLX Gesamtscore Korrelationen von  $r=.78$  und für die NASA-TLX Subskala Anstrengung und die SEA-Skala Korrelationen von  $r=.81$  (Korrelationsmatrix siehe Anhang 1). Da der NASA-TLX einige für die Evaluation von Computersystemen weniger bedeutsame Subskalen enthält und eine mehrstufige, auf einem eingangs durchzuführenden Paarvergleich beruhende Beurteilung verlangt, wurde für die zeitkritische experimentelle Untersuchung zur summativen Bewertung (siehe 5.5) die effizientere SEA-Skala eingesetzt.

Das semantische Differential (Osgood, Suci & Tannenbaum 1957) ist eine Methode, die die konnotative Bedeutung bzw. die affektiven Qualitäten von Begriffen oder Gegenständen erfasst. Für die Evaluation eines multimodalen Systems kann das semantische Differential wesentliche Hinweise auf dessen generelle affektive Wirkung auf die potentiellen Benutzer erbringen. Die Bedeutung dieser affektiven Bewertungsebene liegt insbesondere darin begründet, dass positiv wirkende Produkte, so auch Computersysteme, die Motivation von Benutzern erhöhen, sich mit dem Produkt vertraut zu machen und es in Gebrauch zu nehmen. Für die Einschätzung von interaktiven Produkten wurde von Hassenzahl und anderen (2000) ein kontextspezifisches semantisches Differential entworfen, mit dem über 23 bipolare verbale Anker von Adjektiven das System mittels einer siebenstufigen Ratingskala beschrieben werden kann. Dabei werden die pragmatische, aufgabenbezogene und die hedonistische, aufgabenunabhängige Qualität sowie die Attraktivität erfasst. Pragmatische und hedonistische Qualität sind zwei unterschiedlich wahrgenommene Dimensionen des beurteilten Gegenstandes, die zu gleichen Anteilen die eingeschätzte Attraktivität bestimmen (Hassenzahl, im Druck).

Ein System mit positiv eingeschätzter pragmatischer Qualität unterstützt den Benutzer funktional hinsichtlich der Erreichung seiner aufgabenbezogenen Ziele. Relevante Systemattribute, die die positiven Pole der Beurteilungsskalen des semantischen Differentials bilden, sind beispielsweise „klar“, „einfach“ oder „unterstützend“. Die pragmatische Qualität bezieht sich auf die grundlegenden menschlichen Bedürfnisse nach Sicherheit, Kontrolle und Vertrauen. Die hedonistische Qualität bezieht sich auf Qualitätsaspekte, die nicht in direkter Weise mit der Aufgabenerfüllung verknüpft sind. Allerdings können die hedonistischen Eigenschaften eines Systems bzw. Produktes die Motivation, das System zu benutzen oder zu besitzen, stärken, da

sie Anregung bieten und im sozialen Kontext dazu verhelfen zu beeindrucken. Für die hedonistische Qualität bedeutsame Beurteilungsattribute sind beispielsweise „originell“, „aufregend“ oder „exklusiv“.

### **3.4.3 Experimentelle Methoden**

Die experimentelle Vorgehensweise ist für die Evaluation dann geeignet, wenn der infragestehende Gegenstand soweit eingegrenzt werden kann, dass die Wirkzusammenhänge in Hypothesen gefasst, störende Einflüsse kontrolliert und Zufallseinflüsse minimiert werden können. Konstituierende Merkmale des Experiments sind die systematische Beobachtung des Untersuchungsgegenstandes und dessen Manipulation, d.h. die systematische Variation der Versuchsbedingungen. Die so gefundenen experimentellen Effekte können dann im Zusammenhang mit den eingeführten Variationen erklärt werden. Die Hypothesen über das betrachtete Phänomen können anhand der erzeugten Daten statistisch getestet und ihre Relevanz kann somit über das Experiment hinaus abgeschätzt werden. Das Experiment wird weit öfter für die summative Evaluation eingesetzt als für die formative (Bortz & Döring 1995). Allerdings erscheint es für spezifische Entscheidungsprobleme in der Systementwicklung sinnvoll, experimentell zu prüfen, wie sich eine Designalternative auf zu definierende Zielgrößen auswirkt. Experimentelle Untersuchungen haben eingeführte Gütekriterien nach Validität, Reliabilität und Objektivität zu erfüllen (Sarris 1992). Zur Sicherung der Objektivität, d.h. die intersubjektive Gültigkeit der Resultate, ist die Experimentalsituation zu standardisieren. Dies ist für eine Systemevaluation dann unproblematisch, wenn die Testung im Labor stattfindet. Zur Absicherung der Reliabilität sind die eingesetzten Messinstrumente auf ihre Sensitivität für unterschiedliche Ausprägungen des gemessenen Merkmals zu untersuchen. Dafür stehen verschiedene Verfahren zur Verfügung (vgl. Lienert & Ratz 1994). Sofern bei der Evaluation auf geprüfte Messinstrumente, wie standardisierte Fragebögen, zurückgegriffen wird, ist die Reliabilität bekannt. Die Validität, d.h. die Aussagekraft der Ergebnisse bzgl. der theoretischen Vorannahmen, bezieht sich auf zwei Bereiche: die interne und die externe Validität. Eine Untersuchung ist dann intern valide, wenn sie eindeutig interpretierbar ist, d.h. die Ergebnisse nicht plausibel alternativ erklärt werden können. Die externe Validität ist dann hoch, wenn die Untersuchungsergebnisse verallgemeinerbar sind. Für die experimentelle Untersuchung im Anwendungskontext ist abzuwägen, wie stark die interne Validität durch die Eliminierung aller Störvariablen maximiert werden sollte, da darunter u.U. die Generalisierbarkeit auf lebensnähere Situationen leidet. Eine besonders in der Anwendungsforschung relevante Störgröße ist der Freiheitsgrad des Versuchsteilnehmers bei der Erledigung der ihm gestellten Aufgaben, der bei komplexeren Handlungen kaum vermeidbar ist. Da für die summative Evaluation eines Prototypen die Ergebnisse über die Untersuchungssituation und die untersuchte Stichprobe hinaus gültig sein soll, ist eine natürlichere Untersuchungssituation zur Erhöhung der ökologischen Validität von großer Bedeutung. Stärker fokussierte Fragestellungen während des Systementwicklungsprozesses sind u.U. auch mit einem geringen Maß an Verhaltensfreiheit für die Partizipanten und mit einer kleineren Zahl von Kriterien experimentell untersuchbar.

## **4 Gegenstand der Systementwicklung: mUltimo3D**

### **4.1 Entwicklungsziele von mUltimo3D**

#### **4.1.1 Generelle Entwicklungsziele**

Das Projekt mUltimo3D der Abteilung Interactive Media – Human Factors des Heinrich-Hertz-Institut für Nachrichtentechnik Berlin GmbH unter Leitung von Dr.-Ing. Siegmund Pastoor, in dessen Rahmen die hier vorgestellten Evaluationsuntersuchungen durchgeführt wurden, verfolgt die prototypische Entwicklung eines innovativen multimodalen Rechnerarbeitsplatzsystems mit autostereoskopischen Display. Der Gestaltungsprozess des Systems folgt dem vertikalen Prototypenmodell (Floyd 1984). Die multimodalen Interaktionskanäle, die im mUltimo3D-System angeboten werden, sollen insbesondere die Direktmanipulation von dreidimensionalen Objekten der graphischen Benutzungsoberfläche unterstützen. Auch für das mUltimo3D-System wird angestrebt, den Lernaufwand für die Nutzung der systemseitigen Interaktionsangebote zu minimieren. Die Systemfunktionalität der entwickelten Testanwendungen soll sich dem Benutzer durch den Rückgriff auf Interaktionsabläufe, die aus der zwischenmenschlichen Kommunikation vertraut sind, leicht erschließen. Die Entwicklung videobasierter kontaktfreier Interaktionstechniken und die Integration automatischer Spracherkennung und -verarbeitung dienen dazu, dass die Benutzer gleichermaßen natürlich wie komfortabel mit dem System interagieren können. Das mUltimo3D-System kann als ein Prototyp für zukünftige Arbeitsplatzsysteme betrachtet werden, in dem innovative Schnittstellenkonzepte mit herkömmlichen Gestaltungslösungen kombiniert und breite Anwendungsbereiche mit vielfältigen Applikationen abgedeckt werden können. Obwohl sich das System gegenwärtig noch nicht im Produktstatus befindet, wurde durch die Realisierung des Systems auf PC-Plattform eine für viele kommerzielle Anwendungen kompatible Plattform ausgewählt, die eine produktnahe Weiterentwicklung erlaubt.

#### **4.1.2 Spezifische Gestaltungsbereiche**

Die Vorgehensweise bei der Gestaltung des mUltimo3D-Systems bestand darin, ein proprietäres Betriebssystem VOS (*visual operating system*) zu entwickeln, das in der Lage ist, Eingaben in unterschiedlichen Modalitäten zu verarbeiten und eine für die stereoskopische Bildschirmdarstellung aufbereitete Grafik zu erzeugen. Die multimodalen Systemkomponenten wurden vorerst separat realisiert. Vorleistungen aus vorangegangenen Projekten, auf die zurückgegriffen werden konnte, betrafen insbesondere die Entwicklung des autostereoskopischen Bildschirms mit der Kopf- und Augenpositionserfassung. Im Rahmen des aktuellen Projektes wurden die Leistungsmerkmale der einzelnen Komponenten kontinuierlich bezüglich der Erkennungsrate und -zuverlässigkeit verbessert.

Die spezifischen Gestaltungsbereiche für das mUltimo3D-System umfassen das autostereoskopische Display, die Erkennung der Kopfposition, des Blickortes, der Sprache und der Ges-

tik. In der Abbildung 14 wird ein Überblick über die multimodalen Interaktionskomponenten des mUltimo3D-Systems gegeben. Für alle Interaktionsmodalitäten ist angestrebt, dass das natürliche, situationsbezogene Verhaltensrepertoire von Benutzern möglichst wenig eingeschränkt wird. Dies erfordert in der Regel die Erfassung der Verhaltensäußerungen in angemessener zeitlicher und örtlicher Auflösung und Genauigkeit innerhalb eines möglichst großen Messbereiches.

Die videobasierte Kopfpositionserkennung dient dazu, das autostereoskopische Display parallel zum Kopf auszurichten, um eine optimale 3D-Sicht auf das GUI bei größtmöglicher Kopfbewegungsfreiheit zu haben. Weiterhin speisen die Kopfpositionsdaten die Nachführautomatik der Kamera für die Augenpositionserkennung und steuern die blickwinkelabhängige Darstellung der 3D-Objekte auf dem Display. Für die Kopfpositionserkennung, die mit einer Frequenz von 50 Hz arbeitet, wird die Position beider Augen im Videobild mit benutzerspezifischen Referenzmustern verglichen. Die Augen werden im Videobild zur Kalibrierung und Rekalibrierung aufgefunden, in dem Bewegungsinformationen ausgewertet werden, die für Lidschläge typisch sind.

Die Augenpositionserkennung erfolgt kontaktfrei über eine am Displayrand angebrachte Videokamera und arbeitet nach der PC-Portierung mit einer Messfrequenz von 50 Hz. Die Messmethode basiert darauf, die individuelle, jedoch systematische Abweichung des Pupillenmittelpunktes von einem mit einer Infrarotlichtquelle erzeugten Reflex auf der Cornea zu ermitteln (vgl. auch Liu & Pastoor 1999). Aus den Augenpositionsdaten wird nach der Kalibrierung der entsprechende Blickort auf dem Display errechnet, mittels dessen dann Objekte der graphischen Benutzungsoberfläche adressiert und weitere Interaktionsschritte initialisiert werden können. Die Auswertung von Blickmustern i.S. eines Kurzzeitgedächtnisses für angeblickte oder nicht beobachtete Objekte unterstützt adaptive Funktionen in der graphischen Benutzungsoberfläche.

Die kontaktfreie Erkennung der Raumposition der Hand wird im mUltimo3D-System mit Hilfe von zwei infrarotsensitiven Kameras und der dazugehörigen Infrarotquelle in einer Frequenz von 10 Hz realisiert. Die örtliche Auflösung der Gestenerkennung ist in horizontaler Richtung höher als in vertikaler Richtung, was der Anzahl bzw. horizontalen Position der Kameras geschuldet ist. Die beiden Kameras sind von unten auf die Hand ausgerichtet, so dass besonders gut Zeige- und ikonische Gesten mit Bezug auf virtuelle Objekte der graphischen Benutzungsoberfläche erkannt werden können. Die Erkennung der Geste, die sich auf ein Objekt bezieht, erfolgt über die Segmentierung des Videobilds in die Hand und den Hintergrund, die Bildung von Handvektoren aus dem ermittelten Schwerpunkt der Hand und dem Mustervergleich mit gespeicherten Vektorbündeln.

Die sprecherunabhängige Spracherkennung basiert auf dem kommerziellen ViaVoice'98 Software Developers Kit von IBM. Der Befehlswoortschatz der Spracherkennung wurde für die verschiedenen Testanwendungen in separat zu ladenden Dateien zusammengefasst und umfasst ca. 90 Wörter pro Testanwendung, die nach Vortests mit Synonymen und verschie-

denen Sprechern und Sprecherinnen als gut erkennbar eingeschätzt wurden. Der Wortschatz enthält weiterhin gut erkennbare Synonyme, um den Benutzerinnen einen größeren Freiraum für die Sprachinteraktion einzuräumen. Die Erkennung der eingehenden Sprachbefehle beruht auf der *keyword-spotting technique*, mittels derer interaktionsrelevante Schlüsselbegriffe aus dem eingehenden Signal herausgefiltert werden, um in den Applikationen die mit ihnen verbundenen Funktionen auszulösen.

Ausgabeseitig ist im mUltimo3D-System eine kommerzielle *Text-to-Speech*-Software integriert, die Logox3-Software von G-Data. Sie ermöglicht, Rückmeldungen oder Anfragen des Systems an den Benutzer in Form gesprochener Sprache auszugeben. Die Synthese der Sprache erfolgt aus Mikrosegmenten natürlich gesprochener Sprache, wobei männliche und weibliche Stimmen ausgewählt sowie weitere Stimmerkmale, wie Tonhöhe oder Intonation, variiert werden können.

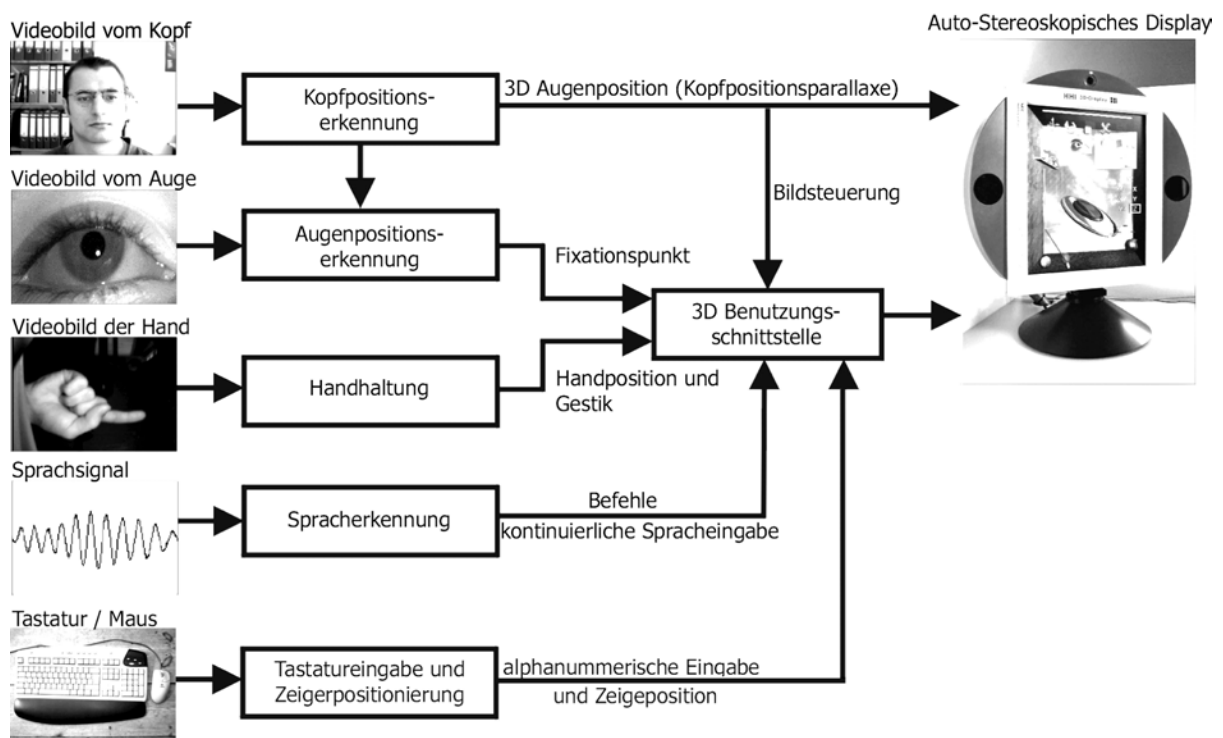


Abb. 14: Multimodale Interaktionskomponenten des mUltimo3D-Systems ©HHI

Das autostereoskopische Display verschafft dem Benutzer einen dreidimensionalen Eindruck der graphischen Benutzungsoberfläche, wobei die eingesetzte Technologie das Tragen von Brillen für die Tiefenwahrnehmung überflüssig macht. Die 3D-Visualisierung basiert auf der Linsenrastermethode, dabei werden die beiden stereoskopischen Teilbilder so getrennt, dass sie jeweils passend in das linke und rechte Auge des Benutzers projiziert werden. Das Gehirn integriert die beiden, leicht unterschiedlichen Teilbilder zu einem räumlichen visuellen Eindruck. Eine umfassende Beschreibung der autostereoskopischen Displaytechnologie findet sich u.a. in Pastoor und Skerjanc (1997). Die 3D-Visualisierung des GUI kann einerseits zur natürli-

chen Darstellung dreidimensionaler Objekte, auf die sich Manipulationen beziehen, und andererseits als Metapher für die Verwaltung von Datenbeständen genutzt werden. Eine detaillierte Darstellung der technischen Systemkomponenten wird in Liu et al. (2000) gegeben.

## **4.2 Anwendung des Evaluationsmodells auf die Entwicklung von mUltimo3D**

Die Evaluation des mUltimo3D-Systems verfolgte das Ziel, ein software-ergonomischen Anforderungen gerechtes Testsystem zu entwickeln und die in Abschnitt 3.3.2 vorgestellte Matrix, die über prototypische Aufgaben und Modalitäten aufgespannt ist, mit empirischen Daten zu füllen.

Die für die Evaluation auszuwählenden Methoden richten sich nach dem Kenntnisstand über die benutzungsbezogene Wirkung des Systems. Mit zunehmenden Wissen über die software-ergonomischen Eigenschaften des multimodalen Systems kann von explorativen Methoden zu solchen zur Testung spezifischer Hypothesen übergegangen werden.

Da bisher keine systematischen empirischen Untersuchungen am mUltimo3D-System stattgefunden hatten und wenig Informationen vorlagen, konnten noch keine Hypothesen über die Interaktion mit dem System aufgestellt und getestet werden. Um den weiteren Gestaltungsprozess möglichst kurzfristig mit umfangreichen Informationen zu Defiziten in der Gestaltung und möglichen Lösungsansätzen zu unterstützen, wurde eine explorative Expertenevaluation durchgeführt. Ein Teil der Ergebnisse aus dieser ersten empirischen Bewertung des Systems wurden aufbereitet und den Entwicklern zurückgemeldet, um in der Überarbeitung der Systemversion berücksichtigt zu werden.

Ein anderer Teil der Ergebnisse, der sich auf besondere Aspekte der Blick- und Sprachinteraktion bezog, konnte nicht in Gestaltungsempfehlungen für die Entwickler umgesetzt werden. Es bestanden mehrere, nicht ohne weitere Empirie auszuwählende Gestaltungsalternativen. Für jede der denkbaren Gestaltungsvarianten konnten theoretisch abgeleitete Hypothesen über ihre Effektivität und die mit ihnen zu erreichende Zufriedenheit der Benutzer aufgestellt werden, die dann empirisch zu prüfen waren. In Zusammenarbeit mit den Systementwicklern wurden verschiedene Gestaltungsvarianten für die Blick- und Sprachinteraktion in modalitätsspezifischen Testanwendungen realisiert, die die jeweiligen Gestaltungsvarianten der Modalität als experimentelle Variationen zu untersuchen erlaubte.

Die detaillierte experimentelle Untersuchung der Interaktionsmodalitäten in mUltimo3D diente dazu, die Spezifika der einzelnen Modalitäten herauszuarbeiten, um sie in ein multimodales Interaktionskonzept zu integrieren. Der Rückgriff auf rudimentäre Testanwendungen mit eingeschränktem Funktionsumfang konnte sehr schnell programmiert werden und erleichterte die Experimentalauswertung und -interpretation. Die Ergebnisse dieser Experimente führten zu einem konkreten Gestaltungshinweis für die multimodale Interaktion und die graphische Benutzungsoberfläche. Gemeinsame Diskussionen von Systementwicklern und -evaluatoren

verhelfen dazu, ein empirisch fundiertes Konzept für die weitere Gestaltung der Interaktion für die Testanwendungen zu erarbeiten.

Der überarbeitete Prototyp wurde zur abschließenden Bewertung der multimodalen Interaktion experimentell untersucht, um zu prüfen, wie die Entwicklungsziele von mUltimo3D erreicht wurden und wie sich die darin gestaltete multimodale Interaktion auf die untersuchten Anforderungen auswirkt. Der letztere Aspekt der summativen Evaluation verweist auf zukünftige Anwendungsdomänen für multimodale Systeme mit den getesteten Merkmalen und füllt damit auch weitere Zellen der System-Aufgaben-Eigenschaftsmatrix.

In Abbildung 15 wird der dreischrittige Evaluationsprozess von mUltimo3D schematisch dargestellt. Die ersten beiden Schritte dienen dazu, bewertende Informationen über Gestaltungsalternativen zu generieren und sind somit als formative Evaluationen zu bezeichnen, während die abschließende Untersuchung auf die summative Bewertung des überarbeiteten Systems abzielt.

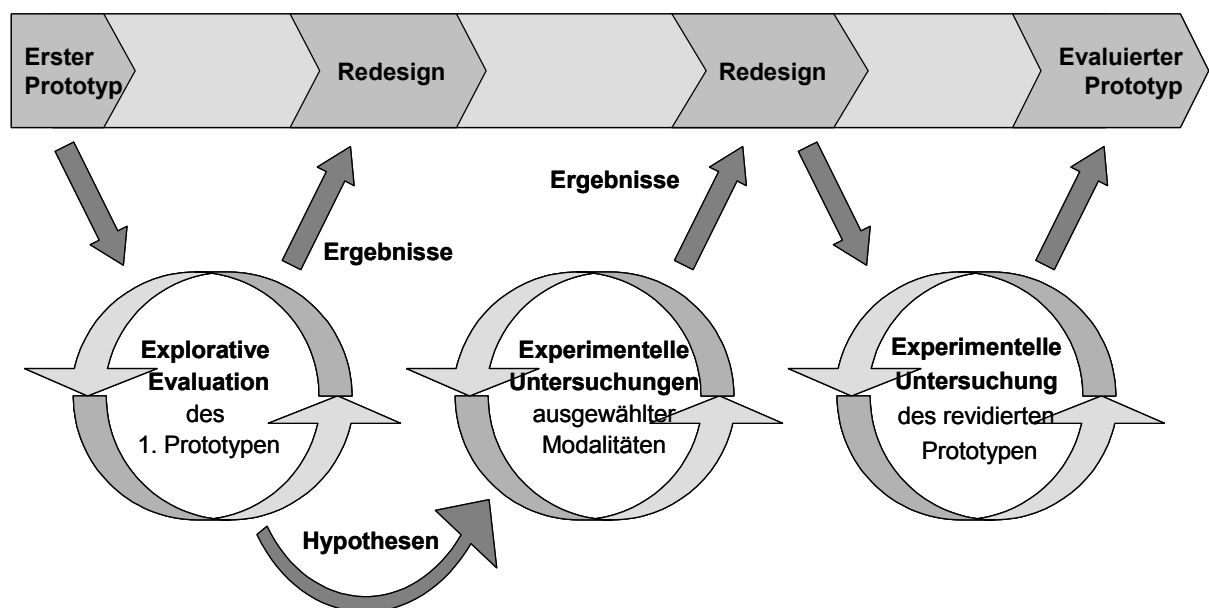


Abb. 15: Formative und summative Evaluation von mUltimo3D

Der Evaluationsprozess von mUltimo3D war auf die multimodale Interaktion mit dem System fokussiert. Dafür war es notwendig, die multimodalen Systemkomponenten soweit realisiert und mit VOS verknüpft zu haben, dass an ihnen Benutzungstests mit prototypischen Aufgaben durchführbar waren. Die Leistungsmerkmale der einzelnen multimodalen Komponenten konnten sukzessive in Abhängigkeit von den Entwicklungsarbeiten geprüft werden, inwiefern sie den Anforderungen für den folgenden Evaluationsschritt entsprachen, z.B. die Genauigkeit der Blickortbestimmung oder die Erkennungsleistung der Spracherkennung.

Da die Erkennungsleistung der automatischen Spracherkennung anfänglich sehr schwankte und für jede Person eine mindestens 45-minütige Trainingszeit erforderlich war, wurde für die



Untersuchung von Gestaltungsalternativen der Sprachinteraktion nicht mit dem System gearbeitet. Es wurde ein Simulationsexperiment, ein sogenanntes *Wizard-of-Oz-Experiment* (vgl. Dahlbäck, Jönsson & Ahrenberg 1993) durchgeführt, bei dem ein technisches System, beispielsweise eines für automatische Spracherkennung (Detmer et al. 1995), durch einen zweiten Versuchsleiter simuliert wird. Die Ergebnisse der Untersuchung können damit eindeutiger auf die experimentellen Variationen zurückgeführt werden, da unsystematische Einflüsse durch mangelnde Systemperformanz ausgeschlossen sind. Die Gestenerkennung konnte in den experimentellen Untersuchungen noch nicht berücksichtigt werden, da sie als völlig neue Systemkomponente den notwendigen Entwicklungsstand noch nicht erreicht hatte.

Die Gestaltung des Funktionsumfangs der Testapplikationen wurde zwischen den Evaluatoren und den Systementwicklern im Hinblick auf die zu untersuchenden Aufgaben abgestimmt. Die Gestaltung der grafischen Benutzungsoberflächen für den revidierten Prototyp wurde teilweise von empirischen Untersuchungen, beispielsweise einer Ikonanalyse, flankiert. Diese Vorgehensweise stellte sicher, dass die Systemelemente, die die Untersuchungsergebnisse zur multimodalen Interaktion indirekt beeinflussen könnten, nach den geltenden softwareergonomischen Gestaltungsrichtlinien ausgelegt wurden.

In Abbildung 16 ist das mUltimo3D-System mit dem Stand der Implementation dargestellt, der für die Evaluation genutzt wurde.

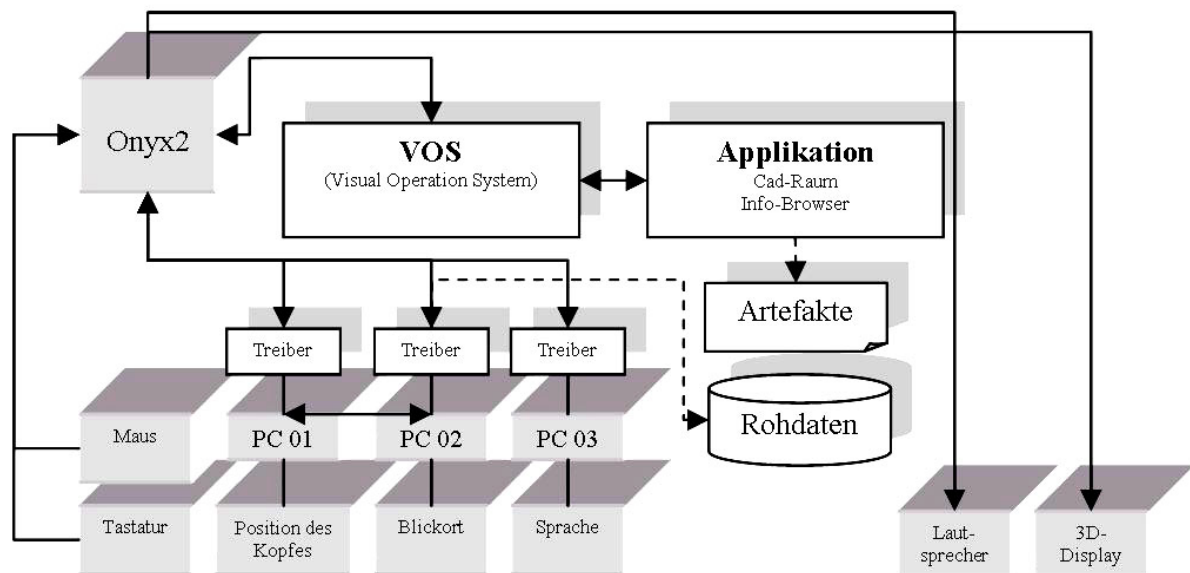


Abb. 16: Aufbau des Testsystems für die Evaluation. Graue Kästchen symbolisieren Hardware-Komponenten, weiße Kästchen symbolisieren Software-Komponenten.

Die multimodalen Komponenten für die Kopf- und Blickpositions- sowie die Spracherkennung wurden auf separaten PC's implementiert, die die Signale von den Kameras bzw. vom Mikrophon verarbeiten. Die PC's kommunizieren über spezifische Software (Treiber) mit dem

auf der Onyx2-Workstation von SGI laufenden Betriebssystem VOS. Die Messwerte der Blickpositionserkennung (Rohdaten) können für post-experimentelle Analysen gespeichert werden. Mit der Onyx2 sind auch die herkömmlichen Eingabegeräte Maus und Tastatur verknüpft. Über sie werden auch die akustischen Signale, beispielsweise zur Rückmeldung des Empfanges eines Befehles, durch die Lautsprecher ausgegeben. Für die Ausgabe von Sprache wird jedoch der PC, auf dem auch die Spracherkennung läuft, genutzt. Die Steuerung der Display-Motorik und die Berechnung der beiden Bilder für die stereoskopische Bildschirmdarstellung werden ebenfalls von der Onyx2 geleistet. Das Betriebssystem VOS unterstützte die untersuchten Testapplikationen „3D-Konstruktionsraum“ (CAD-Raum) und „Informationsbrowser“ (Info-Browser), bei deren Benutzung auch automatisch Interaktionsprotokolle erstellt und manuell die in der Interaktion erstellten Datenobjekte (Artefakte) gesichert wurden.

## **5 Empirische Untersuchung zur Evaluation von mUltimo3D**

### **5.1 Ableitung von Aufgaben für die Untersuchung**

Für die Bereiche des Informationsmanagements als auch der Produktentwicklung wird unter den Leitworten „Informationsgesellschaft“ und „Virtualisierung der Produktentwicklung“ weiterhin steigende volkswirtschaftliche Bedeutung prognostiziert. Multimodale Systeme können als besonders geeignet für diese potentiellen Anwendungsdomänen eingestuft werden, befinden sich jedoch gegenwärtig noch im Stadium prototypischer Realisierungen. Deshalb sollte das mUltimo3D-System an Testaufgaben erprobt werden, die auch zukünftig in verschiedensten Tätigkeiten wichtig und informationstechnisch zu unterstützen sind. Für die Evaluation des mUltimo3D-Systems wurden daher Testapplikationen entwickelt, die typische Aufgaben für die Suche, die Übermittlung, die Manipulation und den Abruf von Informationen aus umfangreichen Datenmengen und die Erzeugung virtueller Produkte unterstützen sollen.

Die zwei Testapplikationen, die sukzessive entwickelt wurden, waren der „3D-Informationsbrowser“ und der „3D-Konstruktionsraum“. Der 3D-Informationsbrowser unterstützt nach dem Berliner Klassifikationsschema die Suche und den Empfang von Informationen sowie die Informationsverarbeitung. Die Aufgaben sind dadurch geprägt, dass Informationen identifiziert, kodiert, kategorisiert, geordnet und spezifiziert werden müssen. Der 3D-Konstruktionsraum hilft nach dem Berliner Klassifikationsschema dabei, komplexe kontinuierliche motorische Leistungen sowie die dafür notwendigen Problemlöse- und Entscheidungsprozesse zu erbringen. Die Konstruktionsaufgaben erfordern es, komplexe räumliche Strukturen zu analysieren, ihre Relationen zu schätzen, die konstruktiven Schritte zu planen, motorisch umzusetzen und die entstehenden Strukturen mit den geplanten zu vergleichen.

Für die Untersuchung von Gestaltungsalternativen für die einzelnen Modalitäten Blick und Sprache wurden reduzierte Testapplikationen gewählt für Aufgaben, für welche die jeweiligen Einzelmodalitäten besondere Vorteile bzgl. der Leistung und der Akzeptanz bieten könnten, die jedoch nur Aspekte umfassenderer Aufgaben abbilden. Die Untersuchung der Blickinteraktion wurde an einer Aufgabe durchgeführt, die nach dem Berliner Klassifikationsschema als Identifikation von Objekten, Handlungen und Ereignissen beschrieben wird. Bei dieser Aufgabe müssen die Versuchsteilnehmer Objekte lokalisieren, identifizieren und in relevante bzw. irrelevante unterscheiden. Für die Untersuchung der Sprachinteraktion wurde eine Aufgabe entwickelt, die kommunikative Prozesse von den Probanden erfordert. Die Versuchsteilnehmerinnen mussten mit dem System kommunizieren, sie hatten Systemanfragen zu beantworten und Informationen zu übermitteln sowie weitere Bearbeitungsschritte anzuweisen.

## 5.2 Explorative Experten-Evaluation

### 5.2.1 Problemstellung der Experten-Evaluation

Die explorative Untersuchung hat zum Ziel, die Gestaltungsprobleme der experimentellen Anwendungsschnittstelle, die zu Einschränkungen der Interaktionsgüte führen, aufzudecken. Diese sollen dann im Rahmen des iterativen Designprozesses überarbeitet werden. Die Untersuchung der ersten Realisierung von zwei multimodalen Testapplikationen wurde mit den in Kapitel 3.4.1 näher beschriebenen Methoden der Heuristischen Evaluation nach Nielsen (1993) und der Methode des Lauten Denkens (vgl. Dix et al. 1995) durchgeführt. Je nach Zielstellung der Systembewertung kann das Auffinden möglichst vieler Probleme oder die Erkennung von Gestaltungsmängeln, die sich auf spezifische Kriterien wie die Effizienz auswirken, im Mittelpunkt stehen. Im Falle der Expertenevaluation von mUltimo3D sollten möglichst alle Benutzungsprobleme identifiziert werden, die mit der Gestaltung der Benutzungsoberfläche verbunden sind.

Die Untersuchung der multimodalen Interaktion mit dem Testsystem setzte voraus, dass grobe Mängel in der graphischen Benutzungsoberfläche der Experimentalanwendungen behoben sind. Nur so können Interferenzen zwischen solchen Gestaltungsfehlern und der multimodalen Interaktion, auf welche die Evaluation fokussiert ist, vermieden werden. Ein weiteres Ziel der explorativen Untersuchung bestand darin, Hypothesen für die experimentelle Untersuchung der multimodalen Interaktion zu generieren.

### 5.2.2 Durchführung der Experten-Evaluation

#### 5.2.2.1 Beschreibung der Testanwendungen

Zur Evaluation des multimodalen Experimentalsystems standen zwei Applikationen (Abbildung 17) zur Verfügung:

- der 3D-Informationsbrowser und
- der 3D-Konstruktionsraum.

Diese Testapplikationen wurden von Software-Entwicklern erstellt, die dabei auf ihre Expertise für die Gestaltung herkömmlicher Software zurückgriffen. Der Informations-Browserraum (Abbildung 17 links) enthielt eine rudimentäre Filmdatenbank mit Informationen über Filmtitel, Schauspieler, Kinos und Aufführungszeiten. Die aktuelle Seite erschien in voller Größe im Vordergrund, während die weiter zurückliegenden Seiten entsprechend gestaffelt und durch Linien (engl. *links*) verbunden dahinter lagen. Die an der linken Wand angebrachten Schaltflächen simulierten eine vom Nutzer definierbare Ablage häufig benötigter Informationen nach selbstgewählten Kategorien. Einige „Favoriten“ hingen in Form von verkleinerten Vorschaubildern an der rechten Wand. Unten links im Informationsraum veranschaulichte eine miniaturisierte Seitendarstellung, sogenannte *thumbnails*, der bereits besuch-

ten Seiten den bisher zurückgelegten Weg durch das Informationsangebot als Navigationshilfe (Bezeichnung der Funktion: engl. *history*). Sie baute sich im Laufe der Navigation auf. Alle nicht aktivierten Seiten konnten einzeln per Blicksteuerung oder per Sprache aktiviert werden. Diese „drängten“ dann in den Vordergrund, während die vorherige Seite automatisch in den Hintergrund verschwand. Die unter einem Blickwinkel verdeckten Wandflächen und Webseiten konnten bei Bedarf aufgedeckt und für den Nutzer sichtbar werden, indem man sich etwas zur Seite bewegte (dynamische Perspektive).

Der Konstruktionsraum (Abbildung 17 rechts) bot Funktionalitäten, die dem Benutzer einen realen 3D-Eindruck der von ihm erstellten und bearbeiteten Objekte vermittelten sowie durch Sprache einen direkten Zugriff auf die verfügbaren Operationen und Objekte gewährte. Gewünschte Objekte oder Operationen konnten wiederum per Blick selektiert und mit einem zusätzlichen Sprachbefehl aktiviert werden. Vielfältige geometrische Grundformen konnten aus einem in den Vordergrund klappbaren Objektregal ausgewählt und im zentralen Arbeitsbereich bearbeitet werden. Für die Manipulation der Objekte standen Verformungsoperationen zur Verfügung und Operationen zur Oberflächengestaltung. Die Objekte konnten sowohl in diskreter (Normperspektiven) als auch kontinuierlicher Veränderung der Perspektive (beliebige Drehung des Arbeitsraumes) betrachtet und bearbeitet werden.



Abb. 17: Screenshot des multimodalen 3D-Informationsbrowsers (links) und des multimodalen 3D-Konstruktionsraumes (rechts)

Das Selektieren und Aktivieren von Interface-Elementen wurde durch deren Aufhellen oder Drehen oder ein akustisches Signal markiert, was der Rückmeldung an die Benutzer dienen sollte. Insbesondere im Informationsbrowser, wo es mehrere Repräsentanten eines Objektes gab, konnten zu einem Zeitpunkt mehrere Elemente markiert sein, z.B. die zentrale Seite, der entsprechende Link und die *thumbnail*-Repräsentation des Dokumentes.

### 5.2.2.2 Auswahl der Evaluatoren und Durchführung der Evaluation

Es wurden insgesamt sieben männliche Experten eingeladen. Diese Anzahl sollte nach Nielsen (1993) sicherstellen, dass ca. 80 Prozent der auffindbaren Fehler entdeckt würden. Leider konnten keine Expertinnen für die Beurteilung des Systems gewonnen werden. Die Experten wurden aus den Professoren und Wissenschaftlichen Mitarbeitern der einschlägig arbeitenden Institute der TU Berlin und der Humboldt-Universität Berlin ausgewählt. Sie verfügten über langjährige Expertise in der Software-Ergonomie, der Gestaltung graphischer Benutzungsoberflächen und der Informatik. Vier der Evaluatoren waren informatik- und evaluationserfahrene Diplom-Psychologen, drei waren informatik- und evaluationserfahrene Diplom-Ingenieure.

Zu Beginn wurden die Evaluatoren über den Zweck der Untersuchung informiert und in das System eingeführt. Anschließend mussten die Spracherkennungssoftware, die im ersten Prototypen noch nicht sprecherunabhängig war, trainiert und der *Gaze-Tracker* auf den jeweiligen Beurteiler kalibriert werden. Danach lernten die Evaluatoren die Interaktionsmodalitäten zu benutzen. Die Heuristische Evaluation dauerte etwa zwei Stunden. In einigen Fällen verlängerte sich die Sitzung durch systembedingte Instabilitäten und die Menge der Äußerungen der Evaluatoren.

Um Benutzungsprobleme systematisch aufzudecken, wurde allen Evaluatoren in jeder der beiden Anwendungen jeweils eine Aufgabe zur Bearbeitung gegeben, zu deren Ausführung beinahe alle implementierten Funktionen benutzt werden mussten. Im 3D-Objektmanipulationsraum sollte eine Figur gebaut werden und im 3D-Informationsbrowser sollten verschiedene Informationen zu einem aktuellen Kinofilm gesammelt werden.

Bereits während der Bearbeitung wurden Äußerungen zu Gestaltungsschwierigkeiten und Vorschläge zu deren Behebung von der Versuchsleiterin protokolliert. Außerdem wurde die Bildschirmdarstellung und das Gesicht des Experten auf Video mit Ton aufgezeichnet. Nach der Aufgabenbearbeitung wurden die Heuristiken in Papierform vorgelegt, um Einschätzungen zu allen zehn Regeln einer benutzerorientierten, fehlervermeidenden und transparenten Interaktionsgestaltung zu erhalten, die in Tabelle 3 unter 3.4.1 aufgeführt und beschrieben wurden. Am Ende der Sitzung wurde den Teilnehmern eine Aufwandsentschädigung von 80,00 DM ausgezahlt. Nach Abschluss der Untersuchungen wurden die Protokolle über die Kommentare der Evaluatoren während der Interaktion mit dem mUltimo3D-System ausgewertet und die Videobänder analysiert, um auch die während der Aufgabenbearbeitung gemachten Aussagen über die Verletzung von Gestaltungsregeln und ihre Verbesserungsvorschläge in eine schriftliche Form zu bringen.

### 5.2.3 Ergebnisse der Experten-Evaluation

Den einzelnen Protokollen der Evaluatoren konnten post hoc drei Mängelbereichen für jede der beiden Anwendungen zugeordnet werden, die anschließend zusammengefasst wurden.

1. *Fehlende Funktionen* zur Systemrückmeldung, Fehlervermeidung oder Fehlerbehebung
2. *Zuverlässigkeit sowie die Konsistenz der Interaktion und die Art der Rückmeldung* entsprachen nicht den Erwartungen der Experten und die
3. *graphische Gestaltung der Oberfläche* zeigte Inkonsistenzen in der Symbolisierung von Funktionen und die farbliche Gestaltung führte zu eingeschränkter Wahrnehmbarkeit.

Zur Gewichtung eines spezifischen Mangels wurde die Anzahl der Nennungen und eine Einstufung der Bedeutung des Mangels für zukünftige experimentelle Studien herangezogen. Die Mängel, die als funktionale Einschränkungen für die weiterhin untersuchten Aufgaben und die Interaktion eingestuft wurden, waren bedeutsamer als ästhetische Mängel im GUI-Design und Mängel im wünschenswerten Funktionsumfang. Die sich daraus ergebende Rangreihe der Gestaltungserfordernisse wurde unterschiedlich bearbeitet. Zum einen konnten die Evaluatoren bereits konkrete Vorschläge zur Behebung von Mängeln machen, die für ein Redesign nutzbar waren. Zum anderen wurden die Äußerungen der Evaluatoren als Hypothesen formuliert, die weitere Untersuchungen zu Gestaltungsalternativen der Blick- und Sprachinteraktion anregten. In Tabelle 4 und 5 sind die bedeutsamsten und weiterhin gestaltungsrelevanten Aussagen für den Informationsbrowser und den Konstruktionsraum enthalten. Insgesamt wurden zu den Funktionen 113 Aussagen gemacht, zu Interaktionsproblemen gab es 104 und zum Design 133 Äußerungen. Neben den genannten Mängeln wurden auch die Gestaltungshinweise und positive Kritik berücksichtigt.

Tab. 4: Ausgewählte Ergebnisse der Experten Evaluation des Informationsbrowsers in den drei Mängelbereichen

Funktionen	Interaktion	Design
<ul style="list-style-type: none"> <li>• Neu laden / Aktualisieren fehlt</li> <li>• Mehr Schritte für Zurück-Button ermöglichen</li> <li>• Fehlererkennung und -vermeidung ist nicht unterstützt</li> <li>• Linkleiste manchmal und unvorhergesehen nicht aktivierbar</li> <li>• Navigationsbaum baut sich inhomogen auf</li> <li>• Linkleiste erscheint inhomogen: ober- und nebengeordnete Begriffe vermischt</li> <li>• Funktionsaufteilung zwischen Linkleiste und Favoritentafeln ist unklar</li> <li>• Hilfe zur Funktion einzelner Interaktionselemente fehlt</li> <li>• Zoom für Dokumentenvergrößerung fehlt</li> <li>• Lupe funktioniert nicht</li> <li>• <i>Shortcuts</i> für Verwaltungsfunktionen fehlen</li> <li>• Browser ist eher für Sites mit konstantem Material und konstanten Kategorien</li> <li>• Browser ist auch gut für Lernsoftware zur Verdeutlichung der Struktur des Lernstoffes</li> <li>• Linkliste ist gut und wichtig</li> </ul>	<ul style="list-style-type: none"> <li>• Blick springt im Navigationsbaum, bei der Linkleiste und bei den Favoriten permanent um</li> <li>• Ausgewählten Fokus halten und nicht wieder verlieren ist schwer</li> <li>• Akustische Rückmeldung über Blickauswahl stört</li> <li>• Rate der Spracherkennung zu niedrig</li> <li>• Rückmeldung für nicht erkannte Sprachbefehle fehlt</li> <li>• Aufbau der Seitenstruktur des Navigationsbaumes ist schlecht aus der eigenen Navigation nachvollziehbar</li> <li>• Zu viele Dinge sind gleichzeitig optisch markiert</li> <li>• Zu lange Rückmeldezeiten</li> <li>• Erscheinen des Navigationsbaumes wurde nicht erkannt</li> <li>• 3D nicht gut genutzt: Tiefe ohne Funktion</li> <li>• Alle Interaktionselemente direkt per Benennung der Funktion adressieren</li> <li>• Spracherkennung sollte man adäquat zu den Funktionen machen</li> <li>• Links sollten per Sprache direkt wählbar sein</li> <li>• Sprache als <i>shortcut</i> für Verwaltungsoperationen einsetzen</li> </ul>	<ul style="list-style-type: none"> <li>• Navigationsbaum: Seitenstruktur wird nicht so dargestellt wie empfunden</li> <li>• Buttons zu groß</li> <li>• Anordnung der Elemente ist uneinheitlich, zu verstreut über den Bildschirm</li> <li>• Beschriftungen und Schriften sind zu klein</li> <li>• Durchgängige Metapher fehlt</li> <li>• Dokumentenfenster: Schrägstellung ist zum Lesen ungeeignet</li> <li>• Ikons sind ohne Beschriftung nicht zu interpretieren wg. ungünstiger Metaphern</li> <li>• Startseite: bogenförmige Beschriftung ist schwer zu lesen</li> <li>• Hauptfenster: <i>Scrollbar</i> wird mit Stift verwechselt</li> <li>• Hauptfenster gerade in 2D darstellen wegen Lesbarkeit</li> <li>• Für's Kino wäre besser, Plakate oder Leinwände als Metapher zu nehmen</li> <li>• 3D besser auszunutzen: Navigationsbaum jederzeit zugreifbar in den Hintergrund verschieben</li> <li>• Zurück-Button: gut, dass es ihn gibt</li> <li>• Ausgang: gut, dass es ihn gibt</li> </ul>



Tab. 5: Ausgewählte Ergebnisse der Experten Evaluation des Konstruktionsraumes in den drei Mängelbereichen

Funktionen	Interaktion	Design
<ul style="list-style-type: none"> <li>• Rückmeldung, bevor ganze Szene gelöscht wird</li> <li>• Fehlende „Eingabe - rückgängig“-Funktion</li> <li>• Fehlende Funktion für dimensionales Skalieren von Objektgrößen</li> <li>• Verschieben von Objekten in 3D fehlt</li> <li>• Hilfefunktion fehlt</li> <li>• Kopierfunktion fehlt</li> <li>• Griffmarken an Objekten anfügen zum Skalieren durch Ziehen der Maus</li> <li>• Beim Laden von Grundobjekten Überlagerung vermeiden</li> <li>• Koordinaten des Objektes zurückmelden</li> <li>• Kritische Funktionen (z.B. Verschieben des Hintergrunds) von Maustasten weg verlegen</li> <li>• Gruppierungen ermöglichen</li> <li>• Fehlerrückmeldung fehlt</li> <li>• Fehlende Rückmeldung im Arbeitsbereich über Ebenenauswahl</li> <li>• 3D-Visualisierung ist für die Konstruktion hilfreich</li> </ul>	<ul style="list-style-type: none"> <li>• Stoppen ausgewählter Aktionen unklar</li> <li>• Interaktionssequenzen zu lang</li> <li>• Blickinteraktion schwierig, besonders bei kleinen Objekten</li> <li>• Systemantwortzeiten zu lang</li> <li>• Interaktionssyntax nicht eingängig</li> <li>• Interaktionssyntax beim Löschen verändert</li> <li>• Objekte per Sprache direkt adressierbar machen</li> <li>• Ständige akustische Auswahlbestätigung stört</li> <li>• Blickauswahl von Funktionen schlecht erkennbar</li> <li>• Interferenz mit Blick führt zu falschen Annahmen über mögliche Sprachinteraktion</li> <li>• Unklarheit über zulässige Sprachbefehle</li> <li>• Zuverlässigkeit der Blickinteraktion erhöhen</li> <li>• Zuverlässigkeit der Sprachinteraktion erhöhen</li> <li>• Hierarchisierung der Interaktionskanäle Blick und Maus nötig wg. Interferenz</li> </ul>	<ul style="list-style-type: none"> <li>• Zuviel Animation als Auswahlrückmeldung stört</li> <li>• Ikon „Alles Löschen“: leeres Gitternetz bedenklich</li> <li>• Ikon für Löschen verändert sich grafisch nicht bei Aktivierung</li> <li>• Wiedererkennen schlecht, da Ikon unbekannt</li> <li>• Interaktionsobjekte benötigen zuviel Platz</li> <li>• Ikon „Ansichten“ zu unauffällig</li> <li>• Ikon für xyz-Ebenenwahl nicht alphanumerisch visualisieren</li> <li>• Helle Interface-Objekte konfundieren mit Hintergrund</li> <li>• Ansichts- und Manipulationsfunktionen besser gruppieren</li> <li>• Bedeutung gestalterischer Elemente im Interface unklar (z.B. Balken über den Funktionen)</li> <li>• Auswahlrückmeldung über rotierende Objekte unklar</li> <li>• Interface-Metapher nicht erkennbar</li> <li>• Design zu verspielt und ablenkend</li> <li>• Farben im Symbol für xyz-Ebenenwahl und im Arbeitsbereich stimmen nicht überein</li> </ul>

#### **5.2.4 Interpretation der Ergebnisse und Konsequenzen für die Systementwicklung**

Die unter „Fehlende Funktionen“ subsummierten Mängel wurden korrigiert, indem im überarbeiteten Prototypen von mUltimo3D Funktionen hinzugefügt wurden, welche die letzte Aktion rückgängig zu machen erlaubten und außerdem wurde ein Hilfesystem über die Systemfunktionalitäten integriert.

Die unter „Zuverlässigkeit, Konsistenz und Rückmeldung der Interaktion“ genannten Probleme der Zuverlässigkeit der Interaktion führte dazu, die Blickinteraktion durch eine Erhöhung und Stabilisierung der Messrate von Blickkoordinaten durch Portierung der Blickbewegungsmesstechnik auf eine PC-Plattform zu beschleunigen. Außerdem wurden die blicksensitiven Elemente der grafischen Benutzungsschnittstelle bzgl. ihrer Größe optimiert, um den jeweiligen Blickort mit der gegebenen, in Tests erhobenen Messsystemgenauigkeit von ca. 0,6° Schwinkel sicher zu erkennen.

Für die Sprachinteraktion wurde ein sprecherunabhängiger Modus des Spracherkennungssoftware ViaVoice'98 Software Developers Kit von IBM implementiert. Es zeigte sich in der Heuristischen Evaluation, dass die Sprachinteraktion mit den Testanwendungen kommando-basiert geführt wurde und nicht notwendigerweise fließend gesprochene Sprache erkannt werden muss, wofür ein sprecherabhängiges Training der Software erforderlich ist. Zur Verbesserung der Erkennungsleistung wurden Tests mit Synonymen durchgeführt. Für alle Systemfunktionen wurde mit männlichen und weiblichen Sprechern die höchste Erkennungsrate pro Phrase erhoben. Die zuverlässigsten Phrasen wurden den Benutzern in den späteren Experimenten zur Auswahl gestellt.

Gleiche Funktionen in den beiden Testanwendungen wurden zur Erhöhung der Interaktionskonsistenz hinsichtlich ihrer Darstellung und ihrer Aktivierung angepasst. Für die Gestaltung der Rückmeldung der Blickinteraktion konnte aus der explorativen Untersuchung noch kein konkreter Gestaltungshinweis gezogen werden, deshalb wurde die unter 5.3 dargestellte experimentelle Untersuchung durchgeführt.

Die Protokolle über die bewertenden Äußerungen unterstrichen, dass die Experten auch mit der Sprachinteraktion unzufrieden waren. Es traten Unsicherheiten auf, wie das System anzusprechen sei. Außerdem wurde es als unangenehm empfunden, dass „stumme“ System aktiv anzusprechen. Unter anderem auch zur Frage, wie diese Interaktionsprobleme für diese Modalität zu lösen sind, wurde das unter 5.4 dokumentierte Experiment durchgeführt.

Die dritte Problemkategorie bezog sich auf die Gestaltung der grafischen Benutzungsoberfläche. Da die Experten die Funktionen, die die Ikonen symbolisierten, öfter nicht erschließen konnten, wurde für die Überarbeitung des Prototypen eine zweistufige Ikonanalyse durchgeführt. So sollte sicher gestellt werden, dass die verwendeten Ikonen die an sie geknüpften Funktionen möglichst gut repräsentieren. Den Ikonen wurden zusätzlich die Funktionsnamen hinzugefügt, was auch die Sprachinteraktion unterstützt, da die Benutzer auf die vorgegebenen Be-

zeichnungen zurückgreifen können (vgl. auch Revels et al. 2001). Farblich wurde der Hintergrund des GUI in der überarbeiteten Version dezenter und weniger unstet gestaltet, um die interaktiven Elemente sichtbar davon abzuheben. Die interaktiven Elemente selbst wurden ebenfalls farblich konsistenter dargestellt. Ein weiteres Problem bestand darin, dass der Aufbau der History, die die Orientierung im Informationsangebot unterstützen sollte, nicht wahrgenommen wurde. Hierfür schienen zwei unterschiedliche Hypothesen naheliegend:

1. Das Problem wurde durch die grelle grafische Gestaltung des GUI hervorgerufen, die stark als Distraktor auf die Aufmerksamkeitsverteilung wirkte.
2. Die Rückmeldung der Blickinteraktion durch das Aufleuchten der interaktiven Elemente verursachte eine Aufmerksamkeitsfokussierung, die das funktionale Gesichtsfeld einschränkte.

Zur Beantwortung dieser Frage wurde in die Untersuchung zur Gestaltung der Blickinteraktion eine Zweitaufgabe eingefügt. Da das GUI bei diesem Experiment über alle Variationen konstant gestaltet war, sollten eventuell auftretende Leistungsunterschiede in der Wahrnehmung eines peripher auftauchendes Interface-Objektes allein auf die Rückmeldevariante der Blickinteraktion zurückführbar sein.

## **5.3 Modalitätsspezifische Untersuchungen der Blickinteraktion**

### **5.3.1 Problemstellung und Hypothesen der Untersuchung der Blickinteraktion**

Die Gestaltung der blickgestützten Interaktion in Ultimo3D warf das Problem einer geeigneten Darstellung der vom System interpretierten Eingabe auf. Die Rückmeldung des Blickortes sollte die visuelle Wahrnehmung nicht beeinträchtigen bzw. zu stark fokussieren. Zur Entscheidung über eine Gestaltungsalternative für die Rückmeldung des vom System interpretierten Blickortes wurde ein Experiment durchgeführt.

Für die Gestaltung der Rückmeldung des Blickortes können sowohl Hinweise aus der software-ergonomischen als auch der kommunikationspsychologischen Forschung herangezogen werden. Die Entscheidung für eine aus den beiden Perspektiven ableitbaren Gestaltungsvariante soll auf der Basis der effizientesten Aufgabenbearbeitung, der geringsten subjektiv eingeschätzten Beanspruchung und der höchsten subjektiven Bewertung durch die Versuchsteilnehmer erfolgen.

Die software-ergonomische Forschung hat *Rückmeldung* bzw. *Feedback* als ein wichtiges Gestaltungskriterium für die Gestaltung von *graphical user interfaces* (GUIs) identifiziert (vgl. Mayhew 1992, Nielsen 1993, Norman 1988, Oppermann et al. 1992, Preece et al. 1994, Shneiderman 1992, Smith & Mosier 1986, Wandmacher 1993). Meist wird allgemein die Rückmeldung des aktuellen Systemzustandes und der vom Benutzer ausgelösten Aktionen

gefordert. Die Rückmeldung eines interaktiven Systems dient allgemein der „Sichtbarkeit des Systemzustandes“ (Nielsen 1993), spezifischer der Darstellung der „Ansprechensitivität“ (engl. *responsiveness*) (Mayhew 1992). Wandmacher (1993) und Oppermann et al. (1992) stellen die Verbindung zur ISO-Norm 9241-10 her, indem sie die Rückmeldung unter dem Normkriterium Selbstbeschreibungsfähigkeit diskutieren. Für die Entwicklung der Gestaltungsvarianten lassen sich die Gestaltungsrichtlinien von Smith und Mosier (1986) heranziehen, die vorschlagen, einen beweglichen Cursor mit klar erkennbaren visuellen Eigenschaften zu benutzen, um Positionseingaben zurückzumelden. Für die Maus ist daher normkonform im GUI in Abhängigkeit von ihrem jeweiligen Funktionsmodus ständig eine grafische Repräsentation sichtbar, beispielsweise ein Pfeil. Wie aber sollten in einem multimodalen System die verschiedenen Eingabegeräte bzw. -kanäle im GUI symbolisiert werden? Sollten neben dem Mauscursor auch Blickcursor in der grafischen Benutzungsoberfläche integriert werden?

Unter der Prämisse, dass die eingeführte software-ergonomischen Norm für diesen Anwendungsfall heranzuziehen ist, lassen sich die folgenden Hypothesen aufstellen:

Hypothese 1: Die Blickpositionsrückmeldung mittels eines kontinuierlichen Blickcursors unterstützt die höchste Leistung.

Hypothese 1a: Die Blickpositionsrückmeldung mittels eines kontinuierlichen Blickcursors unterstützt die kürzesten Reaktionszeiten.

Hypothese 1b: Die Blickpositionsrückmeldung mittels eines kontinuierlichen Blickcursors unterstützt die geringste Fehlerzahl.

Hypothese 2: Die Blickpositionsrückmeldung mittels eines kontinuierlichen Blickcursors unterstützt die geringste subjektiv eingeschätzten Beanspruchung.

Die Normen sehen auch vor, eine diskrete, visuelle Hervorhebung am interaktiven Objekt anzubringen, wenn sich die Positionseingabe nur auf die Auswahl von auf dem Bildschirm dargestellten Objekten beschränkt. Auch dies ist eine mögliche Gestaltungsvariante für die Blickrückmeldung, für die dann folgende Hypothesen gelten:

Hypothese 4: Die Blickpositionsrückmeldung mittels einer optischen Hervorhebung unterstützt die höchste Leistung.

Hypothese 4a: Die Blickpositionsrückmeldung mittels einer optischen Hervorhebung des interaktiven Objektes im GUI unterstützt die kürzesten Reaktionszeiten.

Hypothese 4b: Die Blickpositionsrückmeldung mittels einer optischen Hervorhebung des interaktiven Objektes im GUI unterstützt die geringste Fehlerzahl.

Hypothese 5: Die Blickpositionsrückmeldung mittels einer optischen Hervorhebung des interaktiven Objektes im GUI unterstützt die geringste subjektiv eingeschätzten Beanspruchung.

Die Betrachtung kommunikationspsychologischer Befunde zur Blickinteraktion im natürlichen Kontext lassen eine andere Gestaltungsvariante sinnvoll erscheinen. Natürlichkeit bedeutet bei der blickgestützten Interaktion, dass durch den Blick des Betrachters keine Veränderung am betrachteten Objekt ausgelöst wird. Der Blick auf ein Objekt gilt als Voraussetzung, gezielte Handlungen am Objekt ausführen zu können (Neumann 1992), manipuliert Gegenstände jedoch nicht eigenständig. Nicht jede Veränderung des Blickortes steht unter der bewussten Kontrolle des Betrachtenden, da Merkmale der Umgebung wie Bewegung oder Auftauchen neuer Objekte unwillkürlich Blickzuwendungen auslösen können. Dadurch wird blickgesteuerte Mensch-Computer-Interaktion erschwert und führt zu Problemen wie dem vielzitierten *Midas-Touch*-Problem (Jacob 1995). In der zwischenmenschlichen Kommunikation wird der Blick vom Interaktionspartner als Indikator für den Ort der visuellen Aufmerksamkeit interpretiert, der jedoch, verglichen mit einem kontaktherstellenden Zeigeelement, nur vage umschrieben ist. Der Blick wird in natürlichen Kommunikationssituationen als zusätzliche, aber für die Ausrichtung der sozialen Aufmerksamkeit wichtige nonverbale Verhaltensäußerung interpretiert, die üblicherweise von anderen verbalen und nonverbalen Ausdrucksformen begleitet wird (Langton 2000). Die Rückmeldung des Blickes durch eine Veränderung des betrachteten Objektes findet also keine Parallele in der physikalischen bzw. sozialen Welt. Die Wirksamkeit der Blickinteraktion zeigt sich im sozialen Kontext am Verhalten des Gegenübers.

Diese Befunde sprechen dafür, die Blickposition nicht zurückzumelden. Wenn diese Gestaltungsvariante die günstigste ist, sollten sich folgende Hypothesen bestätigen:

Hypothese 6: Keine grafische Blickpositionsrückmeldung unterstützt die höchste Leistung.

Hypothese 6a: Keine grafische Blickpositionsrückmeldung unterstützt die kürzesten Reaktionszeiten.

Hypothese 6b: Keine grafische Blickpositionsrückmeldung unterstützt die geringste Fehlerzahl.

Hypothese 7: Keine grafische Blickpositionsrückmeldung unterstützt die geringste subjektiv eingeschätzten Beanspruchung.

Weiterhin wird angenommen, dass die Rückmeldungsvariante, in welcher die besten Leistungen und die geringsten Beanspruchungen auftreten, von den Untersuchungsteilnehmerinnen auch subjektiv bevorzugt wird.

### 5.3.2 Versuchsplan und Aufgabe für die Untersuchung der Blickinteraktion

Das Experiment baute auf einem einfaktoriellen Versuchsplan mit Messwiederholung auf drei vollständig permutierten Faktorstufen auf. Der Faktor wird mit Blickpositionsrückmeldung bezeichnet und umfasst drei Faktorstufen (vgl. Tabelle 6):

1. Der Blickort wird kontinuierlich durch einen Blickcursor auf dem Display angezeigt (Cursorbedingung) (siehe Abbildung 18 links),
2. Ein interaktives Interface-Element wird optisch hervorgehoben, sobald das System den Blick als darauf ausgerichtet interpretiert (Highlight-Bedingung) (siehe Abbildung 18 mitte),
3. der Blickort wird nicht angezeigt, er wird nur systemintern zur Zuweisung eines Tastendrucks zum angeblickten interaktiven Interface-Element genutzt (Bedingung ohne Blickrückmeldung) (siehe Abbildung 18 rechts).

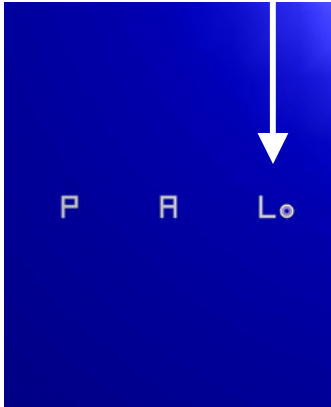
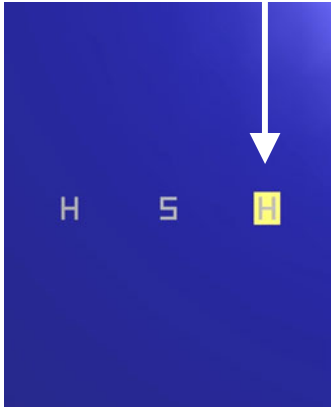
Beschreibung	Cursorbedingung: Berechneter Blickort wird durch die runde Scheibe kontinuierlich rückgemeldet	Highlight-Bedingung: Blickinteraktion mit Rückmeldung über das Aufleuchten der spezifischen Position	Bedingung ohne Blickrückmeldung: Berechneter Blickort wird nicht rückgemeldet
Abbildung			

Abb. 18: Screenshots der Untersuchungsbedingungen

Als Kontrollbedingung wurde ein Versuchsdurchlauf ohne Blickinteraktion durchgeführt, der die Auswirkung der blickgestützten Interaktion auf die gemessenen Werte generell darstellen sollte. Wie in den Ergebnissen der explorativen Untersuchung berichtet, fanden sich Hinweise, dass selbst größere Veränderungen des GUI im peripheren Sichtfeld bei der blickgestützten Interaktion nicht wahrgenommen wurden. Diesem Problem sollte unter experimentellen Bedingungen mit konstant gestaltetem GUI in allen Blickinteraktionsbedingungen und der Kontrollbedingung nachgegangen werden. Deshalb wurde zur Abschätzung der Einschrän-

kung des visuellen Feldes bei den verschiedenen Rückmeldebedingungen neben der Hauptaufgabe ein zu quittierender Stimulus als Zweitaufgabe eingefügt (siehe Abbildung 18 rechts).

Tab. 6: *Versuchsplan für die Untersuchung zur Blickrückmeldung*

Versuchsbedingungen	Cursor	Highlight	Ohne Blickrückmeldung	Kontrollbedingung
Stichprobe	S <sub>1</sub>	S <sub>1</sub>	S <sub>1</sub>	S <sub>1</sub>

Als abhängige Variablen wurden gemessen:

1. die Leistung in der Erstaufgabe (Reaktionszeiten, falsche Alarmerne [false alarms] und übersehene Stimuli [misses]),
2. die Leistung in der Zweitaufgabe (Reaktionszeiten, falsche Alarmerne [false alarms] und übersehene Stimuli [misses]),
3. die subjektiv empfundene mentale Beanspruchung (NASA-TLX dt. Version, SEA-Skala) und
4. die subjektive Präferenz für eine der Bedingungen.

In jeder der vier Versuchsbedingungen mit mehreren Durchgängen war die Aufgabenstellung grundsätzlich gleich. Auf blauem Bildschirmhintergrund wechselte an drei festen Positionen jeweils ein Buchstabe aus der Menge [T,A,C,U,F,E,H,P,S,L] mit annähernd konstanter Frequenz alle 1,7 Sekunden. Die erscheinenden Buchstaben wurden zufällig aus der Menge der zehn Buchstaben gezogen, wobei mindestens jeder zehnte Buchstabe ein Zielbuchstabe war. Die Erstaufgabe der Versuchsperson bestand darin, einen auditiv vorgegebenen Zielbuchstaben („H wie Heinrich“) auf allen drei Positionen zu stoppen. Wurde in den Blickinteraktionsbedingungen auf einer Buchstabenposition ein falscher Buchstabe angehalten, dann blieb dieser solange stehen, bis die Probandin ihn durch nochmaliges Anschauen und Tastendruck wieder zum Wechseln brachte. Als Zweitaufgabe tauchte innerhalb eines zeitlich begrenzten Intervalls zufällig ein zum momentanen Blickort peripherer Stimulus auf, der mit einer definierten Taste zu quittieren war. Die Verweildauer der peripheren Stimuli betrug zwei Sekunden. Der periphere Stimulus hatte einen Durchmesser von 11 Pixel. Ein Pixel entspricht 0,03° bei der gegebenen Bildschirmauflösung von 1280x1024 Pixel, einer Bilddiagonale von 20,1" und einem mittleren Abstand zwischen Bildschirm und Benutzer von 60 cm. Der periphere Stimulus erschien auf einer zufälligen Position in einem Radius von 6,1° um den aktuellen Blickort, jedoch nicht auf oder zwischen den blicksensitiven Feldern der Buchstabenpositionen. Ein Durchgang war beendet, wenn der Proband auf allen drei Positionen den Zielbuchstaben angehalten hat.

Um einen Buchstaben auf einer spezifischen Position zu stoppen, musste die Versuchsperson in allen drei Blickinteraktionsbedingungen auf die entsprechende Position schauen und gleichzeitig eine der vereinbarten Tasten drücken. Die Buchstaben hatten eine Größe von 53x73 Pixel ( $1,6^\circ \times 2,2^\circ$ ). Die erkannte Blickposition wurde einem der Buchstaben zugeordnet, wenn diese innerhalb eines sensitiven Feldes von 246x246 Pixeln ( $7,4^\circ \times 7,4^\circ$ ) lag.

Der Blickcursor (siehe Abbildung 198 links) war ein im Durchmesser 44 Pixel großer Kreis, der kontinuierlich am berechneten Blickort der Versuchsperson auf dem Monitor dargeboten wurde. Ein Highlight (siehe Abbildung 18 mitte) hingegen war immer nur dann auf dem Bildschirm zu sehen, wenn der berechnete Blick der Versuchsperson auf ein sensitives Feld einer Buchstabenposition fiel. Dieser Buchstabe wurde dann mit einem Highlight von 81x101 Pixel ( $2,4^\circ \times 3,0^\circ$ ) gelb hinterlegt. Bei der Bedingung ohne Rückmeldung (siehe Abbildung 198 rechts) gab es keine allein blickbedingte Veränderung der Darstellung auf dem Monitor.

In der Kontrollbedingung wurden die Buchstaben nicht durch die Referenzierung des Blickes auf spezifischen Positionen angehalten, sondern sobald die vereinbarte Taste gedrückt wurde. Es wurde automatisch immer die Position angehalten, auf welcher der Zielbuchstabe zuletzt aufgetreten war. Die Zweitaufgabe blieb über die vier Bedingungen des Experimentes gleich.

### **5.3.3 Durchführung der Untersuchung der Blickinteraktion**

#### **5.3.3.1 Versuchsteilnehmer**

An der experimentellen Untersuchung nahmen achtzehn bezahlte Personen teil. Sie waren im Mittel 24,6 Jahre alt (zwischen 22 und 26 Jahre) und studierten in unterschiedlichen Fachrichtungen an der TU Berlin bzw. der Humboldt-Universität zu Berlin. Zwei Drittel der Versuchspersonen waren weiblich, ein Drittel männlich. Sie besaßen im Mittel 66 Monate Computererfahrung, 6 Versuchsteilnehmer hatten bis zu zwei Jahre, sechs weitere mindestens zehn Jahre Erfahrung mit Computern. Die Versuchsteilnehmer nutzten den Computer durchschnittlich 13 Stunden wöchentlich, acht Personen bis zu fünf Stunden wöchentlich, vier Personen 30 und mehr Stunden. Alle Probanden benutzen Schreibprogramme, 83 Prozent benutzen das Internet und E-mail, 29 Prozent spielen und 12 Prozent programmieren am Computer. Da Mehrfachnennungen möglich waren, beträgt die Summe aller Nennungen mehr als 100 Prozent. Es wurde vorausgesetzt, dass die Untersuchungsteilnehmer weder Brillen noch Kontaktlinsen trugen, um eine möglichst zuverlässige Blickbewegungsregistrierung zu gewährleisten.

#### **5.3.3.2 Versuchsablauf**

Die 18 bezahlten Versuchspersonen wurden von den Versuchsleitern über den Zweck der Untersuchung informiert und nahmen Platz. Der Stuhl wurde so eingestellt, dass die *head*-Kamera den Kopf-Schulterbereich der Versuchsperson gut erfassen konnte. Die Versuchspersonen füllten einen Fragebogen mit demografischen Angaben und zur Vorerfahrung in der Computernutzung aus. Dann wurden *head*- und *gaze-tracker* konfiguriert und kalibriert.



In der Abbildung 19 ist der Versuchsaufbau dokumentiert. Die Probandin sitzt vor dem System, mit beiden Händen auf der Tastatur, um die vereinbarten Tasten entsprechend der Instruktion zu drücken. Ihre Blickposition wird kontaktlos mit dem unter dem Monitor stehenden Blickregistrierungssystem gemessen. Die Kopfposition wird von einer Kamera, die zentral über dem Display im Rahmen des Monitors integriert ist, erfasst.



Abb. 19: Versuchsaufbau für die Untersuchung der Blickinteraktion

Die Versuchspersonen begannen immer mit der Kontrollbedingung. Die Abfolge der drei Versuchsbedingungen wurde vollständig permutiert. Die Anzahl der Durchgänge pro Version richtet sich nach der Anzahl der quitierten peripheren Stimuli der Signalentdeckungsaufgabe. Vier Reaktionen von Personen sind im Signalentdeckungsparadigma zu unterscheiden: *hit* = Erkennen eines vorhandenen Signals, *miss* = Übersehen eines vorhandenen Signals, *false alarm* = fälschliches Erkennen eines nichtvorhandenen Signals, *correct rejection* = richtige Zurückweisung bei nichtvorhandenem Signal. Damit eine ausreichende Anzahl von Messwerten für die vier Kategorien der Signalentdeckung (vgl. Green & Swets 1966) generiert werden konnte, sollten mindestens 20 periphere Stimuli quitiert werden. Dadurch wurde eine Durchgangszahl von 12 bis 18 erreicht. Die ersten drei Durchgänge wurden als Übungsdurchgänge gewertet und von den Datenanalysen, in welche die folgenden neun Durchgänge eingingen, ausgenommen. Nach Beendigung einer Variation wurde den Probanden der NASA-TLX (Hart & Staveland 1988) in deutscher Version (vgl. Unema et al. 1988, siehe Anhang 5) und der SEA-Fragebogen (Eilers, Nachreiner & Hänecke 1986) zur Einschätzung der subjektiv empfundenen Beanspruchung vorgelegt. Zuletzt wurde mittels Befragung die subjektive Präferenz für eine der Blickrückmeldungsvarianten erhoben. Die Koordinaten der berechneten Blickorte, die gedrückten Tasten und Reaktionszeiten wurden zusammen mit den Informationen über die momentanen Buchstaben auf den Buchstabenpositionen in getrennte Protokolldateien gespeichert. Die einzelnen Protokolldateien wurden automatisch personen- und variationspezifisch sortiert, zusammengefasst, geschnitten und in eine SPSS lesbare Datenmatrix exportiert.

### 5.3.4 Ergebnisse der Untersuchung der Blickinteraktion

Die Ergebnisse der varianzanalytischen Auswertung der drei Blickinteraktionsbedingungen und der Kontrollbedingung sind in Tabelle 7 dargestellt.

Tab. 7: Ergebnisse der Blickinteraktionsbedingungen und der Kontrollbedingung (Signifikanzniveau: \*\*= $p < 0,01$ ; \*= $p < 0,05$ ; n.s. = nicht signifikant)

Abhängige Variablen		Unabhängige Variablen				Faktorwirkung F-Wert (df <sub>z</sub> /df <sub>n</sub> )	Kontrast H vs C F-Wert (df <sub>z</sub> /df <sub>n</sub> )	Kontrast (H,C) vs O F-Wert (df <sub>z</sub> /df <sub>n</sub> )	Kontrast (H,C,O) vs K F-Wert (df <sub>z</sub> /df <sub>n</sub> )
		Cursor (C)	High- light (H)	Ohne Rück- mel- dung (O)	Kon- troll- bedin- gung (K)				
Erstaufgabe	Mittlere Reaktionszeit (ms)	924	926	792	649	(2,27/43,5) 90,322 **	(1/16) 0,014 n.s.	(1/16) 86,735 **	(1/16) 169,313 **
	Misses (%)	23,88	27,58	17,49	2,59	(3/48) 26,649 **	(1/16) 1,148 n.s.	(1/16) 9,079 **	(1/16) 113,964 **
	False Alarms (je Durchgang)	0,78	1,15	0,91	0,12	(2,49/38,77) 12,074 **	(1/15) 4,164 n.s.	(1/15) 0,087 n.s.	(1/15) 81,549 **
Zweitaufgabe	Mittlere Reaktionszeit (ms)	862	841	777	703	(2,32/37,04) 14,369 **	(1/16) 0,425 n.s.	(1/16) 21,440 **	(1/16) 30,290 **
	Misses (%)	9,18	9,81	4,13	1,80	1,97/31,52) 6,459 **	(1/16) 0,137 n.s.	(1/16) 12,150 **	(1/16) 6,499 *
	False Alarms (je Durchgang)	0,12	0,13	0,05	0,06	(3/46) 3,113 *	(1/14) 0,236 n.s.	(1/15) 6,275 *	(1/15) 2,769 n.s.
Subjektive Daten	NASA-TLX	4,6	4,8	3,8	3,5	(3/48) 8,121 **	(1/16) 0,935 n.s.	(1/16) 16,859 **	(1/16) 10,948 **
	SEA-Skala	112	109	78	51	(3/47) 23,395 **	(1/16) 0,120 n.s.	(1/16) 21,265 **	(1/15) 54,776 **

Niemand bevorzugte in der abschließenden Befragung die Highlight-Bedingung, 22% der Versuchsteilnehmerinnen bevorzugten die Cursor-Variante der Blickpositionsrückmeldung, 67% wollten keine Blickpositionsrückmeldung und 11% der Probanden entschieden sich für die Kontrollbedingung ohne Blickinteraktion. Die Blickinteraktion ohne Rückmeldung wird von den Probanden eindeutig bevorzugt. Dies steht in erwartungskonformer Übereinstimmung mit den Leistungs- und Beanspruchungsdaten der Blickinteraktionsvarianten.

Falsche Positivantworten (*false alarms*) waren sehr seltene Ereignisse, so dass Fehler überwiegend durch das Übersehen eines Zielreizes in der Erst- und Zweitaufgabe (*misses*) entstanden. Insgesamt erscheint die Bedeutung von *false alarms* für die Bewertung der Aufgabenerfüllung deshalb geringer.

Die Resultate der Untersuchung legen für die Blickinteraktionsbedingungen nahe, dass die Hypothesen 1 bis 5 zu verwerfen sind. Bestätigt wurde hingegen die Hypothese 6, da die höchsten Leistungen in der Bedingung ohne Blickpositionsrückmeldung erzielt wurden. In der Bedingung ohne Rückmeldung wurden die kürzesten Reaktionszeiten erreicht, wie in Hypothese 6a angenommen. Die geringste Fehlerzahl (Hypothese 6b) trat ebenfalls in der Bedingung ohne Blickrückmeldung hinsichtlich der *misses* in beiden Aufgaben und den *false alarms* in der Zweitaufgabe auf. In der Erstaufgabe gab es keinen signifikanten Unterschied in den *false alarms* über alle Blickinteraktionsbedingungen. Auch die Hypothese 7, dass die geringste subjektiv eingeschätzte Beanspruchung (NASA-TLX und SEA-Skala) in der Blickinteraktionsbedingung ohne Rückmeldung auftreten würde, konnte bestätigt werden. Die Highlight- und die Cursor-Bedingungen unterschieden sich weder in den Leistungs- noch in den subjektiven Beanspruchungsdaten. Die Cursor-Bedingung wurde allerdings von den Untersuchungsteilnehmern subjektiv gegenüber der Highlight-Bedingung bevorzugt.

Die Kontrollbedingung ist den Blickinteraktionsbedingungen in den Leistungs- und in den subjektiven Beanspruchungsdaten überlegen. Allerdings sind zumindest die Leistungsdaten nicht eindeutig zu bewerten, weil es eine bestimmte Fehlerart in der Kontrollbedingung nicht gab: das Anblicken einer falschen Position beim Vorhandensein eines Zielbuchstabens.

Die post hoc Analyse der Blickverteilungen auf dem Display unter allen Bedingungen zeigte, dass die Versuchsteilnehmerinnen im Vergleich zu der Kontrollbedingung in den Blickinteraktionsbedingungen den Blick stärker auf die möglichen Zielpositionen fokussierten (vgl. Abbildung 20).

### **5.3.5 Interpretation der Ergebnisse der Untersuchung der Blickinteraktion**

Zwischen den Rückmeldebedingungen mit Cursor oder Highlight konnten keine quantitativen Unterschiede gefunden werden. Dies lässt darauf schließen, dass sowohl die bewusste Steuerung des Blickcursors als auch des Highlights die Versuchsteilnehmer bei der Aufgabenbearbeitung zusätzlich beanspruchte. Auch zeigen sich in den Leistungsdaten (Reaktionszeit und Fehler) und den Beanspruchungsdaten für die Bedingung ohne Rückmeldung klare Vorteile. Ebenso bevorzugten zwei Drittel der Probandinnen in ihrer subjektiven Beurteilung die Bedingung ohne Rückmeldung. Die Daten sprechen dafür, dass bei ausreichender technischer Zuverlässigkeit der blickgestützten Interaktion die Benutzer besser mit einer der physikalischen Welt bzw. der zwischenmenschlichen Kommunikation ähnlicheren Gestaltungslösung arbeiten können. Wird die erwartete Reaktion nach kurzer Latenz vom System hervorgebracht, benötigen die Benutzer keine weiteren Positionsrückmeldungen ihres Blickes. Die

Blickrückmeldung kann u.U. eher unerwünschte Effekte nach sich ziehen, etwa verzögerte Blickwechsel (auch als erhöhte Sakkadenlatenz bezeichnet, vgl. Rötting 2001): Diese Annahme deckt sich auch mit den in psychologischen Untersuchungen beobachteten verkürzten Sakkadenlatenzen, wenn ein visueller Stimulus, auf den der Blick fixiert ist, gelöscht wird, bevor ein neuer Stimulus an einer anderen Position erscheint (*gap paradigm*, Saslow 1967). Für die Gestaltung der Blickinteraktion in mUltimo3D wurde daraufhin auf einen Blickcursor verzichtet.

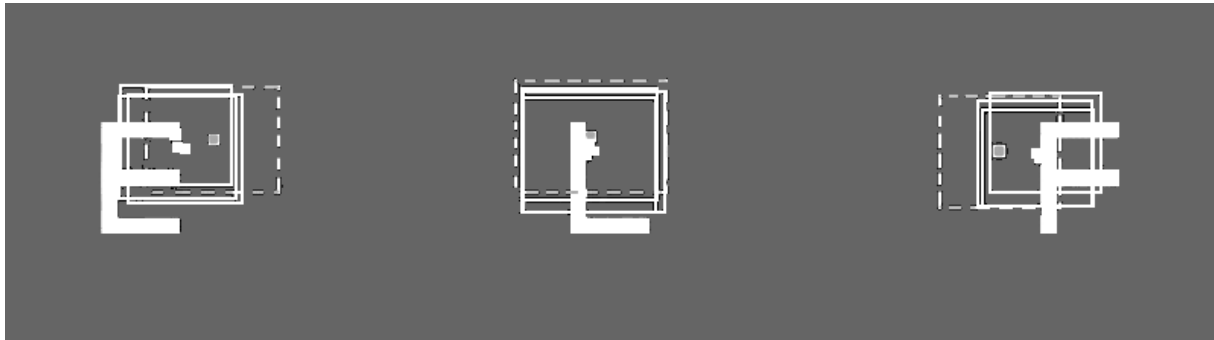


Abb. 20: Bereiche der höchsten Blickintensität über Buchstabenpositionen (Kästchen) und Schwerpunkte über Blickinteraktions- und Kontrollbedingung. Die gestrichelte Linie markiert die Bedingung ohne Blickinteraktion (Kontrollbedingung), bei der die linke und rechte Targetposition weniger korrekt fokussiert wird als über alle Blickinteraktionsbedingungen (durchgehende Linien).

Die niedriger eingeschätzte subjektive Beanspruchung in der Kontrollbedingung kann darauf zurückgeführt werden, dass allein das Bewusstmachen der „Wirksamkeit“ ihres Blickes in den Blickinteraktionsbedingungen die Versuchsteilnehmer eine erhöhte Beanspruchung wahrnehmen ließ. Dafür sprechen auch die Blickmuster der Untersuchungsteilnehmerinnen, die unter den Blickinteraktionsbedingungen korrekter auf die jeweiligen Zielpositionen ausgerichtet waren (siehe Abbildung 20), während in der Kontrollbedingung der Schwerpunkt der Blicke zur mittleren Position hin verschoben war. Das veränderte Blickverhalten könnte sich in der subjektiv eingeschätzten Beanspruchung niedergeschlagen haben.

## 5.4 Modalitätsspezifische Untersuchungen der Sprachinteraktion

### 5.4.1 Problemstellung und Hypothesen der Untersuchung der Sprachinteraktion

Die Evaluatoren wiesen in der Heuristischen Evaluation darauf hin, dass sie nicht sicher waren, wie sie mit dem System sprechen sollten und empfanden es als unangenehm, mit einem „stummen System“ sprachlich zu kommunizieren. Deshalb wurde für die Gestaltung der

Sprachinteraktion untersucht, ob man die systemseitige Sprachausgabe nutzen kann, den Benutzern die Regeln der Sprachinteraktion leicht nachvollziehbar zu verdeutlichen und sie zur Benutzung dieser Interaktionsform anzuregen.

In der Konzeption des *Ecological Interface Design* (Vicente & Rasmussen 1992) erlangte das Gestaltungskriterium „Aufforderungscharakter“ (engl. *affordance*), ein ursprünglich von Kurt Lewin 1926 geprägter Begriff zur Erklärung von zielgerichteten Handlungen, große Bedeutung. In der ökologischen Wahrnehmungstheorie von Gibson (1979) wird die *affordance* zum verbindenden Konstrukt zwischen Wahrnehmung und Handlung. Unter *affordance* wird die auffordernde Qualität eines Gegenstandes oder eines Ereignisses verstanden, die eine inhärente Eigenschaft von Items oder Ereignissen darstellt. In Abgrenzung zu Gibson, der die Umgebungseigenschaften in ihrer Funktionalität für ein mit einem bestimmten Verhaltensspektrum ausgestatteten Lebewesen fokussiert, ist der Aufforderungscharakter einer Entität der äußeren Welt bei Lewin eine temporäre Eigenschaft (Warren 1984). Er ist an die spezifische Motivlage der Person gebunden und beeinflusst deren jeweilige Handlungsauswahl. Das Konzept der *affordance* kann auf die Gestaltung von Technik allgemein übertragen werden (vgl. Norman 1993, Norman 1998), um Konstrukteure oder Entwickler darauf aufmerksam zu machen, die Funktionalität des jeweiligen Artefakts durch die Wahl der Gestaltungsvariante zu explizieren. Für Computersysteme wird häufig die graphische Gestaltung der interaktiven Objekte in der Benutzungsschnittstelle unter dem Aspekt ihrer *affordance* diskutiert (Preim 1999). Im Rahmen der Gestaltung multimodaler Mensch-Computer-Interaktion wird hier das Lewinsche Konzept dazu verwendet zu untersuchen, welche Gestaltungsvarianten einer Modalität den Aufforderungscharakter des Systems erhöhen. Der Aufforderungscharakter einer Modalität zur Mensch-Computer-Interaktion kann darin gesehen werden, dass der Computer dem Benutzer eine Modalität der Interaktion in einem spezifischen Handlungskontext anbietet. Indem der Computer Informationen nicht auf einem fixen Kanal ausgibt, sondern dem Benutzer eine für seine aktuelle Tätigkeit geeignete Interaktionsmodalität vorgibt, fordert der Computer zur Benutzung der jeweiligen Modalität auf.

Die experimentelle Untersuchung sollte dem Benutzer einen für Spracheingabe sinnvollen Aufgabenkontext anbieten. Spracheingabe scheint dann vorteilhaft zu sein, wenn neue oder an die Handlungsabsichten der Person gebundene Information in das Computersystem eingegeben werden soll. Es sollte nun geprüft werden, ob die Sprachausgabe am Rechner die Interaktion mittels Sprache erhöht, während eine visuelle Informationsausgabe über geschriebenen Text die Benutzung der Tastatur zur weiteren Interaktion nahe legt. Hieraus leitete sich die folgende Hypothese ab:

Hypothese 1: Die relative Häufigkeit der Eingabe per Sprache ist bei der Sprachausgabe des Computers größer, als bei der Textausgabe des Computers.

Für die Benutzung einer Interaktionsmodalität sind jedoch auch die Eigenschaften der zu übertragenden Information von Bedeutung (Oviatt 1999). Personen benutzen beispielsweise bei herkömmlichen Computern die Tastatur überwiegend für die Texteingabe, während die

Maus häufiger zur Auswahl einer Funktion eingesetzt wird. In einer multimodalen Mensch-Computer-Interaktion ist davon auszugehen, dass Benutzer - je nach Art der ausgeführten Teilaufgaben und damit verbundener Informationseingabe - unterschiedliche Modalitäten benutzen. Es ist anzunehmen, dass für die Eingabe neuer textueller Informationen und für Steuerungseingaben (Eingabebestätigung, Fortführung der Aufgabe) unterschiedliche Modalitäten benutzt werden. Letztere sind stark repetitiv und werden wahrscheinlich weniger gern ständig gesprochen. Auf die Überprüfung dieser Annahmen beziehen sich Hypothese 2 und 3:

Hypothese 2: Die relative Häufigkeit der Eingabe per Sprache ist höher als die relative Häufigkeit der Eingabebestätigung per Sprache.

Hypothese 3: Die relative Häufigkeit der Eingabe per Sprache ist höher als die relative Häufigkeit der Fortführung der Aufgabe per Sprache.

#### **5.4.2 Versuchsplan und Aufgabe für die Untersuchung der Sprachinteraktion**

Die experimentelle Untersuchung folgte einem zweifaktoriellen Versuchsplan: Art der Informationsausgabe des Computers und Art der einzugebenden Information. Der erste Faktor (Art der Informationsausgabe des Computers) unterteilte sich in drei Faktorstufen, bei denen den Versuchsteilnehmern drei unterschiedliche Varianten der Ausgabe des Computers angeboten wurden:

1. Der Computer forderte den Benutzer durch Sprachausgabe dazu auf, eine Eingabe vorzunehmen (Sprachinteraktionsbedingung) ,
2. Der Computer forderte den Benutzer durch Textausgabe dazu auf, eine Eingabe vorzunehmen (Textinteraktionsbedingung),
3. Der Computer forderte den Benutzer durch Text- und Sprachausgabe dazu auf, eine Eingabe vorzunehmen (kombinierte Interaktionsbedingung)

Der zweite Faktor (Art der einzugebenden Information) unterteilte sich in zwei Faktorstufen: den Datenbank-Eingaben (Namen, Jahreszahlen) und den Eingaben zur Programmsteuerung (Eingabebestätigung, Fortführung der Aufgabe).

Tab. 8: *Versuchsplan für die Untersuchung zur Sprachinteraktion. Mit  $S_i$  wird die Versuchspersonengruppe bezeichnet, gleiche Indexnummer  $i$  kennzeichnet die gleichen Personen.*

Ausgabe \ Eingabe	Datenbank-Eingaben	Eingaben zur Programmsteuerung
Sprache	$S_1$	$S_1$
Text	$S_2$	$S_2$
Sprache und Text kombiniert	$S_3$	$S_3$

Als abhängige Variable wurde die relative Häufigkeit der Eingabe per Sprache erhoben.

Die Versuchsteilnehmer hatten fehlende Eintragungen in einer fiktiven Filmdatenbank zu ergänzen. Dazu bekam jeder Proband Listen mit Angaben zu sechs Kategorien, nach denen die Datenbank strukturiert wurde: Titel des Filmes, Hauptdarsteller, Regisseur, Genre, Prämierung und Produktionsjahr. Alle Angaben über die Filme waren fiktiv, dadurch konnte das Vorwissen von Versuchsteilnehmern über Filme vernachlässigt werden. Der Titel der Filme wurde zur Kennzeichnung des Films, auf den sich die Anfrage bezog, benutzt. Das Computersystem forderte den Probanden auf, für einen Film eine fehlende Information für eine der fünf verbleibenden Kategorien einzugeben. Nach jeder Eingabe musste diese Eingabe, die bei Tastatur- ebenso wie bei Spracheingabe am Bildschirm dargestellt wurde, bestätigt werden und die Fortführung der Bearbeitung angezeigt werden. Die Eingabebestätigung konnte per Sprache mit einem „okay“ oder durch Drücken der Enter-Taste erfolgen. Die Programmsteuerung konnte per Sprache mit „weiter“ oder ebenfalls durch Drücken der Enter-Taste erfolgen. Für jede Eingabe des Probanden wurde die benutzte Modalität automatisch protokolliert. Um die experimentellen Bedingungen hinsichtlich der Systemmerkmale konstant zu halten, wurde die Spracherkennung durch einen technischen Versuchsleiter simuliert. Es wurde ein Spracherkennungssystem simuliert, dass keine 100-prozentige Erkennungsleistung aufweist. Bei definierten Teilaufgaben wurde eine „Nichtererkennung“ simuliert, die erst nach Wiederholung der Eingabe aufgehoben wurde. Damit konnte die bei realer Spracherkennungssoftware auftretende Variation in der Erkennungsleistung für unterschiedliche Sprecher verhindert und dennoch ein realistischer Eindruck der Funktionstüchtigkeit heutiger Spracherkennungssoftware vermittelt werden.

### 5.4.3 Durchführung der Untersuchung der Sprachinteraktion

#### 5.4.3.1 Versuchsteilnehmer

An der experimentellen Untersuchung nahmen die gleichen achtzehn bezahlten Personen teil, wie an der Untersuchung zur Blickinteraktion (siehe Kapitel 5.3.3.1). Sie waren im Mittel 24,6 Jahre alt (zwischen 22 und 26 Jahre) und studierten in unterschiedlichen Fachrichtungen an der TU Berlin und der Humboldt-Universität zu Berlin. Zwei Drittel der Versuchspersonen waren weiblich, ein Drittel männlich. Die Untersuchungsteilnehmer schätzten sich selbst als

mittelmäßig geübt im Umgang mit der Tastatur ein. Der überwiegende Teil der Probanden (87 Prozent) besaß keine Erfahrung mit Spracherkennung im Computer, 13 Prozent hingegen hatten eher schlechte Erfahrungen damit gemacht.

#### 5.4.3.2 Versuchsablauf

Nach der Begrüßung durch die Versuchsleiter und der Erläuterung zum Zweck der Untersuchung füllte der Proband einen Fragebogen zu demographischen Angaben und zu seiner Erfahrung im Umgang mit Computer-Anwendungen aus. Dann wurden die Untersuchungsteilnehmer durch fünf Probedurchgänge mit ihrer Aufgabe vertraut gemacht, dabei wurden sie durch den Versuchsleiter angehalten, sowohl die Tastatur- als auch die Spracheingabe auszuprobieren. Danach wurden zwei Listen mit jeweils zehn Filmen bearbeitet. Anschließend wurde die randomisiert zugewiesene Versuchsbedingung gestartet: Sprachinteraktionsbedingung, Textinteraktionsbedingung oder Sprache und Text Bedingung (vgl. Abbildung 21). Dem Probanden wurde nun die erste Liste mit zehn Filmtiteln und nach deren Abarbeitung die zweite Liste mit weiteren zehn Filmtiteln mit den dazugehörigen kategorisierten Angaben vorgelegt. Das System fragte für jeden der aufgelisteten Filmtitel nach einer fehlenden Eintragung in einer zufällig gewählten Kategorie der Filmdatenbank, die der Versuchsteilnehmer aus der Liste auszusuchen und in das System in einer von ihm gewählten Modalität (Tastatur oder Sprache) einzugeben hatte. Nach dem Ende der experimentellen Untersuchung wurde der Versuchsteilnehmer gebeten, seine persönliche Präferenz für die Informationsausgabe des Computers anzugeben. Vor der Verabschiedung erhielt jeder Versuchsteilnehmer seine Bezahlung.

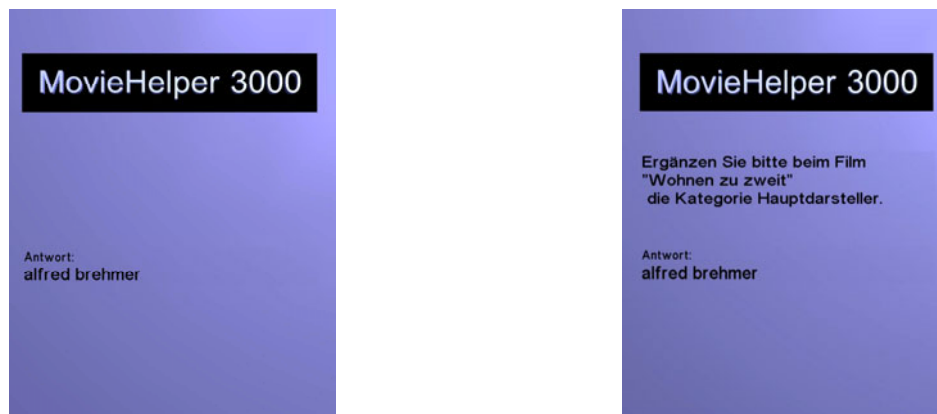


Abb. 21: Interface der fiktiven Filmdatenbank für die Sprachinteraktionsbedingung (die Aufforderung wurde nur gesprochen und nicht als Text angezeigt) (links) und für die Textbedingung bzw. für die kombinierte Bedingung (rechts)

#### 5.4.4 Ergebnisse der Untersuchung der Sprachinteraktion

Die Hypothese 1, dass die relative Häufigkeit der Eingabe per Sprache bei Sprachausgabe des Computers höher ist, als die relative Häufigkeit der Eingabe per Sprache bei Textausgabe des



Computers, konnte statistisch nicht bestätigt werden. Die in den deskriptiven Daten ersichtlichen Unterschiede wurden aufgrund der großen Streuungen in der kleinen Stichprobe nicht statistisch signifikant (Kruskal-Wallis-Test:  $\chi^2=1,779$ ;  $df= 2$ ;  $p=0,2$ ). Die Medianwerte der relativen Häufigkeit der Spracheingabe liegen jedoch in der erwarteten Rangreihe (siehe Abbildung 22): am häufigsten wurde die Sprachinteraktion bei Sprachausgabe des Computers gewählt ( $Md=0,81$ ), am seltensten wurde die Sprachinteraktion bei Textausgabe des Computers benutzt ( $Md=0,59$ ), in der neutralen Bedingung wurde die Sprachinteraktion mehr genutzt als in der Tastaturbedingung ( $Md=0,7$ ).

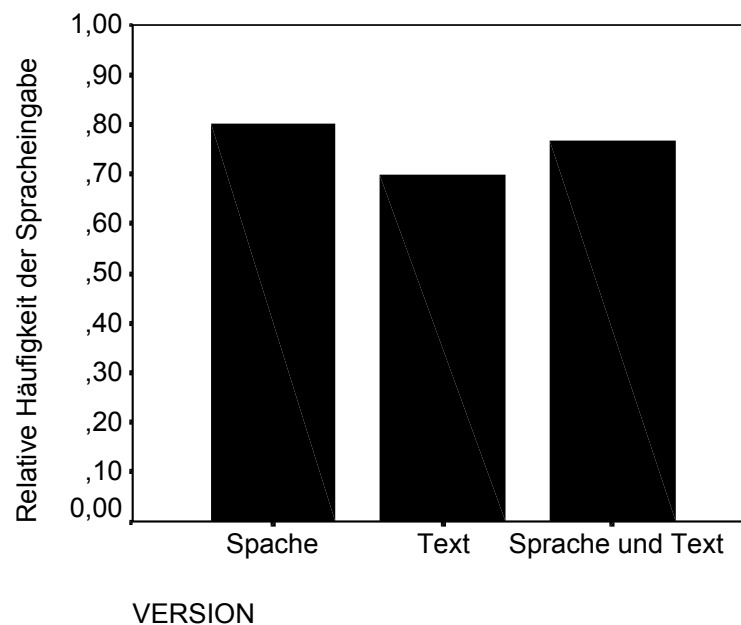


Abb. 22: Medianwerte der relativen Häufigkeit Spracheingabe über die drei Interaktionsbedingungen.

Die Hypothese 2, dass die relative Häufigkeit der Eingabe per Sprache höher ist als die relative Häufigkeit der Eingabebestätigung per Sprache, konnte bestätigt werden (Wilcoxon-Test:  $z= -3,314$ ;  $p=0,001$ ). Die Probanden benutzten für die Eingabebestätigung signifikant häufiger die Tastatur.

Die Hypothese 3, dass die relative Häufigkeit der Eingabe per Sprache höher ist als die relative Häufigkeit der Programmsteuerung per Sprache, konnte ebenfalls bestätigt werden (Wilcoxon-Test:  $z= -3,362$ ;  $p=0,001$ ). Auch für den Aufruf der nächsten Teilaufgabe benutzten die Versuchsteilnehmer signifikant häufiger die Tastatur.

Post hoc wurde ermittelt, dass die Benutzer die Modalität auch für unterschiedliche Datenarten bei der Aufgabenbearbeitung gewechselt haben. Sie sprachen Begriffe signifikant häufiger als Jahreszahlen ( $z = -2,7$ ;  $\alpha=0,007$ ). Obgleich die Spracheingabe signifikant länger dauerte als die Tastatureingabe ( $T= -7,007$ ;  $df= 17$ ;  $p= 0,000$ ), benutzten die Probanden häufiger die

Sprachinteraktion. Als bevorzugte Modalität für die Informationsausgabe durch den Computer wurden bei der Befragung von 71,4% der Probanden Text und Sprache angegeben, von 19% nur Text und von 9,5% nur Sprache.

#### **5.4.5 Interpretation der Ergebnisse**

Die Ergebnisse der experimentellen Untersuchung und der Benutzerbefragung geben Hinweise für die Gestaltung der Sprachinteraktion im Rahmen multimodaler Mensch-Computer-Interaktion. Tendenziell (deskriptive Medianvergleiche) sollten die verfügbaren Eingabemodalitäten für den Benutzer den Ausgabemodalitäten des Computers entsprechen, da dies dem Benutzer die verfügbaren Interaktionskanäle stets veranschaulicht. Für die Wahl einer Interaktionsmodalität hat sich die Funktionalität als entscheidend erwiesen: wenn die Spracherkennung technisch hinlänglich zuverlässig ist, werden einmalig einzugebende Wortgruppen (Namen) bevorzugt gesprochen, wobei als Auswahlkriterium nicht die benötigte Zeit, sondern eher der Komfort der Eingabemodalität im jeweiligen Aufgabenkontext maßgeblich zu sein scheint. In der experimentellen Aufgabe waren Informationen aus einer Liste herauszufinden, diese konnten durch lautes Vorlesen komfortabel in die Datenbank eingegeben werden. Jahreszahlen, die ihre symbolische Entsprechung auf dem Nummernblock der Tastatur haben, und repetitive Befehle, die mit einer Funktionstaste ausgelöst werden können, wurden eher per Tastatur eingegeben. Dies deckt sich auch mit Untersuchungsergebnissen von Oviatt (1997) und Oviatt und Olson (1994) zur Nutzung von unterschiedlichen Modalitäten zur Text und Zifferneingabe. Für die Ausgabe des Computers bevorzugten die Versuchsteilnehmer die kombinierte visuelle und akustische Informationsausgabe. Dies ist vermutlich darauf zurückzuführen, dass dem Benutzer dadurch ein größerer Verhaltensspielraum eingeräumt wird als bei der Textanzeige allein, z.B. ist das Wegschauen vom Bildschirm möglich. Allerdings bietet die Bildschirmdarstellung auch jederzeit den erneuten Informationsabruf an. Schulmeister (1997) verweist darauf, dass die Darbietung von Informationen auf mehreren Kanälen dazu verhelfen kann, die Kodierung der Information über mehrere Sinne zu verankern. Für die Gestaltung der multimodalen Mensch-Computer-Interaktion sollten die zu bewältigenden Aufgaben und Teilaufgaben dahingehend analysiert werden, welche Modalitäten sich als funktional bzw. komfortabel erweisen. Dabei scheint der interne Umkodierungsaufwand, der bei der Transformation einer Information in eine andere Modalität anzunehmen ist, für die Benutzer von Bedeutung zu sein.

### **5.5 Untersuchung prototypischer Aufgaben**

#### **5.5.1 Problemstellung und Hypothesen der Untersuchung prototypischer Aufgaben**

Eine summative Bewertung des überarbeiteten Prototypen mit additiver Blick- und Sprachinteraktion schloss den Designzyklus des mUltimo3D-Prototypen mit Blick- und Sprachinteraktion ab. Es war zu prüfen, wie sich ein multimodales System mit Sprach- und Blickinteraktion

neben den konventionellen Eingabegeräten Maus und Tastatur auf die Effizienz und die Beanspruchung, aber auch auf die subjektive Bewertung der Attraktivität auswirkt. Auf der Basis der experimentellen Befunde sollten dann Aussagen getroffen werden, inwiefern die Entwicklungsziele von mUltimo3D erfüllt wurden und darüber hinaus, ob sich ein System mit den gegebenen Interaktionsmodalitäten generell eignet, Aufgaben mit den spezifizierten Anforderungen zu unterstützen. Da die Multimodalität dem Benutzer Wege zur effektiveren Interaktion eröffnet, wurde davon ausgegangen, dass die Leistung in den prototypischen Aufgaben mit häufiger Nutzung der multimodalen Interaktionsmöglichkeiten generell unterstützt wird. Für jede der prototypischen Aufgaben, „Konstruktion“ und „Informationssuche“, wurden Aufgabenvarianten unterschiedlicher Schwierigkeit entwickelt. Sie sollten es möglich machen einzuschätzen, ob die erwarteten Vorteile multimodaler Interaktion sich bei zunehmend schwierigen, ressourcenzehrenden Aufgaben verändern.

Die Benutzung der Modalitäten Blick und Sprache wird sich vermutlich für die beiden Aufgaben „Informationssuche“ und „Konstruktion“ unterscheiden. Bei der Suchaufgabe dürften Informationsangebote per Blick bedeutsamer sein als bei der Konstruktion, da hier im Blickfokus nach kurzer Verweildauer zusätzliche Informationen eingeblendet werden, beispielsweise zur Bezeichnung der dargestellten Objekte. Die Sprachinteraktion sollte wichtiger für die Manipulation von Objekten sein, da über sie Aktionsmodi ohne beanspruchendere Mausbewegungen und Funktionssuche in Menüs schnell gewechselt werden können.

Für die Akzeptanz eines innovativen Systems auf Seiten potentieller Anwender ist die Anmutung des Systems eine bedeutsame Komponente. Daher wurde geprüft, wie die Untersuchungsteilnehmer die Qualität des Systems bezüglich seiner pragmatischen und seinen hedonistischen Eigenschaften (vgl. 3.4.2) wahrnehmen.

Die wahrgenommene Beanspruchung in den beiden prototypischen Aufgaben wird nicht unterschiedlich erwartet, wobei mit ansteigender Aufgabenschwierigkeit vermutlich auch die Beanspruchung leicht ansteigt. Als Kontrollvariablen wurden weiterhin die subjektiven Schwierigkeitseinstufungen für die Aufgaben einer Applikation, Computererfahrung und relevante Intelligenzmaße erhoben.

Es wird davon ausgegangen, dass das Gestaltungsziel von mUltimo3D erreicht wurde, wenn die Multimodalität von den Probanden genutzt wird und eine hohe Nutzungshäufigkeit mit besseren Leistungen einhergeht. Die Beanspruchung der Teilnehmer sollte keine überdurchschnittlichen Werte aufweisen, darüber hinaus sollte die Beurteilung des Systems bzgl. seiner pragmatischen und hedonistischen Qualität sowie seiner Attraktivität zumindest tendenziell positiv ausfallen.

Folgende Hypothesen zur multimodalen Interaktion wurden im Einzelnen geprüft:

Hypothese 1: Je häufiger die multimodale Interaktion genutzt wird, umso besser sind die Leistungen in den Konstruktionsaufgaben.

Hypothese 2: Je häufiger die multimodale Interaktion genutzt wird, umso besser sind die Leistungen in den Aufgaben zur Informationssuche.

Hypothese 3: Die Modalitäten werden in den beiden Aufgabentypen unterschiedlich genutzt.

Hypothese 3a: Sprachinteraktion wird häufiger als Blickinteraktion beim Konstruieren genutzt.

Hypothese 3b: Blickinteraktion wird häufiger als Sprachinteraktion beim Informationssuchen genutzt.

Weitere Hypothesen wurden zur Beurteilung der Qualität und Attraktivität der Testanwendungen geprüft:

Hypothese 4: Die pragmatische Qualität wird für die Testapplikation „Konstruktion“ positiv eingeschätzt.

Hypothese 5: Die pragmatische Qualität wird für die Testapplikation „Informationssuche“ positiv eingeschätzt.

Hypothese 6: Die hedonistische Qualität wird für die Testapplikation „Konstruktion“ positiv eingeschätzt.

Hypothese 7: Die hedonistische Qualität wird für die Testapplikation „Informationssuche“ positiv eingeschätzt.

Hypothese 8: Die Attraktivität wird für die Testapplikation „Konstruktion“ positiv eingeschätzt.

Hypothese 9: Die Attraktivität wird für die Testapplikation „Informationssuche“ positiv eingeschätzt.

Auch für die empfundene Anstrengung wurde eine spezifische Hypothese aufgestellt:

Hypothese 10: Die Anstrengung wird durchschnittlich geringer als „stark anstrengend“ (< 8 auf der SEA-Skala) eingeschätzt.

## 5.5.2 Versuchsplan und Aufgabe der Untersuchung prototypischer Aufgaben

### 5.5.2.1 Unabhängige Variablen

Der Versuchsplan wurde zweifaktoriell mit vollständiger Messwiederholung auf beiden Faktoren und vollständiger Permutation der Untersuchungsabfolge konzipiert. Die erste, zweifach gestufte, unabhängige Variable bildeten die Aufgaben „Konstruktion“ und „Informationssuche“, die durch eine entsprechende Testapplikation zur Konstruktion und zur Informationssuche unterstützt wurden. Die zweite, dreifach gestufte, unabhängige Variable bildete die Aufgabenschwierigkeit mit den Ausprägungen „leicht“, „mittel“ und „schwer“.

Die unabhängige Variable „Aufgabe“ forderte von den Untersuchungsteilnehmern in der Stufe „Konstruktion“, dreidimensionale Figuren, die als Papierausdrucke vorgegeben waren, nachzubauen. In der Stufe „Informationssuche“ wurden die Versuchsteilnehmer gebeten, die in der Testapplikation „Informationssuche“ enthaltenen Informationen zum mUltimo3D-System zu finden und wiederzugeben.

Tab. 9: Versuchsplan für die Untersuchung prototypischer Aufgaben

UV Aufgabe	Konstruktion			Informationssuche		
UV Schwierigkeit	Leicht	Mittel	Schwer	Leicht	Mittel	Schwer
Stichprobe	S <sub>1</sub>	S <sub>1</sub>	S <sub>1</sub>	S <sub>1</sub>	S <sub>1</sub>	S <sub>1</sub>

Die Aufgabenschwierigkeit wurde bei der Aufgabe „Konstruktion“ über die Anzahl zu bearbeitender Grundobjekte operationalisiert. Es wurden jeweils drei Stufen der Aufgabenschwierigkeit definiert (siehe Abbildung 23): leicht (7 Objekte), mittel (15 Objekte) und schwierige Aufgabe (30 Objekte).



Abb. 23: Leichte (links), mittlere (Mitte) und schwere Konstruktionsaufgabe (rechts)

In der Testapplikation „Informationssuche“ wurde die Aufgabenschwierigkeit über die für die Aufgabe relevante Grundmenge verfügbarer Informationen operationalisiert. Dabei beinhaltete die leichte Aufgabe das schnellstmögliche Auffinden einer auf Papier abgebildeten Komponente, um sie zu identifizieren und ihre Zugehörigkeit zu einer Komponentengruppe zu nennen. In der mittelschweren Aufgabe wurden die Test-Benutzer gebeten, sich mit dem Aufbau des Systems vertraut zu machen, um es nachher mit vorgegebenen Kärtchen zu rekonstruieren. Die Informationsgrundmenge bildeten 35 Systemkomponenten, vier logische Verknüpfungen und sieben Gruppen. Die schwere Aufgabe bestand im Finden und Wiedergeben von allgemeinen Funktionsinformationen und den fiktiven technischen Informationen zu den 35 Komponenten (pro Komponente ca. 6 allgemeine und 6 technische Informationen).

### **5.5.2.2 Testapplikation zur Konstruktion**

Mit der Testapplikation zur Konstruktion (vgl. Abbildung 24; die im Folgenden in Klammern angegebenen Nummern verweisen auf das entsprechende Detail der Abbildung) können aus geometrischen Grundobjekten komplexere Objekte (1) erstellt werden. Von einem Objektregal lassen sich Grundobjekte wie Würfel, Kugel, Zylinder, Ring, Kegel, Hohlzylinder und Pyramide in den Arbeitsraum auswählen. Mit den Grundfunktionen Bewegen, Rotieren und Skalieren (7) können diese Grundobjekte bearbeitet werden. Einzelnen Objekten oder ganzen Objektgruppen können dann Farben und Texturen zugewiesen werden. Zur Auswahl von Objekten stehen auch Gruppierungsfunktionen zur Verfügung, mit denen alle Objekte in der Szene ausgewählt oder abgewählt oder die momentane Auswahl invertiert werden kann. Die 3D-Szene (2) kann in x- und y-Richtungen der Displayebene gedreht werden, damit sie besser zu überschauen ist (5). Mittels der Funktion zur diskreten Perspektivenauswahl können auch orthogonale Ansichten angezeigt werden (6). Die vier Hauptmenüs gruppieren die Funktionen in Verwaltungs-, Objekt-, Werkzeug- und Hilfsfunktionen (4). Ein Hauptmenü ähnelt einer Schachtel, in die eine Funktionsgruppe einsortiert ist. Die Hauptmenüs werden bei Aktivierung transparent und öffnen sich mit einer Bewegung in den Vordergrund. Dadurch erhält der Nutzer Zugriff auf die enthaltenen Funktionen. Alle Funktionen sind sowohl als Text als auch als *icon* repräsentiert und dadurch leicht zu unterscheiden. Wird eine der angezeigten Funktionen selektiert, schließt sich das Menü wieder und ausgewählte Manipulationsfunktionen werden mit der Bearbeitungsachse oder -ebene angezeigt (7). Während die meisten Funktionen aus dem Menü direkte Auswirkungen auf die Szene oder die ausgewählten Objekte haben, gibt es drei Ausnahmen. Das Objektregal, mit dem die Grundobjekte der Szene beigelegt werden, der Farbwürfel sowie der Textureditor können zusätzlich zu den Grundfunktionen benutzt werden. Diese drei Werkzeuge werden durch die Auswahl eines Menüs automatisch wieder geschlossen.

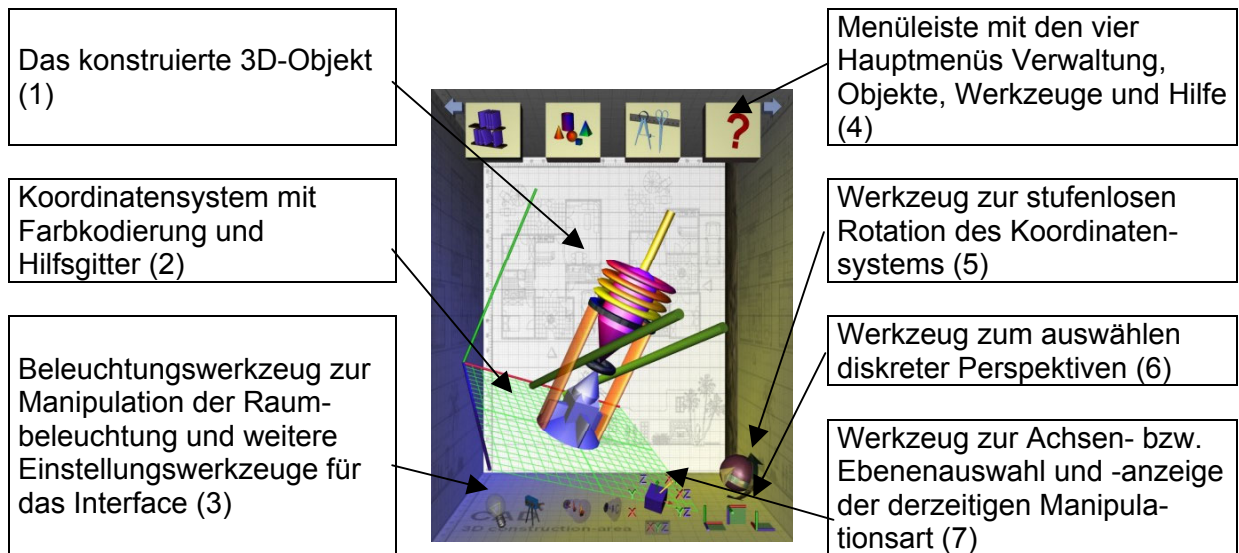


Abb. 24: Screenshot der Testapplikation „Konstruktion“

Bei der Gestaltung der multimodalen Interaktion wurde festgelegt, mit welchen Modalitäten eine spezifische Funktion ausgelöst werden kann. Dabei entstanden Gruppen von Funktionen, die drei, zwei oder keine alternativen Auslöser bekamen. Für die Anwendung „Konstruktion“ gab es grundsätzlich drei Arten der Interaktion, mit der Funktionen aktiviert werden konnten:

Maus: Funktionen wurden ausgelöst, wenn auf Elemente des Interfaces mit dem Mauszeiger gezeigt wurde und dabei die linke Maustaste gedrückt wurde.

Sprache: Funktionen wurden ausgelöst, wenn ihre zugehörigen Kommandos von der Spracherkennung erkannt wurden.

Blick & Maus: Funktionen wurden ausgelöst, wenn der interpretierte Blick mit einem blicksensitiven Element des Interfaces übereinstimmte und gleichzeitig die rechte Maustaste gedrückt wurde,

Blick & Sprache: Funktionen wurden ausgelöst, wenn der interpretierte Blick mit einem blicksensitiven Element des Interfaces übereinstimmte und die Spracherkennung gleichzeitig die Kommandos „öffnen“ oder „aktivieren“ erkannte.

Zu der Gruppe der Funktionen, die über drei Auslöser verfügten (G3), gehörten die Hauptmenüs (4). Sie konnten mittels Maus, Sprache und Blick geöffnet werden. Alle Funktionen, die in den Hauptmenüs zu finden waren, und alle Objekte konnten mit Maus und Sprache ausgewählt werden und verfügten über zwei Auslöser (G2). Eine Ausnahme bildete die Rotationskugel (5) zur kontinuierlichen Perspektivenwahl, die nur mit der Maus gesteuert werden konnte (G1).

### 5.5.2.3 Testapplikation zur Informationssuche

Mit der Testapplikation zur Informationssuche (siehe Abbildung 25; die im Folgenden in Klammern angegebenen Nummern verweisen auf das entsprechende Detail der Abbildung) lassen sich umfangreiche strukturierte Informationen ganzheitlich betrachten. Für das Experiment wurde der mUltimo3D-PC mit seinen einzelnen Komponenten und deren wechselseitigen Verknüpfungen modelliert. Damit können die technischen Komponenten des mUltimo3D-Systems als 3D-Objekte mit unterschiedlichem Vergrößerungsfaktor betrachtet werden. Der Vergrößerungsfaktor für den dargestellten Ausschnitt kann über verschiedene Interaktionen verändert werden. So können Komponenten direkt über eine dynamische Linkliste (1) angesteuert werden. Weiterhin kann eine „Lupe“ (2) auf einem Übersichtsmodell des Gesamtsystems verschoben werden. Deren Vergrößerungsfaktor kann mit einem „zoomslider“ (4) eingestellt werden. Die angezeigte Detailansicht der Szene kann frei im Raum rotiert werden. Zu den im Detailfenster (7) angezeigten Objekten können Informationstexte (5) eingeblendet werden, die eine Beschreibung des selektierten Objektes enthalten. Die Funktionsweise der Hauptmenüs entspricht den Menüs der Testapplikation zur Konstruktion.

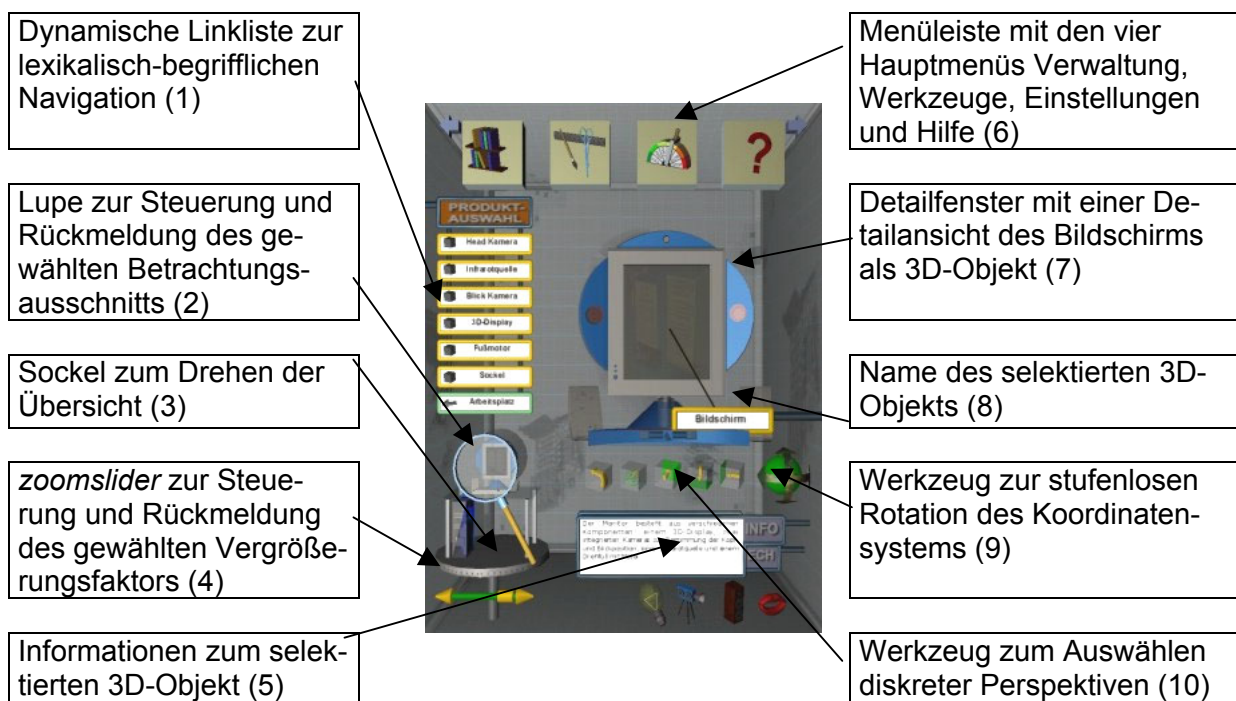


Abb. 25: Screenshot der Testapplikation „Informationssuche“

Zu den Auslösemechanismen von Funktionen, die in der Testanwendung „Konstruktion“ gelten, kommt bei der Informationssuche-Anwendung noch eine dem Blick untergliederte Auslösefunktion hinzu:

Maus: Funktionen wurden ausgelöst, wenn auf Elemente des Interfaces mit dem Mauszeiger gezeigt wurde und dabei die linke Maustaste gedrückt wurde.

Sprache: Funktionen wurden ausgelöst, wenn ihre zugehörigen Kommandos von der Spracherkennung erkannt wurden.



Blick & Maus: Funktionen wurden ausgelöst, wenn der interpretierte Blick mit einem blicksensitiven Element des Interfaces übereinstimmte und gleichzeitig die rechte Maustaste gedrückt wurde.

Blick & Sprache: Funktionen wurden ausgelöst, wenn der interpretierte Blick mit einem blicksensitiven Element des Interfaces übereinstimmte und die Spracherkennung gleichzeitig die Kommandos „öffnen“ oder „aktivieren“ erkannte.

Blick und Zeit: Funktionen wurden ausgelöst, wenn der interpretierte Blickort auf einem Element des Interfaces über eine definierte Schwellenzeit (200 ms) hinaus verweilte.

Auch in der Testapplikation zur Informationssuche gehörten die Hauptmenüs zu der Gruppe von Funktionen mit drei Auslösern (G3). Die Gruppe (G2) der Funktionen mit zwei Auslösern besteht hier nicht nur aus den Auslöserpaar Maus und Sprache wie beim CAD-Raum, sondern auch aus Maus oder Blick mit Mausklick sowie aus Maus oder Blick mit Schwellzeit. Zu der Gruppe G2 gehören alle Funktionen aus den Menüs und alle weiteren Funktionen. Die Rotationskugel, die Lupe, der *zoomslider* (4) und der Sockel (3) bilden eine Ausnahme, da sie nur mit der Maus bedient werden können. Sie bildeten die Gruppe G1.

#### **5.5.2.4 Abhängige Variablen**

Als eine abhängige Variable wurde die Leistung gemessen. In einem Rating vergaben fünf trainierte Beurteiler Punktwerte nach dem im folgenden beschriebenen Schema. Die Leistungsdaten für die Konstruktionsaufgaben wurden additiv aus der Anzahl der bearbeiteten Elemente, der Formähnlichkeit und der Position der Teile zueinander sowie der Oberflächenbearbeitung (Farbe und Textur) an den *screenshots* der konstruierten Figuren aus drei Perspektiven ermittelt (siehe Abbildung 26). Jedes zur Figur gehörige Element, das bearbeitet wurde, bekam einen Punkt. Für die Formähnlichkeit wurden pro Element null bis zwei Punkte vergeben, deren Summe dann an der Anzahl der Teile relativiert wurden. Die Position der Elemente zueinander wurde ebenso wie die richtige Zuweisung von Textur bzw. Farbe zu den Elementen mit null bis zwei Punkten bewertet.



*Abb. 26: Screenshot einer Konstruktion mit mittlerer Aufgabenschwierigkeit von Versuchsteilnehmer 10, der eine sehr gute Leistung erbrachte.*

Die Leistungsdaten für die Informationssuche und -wiedergabe für die leichte Informationsaufgabe wurde durch die Division der Antwortrichtigkeit (von 1= Komponente oder Gruppe gefunden bis 2= Komponente und Gruppe gefunden) durch die benötigte Zeit in Millisekunden berechnet, um eine gewichtete Relation zwischen Antwortrichtigkeit und der dafür benötigten Zeit zu bilden. Vollständig nicht erbrachte Leistungen wurden mit Null bewertet. Durch eine Subtraktion von einem fixen Wert (80) wurde die Skalenrichtung in Übereinstimmung mit den übrigen Leistungsdaten (je höher der Zahlenwert, umso besser die Leistung) gebracht. Die Wahl des fixen Wertes für die Skalenumkehrung ist für die weitere korrelationsstatistische Auswertung unbedeutend, da intervallskalierte Daten z-transformiert und Ordinaldaten als Ränge in die Korrelationsrechnung eingehen. Bei der Rekonstruktion der Systemstruktur wurde die Reproduktionsleistung durch die örtliche Nähe der Komponenten (mindestens ein Objekt der Gruppe in unmittelbarer Nähe angeordnet), die zu einer Gruppe gehören, betrachtet. Zusätzlich wurden die eingezeichneten logischen Verbindungen, die im Informationsraum angegeben wurden, und die logischen Gruppenumrahmungen (sofern alle Objekte korrekt in die Gruppe eingeordnet wurden) mit je einem Punkt gewertet (siehe Abbildung 27).



Abb. 27: Strukturrekonstruktion des Informationsobjektes „mUltimo3D“ von Versuchsteilnehmer 19

Bei der Reproduktionsleistung zu den allgemeinen und technischen Komponenteninformationen wurde jede richtige Reproduktion mit einem Punkt bewertet, wobei mUltimo3D-spezifische Komponenten, die nicht in einem aktuellen Standard-PC vorhanden sind, mit einem Bonuspunkt gewertet wurden.

Die multimodale Interaktionshäufigkeit war eine weitere abhängige Variable, die aus der an den objektiven Wahrscheinlichkeiten (G2 und G3) gewichteten Summe von Blick- und Sprachinteraktion gebildet wurde. Aus den Interaktions-Protokollen, in denen über die gesamte Aufgabenbearbeitungszeit für jede aktivierte Funktion die auslösende Modalität gespeichert wurde, wurden die Häufigkeiten jeder Interaktionsmodalität für die Gruppen G1 bis G3 ausgelesen. Anschließend wurden die Häufigkeiten an den Interaktionsoptionen gewichtet, d.h. die Häufigkeiten der Gruppe G2 wurden mit  $\frac{1}{2}$ , die Häufigkeiten der Gruppe G3 wurden mit  $\frac{1}{3}$  multipliziert. Aktionen, bei denen Funktionen aus der G1-Gruppe eingesetzt wurden, gingen nicht mit in die Auswertung ein, da sie keine multimodalen Interaktionsmöglichkeiten enthielten.

Die gesetzten Markierungen auf der SEA-Skala wurden mit dem Lineal in Millimetergenauigkeit ausgemessen und anschließend in der Maßeinheit Zentimeter weiter analysiert.

Weiterhin wurden die subjektiven Urteile über das System aus Benutzerperspektive erhoben. Dazu wurde das unter 3.4.2 beschriebene semantisches Differential verwendet. Für die 23 bipolaren verbalen Anker (z.B. gut - schlecht, übersichtlich - verwirrend) wurden die Urteile

jeder Person aus den siebenstufigen Ratingskalen ermittelt. Diese wurde anschließend in einem Personenwert für jede der drei Skalen zusammengefasst: wahrgenommene pragmatische Qualität, wahrgenommene hedonistische Qualität und Attraktivität des Systems.

### **5.5.3 Durchführung der Untersuchung prototypischer Aufgaben**

#### **5.5.3.1 Versuchsablauf**

Vor der Einführung in das mUltimo3D-System wurden der Fragebogen zur Person und die Tests zum stereoskopischen Sehen mit dem Zeiss Nahprüfgerät der Firma Carl Zeiss in Jena vorgelegt und unmittelbar ausgewertet. Außerdem wurden Subtests aus HAWIE (Wechsler 1991) und LPS (Horn 1962) zum räumlichen Vorstellungsvermögen und zur Arbeitsgedächtnisspanne sowie zur verbal-logischen Intelligenz vorgelegt, um interindividuell unterschiedliche Ausgangsvoraussetzungen für die Aufgabenbearbeitung zu kontrollieren. Nach der Kalibrierung des Systems erhielten die Versuchsteilnehmer eine allgemeine Einführung zum mUltimo3D-System und der ersten Applikation. Die anschließende Übungsaufgabe diente dazu, sich mit der Interaktion vertraut zu machen. Dabei konnten auch Fragen an die Versuchsführerin gestellt werden, um mehr Sicherheit im Umgang mit dem System zu erlangen. In der Übungsaufgabe für die Informationssuche hatten sich die Teilnehmer über eine spezifische Systemkomponente zu informieren, um sich mit den Navigationsmöglichkeiten vertraut zu machen. In der Übungsaufgabe zur Konstruktion modellierten die Probanden eine nur aus drei Grundobjekten bestehende Hantel, um die Manipulationsmöglichkeiten auszuprobieren und zu üben. Die Einführungsphase dauerte durchschnittlich 45 Minuten. Für die einzelnen Aufgaben hatten die Teilnehmer jeweils maximal 10 Minuten Zeit. Die Konstruktionsaufgaben waren so ausgelegt, dass sie in dieser Zeit nicht komplett bewältigt werden konnten, um Deckeneffekte zu vermeiden. Dies trifft auch auf die mittlere und letzte Aufgabe im Info-Browser zu. Nach jeder Aufgabe wurden die Versuchsteilnehmer gebeten, ihre Beanspruchung auf der SEA-Skala (Eilers, Nachreiner & Hänecke 1986) einzuschätzen. Am Ende der Aufgabenbearbeitung wurde das semantische Differential zur Einschätzung der hedonistischen und pragmatischen Qualität sowie der Attraktivität für die entsprechende Applikation vorgelegt (Hassenzahl et al. 2000). Außerdem waren die bearbeiteten Aufgaben in ihrer Schwierigkeit einzuschätzen und in eine Rangreihe zu bringen. Nach einer Pause folgte die zweite Applikation mit einer Einführung in die Testapplikation und der Übungsaufgabe. Abbildung 28 veranschaulicht den Untersuchungsablauf.

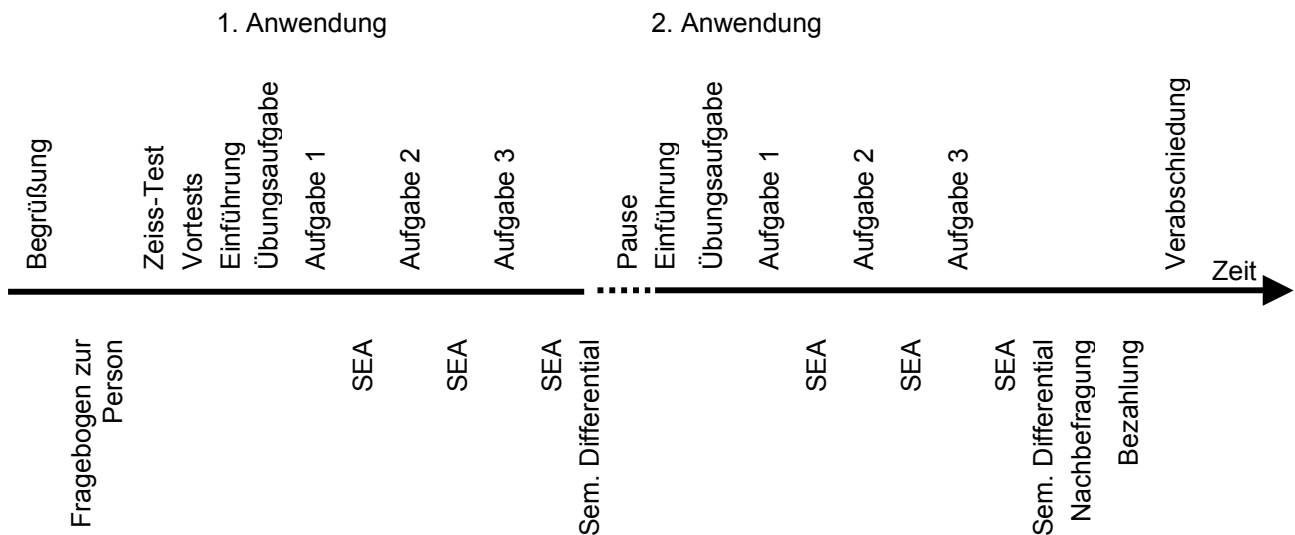


Abb. 28: Versuchsablauf bei der Untersuchung prototypischer Aufgaben. Die Untersuchungszeit variierte für die einzelnen Versuchsteilnehmer in der Einführung in Abhängigkeit von der Kalibrierungsdauer sowie bei den Übungsaufgaben durch die unterschiedliche Anzahl von Nachfragen und der Lerngeschwindigkeit.

Danach wurden die drei Aufgaben in der Reihenfolge des Permutationsplanes bearbeitet und die SEA-Skala, das semantische Differential sowie ein Ranking der Aufgabenschwierigkeit ausgefüllt. Abschließend wurden die Teilnehmer gebeten, ihre Meinung zum System in einer Nachbefragung zum Experiment zu äußern.

### 5.5.3.2 Versuchsteilnehmer

Insgesamt wurden 24 Personen (4 Frauen, 20 Männer) untersucht. Das durchschnittliche Alter lag bei 27 Jahren (Minimum 13 Jahre, Maximum 42 Jahre). Die durchschnittliche Computerefahrung lag bei 8,6 Jahren, die durchschnittliche Computernutzung bei 24 Stunden pro Woche. Genutzte Computeranwendungen waren übliche Anwendungsprogramme (z.B. MS Word), 3-D Programme, Datenbanken, www/E-Mail, Spielen und Programmieren, wobei Anwendungsprogramme und WWW bei den Nennungen an erster Stelle standen. Die Hälfte der Personen hatte Erfahrung mit 3D-Anwendungen. Die Erfahrung mit Spracherkennungssoftware war eher gering, da 16 Personen keine und eine Person wenig Erfahrungen besaßen. Von den übrigen sieben Versuchsteilnehmer werteten sechs ihre Erfahrungen als schlechte.

Bei der Frage nach dem Lerntyp gaben die meisten Personen an, am besten durch eigenes Ausprobieren zu lernen. Als Gründe für die Versuchsteilnahme dominierten Interesse an Wissenschaft und Technik sowie Neugier.

Mit dem Zeiss Nahprüfgerät wurde sichergestellt, dass alle Versuchsteilnehmer über ein gutes Stereosehen verfügen, was für die beanspruchungsarme Arbeit mit dem autostereoskopischen Display vorauszusetzen ist. Dabei waren in einer unregelmäßig strukturierten Fläche mindes-

tens drei, maximal fünf Stufen einer Treppe, die mit stereoskopischen Parallaxe verborgen war, zu erkennen. Die Subtests aus HAWIE (Wechsler 1991) und LPS (Horn 1962) zum räumlichen Vorstellungsvermögen und zur Arbeitsgedächtnisspanne sowie zur verballogischen Intelligenz zeigten keine stark abweichenden Ergebnisse. Alle Teilnehmer lagen hier im durchschnittlichen Bereich. Es wurden keine signifikanten Zusammenhänge zwischen den Intelligenzmaßen und den erbrachten Leistungen gefunden.

#### 5.5.4 Ergebnisse der Untersuchung prototypischer Aufgaben

Die intendierten Schwierigkeitsstufen im Untersuchungsdesign wurden mit der wahrgenommenen Aufgabenschwierigkeit, die durch die Aufgabenrankings erhoben wurden, geprüft. Hier zeigte sich für die Konstruktionsaufgaben, dass 46% der Untersuchungsteilnehmer die Aufgabenschwierigkeit übereinstimmend mit den im Versuchsdesign angelegten Abstufungen wahrgenommen hatten (vgl. Abbildung 29).

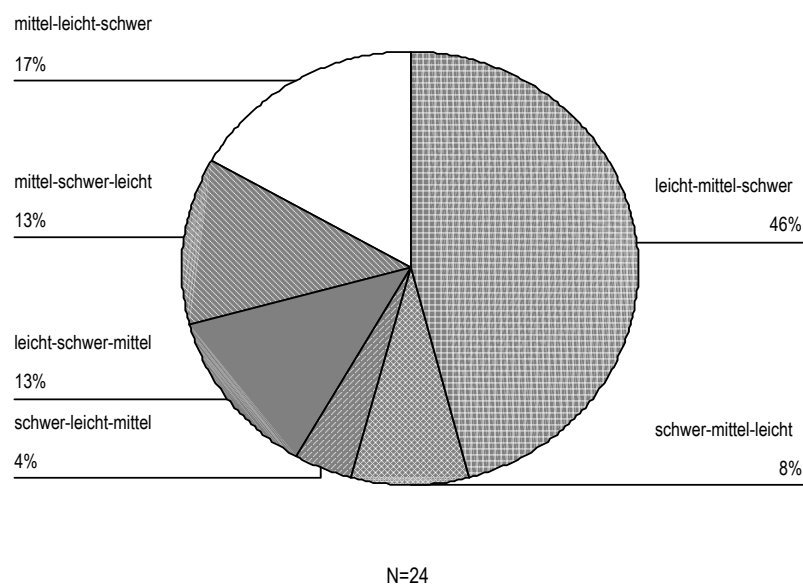


Abb. 29: Schwierigkeitsranking für die Aufgaben zur Konstruktion

Bei den Informationssuche beurteilten nur 33% der Probanden die Schwierigkeit der Aufgaben in der vorgesehenen Reihenfolge, 29% der Teilnehmer hielten die als schwerste konzipierte Aufgabe für die leichteste (vgl. Abbildung 30).

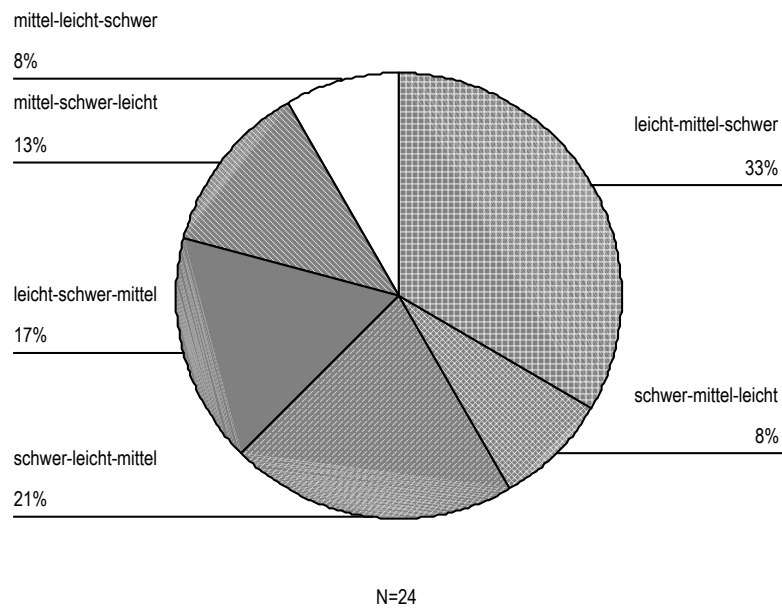


Abb. 30: Schwierigkeitsranking für die Aufgaben zur Informationssuche

Die Mediane des Schwierigkeitsrankings für die Konstruktionsaufgaben entsprechen der angestrebten experimentellen Manipulation (vgl. Tabelle 10). Die Mediane des Schwierigkeitsrankings für die Informationssucheaufgaben differenzieren nicht zwischen leichter und schwerer Aufgabe (vgl. Tabelle 10). Die Aufgabenschwierigkeit wird deshalb nur für die Konstruktionsaufgaben in der geplanten Weise interpretiert.

Tab. 10: Schwierigkeitsranking für die Aufgaben zur Konstruktion und Informationssuche

	Leicht (Normwert 3)	Mittel (Normwert 2)	Schwer (Normwert 1)
Mediane für die Konstruktion	3	2	1
Mediane für die Informationssuche	2,5	2	2

Da die Leistungen beim Konstruieren und beim Informationssuchen nach dem Kolmogorov-Smirnov-Test nicht normalverteilt sind, wurden für die deskriptive Darstellung als Maße der zentralen Tendenz Mediane (Tabelle 11 und 12) und für die inferenzstatistische Auswertung der Leistungsdaten nonparametrische Verfahren verwandt. Die Leistungen in den Aufgaben zur Konstruktion unterscheiden sich signifikant über die Aufgabenschwierigkeit (Friedman-Test, N=24;  $\chi^2=16,967^{**3}$ ; df=2), wobei der Unterschied auf die höhere Leistung in der mittleren Schwierigkeitsstufe zurückgeht. Es gibt keine signifikanten Leistungsunterschiede in

3 \*\* Ergebnis ist auf dem Niveau von 0,01 signifikant; \* Ergebnis ist auf dem Niveau von 0,05 signifikant

den drei Schwierigkeitsabstufungen bei der Konstruktion in Abhängigkeit von der Position der jeweiligen Aufgabe. Die Leistungen zwischen den Aufgaben, die zuerst, an zweiter oder an dritter Stelle bearbeitet wurden, unterscheiden sich bei der Konstruktion mit einer Irrtumswahrscheinlichkeit von  $p=.05$  grenzwertig nicht statistisch bedeutsam.

Tab. 11: *Mediane der Leistung in den Aufgaben zur Konstruktion*

Aufgaben-schwierig-keit / Position	Leicht	Mittel	Schwer	Mediane nach Position
Erste	8,5	10,5	9	9
Zweite	10	14	8,5	10
Dritte	11,5	13,5	11	11
Mediane nach Aufga-ben	9	13	9	

Die Leistungen in den Aufgaben zur Informationssuche unterscheiden sich nicht über die Aufgabenschwierigkeit. Die Position der Bearbeitung für die leichte, die mittelschwere und die schwere Aufgabe zur Informationssuche verursacht keine signifikanten Leistungsunterschiede. Die Leistungen zwischen den Aufgaben bei der Informationssuche, die an erster, an zweiter und an dritter Stelle bearbeitet wurden, unterscheiden sich jedoch signifikant (Friedman-Test,  $N=24$ ,  $\chi^2=7,853^*$ ,  $df=2$ ).

Tab. 12: *Mediane der Leistung in den Aufgaben zur Informationssuche*

Aufgaben-schwierig-keit / Position	Leicht	Mittel	Schwer	Mediane nach Position
Erste	45	44	48	45
Zweite	48,18	47	54,5	48,5
Dritte	46,95	49,5	54	52,5
Mediane nach Aufga-ben	45	47,5	53	

Da auch die multimodale Interaktionshäufigkeit nicht dem Normalverteilungskriterium genügt, wurden auch dafür Mediane als deskriptives Maß (Tabelle 13 und 14) und nonparametrische Verfahren zur Signifikanzprüfung eingesetzt. Die multimodale Interaktionshäufigkeit beim Konstruieren unterscheidet sich signifikant zwischen den drei Schwierigkeitsstufen (Friedman-Test,  $N=24$ ,  $\chi^2=6,583^*$ ,  $df=2$ ), wobei die Unterschiedlichkeit durch die höheren Werte in der mittelschweren Aufgabe erzeugt werden. Die Position der Bearbeitung für die leichte, die mittelschwere und die schwere Aufgabe zur Konstruktion verursacht keine signifikant unterschiedlichen multimodalen Interaktionshäufigkeiten. Die multimodale Interakti-



onshäufigkeit verändert sich nicht bedeutsam über die durch die Position widergespiegelten Messzeitpunkte.

Tab. 13: *Mediane der multimodalen Interaktionshäufigkeit beim Konstruieren*

Aufgabenschwierigkeit \ Position	Leicht	Mittel	Schwer	Mediane nach Position
Erste	27	37	52,5	37
Zweite	31,5	57,5	31,5	37,5
Dritte	40,5	46,5	33,5	40,5
Mediane nach Aufgaben	35,5	42,5	35	

Die multimodale Interaktionshäufigkeit beim Informationssuchen unterscheidet sich signifikant zwischen den drei Schwierigkeitsstufen (Friedman-Test,  $N=24$ ,  $\chi^2=9,083^*$ ,  $df=2$ ). Die Position, an der die leichte, die mittelschwere und die schwere Aufgabe jeweils bearbeitet wurde, führt nicht zu bedeutsam unterschiedlichen multimodalen Interaktionshäufigkeiten. Die multimodale Interaktionshäufigkeit ist für die drei Messzeitpunkte nicht signifikant unterschiedlich.

Tab. 14: *Mediane der multimodalen Interaktionshäufigkeit beim Informationssuchen*

Aufgabenschwierigkeit \ Position	Leicht	Mittel	Schwer	Mediane nach Position
Erste	40	64	52,5	52
Zweite	32,5	71,5	54,5	59,5
Dritte	42	63	63	63
Mediane nach Aufgaben	40	68,5	56	

Die Maße zur Anstrengungseinschätzung, die mit der SEA-Skala gemessen wurden, sind normalverteilt, so dass Mittelwerte zur Beschreibung (Tabelle 15 und 16) und parametrische Verfahren zur inferenzstatistischen Absicherung der Daten genutzt wurden. Die Anstrengungsmaße unterscheiden sich zwischen den drei Schwierigkeitsstufen beim Konstruieren nicht bedeutsam. Die Position, an der die einzelnen Aufgaben bearbeitet wurden, beeinflusst die Anstrengungsmaße beim Konstruieren nicht. Die Anstrengungsmaße werden nicht bedeutsam durch den Messzeitpunkt beeinflusst.

Tab. 15: Mittlere Anstrengung beim Konstruieren

Aufgabenschwierigkeit \ Position	Leicht	Mittel	Schwer	Mittelwerte nach Position
Erste	5,32	3,80	4,20	4,44
Zweite	3,25	3,31	5,60	4,05
Dritte	3,72	5,64	4,86	4,74
Mittelwerte nach Aufgaben	4,1	4,25	4,89	

Die eingeschätzte Anstrengung bei der Informationssuche unterscheidet sich nicht in den drei Aufgaben. Die Position, an der die leichte, mittelschwere und schwere Aufgabe jeweils bearbeitet wurde, führt nur bei der schweren Aufgaben zu signifikanten Unterschieden in den Anstrengungsmaßen (Anova,  $N=24$ ,  $F=4,619^*$ ,  $df=2$ ). Dieser Unterschied in den Anstrengungsmaßen geht darauf zurück, dass die schwere Aufgabe, wenn sie zuletzt bearbeitet wurde, als bedeutend weniger anstrengend eingeschätzt wurde (vgl. Tabelle 16, Spalte 4). Die Anstrengungswerte werden nicht von der Position, an der sie erhoben wurden, beeinflusst.

Tab. 16: Mittlere Anstrengung beim Informationssuchen

Aufgabenschwierigkeit \ Position	Leicht	Mittel	Schwer	Mittelwerte nach Position
Erste	3,81	2,94	4,94	3,90
Zweite	3,37	3,23	4,13	3,58
Dritte	1,53	4,56	2,24	2,78
Mittelwerte nach Aufgaben	2,90	3,58	3,77	

Insgesamt wurde die Anstrengung sowohl beim Konstruieren als auch beim Informationssuchen als relativ niedrig eingeschätzt. Die gemittelte Anstrengung für das Konstruieren lag bei 4,41 und für die Informationssuche bei 3,42 auf einer Skala von 0 bis 11. Die Hypothese 10, dass die durchschnittliche Beanspruchung für beide Aufgaben signifikant geringer als „stark anstrengend“ eingeschätzt wird, konnte bestätigt werden (vgl. Tabelle 17). Auch bei einem Testwert von 6, der der Wortmarke „ziemlich anstrengend“ auf der SEA-Skala entspricht, sind die gemessenen Werte noch signifikant kleiner. Die verbalen Anker für die Beurteilung sind aus der im Anhang 3 eingefügten SEA-Skala zu entnehmen.

Tab. 17: T-Tests zur Beanspruchungshöhe

	Testwert = 8 (SEA-Skala „stark anstrengend“)					
	T	df	Sig. (2-seitig)	Mittlere Differenz	95% Konfidenzintervall der Differenz	
					Untere	Obere
Informationssuche Leicht	-9,835	23	,000	-5,0979	-6,1702	-4,0256
Informationssuche Mittel	-13,049	23	,000	-4,4229	-5,1241	-3,7217
Informationssuche Schwer	-9,922	23	,000	-4,2333	-5,1160	-3,3507
Konstruktion Leicht	-10,067	23	,000	-3,9042	-4,7064	-3,1019
Konstruktion Mittel	-8,235	23	,000	-3,7500	-4,6920	-2,8080
Konstruktion Schwer	-6,478	23	,000	-3,1146	-4,1092	-2,1200

Die Anstrengungsunterschiede zwischen den Konstruktionsaufgaben und den Aufgaben zur Informationssuche sind jedoch signifikant (Wilcoxon,  $N=24$ ,  $Z=-2,20^*$ ), d.h. das Konstruieren wurde als anstrengender eingeschätzt als die Informationssuche. Abbildung 31 veranschaulicht die unterschiedlichen Beurteilungen der Beanspruchung über die Aufgabenvarianten beim Konstruieren und beim Informationssuchen.

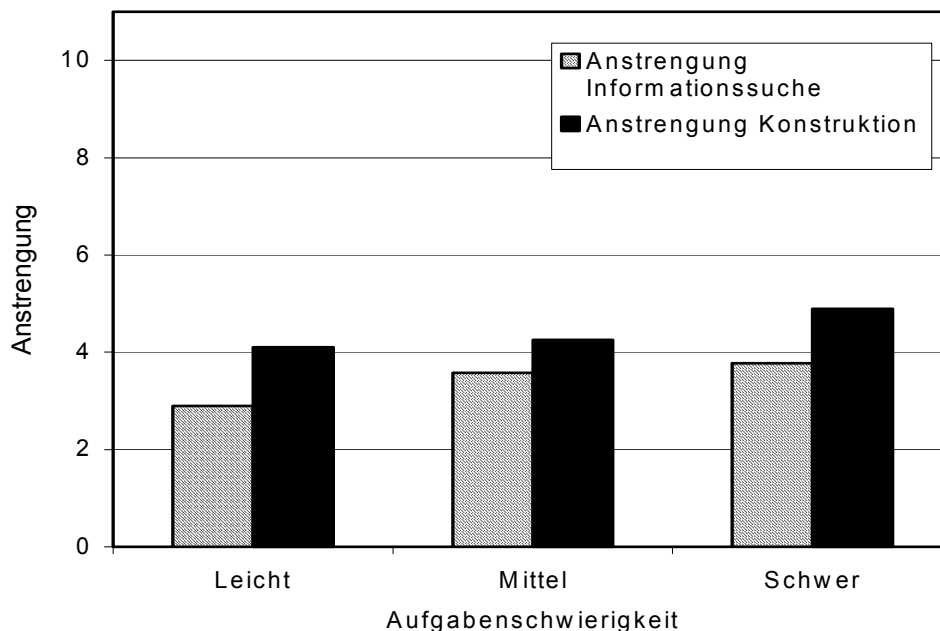


Abb. 31: Mittlere Anstrengung beim Konstruieren und beim Informationssuchen

Zur Prüfung der Zusammenhangshypothesen 1 und 2, dass mit zunehmender Nutzung der multimodalen Interaktion die Leistung ansteigt, wurden Korrelationen berechnet. Da nicht alle eingehenden Variablen der Normalverteilungsvoraussetzung entsprachen, werden die auf Rangkorrelationen beruhenden Ergebnisse dargestellt.

Für das Konstruieren zeigte sich ein hochsignifikanter positiver Zusammenhang zwischen der erbrachten Leistung und der multimodalen Interaktionshäufigkeit (Spearman-Rho,  $N=72$ ,  $r=0,373^{**}$ ). Diese Korrelation ist auch signifikant bei der Verwendung von gemittelten Werten für die Leistung und die multimodale Interaktionshäufigkeit beim Konstruieren bei einer Stichprobengröße von  $N=24$ . Werden die Computererfahrung und Nutzung, sowie die erhobenen Intelligenzmaße und die Anstrengung mit Hilfe einer Partialkorrelation kontrolliert, können die hochsignifikanten Zusammenhänge zwischen der Leistung und der multimodalen Interaktionshäufigkeit bestätigt werden (Partialkorrelation,  $r=0,5583^{**}$ ,  $df=16$ ). Damit kann die Hypothese 1 angenommen werden.

Für die Informationssuche konnte kein statistisch bedeutsamer Zusammenhang zwischen der erbrachten Leistung und der multimodalen Interaktionshäufigkeit gefunden werden, so dass die Hypothese 2 abgelehnt werden muss.

Werden die einzelnen Konstruktionsaufgaben getrennt für jede Schwierigkeitsstufe ausgewertet, ergeben sich die in Tabelle 18 dargestellten Korrelationen.

*Tab. 18: Leistung und multimodale Interaktionshäufigkeit über die Schwierigkeitsstufen beim Konstruieren*

Aufgabenschwierigkeit	Leicht	Mittel	Schwer
Korrelationskoeffizient	0,443	0,359	0,243
Signifikanzniveau (1-seitig)	0,015*	0,042*	0,126

Bei der leichten und mittelschweren Aufgabe ergeben sich die erwarteten signifikant positiven Korrelationen zwischen der Leistung und der multimodalen Interaktionshäufigkeit, während es in der schweren Aufgabe keinen signifikanten Zusammenhang gibt. Die Aufgabenschwierigkeit beeinflusst also die Enge des Zusammenhanges zwischen den korrelierten Variablen. Auch bei der Partialisierung von Computererfahrung, Intelligenzmaßen und Anstrengung bleiben die Signifikanzniveaus für die drei unterschiedlich schwierigen Aufgaben erhalten.

Die Analyse der einzelnen Aufgabenvarianten zur Informationssuche führen zu den in Tabelle 19 dargestellten Korrelationen.

*Tab. 19: Leistung und multimodale Interaktionshäufigkeit über die Schwierigkeitsstufen beim Informationssuchen*

Aufgabenschwierigkeit	Leicht	Mittel	Schwer
Korrelationskoeffizient	-0,623	0,441	0,464
Signifikanzniveau (1-seitig)	0,001**	0,015*	0,011*

Für die beiden Aufgaben (präskriptiv als mittelschwere und schwere Aufgabe bezeichnet), in denen innerhalb eines feststehenden Zeitraumes verschiedene Informationen gesucht und behalten werden sollten, ergeben sich ähnlich signifikant positive Zusammenhänge zwischen Leistung und multimodaler Interaktionshäufigkeit wie in der leichten Konstruktionsaufgabe. Für die erste Aufgabenvariante, bei der für eine bildlich dargestellten Komponente so schnell als möglich der Name herauszufinden war, wurde ein hochsignifikant negativer Zusammenhang zwischen der Leistung und der multimodalen Interaktionshäufigkeit gefunden. Die zeitkritische Suche nach einer nicht konkret-begrifflich repräsentierten Information kann effektiver ohne Modalitätswechsel vorgenommen werden. Die Kontrolle von Computererfahrung bzw. -nutzung, Intelligenzmaßen und der Anstrengung durch Partialkorrelationen für die einzelnen Aufgaben veränderten für die als schwer operationalisierte Aufgabe das Signifikanzniveau auf  $p=0,076$ . Die Anstrengung beeinflusst bei dieser Aufgabe die Enge des Zusammenhang zwischen Leistung und multimodaler Interaktionshäufigkeit. Sie beeinflusst die Leistung in der schweren Aufgabe ebenfalls in positiver Richtung.

Es wurde angenommen, dass die Modalitäten in den beiden Aufgabentypen „Konstruktion“ und „Informationssuche“ unterschiedlich genutzt werden. Abbildung 32 veranschaulicht den Unterschied.

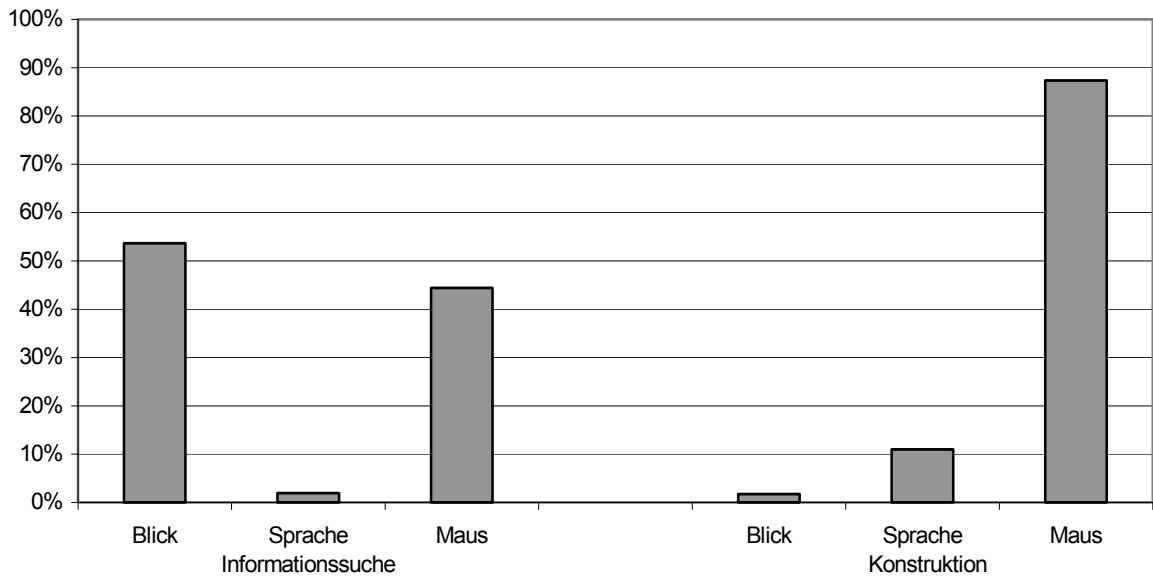


Abb. 32: Relativer Anteil der Interaktionsmodalitäten für die beiden prototypischen Aufgaben Informationssuche und Konstruktion

Es zeigte sich, wie in Hypothese 3 angenommen, dass die Blickinteraktion und die Sprachinteraktion beim Konstruieren und bei der Informationssuche unterschiedlich benutzt wurden, nicht jedoch die Mausinteraktion. Die Blickinteraktion wird signifikant häufiger bei der Informationssuche, die Sprachinteraktion signifikant häufiger beim Konstruieren benutzt (vgl. Tabelle 20). Somit bestätigten sich auch die Hypothesen 3a und 3b.

Tab. 20: Unterschiedliche Nutzung der Modalitäten beim Konstruieren und beim Informationssuchen

	Konstruktion alle Blickinteraktionshäufigkeit – Informationssuche alle Blickinteraktionshäufigkeit	Konstruktion alle Sprachinteraktionshäufigkeit – Informationssuche alle Sprachinteraktionshäufigkeit	Konstruktion alle Mausinteraktionshäufigkeit – Informationssuche alle Mausinteraktionshäufigkeit
Z	-2,429	-2,387	-1,772
Signifikanz (2-seitig)	,015	,017	,076

Die Auswertung des Semantischen Differentials basiert auf der Einstufung der Urteile der Probanden auf der siebenstufigen Ratingskala zwischen den Polen der Urteilsdimensionen. Der Wert 4 stellt die unentschiedene Beurteilung des Systems im internalisierten Abgleich mit bekannten Systemen dar. Werte, die unterhalb des Wertes 4 liegen, sprechen für eine schlechtere Beurteilung des eingeschätzten Systems gegenüber den als ähnlich wahrgenommenen, bekannten Systemen. Werte über 4 belegen eine positivere Einschätzung gegenüber den Vergleichsobjekten.

Die pragmatische Qualität der Testapplikation „Konstruktion“ wurde, wie in Hypothese 4 angenommen, „tendenziell positiv“ bewertet mit einem Wert über 4 im Semantischen Differential. Dies gilt auch für die Testapplikation „Informationssuche“, so dass die Hypothese 5 ebenfalls bestätigt werden konnte.

Die hedonistische Qualität für die Testapplikation „Konstruktion“ wurde darüber hinaus, wie in Hypothese 6 erwartet, als „definitiv positiv“ eingeschätzt mit einem signifikant über dem Testwert 5 liegenden Score (t-Test, Testwert 5,  $t=2,783^*$ ,  $df=23$ ). Die hedonistische Qualität und die Attraktivität für die Testapplikation „Informationssuche“ wurde „tendenziell positiv“ beurteilt, damit gelten die Hypothesen 7 und 9 als bestätigt.

Die Attraktivität der Testapplikation „Konstruktion“ wurde als „definitiv positiv“ bewertet (t-Test, Testwert 5,  $t=2,230^*$ ,  $df=23$ ), wodurch auch die Hypothese 8 angenommen werden kann. Tabelle 21 zeigt die Testergebnisse im Überblick.

Tab. 21: Pragmatische und hedonistische Qualität sowie Attraktivität der Testapplikationen

		Testwert = 4					
		T	df	Sig. (2-seitig)	Mittlere Differenz	95% Konfidenzintervall der Differenz	
						Untere	Obere
Informations- suche	Pragmatische Qualität	7,318	22	,000	1,1509	,8247	1,4770
	Hedonistische Qualität	8,159	22	,000	1,3304	,9923	1,6686
	Attraktivität	8,346	22	,000	1,2152	,9132	1,5172
Konstruktion	Pragmatische Qualität	5,741	23	,000	,9104	,5823	1,2385
	Hedonistische Qualität	9,350	23	,000	1,4237	1,1088	1,7387
	Attraktivität	9,074	23	,000	1,3258	1,0236	1,6281

Insgesamt wird das System mit den beiden Testapplikationen als tendenziell positiv beurteilt gegenüber vergleichbaren Systemen, obwohl es sich bisher um eine prototypische Realisierung handelt. Stärker positiv hervorgehoben wurde von den Beurteilern die hedonistische, innovative Komponente des Systems, die auch bei der Testapplikation „Informationssuche“ nur knapp das Signifikanzniveau für „definitiv positiv“ verfehlte. Die tendenziell positive Bewertung der pragmatischen Qualität belegt, dass die in der Interaktion mit dem System erzielten Leistungen sinnvoll im Rahmen der Evaluation interpretiert werden können.

In der Nachbefragung beurteilten die Untersuchungsteilnehmer die neuen Interaktionsmodalitäten Blick und Sprache. Dabei wurden die Verzögerungszeiten bei der systemseitigen Verarbeitung der neuen Eingabemodalitäten und ihre Funktionalität eingeschätzt. Fehlende Prozentwerte gehen auf diejenigen Befragten zurück, die keine Angaben machen konnten.

Die Blickinteraktion wurde von 44% der Probanden als zu langsam, von 7% als zu schnell und von 41% als angemessen schnell empfunden. 63% der Befragten wünschten sich keine zusätzliche Rückmeldung für die Blickinteraktion, was die Ergebnisse aus dem Blickrückmeldungsexperiment (vgl. 5.3.1) weiterhin untersetzt. Die Blickinteraktion wurde von 44% der Teilnehmer als weniger funktional, von 52% als hilfreich eingeschätzt.

Die Sprachinteraktion wurde von 63% der Teilnehmer als zu langsam und von 30% als angemessen schnell beurteilt. 96% der Untersuchungsteilnehmer empfanden die akustische Rückmeldung der Erkennung als hilfreich, nur 4% empfanden sie als überflüssig. Die Qualität der Spracherkennung wurde von 52% der Befragten als gut, von 41% als schlecht eingeschätzt. Die Funktionalität der Sprachinteraktion wurde von 85% der Probanden als hilfreich und nur von 4% als wenig funktional wahrgenommen. Als Interaktionsformen bevorzugten 54% der Befragten eine Kombination aus Maus, Sprach- und Blickinteraktion. 31% präferierten die Sprachinteraktion in Kombination mit der Maus. Jeweils 8% der Befragten sprachen sich für die Maus allein bzw. für die kombinierte Blick- und Mausinteraktion aus.

Einzelne weitere Aussagen schätzten die Blickinteraktion widersprüchlich ein: Während zwei Versuchsteilnehmer sie besonders wenig anstrengend empfanden, bemängelten zwei andere, dass sie anstrengend für die Augen wäre. Die Kombination von Blickinteraktion zur Adressierung eines interaktiven Objektes im GUI und einem Mausklick zur Bestätigung wurde als innovativ gelobt. Die Interpretation des Blickes wurde für die Informationssuche von einer Person als passend eingeschätzt, eine andere fand sie eher störend. Die Ungenauigkeit der Messung über die Dauer der Untersuchung wurde von mehreren Personen kritisiert.

Die Sprachinteraktion wurde von neun Probanden in unterschiedlichen Aspekten positiv hervorgehoben. Die Erkennungsrate wurde von fünf Personen bemängelt, die jedoch die Möglichkeit zur Sprachinteraktion dennoch befürworteten. Vier Personen fanden es anstrengend, deutlich sprechen zu müssen.

### **5.5.5 Interpretation der Ergebnisse der Untersuchung prototypischer Aufgaben**

Die Aufgabenschwierigkeit wurde nur für die Konstruktionsaufgaben, nicht jedoch bei der Informationssuche und -wiedergabe, von den Untersuchungsteilnehmern entsprechend der operationalisierten Schwierigkeitsabstufung beurteilt.

Die nach der Operationalisierung leichte Aufgabe für die Informationssuche und -wiedergabe kann in einem anderen Anforderungsfeld des Berliner Schemas eingeordnet werden als die



mittelschwere und schwere Aufgabe. Während letztere Leistungsanforderungen insbesondere in der Informationsverarbeitung, dem Kodieren und Ordnen von Informationen für ihre gezielte Speicherung und Wiedergabe stellten, erforderte die leichte Aufgabe das Entdecken eines Objektes. Die mittelschwere und schwere Aufgabe konnten von den Versuchsteilnehmern bzgl. ihrer Schwierigkeit nicht unterschieden werden, da die gleichen Interaktionserfordernisse in beiden Aufgaben bestanden und offenbar die absolut enthaltenen Informationen von den Probanden als Beurteilungskriterium weniger stark gewichtet wurden.

Die Leistungen in den Konstruktionsaufgaben unterschieden sich nicht für die leichte und schwere Aufgabe, die mittelschwere Aufgabe wurde mit besserer Leistung bewältigt. Die größeren Unterschiede in der Leistung konnten dann erzielt werden, wenn die mittelschwere Aufgabe an zweiter und dritter Position bearbeitet wurde. Beim Vergleich der deskriptiven Maße der multimodalen Interaktionshäufigkeit zeigt sich, dass die multimodale Interaktion besonders häufig in der mittelschweren Aufgabe genutzt wurde, wenn diese an zweiter oder dritter Position bearbeitet wurde. Die multimodale Interaktionshäufigkeit wird nicht durch die drei möglichen Positionen einer Aufgabe beeinflusst und verändert sich auch nicht bedeutsam über die Bearbeitungszeit. Diese Übereinstimmungen erhärten die Hypothese des Zusammenhanges von Konstruktionsleistung und multimodaler Interaktionsnutzung. Da Personenvarianzen im Messwiederholungsplan keinen Einfluss haben, und die Aufgabenschwierigkeit von 71% der Versuchsteilnehmer als mittelschwer oder schwer beurteilt wurde, gehen die Leistungsunterschiede eher auf die Interaktionsnutzung zurück als auf die Aufgabenschwierigkeit. Dies zeigen auch die Korrelationsmaße, die auch bei der Kontrolle von Personen- und Anstrengungsunterschieden einen hochsignifikanten Zusammenhang zwischen der Leistung und der multimodalen Interaktionshäufigkeit aufweisen. Nur bei der schweren Konstruktionsaufgabe, die von 73% der Probanden als schwerste Aufgabe beurteilt wurde, konnte die Leistung nicht mehr von der Nutzung der multimodalen Interaktionsangebote profitieren. Insgesamt zeigte sich jedoch ein hochsignifikanter positiver Zusammenhang zwischen der Leistung und der multimodalen Interaktionshäufigkeit in den Konstruktionsanforderungen. Daraus ist zu folgern, dass ein multimodales System wie mUltimo3D Leistungen in Aufgaben, die komplexe motorische und Problemlöseanforderungen stellen, unterstützen kann. Die Position der einzelnen Aufgaben hatte keinen bedeutsamen Einfluss auf die Konstruktionsleistung. Es gibt einen tendenziellen Lerneffekt über die Bearbeitungszeit, der jedoch durch das vollständig permutierte Untersuchungsdesign für die Aufgabenschwierigkeit kontrolliert wurde.

Die Leistung in den Aufgaben zur Informationssuche und -wiedergabe unterscheiden sich nicht über die Aufgabenschwierigkeit, was sich auch in der uneinheitlichen Beurteilung der Aufgabenschwierigkeit durch die Versuchsteilnehmer widerspiegelt. Auch die Bearbeitungsposition der Aufgaben hat keinen Einfluss auf die Leistung, damit ist abgesichert, dass die Leistungen nicht durch einen Wissenstransfer in einer einzigen Aufgabe bedeutsam beeinflusst wurde. Alle Aufgaben werden mit zunehmender Bearbeitungszeit besser bewältigt.

Auch für die Informationssuche und -wiedergabe sind mittels Permutation in jeder operationalisierten Schwierigkeitsstufe der Aufgaben gleich viele Fälle mit gleichlanger vorangegangener Bearbeitungszeit enthalten. Die multimodale Interaktionshäufigkeit unterscheidet sich zwischen den Schwierigkeitsstufen der Aufgaben, wobei die größte Unterschiedlichkeit zwischen der leichten und mittleren Schwierigkeitsstufe zu verzeichnen ist. Die Position der einzelnen Aufgaben und die Bearbeitungszeit haben keinen bedeutsamen Einfluss auf die multimodale Interaktionshäufigkeit. Die Korrelationen zwischen der Leistung und der Nutzung multimodaler Interaktion in den drei als unterschiedlich schwierig operationalisierten Aufgaben zur Informationssuche und -wiedergabe zeigen ein differenziertes Bild. Während es sowohl in der Aufgabe mittlerer Schwierigkeit und der schweren Aufgabe einen signifikanten positiven Zusammenhang zwischen der Leistung und der multimodalen Interaktionshäufigkeit gibt, besteht in der leichten Aufgabe ein hochsignifikant negativer Zusammenhang zwischen Leistung und multimodaler Interaktionshäufigkeit. Aufgrund dieser gegenläufigen Zusammenhänge konnte für die Informationssuche und -wiedergabe insgesamt kein bedeutsamer positiver Zusammenhang zwischen Leistung und multimodaler Interaktionshäufigkeit belegt werden. Die Auswahl einer geeigneten Interaktionsstrategie scheint bei der Informationssuche und -wiedergabe ein ausgeglichenes Leistungsvermögen über alle Aufgabenschwierigkeiten zu ermöglichen. Für die leichte Aufgabe, deren Leistung durch die Geschwindigkeit mit der eine abgebildete Komponente erfolgreich identifiziert wurde, war die Auswahl einer einzigen Modalität für die zügige Navigation durch den Datenraum vorteilhaft. Zur Erschließung möglichst vieler spezifischer Informationen im Datenraum mit gleich-bleibendem zeitlichen Rahmen, wie es in den beiden anderen Aufgaben gefordert war, bietet die multimodale Interaktion über Sprache und Blick günstigere Voraussetzungen.

Die Analyse der Interaktionsdaten zeigte, dass sich die Nutzung der angebotenen Modalitäten in den Konstruktionsaufgaben von der in den Aufgaben zur Informationssuche und -wiedergabe unterschied. Die Blickinteraktion spielte eine bedeutendere Rolle bei letzteren, wohingegen die Sprachinteraktion häufiger bei den Konstruktionsaufgaben genutzt wurde. Die Mausinteraktion unterschied sich nicht bedeutsam in den beiden Aufgabenbereichen. Dieses Ergebnis weist darauf hin, dass die Bedeutung, die eine Modalität für die Aufgabenbearbeitung hat, unterschiedlich sein kann. Dies kann im Rahmen der Entwicklung multimodaler Systeme durch Evaluationsstudien geprüft werden, um die jeweils zweckmäßigen Interaktionsmodalitäten aus der Vielzahl denkbarer auszuwählen.

Die subjektiven Beanspruchungsdaten wurden in den Konstruktionsaufgaben nicht durch die Aufgabenschwierigkeit, die Position oder Bearbeitungszeit beeinflusst. Sie lag durchschnittlich etwas über dem verbalen Anker von „einigermaßen anstrengend“ und waren damit noch signifikant größer als bei den Aufgaben zur Informationssuche und -wiedergabe. Die subjektiven Beanspruchungsdaten wurden in den Aufgaben zur Informationssuche und -wiedergabe nicht von der Aufgabenschwierigkeit und der Bearbeitungszeit beeinflusst. Nur die schwere

Aufgabe wurde als weniger anstrengend erachtet, wenn sie an letzter Position bearbeitet wurde. Für die Informationssuche und -wiedergabe wurde die Anstrengung durchschnittlich unter der Wortmarke „einigermaßen anstrengend“ beurteilt. Den Versuchsteilnehmern war es demnach nach einer relativ kurzen Übungsphase von 45 Minuten möglich, mit einem moderaten Anstrengungsniveau die geforderten Leistungen unter Ausnutzung der neuen Interaktionsmodalitäten zu erbringen. Es kann bereits für das Prototypen-System bestätigt werden, dass die darin gestaltete multimodale Interaktion von den Benutzern ohne besonderen Aufwand eingesetzt werden kann, da auch ein ex post Test keinen Zusammenhang zwischen Anstrengung und multimodaler Interaktionshäufigkeit aufwies.

Die Beurteilung der beiden Testapplikationen durch das semantische Differential fiel für die pragmatische Qualität tendenziell positiv aus. Das Ergebnis kann jedoch aus zwei Perspektiven interpretiert werden. Einerseits konnten die Probanden die Systemeigenschaften für die jeweiligen Aufgaben als hilfreich erfahren, andererseits kann die Beurteilung auch prospektive Meinungen über den zukünftigen Vorteil solcher Systeme enthalten, die ein positiveres Urteil verursachen. Die pragmatische Qualität des Systems und der einzelnen Applikationen muss mit zunehmender Produktreife stärker im Fokus sowohl der Systementwicklung als auch der Evaluation stehen, als es für den bisherigen Entwicklungsstand angebracht erscheint. Dafür ist es notwendig, die Funktionalität der Anwendungen zu erhöhen. Darüber hinaus ist für die weitere Entwicklung großer Wert auf die Stabilität der Interaktionskanäle zu legen, damit den Benutzern ein zuverlässig funktionierendes, intelligentes Werkzeug an die Hand gegeben werden kann. Auch die hedonistische Qualität wurde als tendenziell oder definitiv positiv beurteilt. Auch auf dieser Beurteilungsdimension kann dem mUltimo3D-System in der ersten prototypischen Realisation eine gelungene Gestaltung bestätigt werden. Die Neuartigkeit des Systems, seine „überraschenden“ Charakteristika, sind ein wichtiger Anreiz für Benutzer, sich mit dem System zu beschäftigen und die notwendigen Lernschritte zu durchlaufen. Daher ist im weiteren Entwicklungsprozess ebenfalls auf Erhaltung und Verbesserung der hedonistische Qualität zu achten. Die Akzeptanz zukünftiger multimodaler Systeme wird von der Beachtung beider Qualitäten im Entwicklungsprozess profitieren.

Wesentliche Hinweise für weitere Gestaltungserfordernisse erbrachte die Nachbefragung der Test-Benutzer. Die Erkennungsgenauigkeit und -geschwindigkeit der Sprach- und Blickinteraktion sollte zukünftig verbessert werden. Eine Anreicherung an Funktionalität wurde besonders für die Blickinteraktion als wünschenswert erachtet. Bei einer verbesserten Erkennungsleistung der Spracheingaben und höherer Verarbeitungsgeschwindigkeit ist zu prüfen, ob die bisher von den Untersuchungsteilnehmern als hilfreich eingeschätzte akustische Rückmeldung der Erkennung einer Spracheingabe noch notwendig ist. Die Möglichkeit, über die angebotenen Interaktionskanäle verfügen zu können, wurde von den Befragten überwiegend positiv eingeschätzt. Die Präferenzen für die Kombinationen von Interaktionsformen unterschieden sich insofern, dass die Sprachinteraktion besonders häufig als wünschenswert angegeben wurde. Multimodale Systeme treffen demnach bei angemessener Gestaltung die Wünsche der Benutzer.

## 6 Integration der Ergebnisse und Diskussion

Aus dem Evaluationsprozess am mUltimo3D-System lassen sich die in Tabelle 22 dargestellten gestalterischen Schlussfolgerungen ziehen.

Tab. 22: Gestalterische Empfehlungen für die Blick- und Sprachinteraktion

Modalität Funktionsbereich	Blick	Sprache
<b>Erfassung</b>	<p>Tageslichttauglich</p> <p>Applikationsabhängig hinlänglich schnell</p> <p>Applikationsabhängig hinlänglich genau</p> <p>Kontaktfrei</p> <p>Nachführung bei und Kompensation von Kopf- bzw. Körperbewegungen mindestens in der x-, y- und z-Achse</p>	<p>Applikationsabhängig sprecherunabhängig oder sprecherabhängig</p> <p>Hohe Erkennungszuverlässigkeit</p> <p>Schnelle Analyse</p> <p>Applikationsabhängige Variation des Vokabularumfangs</p> <p>Applikationsabhängige Wahl der Analyseeinheiten</p> <p>Möglichst kontaktfrei (Abwägung bzgl. Zuverlässigkeit der Erkennung)</p>
<b>Kalibrierung/ Rekalibrierung</b>	<p>Einfach vom Benutzer durchzuführende Kalibrierung</p> <p>Automatische Augenfindung</p> <p>Angaben über die Kalibrierungsgüte</p> <p>Speicherung von Kalibrierungsdaten der Systembenutzerinnen</p> <p>Leichte, in jeder Applikation auslösbbare Rekalibrierung</p>	<p>Skalierbar bzgl. Umgebungsgeräuschen</p> <p>Applikationsabhängig Trainierbarkeit</p> <p>Einfaches Hinzufügen von Vokabular</p> <p>Speicherung von sprachinteraktionsbezogenen Benutzerprofilen</p>
<b>Applikationsbezogene Funktion</b>	<p>Systemseitige Interpretation als Gebiet visueller Aufmerksamkeit</p> <p>Verzicht auf permanente Rückmeldung des Blickortes</p> <p>Verknüpfung mit aktiv erzeugten Interaktionen zur Auslösung von Manipulationen (z.B. über Maus, Tastatur, Sprache, Geste)</p> <p>Anreicherung/Detaillierung der grafischen Darstellung am Blickort</p> <p>Rückkehr zum Ausgangszustand der Darstellung bei Blickortwechsel (z.B. durch Verblässen von Informationen)</p> <p>Schwellzeitabhängige Darstellungsänderungen in adäquater Geschwindigkeit (Vermeidung unkontrollierbaren Aufklackerns)</p>	<p>Rückmeldung der Signalgüte (beim derzeitigen Stand der Technik)</p> <p>Kombination mit Sprachausgabe günstig</p> <p>Abkürzung von Interaktionswegen durch die Menüstrukturen</p> <p>Kombination mit Blick, Geste oder Maus für die Auswahl von Orten bzw. Objekten mit unterschiedlicher Präzision</p> <p>Kombination mit Schreibeingabe (Tastatur, Stift) für spezifische Eingaben (z.B. Zahlen)</p> <p>Applikationssteuerung eher kommandobasiert mit umschriebenem Vokabular</p> <p>Synonymerkennung bei befehlsbasierter Spracheingabe</p>

Die in den vorhergehenden Kapiteln ausführlich dargestellten Evaluationsuntersuchungen unterstützten den Systementwicklungsprozess von mUltimo3D dadurch, dass stufenweise Gestaltungsvarianten empirisch geprüft wurden. So konnten letztlich funktionierende unimodale Interaktionsformen in ein multimodales Gesamtsystem integriert werden, welches ebenfalls evaluiert wurde.

Im Folgenden sollen die aus den empirischen Studien gewonnen Erkenntnisse für die Gestaltung eines multimodalen Systems mit Blick- und Sprachinteraktion in der unter 3.3.2.2 eingeführten System-Aufgaben-Eigenschaftsmatrix zusammengefasst werden. Für die Systemeigenschaften wird eine Unterteilung in den Informationsfluss vom Menschen zur Maschine und von der Maschine zum Menschen vorgenommen. Zum einen werden dabei die multimodalen Eingaben nach den vom System erfassten und weiterverarbeiteten Datenniveau der multimodalen Benutzereingaben charakterisiert, z.B. Rohdaten der Blickbewegungsmessung, identifizierte Fixationen, Kommandos oder fließend gesprochene Sprache. Andererseits wird die vom System erzeugte Rückmeldung bzw. Systemantwort für diese Benutzereingaben, die über verschiedene Ausgabemedien erfolgte, aufgeführt. Der Erfolg der so skizzierten unterschiedlichen Gestaltungsvarianten lässt sich jedoch nicht absolut formulieren, sondern kann nur auf solche multimodalen Systeme übertragen werden, die Aufgaben mit ähnlichen Aufgabeneigenschaften wie die untersuchten unterstützen. Deshalb werden die Aufgaben, die in den vorliegenden Untersuchungen eingesetzt wurden, entsprechend dem in Abschnitt 3.3.2.1 entwickeltem Schema klassifiziert. Die Aufgaben wurden daraufhin geprüft, inwiefern sie durch die den spezifischen Aktivitäten zugeordneten Verhaltensformen beschrieben werden können und danach den dominierenden Aktivitäten bzw. Prozessen zugeordnet.

Die Beurteilung der empirisch geprüften Gestaltungsvarianten wird durch eine Skala veranschaulicht, welche die Eignung einer Gestaltungsvariante auf der Basis der deskriptiv-statistischen und inferenzstatistischen Evidenzen für die untersuchten Aufgabeneigenschaften einschätzt. Die Skala unterscheidet fünf Urteile, die durch die in Klammern stehenden Zeichen in der Matrix symbolisiert werden: ungeeignet (--), eher ungeeignet (-), unentschieden (0), eher geeignet (+) und geeignet (++). In die Beurteilung gehen häufige Aussagen aus der Heuristischen Evaluation (eine detaillierte Darstellung ist in Kapitel 5.2.3 zu finden), Leistungsdaten und Beurteilungen aus der Untersuchung der Blickinteraktion (eine detaillierte Darstellung ist in Kapitel 5.3.4 zu finden), Leistungsdaten und Beurteilungen aus der Untersuchung der Sprachinteraktion (eine detaillierte Darstellung ist in Kapitel 5.4.4 zu finden) und Leistungsdaten und Beurteilungen der Untersuchung prototypischer Aufgaben (eine detaillierte Darstellung ist in Kapitel 5.5.4 zu finden) ein. In den Anmerkungen zur Matrix, auf die durch hochgestellte Zahlen an den Beurteilungen verwiesen wird, werden noch weitere Ausführungen dazu gemacht, welche Ergebnisse die Beurteilung bestimmten.

Die hier vorgenommenen Beurteilungen über die Gestaltungsvarianten sollten durch weitere empirische Untersuchungen abgesichert und durch den Einbezug weiterer Aufgabenmerkmale vertieft werden. Die Untersuchung verschiedener multimodaler Systeme mit unterschiedlichen technischen Parametern für die Erfassung und Verarbeitung multimodaler Eingaben

trägt ebenfalls zur Erhärtung von Gestaltungsheuristiken bei. Auch eine erhöhte Erkennungszuverlässigkeit und -geschwindigkeit für Benutzereingaben können einige aktuelle Gestaltungsrestriktionen aufheben oder aus software-ergonomischer Sicht nicht mehr zwingend erfordern. Ein Beispiel dafür bietet die Sprachinteraktion in mUltimo3D, für die es sich bei der gegenwärtigen Erkennungszuverlässigkeit und -geschwindigkeit empfiehlt, der Benutzerin den Empfang einer sprachlichen Eingabe akustisch zurückzumelden.

### ***Anmerkungen zur Tabelle***

- zu 1: Die Beurteilung beruht auf den hochsignifikanten Leistungs- und Anstrengungsunterschieden und der subjektiven Präferenz zwischen den Bedingungen mit Blickrückmeldung und ohne Blickrückmeldung im Blickinteraktionsexperiment.
- zu 2: Die Beurteilung beruht darauf, dass die akustische Rückmeldung eines Blickes auf einem interaktiven Element im GUI in der Konstruktionsaufgabe und der Informationsaufgabe während der Heuristischen Evaluation von mehreren Experten als störend empfunden wurde; sie erfolgt mit Vorsicht, da bei dieser Methode keine Leistungs- oder Anstrengungsdaten erhoben werden.
- zu 3: Die Beurteilung beruht darauf, dass in der leichten Aufgabe zur Suche nach Informationen, die stärker perzeptive Leistungen verlangte, bei der summativen Evaluation ein negativer Zusammenhang zwischen multimodaler Interaktion und Leistung gefunden wurde und auch keine erhöhten Anstrengungswerte angegeben wurden. In der mittleren und schweren Aufgabe zur Informationssuche, in denen es mehr auf die Kategorisierung und Kodierung der Informationen ankam, wurde jedoch ein signifikant positiver Zusammenhang mit der multimodalen Interaktion gefunden. Die positive Einschätzung bezieht weiterhin die Nutzungshäufigkeit der Blickinteraktion bei diesen Aufgabentyp mit ein. Für die Konstruktionsaufgaben, die sowohl Problemlösen und Entscheiden als auch komplex-kontinuierliche motorische Anforderungen stellten, konnte ein positiver Zusammenhang zwischen der Leistung und der multimodalen Interaktion gefunden werden, allerdings war die Blickinteraktion dabei weniger bedeutsam als die Sprachinteraktion.
- zu 4: Die Beurteilung beruht auf der Untersuchung zur Sprachinteraktion, bei der die natürliche Sprache am wenigsten genutzt wurde, wenn nur textuelle Informationen ausgegeben wurden und am häufigsten, wenn das System verbal antwortete. Da die Nutzungsunterschiede jedoch nur deskriptiv, nicht inferenzstatistisch belegt werden konnten, werden diese Gestaltungsvarianten nur als „eher geeignet“ beurteilt. Die Bedingung, in der Text und Sprache angeboten wurden, wurde von den Testbenutzern bevorzugt, da der Text eine Erinnerungshilfe bot.
- zu 5: Die Beurteilung beruht darauf, dass die multimodale Interaktion die Leistung in den durch Problemlösen und Entscheiden sowie durch komplex-kontinuierliche motori-

sche Anforderungen geprägte Konstruktionsaufgaben unterstützt und die Sprachinteraktion bei diesen Aufgaben häufiger genutzt wurde als die Blickinteraktion. Die Sprachinteraktion wurde als eine hilfreiche Abkürzung von Interaktionswegen von 85% der Untersuchungsteilnehmer empfunden. Auch für die Aufgaben, die die Kategorisierung und Kodierung von Informationen bei der Informationssuche verlangten, wurde dieser positive Zusammenhang gefunden. Die Sprachinteraktion wurde dabei aber weniger oft benutzt als die Blickinteraktion, so dass sie nur tendenziell positiv beurteilt wurde. Für die perzeptiv geprägte Aufgabe der Informationssuche ergab sich ein negativer Zusammenhang zwischen Leistung und multimodaler Interaktion, so dass multimodale Interaktionsangebote bei solchen Aufgabentypen eher ungeeignet erscheinen. Die akustische Rückmeldung der Erkennung der Spracheingabe wurde von 96% der Untersuchungsteilnehmer als hilfreich empfunden, dürfte aber bei verbesserter Erkennungsrate und Verarbeitungsgeschwindigkeit überflüssig werden.

Tab. 23: Integration der Evaluationsergebnisse von mUltimo3D in die System-Aufgaben-Eigenschaftsmatrix (nähere Erläuterungen im Text)

Mensch>Maschine		Systemeigenschaften			Aufgabeneigenschaften				
		Maschine>Mensch	Perzeptive Prozesse	Vermittelnde Prozesse	Kommunikative Prozesse	Motorische Prozesse			
Modalitäten	Eingabe	wahrnehmbare Medien	Suche nach und Empfangen von Informationen	Identifizieren von Objekten, Handlungen und Ereignissen	Informationsverarbeitung	Problemlösen und Entscheiden	Komplex-kontinuierlich	Einfach-diskret	
	Blick	Rohdaten der Blickfassung	visuell	- - <sup>1</sup>	- - <sup>1</sup>				
			auditiv	ohne	+ + <sup>1</sup>				
			haptisch	Signalton		- <sup>2</sup>	- <sup>2</sup>		
	Blick	Parametrisierte Rohdaten der Blickfassung	visuell	Reaktion der Interface-Objekte		+ <sup>3</sup>	+ <sup>3</sup>	+ <sup>3</sup>	
			auditiv						
			haptisch						
	Sprache	Fließend gesprochene Sprache	visuell	Text				- <sup>4</sup>	
			auditiv	Text & Sprache				+ <sup>4</sup>	
			auditiv	Text & Sprache				+ <sup>4</sup>	
haptisch			Sprache				+ <sup>4</sup>		



Systemeigenschaften			Aufgabeneigenschaften			
Mensch>Maschine	Maschine>Mensch	Perzeptive Prozesse	Vermittelnde Prozesse	Kommunikative Prozesse	Motorische Prozesse	
Modalitäten	Eingabe	Suche nach und Empfangen von Informationen	Identifizieren von Objekten, Handlungen und Ereignissen	Informationsverarbeitung	Problemlösen und Entscheiden	Komplex-kontinuierlich
Sprache	Wahrnehmbare Medien	Ausgabe	Reaktion der Interface-Objekte	+ <sup>5</sup>	+ <sup>5</sup>	+ <sup>5</sup>
	visuell					
	auditiv	Akustisches Signal bei Erkennung		+ <sup>5</sup>	+ <sup>5</sup>	+ <sup>5</sup>
Gestik	haptisch					
	visuell					
	auditiv					
	haptisch					
	visuell					
	auditiv					
Gestik	haptisch					
	visuell					
	auditiv					
Ggf. weitere Modalitäten	haptisch					

## 7 Verallgemeinerbarkeit des Evaluationsmodells und Ausblick

Das Evaluationsmodell, das für die Bewertung von Designalternativen für multimodale Computersysteme für unterschiedliche Aufgabenmerkmale entwickelt wurde, hat sich in einer ersten Anwendung im Rahmen des Forschungsprojektes mUltimo3D bewährt. Es unterstützt die systematische nutzungsbezogene Erprobung von Gestaltungsalternativen unterschiedlicher Systemkomponenten und deren interaktionsbezogener Vernetzung innerhalb einer multimodalen Architektur. In die System-Aufgaben-Eigenschaftsmatrix lassen sich weitere Informationen aus einem iterativen Design- und Evaluationsprozess anderer Forschungsprojekte zur multimodalen Interaktion integrieren. Dabei ist zumindest die summative Evaluation an prototypischen Aufgaben ein unerlässlicher Bestandteil. Bewährt hat sich hierfür ein experimentelles Vorgehen, um quantitative Maße für die Effizienz und die Effektivität sowie die Zufriedenheit bzw. Akzeptanz des untersuchten Systems zu generieren, die die anschließende Beurteilung von Gestaltungsvarianten für spezifische Aufgabenmerkmale erleichtern. Die getesteten Aufgaben können mit Hilfe der Berliner Taxonomie und den darin enthaltenen Beschreibungskategorien für Aufgabenmerkmale plausibel zugeordnet werden. Die Systematisierung von Evaluationsergebnissen unter der Perspektive der Systemauslegung und der der Aufgabenmerkmale erweitert die Wissensbasis über multimodale Interaktion und kann dazu beitragen, Richtlinien oder Heuristiken für die multimodale Systemgestaltung zu generieren. Außerdem kann die Nützlichkeit verschiedener, die multimodale Interaktion unterstützender Systemkomponenten für unterschiedliche Aufgabenbereiche aus der System-Aufgaben-Eigenschaftsmatrix abgelesen werden. Weitere Arbeitsschritte sollten darin bestehen, Evaluationsergebnisse über andere Modalitäten, beispielsweise zur haptischen und gestischen Interaktion, und Evaluationsergebnisse aus weiteren empirischen Studien anderer Forschungsprojekte in die Matrix einzubringen. Die schrittweise Vervollständigung und Erweiterung der Matrix verhilft dazu, einen software-ergonomischen Wissenstand für die Gestaltung multimodaler Systeme zu erreichen, wie er für die traditionellen Bereiche der Mensch-Computer-Interaktion bereits vorliegt.

Wenn in Zukunft multimodale Systeme einen Standard in der Mensch-Computer-Interaktion darstellen, werden für ein spezifisches multimodales Computersystem, das Kundenanforderungen genügen muss, dennoch zusätzlich Gebrauchstauglichkeitsuntersuchungen notwendig sein. Die an Prototypen gesammelten Erfahrungen genügen nicht, um für ein in einem differenzierten Tätigkeitskontext eingesetztes System alle Gestaltungsanforderungen zu definieren. Ein Prozessmodell, das dem Designprozess in diesen produktorientierten Phasen dient, kann dann der in Kapitel 3.1 vorgestellte *usability life cycle* von Mayhew (1999) sein.

Zum gegenwärtigen Zeitpunkt richtet sich die forschungsbezogene Methodenauswahl für die Evaluation danach, inwiefern software-ergonomisches Wissen bereits bei der ersten Systemrealisierung eingeflossen ist und ab welcher Phase des Designprozesses hypothesengeleitete Untersuchungen von Gestaltungsvarianten möglich sind. Die Anzahl der Iterationen kann in

Abhängigkeit von den detektierten Gestaltungsschwachstellen und Entscheidungsindifferenzen bei der Auswahl spezifischer Gestaltungslösungen für eine Modalität oder dem Zusammenspiel verschiedener Modalitäten variieren. Die frühzeitige Einbeziehung softwareergonomischer und kognitionspsychologischer Kompetenz trägt dazu bei, den Evaluationszyklus dadurch zu verkürzen, dass zielgerichtete formative Evaluationsschritte den Redesignaufwand wesentlich verringern und eine erfolgreiche summative Evaluation vorbereiten. Oviatt und andere (2000) verweisen darauf, dass es für die Entwicklung neuer, an der natürlichen Kommunikation ausgerichteter Interaktionskonzepte erforderlich ist, multidisziplinär zu arbeiten und den Entwicklungsprozess stärker durch interaktive Tests zu unterstützen. Der Einsatz effizienter Testmethoden, die beispielsweise auf Systemsimulationen und automatisierte Datenaufbereitung zurückgreifen, verhelfen dazu, die Evaluation mit dem Designprozess zu verzahnen. Ein angemessenes Kosten-Nutzen-Verhältnis zwischen Aufwänden, die in Entwicklungsprojekten für die Evaluation entstehen, und den aus der Evaluation gewonnenen Erkenntnissen, kann mit Hilfe eines abgestimmten Projektmanagements erreicht werden. Die Restriktionen beim Design- und beim Evaluationsteam müssen in der Planung berücksichtigt und ein gegenseitiges Verständnis für die Tätigkeit der anderen Disziplinen muss geschaffen werden, um Synergieeffekte im interdisziplinären Entwicklungsteam zu erzielen. Die grundsätzliche Einsicht, die technologischen Zielstellungen nicht allein aus technischer Perspektive lösen zu können, bietet für die erfolgreiche interdisziplinäre Arbeit eine nützliche Grundlage.

Die Voraussetzung für die zufriedenstellende multimodale Interaktion besteht darin, dass die jeweiligen Interaktionsmodalitäten zuverlässig funktionieren. Das technische System sollte demnach in der Lage sein, die Benutzereingaben fehlerfrei unter verschiedenen Anwendungs- und Umweltbedingungen zu erkennen. Für die Interpretation von Benutzerintentionen, die aus der Integration der verschiedenen Verhaltensäußerungen vorgenommen wird, ist es notwendig, die Informationen aus den Modalitäten zu gewichten bezüglich ihrer Intentionalität, ihrer Eindeutigkeit, ihrer gegenseitigen Bezogenheit oder auch bezüglich der erwarteten Erkennungsrate. Hierfür ist weitere Forschung notwendig, die auf den Aufbau kognitiver Theorien über multimodale Interaktion abzielt und die Technologien für die semantische Integration von Benutzereingaben, für effektivere Verarbeitungsprozesse sowie eine verbesserte Fehlerbehandlung bereitstellt (Oviatt et al. 2000). Die Qualität multimodaler Mensch-Computer-Interaktion kann gleichermaßen dadurch gewinnen, dass leistungsfähigere Soft- und Hardware für robustere und adaptive multimodale Systeme entwickelt wird und dass humanwissenschaftliche Erkenntnisse über kognitive, soziale sowie kulturelle Faktoren mit in die Systemgestaltung einfließen. Die interdisziplinäre Betrachtung des Gegenstandes „multimodale Mensch-Computer-Interaktion“ ist nicht zuletzt dafür wichtig, die Anforderungen und Grenzen der zu entwickelnden Technik zu bestimmen, die nicht allein technischer Machbarkeit geschuldet sind, sondern auch den Bedürfnissen und Erwartungen der mit der Technik konfrontierten Menschen für ein nützlich Zusammenwirken entsprechen.

## 8 Literatur

- ACM SIGCHI (1992). *Curricula for Human Computer Interaction*. Verfügbar unter: <http://www.acm.org/sigchi/cdg/cdg1.html> (Zugriff am 2.12.2001, 19.00 Uhr).
- Ainsworth, W.A. (1999). Some Approaches to Automatic Speech Recognition. In W.J. Hardcastle & J. Laver (Eds.), *The Handbook of phonetic sciences*. Oxford, Malden: Blackwell.
- Akyol, S., Canzler, U., Bengler, K. & Hahn, W. (2000). Gestengesteuerter Nachrichtenspeicher im Kraftfahrzeug. In K.-P. Gärtner (Hrsg.), *Multimodale Interaktion im Bereich der Fahrzeug- und Prozessführung*, 42. *Fachausschusssitzung Anthropotechnik der Deutschen Gesellschaft für Luft- und Raumfahrt e.V.*, 24.-25.10.2000 München, *DGLR-Bericht 2000-02*. Bonn: DGLR.
- Aldridge, L.C. & Lansdown, T.C. (1999). Driver Preferences For Speech Based Interaction With In-Vehicle Systems. *Proceedings of the Human Factors and Ergonomics Society 43<sup>rd</sup> Annual Meeting*, 977-981.
- Alexandersson, J., Buschbeck-Wolf, B., Fujinami, T., Kipp, M., Koch, S., Maier, E., Reithinger, N., Schmitz, B. & Siegel, M. (1998). *Dialogue Acts in VERBMOBIL-2*. Verbmobil-Report 226. DFKI GmbH, Saarbrücken. Verfügbar unter: <http://www.dfki.de/~kipp/research/publications> (Zugriff am 27.2.2002, 17:55 Uhr).
- Ark, W., Dryer, C. & Lu, D.J. (1999). The Emotion Mouse. In H.-J. Bullinger & J. Ziegler (Eds.), *Proceedings of HCI International '99*, 818-823.
- Austin, J. L. (1962). *How to do things with words*. Oxford: Clarendon Press.
- Baggen, R. & Hemmerling, S. (2000). Evaluation von Benutzbarkeit in Mensch-Maschine-Systemen. In K. P. Timpe, T. Jürgensohn und H. Kolrep (Hrsg.), *Mensch-Maschine-Systemtechnik*. Düsseldorf: Symposion.
- Bahrtdt, H.P. (1984). *Schlüsselbegriffe der Soziologie*. München: Beck.
- Balzert, H. (1998). *Lehrbuch der Software-Technik*. Heidelberg, Berlin: Spektrum.
- Baumgarten, T. (2002). *Ableitung eines noninvasiven Indikators von mental workload für die Implementierung in einem adaptiven multimodalen System*. Unveröffentlichte Diplomarbeit am Institut für Psychologie der TU Berlin.
- Becker, J., Haulsen, I. & Rüsseler, H. (2001). Human Computer Interaction and Virtual Worlds. *ERCIM News*, 46, 27.
- Benington, H.D. (1956). Production of Large Computer Programs, *Proc. ONR Symposium on Advanced Programming Methods for Digital Computers*, 15-27.
- Berliner, D.C., Angell, D. & Shearer, J.W. (1964). *Behaviors, measures, and instruments for performance evaluation in simulated environments*. Paper presented at a symposium and workshop on the quantification of human performance, Albuquerque, New Mexico, August 1964.
- Bengler, K., Geutner, P., Niedermaier, B. & Steffens, F. (2000). „Eyes free – Hands free“ oder „Zeit der Stille“ – Ein Demonstrator zur multimodalen Bedienung im Automobil. In K.-P. Gärtner (Hrsg.), *Multimodale Interaktion im Bereich der Fahrzeug- und Prozessführung*, 42. *Fachausschusssitzung Anthropotechnik der Deut-*

*schen Gesellschaft für Luft- und Raumfahrt e.V.*, 24.-25.10.2000 München, *DGLR-Bericht 2000-02*. Bonn: DGLR.

- Bers, J., Miller, S. & Makhoul, J. (1998). Designing conversational interfaces with multimodal interaction. *DARPA Workshop on Broadcast News Understanding Systems*, 319-321.
- Bierwagen, T. & Vielhauer, J. (2000). Integration von Spracherkennung in die Mensch-Maschine-Interaktion am Fluglotsenarbeitsplatz. In K.-P. Gärtner (Hrsg.), *Multi-modale Interaktion im Bereich der Fahrzeug- und Prozessführung*, 42. *Fachaus-schusssitzung Anthropotechnik der Deutschen Gesellschaft für Luft- und Raum-fahrt e.V.*, 24.-25.10.2000 München, *DGLR-Bericht 2000-02*. Bonn: DGLR.
- Bimber, O. (1999). Gesture Controlled Object Interaction : A Virtual Table Case-Study. *Proceedings of 7<sup>th</sup> International Conference in Central Europe on Computer Graphics, Visualization and Interactive Digital Media'99*. Verfügbar unter: [http://wscg.zcu.cz/wscg99/wscg\\_99\\_program.htm](http://wscg.zcu.cz/wscg99/wscg_99_program.htm) (Zugriff am 19.12.2001, 14:25 Uhr)
- Blanchard, B.S. & Fabrycky, W.J. (1998). *Systems Engineering and Analysis*. Upper Saddle River, New Jersey: Prentice Hall.
- Boehm, B.W. (1981). *Software Engineering Economics*. Upper Saddle River, New Jersey: Prentice Hall.
- Boehm, B.W. (1986). A Spiral Modell of Software Development and Enhancement, *ACM SIGSOFT, Aug. 1986*, 14-24.
- Boehme, H.-J., Brakensiek, A., Braumann, U.-D., Krabbes, M., Gross, H.-M. (1998) Neural Networks for Gesture-based Remote Control of a Mobile Robot. *Proceedings of the World Congress on Computational Intelligence and IEEE International Joint Conference on Neural Networks*, 1, 372-377.
- Bolt, R.A. (1980). Put-that-there: Voice and gesture at the graphics interface. *Computer Graphics*, 14, 262-270.
- Bordegoni, M., Cugini, U., Mussio, P. & Matera, M. (2000). The Role of Continuity in Haptic Interaction Systems. *Workshop Continuity in Human Computer Interaction, April 2 and 3 , 2000, The Hague/Scheveningen, The Netherlands*. Verfügbar unter: <http://kazan.cnuce.cnr.it/TACIT/CHI2000/Contrib/monica.pdf> (Zugriff am 7.3.2002, 15:30 Uhr).
- Bortz, J. & Döring, N. (1995). *Forschungsmethoden und Evaluation*, 2. Auflage. Berlin, Heidelberg, New York: Springer.
- Brenner, W., Zarnekow, R. & Wittig, H. (1998). *Intelligente Software-Agenten*. Berlin, Heidelberg, New York: Springer.
- Burdea, G.C. (1996). *Force and Touch Feedback for Virtual Reality*. New York: Wiley.
- Burkhardt, F. (2000). *Simulation emotionaler Sprechweise mit Sprachsyntheseverfahren*. Dissertation am Fachbereich Kommunikations- und Geschichtswissenschaften an der Technischen Universität Berlin.
- Çakmak, H. K. & Kühnapfel, U. (2000). Animation and Simulation Techniques for VR-Training Systems in Endoscopic Surgery. *Eurographics Workshop on Animation and Simulation '2000 (EGCAS '2000)*, 173-185.

- Card, S., Moran, T. & Newell, A. (1983). *The psychology of human computer interaction*. Hillsdale: Lawrence Erlbaum.
- Caro, P.W. (1988). Flight Training and Simulation. In E.L. Wiener & D.C. Nagel (Eds.), *Human Factors in Aviation*. San Diego, New York, Boston: Academic Press.
- Chen, T. & Rao, R. (1998). Audio-visual integration in multimodal communication. *Proceedings of IEEE, Special Issue on Multimedia Signal Processing*, 837-852.
- Cheyer, A. & Julia, L. (1999a). Designing, Developing & Evaluating Multimodal Applications. *CHI'99, Pittsburgh (USA), workshop 13: Designing the User Interface for Pen and Speech Applications*. Verfügbar unter: <http://www.bravobrava.com/people/julia/articles/chi99.pdf> (Zugriff am 3.3.2002, 17:45 Uhr).
- Cheyer, A. & Julia, L. (1999b). InfoWiz: An Animated Voice Interactive Information System. *Agents'99 : Seattle (USA), workshop 6: Communicative Agents*. Verfügbar unter: <http://bravobrava.com/people/julia/articles/agents99.pdf> (Zugriff am 3.3.2002, 17:35 Uhr).
- Christensen, J.M. & Mills, R.G. (1967). What does the Operators do in Complex Systems. *Human Factors*, 9 (4), 329-340.
- Cohen, P.R., Johnston, M., McGee, D., Oviatt, S., Pittman, J., Smith, I., Chen, L. & Clow, J. (1997). Quickset: Multimodal interaction for distributed applications. *Proceedings of the Fifth ACM International Multimedia Conference*, 31-40.
- Cohn, J.F., Kanade, T., Moriyama, T., Ambadar, Z., Xiao, J., Gao, J. & Imamura, H. (2001). *A Comparative Study of Alternative FACS Coding Algorithms*. Technical Report CMU-RI-TR-02-06, Robotics Institute, Carnegie Mellon University.
- Conrad, W. (1983). Intelligenzdiagnostik. In K.-J. Groffmann (Hrsg.), *Intelligenz- und Leistungsdiagnostik, Enzyklopädie der Psychologie : Themenbereich B, Methodologie und Methoden: Ser. 2, Psychologische Diagnostik ; Bd. 2*. Göttingen u.a.: Hogrefe.
- Cutler, R. & Turk, M. (1998). View-based Interpretation of Real-time Optical Flow for Gesture Recognition. *Proceedings of the 3<sup>rd</sup> IEEE International Conference on Automatic Face and Gesture Recognition*, 416-421.
- Dahlbäck, N., Jönsson, A., and Ahrenberg, L. (1993). Wizard of Oz Studies - Why and How. *Proceedings of the 1993 International Workshop on Intelligent User Interfaces (IUI'93)*, 193-200.
- David, H. (1984). *Wörterbuch der Medizin*, 2.Bd. Berlin: Verlag Volk und Gesundheit.
- Denecke, M. (1997). An Information-based Approach for Guiding Multi-Modal Human-Computer-Interaction. *Proceedings of the 15<sup>th</sup> International Joint Conference on Artificial Intelligence, Nagoya, Japan, August 23-29, 1997*, 1036-1041.
- Detmer, W. M., Shiffman, S., Wyatt, J. C., Friedman, C. P., Lane, C. D. & Fagan, L. M. (1995). A Continuous-Speech Interface to a Decision-Support System: II. An Evaluation Using a Wizard-of-Oz Experiment. *Journal of the American Medical Informatics Association* 2 (1), 46-57.
- DIN EN ISO-Norm 9241-10 (1996-07). *Ergonomische Anforderungen für Bürotätigkeiten mit Bildschirmgeräten*, Teil 10: Grundsätze der Dialoggestaltung – Leitsatz. Berlin: Beuth.

- DIN EN ISO-Norm 9241-11 (1998). *Ergonomische Anforderungen für Bürotätigkeiten mit Bildschirmgeräten*, Teil 11: Anforderungen an die Gebrauchstauglichkeit - Leitsatz. Berlin: Beuth.
- Dix, A., Finlay, J., Abowd, G., & Beale, R. (1995). *Mensch Maschine Methodik*. München: Prentice Hall.
- DLR (2001). *Leitprojekte des BMBF zur Mensch-Technik-Interaktion*. Verfügbar unter: <http://www.dlr.de/IT/IV/MTI> (Zugriff am 27.02.2002, 19:55 Uhr).
- DLR (2002). *Mensch-Maschine-Interaktion*. Verfügbar unter: <http://www.robotic.dlr.de/MMI/> (Zugriff am 7.03.2002, 12:00 Uhr).
- Driver, J. Davis, G., Ricciardelli, P., Kidd, P., Maxwell, E. & Baron-Cohen, S. (1999). Gaze perception triggers reflexive visuospatial orientating. *Visual Cognition*, 6 (5), 509-540.
- Dryer, D.C. & Horowitz, L.M. (1997). When do opposites attract? Interpersonal complementary versus similarity. *Journal of Personality and Social Psychology*, 72, 592-603.
- Dunckel, H. (1999). *Handbuch psychologischer Arbeitsanalyseverfahren*. Zürich: vdf Hochschulverlag an der ETH.
- Dzida, W., Hofmann, B., Freitag, R. & Red, W. (2001). *Gebrauchstauglichkeit von Software ErgoNorm: Ein Verfahren zur Konformitätsprüfung von Software auf der Grundlage von DIN EN ISO 9241 Teile 10 und 11*. Bremerhaven: NW-Verlag.
- Ehrenmann, M., Lütticke, L. & Dillmann, R. (2001). Directing a Mobile Robot with Dynamic Gestures. *Proceedings of the 32<sup>nd</sup> International Symposium on Robotics* (CD-ROM). Verfügbar unter: <http://www.iain.ira.uka.de/publications.bdy.html#2001> (Zugriff am 14.2.2002, 17:25 Uhr).
- Eiff, W. (1991). Prozesse optimieren – Nutzen erschließen, *IBM Nachrichten*, 41, 23-27.
- Eilers, K., Nachreiner, F. & Hänecke, K. (1986). Entwicklung und Überprüfung einer Skala zur Erfassung subjektiv erlebter Anstrengung. *Zeitschrift für Arbeitswissenschaft*, 40, 215-224.
- Ekman, P. (1972). Universals and cultural differences in facial expressions of emotion. In J. Cole (Ed.), *Nebraska Symposium on Motivation*, Vol.19. Lincoln: University Nebraska Press.
- Ekman, P., & Friesen, W. V. (1978). *Facial action coding system: A technique for the measurement of facial movement*. Palo Alto, CA: Consulting Psychologists Press.
- Ekman, P. (1993). Facial Expression and Emotion. *American Psychologist*, 48 (4), 384-392.
- Ekstrom, R.B., French, J.W. & Harman, H.H. (1979). Cognitive Factors: Their Identification And Replication. *Multivariate Behavioral Research Monographs*, No.79-2, 8-84.
- EMBASSI (2001). *Elektronische Multimediale Bedien- und Service-Assistenz*. White Draft. Verfügbar unter: <http://www.embassi.de> (Zugriff am 2.2.2002, 10:15 Uhr).
- Fleishman, E.A. & Quaintance, M.K. (1984). *Taxonomies of Human Performance – The Description of Human Tasks*. Orlando u.a.: Academic Press.
- Floyd, C. (1984). A Systematic Look At Prototyping. In R. Budde, K. Kuhlenkamp, L. Mathiassen & H. Züllighoven (Hrsg.), *Approaches to Prototyping*. Berlin, Heidelberg, New York: Springer.

- Floyd, J. (1999). Haptic Interaction with Three-Dimensional Bitmapped Virtual Environments. *The Fourth PHANTOM Users Group Workshop (PUG99), October 9-12, 1999, Dedham MA, Massachusetts Institute of Technology, Cambridge, MA*. Verfügbar unter: <http://vorlon.mit.edu/~jered/publications/floyd-pug99.pdf> (Zugriff am 2.3.2002, 14:24 Uhr).
- Frieling, E., Facaoaru, C., Benedix, J., Pfaus, H., Sonntag, Kh. (1993). *TAI Tätigkeitsanalyseinventar – Theorie, Auswertung, Praxis, Handbuch und Verfahren*. Landsberg: ecomed.
- Friesen, C.K. & Kingstone, A. (1998). The eyes have it!: Reflexive orienting is triggered by nonpredictive gaze. *Psychonomic Bulletin*, 5, 490-495.
- Funke, J. & Vaterrodt-Plünnecke, B. (1998). *Was ist Intelligenz?* Nördlingen: C.H. Beck.
- Gamma, E., Helm, R., Johnson, R. & Vlissides, J. (1993). Design Patterns: Abstraction and Reuse of Object-Oriented Design. *Proceedings of the ECOOP'93, LNCS 707*, 406-431.
- Gediga, G. & Hamborg, K.-C. (1999). IsoMetrics: Ein Verfahren zur Evaluation von Software nach ISO 9241/10. In H. Holling & G. Gediga (Hrsg.), *Evaluation*. Göttingen u.a.: Hogrefe.
- Geiser, G. (1990). *Mensch-Maschine-Kommunikation*. München, Wien: Oldenbourg.
- Gibson, J.J. (1979). *The Ecological Approach to Visual Perception*. Boston: Houghton-Mifflin.
- Giesa, H.-G. & Köhler, T. (2001). Application and further development of CREAM exemplified by operation and control of a chemical process. In R. Trimpop, B. Zimolong & H.-W. Faulenbach (Hrsg.), *Psychologie der Arbeitssicherheit*. Heidelberg: Asanger.
- Goldberg, J. H., & Schryver, J. C. (1995). Eye-gaze determination of user intent at the computer interface. In J. M. Findiay, R. Walker, & R. W. Kentridge (Eds.), *Eye Movement Research: Mechanisms, Processes, and Applications*. New York: Elsevier Science.
- Goldschen, A.J., Garcia, O.N. & Petajan, E. (1995). Continuous Optical Automatic Speech Recognition by Lipreading. *Proceedings of 28<sup>th</sup> Annual Asilomar Conference on Signals, Systems, and Computers*, 572—577.
- Goldschen, A.J. (1997). Modality Integration: Facial Movement & Speech Recognition. In R.A. Cole, J. Mariani, H. Uszkoreit, A. Zaenen & V. Zue (Eds.), *Survey of the State of the Art in Human Language Technology*. New York: Cambridge University Press.
- Grasso, M.A., Ebert, D.S. & Finin, T.W. (1998). The Integrality of Speech in Multimodal Interfaces. *ACM Transactions on Human-Computer Interaction*, 5 (4), 303-325.
- Green, D.M. & Swets, J. A. (1966). *Signal detection theory and psychophysics*. New York: Wiley.
- Greenstein, J.S. (1997). Pointing Devices. In M. Helander, T.K. Landauer & P.V. Prabhu (Eds.), *Handbook of Human-Computer Interaction*. Amsterdam u.a.: Elsevier.
- Gullberg, M. & Holmqvist, K. (1999). Keeping an eye on gestures: Visual perception of gestures I face-to-face communication. *Pragmatics & Cognition*, 7 (1), 35-63.



- Habermas, J. (1981). *Theorie des kommunikativen Handelns*. Frankfurt a.M.: Suhrkamp.
- Hacker, W. & Richter, P. (1984). *Psychologische Bewertung von Arbeitsgestaltungsmaßnahmen*. Berlin: Springer.
- Hacker, W. (1986). *Arbeitspsychologie – Psychische Regulation von Arbeitstätigkeiten*. Bern, Stuttgart, Toronto: Hans Huber.
- Hacker, W. (1995). *Arbeitstätigkeitsanalyse*. Heidelberg: Asanger.
- Hart, S.G. & Staveland, L.E. (1988). Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. In P.A. Hancock & N. Meshkati (Eds.), *Human Mental Workload*. Amsterdam: Elsevier.
- Hassenzahl, M. (to appear). The Effect of Perceived Hedonic Quality on Product Appealingness. *International Journal of Human-Computer Interaction*.
- Hassenzahl, M., Platz, A., Burmester, M. & Lehner, K. (2000). Hedonic and ergonomic quality aspects determine a software's appeal. *Proceedings of the CHI 2000 Conference on Human Factors in Computing Systems* (201-208). New York: ACM Press, Addison Wesley.
- Hedicke, V. (2000). Multimodalität in Mensch-Maschine-Schnittstellen. In K. P. Timpe, T. Jürgensohn und H. Kolrep (Hrsg.), *Mensch-Maschine-Systemtechnik*. Düsseldorf: Symposium.
- Heeman, P.A., Johnston, M., Denney, J. & Kaiser, E. (1998). Beyond Structured Dialogues: Factoring Out Grounding. *Proceedings of the 5<sup>th</sup> International Conference on Spoken Language Processing (ICSLP-98)*, 863-866.
- Herrmann, T. & Hoppe-Graff, S. (1988). Textproduktion. In H. Mandl & H. Spada (Hrsg.), *Wissenspsychologie*. München, Weinheim: PVU.
- Herrmann, T. (1992). Sprechen und Sprachverstehen. In H. Spada (Hrsg.), *Lehrbuch der Allgemeinen Psychologie*. Bern u.a.: Huber.
- Herrmann, T. (1995). *Allgemeine Sprachpsychologie*, 2. Auflage. Weinheim: PVU.
- Hickmann, M. (2000). Pragmatische Entwicklung. In H. Grimm (Hrsg.), *Enzyklopädie der Psychologie. Sprache, Bd. 3: Sprachentwicklung*. Göttingen: Hogrefe.
- Hollnagel, E. (1998). *Cognitive Reliability and Error Analysis Method – Cream*. Oxford: Elsevier Science.
- Horn, W. (1962). *Leistungsprüfsystem - (LPS)*. Göttingen: Hogrefe.
- House, E.R. (1993). *Professional Evaluation: Social Impact and Political Consequences*. San Francisco: Sage Publication.
- INVITE (2001). *Intuitive Mensch-Technik-Interaktion für die vernetzte Informationswelt der Zukunft*. Verfügbar unter: <http://invite.de/> (Zugriff am 9.3.2002, 17:44 Uhr).
- Jacob, R.J.K. (1995). Eye tracking in advanced interface design. In W. Barfield & T.A. Furness (Eds.), *Virtual environments and advanced interface design*. New York, Oxford: Oxford University Press.
- Jansson, G. (1999). The Importance of Available Exploration Methods for the Efficiency of Haptic Displays. *Symposium on Multimodal Communication, Linköping, Sweden, Oct 15-16, 1999*. Verfügbar unter: <http://www.ida.liu.se/~ssomc/papers/Jansson.pdf> (Zugriff am 27.2.2002, 17:20 Uhr).

- Jörding, T. & Wachsmuth, I. (2001). An Anthropomorphic Agent for the Use of Spatial Language. In K.R. Coventry & P. Olivier (Eds.), *Spatial Language: Cognitive and Computational Aspects*. Dordrecht: Kluwer.
- Jung, B., Kopp, S., Latoschik, M.E., Sowa, T. & Wachsmuth, I. (2000). *Virtuelles Konstruieren mit Gestik und Sprache*. *Künstliche Intelligenz*, 2, 2-11.
- Kamm, C. & Helander, M. (1997). Design Issues for Interfaces using Voice Input. In M. Helander, T.K. Landauer & P.V. Prabhu (Eds.), *Handbook of Human-Computer Interaction*. Amsterdam u.a.: Elsevier.
- Kantner, L. (1997). *Position Paper at the CHI'97 Workshop "Usability Testing World Wide Web Sites", March 23.-24., 1997, Atlanta, USA*. Verfügbar unter: <http://www.acm.org/sigchi/web/chi97testing/kantner.htm> (Zugriff 13.3.2002, 9:25 Uhr)
- Keates, S. & Robinson, P. (1999). Gestures and multimodal input. *Behaviour & Information Technology*, 18 (1), 36-44.
- Kieras, D. (1988): Towards a practical GOMS model technology for user interface design. In M. Helander (Ed.), *Handbook of Human-Computer Interaction*. Amsterdam: North Holland.
- Koch, U. & Wittmann, W.W. (1990). *Evaluationsforschung. Bewertungsgrundlage von Sozial- und Gesundheitsprogrammen*. Heidelberg: Springer.
- Kohler, M. (2001). *ZYKLOP – Visual Human-Computer Interaction through Hand Gestures*. Verfügbar unter: <http://ls7-www.cs.uni-dortmund.de/research/gesture/zyklop/> (Zugriff 19.12.2001, 17:05 Uhr).
- Kopp, S. & Wachsmuth, I. (2000). Planning and Motion Control in Lifelike Gesture: A Refined Approach. *Proceedings of Computer Animation*, 92-97.
- Krauss, R.M. & Hadar, U. (1999). The Role of Speech Related Arm/Hand Gestures in Word Retrieval. In L.S. Messing & R. Campbell (Eds.), *Gesture, Speech, and Sign*. Oxford: Oxford University Press.
- Kurze, M. (1999). Tguide: a guidance system for tactile image exploration. *Behaviour & Information Technology*, 18 (1), 11-17.
- Landauer, T.K. (1988). Research Methods in Human-Computer Interaction. In M. Helander (Ed.), *Handbook of Human-Computer Interaction*. Amsterdam: North Holland, Elsevier.
- Langton, S.R.H. (2000). The mutual influence of gaze and head orientation in the analysis of social attention direction. *The Quarterly Journal of Experimental Psychology*, 53A (3), 825-845.
- Langton, S.R.H., Roger, J.W. & Bruce, V. (2000). Do the eyes have it? Cues to the direction of social attention. *Trends in Cognitive Sciences*, 4 (2), 50-59.
- Lenzmann, B. (1998). *Benutzeradaptive und multimodale Interface-Agenten*. Sankt Augustin: infix.
- Lewin, K. (1926). Vorsatz, Wille und Bedürfnis. *Psychologische Forschung*, 7, 330-385.
- Lewis, C. and Wharton, C. (1997). Cognitive walkthroughs. In M. Helander, T. K. Landauer, & P. Prabhu (Eds.), *Handbook of Human-Computer Interaction*. Amsterdam u.a.: Elsevier.

- Lienert, G.A. & Ratz, U. (1994). *Testaufbau und Testanalyse*, 5. Auflage. Weinheim: Beltz, PVU.
- Liu, J. & Pastoor, S. (1999). Augenmaus & Co – Entwicklung einer neuartigen Anwendungsschnittstelle für Multimedia-Computer. In M. Rötting & K. Seifert (Hrsg.), *Blickbewegungen in der Mensch-Maschine-Systemtechnik*. Sinzheim: Pro Universitate.
- Liu, J., Pastoor, S., Seifert, K. & Hurtienne, J. (2000). Three-dimensional PC: toward novel forms of human-computer interaction, *Critical Review*, vol. CR76, *SPIE Photonics East*, Boston, 2000.
- Luczak, H. (1997). *Menschengerechte Arbeitsplätze durch Spracheingabe (MASEA)*. Aachen: Shaker.
- Lueth, T.C., Laengle, T., Herzog, G., Stopp, E. & Rembold, U. (1994). *KANTRA: Human-Machine Interaction for Intelligent Robots Using Natural Language*. Sonderforschungsbereich 314 – Bericht 104.
- Luettin, J., Thacker, N.A. & Beet, S.W. (1996). Active Shape Models for Visual Speech Feature Extraction. In D.G. Storck & M.E. Hennecke (Eds.), *Speechreading by Humans and Machines*. Berlin, Heidelberg: Springer.
- Maglio, P.P., Matlock, T., Campbell, C.S., Zhai, S. & Smith, B.A. (2000). Gaze and Speech in Attentive User Interfaces. *Lecture Notes in Computer Sciences*, 1948, 1-7.
- Majonica, B. (1996). *Evaluation eines Informations-Systems für die Unterstützung von Instandhaltungsaufgaben*. Münster: Waxmann.
- MAP (2001). *Multimedia Arbeitsplatz der Zukunft*. Verfügbar unter: [www.map21.de](http://www.map21.de) (Zugriff am 25.2.2002, 19:12 Uhr).
- Marrenbach, J. (2000). Konzept eines Werkzeugs zur formalen und empirischen Evaluierung der Gebrauchstauglichkeit interaktiver Endgeräte. In K.-P. Timpe, H.-P. Willumeit, H. Kolrep (Hrsg.), *Bewertung von Mensch-Maschine-Systemen*. Düsseldorf: VDI.
- Massaro, D.W. (1998). *Perceiving Talking Faces: From Speech Perception to a Behavioral Principle*. Cambridge, MA.: MIT Press.
- Massie, T.H. & Salisbury, J.K. (1994). The PHANTOM Haptic Interface: A Device for Probing Virtual Objects. *Proceedings of the ASME Dynamics Systems and Control*, 1, 295-301.
- Matin, E. (1974). Saccadic suppression: A review and an analysis. *Psychological Bulletin*, 81, 899-917.
- Mayhew, D.J. (1992). *Principles and Guidelines in Software User Interface Design*. Englewood Cliffs, NJ: Prentice Hall.
- Mayhew, D.J. (1999). *The Usability Engineering Lifecycle*. San Francisco: Morgan Kaufmann.
- MacDonald, J. & McGurk, H. (1978). Visual influences on speech perception process. *Perception and Psychophysics*, 24, 253-257.
- McNeill, D. & Levy, E. (1982). *Conceptual Representations in Language Activity and Gesture*. New York: Wiley.
- McNeill, D. (1992). *Hand and Mind: What Gestures Reveal about Thought*. Chicago, London: The University of Chicago Press.

- Merton, R.K. & Kendall, P.L. (1979). Das fokussierte Interview. In C. Hopf & E. Weingarten (Hrsg.), *Qualitative Sozialforschung*. Stuttgart: Klett-Cotta.
- Miller, R.B. (1962). Task description and analysis. In R.M. Gagne (Ed.), *Psychological principle in system development*. New York: Wiley.
- Miller, R.B. (1966). Task taxonomy: science or technology. In W.T. Singleton, R.S. Easterby & D.C. Whiffield (Eds.), *The human operator in complex systems*. London: Taylor & Francis.
- Molich, R. & Nielsen, J. (1990). Improving a human-computer dialogue. *Communications of the ACM* 33, 3, 338-348.
- Monk, A., Wright, P., Haber, J. & Davenport, L. (1993). *Improving your human computer interface: a practical approach*. Hemel Hempstead: Prentice Hall.
- Monk, A.F. & Gilbert, N. (1995). *Perspectives on HCI – Diverse Approaches*. London u.a.: Academic Press.
- Morley, S., Petrie, H., O'Neill, A.-M. & McNally, P. (1999). Auditory navigation in hyper-space: design and evaluation of a non-visual hypermedia system for blind users. *Behaviour & Information Technology*, 18 (1), 18-26.
- Morpha (2001). *Intelligente anthropomorphe Assistenzsysteme*. Verfügbar unter: <http://www.morpha.de> (Zugriff am 25.2.2002, 18:55 Uhr).
- Mulder, A. (1996). *Hand Gestures for HCI*. Technical Report 96-1. Verfügbar unter: <http://www.cs-sfi-ca/people/ResearchStaff/amulder/personal/vmi/HCI-gestures.html> (Zugriff am 28.09.2000, 14:12 Uhr)
- Müller, T., Giesa, H.-G. & Anders, G. (1999). Entwicklung und Evaluierung einer bordseitigen Benutzungsoberfläche für Data-Link-Kommunikation zwischen Piloten und Fluglotsen. *MMI Interaktiv*, [http://www.mmi-interaktiv.de/ausgaben/10\\_99/mueller.pdf](http://www.mmi-interaktiv.de/ausgaben/10_99/mueller.pdf).
- Naur, P. (1960). Revised Report on the Algorithmic Language ALGOL 60. *Communications of the ACM*, 3 (5), 299-314.
- Neumann, O. (1992). Theorien der Aufmerksamkeit: von Metaphern zu Mechanismen. *Psychologische Rundschau*, 43, 83-101.
- Nielsen, J. (1993). *Usability Engineering*. Boston: Academic Press.
- Nielsen, J. (1994). Heuristic Evaluation. In J. Nielsen & R.L. Mack (Eds.), *Usability Inspection Methods*. New York: Wiley.
- Nigay, L. & Coutaz, J. (1993). A Design Space For Multimodal Systems: Concurrent Processing and Data Fusion. *Proceedings of the INTERCHI'93*, 172-178.
- Norman, D.A. (1988). *The Design of Everyday Things*. New York: Doubleday/Currency.
- Norman, D.A. (1993). *Things that make us Smart*. Reading u.a.: Addison-Wesley.
- Norman, D.A. (1998). *The invisible Computer*. Cambridge, London: MIT Press.
- O'Shaughnessy, D. (2000). *Speech Communications – Human and Machine*, 2<sup>nd</sup> ed.. Piscataway, NJ: IEEE Press.
- O'Sullivan, C., Janott, M., Watson, M., Dingliana, J. (2000). Level of Detail Control for Real-time Computer Graphics and Virtual Reality: Applications of Eye-tracking. *Abstracts, The First Irish Workshop on Eye-Tracking, May 2000, Trinity College*

Dublin, Ireland. Verfügbar unter: <http://isg.cs.tcd.ie/iwet/Abstract14.htm> (Zugriff am 12.02.2002, 20.17 Uhr).

- O'Sullivan, C. & Dingliana, J. (2001). Collisions and Perception, *ACM Transactions on Graphics*, to appear.
- Opperman, R., Murchner, B., Reiterer, H. & Koch, M. (1992). *Software-ergonomische Evaluation. Der Leitfaden EVADIS II.*, 2. Auflage. Berlin New, York: Walter de Gruyter.
- Opwis, K. (1988). Produktionssysteme. In H. Mandl & H. Spada (Hrsg.), *Wissenspsychologie*. München, Weinheim: PVU.
- Osgood, C.E., Suci, G.J. & Tannenbaum, D.H. (1957). *The Measurement of Meaning*. Urbana, Ill.: University of Illinois Press.
- Oviatt, S.L. & Olsen, E. (1994). Integration themes in multimodal human-computer interaction. In K. Shirai, S. Furui, & K. Kakehi (Eds.), *Proceedings of the International Conference on Spoken Language Processing* (Vol. 2). Yokohama, Japan: Acoustical Society of Japan.
- Oviatt, S.L. (1997). Multimodal interactive maps: Designing for human performance. *Human-Computer Interaction*, 12, 93-129.
- Oviatt, S.L., DeAngeli, A. & Kuhn, K. (1997). Integration and synchronization of input modes during multimodal human-computer interaction. *Proceedings of Conference on Human Factors in Computing Systems (CHI'97)*, 415-422.
- Oviatt, S.L. & Wahlster, W. (1997). Introduction to This Special Issue on Multimodal Interfaces, *Human-Computer Interaction*, 12, 1-5.
- Oviatt, S.L. (1998). User-centered modeling for spoken language and multimodal interfaces. In M.T. Maybury & W. Wahlster (Eds.), *Readings in intelligent user interfaces*. San Francisco: Morgan Kaufmann.
- Oviatt, S.L. (1999). Ten myth of multimodal interaction. *Communications of the ACM*, 42 (11), 74-81.
- Oviatt, S.L., Cohen, P., Wu, L., Vergo, J., Duncan, L., Suhm, B., Bers, J., Holzman, T., Winograd, T., Landay, J., Larson, J. & Ferro, D. (2000). Designing the User Interface for Multimodal Speech and Pen-Based Gesture Applications: State-of-the-Art Systems and Future Research Directions. *Human-Computer Interaction*, 15, 263-322.
- Pahl, G. & Beitz, W. (1997). *Konstruktionslehre*, 4.Auflage. Berlin, Heidelberg: Springer.
- Palm Inc. (2002). *Palm Handhelds™*. Verfügbar unter: <http://www.palm.com/products/> (Zugriff am 14.4.2002, 10:30 Uhr).
- Pastoor, S. & Skerjanc, R. (1997). 3D Displays: Perspektiven für die Neugestaltung der Mensch-Computer-Interaktion (Eingeladener Beitrag). *Konferenzband der 12. Electronic Displays '97*, 120-126.
- Pateau, M. (1990). *Mensch-Maschine-Kommunikation*. Frankfurt a.M., New York: Campus.
- Pavlović, V.I., Sharma, R. & Huang, T.S. (1997). Visual Interpretation of Hand Gestures for Human-Computer Interaction: A Review. *IEEE Transaction on Pattern Analysis and Machine Intelligence*, 19 (7), 677-695.
- Pentland, A. (2000). Perceptual Intelligence. *Communications of the ACM*, 43 (4), 35-44.

- Picard, R.W. (1997). *Affective Computing*. Cambridge, MA.: MIT Press.
- Potts, A. (2000). Phantom-Based Haptic Interaction. *Proceedings of the Computer Science Discipline Seminar Conference (CSCI 3901)*. Verfügbar unter: <http://cda.mrs.umn.edu/~lopezdr/seminar/spring2000/potts.pdf> (Zugriff am 27.2.2002, 10:30 Uhr).
- Preece, J., Rogers, Y., Sharp, H., Benyon, D., Holland, S. & Carey, T. (1994). *Human-Computer-Interaction*. Wokingham, UK: Addison-Wesley.
- Preim, B. (1999). *Entwicklung interaktiver Systeme*. Berlin, Heidelberg: Springer.
- Press, W. H., Flannery, B. P., Teukolsky, S. A., & Vetterling, W. T. (1989). *Numerical Recipes in Pascal*. Cambridge: Cambridge University Press.
- Rabiner, L. & Juang, B.H. (1993). *Fundamentals of Speech Recognition*. Englewood Cliffs, NJ: Prentice Hall.
- Rabideau, G.F. (1964). Field measurement of human performance in man-machine systems. *Human Factors*, 6 (6), 663-672.
- Rasmussen, J., Pejtersen, A.M. & Goodstein, L.P. (1994). *Cognitive Systems engineering*. New York: Wiley.
- Revels, A.R., Kancler, D.E., Quill, L.L. & Donahoo, C.D. (2001). Determining voice strategies for hands-free human-computer interaction. *Proceedings of the Human Factors And Ergonomics Society 45<sup>th</sup> Annual Meeting*, 1704-1708.
- Rezagholi, M. (2000). Software-Entwicklungsprozesse als Evaluationsobjekt. In L.J. Heinrich & I. Häntschel (Hrsg.), *Evaluation und Evaluationsforschung in der Wirtschaftsinformatik*. München, Wien: Oldenbourg.
- Rohmert, W. & Landau, K. (1983). *A new technique for job analysis*. London: Taylor & Francis.
- Rötting, M. (1999). Methoden zur Registrierung von Augenbewegungen. In M. Rötting & K. Seifert (Hrsg.), *Blickbewegungen in der Mensch-Maschine-Systemtechnik*. Sinzheim: Pro Universitate.
- Rötting, M. (2001). *Parametersystematik der Augen- und Blickbewegungen für arbeitswissenschaftliche Untersuchungen*. Aachen: Shaker.
- Roy, D. & Pentland, A. (1997). *Multimodal Adaptive Interfaces*. Vision and Modeling Technical Report #438, MIT Media Lab. Verfügbar unter: [http://www-white.media.mit.edu/cgi-bin/tr\\_pagemaker/](http://www-white.media.mit.edu/cgi-bin/tr_pagemaker/) (Zugriff am 3.3.2002, 15:00 Uhr).
- Royce, W.W. (1970). Managing the development of large software systems: Concepts and Techniques. *Proceedings of the IEEE WESCON*, Technical Papers, A/1-1- A/1-9.
- Ruske, G. (1994). *Automatische Spracherkennung: Methoden der Klassifikation und Merkmalsextraktion*, 2. Auflage. München, Wien: R. Oldenbourg.
- Salvucci, D.D. & Anderson, J.R. (2000). Intelligent Gaze-Added Interfaces. *CHI letters*, 2 (1), 273-280.
- Sanders, A.F. (1983). Towards a model of stress and human performance. *Acta Psychologica*, 53, 61-97.
- Sarris, V. (1992). *Methodologische Grundlagen der Experimentalpsychologie*, 2. Bd. München, Basel: E. Reinhardt.

- Saslow, M. G. (1967). Effects of components of displacement-step stimuli upon latency of saccadic eye movement. *Journal of the Optical Society of America*, 57, 1024-1029.
- Scheele, B. & Groeben, N. (1988). *Dialog-Konsens-Methoden zur Rekonstruktion Subjektiver Theorien*. Tübingen: Francke.
- Schlick, C. (2000). *Modellbasierte Gestaltung der Benutzungsschnittstelle autonomer Produktionszellen*. Aachen: Shaker-Verlag.
- Schomaker, L., Nijtmans, J., Camurri, A., Lavagetto, F., Morasso, P., Benoît, C., Guiard-Marigny, T., Le Goff, B., Robert-Ribes, J., Adjoudani, A., Defée, I., Münch, S., Hartung, K. & Blauert, J. (1995). *A Taxonomy of Multimodal Interaction in the Human Information Processing System*. A Report of the ESPRIT Project 8579 MIAMI. Verfügbar unter: <http://hwr.nici.kun.nl/~miami/taxonomy/taxonomy.html> (Zugriff am 10.3.2002, 15:47 Uhr).
- Schönplflug, W. (1987). Beanspruchung und Belastung bei der Arbeit - Konzepte und Theorien. In U. Kleinbeck & J. Rutenfranz (Hrsg.), *Enzyklopädie der Psychologie. Arbeitspsychologie, Bd. 1*. Göttingen: Hogrefe.
- Schukat-Talamazzini, E.G. (1995). *Automatische Spracherkennung*. Wiesbaden: Vieweg.
- Schulmeister, R. (1997). *Grundlagen hypermedialen Lernsysteme*, 2. Auflage. München, Wien: Oldenbourg.
- Searle, J. (1969). *Speech Acts: An Essay in the Philosophy of Language*. New York, Cambridge: University Press.
- SensAble Technologies (2001). *The PHANTOM™ hardware device*. Verfügbar unter: <http://www.sensable.com/haptics/products.html> (Zugriff am 30.4.2002, 13:15 Uhr)
- Sheridan, T.B. (1997). Task Analysis, Task Allocation and Supervisory Control. In M. Helander, T.K. Landauer & P.V. Prabhu (Eds.), *Handbook of Human-Computer Interaction*. Amsterdam u.a.: Elsevier.
- Sherra, J., Gong, S., Howell, J. and Buxton, H. (2000). Interpretation of group behaviour in visually mediated interaction. *Proceedings of the 15<sup>th</sup> International Conference on Pattern Recognition*, 266-269.
- Shneiderman, B. (1992). *Designing the User Interface. Strategies for Effective Human-Computer Interaction*, 2. Auflage. Reading, MA u.a.: Addison-Wesley.
- SmartKom (2001). *Dialogische Mensch-Technik- Interaktion durch koordinierte Analyse und Generierung multipler Modalitäten*. Verfügbar unter: <http://smartkom.dfki.de/> (Zugriff am 10.3.2002, 9:24 Uhr).
- Smith, S.L. and Mosier, J.N. (1986). *Guidelines for designing user interface software*. Technical Report MTR-10090. Bedford MA: MITRE Corporation.
- Smith, B.A., Ho, J., Ark, W., Zhai, S. (2000). Hand Eye Coordination Patterns in Target Selection. *Proceedings of the ACM Symposium on Eye Tracking Research and Applications*, 117 - 122.
- Soojung-Kim Pang, A. (2002). The making of the mouse. *American Heritage of Invention & Technology*, 17 (3), 48-54.

- Sowa, T. & Wachsmuth, I. (2000). *Coverbal Iconic Gestures for Object Descriptors in Virtual Environments: An Empirical Study*. Technical Report. Verfügbar unter: <http://www.TechFak.Uni-Bielefeld.DE/~tsowa/download/Porto.pdf> (Zugriff am 10.2.2002, 8:30 Uhr).
- Spiegel, M.F. & Streeter, L. (1997). Applying Speech Synthesis to User Interfaces. In M. Helander, T.K. Landauer & P.V. Prabhu (Eds.), *Handbook of Human-Computer Interaction*. Amsterdam u.a.: Elsevier.
- Steininger, S. (2000). Transliteration of Language and Labeling of Emotion and Gestures in SmartKom. *Workshop Proc. of the Second International Conference on Language Resources and Evaluation: Meta-Descriptions and Annotation Schemes for Multimodal/Multimedia Language Resources*, 49-51.
- Steininger, S., Schiel, F. & Louka, K. (2001). Gestures During Overlapping Speech in multimodal Human-Machine Dialogues. *International Workshop on Information Presentation and Natural Multimodal Dialogue 2001, Verona, Italy*. Verfügbar unter: <http://www.phonetik.uni-muenchen.de/Publications/Steininger-Verona-01.pdf> (Zugriff am 1.3.2002, 9:30 Uhr).
- Stolcke, A., Shriberg, E., Bates, R., Coccaro, N., Jurafsky, D., Martin, R. Meteer, M., Ries, K., Taylor, P. & Van Ess-Dykema, C. (1998). Dialog Act Modeling for Conversational Speech. *Papers from the 1998 AAAI Spring Symposium, Technical Report SS-98-01*, 98-105.
- Stufflebeam, D.L. (1972). Evaluation als Entscheidungshilfe. In C. Wulf (Hrsg.), *Evaluation*. München: Piper.
- Thaller, G.E. (1998). *SPICE – ISO 9001 und Software in der Zukunft*. Kaarst: bhv.
- Timpe, K.-P. & Kolrep, H. (2000). Das Mensch-Maschine-System als interdisziplinärer Gegenstand. In K. P. Timpe, T. Jürgensohn und H. Kolrep (Hrsg.), *Mensch-Maschine-Systemtechnik*. Düsseldorf: Symposion.
- TU München (2002). *SFB 453 Wirklichkeitsnahe Telepräsenz und Teleaktion*. Verfügbar unter: <http://www.lsr.ei.tum.de/sfb453/> (Zugriff am 20.2.2002, 16:15 Uhr).
- Turk, M. & Robertson, G. (2000). Perceptual user interfaces. *Communications of the ACM*, 43 (3), 33-34.
- Ulich, E. (1994). *Arbeitspsychologie* (3. Auflage). Stuttgart: Schäfer-Poeschel.
- Unema, P., Rötting, M., Sepher-Willeberg, M., Strümpfel, U. & Kopp, U. (1988). Der NASA Task Load Index: Erste Ergebnisse mit der deutschen Fassung. In Gesellschaft für Arbeitswissenschaft e.V. (Hrsg.). *Jahresdokumentation 1988 der Gesellschaft für Arbeitswissenschaft e.V. - Bericht zum 34. Arbeitswissenschaftlichen Kongress an der RWTH Aachen*. Köln: O. Schmidt.
- Vicente, K. & Rasmussen, J. (1992). Ecological Interface Design: Theoretical Foundations, *IEEE Transactions on Systems, Man, and Cybernetics*, 22 (4), 589-606.
- Virzi, R.A., Sokolov, J.L. & Karis, D. (1996). Usability Problem Identification Using Both Low- and High Fidelity Prototypes. *Proceedings of the ACM Conference on Human Factors in Computing Systems* (236-243). New York: ACM Press.
- Vo, M.T. & Waibel, A. (1993). *Multimodal Human-Computer-Interaction*. Verfügbar unter: <http://www.is.cs.cmu.edu/papers/multimodal/93.issd.ps.gz>. (Zugriff am 28.11.2001, 19:17 Uhr).



- Vo, M.T., Houghton, R., Yang, J., Bub, U., Meier, U., Waibel, A. & Duchnowski, P. (1995). Multimodal Learning Interfaces. *Proceedings of the ARPA Spoken Language Technology Workshop 1995, Barton Creeks*. Verfügbar unter: [http://www.ri.cmu.edu/pub\\_files/pub1/vo\\_m\\_t\\_1995\\_1/vo\\_m\\_t\\_1995\\_1.pdf](http://www.ri.cmu.edu/pub_files/pub1/vo_m_t_1995_1/vo_m_t_1995_1.pdf) (Zugriff am 28.11.2001, 19.20 Uhr).
- Wachsmuth, I. (2000). Kommunikative Rhythmen in Gestik und Sprache. *Kognitionswissenschaft*, 8, 151-159.
- Wachsmuth, I., Voss, I., Sowa, T., Latoschik, M.E., Kopp, S., Jung, B (2001). Multimodale Interaktion in der Virtuellen Realität. In H. Oberquelle, R. Oppermann, J. Krause (Hrsg.), *Mensch & Computer 2001*. Stuttgart: Teubner.
- Wahlster, W. (2000). *Verbmobil: Foundations of Speech-to-Speech Translation*. Berlin, Heidelberg, New York: Springer.
- Waibel, A., Suhm, B., Vo, M.T. and Yang, J. (1997). Multimodal Interfaces for Multimedia Information Agents. *Proceedings of the International Conference on Acoustics, Speech and Signal Processing*, 167-170.
- Waibel, A., Minh Tue Vo, Duchnowski, P. & Manke, S. (1995). Multimodal Interfaces. *Artificial Intelligence Review, Special Volume on Integration of Natural Language and Vision Processing*, 10 (3-4), 299-319.
- Wandmacher, J. (1993). *Software-Ergonomie*. Berlin, New York: Walter de Gruyter.
- Warren, W. H. (1984). Perceiving Affordances: Visual Guidance of Stair Climbing. *Journal of Experimental Psychology: Human Perception and Performance*, 10 (5), 683-703.
- Weber, W.G., Oesterreich, R., Zölch, M. & Leder, L. (1994). *Arbeit an CNC-Werkzeugmaschinen – ein arbeitswissenschaftlicher Leitfadens*. Stuttgart: Teubner.
- Wechsler, D. (1991). *Hamburg-Wechsler-Intelligenztest für Erwachsene - (HAWIE-R)*. Bern, Stuttgart, Toronto: Huber.
- Wieland-Eckelmann, R. (1992). *Kognition, Emotion und psychische Beanspruchung*. Göttingen: Hogrefe.
- Winograd, T.A. (1983). *Language as a cognitive process*. Reading, Mass.: Addison-Wesley
- Wirth, W., Plasa, K. & Schubert, M. (1999). Evaluation der Methode des Lauten Denkens. *1.Tagung der DGPK-Fachgruppe Methoden der empirischen Kommunikationsforschung zum Thema „Kritische Bestandsaufnahme und Evaluation kommunikationswissenschaftlicher Methoden“ am 19. bis 20. Februar 1999 in Leipzig*. Verfügbar unter: [http://www.dgpuk.de/fg\\_meth/](http://www.dgpuk.de/fg_meth/) (Zugriff am 4.9.2001, 13:00 Uhr).
- Wixon, D. & Wilson, C. (1997). The Usability Engineering Framework for Product Design and Evaluation. In M.G. Helander, T.K. Landauer & P.V. Prabhu (Eds.), *Handbook of Human-Computer Interaction*. Amsterdam u.a.: Elsevier.
- Wottawa, H. & Thierau, H. (1990). *Evaluation*. Bern: Huber.
- Yankelovich, N., Levow, G.-A. & Marx, M. (1995). Designing SpeechActs: Issues in Speech User Interfaces. *Proceedings on Human Factors in Computing Systems*, 369-376.
- Yoshida, A., Rolland, J. & Reif, J. (1995). Design and applications of a highresolution insert head-mounted-display. *Proceedings of the Virtual Reality Annual International Symposium '95*, 84-93.

- Zhai, S. (1998). User Performance in Relation to 3D Input Device Design. *Computer Graphics*, 32 (4), 50-54.
- Zhai, S., Morimoto, C. & Ihde, S. (1999). Manual And Gaze Input Cascaded (MAGIC) Pointing. *Proc. CHI '99: ACM Conference on Human Factors in Computing Systems*, 246-253.
- ZMMS Eyes Tea (2002). *Geräte*. Verfügbar unter: <http://www.zmms.tu-berlin.de/Eyes-Tea/geraete.phtml> (Zugriff am 5.1.2002, 18:22 Uhr).
- Zue, V. (1997). Conversational Interfaces: Advances And Challenges. In Kokkinakis, G., Fakotakis, N., Dermatas, E. (eds.), *Eurospeech'97. 5<sup>th</sup> European Conference on Speech Communication and Technology. Rhodes, Greece, 22-25 September 1997*, KN-9 - KN 18.
- Zukerman, I. & Litman, D. (2001). Natural Language Processing and User Modeling: Synergies and Limitations. *User Modeling and User-Adapted Interaction*, 11, 129-158.
- Zwicky, F. (1966). *Entdecken, Erfinden, Forschen im Morphologischen Weltbild*. München: Droemer-Knaur.

## 9 Verzeichnisse

### 9.1 Abbildungsverzeichnis

Abb. 1:	Modell der multimodalen Mensch-Computer-Interaktion .....	5
Abb. 2:	Ziele der Entwicklung multimodaler Mensch-Computer-Interaktion .....	7
Abb. 3:	Blockdiagramm der Sprachsynthese .....	11
Abb. 4:	Allgemeines psychologisches Kommunikationsmodell (nach Hermann 1995) .....	13
Abb. 5:	Blickrichtungsbestimmung aus dem Verhältnis von Pupillenmittelpunkt und cornealem Reflex (links: Fadenkreuze durch Pupillenschwerpunkt und Corneareflex) bei einem kontaktfreien Messsystem (rechts: iview- System der Firma SMI) .....	17
Abb. 6:	Blickrichtungsabhängiges Objekt-Rendering zur Optimierung von Berechnungszeiten für die Computergrafik; links: 358 Polygone gleichmäßig über das gesamte Objekt verteilt, rechts: 358 Polygone detaillierter am Kopf und niedrigauflösend an den Hinterbeinen verteilt (O’Sullivan et al. 2000). .....	20
Abb. 7:	Segmentierte Hand (links) und ihre Handvektoren (rechts) zur Erkennung der Handgesten ©HHI .....	23
Abb. 8:	Das Phantom zur haptischen Interaktion. Der Finger, der die Krafrückmeldung empfindet, steckt in einem beweglichen Fingerhut am Ende des Auslegers, der in den drei Achsen bewegt werden kann (Sens Able Technologies 2001).....	24
Abb. 9:	Gestentaxonomie für die Mensch-Computer-Interaktion (nach Pavlović, Sharma & Huang 1997) .....	25
Abb. 10:	Integration der Eingangssignale verschiedener Modalitäten mit unterschiedlichen Signalcharakteristiken .....	32
Abb. 11:	Das V-Modell (links) und das Spiralmodell (rechts) für die in Phasen gegliederte Systementwicklung (Blanchard & Fabrycky 1998).....	35
Abb. 12:	Realisierungsebenen von vertikalen gegenüber horizontalen Prototypen (Balzert 1998) .....	37
Abb. 13:	Field Medic Information System von NCR Corporation (aus Oviatt et al. 2000): die Field Medic Associate Hardware mit Spracherkennung ist in die Weste des Notarztes integriert (links), das Interface des Field Medic Coordinator mit alternativer Sprach- oder Stifteingabe (rechts). .....	44
Abb. 14:	Multimodale Interaktionskomponenten des mUltimo3D-Systems ©HHI .....	62
Abb. 15:	Formative und summative Evaluation von mUltimo3D.....	64
Abb. 16:	Aufbau des Testsystems für die Evaluation. Graue Kästchen sym- bolisieren Hardware-Komponenten, weiße Kästchen symbolisieren Software-Komponenten. ....	65

Abb. 17:	Screenshot des multimodalen 3D-Informationsbrowsers (links) und des multimodalen 3D-Konstruktionsraumes (rechts) .....	69
Abb. 18:	Screenshots der Untersuchungsbedingungen .....	78
Abb. 19:	Versuchsaufbau für die Untersuchung der Blickinteraktion .....	81
Abb. 20:	Bereiche der höchsten Blickintensität über Buchstabenpositionen (Kästchen) und Schwerpunkte über Blickinteraktions- und Kontrollbedingung. Die gestrichelte Linie markiert die Bedingung ohne Blickinteraktion (Kontrollbedingung), bei der die linke und rechte Targetposition weniger korrekt fokussiert wird als über alle Blickinteraktionsbedingungen (durchgehende Linien). .....	84
Abb. 21:	Interface der fiktiven Filmdatenbank für die Sprachinteraktionsbedingung (die Aufforderung wurde nur gesprochen und nicht als Text angezeigt) (links) und für die Textbedingung bzw. für die kombinierte Bedingung (rechts) .....	88
Abb. 22:	Medianwerte der relativen Häufigkeit Spracheingabe über die drei Interaktionsbedingungen. ....	89
Abb. 23:	Leichte (links), mittlere (Mitte) und schwere Konstruktionsaufgabe (rechts) .....	93
Abb. 24:	Screenshot der Testapplikation „Konstruktion“ .....	95
Abb. 25:	Screenshot der Testapplikation „Informationssuche“ .....	96
Abb. 26:	Screenshot einer Konstruktion mit mittlerer Aufgabenschwierigkeit von Versuchsteilnehmer 10, der eine sehr gute Leistung erbrachte. ....	98
Abb. 27:	Strukturrekonstruktion des Informationsobjektes „mUltimo3D“ von Versuchsteilnehmer 19 .....	99
Abb. 28:	Versuchsablauf bei der Untersuchung prototypischer Aufgaben .....	101
Abb. 29:	Schwierigkeitsranking für die Aufgaben zur Konstruktion .....	102
Abb. 30:	Schwierigkeitsranking für die Aufgaben zur Informationssuche .....	103
Abb. 31:	Mittlere Anstrengung beim Konstruieren und beim Informationssuchen .....	107
Abb. 32:	Relativer Anteil der Interaktionsmodalitäten für die beiden prototypischen Aufgaben Informationssuche und Konstruktion .....	110

## 9.2 Tabellenverzeichnis

Tab. 1:	Berliner Klassifikationsschema .....	49
Tab. 2:	Matrix für die strukturierte Zuordnung von Systemeigenschaften zu Aufgabeneigenschaften .....	52
Tab. 3:	Usability-Heuristiken nach Nielsen (1994) .....	56
Tab. 4:	Ausgewählte Ergebnisse der Experten Evaluation des Informationsbrowsers in den drei Mängelbereichen .....	72
Tab. 5:	Ausgewählte Ergebnisse der Experten Evaluation des Konstruktionsraumes in den drei Mängelbereichen .....	73
Tab. 6:	Versuchsplan für die Untersuchung zur Blickrückmeldung .....	79
Tab. 7:	Ergebnisse der Blickinteraktionsbedingungen und der Kontrollbedingung (Signifikanzniveau: **= $p < 0,01$ ; *= $p < 0,05$ ; n.s.=nicht signifikant) .....	82
Tab. 8:	Versuchsplan für die Untersuchung zur Sprachinteraktion. Mit Si wird die Versuchspersonengruppe bezeichnet, gleiche Indexnummer i kennzeichnet die gleichen Personen. ....	87
Tab. 9:	Versuchsplan für die Untersuchung prototypischer Aufgaben .....	93
Tab. 10:	Schwierigkeitsranking für die Aufgaben zur Konstruktion und Informationssuche .....	103
Tab. 11:	Mediane der Leistung in den Aufgaben zur Konstruktion .....	104
Tab. 12:	Mediane der Leistung in den Aufgaben zur Informationssuche .....	104
Tab. 13:	Mediane der multimodalen Interaktionshäufigkeit beim Konstruieren .....	105
Tab. 14:	Mediane der multimodalen Interaktionshäufigkeit beim Informationssuchen .....	105
Tab. 15:	Mittlere Anstrengung beim Konstruieren .....	106
Tab. 16:	Mittlere Anstrengung beim Informationssuchen .....	106
Tab. 17:	T-Tests zur Beanspruchungshöhe .....	107
Tab. 18:	Leistung und multimodale Interaktionshäufigkeit über die Schwierigkeitsstufen beim Konstruieren .....	108
Tab. 19:	Leistung und multimodale Interaktionshäufigkeit über die Schwierigkeitsstufen beim Informationssuchen .....	109
Tab. 20:	Unterschiedliche Nutzung der Modalitäten beim Konstruieren und beim Informationssuchen .....	110
Tab. 21:	Pragmatische und hedonistische Qualität sowie Attraktivität der Testapplikationen .....	111
Tab. 22:	Gestalterische Empfehlungen für die Blick- und Sprachinteraktion .....	116
Tab. 23:	Integration der Evaluationsergebnisse von mUltimo3D in die System-Aufgaben-Eigenschaftsmatrix (nähere Erläuterungen im Text) .....	120
Tab. A1-1:	Rang-Korrelationen zwischen der SEA-Skala und dem NASA-TLX .....	142

## 10 Anhänge

### Anhang 1: Validierung der SEA-Skala am NASA-TLX

Die Validierung der SEA-Skala am NASA-TLX ist mit einer Korrelation von  $r=.78$  bei  $N=72$  hochsignifikant. Die höchste Korrelation zwischen SEA-Skala und NASA-TLX besteht mit der Subskala „Anstrengung“ ( $r=.81$ ). Die Ergebnisse sind in der Tab. A1-1 nachzulesen:

Tab. A1-1: Rang-Korrelationen zwischen der SEA-Skala und dem NASA-TLX

		tlx_score	TLX_GA	TLX_KA	TLX_ZA	TLX_AA	TLX_A	TLX_F	SEA
tlx_score	Korrelation nach Pearson	1,000	,841	,626	,796	,644	,835	,817	,779
	Signifikanz (2-seitig)	,	,000	,000	,000	,000	,000	,000	,000
	N	114	114	114	114	114	114	114	113
TLX_GA	Korrelation nach Pearson	,841	1,000	,436	,693	,295	,683	,682	,694
	Signifikanz (2-seitig)	,000	,	,000	,000	,001	,000	,000	,000
	N	114	114	114	114	114	114	114	113
TLX_KA	Korrelation nach Pearson	,626	,436	1,000	,509	,311	,538	,470	,580
	Signifikanz (2-seitig)	,000	,000	,	,000	,001	,000	,000	,000
	N	114	114	114	114	114	114	114	113
TLX_ZA	Korrelation nach Pearson	,796	,693	,509	1,000	,454	,637	,668	,678
	Signifikanz (2-seitig)	,000	,000	,000	,	,000	,000	,000	,000
	N	114	114	114	114	114	114	114	113
TLX_AA	Korrelation nach Pearson	,644	,295	,311	,454	1,000	,464	,467	,355
	Signifikanz (2-seitig)	,000	,001	,001	,000	,	,000	,000	,000
	N	114	114	114	114	114	114	114	113
TLX_A	Korrelation nach Pearson	,835	,683	,538	,637	,464	1,000	,712	,811
	Signifikanz (2-seitig)	,000	,000	,000	,000	,000	,	,000	,000
	N	114	114	114	114	114	114	114	113
TLX_F	Korrelation nach Pearson	,817	,682	,470	,668	,467	,712	1,000	,772
	Signifikanz (2-seitig)	,000	,000	,000	,000	,000	,000	,	,000
	N	114	114	114	114	114	114	114	113
SEA	Korrelation nach Pearson	,779	,694	,580	,678	,355	,811	,772	1,000

	Signifikanz (2-seitig)	,000	,000	,000	,000	,000	,000	,000	,
	N	113	113	113	113	113	113	113	113

Die Korrelation sind alle auf dem Niveau von 0,01 (2-seitig) signifikant.

GA – Geistige Anforderungen

KA – Körperliche Anforderungen

ZA – Zeitliche Anforderungen

AA – usführung der Aufgaben

A – Anstrengung

F – Frustration

## Anhang 2: Prioritätsliste für die Gestaltung der Anwendungen

### Allgemeine Gestaltungsanforderungen

**Zuverlässigkeit der Funktionen und Modalitäten** erhöhen für Objektmanipulation und WWW-Browser

**Verzögerungszeiten** minimieren

Alle **Interaktionselemente**, die sichtbar sind, sollten funktionieren

**Schattenwurf** in Objektmanipulation und Browser inkonsistent

#### **Blick:**

Kleine Objekte schwer zu „treffen“ – keine Blickinteraktion oder Objekte (zumindest temporär) in Vordergrund bringen (vergrößern)

Fokus bleibt zu lange auf einem Objekt hängen – Schwellzeiten verändern

Permanente akustische Rückmeldung der Blickauswahl abstellen

#### **Sprache:**

Objekte direkt adressierbar machen

Syntaxfreie Koordinaten- und Metrikeingaben

*shortcuts* per Sprache

Interaktionselemente (Funktionen, Links) sollen mit Funktions- und Metaphernbenennung adressierbar sein

für <Aktivieren> Befehl andere Interaktionsform finden, da zu umständlich

Hilfe per Sprache

Suchfunktionen per Sprache

### Objektmanipulation (CAD)

**Funktionen**, die implementiert werden sollten:

<Hilfefunktion>

<Rückgängig machen von Aktionen>

<Kopieren>

<*drag and drop*>

<Stoppen kontinuierlicher Funktionen per Sprache>

<*handles* an Objekte> zum Anfassen mit der Maus

<Gruppierung von Objekten>

<Fangfunktion>

<Ausrichten am Gitternetz>

<Rotationsachse festlegen>

<dimensionale Skalierung>



<Rückmeldung> bei kritischen Funktionen wie <alles löschen>  
<Rückmeldung von Objekteigenschaften>  
<Rückmeldung> für die Leistung der multimodalen Sensoren (H&E-tracker, Spr.)  
<Lage zweier Objekte zueinander verändern>  
<Bohrung/Fase/Verrundung/Nut anbringen>  
<Stift/Rippe hinzufügen>  
<Bohrung/Stift verschieben>  
<genaue Größen- u. Lagekoordinaten eingeben> von Objekten

### **Interaktionen:**

**Maus:** Auswahl zwischen 3D-Maus und 2D-Maus mit eindeutiger Rückmeldung der möglichen Interaktionsebene (sollte im Bearbeitungsraum markiert werden)

Beim Löschen wird eingeführte Syntax verkehrt - angleichen

Rückmeldung über Systemstatus (Sanduhr o.ä.)

Stoppen von Funktionen transparent machen

### **Design:**

drei Gitternetze für Normansichten besser erkennbar machen

Übereinanderlagerung von Objekten bei Neuauswahl aus dem Regal abstellen (evtl. <default verschieben>)

Vermindern der Animation (ausgewählte Funktion + aktivierte Ebene)

<Skalierwürfel> mit Pfeilen symbolisieren

Funktionen beschriften

Farben weniger gesättigt für große Objekte, stärker gesättigt für kleine Objekte

XYZ-Ebenen-Auswahl direkt am Arbeitsraum symbolisieren

Durchgängige Metapher suchen

### **WWW-Browser**

**Funktionen**, die implementiert werden sollten:

<Blättern/*scrollen* der Seiten>

<Favoriten selbst anlegen/ Seiten speichern>

<Favoriten benennen>

<Zoom für aktuelle Seite>

<Zoom des Navigationsbaumes>

<Hilfefunktion>

<Rückmeldung Ladefortschritt>

## **Interaktionen:**

<Zurück> mit mehreren Schritten ermöglichen

Datenbank erweitern

## **Design:**

Konsistente Metapher finden (Haus oder Raumfahrt)

Zusammenhängende Anordnung der Steuerungsoperationen (Zurück, Home...)

Verringern von Animation und Highlights (Seite im Navigationsbaum + Favorit + aktuelles Dokument + Link in der Linkleiste)

Navigationsbaum abstrakter darstellen, *thumbnail*-Darstellung führt nicht zu guter Lesbarkeit

Funktionsaufteilung zwischen Linkleiste und Favoriten klären

Linkleiste homogener klassifizieren

Linkleiste abhängig vom aktuellen Dokument aufbauen, Metapher angeschraubter Schilder unpassend

Zurückbutton als Pfeil

Favoriten kleiner mit lesbarer, individualisierbarer Bezeichnung

Startseite bekannter als Häuschen symbolisiert, bogenförmige Beschriftung liest sich schwer – geradestellen

Notizblock-Metapher für „Aktuelles Dokument“ suggeriert Beschreibbarkeit,

*Scrollbar* mit Richtungspfeilen symbolisieren (keine Verwechslung mit Schreibstift)

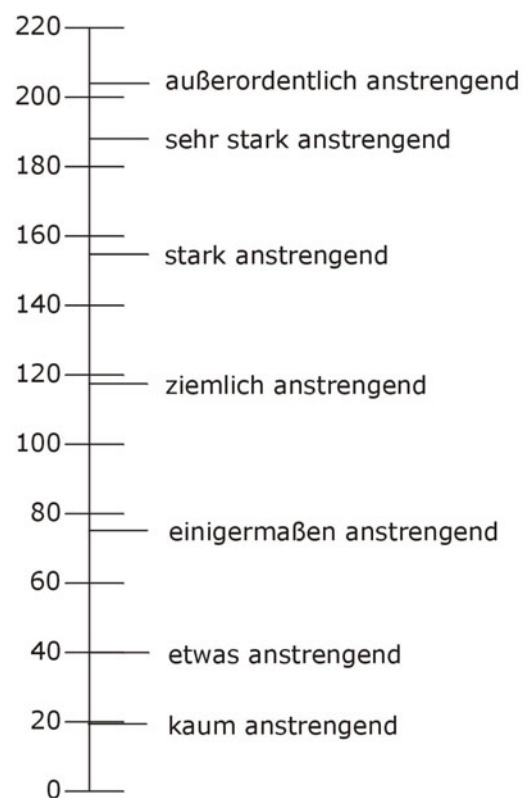
Aktuelles Dokument beim Vergrößern gerade stellen (bessere Lesbarkeit)

Inhalte von Aktuellen Dokumenten strukturieren

### Anhang 3: SEA-Skala

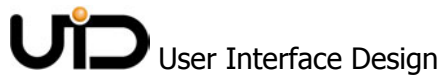
# Beanspruchungshöhe

Bitte kreuzen Sie auf der folgenden Skala Ihre Gesamtbewertung für die gerade absolvierte Versuchsbedingung an.



## Anhang 4: Semantisches Differential

Alle Rechte an diesem Instrument liegen bei der User Interface Design GmbH



Tester	
--------	--

### Ihr Urteil!

Nachfolgend finden Sie Wortpaare, mit deren Hilfe Sie die CAD-Applikation bewerten können. Sie stellen jeweils extreme Gegensätze dar, zwischen denen eine Abstufung möglich ist.

*Ein Beispiel:*

unsympathisch        sympathisch

*Diese Bewertung bedeutet, dass die CAD-Applikation eher sympathisch, aber noch verbesserungsbedürftig ist.*

Denken Sie nicht lange über die Wortpaare nach, sondern geben Sie bitte die Einschätzung ab, die Ihnen spontan in den Sinn kommt.

Vielleicht passen einige Wortpaare nicht so gut auf das Produkt, kreuzen Sie aber trotzdem bitte immer eine Antwort an. Denken Sie daran, dass es keine „richtigen“ oder „falschen“ Antworten gibt - nur Ihre persönliche Meinung zählt!

### Ihr Urteil!

*Bitte geben Sie mit Hilfe der folgenden Wortpaare Ihren Eindruck der CAD-Applikation wieder. Bitte kreuzen Sie nur jeweils ein Kästchen an!*

## **Anhang 5: NASA-TLX deutsche Fassung**

### **Beanspruchungsstruktur**

Geben Sie bitte an, welche relative Bedeutung für die empfundene Gesamtbeanspruchung bei der eben durchgeführten Aufgabe die sechs Beanspruchungsdimensionen

Geistige Anforderung

Körperliche Anforderung

Zeitliche Anforderung

Aufgabenausführung

Anstrengung und

Frustration

für Sie hatten. Lesen Sie dazu bitte zunächst die folgenden Erläuterungen:

#### **Geistige Anforderungen**

Wie viel geistige Anstrengung war bei der Informationsaufnahme und bei der Informationsverarbeitung erforderlich (z.B. Denken, Entscheiden, Rechnen, Erinnern, Hinsehen, Suchen ...)? War die Aufgabe leicht oder anspruchsvoll, einfach oder komplex, erfordert sie hohe Genauigkeit oder ist sie fehlertolerant?

#### **Körperliche Anforderungen**

Wie viel körperliche Aktivität war erforderlich (z.B. ziehen, drücken, drehen, steuern, aktivieren ...)? War die Aufgabe leicht oder schwer, einfach oder anstrengend, erholsam oder mühselig?

#### **Zeitliche Anforderungen**

Wie viel Zeitdruck empfanden Sie hinsichtlich der Häufigkeit oder dem Takt mit dem Aufgaben oder Aufgabenelemente auftraten? War die Abfolge langsam und geruhsam oder schnell und hektisch?

#### **Ausführung der Aufgaben**

Wie erfolgreich haben Sie ihrer Meinung nach die vom Versuchsleiter (oder Ihnen selbst) gesetzten Ziele erreicht? Wie zufrieden waren Sie mit Ihrer Leistung bei der Verfolgung dieser Ziele?

## Anstrengung

Wie hart mussten Sie arbeiten, um Ihren Grad an Aufgabenerfüllung zu erreichen?

## Frustration

Wie unsicher, entmutigt, irritiert, gestresst und verärgert (versus sicher, bestätigt, zufrieden, entspannt und zufrieden mit sich selbst) fühlten Sie sich während der Aufgabe?

Im folgenden werden jeweils zwei der sechs Beanspruchungsdimensionen in verschiedenen Kombinationen gegenübergestellt. Geben Sie jeweils an, welche Beanspruchungsdimension für die Gesamtbeanspruchung, die Sie empfunden haben, bedeutsamer war. Es geht also zunächst nicht darum, wie hoch die Beanspruchung in den einzelnen Dimensionen war, sondern wie wichtig die jeweilige Dimension für das Gesamtempfinden war!

Beispiel: Wenn für Sie die geistigen Anforderungen, die die Aufgabe gestellt hat, bedeutsamer für das Beanspruchungserleben waren, als die Anstrengung, die Sie aufbringen mussten, kreuzen Sie bitte so an:

Anstrengung	<input type="checkbox"/>	<input checked="" type="checkbox"/>	Geistige Anforderungen
Körperliche Anforderungen	<input type="checkbox"/>	<input type="checkbox"/>	Zeitliche Anforderungen
Anstrengung	<input type="checkbox"/>	<input type="checkbox"/>	Geistige Anforderungen
Frustration	<input type="checkbox"/>	<input type="checkbox"/>	Körperliche Anforderungen
Anstrengung	<input type="checkbox"/>	<input type="checkbox"/>	Frustration
Geistige Anforderungen	<input type="checkbox"/>	<input type="checkbox"/>	Zeitliche Anforderungen
Körperliche Anforderungen	<input type="checkbox"/>	<input type="checkbox"/>	Anstrengung
Zeitliche Anforderungen	<input type="checkbox"/>	<input type="checkbox"/>	Ausführung der Aufgaben

Frustration	<input type="checkbox"/>	<input type="checkbox"/>	Geistige Anforderungen
Zeitliche Anforderungen	<input type="checkbox"/>	<input type="checkbox"/>	Frustration
Ausführung der Aufgaben	<input type="checkbox"/>	<input type="checkbox"/>	Anstrengung
Anstrengung	<input type="checkbox"/>	<input type="checkbox"/>	Zeitliche Anforderungen
Frustration	<input type="checkbox"/>	<input type="checkbox"/>	Ausführung der Aufgaben
Ausführung der Aufgaben	<input type="checkbox"/>	<input type="checkbox"/>	Körperliche Anforderungen
Geistige Anforderungen	<input type="checkbox"/>	<input type="checkbox"/>	Ausführung der Aufgaben
Geistige Anforderungen	<input type="checkbox"/>	<input type="checkbox"/>	Körperliche Anforderungen

Kontrollieren Sie bitte, ob Sie auch kein Vergleichspaar vergessen haben!

## Beanspruchungshöhe

Geben Sie jetzt bitte an, wie hoch die Beanspruchung in den einzelnen Dimensionen war. Markieren sie dazu auf den folgenden Skalen bitte, in welchem Maße Sie sich in den sechs genannten Dimensionen von der Aufgabe beansprucht oder gefordert gesehen haben:

Beispiel:



## Geistige Anforderungen

Wie viel geistige Anstrengung war bei der Informationsaufnahme und bei der Informationsverarbeitung erforderlich (z.B. Denken, Entscheiden, Rechnen, Erinnern, Hinsehen, Suchen ...)? War die Aufgabe leicht oder anspruchsvoll, einfach oder komplex, erfordert sie hohe Genauigkeit oder ist sie fehlertolerant?



## Körperliche Anforderungen

Wie viel körperliche Aktivität war erforderlich (z.B. ziehen, drücken, drehen, steuern, aktivieren ...)? War die Aufgabe leicht oder schwer, einfach oder anstrengend, erholsam oder mühselig?



## Zeitliche Anforderungen

Wie viel Zeitdruck empfanden Sie hinsichtlich der Häufigkeit oder dem Takt mit dem Aufgaben oder Aufgabenelemente auftraten? War die Abfolge langsam und geruhsam oder schnell und hektisch?



## Ausführung der Aufgaben

Wie erfolgreich haben Sie ihrer Meinung nach die vom Versuchsleiter (oder Ihnen selbst) gesetzten Ziele erreicht? Wie zufrieden waren Sie mit Ihrer Leistung bei der Verfolgung dieser Ziele?





## Anstrengung

Wie hart mussten Sie arbeiten, um Ihren Grad an Aufgabenerfüllung zu erreichen?



## Frustration

Wie unsicher, entmutigt, irritiert, gestresst und verärgert (versus sicher, bestätigt, zufrieden, entspannt und zufrieden mit sich selbst) fühlten Sie sich während der Aufgabe?



Kontrollieren Sie bitte, ob Sie zu allen Fragen Angaben gemacht haben. Bei Unklarheiten wenden Sie sich bitte an die anwesenden Versuchsleiter.