





Christian Elting

Modellierung, Implementierung und Evaluierung von  
gekoppelten Ausgabegeräten in ubiquitären  
Informationssystemen

Erstgutachterin: Prof. Dr. rer. nat. Elisabeth André  
Zweitgutachter: Prof. Dr. rer. nat. Rainer Malaka  
Tag der mündlichen Prüfung: 21. Januar 2008

# Modellierung, Implementierung und Evaluierung von gekoppelten Ausgabegeräten in ubiquitären Informationssystemen

## **Dissertation**

zur Erlangung des akademischen Grades eines

## **Doktors der Naturwissenschaften**

der Fakultät für Angewandte Informatik

der Universität Augsburg

eingereicht von

Dipl. Inform. Christian Elting



Die Deutsche Bibliothek – CIP-Einheitsaufnahme

Autor: Christian Elting

Modellierung, Implementierung und Evaluierung von gekoppelten Ausgabe-  
geräten in ubiquitären Informationssystemen

Christian Elting – Berlin: uni-edition, 2008

ISBN 978-3-937151-73-1

Informationen über den Verlag und das aktuelle Buchangebot finden Sie im  
Internet unter [www.uni-edition.de](http://www.uni-edition.de)

Gedruckt auf holz- und säurefreiem Papier, 100% chlorfrei gebleicht.

© uni-edition GmbH, Berlin

Zehrendorfer Str. 11, D – 12277 Berlin

Die Verwertung der Texte und Bilder, auch auszugsweise, ist ohne Zustimmung  
des Verlags urheberrechtswidrig und strafbar. Dies gilt auch für Vervielfältigungen,  
Übersetzungen, Mikroverfilmung und für die Verarbeitung mit  
elektronischen Systemen.

Herstellung: Schaltdienst Lange, Berlin

Printed in Germany

ISBN 978-3-937151-73-1

## Danksagung

Diese Arbeit hat von vielerlei Seiten Unterstützung erfahren, ohne die sie nicht in dieser Form hätte realisiert werden können. Ich danke Frau Prof. Dr. Elisabeth André für die Begutachtung der Arbeit, für die vielen interessanten Gespräche, für ihre Unterstützung in fachlichen wie menschlichen Belangen und für die ebenso freundschaftliche wie produktive Zusammenarbeit während meines Aufenthaltes an ihrem Lehrstuhl.

Herrn Prof. Dr. Rainer Malaka danke ich für die Übernahme des Zweitgutachtens. Ich danke ihm als meinem Betreuer vor Ort in Heidelberg für die interessante Themenstellung, seinen Einsatz für diese Arbeit und die Möglichkeit, sechs Jahre unter außergewöhnlich guten Bedingungen forschen zu können. Ich danke ihm für die freundschaftliche und kreative Atmosphäre, die die Arbeit innerhalb der Personal Memory Gruppe ausgezeichnet hat. Ich danke ihm neben seiner fachlichen Unterstützung insbesondere für die Gelegenheit, die Laufstrecken der Stadt Heidelberg gemeinsam kennen zu lernen.

Ich danke meinen Kolleginnen und Kollegen an der European Media Laboratory GmbH, der EML Research gGmbH und der Klaus Tschira Stiftung gGmbH für ihre Freundschaft, ihre Unterstützung und ihre Ermunterung. Insbesondere danke ich Alessandra Moschetti, Annika Scheffler und Jan Zwickel für ihr außergewöhnliches Engagement im Rahmen der Benutzerstudien. Jens Milatz danke ich für seine Hilfe bei der Implementierung von Software für eingebettete Systeme. Diese Arbeit wurde durch die Klaus Tschira Stiftung gGmbH und das Bundesministerium für Bildung und Forschung als Teil der Projekte Embassi und Dynamite (Förderkennzeichen 01 IL 904 D/2 und 01 ISC 27B) gefördert. Ich danke den externen Projektpartnern beider Projekte für die angenehme und erfolgreiche Zusammenarbeit. Der Klaus Tschira Stiftung gGmbH danke ich insbesondere für das Stipendium, das die Fertigstellung dieser Arbeit ermöglicht hat.

Beim Gegenlesen der Arbeit wurde ich immens unterstützt von Inge Elting, Eckart Scheffler, Annika Scheffler, Matthias Merdes, Alessandra Moschetti und Jan Zwickel. Vielen Dank für die zahlreichen sowie detaillierten Kommentare, Korrekturen und Anregungen. Insbesondere möchte ich meiner Familie danken, die mich in allen Belangen unterstützt hat. Besonderer Dank gebührt dabei meiner Partnerin Annika Scheffler, meiner Mutter Inge Elting, meiner Schwester Sabine Elting und meinem Bruder Frank Elting. Ganz herzlich möchte ich mich auch bei Christa Scheffler und Eckart Scheffler bedanken, die mich mit Rat und Tat enorm unterstützt haben.

Gewidmet ist diese Arbeit meiner Partnerin Annika Scheffler, die auch bei hoher See stets mein Ruhepunkt ist und mich auf Kurs hält.



Wir sind Zwerge, die auf den Schultern von Riesen sitzen.  
Wir können weiter sehen als unsere Ahnen,  
und in dem Maß ist unser Wissen größer als das ihrige,  
und doch wären wir nichts,  
würde uns die Summe ihres Wissens nicht den Weg weisen.

Bernhard von Chartres (1080-1167)



## Zusammenfassung

Benutzer sind heutzutage von zahlreichen Ein- und Ausgabegeräten umgeben. Fast jeder Benutzer verfügt über einen PC, einen Fernseher, einen mobilen Kleinst-Computer (PDA) oder eine Digitalkamera. Obwohl viele dieser Geräte bereits über gemeinsame Standards kommunizieren können, fehlen noch Verfahren, um insbesondere die Ausgabe-Software und -Hardware verschiedener Geräte intelligent anzusteuern und effizient zu koppeln.

Es existieren bereits Beispiele, wo eine Kopplung mehrerer Ausgabegeräte erfolgreich verwendet wird. So verwendet z. B. die Berliner U-Bahn zwei gekoppelte Bildschirme zur Darstellung von Nachrichten. Ebenso verwenden viele Bildschirmarbeiter zwei oder mehr gekoppelte Monitore am Arbeitsplatz. Ein weiteres Beispiel ist die Nintendo DS Videospielekonsole, bei der Benutzer mittels zweier gekoppelter Bildschirme interagieren. Trotzdem existieren nur wenige Produkte, die gekoppelte Ausgabegeräte erfolgreich einsetzen, da jede Benutzerschnittstelle aufwendig an die jeweilige Gerätekombination angepasst werden muss. Aus diesem Grund wurde in dieser Arbeit nach Verfahren gesucht, um Ausgabegeräte möglichst einfach anzusteuern und automatisch Inhalt auf sie verteilen zu können.

Zur Umsetzung des Verfahrens wurde zunächst konzeptuell erarbeitet, wie Selbstbeschreibungen von Multimedia-Ausgabekomponenten (z. B. Komponenten zur Sprach- oder Bildausgabe) in Kombination mit einem Planungsansatz dazu verwendet werden können, Präsentationen automatisch auf unterschiedliche Gerätetypen verteilen zu können. Dieser Ansatz wurde in der Implementierung eines Home-Entertainment-Systems, in der ein Fernseher, ein PDA und ein elektronischer Bilderrahmen gekoppelt wurden, validiert.

Weiterhin wurden die Akzeptanz und die kognitiven Effekte von Präsentationen auf gekoppelten Ausgabegeräten untersucht. Das realisierte Home-Entertainment-System wurde zunächst durch eine Fragebogenstudie mit 56 Teilnehmern evaluiert. In der Fragebogenstudie wurde das System als „gut“ bewertet. Als bevorzugte Gerätekombination erwies sich eine Kopplung von PDA und Fernseher. Der bevorzugte Präsentationsmodus war Text-Bild.

Darüber hinaus wurden die vom System generierten Ausgaben in einer Laborstudie mit 36 Teilnehmer evaluiert. In dieser Studie wurden Ausgaben auf einem Gerät mit Ausgaben, die über mehrere Geräte verteilt sind, in Bezug auf ihre Akzeptanz und kognitive Effizienz verglichen. Dabei ergab sich, dass die Hinzunahme des elektronischen Bilderrahmens die Interaktionszeiten signifikant verringert hat, ohne dass die Erinnerungsleistung signifikant abfiel. Die Kopplung von Fernseher, PDA und elektronischem Bilderrahmen wurde gegenüber einem einzelnen Fernseher signifikant bevorzugt.



# Inhaltsverzeichnis

|          |                                                                |           |
|----------|----------------------------------------------------------------|-----------|
| <b>1</b> | <b>Problemstellung und Ansatz</b>                              | <b>27</b> |
| 1.1      | Thema . . . . .                                                | 27        |
| 1.2      | Ziel . . . . .                                                 | 28        |
| 1.3      | Gliederung . . . . .                                           | 30        |
| <b>2</b> | <b>Einleitung</b>                                              | <b>33</b> |
| 2.1      | Ubiquitäre Informationssysteme . . . . .                       | 33        |
| 2.2      | Multi-modale Dialogsysteme . . . . .                           | 36        |
| 2.3      | Gekoppelte und flexible Bildschirme . . . . .                  | 41        |
| 2.4      | Zusammenfassung . . . . .                                      | 43        |
| <b>3</b> | <b>Stand der Wissenschaft</b>                                  | <b>45</b> |
| 3.1      | Terminologie . . . . .                                         | 45        |
| 3.2      | Multi-modale Ausgaben auf einem Gerät . . . . .                | 46        |
| 3.2.1    | Empirische Untersuchungen . . . . .                            | 46        |
| 3.2.2    | Technische Umsetzung . . . . .                                 | 48        |
| 3.3      | Multi-modale Ausgaben in verteilten Geräteumgebungen . . . . . | 51        |
| 3.3.1    | Empirische Untersuchungen . . . . .                            | 52        |
| 3.3.2    | Technische Umsetzung . . . . .                                 | 56        |
| 3.4      | Modellierung multi-modaler Ausgaben . . . . .                  | 60        |
| 3.4.1    | Modalitätentheorien . . . . .                                  | 61        |
| 3.4.2    | Animierte Charaktere . . . . .                                 | 62        |
| 3.5      | Zusammenfassung . . . . .                                      | 64        |
| <b>4</b> | <b>Grundlagen</b>                                              | <b>67</b> |
| 4.1      | Multi-modale Dialogsysteme . . . . .                           | 67        |
| 4.1.1    | Architektur . . . . .                                          | 67        |
| 4.1.2    | Generierung multi-modaler Präsentationen . . . . .             | 69        |
| 4.2      | Planungsansatz . . . . .                                       | 71        |
| 4.3      | Unimodalitätenmodell nach Bernsen . . . . .                    | 72        |
| <b>5</b> | <b>Modellierung multi-modaler Ausgaben</b>                     | <b>75</b> |

|       |                                                                   |     |
|-------|-------------------------------------------------------------------|-----|
| 5.1   | Beispiel . . . . .                                                | 75  |
| 5.2   | Verteilte multi-modale Dialogsysteme . . . . .                    | 78  |
| 5.2.1 | Architektur . . . . .                                             | 79  |
| 5.2.2 | Modelle und Abhängigkeiten der Präsentationsgenerierung . . . . . | 81  |
| 5.2.3 | Formulierung Präsentationsproblem . . . . .                       | 83  |
| 5.3   | Ad-hoc-Interoperabilität von Ausgabegeräten . . . . .             | 86  |
| 5.3.1 | Koordinationsansatz . . . . .                                     | 87  |
| 5.3.2 | Heuristischer Ansatz . . . . .                                    | 89  |
| 5.3.3 | Konfigurierbarer Ansatz . . . . .                                 | 91  |
| 5.3.4 | Zusammenfassung und Ansatz . . . . .                              | 92  |
| 5.4   | Modellierung multi-modaler Ein-/Ausgabekomponenten . . . . .      | 93  |
| 5.4.1 | Beispiele . . . . .                                               | 94  |
| 5.4.2 | Vorteile des Ansatzes . . . . .                                   | 97  |
| 5.4.3 | Domänenunabhängiges Modell . . . . .                              | 100 |
| 5.4.4 | Modellierung einer Home-Entertainment Umgebung . . . . .          | 105 |
| 5.4.5 | Beispiele für Selbstbeschreibungen . . . . .                      | 115 |
| 5.5   | Präsentationsplanung . . . . .                                    | 122 |
| 5.5.1 | Integration Dienst-, Geräte- und Ressourcenmodell . . . . .       | 123 |
| 5.5.2 | Vorteile des Ansatzes . . . . .                                   | 126 |
| 5.6   | Umsetzung des Beispielszenarios . . . . .                         | 128 |
| 5.7   | Zusammenfassung . . . . .                                         | 131 |

## **6 Implementierung 133**

|       |                                                             |     |
|-------|-------------------------------------------------------------|-----|
| 6.1   | Prototypimplementierung . . . . .                           | 133 |
| 6.1.1 | Überblick . . . . .                                         | 133 |
| 6.1.2 | Modelle . . . . .                                           | 135 |
| 6.1.3 | Präsentationsplanung . . . . .                              | 139 |
| 6.2   | Home-Entertainment mit gekoppelten Ausgabegeräten . . . . . | 140 |
| 6.2.1 | Überblick . . . . .                                         | 140 |
| 6.2.2 | Präsentationen . . . . .                                    | 142 |
| 6.2.3 | Architektur . . . . .                                       | 144 |
| 6.2.4 | Modelle . . . . .                                           | 146 |
| 6.2.5 | Präsentationsplanung . . . . .                              | 152 |

|          |                                                                       |            |
|----------|-----------------------------------------------------------------------|------------|
| 6.3      | Vorteile des gewählten Ansatzes . . . . .                             | 156        |
| 6.4      | Zusammenfassung . . . . .                                             | 156        |
| <b>7</b> | <b>Fragebogenstudie</b>                                               | <b>159</b> |
| 7.1      | Problemstellung und Hypothesen . . . . .                              | 159        |
| 7.1.1    | Problemstellung . . . . .                                             | 159        |
| 7.1.2    | Akzeptanz von gekoppelten Ausgabegeräten . . . . .                    | 159        |
| 7.1.3    | Hypothesen . . . . .                                                  | 162        |
| 7.2      | Methode . . . . .                                                     | 163        |
| 7.2.1    | Variablen . . . . .                                                   | 163        |
| 7.2.2    | Teilnehmer . . . . .                                                  | 164        |
| 7.2.3    | Materialien . . . . .                                                 | 164        |
| 7.2.4    | Durchführung . . . . .                                                | 167        |
| 7.2.5    | Statistische Auswertung . . . . .                                     | 167        |
| 7.3      | Ergebnisse . . . . .                                                  | 168        |
| 7.3.1    | Beschreibung der Stichprobe . . . . .                                 | 169        |
| 7.3.2    | Altersgruppen . . . . .                                               | 170        |
| 7.3.3    | Art des Messebesuchs . . . . .                                        | 171        |
| 7.3.4    | Erfahrungen mit Computern, Fernsehern und PDAs . . . . .              | 171        |
| 7.3.5    | Kinofilme pro Monat . . . . .                                         | 172        |
| 7.3.6    | Eingabemodalitäten . . . . .                                          | 172        |
| 7.3.7    | Ausgabemodalitäten . . . . .                                          | 172        |
| 7.3.8    | Gerätekombinationen . . . . .                                         | 176        |
| 7.3.9    | Bewertung des Gesamtsystems . . . . .                                 | 177        |
| 7.3.10   | Kommentare . . . . .                                                  | 181        |
| 7.4      | Diskussion . . . . .                                                  | 182        |
| 7.5      | Zusammenfassung . . . . .                                             | 184        |
| <b>8</b> | <b>Laborstudie</b>                                                    | <b>187</b> |
| 8.1      | Problemstellung und Hypothesen . . . . .                              | 187        |
| 8.1.1    | Problemstellung . . . . .                                             | 187        |
| 8.1.2    | Kognitive Last und Effizienz von gekoppelten Ausgabegeräten . . . . . | 187        |
| 8.1.3    | Hypothesen . . . . .                                                  | 191        |
| 8.2      | Methode . . . . .                                                     | 192        |

|          |                                                           |            |
|----------|-----------------------------------------------------------|------------|
| 8.2.1    | Variablen . . . . .                                       | 192        |
| 8.2.2    | Teilnehmer . . . . .                                      | 195        |
| 8.2.3    | Materialien . . . . .                                     | 195        |
| 8.2.4    | Durchführung . . . . .                                    | 205        |
| 8.2.5    | Bewertung der erinnerten Filminhalte . . . . .            | 208        |
| 8.2.6    | Statistische Auswertung . . . . .                         | 209        |
| 8.3      | Ergebnisse . . . . .                                      | 210        |
| 8.3.1    | Beschreibung der Stichprobe . . . . .                     | 211        |
| 8.3.2    | Eingangsfragebogen . . . . .                              | 211        |
| 8.3.3    | Erinnerung von Filminhalten . . . . .                     | 212        |
| 8.3.4    | Betrachtungszeit . . . . .                                | 221        |
| 8.3.5    | Erinnerungseffizienz . . . . .                            | 227        |
| 8.3.6    | Attraktivität . . . . .                                   | 230        |
| 8.3.7    | Benötigte Konzentration . . . . .                         | 233        |
| 8.3.8    | Häufigkeit der Betrachtung . . . . .                      | 235        |
| 8.3.9    | Interesse . . . . .                                       | 236        |
| 8.3.10   | Beurteilung des Gesamtsystems . . . . .                   | 238        |
| 8.3.11   | Störfaktoren . . . . .                                    | 240        |
| 8.3.12   | Kommentare . . . . .                                      | 244        |
| 8.4      | Diskussion . . . . .                                      | 245        |
| 8.5      | Zusammenfassung . . . . .                                 | 253        |
| <b>9</b> | <b>Wissenschaftlicher Beitrag und Ausblick</b>            | <b>255</b> |
| 9.1      | Zusammenfassung der Ergebnisse . . . . .                  | 255        |
| 9.2      | Bearbeitung der Forschungsfragen . . . . .                | 258        |
| 9.3      | Ausblick . . . . .                                        | 261        |
| <b>A</b> | <b>Implementierung</b>                                    | <b>265</b> |
| A.1      | Präsentierte Inhalte . . . . .                            | 265        |
| A.2      | Dienst-, Geräte- und Ressourcenmodell . . . . .           | 267        |
| A.3      | Startzustand und Operatoren des Ckuckuk-Planers . . . . . | 268        |
| A.4      | Beispielpräsentationen . . . . .                          | 277        |
| <b>B</b> | <b>Materialien Fragebogenstudie</b>                       | <b>283</b> |

|                                                   |                |
|---------------------------------------------------|----------------|
| <b>C Materialien Laborstudie</b>                  | <b>289</b>     |
| C.1 Eingangsfragebogen . . . . .                  | 289            |
| C.2 Einweisung in die Gerätekombination . . . . . | 290            |
| C.3 Zwischenfragebogen . . . . .                  | 291            |
| C.4 Abschlussfragebogen . . . . .                 | 294            |
| <br><b>Literaturverzeichnis</b>                   | <br><b>295</b> |



## Tabellenverzeichnis

|      |                                                                                        |     |
|------|----------------------------------------------------------------------------------------|-----|
| 3.1  | Übersicht über Systeme mit gekoppelten Ausgabegeräten. . . .                           | 65  |
| 4.1  | Generische Ausgabeunimodalitätenklassen. . . . .                                       | 73  |
| 5.1  | Domänenunabhängiges Modell für Parameter und Status von<br>Ausgabekomponenten. . . . . | 101 |
| 5.2  | Domänenunabhängiges Modell für Parameter und Status von<br>Eingabekomponenten. . . . . | 104 |
| 5.3  | Multi-modale Parameter und Status für <i>agent</i> und <i>player</i> . . .             | 105 |
| 5.4  | Verhaltensweisen von animierten Charakteren vom Typ <i>agent</i> . . .                 | 108 |
| 5.5  | Animationen für Komponenten vom Typ <i>agent</i> . . . . .                             | 109 |
| 5.6  | Ausgabemedium für Komponenten vom Typ <i>agent</i> . . . . .                           | 110 |
| 5.7  | Atomare Ausgabeunimodalitäten: Bernsen-Kategorien. . . . .                             | 111 |
| 5.8  | Atomare Ausgabeunimodalitäten: Parameter, Status und Geräte. . .                       | 112 |
| 5.9  | Atomare Eingabeunimodalitäten: Bernsen-Kategorien. . . . .                             | 114 |
| 5.10 | Atomare Eingabeunimodalitäten: Parameter, Status und Geräte. . .                       | 114 |
| 5.11 | Selbstbeschreibung eines Microsoft Agent-Charakters. . . . .                           | 116 |
| 5.12 | Selbstbeschreibung des TU Darmstadt Charakters. . . . .                                | 117 |
| 5.13 | Selbstbeschreibung des DFKI-Charakters Herr Kaufmann. . . . .                          | 118 |
| 5.14 | Liste der Ausgabemedien des DFKI-Charakters Herr Kaufmann. . . .                       | 119 |
| 5.15 | Selbstbeschreibung einer Sprachausgabekomponente. . . . .                              | 119 |
| 5.16 | Liste der Ausgabemedien einer Sprachausgabekomponente. . . . .                         | 120 |
| 5.17 | Selbstbeschreibung einer Komponente zur SMIL-Wiedergabe. . . . .                       | 120 |
| 5.18 | Selbstbeschreibung einer GUI-Eingabekomponente. . . . .                                | 121 |
| 5.19 | Selbstbeschreibung einer Spracherkennungskomponente. . . . .                           | 122 |
| 5.20 | Ausgabekomponenten im Beispielszenario. . . . .                                        | 127 |
| 6.1  | Ausgabeziele und Verhaltensweisen des animierten Charakters. . . .                     | 138 |
| 7.1  | Abhängige Variablen der Fragebogenstudie. . . . .                                      | 163 |
| 8.1  | Abhängige Variablen der Laborstudie. . . . .                                           | 193 |

|     |                                                                                                                          |     |
|-----|--------------------------------------------------------------------------------------------------------------------------|-----|
| 8.2 | Nachrichtenbeitrag getrennt nach Segmenten (Quelle: Tageschau.de). . . . .                                               | 196 |
| 8.3 | Anzahl gleichzeitig präsentierte Segmente getrennt nach Gruppe und Bedingung. . . . .                                    | 201 |
| 8.4 | Benötigte Anzahl von Ausgabeschritten zur Präsentation eines Beitrages getrennt nach Gruppe und Bedingung. . . . .       | 201 |
| 8.5 | Reihenfolge der Präsentationsbedingungen (TB, BS, BSC) und der Folgen von Nachrichtenbeiträgen (1, 2, 3) im Versuch. . . | 206 |
| 8.6 | Maximal mögliche Bewertung der Erinnerungen pro Präsentationsbedingung. . . . .                                          | 208 |

## Abbildungsverzeichnis

|      |                                                                                                                              |    |
|------|------------------------------------------------------------------------------------------------------------------------------|----|
| 2.1  | Anzahl von Transistoren in einer integrierten Schaltung eines Intel Prozessors (Copyright © 2005 Intel Corporation). . . . . | 34 |
| 2.2  | Digitaler Bilderrahmen Philips 9FF2CWO/00 (Foto: Philips). . . . .                                                           | 35 |
| 2.3  | Philips HomeLab in Eindhoven (Foto: Philips). . . . .                                                                        | 35 |
| 2.4  | Prinzip von multi-modalen Dialogsystemen (Vorlage: Stefan Kopp). . . . .                                                     | 36 |
| 2.5  | Put-that-there System (Foto: [Bolt, 1980], Copyright © 1980 ACM Inc., genehmigter Nachdruck). . . . .                        | 38 |
| 2.6  | Ubiquitäres Informationssystem: Embassi Home-Entertainment-Demonstrator (Foto: European Media Laboratory). . . . .           | 39 |
| 2.7  | Ubiquitäres Informationssystem: Intelligenter Einkaufsführer (Foto: [Wahlster & Wasinger, 2006]). . . . .                    | 40 |
| 2.8  | Berliner Fenster: Wettervorhersage auf zwei gekoppelten Monitoren. . . . .                                                   | 41 |
| 2.9  | Nintendo DS Videospielekonsole (Foto: Wikipedia). . . . .                                                                    | 42 |
| 2.10 | Rollbarer Bildschirm der Firma Polymer Vision (Foto: Polymer Vision). . . . .                                                | 43 |
| 4.1  | Beispiel multi-modales Dialogsystem (Vorlage: [Elting et al., 2003]). . . . .                                                | 68 |
| 4.2  | Präsentationsoperator. . . . .                                                                                               | 72 |
| 5.1  | Präsentationen auf den Gerätekombinationen Fernseher-Bilderrahmen (linke Seite) und PDA (rechte Seite). . . . .              | 76 |
| 5.2  | Präsentationen auf den Gerätekombinationen PDA-Fernseher (linke Seite) und Mobiltelefon (rechte Seite). . . . .              | 78 |
| 5.3  | Beispiel multi-modales Dialogsystem in verteilter Geräteumgebung. . . . .                                                    | 80 |
| 5.4  | Abhängigkeiten von multi-modalen Präsentationen in verteilten Geräteumgebungen. . . . .                                      | 81 |
| 5.5  | Datenfluss in einem verteilten multi-modalen Dialogsystem. . . . .                                                           | 84 |
| 5.6  | Präsentationsproblem in verteilten multi-modalen Dialogsystemen. . . . .                                                     | 85 |

|      |                                                                                                                                  |     |
|------|----------------------------------------------------------------------------------------------------------------------------------|-----|
| 5.7  | Statisches Modell für Ausgabedienste. . . . .                                                                                    | 87  |
| 5.8  | Dynamisches Modell für Ausgabedienste. . . . .                                                                                   | 89  |
| 5.9  | Konfigurierbares Modell für Ausgabedienste. . . . .                                                                              | 91  |
| 5.10 | Statisches Modell für Ausgabedienste mit Trennung nach domänenunabhängigen Eigenschaften. . . . .                                | 93  |
| 5.11 | Modellierung eines animierten Charakters (Bild Charakter: TU Darmstadt). . . . .                                                 | 95  |
| 5.12 | Drei Ansätze zur Modellierung der Ausgabedienste einer Rendering-Komponente $R_i$ . . . . .                                      | 98  |
| 5.13 | Microsoft Agent-Charakter Merlin (Nachdruck der Screen Shots mit freundlicher Erlaubnis der Microsoft Corporation). . . . .      | 115 |
| 5.14 | Animierter Charakter der TU Darmstadt. . . . .                                                                                   | 116 |
| 5.15 | DFKI-Charakter Herr Kaufmann. . . . .                                                                                            | 118 |
| 5.16 | GUI-Eingabekomponenten aus Dynamite ([Elting & Hellenschmidt, 2004], links) und Embassi ([Herfet et al., 2001], rechts). . . . . | 121 |
| 5.17 | Präsentationsplanung in verteilten multi-modalen Dialogsystemen. . . . .                                                         | 123 |
| 5.18 | Ablauf der Präsentationsplanung. . . . .                                                                                         | 125 |
| 6.1  | Mobiles GUI in Embassi (Foto: European Media Laboratory). . . . .                                                                | 134 |
| 6.2  | GUI und animierter Charakter in Embassi (Foto: European Media Laboratory). . . . .                                               | 135 |
| 6.3  | Selbstbeschreibung der Charakterkomponente. . . . .                                                                              | 136 |
| 6.4  | Selbstbeschreibung der Programmauswahl-GUI auf dem Fernseher. . . . .                                                            | 137 |
| 6.5  | Selbstbeschreibung der Sprachausgabekomponente. . . . .                                                                          | 137 |
| 6.6  | Ausgabeinhalt mit Dialogakten. . . . .                                                                                           | 138 |
| 6.7  | GUI-Eingabe und animierter Charakter auf PDA (Fotos: European Media Laboratory). . . . .                                         | 140 |
| 6.8  | Elektronischer Bilderrahmen und Fernseher mit animiertem Charakter. . . . .                                                      | 141 |
| 6.9  | PDA und Fernseher mit animiertem Charakter. . . . .                                                                              | 142 |
| 6.10 | Bilderrahmen, PDA und Fernseher im Text-Layout. . . . .                                                                          | 143 |
| 6.11 | Architektur des Home-Entertainment-Systems. . . . .                                                                              | 144 |
| 6.12 | Selbstbeschreibung der Komponente zur SMIL-Wiedergabe auf dem PDA. . . . .                                                       | 147 |

|      |                                                                                                         |     |
|------|---------------------------------------------------------------------------------------------------------|-----|
| 6.13 | Selbstbeschreibung der Wiedergabekomponente des elektronischen Bilderrahmens. . . . .                   | 148 |
| 6.14 | Selbstbeschreibung der Sprachausgabekomponente. . . . .                                                 | 149 |
| 6.15 | Selbstbeschreibung der Charakterkomponente. . . . .                                                     | 150 |
| 6.16 | Selbstbeschreibung der Komponente zur Spracheingabe. . . . .                                            | 151 |
| 6.17 | Ausgabehalt und -ziel. . . . .                                                                          | 152 |
| 6.18 | Ausgabedienste der Sprachausgabe im Startzustand $s_0$ . . . . .                                        | 153 |
| 6.19 | Planungsbaum für Präsentation mit PDA, Fernseher und Bilderrahmen. . . . .                              | 154 |
|      |                                                                                                         |     |
| 7.1  | Vp während der Demonstration am Dynamite-Stand (Bild: European Media Laboratory). . . . .               | 165 |
| 7.2  | Histogramm für Altersintervalle von fünf Jahren. . . . .                                                | 169 |
| 7.3  | Bewertungen Eingabemodalitäten für Gruppen jünger/älter 20. . . . .                                     | 173 |
| 7.4  | Bewertungen Ausgabemodalitäten. . . . .                                                                 | 173 |
| 7.5  | Bewertungen Ausgabemodalitäten für Gruppen jünger/älter als 20 Jahre. . . . .                           | 174 |
| 7.6  | Bewertungen Ausgabemodalitäten getrennt nach Art des Mesesebesuchs. . . . .                             | 175 |
| 7.7  | Bewertungen Ausgabemodalitäten getrennt nach PC-Nutzung. . . . .                                        | 176 |
| 7.8  | Bewertungen Gerätekombinationen. . . . .                                                                | 177 |
| 7.9  | Bewertungen Gerätekombinationen für Gruppen bis zu drei bzw. mehr als drei Kinofilme pro Monat. . . . . | 178 |
| 7.10 | Histogramm Bewertung der Gesamtdemonstration. . . . .                                                   | 179 |
| 7.11 | Histogramm Nützlichkeit der Kombination mehrerer Geräte. . . . .                                        | 180 |
| 7.12 | Histogramm Nützlichkeit der Wahl zwischen Text-/Charakter-Layout. . . . .                               | 181 |
| 7.13 | Ausgabebereitschaft für System in Euro. . . . .                                                         | 182 |
|      |                                                                                                         |     |
| 8.1  | Geräte in der Tastatur-TV-Gruppe. . . . .                                                               | 197 |
| 8.2  | Geräte in der Tastatur-TV-Bilderrahmen-Gruppe. . . . .                                                  | 198 |
| 8.3  | Geräte in der PDA-TV-Gruppe. . . . .                                                                    | 198 |
| 8.4  | Geräte in der PDA-TV-Bilderrahmen-Gruppe. . . . .                                                       | 199 |
| 8.5  | TB-Bedingung in der Tastatur-TV-Gruppe. . . . .                                                         | 200 |
| 8.6  | TB-Bedingung in der Tastatur-TV-Bilderrahmen-Gruppe. . . . .                                            | 200 |
| 8.7  | BS-Bedingung in der PDA-TV-Bilderrahmen-Gruppe. . . . .                                                 | 202 |

|      |                                                                                                                    |     |
|------|--------------------------------------------------------------------------------------------------------------------|-----|
| 8.8  | TB-Bedingung in der PDA-TV-Bilderrahmen-Gruppe. . . . .                                                            | 203 |
| 8.9  | BS-Bedingung in der Tastatur-TV-Gruppe. . . . .                                                                    | 203 |
| 8.10 | Histogramm für Altersintervalle von zehn Jahren. . . . .                                                           | 211 |
| 8.11 | Erfahrungen mit PDAs, Fernsehern, Computern und elektro-<br>nischen Bilderrah- men. . . . .                        | 212 |
| 8.12 | Gesamtbewertung erinnerter Beiträge getrennt nach Geschlecht<br>und Gerätekombination. . . . .                     | 213 |
| 8.13 | Gesamtbewertung erinnerter Beiträge getrennt nach Zeitpunk-<br>ten der Präsentation. . . . .                       | 214 |
| 8.14 | Gesamtbewertung erinnerter Beiträge getrennt nach Geschlecht<br>und Präsentationsbedingungen. . . . .              | 215 |
| 8.15 | Gesamtbewertung erinnerter Beiträge getrennt nach Alter und<br>Präsentationsbedingungen. . . . .                   | 216 |
| 8.16 | Bildbewertung erinnerter Beiträge getrennt nach Gerätekom-<br>binationen. . . . .                                  | 217 |
| 8.17 | Bildbewertung erinnerter Beiträge getrennt nach Präsentati-<br>onsbedingungen. . . . .                             | 218 |
| 8.18 | Lebhaftigkeit der Erinnerungen an Beiträge getrennt nach Zeit-<br>punkten der Präsentation. . . . .                | 219 |
| 8.19 | Lebhaftigkeit der Erinnerungen an Beiträge getrennt nach Zeit-<br>punkten der Präsentation eines Beitrags. . . . . | 220 |
| 8.20 | Zeiten pro Nachrichtenbeitrag mit Extremwerten. . . . .                                                            | 222 |
| 8.21 | Zeiten pro Nachrichtenbeitrag getrennt nach Gerätekombina-<br>tionen. . . . .                                      | 223 |
| 8.22 | Zeiten pro Nachrichtenbeitrag getrennt nach Zeitpunkten der<br>Präsentation. . . . .                               | 224 |
| 8.23 | Zeiten pro Nachrichtenbeitrag getrennt nach Präsentationsbe-<br>dingungen. . . . .                                 | 225 |
| 8.24 | Zeiten pro Nachrichtenbeitrag getrennt nach Gerätekombina-<br>tionen und Präsentationsbedingungen. . . . .         | 226 |
| 8.25 | Erinnerungseffizienz getrennt nach Zeitpunkten der Präsentati-<br>on. . . . .                                      | 228 |
| 8.26 | Erinnerungseffizienz getrennt nach Präsentationsbedingungen.                                                       | 229 |
| 8.27 | Erinnerungseffizienz getrennt nach Präsentationsbedingungen<br>und Geschlecht. . . . .                             | 230 |
| 8.28 | Attraktivität der Ausgabemodalitäten in der BSC-Bedingung. .                                                       | 231 |

|      |                                                                                                     |     |
|------|-----------------------------------------------------------------------------------------------------|-----|
| 8.29 | Attraktivität der Darstellung getrennt nach Präsentationsbedingungen. . . . .                       | 232 |
| 8.30 | Benötigte Konzentration getrennt nach Präsentationsbedingungen. . . . .                             | 233 |
| 8.31 | Benötigte Konzentration getrennt nach Zeitpunkten der Präsentation. . . . .                         | 234 |
| 8.32 | Häufigkeit der Betrachtung des Fernseherers getrennt nach Gerätekombinationen. . . . .              | 235 |
| 8.33 | Interesse am Beitrag getrennt nach Zeitpunkten der Präsentation.                                    | 237 |
| 8.34 | Interesse am Beitrag getrennt nach Nachrichtenfolgen. . . . .                                       | 238 |
| 8.35 | Attraktivität der Gerätekombinationen. . . . .                                                      | 239 |
| 8.36 | Histogramm bevorzugte Präsentationsmodi. . . . .                                                    | 240 |
| 8.37 | Systemfehler pro Präsentationsbedingung getrennt nach Gerätekombinationen. . . . .                  | 241 |
| 8.38 | Systemfehler pro Präsentationsbedingung getrennt nach Präsentationsbedingungen. . . . .             | 242 |
|      |                                                                                                     |     |
| A.1  | UML-Diagramm Implementierung Ausgabekomponentenmodell. . . . .                                      | 269 |
| A.2  | UML-Diagramm Implementierung Eingabekomponentenmodell. . . . .                                      | 270 |
| A.3  | Beispiel für Startzustand $s_o$ des Ckuckuk-Planers. . . . .                                        | 271 |
| A.4  | Operator <code>BuildPresentation</code> des Ckuckuk-Planers. . . . .                                | 272 |
| A.5  | Operator <code>BuildSmilPresentationSingleImage</code> des Ckuckuk-Planers. . . . .                 | 273 |
| A.6  | Operator <code>PresentStaticImage-640-480</code> des Ckuckuk-Planers. . . . .                       | 273 |
| A.7  | Operator <code>BuildSmilPresentation</code> des Ckuckuk-Planers.                                    | 274 |
| A.8  | Operator <code>PresentStaticImageWithSpeechAndCharacter</code> des Ckuckuk-Planers. . . . .         | 275 |
| A.9  | Operator <code>PresentStaticImageWithSpeechAndCharacter-240-290</code> des Ckuckuk-Planers. . . . . | 276 |
| A.10 | Operator <code>BuildDigiframePresentationSingleImage</code> des Ckuckuk-Planers. . . . .            | 277 |
| A.11 | Präsentation auf dem PDA mit Charakter (Bild: European Media Laboratory). . . . .                   | 277 |

|                                                                                            |     |
|--------------------------------------------------------------------------------------------|-----|
| A.12 Präsentation auf dem PDA mit visuellem Text. . . . .                                  | 278 |
| A.13 Präsentation auf dem Fernseher mit Charakter. . . . .                                 | 278 |
| A.14 Präsentation auf dem Fernseher mit visuellem Text. . . . .                            | 279 |
| A.15 Präsentation auf PDA und Fernseher mit Charakter. . . . .                             | 279 |
| A.16 Präsentation auf PDA und Fernseher mit visuellem Text. . . .                          | 280 |
| A.17 Präsentation auf Fernseher und Bilderrahmen mit Charakter. .                          | 280 |
| A.18 Präsentation auf Fernseher und Bilderrahmen mit visuellem<br>Text. . . . .            | 281 |
| A.19 Präsentation auf PDA, Fernseher und Bilderrahmen mit Cha-<br>rakter. . . . .          | 281 |
| A.20 Präsentation auf PDA, Fernseher und Bilderrahmen mit visu-<br>ellem Text. . . . .     | 282 |
|                                                                                            |     |
| C.1 Eingangsfragebogen, Seite 1: Einleitung und Fragen zu de-<br>skriptiven Daten. . . . . | 289 |
| C.2 Eingangsfragebogen, Seite 2: Instruktionen. . . . .                                    | 290 |
| C.3 Einleitung auf Fernseher, Seite 1. . . . .                                             | 290 |
| C.4 Einleitung auf Fernseher, Seite 2. . . . .                                             | 291 |
| C.5 Einleitung auf Fernseher, Seite 3. . . . .                                             | 291 |
| C.6 Zwischenfragebogen, Teil A: Fragen zur Präsentationsbedin-<br>gung (hier BSC). . . . . | 292 |
| C.7 Zwischenfragebogen, Teil B: Abfrage der erinnerten Nach-<br>richtentitel. . . . .      | 292 |
| C.8 Zwischenfragebogen, Teil C: Abfrage der erinnerten Nach-<br>richteninhalte. . . . .    | 293 |
| C.9 Zwischenfragebogen, Teil D: Abfrage der Gerätenutzung. . . .                           | 293 |
| C.10 Abschlussfragebogen in der PDA-TV-Bilderrahmen-Gruppe. .                              | 294 |

# 1 Problemstellung und Ansatz

## 1.1 Thema

Das Eingangszitat von Bernhard von Chartres beschreibt die Situation von Menschen in Zeiten der individuellen, gesellschaftlichen oder politischen Weiterentwicklung. Von Chartres Gleichnis trifft ebenso auf die heutige Informationsgesellschaft zu, die sich in einem permanenten Wandel befindet und gleichsam von Chartres Zwergen auf eine Vielzahl von informationstechnologischen Meilensteinen zurückblickt. Diese Entwicklung erfolgte in sehr kurzer Zeit. Von dem ersten elektronischen Digitalrechner (1942, [Mollenhoff, 1988]) bis zum Versenden der ersten Email (1971, [Coffman & Odlyzko, 2002]) und der Vernetzung von 400 Millionen Rechnern im Internet<sup>1</sup> (2006) vergingen nur wenige Jahrzehnte.

Für die Menschen im 21. Jahrhundert ging mit dieser Entwicklung insbesondere eine starke Informationsüberflutung einher, die durch die ständig wachsenden Bandbreiten, Bildschirmauflösungen und Ausgabemöglichkeiten gefördert wurde. An Wendepunkten dieser Entwicklung mussten neue Interaktionsmöglichkeiten geschaffen werden, die es den Benutzern erlauben, trotz der Informationsflut effizient und produktiv arbeiten zu können.

Ein Beispiel hierfür ist die Entwicklung des Fenster/Symbol/Menü/Zeigegerät-Paradigmas durch Xerox Parc, das in den Siebziger Jahren des 20. Jahrhunderts die intuitive Verarbeitung großer Datenmengen auf grafischem Wege ermöglichte [van Dam, 1997].

Darauf aufbauend erfolgte in den frühen Neunziger Jahren die Entwicklung des World Wide Webs mittels Hypertext-Oberflächen am Kernforschungszentrum CERN, um Versuchsdaten zwischen Computern auf einfache Weise verfügbar zu machen [Berners-Lee & Fischetti, 1999].

Zurzeit befindet sich die Informationsgesellschaft an einem weiteren Wendepunkt. In der Phase der Lochkartenrechner und Mainframes haben viele Benutzer einen Computer geteilt. Heutzutage verwendet ein Benutzer bereits mehrere Computer parallel und verbindet PCs mit MP3 Playern, Digitalkameras, Fernsehern und persönlichen digitalen Assistenten (PDAs). Die Informationsgesellschaft befindet sich an der Schwelle zum Zeitalter der ubiquitären Informationssysteme [Weiser, 1991], in der Menschen permanent von Computern umgeben sein werden. Dieser technologiegetriebenen Entwicklung fehlt es zurzeit jedoch an Modellen, Werkzeugen und humanwissenschaftlichem Grundlagenwissen, um effiziente Interaktionen mit diesen Systemen realisie-

---

<sup>1</sup>ISC Domain Survey, Juli 2006, <http://www.isc.org>

ren zu können [Nijholt et al., 2004]. Dies betrifft insbesondere die Generierung von Ausgaben in ubiquitären Informationssystemen. Benutzer sind im Büro, in öffentlichen Umgebungen oder im Privathaushalt bereits von zahlreichen Bildschirmen und Lautsprechern umgeben. Bisher wurden die Ausgabemöglichkeiten jedes Gerätes jedoch nur einzeln verwendet. Der PC-Bildschirm wurde zum Arbeiten im Internet genutzt. Der PDA wurde für die Terminplanung eingesetzt. Der Fernseher wurde zur Wiedergabe des Fernsehsignals verwendet. Verschiedene Untersuchungen haben jedoch gezeigt, dass die gleichzeitige Verwendung mehrerer Ausgabegeräte eine effizientere und angenehmere Interaktion als mit einzelnen Geräten erlaubt [Czerwinski et al., 2003; Grudin, 2001].

Erste Prototypen von Systemen, die mehrere Ausgabegeräte koppeln, existieren bereits [Kruppa, 2006; Myers et al., 2004; Rekimoto, 2000]. Der Nachteil dieser Systeme ist, dass sie für bestimmte Ausgabegeräte konzipiert sind. Wird dem ubiquitären Informationssystem ein neues Gerät hinzugefügt oder ein existierendes Gerät entfernt, so können sich diese Systeme nicht anpassen. Eine Lösung bieten hier Ansätze zur automatisierten Generierung von multimedialen Ausgaben mittels eines Planungsansatzes [André et al., 2005; Zancanaro et al., 2003; Zhou & Feiner, 2001], der auf einzelnen Geräten Inhalt und Ausgabeform automatisch an Ressourcen wie die Bildschirmauflösung anpasst.

### 1.2 Ziel

Das Ziel der vorliegenden Arbeit ist die Entwicklung von Verfahren, die es erlauben, vorher unbekannte Ausgabegeräte automatisch durch ein ubiquitäres Informationssystem anzusteuern. Diese Verfahren sollen es ermöglichen, Informationen auf mehrere Ausgabegeräte unter Verwendung verschiedener Ausgabeformen zu verteilen. Hierbei wird auf bereits existierenden Ergebnissen zur automatisierten Generierung von Einzelgeräteausgaben mittels eines Planungsansatzes aufgebaut.

Die Verfahren sollen als Teil eines Home-Entertainment-Systems, das einen Fernseher, einen PDA und einen Zusatzbildschirm als Ausgabegeräte koppelt, umgesetzt werden. Anschließend soll das System in Bezug auf die Akzeptanz und die Effizienz der generierten Präsentationen auf gekoppelten Ausgabegeräten untersucht werden. Hierbei sollen insbesondere die folgenden Forschungsfragen beantwortet werden:

1. **Welche Eigenschaften von Ausgabegeräten müssen modelliert werden, um sie automatisch in ubiquitäre Informationssysteme einbinden zu können?**

## 1.2. ZIEL

Ausgabegeräte lassen sich zunächst anhand der Medien (z. B. visuell oder auditoriell) unterscheiden, in denen sie Ausgaben darstellen. Ein weiteres Unterscheidungsmerkmal für Ausgabegeräte stellen die Ressourcen dar, die auf den Geräten verfügbar sind. Dies betrifft bei visuellen Ausgabegeräten die Auflösung des Bildschirms. Bei auditoriellen Ausgabegeräten betrifft dies die Leistung der Lautsprecher, mit denen eine Audioausgabe erfolgt. Neben der Hardware-Ressourcen unterscheiden sich Ausgabegeräte hinsichtlich der Ausgabe-Software, die auf ihnen vorhanden sind. So könnte z. B. auf einem leistungsstarken PC eine Sprachausgabe-Komponente vorhanden sein, die Text in gesprochene Sprache umwandeln kann, während auf einem mobilen PDA lediglich ein Browser zum visuellen Darstellen von Text verfügbar ist. In dieser Arbeit wird zunächst ein allgemeines Modell für Ausgabegeräte definiert, das unabhängig von einer konkreten Domäne ist. Dieses Modell wird anschließend für eine Home-Entertainment-Domäne erweitert und in einer Taxonomie für Ausgabedienste, -geräte und -ressourcen formalisiert.

### 2. **Ist ein Planungsansatz ebenfalls zur Generierung von Präsentationen auf gekoppelten Ausgabegeräten geeignet?**

Bei der automatisierten Generierung von multimedialen Präsentationen auf einzelnen Geräten hat sich ein Planungsansatz als effiziente Methode erwiesen. Es stellt sich die Frage, ob dieser Ansatz auch für Präsentationen auf gekoppelten Ausgabegeräten geeignet ist. Der Planungsansatz wird daher in einem Home-Entertainment-System, das sich aus einem PDA, einem Fernseher und einem elektronischen Bilderrahmen zusammensetzt, realisiert. Anschließend wird der Planungsansatz innerhalb von zwei Benutzerstudien evaluiert.

### 3. **Können multimediale Präsentationen auf gekoppelten Ausgabegeräten mit wenigen Ressourcen effizient umgesetzt werden?**

Komplexe multimediale Präsentationen wie z. B. die Sprach-/Grafik-Ausgaben eines animierten Charakters erfordern genügend Rechenkapazität auf den Ausgabegeräten. In verteilten Systemen werden zusätzlich genügend hohe Netzwerkbandbreiten benötigt, um die Medienobjekte einer Präsentation (z. B. Bild- oder Audiodateien) auf das Ausgabegerät zu laden. Anhand des in dieser Arbeit implementierten Home-Entertainment-Systems wird insbesondere untersucht, inwiefern sich umfangreiche multimediale Ausgaben im SMIL-Format [W3C Recommendation, 2005] auf einem Windows Mobile-PDA und einem Fernseher mit eingebettetem Linux-System umsetzen lassen.

#### 4. **Werden von Benutzern des Home-Entertainment-Systems Präsentationen auf gekoppelten Ausgabegeräten gegenüber konventionellen Einzelgeräteausgaben bevorzugt?**

Bisher existieren nur wenige Untersuchungen zu Präsentationen auf gekoppelten Ausgabegeräten. Es wurde beobachtet, dass Multi-Monitor-Systeme sehr gut durch die Benutzer akzeptiert werden. In dieser Arbeit wird das implementierte Home-Entertainment-System in Bezug auf seine Akzeptanz untersucht und mit den Ergebnissen der existierenden Studien verglichen. Insbesondere werden dabei unterschiedliche Konfigurationen der Ausgabegeräte (z. B. ein, zwei oder drei Bildschirme) sowie unterschiedliche Präsentationsformate (z. B. Sprachausgabe oder grafischer Text) einander gegenübergestellt. Zur Untersuchung dieser Forschungsfragen erfolgten zwei Studien, eine Fragebogenstudie und eine Laborstudie.

#### 5. **Arbeiten Benutzer mit gekoppelten Ausgabegeräten effizienter als mit Ausgaben auf einem einzelnen Gerät?**

Neben der Akzeptanz wird eine Präsentation ebenfalls durch ihre Effizienz charakterisiert. Die Effizienz einer Präsentation kann z. B. dadurch gekennzeichnet werden, wie einprägsam sie ist und wie viel Zeit sie benötigt, um alle Informationen darzustellen. Beide Effizienzkriterien sind umgekehrt proportional, d. h. typischerweise wird eine Präsentation weniger einprägsam, je schneller Informationen präsentiert werden. Diese Aspekte werden in Abhängigkeit der Gerätekonfigurationen und der Präsentationsformate in der Laborstudie anhand des implementierten Home-Entertainment-Systems untersucht.

### 1.3 Gliederung

In Kapitel 2 wird zunächst ein Überblick über ubiquitäre Informationssysteme und multi-modale Benutzerschnittstellen gegeben. Anschließend werden in Kapitel 3 existierende Ansätze für die Generierung von Einzelgeräteausgaben und Multi-Geräteausgaben verglichen und die Unterschiede sowie Gemeinsamkeiten zu dem in dieser Arbeit realisierten System analysiert. Kapitel 4 führt in die Grundlagen von multi-modalen Dialogsystemen ein.

In Kapitel 5 wird ein Dienst-, Geräte- und Ressourcenmodell für Ein-/Ausgabekomponenten definiert, das es ermöglicht, in verteilten multi-modalen Dialogsystemen Ausgabekomponenten automatisch in die Präsentationsgenerierung zu integrieren. Mit Hilfe des Modells wurde innerhalb von zwei Dialogsystemen eine multi-modale Präsentationsgenerierung realisiert (Kapitel 6). Abschnitt 6.1 beschreibt die erste Prototypimplementierung der Präsen-

### 1.3. GLIEDERUNG

tationsgenerierung. Abschnitt 6.2 beschreibt die Implementierung einer Präsentationsgenerierung für Multi-Geräteausgaben innerhalb eines ubiquitären Home-Entertainment-Systems.

In Kapitel 7 wird der Aufbau, die Durchführung und die Auswertung der ersten Studie zur Evaluierung der Akzeptanz des ubiquitären Home-Entertainment-Systems mittels einer Fragebogenstudie beschrieben. Im folgenden Kapitel 8 wird eine Laborstudie beschrieben, in der die Effizienz und Akzeptanz der durch den Home-Entertainment-Demonstrator generierten Multi-Geräteausgaben in Bezug auf ihre Erinnerungseffizienz untersucht wurden. Die Arbeit schließt mit einer Zusammenfassung und einem Ausblick in Kapitel 9.



## 2 Einleitung

Die zentralen Fragen, die in dieser Arbeit behandelt werden, betreffen ubiquitäre Informationssysteme, multi-modale Dialogsysteme und gekoppelte Ausgabegeräte. In diesem Kapitel werden die Begriffe anhand eines Überblicks über ihre Entwicklung und ihre Zusammenhänge eingeführt und in Beziehung gesetzt.

### 2.1 Ubiquitäre Informationssysteme

Der Ausgangspunkt für ubiquitäre Informationssysteme ist die äußerst schnelle Entwicklung leistungsfähigerer und kompakterer Computer. Der Intel-Mitarbeiter Gordon Moore beobachtete 1965, dass sich die Dichte von Transistoren in integrierten Schaltungen beim Ansatz minimaler Herstellungskosten jedes Jahr ungefähr verdoppelt [Moore, 1965]. In den folgenden Jahren wurde der Zeitrahmen der Verdopplung von einem Jahr auf zwei Jahre korrigiert. Diese Gesetzmäßigkeit ist bis zum Verfassen dieser Arbeit gültig geblieben und wurde unter dem Namen „Moore'sches Gesetz“ bekannt.

Abbildung 2.1 illustriert das Gesetz anhand der Transistorendichte von integrierten Schaltungen in Intel Prozessoren. Aus Darstellungsgründen wachsen die Abstände auf der vertikalen Achse nicht linear mit der Anzahl der Transistoren. Die Darstellung zeigt, wie sich nach etwa sieben Jahren die Transistorendichte verzehnfacht. Im Jahr 2000 blieb die Transistorendichte hinter der Vorhersage der Faustregel zurück. In den Jahren 2002 und 2005 erfolgte jedoch ein Anstieg, der die Vorhersagen übertraf. Ein Ende der Gültigkeit der Mooreschen Faustregel ist somit nicht absehbar.

Eine Folge dieser Entwicklung war, dass Ende der Siebziger Jahre der Personal Computer (PC) entwickelt wurde, der erstmals einen Computer für den Massenmarkt darstellte. Diese Entwicklung löste die Mainframe-Zeit ab, in der Computer mit den Ausmaßen eines Raumes mit Lochkarten bedient und von zahlreichen Benutzern geteilt wurde. Der PC ermöglichte es, dass jeder Benutzer über einen einzelnen Computer verfügen konnte. Heutzutage besitzen viele Benutzer bereits mehr als einen Computer, z. B. einen PC, ein Mobiltelefon und einen Laptop. Weiterhin sind viele Computer in Alltagsgegenständen unsichtbar eingebettet. Dies ist z. B. bei der Bremssteuerung des Antiblockiersystems im Auto der Fall.

Mark Weiser von Xerox Parc sah eine Entwicklung voraus, in der wir von unsichtbaren, in die Umgebung eingebetteten Computern umgeben sein werden [Weiser, 1991]. Er bezeichnete dies als ubiquitäre Informationsverarbeitung („Ubiquitous Computing“). Von Industrieseite wurde durch IBM der Begriff

## KAPITEL 2. EINLEITUNG

„Pervasive Computing“ geprägt. Innerhalb des Europäischen Forschungsprogramms IST (Information Society Technologies) wurde der stärker nutzerzentrierte Begriff „Ambient Intelligence“ formuliert. In dieser Arbeit verwenden wir den Begriff *ubiquitäre Informationssysteme* für vernetzte und verteilte Computersysteme, die durch eine gemeinsame Kommunikationsstruktur verbunden sind.

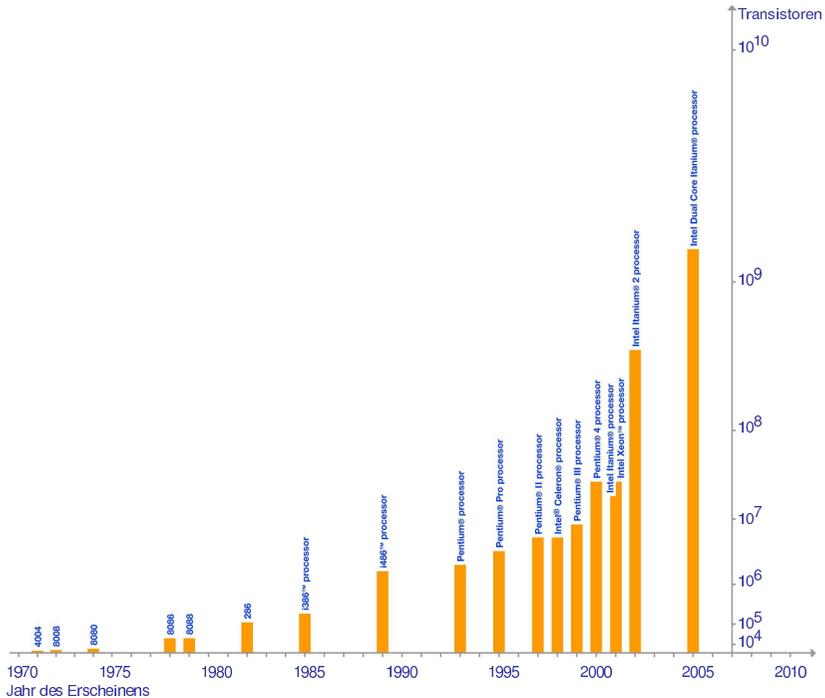


Abbildung 2.1: Anzahl von Transistoren in einer integrierten Schaltung eines Intel Prozessors (Copyright © 2005 Intel Corporation).

In der vorliegenden Arbeit sind insbesondere die Auswirkungen der ubiquitären Informationsverarbeitung auf den Home-Entertainment-Bereich relevant. Hier findet ebenfalls eine starke Einbettung von Informationstechnologie in Alltagsgegenstände statt. Im Zuge der Digitalisierung der Fototechnik haben digitale Bilderrahmen den Einzug in den Massenmarkt gefunden. Abbildung 2.2 zeigt ein Modell der Firma Philips.

## 2.1. UBIQUITÄRE INFORMATIONSSYSTEME



Abbildung 2.2: Digitaler Bilderrahmen Philips 9FF2CWO/00 (Foto: Philips).

Ausdruck der Verschmelzung von Informationstechnologie und Home-Entertainment sind ebenfalls die „intelligenten Häuser,, wie z. B. das inHaus in Duisburg, das T-Com Haus in Berlin oder das Haus der Gegenwart in München. In diesen Häusern werden durch die Verwendung gemeinsamer Standards (z. B. für drahtlose Funknetzwerke) Haussteuerung und Unterhaltungselektronik vernetzt.



Abbildung 2.3: Philips HomeLab in Eindhoven (Foto: Philips).

Dies ermöglicht eine einheitliche Bedienung elektronischer Geräte. So ist es z. B. möglich, über den PC sowohl die Wiedergabe von Filmen zu steuern als auch die Haustür zu überwachen. Abbildung 2.3 zeigt das Philips HomeLab in Eindhoven, in dem Benutzer bei der Interaktion mit einer intelligenten Wohnung von den Forschern mittels Kameras beobachtet werden können.

## 2.2 Multi-modale Dialogsysteme

Bei der Verwendung von ubiquitären Informationssystemen stellt sich die Frage nach einer geeigneten Bedienung. Computer mit einem Bildschirm werden typischerweise durch das von Xerox Parc entwickelte Fenster-Symbol-Menü-Zeigergerät-Paradigma („Window-Icon-Menu-Pointing Device“, WIMP) bedient, das bis heute die Grundlage der Microsoft Windows- und X Window-Benutzerschnittstellen bildet. Für ubiquitäre Informationssysteme, die sich aus mehreren Interaktionsgeräten zusammensetzen können, sind jedoch flexiblere Benutzerschnittstellen nötig.

*Multi-modale Dialogsysteme* stellen intuitivere, flexiblere und fehlertolerantere Schnittstellen zu Informationstechnologie dar. Sie verallgemeinern das WIMP-Paradigma, indem sie weitere Modalitäten unterstützen, die von den Benutzern alternativ zu oder in Kombination mit den WIMP-Modalitäten verwendet werden können.

Eine Modalität bezeichnet dabei eine Interaktionsform, durch die Informationen zwischen Mensch und Computer ausgetauscht werden können. Dabei wird zwischen Eingabemodalitäten (z. B. Spracheingaben oder Gesteneingaben) und Ausgabemodalitäten (z. B. Bildausgaben oder Textausgaben) unterschieden.

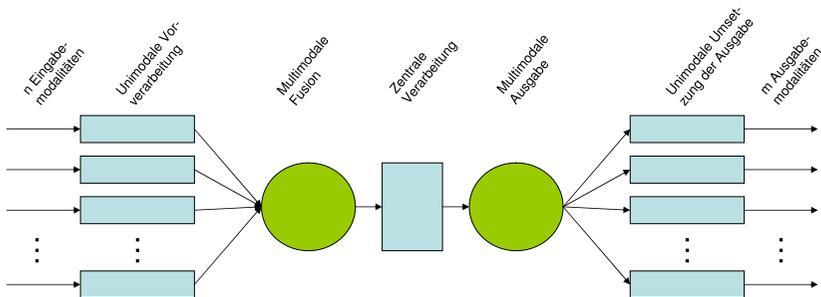


Abbildung 2.4: Prinzip von multi-modalen Dialogsystemen (Vorlage: Stefan Kopp).

## 2.2. MULTI-MODALE DIALOGSYSTEME

Abbildung 2.4 illustriert das Prinzip eines multi-modalen Dialogsystems. Die Eingabemodalitäten werden zunächst vorverarbeitet. Dies kann z. B. durch einen Parser geschehen, der die von einem Spracherkenner erkannte Sprache oder die von einem Gestenerkenner erkannten Gesten analysiert und klassifiziert. Anschließend erfolgt die Fusion der einzelnen Modalitäten. In diesem Schritt wird die sprachliche Benutzeranfrage „Einschalten!“ mit der korrespondierenden Zeigegeste auf eine Lampe zusammengeführt und daraus die Benutzerintention, die Lampe einzuschalten, abgeleitet.

Daraufhin erfolgt eine zentrale Verarbeitung, in der zu der Anfrage des Benutzers eine passende Reaktion des Systems ermittelt wird. Die multi-modalen Ausgaben des Systems werden anschließend auf die Ausgabemodalitäten verteilt, die die einzelnen Ausgaben uni-modal umsetzen. Dabei müssen die Ausgabemodalitäten bzgl. ihres grafischen und zeitlichen Layouts koordiniert werden. So ist es i. Allg. nicht sinnvoll, zwei Bilder auf demselben Bildschirm so darzustellen, dass das eine Bild durch das andere Bild verdeckt wird.

Das Ziel multi-modaler Dialogsysteme ist eine natürlichere Interaktion zwischen Mensch und Computer, die die volle Bandbreite der Modalitäten unterstützt, die von Menschen intuitiv verwendet werden. Die Kommunikation über mehrere Modalitäten hinweg bietet verschiedene Vorteile [Oviatt, 2003]. Multi-modale Systeme überlassen dem Benutzer die Wahl der Modalitäten, die für ihn am besten geeignet sind. Benutzer, die viel Erfahrung mit Informationstechnologie haben, können das System wie bisher mit Tastatur und Maus bedienen, während Benutzer mit wenig Erfahrung das System effizienter mit Sprache bedienen können. In der Tat ziehen Benutzer häufig die Kombination mehrerer Modalitäten gegenüber der Interaktion mit einer einzigen Modalität vor [Oviatt, 1997].

Speziell mobile Anwendungen können von der flexiblen Verwendung von mehreren Modalitäten profitieren, da sich die Umgebungsbedingungen sehr stark ändern können [Jöst et al., 2005]. Innerhalb der abgeschlossenen Umgebung eines Fahrzeugs kann ein PDA zur Routennavigation per Sprache bedient werden. Ein Fußgänger in einer Einkaufsstraße wird die Routennavigation jedoch eher mit dem Stift des PDAs durchführen, um seine Privatsphäre zu wahren. Ein besonderer Vorteil multi-modaler Benutzerschnittstellen ist ihre Fehlertoleranz. Dies liegt darin begründet, dass Benutzer bei der Wahl zwischen mehreren Modalitäten intuitiv die auswählen, die die geringste Fehlerwahrscheinlichkeit hat. So wurde in einem Experiment beobachtet, dass Teilnehmer für die Eingabe von komplexen Fremdwörtern in ein Informationssystem handschriftliche Eingaben bevorzugten. Für alle anderen Wörter verwendeten sie jedoch die schnellere Spracheingabe [Oviatt & van Gent, 1996].

## KAPITEL 2. EINLEITUNG

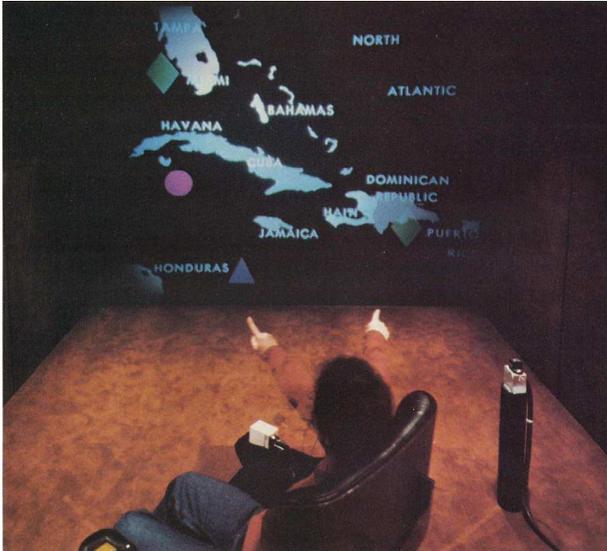


Abbildung 2.5: Put-that-there System (Foto: [Bolt, 1980], Copyright © 1980 ACM Inc., genehmigter Nachdruck).

Die ersten multi-modalen Systeme hatten multi-modale Eingaben als Schwerpunkt und entstanden im Zuge der technischen Fortschritte in der Spracherkennung in den Siebziger und Achtziger Jahren des 20. Jahrhunderts. Das Scholar-System [Carbonell, 1970] unterstützte einfache Zeigegesten auf einen Bildschirm in Kombination mit natürlicher Sprache. Das System stellte ein Lernprogramm für Geographie dar.

Das Put-That-There-System [Bolt, 1980], das Xtra-System [Wahlster, 1988] und das Cubricon-System [Neal & Shapiro, 1991] stellten weitere Systeme dar, die Sprache und Gesten auf unterschiedlichen Detailebenen unterstützten. Im Unterschied zum Scholar-System können Sprach- und Gesteneingaben in diesen Systemen parallel erfolgen. Abbildung 2.5 zeigt eine Interaktion mit dem Put-That-There System von Bolt. Die Anfrage des Benutzers lautet „Verschiebe das blaue Dreieck an die Position rechts vom grünen Quadrat“. Das Bild zeigt die fotografische Überlagerung der zwei Zeigegesten, die parallel zur Spracheingabe erfolgen.

Neben der Fusion multi-modaler Eingaben wurde ebenfalls die Generierung multi-modaler Ausgaben untersucht. Das Ziel war hier, Informationen mittels der Modalitäten zu präsentieren, die vom Benutzer bevorzugt werden und

## 2.2. MULTI-MODALE DIALOGSYSTEME

in denen sich die auszugebenden Daten gut repräsentieren lassen. So können z. B. geographische Daten gut auf grafischem Wege mittels einer zweidimensionalen Landkarte dargestellt werden. Eine rein akustische Ausgabe komplexer geographischer Informationen gestaltet sich jedoch schwierig.

Sprache kann hingegen sowohl visuell als grafischer Text als auch auditoriell als gesprochene Sprache wiedergegeben werden. Der Schwerpunkt der ersten Präsentationssysteme war die Koordination von Sprache mit Bildern oder Animationen. Die Präsentationen konnten in ihrer Größe an unterschiedliche Formate angepasst werden. Weiterhin konnte der Benutzer Präferenzen für das Präsentationsformat angeben.



Abbildung 2.6: Ubiquitäres Informationssystem: Embassi Home-Entertainment-Demonstrator (Foto: European Media Laboratory).

Prototypen dieser Präsentationssysteme sind AIMI [Burger & Marshall, 1993] und Cubricon, die Sprache mit Kartendarstellungen als Teil von militärischen Missionsplanungen koordinierten. Im SAGE-System [Roth et al., 1991] werden Texte und Grafiken zur Darstellung quantitativer Daten wie z. B. Finanzdaten verwendet. Das WIP-System [André et al., 1993] generiert Präsentationen aus statischen Bildern, grafischem Text, gesprochener Sprache und 3D-Animationen zur Illustration der Verwendung oder Reparatur von technischen Geräten wie Espressomaschinen oder Modems. Das System passte die Präsentationen in Abhängigkeit von zwei Benutzerkategorien (Anfänger oder Experte) sowie vier bevorzugte Präsentationsmodi (Text, Grafik, Animationen, keine) an. Das Präsentationssystem COMET („Coordinated Multimedia Ex-

## KAPITEL 2. EINLEITUNG

planung Testbed“, [Feiner & McKeown, 1998]) generiert Text-Bild-Präsentationen, die die Reparatur einer Sende-/Empfangsanlage beschreiben.

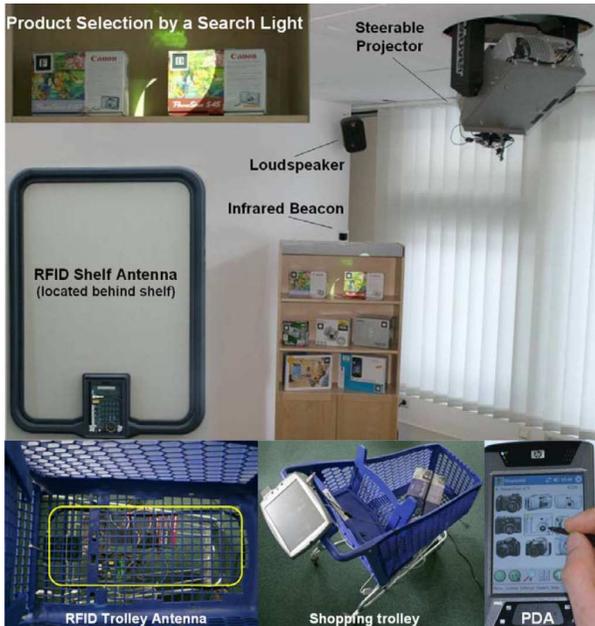


Abbildung 2.7: Ubiquitäres Informationssystem: Intelligenter Einkaufsführer (Foto: [Wahlster & Wasinger, 2006]).

Mit dem Erfolg von ubiquitären Informationssystemen und der fortschreitenden Dezentralisierung von Applikationen durch tragbare Computer und drahtlose Kommunikation wurden multi-modale Benutzerschnittstellen auch zunehmend in ubiquitären Informationssystemen eingesetzt. Im Embassi-Projekt [Herfet et al., 2001] wurde ein multi-modales Dialogsystem zur Steuerung eines Home-Entertainment-Systems realisiert. Das System umfasst Eingaben per Sprache und per Gestik sowie Ausgaben mittels Sprache, Grafik sowie eines animierten Charakters (Abbildung 2.6). Mit dem System können ein Ventilator, ein Festplattenrecorder, ein Fernseher und Fensterläden gesteuert werden.

Ein weiteres Beispiel für ein ubiquitäres Informationssystem, das multi-modale Interaktionen unterstützt, ist das SmartKom-Projekt [Wahlster, 2006], das ebenfalls ein Home-Entertainment-Szenario mit Sprach- und Gestikeingaben

### 2.3. GEKOPPELTE UND FLEXIBLE BILDSCHIRME

realisiert. Innerhalb des Specter-Projekts wurde ein multi-modaler Einkaufsführer realisiert [Wahlster & Wasinger, 2006]. Der Benutzer kann auf einem PDA Spracheingaben, handschriftliche Eingaben sowie Zeigegesten kombinieren, um sich Informationen über ein Produkt anzeigen zu lassen. Das System ist ebenfalls in der Lage zu erkennen, ob der Benutzer ein Produkt in den Händen hält und um welches Produkt es sich handelt. Das System ermittelt die Anwesenheit eines Benutzers mittels Infrarotbaken (Infrared Beacons) sowie durch Funkerkennung (RFID). Als zusätzliche Hervorhebung von gesuchten Produkten wird ein Projektor (SearchLight) verwendet. Abbildung 2.7 zeigt den Aufbau des ubiquitären Informationssystems.

#### 2.3 Gekoppelte und flexible Bildschirme

Bildschirme spielen in ubiquitären Informationssystemen eine besondere Rolle. Ähnlich wie integrierte Schaltungen wurden Bildschirme in Folge der Verwendung von Projektions-, Plasma- oder LCD-Technologien preisgünstiger und leistungsfähiger.



Abbildung 2.8: Berliner Fenster: Wettervorhersage auf zwei gekoppelten Monitoren.

Als Folge dieser Entwicklung sind heutzutage Zusatzbildschirme in zahlreiche Gegenstände eingebettet wie z. B. in Fernbedienungen oder Drucker. Ebenso werden mehrere Bildschirme kombiniert, um die Bildschirmfläche zu vergrößern und um Applikationen besser auf der Arbeitsfläche verteilen zu können. Dies wird in dieser Arbeit als *Kopplung* von Ausgabegeräten bezeichnet. Allgemein wird von einer Kopplung von Ausgabegeräten gesprochen, wenn Inhalte auf mehrere Ausgabegeräte verteilt und koordiniert ausgegeben werden können. Ein Beispiel für die Kopplung von Ausgabegeräten ist eine Fernbedienung mit einem integrierten Bildschirm, auf dem Zusatzinformationen zum

## KAPITEL 2. EINLEITUNG

aktuellen Programm auf dem Fernseher angezeigt werden (z. B. der Name des Fernsehkanals). Ein weiteres Beispiel ist eine Videowand, bei der mehrere Bildschirme zu einem großen Bild kombiniert werden.



Abbildung 2.9: Nintendo DS Videospielekonsole (Foto: Wikipedia).

Die Kopplung von Bildschirmen wird heutzutage bereits erfolgreich eingesetzt. An IT-Arbeitsplätzen werden zwei bis drei Monitore parallel eingesetzt, um die Bildschirmfläche zu vergrößern. Gekoppelte Bildschirme finden sich ebenfalls in öffentlichen Bereichen und im Privathaushalt. In der Berliner U-Bahn wird eine Kopplung von zwei Monitoren verwendet, um Werbe- und Informationssendungen mittels Texten, Bildern und Animationen darzustellen (Abbildung 2.8). In Videospielsystemen wie dem Nintendo DS werden zwei Bildschirme gleichzeitig verwendet, um Daten darzustellen (Abbildung 2.9). So wird bei Spielen häufig ein Bildschirm für eine zweidimensionale Übersichtskarte und ein Bildschirm für eine dreidimensionale Ansicht aus dem Blickwinkel des Spielcharakters verwendet.

Die Flexibilisierung der Bildschirmtechnologien zeigt sich nicht nur durch die Herstellung von kompakteren, flacheren oder größeren Bildschirmen. Die Entwicklung von rollbaren Bildschirmen weist darauf hin, dass in Zukunft Bildschirme Eingang in viele Lebensbereiche finden werden. Abbildung 2.10 zeigt den Prototypen eines rollbaren Bildschirms der Firma Polymer Vision. Im Zusammenhang mit ubiquitären Informationssystemen lässt sich die Hypothese aufstellen, dass in Zukunft nicht nur die Informationsverarbeitung, sondern auch die Ausgabegeräte ubiquitär sein werden.

## 2.4. ZUSAMMENFASSUNG



Abbildung 2.10: Rollbarer Bildschirm der Firma Polymer Vision (Foto: Polymer Vision).

### 2.4 Zusammenfassung

In diesem Kapitel wurden die Grundkonzepte, auf denen diese Arbeit aufbaut, eingeführt. Die Informationstechnologie entwickelt sich zu Systemen hin, die sehr stark verteilt und vernetzt sind. Teilweise sind sie unsichtbar in die Umgebung eingebettet. Diese Systeme werden unter dem Überbegriff „ubiquitäre Informationssysteme“ zusammengefasst. Das Konzept der ubiquitären Informationssysteme wurden zu multi-modalen Dialogsystemen in Beziehung gesetzt, die eine natürlichere und effizientere Kommunikation zwischen Benutzern und Computern als bei den traditionellen grafischen Benutzeroberflächen zum Ziel haben.

Der Schwerpunkt dieser Arbeit ist die Betrachtung gekoppelter Ausgabegeräte. Im Bereich gekoppelter Bildschirme existieren bereits erste Produkte, jedoch stand bisher eine vollständige Betrachtung gekoppelter Ausgabegeräte im Zusammenhang mit ubiquitären Informationssystemen und multi-modalen Dialogsystemen noch aus.



### 3 Stand der Wissenschaft

Die bisherigen Arbeiten zur automatisierten Generierung von multi-modalen Präsentationen auf gekoppelten Ausgabegeräten lassen sich anhand von drei Forschungsrichtungen illustrieren. Nach der Einführung der Grundbegriffe (Abschnitt 3.1) erfolgt in Abschnitt 3.2 eine Gegenüberstellung von Studien und Ansätzen zur automatisierten Generierung von Präsentationen auf einem Gerät. Anschließend werden empirische Untersuchungen und Ansätze zur Generierung von Präsentationen mit gekoppelten Ausgabegeräten verglichen (Abschnitt 3.3). Da für die Ad-hoc-Integration von neuen Ausgabegeräten insbesondere geeignete Selbstbeschreibungen der multi-modalen Ausgabemöglichkeiten nötig sind, wird in Abschnitt 3.4 der Stand der Forschung bzgl. der Modellierung multi-modaler Ausgaben untersucht.

#### 3.1 Terminologie

Da die Begriffe „Modalität“ und „Medium“ in verschiedenen Forschungsbereichen unterschiedlich verwendet werden, erfolgt in diesem Abschnitt zunächst eine Definition der grundlegenden Begriffe, die für diese Arbeit relevant sind. Laut Bernsen bezeichnet der Begriff *Modalität* einen Kanal, durch den Informationen zwischen Menschen oder zwischen Mensch und System ausgetauscht werden können [Bernsen, 2001]. Eine Modalität stellt somit ein Repräsentationsformat dar. Beispiele für Eingabemodalitäten sind gesprochene Sprache oder Zeigegesten. Beispiele für Ausgabemodalitäten sind Sprachausgaben, Bilder, Tabellen oder Braille-Schrift. Dieser Informationsaustausch ist dabei an ein *Medium* gebunden.

Das Medium stellt die physikalische Umsetzung der Informationen dar. Diese kann z. B. im visuellen, auditoriellen oder taktilen Medium erfolgen. Im Bezug auf Medien verwenden wir im Unterschied zu [Bernsen, 2001] die Begriffe „visuell“, „auditoriell“ und „taktile“, da die von Bernsen verwendeten Begriffe „grafisch“, „akustisch“ und „haptisch“ technikbezogen sind und die menschlichen Sinneswahrnehmungen nicht erfassen.

Maybury und Wahlster [Maybury & Wahlster, 1998] hingegen unterscheiden zwischen der Modalität als Interaktionsstil (z. B. visuell, auditoriell oder taktil) und dem Code als Kodierungsformat (z. B. Sprache, Grafik oder Haptik). Dabei kann derselbe Code (z. B. Sprache) in unterschiedlichen Modalitäten übertragen werden (z. B. visuell durch Zeichensprache oder auditoriell durch gesprochene Sprache). Das Medium der Interaktion wird dabei als technischer Informationsträger (z. B. Papier oder CD-ROM) verstanden, auf dem der Code in einer Modalität umgesetzt wird.

Die Begriffe „Modalität“ bei Maybury und Wahlster sowie „Medium“ bei Bernsen haben dieselbe Bedeutung. Der Begriff „Modalität“ bei Bernsen entspricht einem „Code“ bei Maybury und Wahlster, der auf eine bestimmte „Modalität“ bezogen ist, z. B. gesprochene Sprache, die Informationen mittels einer Sprachkodierung auditoriell übermittelt.

Da im Verlauf dieser Arbeit auf weitere Arbeiten von Bernsen Bezug genommen wird verwenden wir im Folgenden Bernsens Definitionen von Modalität und Medium. In Analogie zu Modalitäten werden in dieser Arbeit darüber hinaus Software-Komponenten oder Geräte ebenfalls als visuell, auditoriell oder taktil bezeichnet, falls sie Modalitäten im jeweiligen Medium unterstützen. Eine Komponente zur Ausgabe von grafischem Text stellt somit eine visuelle Ausgabekomponente dar. Ein Lautsprecher repräsentiert ein auditorielles Ausgabegerät.

Bernsens Taxonomie unterscheidet zwischen den folgenden Modalitäteneigenschaften. *Linguistische* Modalitäten sind Modalitäten, in denen Informationen mittels einer Sprache kodiert sind. Beispiele für linguistische Modalitäten sind Schriftsprache oder Gestensprache. Weiterhin unterscheidet Bernsen zwischen *statischen* Modalitäten (z. B. einem Bild) und *dynamischen* Modalitäten (z. B. einem Film).

*Arbiträre Modalitäten* sind Modalitäten, die sich nicht auf eine gegebene Semantik angewiesen sind. Zum Beispiel ist beim Klingeln eines Weckers weniger entscheidend, in welcher Tonhöhe oder Tonfolge der Wecker klingelt. Das Klingeln an sich wird eher als auditorielle Stimulation verwendet. Dasselbe gilt für das grafische Hervorheben eines Wortes durch eine Hintergrundfarbe. *Analoge* Modalitäten sind Modalitäten, die Informationen durch visuelle oder auditorielle Ähnlichkeit repräsentieren. Nicht-analoge Modalitäten wie z. B. Text repräsentieren Konzepte hingegen durch eine feste Semantik.

### 3.2 Multi-modale Ausgaben auf einem Gerät

Die Effekte von multi-modalen Präsentationen auf einem Ausgabegerät sind bereits sehr gut untersucht. Dies betrifft sowohl die Effekte der Wahl unterschiedlicher Modalitäten auf die kognitive Last als auch die Verwendung von animierten Charakteren. Im Folgenden wird zunächst ein Überblick über die für die Arbeit relevanten Ergebnisse gegeben. Anschließend werden Ansätze zur automatisierten Generierung von Präsentationen für ein Ausgabegerät betrachtet.

#### 3.2.1 Empirische Untersuchungen

[Tindall-Ford et al., 1997] haben Text-Bild- und Sprache-Bild-Ausgaben in Bezug auf ihren Lerneffekt untersucht. Hierbei wurde der „Modalitäteneffekt“

### 3.2. MULTI-MODALE AUSGABEN AUF EINEM GERÄT

beobachtet, der eine Überlegenheit von Sprache-Bild-Ausgaben gegenüber Text-Bild-Ausgaben zeigte. Tindall-Ford, Chandler und Sweller begründen dies mit Baddeleys Modell des Arbeitsgedächtnisses [Baddeley, 2001]. In diesem Modell unterscheidet Baddeley zwischen drei voneinander unabhängigen Systemen, einem visuell-räumlichen System („Visuo-Spatial Sketchpad“), einem auditoriellen System („Phonological Loop“) und einem System für übergeordnetes Wissen („Episodic Buffer“). Das visuell-räumliche System verarbeitet sowohl Text- als auch Bildinformationen. Dadurch entsteht bei Text-Bild-Ausgaben der „Split-Attention-Effekt“, da das visuell-räumliche System Text und Bild nicht parallel verarbeiten kann. Der Betrachter der Präsentation ist gezwungen, seine Aufmerksamkeit zwischen Text und Bild aufzuteilen.

Bei den Sprache-Bild-Ausgaben wird hingegen durch Sprache das auditorielle System angesprochen. Dadurch können Sprache und Bild parallel verarbeitet werden und die kognitive Belastung sinkt. Sweller hat gezeigt, dass dieser Effekt jedoch nur bei Material mit hohem Schwierigkeitsgrad beobachtet werden kann. Dieser Effekt wurde in weiteren Studien bestätigt, z. B. [Mayer & Moreno, 1998].

[Tabbers et al., 2001] haben gezeigt, dass der Modalitäteneffekt nicht beobachtet werden kann, falls in den Text-Bild-Ausgaben genügend Verbindungen zwischen Text und Bild existieren, z. B. indem die Textstellen direkt an die Teile eines Diagramms geschrieben werden, worauf sie sich beziehen. Hierdurch wird der Split-Attention-Effekt gering gehalten.

In einer dieser Arbeit vorhergehenden Studie [Elting et al., 2002] wurden unterschiedliche Kombinationen von Ausgabemodalitäten auf drei Geräten (Desktop-PC, Fernseher und PDA) untersucht. Das Ergebnis war, dass sich der Modalitäteneffekt auf den drei Geräten unterschiedlich auswirkt. Besonders deutlich war der Unterschied zwischen der kognitiven Last von Text-Bild- und Bild-Sprache-Ausgaben auf dem PDA, mit dessen Handhabung die meisten Teilnehmer nicht vertraut waren. Dies korrespondiert mit der Beobachtung von Tabbers und Kollegen, dass der Modalitäteneffekt nur unter kognitiver Last auftritt. In [Oviatt et al., 2004] findet sich ein Überblick über Theorien zur kognitiven Last mit Bezug zu multi-modalen Dialogsystemen. Der Schwerpunkt liegt hierbei auf multi-modalen Eingaben.

Ein weiteres Resultat aus dem Bereich der Sozialwissenschaften ist, dass Menschen Medien unbewusst wie soziale Akteure behandeln. Dies dokumentieren [Reeves & Nass, 1999] mit Hilfe von mehreren Studien für Computer, Fernseher und Kinofilme. Die Ergebnisse von Reeves und Nass sind ein Hinweis darauf, dass animierten Charakteren als Ansprechpartner in ubiquitären Informationssystemen eine wichtige Rolle zukommt.

[Van Mulken et al., 1998] hat die Auswirkung eines animierten Charakters auf die Lerneffizienz untersucht. Das Ergebnis des Experiments ist, dass die Lerneffizienz durch den Charakter nicht signifikant gesteigert werden konnte. Jedoch haben die Versuchspersonen (Vpn) Präsentationen mit dem Charakter als unterhaltsamer eingestuft. Die Autoren begründen dies damit, dass der zusätzliche Lerneffekt, der durch die motivierenden Erläuterungen des animierten Charakters entsteht, durch die Ablenkung kompensiert wird, die ein animierter Charakter erzeugt.

### 3.2.2 Technische Umsetzung

Template-basierte Ansätze stellen die einfachste Methode dar, um multi-modale Präsentationen automatisch anzupassen. Bei Webseiten werden hierzu typischerweise XSLT [W3C Recommendation, 1999], JSP<sup>1</sup> oder PHP<sup>2</sup> verwendet, um Webseiten dynamisch an verschiedene Kontexte anpassen zu können. Diese Ansätze haben jedoch den Nachteil, dass jedes einzelne Layout manuell definiert werden muss. Möchte man das zugrunde liegende Layout ändern, so muss jedes einzelne Template manuell modifiziert werden, da die Gemeinsamkeiten der Templates nicht erfasst werden können.

Ein flexiblerer Ansatz ist die wissensbasierte Generierung von Präsentationen, die durch die Modellierung von abstraktem Präsentationswissen in der Lage ist, Gemeinsamkeiten von unterschiedlichen Präsentationen zu erfassen. Ein weiterer Ansatz ist die Formulierung des Präsentationsproblems als numerisches Optimierungsproblem. Beide Ansätze werden im Folgenden gegenüber gestellt.

#### Wissensbasierte Ansätze

Ein weit verbreiteter Ansatz zur Generierung von multi-modalen Präsentationen ist ein Planungsansatz [Towns et al., 1998; Kerpedjiev et al., 1997; André et al., 1993]. Typischerweise entsprechen bei den Planungsansätzen die Planungsoperatoren multimedialen Akten, die den kommunikativen Akten in der Textgenerierung entsprechen [André, 2003]. Der Vorteil eines Planungsansatzes ist, dass das Präsentationswissen transparent aus den Operatoren hervorgeht. Diese explizite Repräsentation vereinfacht die Pflege und Wiederverwendung von Präsentationsstrategien.

Das Autobriefer System [André et al., 2005] erstellt multi-modale Präsentationen von Briefings. Mittels eines GUI kann der Benutzer die strukturelle und zeitliche Ordnung der Präsentation festlegen. Als Präsentationsformate werden Power Point und SMIL [W3C Recommendation, 2005] unterstützt. Die

---

<sup>1</sup>Java Server Pages, <http://java.sun.com/products/jsp/>

<sup>2</sup>PHP Hypertext Preprocessor, <http://www.php.net/>

### 3.2. MULTI-MODALE AUSGABEN AUF EINEM GERÄT

Komposition der Präsentation erfolgt mittels eines hierarchischen Planungsansatzes in Verbindung mit einem linearen Constraint-Löser. Dabei werden zeitliche Constraints und Layout-Constraints mittels einer Intervallalgebra [Allen, 1983] formuliert und durch den Constraint-Löser in ein konkretes Layout umgesetzt.

Das Cuypers System [van Ossenbruggen et al., 2000] verwendet ähnlich zu Autobriefer einen Constraint-basierten Ansatz, zur adaptiven Generierung von SMIL-Präsentationen. Wie Autobriefer werden dabei quantitative Constraints und qualitative Constraints unterstützt, um ein grafisch-zeitliches Layout zu definieren. Das Präsentationswissen und die Constraints werden in Prolog formuliert.

Das in [Zancanaro et al., 2003] beschriebene System komponiert in einem Museumskontext Präsentationen von Standbildern, die durch Sprache sowie kinematografische Übergänge (z. B. Zoom oder Überblenden) ergänzt werden. Die Autoren verwenden dazu ein regelbasiertes System, in dem kinematografische Regeln sowie die zur Verfügung stehende Medienobjekte gespeichert werden. Das Ergebnis ist eine Video-ähnliche Macromedia Flash Präsentation, die auf einem PDA abgespielt wird. Weitere Gerätetypen werden jedoch nicht unterstützt.

Das Improve System [Zhou & Feiner, 2001] wird verwendet, um personalisierte medizinische Briefings zu erstellen. Das System generiert ebenfalls Statistiken über Computer-Netzwerkverkehr. Die Präsentationen bestehen aus einer Sequenz von Animationen, die als „visueller Diskurs“ bezeichnet werden. Dies geschieht in Improve durch einen hierarchischen Planungsalgorithmus. Der Schwerpunkt des Systems liegt in der Berechnung von grafischen Layouts sowie von kohärenten räumlichen und zeitlichen Übergängen zwischen den Präsentationen.

Das Multiface System [Channarukul et al., 2004] verwendet STRIPS-Planung (vgl. [Nau et al., 2004]), um SMIL-Präsentationen zu generieren. Das System wurde in Java realisiert und basiert auf dem Xsmiles-Tool zur Ausgabe von SMIL-Inhalten. Das System verwendet Selbstbeschreibungen, um SMIL-Präsentationen an mehrere Gerätetypen (z. B. Desktop-PC, PDA oder Mobiltelefon) anzupassen. Die Autoren geben jedoch keine weiteren Informationen dazu, wie die Planungsoperatoren aussehen, wie die Selbstbeschreibungen definiert sind oder wie die Anpassung der SMIL-Präsentationen erfolgt. Die Präsentationsstrategien werden in Prolog formuliert.

Im Projekt SmartKom [Wahlster, 2006] wurden ein multi-modales Dialogsystem realisiert, das Eingaben durch Sprache und Zeigegesten unterstützt. Ausgaben werden realisiert durch einen animierten Agenten und ein GUI [Reit-

hinger et al., 2003]. Es existieren vier Szenarien, die jeweils unterschiedliche Ausgabebeugeräte unterstützen. Im Home-Entertainment-Szenario wird ein Touchpad zur Interaktion mit dem System verwendet. Die Ausgabe des Systems wird koordiniert durch den PrePlan-Präsentationsplaner [André, 1995]. Die Sprachausgabe wird dabei mit den Lippenbewegungen des animierten Charakters synchronisiert. Ebenso werden die Gesten des animierten Agenten mit Elementen in dem GUI (z. B. Icons auf einem Stadtplan) koordiniert. Die multi-modalen Ausgaben des Systems sind allerdings stets auf ein Ausgabebeugerät und auf feste Layouts beschränkt.

#### Numerische Ansätze

Im Supple-System [Gajos et al., 2005] wird ein modellbasierter Ansatz mit einem numerischen Optimierungsverfahren kombiniert, um für unterschiedliche Applikationen dynamisch ein GUI zu generieren. Der Vorteil des numerischen Ansatzes ist, dass im Unterschied zu rein wissensbasierten Ansätzen mittels Kostenfunktionen eine große Anzahl von Applikationen unterstützt werden kann. Der Nachteil ist jedoch, dass bereits kleine Änderungen in der Kostenfunktion große Unterschiede in dem erstellten GUI zur Folge haben können. Dies ist durch numerische Instabilitäten des Optimierungsalgorithmus bedingt. Aus diesem Grund verfügt Supple über ein Werkzeug, mit dem ein automatisch generiertes GUI interaktiv durch den Benutzer angepasst werden kann, falls der Algorithmus versagt.

Das RIA-System [Zhen et al., 2005] ist ein Dialogsystem, das Benutzer bei der Suche nach Häusern, Grundstücken und Hotels unterstützt. Das System erstellt je nach Benutzeranfrage fließende Übergänge zwischen zwei Ausgaben, was als „visuelle Kontinuität“ bezeichnet wird. Ein weiterer Schwerpunkt des Systems ist das Ausblenden unnötiger Informationen, die für den Benutzer nicht interessant scheinen sowie das Vergrößern, Hervorheben und Annotieren von Objekten, an denen der Benutzer besonders interessiert ist.

Zhen und Kollegen verwenden dazu einen numerischen Optimierungsalgorithmus. Dieser Algorithmus birgt jedoch zwei Nachteile. Zum einen müssen die Werte der Kostenfunktion sorgfältig gewählt werden. In RIA betrifft dies z. B. die Relevanz von grafischen Objekten im aktuellen Dialogschritt. Andererseits lässt sich bei Fehlern des Systems nur schwer nachvollziehen, welche Kosten zu einer Entscheidung beigetragen haben und modifiziert werden müssen.

Das System von Stent und Guo [Stent & Guo, 2005] verwendet einen datengetriebenen Ansatz, um Text-Bild-Präsentationen in HTML zu generieren. Im Unterschied zu anderen Systemen existieren keine vordefinierten Präsentationsstrategien. Diese werden aus einem Korpus von existierenden Präsentationen gelernt. Das System verwendet dazu einen Korpus von 100 Webseiten,

### 3.3. MULTI-MODALE AUSGABEN IN VERTEILTEN GERÄTEUMGEBUNGEN

auf denen Computer zum Verkauf angeboten werden.

Diese Webseiten werden zunächst nach Inhaltstypen klassifiziert, was mittels einer Ontologie und eines Lexikons erfolgt. So wird z. B. der Wortstamm „GB“ als Abkürzung der Einheit „Gigabyte“ dem Konzept Speicherplatz zugeordnet. Anschließend werden die Wahrscheinlichkeiten von Präsentationselementen und deren Kontext berechnet, z. B. wie wahrscheinlich eine Tabelle der Produkteigenschaften zwischen einem Bild des Produkts und einer textuellen Beschreibung platziert wird. Mittels dieses Verfahrens können Präsentationsstrategien gelernt werden.

In einem zweiten Schritt werden die vom System generierten Präsentationen durch drei menschliche Experten bewertet. Diese Bewertungen werden ebenfalls bei der endgültigen Auswahl der zu verwendenden Präsentationsstrategien berücksichtigt und werden mit den aus dem Korpus abgeleiteten Bewertungen kombiniert. Trotz der Neubewertung durch die Experten ergaben sich jedoch numerische Unregelmäßigkeiten. So wurden durch das System etwa 25% der durch die Experten als „schlecht“ bewerteten Präsentationen als „schlecht“ bewertet.

#### **3.3 Multi-modale Ausgaben in verteilten Geräteumgebungen**

Bei multi-modalen Ausgaben in verteilten Geräteumgebungen ergeben sich im Vergleich zu Ausgaben auf einem einzelnen Gerät neue Herausforderungen. Bei verteilten Systemen können einzelne Geräte (z. B. ein mobiler PDA) aus dem System entfernt oder dem System hinzugefügt werden. Hier stellt sich die Frage, wie ein System auf solche Ereignisse reagieren soll. Weiterhin ist es möglich, denselben Inhalt auf verschiedenen Geräten darzustellen. So kann z. B. die multi-mediale Präsentation von Filminformationen sowohl auf einem PDA als auch auf einem Tablet PC erfolgen. Da der PDA jedoch über wesentlich weniger Rechen- und Netzwerkkapazität verfügt sowie eine geringere Auflösung als der Tablet PC besitzt, stellt sich hier die Frage, wie die Präsentation geeignet an die Geräteeigenschaften des Ausgabegeräts angepasst werden kann.

Auch die Interaktionsmöglichkeiten mit dem System bekommen eine neue Dimension. So ist es jetzt möglich, Informationen für mehr als einen Benutzer zu präsentieren. Dabei stellt sich insbesondere die Frage, welche Informationen auf privaten Ausgabegeräten (z. B. auf einem Bildschirm eines PDAs oder auf Kopfhörern) und welche auf öffentlichen Ausgabegeräten (z. B. einem Wandbildschirm) dargestellt werden. Um diese Fragen in Bezug auf die aktuellen Forschungsarbeiten zu untersuchen, werden zunächst die relevanten empirischen Untersuchungen bzgl. Ausgaben in verteilten Geräteumgebungen untersucht. Anschließend wird ein Überblick über existierende Systeme gegeben.

Zuletzt erfolgt ein Diskurs über Ansätze zur Modellierung von Ausgabegeräten und deren Rolle in Bezug auf ubiquitäre Informationssysteme.

### 3.3.1 Empirische Untersuchungen

Bei Systemen, die über mehrere Geräte verteilt sind, ist die Frage nach einer geeigneten Benutzerschnittstelle noch schwieriger zu beantworten als bei traditionellen Benutzerschnittstellen, die nur ein einzelnes Gerät umfassen. Der Grund hierfür ist, dass die Bedienbarkeit von Systemen mit mehreren Ausgabegeräten stark von der Konfiguration der Geräte untereinander abhängt, z. B. der Größe und der räumlichen Anordnung der Bildschirme. Aus diesem Grund lassen sich experimentelle Ergebnisse und Werkzeuge nicht ohne weiteres auf andere Systeme anwenden. Im folgenden Abschnitt wird ein Überblick über die existierenden Studien und Systeme gegeben. Dabei wird unterschieden zwischen den Bereichen Multi-Monitor-Systeme, ubiquitäre Informationssysteme und vernetzte Home-Entertainment-Anwendungen.

#### Multi-Monitor-Systeme

Es gibt verschiedene Hinweise auf die effiziente Verwendung multipler Ausgabegeräte. Eine von Microsoft in Auftrag gegebene Studie [Robertson et al., 2005] mit 1197 Vpn ergab, dass bei 20% aller Windows-Arbeitsplätze mehrere Monitore (bis zu drei) gekoppelt als ein großer Bildschirm eingesetzt wurden. Weitere 70% aller Benutzer verwendeten mehrere Monitore parallel, die aber an verschiedene Computer angeschlossen waren. Als wesentliche Gründe für die Beschränkung auf einen einzelnen Monitor wurden von den Vpn der beschränkte Platz am Arbeitsplatz sowie die Kosten der Bildschirme angegeben. [Grudin, 2001] hat in einer Feldstudie 18 Anwender von Multi-Monitor-Systemen befragt. Hierbei hat Grudin die Hypothese aufgestellt, dass Multi-Monitor-Systeme beim produktiven Arbeiten helfen können, indem sie das Verteilen von Applikationen zulassen sowie eine periphere Wahrnehmung von Ereignissen (z. B. Email) ermöglichen.

In [Oonk et al., 2003] wurden verschiedene Konfigurationen eines sechs Monitor-Systems in Form einer Videowand für eine militärische Kommandozentrale untersucht. Dabei sollten Applikationen wie Email, Chat oder Word auf mehrere Monitore verteilt werden. Parallel sollten die Vpn ein Dokument erstellen und auf eingehende Anfragen per Chat reagieren. Untersucht wurden Anordnungen mit ein, zwei, drei, vier und sechs Monitoren. Die Ergebnisse zeigen, dass die optimale Monitoranordnung von der Applikation abhängt. Die Vpn bevorzugten in den meisten Fällen eine vier Monitor-Anordnung. Bezüglich der Effizienz haben sich vier Monitore in allen Experimenten deskriptiv als überlegen gezeigt jedoch nie signifikant.

### 3.3. MULTI-MODALE AUSGABEN IN VERTEILTEN GERÄTEUMGEBUNGEN

[Czerwinski et al., 2003] hat die Effizienz und Akzeptanz von Benutzern bei Verwendung eines 15 Zoll großen LCD-Monitors und eines 46,5 Zoll großen Rückprojektionsmonitor verglichen. Der große Bildschirm bestand dabei aus drei nebeneinander projizierten Bildschirmsignalen, die einen großen Desktop ergaben. Es wurde dabei sowohl die Erinnerungsleistung untersucht als auch die Zeit, die die Vpn zur Ausführung der gestellten Aufgaben benötigten. Die Erinnerungsleistung wurde dabei zur Untersuchung der kognitiven Last verwendet, die aus den Arbeiten mit dem jeweiligen Bildschirm resultiert.

Bei der Durchführung der Studie wurden die Vpn gebeten, Aufgaben im Kontext von Office- und Internet-Applikationen zu lösen. Das Ergebnis der Studie war, dass sich beim großen Bildschirm die Erinnerungsleistung nicht signifikant von der Erinnerungsleistung bei Verwendung des kleineren Bildschirms unterscheidet. Jedoch waren die Vpn im Mittel auf dem großen Bildschirm 9% schneller und beurteilten den großen Bildschirm signifikant besser.

In einer Studie von [Hutchings et al., 2004] wurden die Interaktionen von 39 Teilnehmern an ihren Arbeitsplatz-PCs untersucht. Hierbei wurden Einzelmonitorkonfigurationen mit Doppelmonitorkonfigurationen sowie Tripelmonitorkonfigurationen verglichen. Hutchings und Kollegen beobachteten, dass Nutzer von Multimonitorssystemen signifikant mehr Fenster parallel sichtbar haben. Weiterhin waren signifikant weniger Interaktionen mit dem Fenster der Email-Software notwendig, da aufgrund des erhöhten Layout-Platzes das Fenster i. Allg. sichtbar war und nicht in den Vordergrund gewechselt werden musste. Benutzer von Tripelmonitorsystemen nutzten jedoch den erhöhten Layout-Platz ebenso aus wie Benutzer von Einzelmonitorsystemen (in 81% der Zeit ist weniger als 20% Layout-Platz verfügbar). Dies ist ein Hinweis darauf, dass Benutzer von Multimonitorssystemen den gewonnenen Layout-Platz auch tatsächlich verwenden.

[Tan & Czerwinski, 2003] haben die Effizienz von 24 Benutzern bei der Verwendung von sechs Kombinationen von zwei 22" Monitoren sowie eines Projektionsbildschirms untersucht. Ziel der Studie war es, die physikalische Trennung von Informationen über mehrere Displays hinweg zu untersuchen. Dabei ergaben sich entgegen den Erwartungen nur ein geringer Performance-Verlust aufgrund der Trennung der Informationen. Allerdings zogen 14 von 24 Teilnehmern die 22 Zoll Monitore vor und 10 Teilnehmer bevorzugten, mit nur einem Display zu arbeiten.

[Kruppa, 2006] hat die Migration eines animierten Charakters zwischen einem PDA und einem Wandbildschirm in einem Erinnerungsexperiment untersucht. Als Lerninhalte wurden dabei Präsentationen über historische Persönlichkeiten verwendet. Dabei fanden Präsentationen auf beiden Ausgabegeräten statt.

Die Kombination PDA und Wandbildschirm wurde mit einer Bedingung verglichen, in der nur der Wandbildschirm verwendet wurde. Die Teilnehmer in der PDA-Wandbildschirm-Gruppe zeigten im Laufe des Experiments eine signifikante Verbesserung ihrer Erinnerungseffizienz. Nach drei Präsentationen zeigten diese Vpn dieselbe Erinnerungseffizienz wie Vpn, die nur mit dem Wandbildschirm gearbeitet haben.

Neben dem Charakter wurden weitere Methoden untersucht, um den Benutzer über einen Wechsel der Präsentation vom Wandbildschirm auf den PDA zu informieren (kein Signal oder ein animiertes Symbol). Es ergaben sich zwischen den Methoden jedoch keine signifikanten Unterschiede bzgl. der Erinnerungseffizienz. Die Mehrheit der Benutzer zog jedoch den animierten Charakter vor.

#### Ubiquitäre Informationssysteme

In [Meyer zu Kniendorf, 2004] wurde ein Experiment bezüglich der Effizienz und Akzeptanz von Benutzerschnittstellen eines ubiquitären Home-Entertainment-Systems durchgeführt. Dabei wurde das zu untersuchende System in einem Wizard-of-Oz-Experiment [Dahlbäck et al., 1993] simuliert. Das System setzte sich aus einem Fernseher, einem Videorecorder und einer Stereoanlage zusammen. Es wurden drei generelle Typen von Benutzerschnittstellen untersucht. In dem Experiment wurde verschiedene zentrale Kontrollmöglichkeiten für die verteilte Geräteumgebung untersucht und dezentralen Kontrollmöglichkeiten gegenübergestellt. Als zentrale Kontrollinstanzen wurde ein Tablet PC mit einer GUI sowie einem animierten Charakter untersucht. Als dezentrale Kontrollmöglichkeit wurde eine Bedienung des Systems per Sprache untersucht.

Das Ergebnis der Studie war, dass Benutzer in der Interaktionen mit einem bisher unbekanntem ubiquitären Informationssystem dazu tendieren, ihre traditionelle Einzelgerätekontrolle beizubehalten und in Folge dessen Systemfunktionalitäten nicht ausnutzen. Dies zeigt, dass insbesondere in Systemen, bei denen Geräte in die Umgebung integriert sind, geeignete Möglichkeiten zur Erforschung des Systems nötig sind. Die Studie hat hierzu gezeigt, dass ein GUI als zentrale Kontrollinstanz gut geeignet ist, da es die gesamte Funktionalität des Systems grafisch und statisch präsentiert, was zu deren Erforschung einlädt. Der animierte Charakter wurde als zentrale Kontrollinstanz gut akzeptiert.

Neben der Frage der Benutzerschnittstelle gibt es weitere Aspekte von ubiquitären Informationssystemen, die untersucht wurden. Dryer hat in zwei Studien gezeigt, dass sichtbare IT Geräte negative Stereotypen hervorrufen können [Dryer et al., 1999]. So können technische Geräte, die nur schwer gemeinsam zu verwenden sind, Benutzer unangenehm und zurückgezogen erscheinen

### 3.3. MULTI-MODALE AUSGABEN IN VERTEILTEN GERÄTEUMGEBUNGEN

lassen. Insbesondere können somit sichtbare technische Geräte sehr negative Einflüsse auf die Mensch-Mensch-Kommunikation haben. Dies ist insbesondere kritisch im Privathaushalt, der i. Allg. von mehreren Personen gemeinsam benutzt wird.

Neben der Benutzerschnittstelle stellt sich auch die Frage nach der Vorhersagbarkeit und des Datenschutzes [Weiser et al., 1999]. Die Frage nach der Vorhersagbarkeit eines solchen Systems lässt sich anhand des folgenden Beispiels illustrieren. Falls der Benutzer einem ubiquitären Informationssystem die sprachliche Anfrage „Ich möchte eine Komödie sehen“ stellt, so könnte das System z. B. reagieren, indem es die Jalousien schließt, das Licht dimmt und das Fernsehen anschaltet. Der Benutzer würde erwarten, dass das System identisch reagiert, falls er einen Thriller sehen möchte. Andererseits würde er erwarten, dass die Jalousien nicht heruntergelassen werden, wenn es draußen bereits dunkel ist. Diesen Erwartungen des Benutzers zu entsprechen ist für ubiquitäre Informationssysteme eine große Herausforderung.

Für ubiquitäre Informationssysteme ist ebenfalls die Frage des Datenschutzes von Relevanz. In Systemen, die den Benutzer beobachten sowie protokollieren, was der Benutzer sagt, stellt sich die Frage, wie öffentlich diese Daten sind. Können andere Benutzer auf diese Daten zugreifen und sind sie vor unerlaubten Zugriffen sicher? Diese Frage stellt sich insbesondere in privaten Umgebungen wie z. B. im Privathaushalt. So haben erste Experimente mit ubiquitärer Informationstechnologie in Form von RFID-Funkchips im Konsumgüterbereich bereits zu massiven Protesten von Kunden geführt, die sich überwacht fühlen [Fischermann, 2004].

Weiterhin haben Hilbert und Trevor beobachtet, dass Benutzer in formellen Situationen wie z. B. einem Mitarbeitervortrag Interaktionen mit Geräten bevorzugen, deren Bildschirme nicht für jeden sichtbar sind (z. B. einem mobilen Gerät) [Hilbert & Trevor, 2004]. In einer informellen Situation wie z. B. einer internen Gruppenbesprechung trat dieses Phänomen nicht auf.

#### Vernetzung im Home-Entertainment

Ein Bereich, in dem die zunehmende Vernetzung von Geräten und die Integration von PC-Funktionalitäten besonders deutlich wird, ist das Home-Entertainment. Fernseher können dazu verwendet werden, um im Internet zu surfen<sup>3</sup>. Digitale Bilderrahmen können per Netzwerk angesteuert werden<sup>4</sup>. DVD-Recorder verbinden sich mit dem Internet, um den elektronischen Programmführer zu aktualisieren<sup>5</sup>. Digitale Filme auf dem PC können drahtlos

---

<sup>3</sup>Loewe Online Plus Erweiterung, <http://www.loewe.de>.

<sup>4</sup>Digi-Frame DF-1710, <http://www.digi-frame.com>.

<sup>5</sup>KiSS Technology DVD Recorder DP-558, <http://www.kiss-technology.com>.

auf den Fernseher übertragen werden<sup>6</sup>. Selbst Kühlschränke können sich mit dem Internet verbinden, damit der Benutzer den Bestand an Lebensmitteln direkt per Online-Shopping aufstocken kann<sup>7</sup>. Ein weiterer Trend ist die Integration zusätzlicher Bildschirme in Geräte, um der zunehmenden Funktionsvielfalt der Geräte Herr zu werden. So zeichnen sich z. B. die programmierbaren Harmony-Fernbedienungen<sup>8</sup> durch einen zusätzlichen Farb-Display aus, der ähnlich einem Mobiltelefon Zusatzinformationen anzeigen kann.

Insbesondere im Home-Entertainment-Bereich wurde die Entwicklung durch die Ergebnisse mehrerer medienwissenschaftlicher Studien motiviert. Diese Studien haben gezeigt, dass in den USA mehr als die Hälfte der Internet-Benutzer einen Fernseher und einen PC im selben Raum haben und beide Medien auch parallel nutzen [Pilotta et al., 2004; ComScore Presseerklärung, 2002]. Eine Online-Studie mit 700 deutschen Teilnehmern zeigt weiterhin, dass der Bildschirm des Fernsehers als universelles Display akzeptiert wird [Becker et al., 2004]. So haben rund 80% der Befragten auf die Frage, ob sie sich vorstellen können, vom Fernseher aus auf den PC frei zugreifen zu können, zustimmend reagiert. Rund 72% konnten sich vorstellen, Urlaubsvideos oder MP3s auf dem Fernsehbildschirm zu zeigen.

Diese Studien zeigen, dass sich PC-Funktionalität und Heimelektronik nicht ausschließen und dass der PC Einzug ins Wohnzimmer gehalten hat. Jedoch existieren auch fundamentale Unterschiede zwischen der Verwendung eines Fernsehers und der Verwendung eines Computers. So stellt Jakob Nielsen im Zuge einer Analyse eines Fernsehgerätes mit Internet-Browser fest, dass Webseiten auf Fernsehern aufgrund der niedrigeren Auflösung oft nur unzureichend dargestellt werden<sup>9</sup>. Weiterhin wird beim Fernsehen eher passiv konsumiert, während der Benutzer beim Computer eher aktiv die darzustellenden Inhalte bestimmt. Außerdem werden Computer meist von einem Benutzer alleine verwendet, während beim Fernsehen sich oftmals mehrere Benutzer ein Gerät teilen. Nielsen schließt daraus, dass die Nutzung des Internets auf dem Fernseher eher in die Sendungen integriert werden sollte und nicht zu viele Interaktionen der Benutzer voraussetzen sollte.

### 3.3.2 Technische Umsetzung

Es existieren verschiedene Ansätze zur Generierung von Ausgaben auf Systemen mit gekoppelten Ausgabegeräten. Im Folgenden werden die wichtigsten Arbeiten vorgestellt und verglichen.

---

<sup>6</sup>Philips Streamium SL400i, <http://www.streamium.com>.

<sup>7</sup>Samsung HomePAD Refrigerator RH2777AT, <http://www.samsung.com>.

<sup>8</sup>Logitech Harmony Remote 655, <http://www.logitech.com>.

<sup>9</sup>Jakob Nielsen's Alertbox for February 15, 1997, <http://www.useit.com/alertbox/9702b.html>.

### 3.3. MULTI-MODALE AUSGABEN IN VERTEILTEN GERÄTEUMGEBUNGEN

Das Peach-System [Kruppa, 2004] erstellt gekoppelte Bildschirmausgaben mittels eines PDAs und eines Wandbildschirms, den sich mehrere Benutzer gleichzeitig teilen. Animierte Charaktere werden eingesetzt, um zwischen den beiden Ausgabegeräten zu vermitteln. Aufgrund der statischen Museumsumgebung werden jedoch neben PDAs und Wandbildschirmen keine weiteren Ausgabegeräte unterstützt.

Rekimoto [Rekimoto, 2000] beschreibt den Einsatz von Benutzerschnittstellen, die sich über mehrere physikalische Interaktionsgeräte hinweg erstrecken, z. B. für Interaktionen zwischen PDAs und einem Wandbildschirm. Die Interaktionstechniken beschränken sich hier jedoch lediglich darauf, Daten von einem Gerät zum anderen zu bewegen. Systemausgaben, die mehrere Geräte gleichzeitig umfassen, werden nicht generiert.

[Heider & Kirste, 2005] beschreiben einen Ansatz für spontane Gerätekooperation in ubiquitären Informationssystemen. Der Ansatz verwendet das Konzept von zielbasierten Interaktionen. Bei zielbasierten Interaktionen wird die Systemfunktionalität nicht funktional beschrieben (z. B. das Einlegen, Abspielen und Auswerfen einer CD) sondern mittels übergeordneter Ziele (z. B. „Musik abspielen“), die die Interaktion des Benutzers mit dem System erfüllen soll. Diese Ziele werden aus den Benutzeranfragen abgeleitet und an einen Planer übergeben, der sie mittels geeigneter Operatoren in gerätespezifische, funktionale Anfragen übersetzt. Als Beispiel wird ein Szenario beschrieben, in dem der Benutzer, die Helligkeit des Fernsehbildes erhöhen möchte. Das System setzt dies durch das Dimmen einer Lampe und das Schließen von Jalousien um. Die dazu notwendigen Strategien im Planer werden über einer geräteübergreifenden Ontologie definiert, die Konzepte wie die Umgebungshelligkeit umfasst.

Heider und Kirste argumentieren, dass die Ontologie nur aus geräteunabhängigen Konzepten wie Helligkeit oder Sichtbarkeit besteht und dass die Definition der Präsentationsoperatoren über den Konzepten der Ontologie keine Einschränkung der Allgemeingültigkeit der Planungsoperatoren darstellt. Die Planungsoperatoren werden von jedem Gerät, das dem System neu hinzugefügt wird, an den Planer gesendet. Auf diese Weise können theoretisch beliebig viele Geräte durch das System unterstützt werden.

Diese Vorgehensweise ist jedoch nur bei einfachen Szenarien wie dem oben beschriebenen praktikabel. Bei der Koordination von gekoppelten Ausgabegeräten müssen die Ausgaben zwischen den Geräten auf komplexe Weise koordiniert werden. So müsste eine Sprachausgabekomponente Operatoren liefern, die angeben, wie sie mit anderen Komponenten (z. B. den Lippenbewegungen eines animierten Charakters) koordiniert werden kann. Diese Komponenten

und deren Funktionalitäten müssten durch die Sprachausgabe vorausgesehen werden, damit sie in den Operatoren, die an den Planer gesendet werden, berücksichtigt werden können. Da dies i. Allg. ein zu komplexes Problem darstellt, ist der von Heider und Kirste gewählte Ansatz für gekoppelte Ausgabegeräte nicht praktikabel.

Das System aus [Braun et al., 2004] ist in der Lage, ein GUI mit Sprachausgabe zu generieren, das über mehrere eingebettete Displays und PDAs verteilt ist. Dabei wird die Benutzerschnittstelle zunächst geräteunabhängig in einer auf XHTML und XForms basierenden Sprache definiert. Anschließend wird sie von einem Java-Programm in mehrere GUIs aufgespalten. Dies geschieht in Abhängigkeit der Position und Orientierung des Benutzers, dessen Position mit Hilfe von aktiver Sensorik überwacht wird. Die Aufspaltung des geräteunabhängigen Interfaces geschieht allerdings mit Hilfe einer festen Menge von Java-Regeln [Braun & Mühlhäuser, 2004]. Dadurch ist das Regelwerk nur schwer erweiterbar und die betroffene Komponente nicht transparent. Außerdem werden in der geräteunabhängigen Beschreibung Werte bzgl. der Wichtigkeit von Interface-Teilen zwischen 0.0 und 1.0 vergeben, die anschließend bevorzugt auf Geräten mit guten Interaktionsmöglichkeiten dargestellt werden. Der Nachteil dieses Ansatzes ist, dass die zugewiesenen Werte nur wenig Aussagekraft besitzen und dass ihre konkrete Auswirkung auf die Aufspaltung des Interfaces nur schwer reproduzierbar ist.

Der Pebbles Personal Remote Controller [Myers et al., 2004] verwendet Beschreibungen von Applikationen, um automatisch ein PDA-basiertes GUI sowie eine Spracheingabeschnittstelle zu erstellen. Die Applikationen (z. B. die Aufnahme eines Films mittels eines Videorecorders) müssen jedoch manuell durch den Benutzer koordiniert werden. Diese manuelle Koordination ist jedoch nicht praktikabel für multimediale Ausgabeapplikationen, die in ein komplexes räumlich-zeitliches Layout integriert werden müssen.

Innerhalb des Everywhere Displays-Projektes von IBM [Pinhanez & Podlaseck, 2005] werden projizierte Bilder in Kombination mit Gestenerkennung dazu verwendet, um Benutzerschnittstellen ubiquitär erscheinen zu lassen und um sie an den Kontext der Benutzer anzupassen. Ein Szenario, in dem diese projizierten Benutzerschnittstellen eingesetzt werden ist ein Supermarkt, in dem Kunden durch die Projektionen auf Produkte aufmerksam gemacht werden sollen und Zusatzinformationen abrufen können. Hierbei werden Informationen zu Produkten auch auf mehreren Projektionen parallel um das Produkt gruppiert dargestellt. In einer Benutzerstudie [Sukaviriya et al., 2003] ergaben sich jedoch Probleme der Versuchspersonen mit der Handhabung der projizierten Benutzerschnittstellen. So waren Versuchspersonen u. a. verwirrt

### 3.3. MULTI-MODALE AUSGABEN IN VERTEILTEN GERÄTEUMGEBUNGEN

durch die Migration eines Displays von einem Ort zu einem anderen.

Das u-Texture-Projekt [Yanagihara et al., 2005] verwendet multiple Bildschirme im Bereich von intelligenten Möbeln. Eine u-Texture besteht aus einem Bildschirm mit einem eingebettetem PC, der vom Benutzer vertikal oder horizontal ausgerichtet wird. Ebenso können mehrere u-Textures durch den Benutzer verbunden werden, wodurch sich z. B. eine Videowand ergibt. Die Neigung der u-Textures wird mittels Beschleunigungssensoren gemessen und Daten werden über WLAN oder Infrarot ausgetauscht. Die resultierenden Anwendungen beschränken sich jedoch auf einfachen Datenaustausch zwischen den u-Textures, z. B. durch Kopieren der Daten von einer u-Texture auf eine andere.

Das Gaia-Projekt implementiert eine Middleware für ubiquitäre Informationssysteme, die beliebige Anzahlen von Displays als Ausgabegeräte erkennt und ansteuern kann. Der Presentation Manager ist eine auf Gaia aufsetzende Applikation, die dazu dient Power Point-Präsentationen synchron auf mehreren Bildschirmen gleichzeitig darzustellen. Hierbei muss allerdings der Benutzer vor der Präsentation festlegen, welcher Inhalt auf welchem Bildschirm dargestellt werden soll [Hess et al., 2002].

Im Projekt EasyLiving von Microsoft Research [Brumitt et al., 2000] wurde ein ubiquitäres Informationssystem im Home-Entertainment-Bereich realisiert. Innerhalb des Projektes wurde ein Raum mit Kameras ausgestattet, die bis zu drei Benutzer gleichzeitig beobachten können. Neben Personen beobachtet das System die Positionen von Eingabegeräten wie z. B. einer drahtlosen Tastatur. Die Ausgaben des Systems sind an den Kontext des Benutzers angepasst. Wechselt der Benutzer seine Position vom PC-Arbeitsplatz zur Couch, so wird der Desktop des Computers automatisch auf die gegenüberliegende Wand projiziert. Somit passt das EasyLiving System Ausgaben an die Benutzerposition innerhalb des Raumes an, was durch ein Geometriemodell des Raumes ermöglicht wird. Es wird jedoch lediglich der gesamte Windows-Desktop migriert. Es ist insbesondere nicht möglich in Abhängigkeit der vorhandenen Ausgabegeräte andere Kombinationen von Ausgabemodalitäten zu verwenden.

Das WWICE-System von Philips [Baldus et al., 2000] realisiert eine Middleware für verteilte Multimedia-Anwendungen. Mit Hilfe dieses Systems ist es Benutzern im Philips HomeLab möglich, von jedem Zimmer des Labs aus denselben Videofilm anzusehen. Wechselt der Benutzer den Raum, so wird der Film auf das nächste TV in seiner Nähe umgeschaltet. Durch das WWICE werden die im Labor verteilten Multimedia-Geräte und -Applikationen innerhalb eines Netzwerkes angesteuert und verknüpft. Dabei unterstützt das WWI-

CE dynamische Veränderungen des Systems (z. B. das Entfernen eines Gerätes). Allerdings werden durch das WWICE keine Interaktionen unterstützt, die mehrere Displays in einem Raum umfassen.

Es existieren verschiedene modellbasierte Ansätze, die es erlauben aufgrund der Modellbeschreibungen Benutzerschnittstellen von einem Gerät zu einem anderen Gerät zu migrieren und automatisch anzupassen [Ding & Litz, 2006; Massó et al., 2006; Puerta & Eisenstein, 2002]. Jedoch werden dabei aufgrund der Komplexität der Modelle nur grafische Benutzerschnittstellen migriert. Es existieren bisher keine modellbasierten Systeme, die komplexe multi-modale Benutzerschnittstellen automatisch generieren und migrieren können.

Der in [Bandelloni & Paternò, 2004] beschriebene modellbasierte Ansatz erlaubt es, die Navigationsobjekte und die Visualisierungsobjekte von Webseiten zu trennen und auf mehrere Geräte zu verteilen. So kann z. B. ein PDA verwendet werden, um eine Webseite auf einem PC zu kontrollieren. Die Autoren sprechen hier von „partieller Migration“. Der Ansatz ist jedoch auf HTML-Webseiten und VoiceXML-Interaktionen beschränkt und unterstützt keine komplexen Synchronisationen zwischen Medienobjekten, wie es z. B. der SMIL-Standard ermöglicht.

### **3.4 Modellierung multi-modaler Ausgaben**

In ubiquitären Informationssystemen, die sich aus einer Vielzahl von unterschiedlichen Geräten zusammensetzen, können Geräte dynamisch aus dem System hinzugefügt und entfernt werden. Dies kann geschehen, indem ein Gerät (z. B. ein Computermonitor) manuell durch den Benutzer ausgeschaltet wird oder durch einen technischen Fehler ausfällt. Ubiquitäre Informationssysteme, die sich typischerweise aus einer Vielzahl von einzelnen Geräten zusammensetzen, müssen in der Lage sein, auf solche Ergebnisse zu reagieren [Edwards & Grinter, 2001].

Um eine ausgefallene oder entfernte Komponente ersetzen zu können, benötigt das System Informationen darüber, welche Dienste die Komponente ausgeführt hat. Mit dieser Information kann das System versuchen, mit den verbleibenden Ressourcen den verloren gegangenen Dienst zu ersetzen. So könnte z. B. das Ausschalten eines Computermonitors zur Folge haben, dass bei der nächsten Ausgabe stattdessen ein Projektor verwendet wird. Fällt eine Sprachausgabekomponente aus, so könnte sie z. B. durch eine grafische Textausgabe ersetzt werden.

In dieser Arbeit wird insbesondere untersucht, wie ein ubiquitäres Informationssystem flexibel auf das Hinzufügen und Entfernen von Ausgabegeräten reagieren kann. Aus diesem Grund werden in diesem Abschnitt Ansätze zur Modellierung der multi-modalen Ausgaben dieser Geräte untersucht. Die all-

### 3.4. MODELLIERUNG MULTI-MODALER AUSGABEN

gemeine Modellierung von multi-modalen Ausgaben wird hierbei als Modalitätentheorie bezeichnet [Bernsen, 2001]. Neben Modalitätentheorien wird ebenfalls auf die Modellierung von animierten Charakteren als spezielle Form von multi-modalen Ausgaben eingegangen.

#### 3.4.1 Modalitätentheorien

Die von Bernsen definierte Taxonomie von Ausgabeunimodalitäten [Bernsen, 2001] umfasst 30 Basisklassen von Unimodalitäten. Mit Ausgabeunimodalität bezeichnet Bernsen eine atomare Ausgabemodalität. Bernsen zeigt, dass diese Darstellung vollständig und eindeutig ist. Die 30 Basisklassen werden auf der Grundlage von fünf Basiseigenschaften von Ausgabeunimodalitäten definiert. Bernsen unterscheidet zwischen (nicht) linguistischen, (nicht) analogen, (nicht) arbiträren sowie statischen und dynamischen Ausgabeunimodalitäten. Weiterhin unterscheidet Bernsen drei Typen von Medien: visuell, auditoriell und taktil. Einige der Basisklassen werden exemplarisch in eine „subatomare“ Ebene erweitert. Diese Erweiterung findet aber nur exemplarisch statt, da sie laut Bernsen schnittstellenspezifisch durchgeführt werden muss.

Bernsens Modell wird in verschiedenen multi-modalen Benutzerschnittstellen eingesetzt. [Vernier & Nigay, 2000] verwendet die fünf Basiseigenschaften der Ausgabeunimodalitäten zur automatisierten Generierung von Tabellen. Innerhalb des Amebica-Projekts [Vaudry & Crampes, 2001], in dem Visualisierungen zur industriellen Ablaufkontrolle erstellt werden, werden Bernsens Ausgabeunimodalitäten innerhalb der Multimediaantologie des Systems eingesetzt.

In [Obrenovic & Starcevic, 2004] werden multi-modale Ein- und Ausgaben mittels UML modelliert. Das Ziel von Obrenovic und Starcevic ist es, eine Grundlage für die Entwicklung von Tools für multi-modale Dialogsysteme zu bieten. Grundsätzlich unterscheidet das Modell zwischen dynamischen und statischen Ausgabemodalitäten. Es werden die sensorischen Effekte auf den Benutzer beschrieben (visuelle, auditorielle oder taktile Stimuli). Weiterhin wird die Verbindung einer Ausgabemodalität zu einem konkreten Ausgabegerät modelliert. Weitere Unterscheidungen wie z. B. die Unterscheidung zwischen linguistischen und nicht-linguistischen Ausgabemodalitäten werden nicht modelliert. Im Unterschied zu Bernsen ist das Modell nicht als vollständige und eindeutige Beschreibung von Ausgabemodalitäten gedacht, sondern als Unterstützung für Designer von Benutzerschnittstellen. Jedoch verliert das Modell durch die ungenaue Modellierung an Ausdruckskraft.

Das in [Arens et al., 1993] definierte Modalitätenmodell unterscheidet Modalitäten anhand von sieben Kriterien. Die relevantesten Kriterien sind die Dimension des Trägers („carrier dimension“), die interne semantische Dimensi-

on („internal semantic dimension“), sowie der Typ des Mediums (visuell oder auditoriell). Der Typ des Mediums entspricht hier der Definition von Bernsen. Der Nachteil von Arens und Hovys Modell ist jedoch, dass Arens und Hovys Modell Eigenschaften verwendet, mit denen sich Ausgabemodalitäten nur schwer kategorisieren lassen. So unterscheiden Arens und Hovy zwischen der Fähigkeit der Ausgabemodalität, die Wahrnehmung des Benutzers zu erreichen („default detectability“) unter Verwendung der Abstufungen sehr niedrig, niedrig, hoch und sehr hoch. Arens und Hovy modellieren ebenfalls die Zusatzlast einer Ausgabemodalität („baggage“) über dem Wertebereich niedrig und hoch.

Mit diesen Abstufungen ist jedoch eine eindeutige Charakterisierung von Ausgabemodalitäten schwierig. So charakterisieren Arens und Hovy die Wahrnehmungsfähigkeit von gesprochener Sprache als mittel bis hoch. Gesprochene Sprache ließe sich jedoch ebenfalls als sehr hoch modellieren. Bei der Zusatzlast wird Musik von Arens und Hovy als niedrig eingestuft. Genauso ließe sich aber auch argumentieren, dass Musik neben dem Kernthema sehr viele zusätzliche Informationen übermittelt, z. B. durch begleitende Instrumente im Hintergrund.

Arens und Hovy trennen die Beschreibung von Ausgabemodalitäten ebenfalls nicht von ihrer Wirkung auf den Benutzer, was ein weiterer Unterschied zu Bernsens Modell ist. Arens und Hovy schränken ihr Modell auch nicht auf die für die Mensch-Maschine-Kommunikation relevanten Kombinationen der sieben Modelleigenschaften ein.

#### 3.4.2 Animierte Charaktere

Animierte Charaktere stellen eine spezielle Kategorie multi-modaler Ausgaben dar. Im Unterschied zu den in Abschnitt 3.4.1 vorgestellten Ansätzen, die eine allgemeine Modellierung multi-modaler Ausgaben zum Ziel haben, existieren für animierte Charaktere bereits zahlreiche detaillierte Modelle. Die Ansätze unterschieden sich jedoch durch die Ebene, auf der sie das Verhalten von animierten Charakteren charakterisieren. Im Folgenden wird ein Überblick über die relevanten Ansätze gegeben.

Bei der Modellierung von animierten Charakteren können zwei Ebenen unterschieden werden [Kopp et al., 2006]. Die erste Ebene beinhaltet Modelle, die von den Verhaltensweisen des Charakters (z. B. Gestik, Mimik oder Körperhaltung) unabhängig sind. So kann z. B. die Begrüßung des Benutzers sowohl durch einen Charakter ohne Gestik erfolgen als auch durch einen Charakter mit Gestik. Diese Ebene wird als abstrakte Modellierungsebene bezeichnet.

Die zweite Ebene beinhaltet Modelle, die die konkreten Verhaltensweisen eines Charakters beschreiben. Modelle, die dieser Ebene zugeordnet werden, be-

### 3.4. MODELLIERUNG MULTI-MODALER AUSGABEN

schreiben detailliert, wie die Verhaltensweisen des Charakters umgesetzt und parametrisiert werden können. Bei einem Charakter, der neben einer Sprachausgabe auch über Gestik verfügt, könnte z. B. die Synchronisation zwischen Sprache und Gestik modelliert werden. Diese Ebene wird als physikalische Modellierungsebene bezeichnet.

In [Kopp & Wachsmuth, 2004] wird die Sprache MURML (Multi-modal Utterance Representation Markup Language) beschrieben. Die Sprache hat das Ziel, Sprache und Gesten sowie deren Synchronisation auf XML-Basis zu beschreiben. Der Schwerpunkt der Arbeit ist die Beschreibung und die darauf aufbauende Generierung von natürlich wirkenden Gestenanimationen. Das Modell kann somit in die physikalische Modellierungsebene eingeordnet werden. Neben Gesten und Sprache werden jedoch keine weiteren Modalitäten durch MURML beschrieben.

Die Sprache MPML 3.0 (Multi-modal Presentation Markup Language) [Prenninger et al., 2004] ist eine SMIL-ähnliche Sprache zur Modellierung von Präsentationen mit mehreren Charakteren. MPML konzentriert sich dabei auf die Microsoft Agent-Plattform<sup>10</sup> und deren Animationen. Um die Benutzereingaben sowie Interaktionen mit externen Medienobjekten zu ermöglichen, führt MPML weiterhin Java-Systemaufrufe ein. Dieser Ansatz birgt jedoch das Risiko, die Transparenz des XML-basierten Ansatzes zu verlieren. Dieser Ansatz ist ebenfalls der konkreten Modellierungsebene zuzuordnen.

In [DeCarolis et al., 2004] wird die Sprache APML (Affective Presentation Markup Language) beschrieben. Die Sprache beschreibt die Verhaltensweisen eines animierten Charakters unabhängig von deren Umsetzung und kann somit der abstrakten Modellierungsebene zugeordnet werden. APML wird verwendet, um einen Charakter zu steuern, der nur aus einem Gesicht besteht. Für diesen Charakter wird ebenfalls eine Sprache zur Beschreibung der Gesichtsausdrücke des Charakters definiert, die der physikalischen Modellierungsebene zugeordnet werden kann.

Die BEAT-Software (Behavior Expression Animation Toolkit) [Cassell et al., 2001] berechnet durch Analyse einer Texteingabe automatisch die nichtverbalen Verhaltensweisen eines animierten Charakters. Dabei kommt zur Beschreibung der Animationen des Charakters ebenfalls eine XML-basierte Sprache zum Einsatz, die die Gestik und die Mimik des Charakters sowie die Pausen und Intonation in der Sprache des Charakters beschreibt. Diese Sprache lässt sich der physikalischen Modellierungsebene zuordnen.

Von Kopp und Kollegen [Kopp et al., 2006] wurde ein Vorschlag zur Vereinheitlichung der unterschiedlichen Ansätze auf beiden Modellierungsebenen

---

<sup>10</sup>Microsoft Agent Homepage, <http://www.microsoft.com/msagent>.

erbracht. Das resultierende SAIBA-Modell für animierte Charaktere unterscheidet dabei zwischen drei Schichten der Charaktergenerierung. Während der Intensionsplanung wird die kommunikative Absicht der Ausgabe, die unabhängig von dem konkreten Erscheinungsbild des Charakters ist, berechnet. Das Ergebnis wird in SAIBA durch die (in der Veröffentlichung noch undefinierten) Sprache FML (Functional Markup Language) formuliert, die das Verhalten eines Charakters auf der abstrakten Modellierungsebene beschreibt.

Durch die Verhaltensplanung wird dieses Ergebnis in konkrete Verhaltensweisen umgesetzt, die den Ausdrucksmöglichkeiten des vorhandenen animierten Charakters entsprechen. So könnte z. B. ein Charakter, der über Gestik verfügt, bei der Begrüßung des Benutzers mit der Hand winken. Das Ergebnis wird in der Sprache BML (Behaviour Markup Language) formuliert, die in der Arbeit von Kopp und Kollegen definiert wird.

Das BML-Skript wird durch die Verhaltensplanung anschließend in konkrete Animationen umgesetzt. Eine komplette Definition von BML als Standard ist jedoch noch nicht erfolgt. Durch BML wird die physikalische Modellierungsebene beschrieben.

BML definiert verschiedene Kategorien von Verhaltensweisen von animierten Charakteren. Es wird zwischen den Bewegungen des Kopfes, des Torsos, des Gesichts und der Beine sowie den Körperbewegungen unterschieden. Weiterhin werden die Blickrichtung des Charakters, die Gesten, die Sprache und die Lippenbewegungen modelliert. Jede dieser Animationen kann in verschiedenen Varianten ausgeführt werden und durch Synchronisationspunkte mit anderen Animationen koordiniert werden. BML stellt den allgemeinsten Ansatz der Charaktermodellierung auf der physikalischen Modellierungsebene dar.

### **3.5 Zusammenfassung**

Insgesamt lässt sich feststellen, dass bisher kein System existiert, das eine automatisierte Generierung von multi-modalen Präsentationen mit gekoppelten Ausgabegeräten unterstützt. Andererseits geben bisherige Untersuchungen Hinweise darauf, dass mit solchen Systemen ein effizienteres Arbeiten als mit einzelnen Ausgabegeräten möglich ist. Bisherige Untersuchungen zur Effizienz und Akzeptanz bei gekoppelten Ausgabegeräten haben jedoch die Effekte unterschiedlicher Modalitätenkombinationen nicht untersucht.

Weiterhin existieren Hinweise darauf, dass ein GUI als zentrale Kontrollinstanz zum Erforschen der Systemfunktionalität anleiten kann. Es gibt ebenfalls Hinweise darauf, dass ein animierter Charakter als Verkörperung des Systems einen positiven Effekt auf die Akzeptanz des Systems hat. Im Bereich der Modalitätentheorie wurde von Bernsen die bisher umfassendste Beschreibung von Unimodalitäten definiert.

### 3.5. ZUSAMMENFASSUNG

Tabelle 3.1: Übersicht über Systeme mit gekoppelten Ausgabegeräten.

| System     | dyn.<br>Layout | dyn.<br>Ensemb. | Dienst-<br>modell | multi-<br>modale<br>Ausgabe | ress.-<br>adaptiv | Multi-<br>Geräte-<br>Ausgabe |
|------------|----------------|-----------------|-------------------|-----------------------------|-------------------|------------------------------|
| Bandelloni | ×              | ×               | —                 | —                           | ×                 | ×                            |
| Braun      | ×              | ×               | —                 | (×)                         | ×                 | ×                            |
| EasyLiving | —              | ×               | —                 | —                           | —                 | —                            |
| Heider     | ×              | ×               | (×)               | ×                           | ×                 | —                            |
| Gaia       | ×              | ×               | —                 | ×                           | —                 | ×                            |
| IBM ED     | ×              | ×               | —                 | —                           | —                 | ×                            |
| Peach      | ×              | ×               | —                 | ×                           | —                 | ×                            |
| Pebbles    | ×              | —               | —                 | (×)                         | ×                 | —                            |
| Rekimoto   | —              | ×               | —                 | —                           | —                 | ×                            |
| u-Texture  | —              | ×               | —                 | —                           | —                 | ×                            |
| WWICE      | —              | ×               | —                 | —                           | —                 | —                            |

Die in Abschnitt 3.3.2 gegenüber gestellten Systeme mit gekoppelten Ausgabegeräten werden in Tabelle 3.1 mit Hilfe einer Taxonomie kategorisiert. Zunächst unterscheiden sich die Systeme in der Berechnung des Layouts, die bei Multi-Geräteausgaben dynamisch oder statisch erfolgen kann. Weiterhin unterscheiden sich die Systeme darin, ob sie das dynamische Hinzufügen oder Entfernen von Ausgabegeräten aus einem Geräteensemble unterstützen oder eine statische Geräteumgebung voraussetzen.

Darüber hinaus lassen sich die Systeme anhand ihrer Unterstützung von Ausgabediensten unterscheiden. Die Modellierung von Ausgabediensten geht über die Modellierung von Ausgaberesourcen hinaus und beschreibt die Ausgabe-Software, die auf den Geräten vorhanden ist. Dies kann z. B. eine Komponente zur Generierung von Sprachausgaben sein, die WAV-Dateien mit Sprachdaten aus einem ASCII-Text generiert.

Weiterhin können Systeme, die nur rein visuelle Ausgaben generieren, von Systemen unterschieden werden, die multi-modale Ausgaben generieren. Die Systeme unterscheiden sich ebenso darin, ob sie ressourcenadaptiv sind oder nicht, d. h. ob sie zu erstellenden Präsentationen an Gerätesourcen wie z. B. die Bildschirmauflösung anpassen. Schließlich unterscheiden sich die Präsentationssysteme noch darin, ob Multi-Geräteausgaben oder nur Ausgaben auf einem einzelnen Gerät unterstützt werden.

Für die relevantesten vorgestellten Systeme ergibt sich die in Tabelle 3.1 dargestellte Aufteilung. Somit existiert zurzeit kein System, das alle Kriterien erfüllt. Das Ziel dieser Arbeit ist es, ein Verfahren zu definieren, das allen

### *KAPITEL 3. STAND DER WISSENSCHAFT*

Kriterien genügt. Das System soll in einer Prototypimplementierung verifiziert werden und die Effizienz des Systems in einer Benutzerstudie untersucht werden. Insbesondere soll auch der Effekt der Verwendung unterschiedlicher Ausgabemodalitäten bei gekoppelten Ausgabegeräten untersucht werden.

Für die Definition des geplanten Verfahrens ist eine Modellierung von Ausgabegeräten und Ausgabediensten erforderlich. Dabei können die in Abschnitt 3.4.1 vorgestellten Modalitätentheorien zur abstrakten Beschreibung der Ausgabedienste verwendet werden. Diese können mit den in Abschnitt 3.4.2 vorgestellten Ansätzen zur Modellierung von animierten Charakteren kombiniert werden, um ihre Ausdruckskraft zu erhöhen.

## 4 Grundlagen

In diesem Kapitel werden die Grundbegriffe der drei wesentlichen Teilgebiete eingeführt, auf denen diese Arbeit aufbaut. Multi-modale Dialogsysteme bilden in dieser Arbeit die Grundlage der Mensch-Technik-Interaktion. Diese werden in Abschnitt 4.1 betrachtet. Der Schwerpunkt dieser Arbeit liegt bei der Ausgabe von multi-modalen Dialogsystemen. Hierbei wird insbesondere ein Planungsansatz verwendet, der in Abschnitt 4.2 näher beschrieben wird. In Abschnitt 4.3 wird die Modalitätentheorie nach Bernsen eingeführt, die die Grundlage für die Selbstbeschreibung von Ausgabegeräten bildet.

### 4.1 Multi-modale Dialogsysteme

In diesem Abschnitt wird ein Überblick über multi-modale Dialogsysteme gegeben, die die Grundlage der in dieser Arbeit behandelten gekoppelten Ausgabegeräte darstellen. Dabei wird zunächst die Architektur von multi-modalen Dialogsystemen behandelt. Anschließend wird die Ausgabegenerierung in Bezug auf ihre Teilaufgaben näher betrachtet.

#### 4.1.1 Architektur

Laut [Wahlster, 1998] lassen sich Dialogsysteme in Medienanalyse, Interaktionsverarbeitung, Applikation, und Mediendesign aufteilen. In multi-modalen Dialogsystemen erfolgt im Rahmen der Medienanalyse eine Fusion der Eingabemodalitäten. So muss z. B. die Äußerung „Einschalten, bitte.“ zusammen mit der Zeigegeste des Benutzers auf einen Ventilator analysiert werden, um die Benutzerabsicht, den Ventilator einzuschalten, vollständig verstehen zu können.

Umgekehrt ist auf der Ausgabeseite im Rahmen des Mediendesigns eine Bündelung der vorhandenen Ausgabekomponenten notwendig, um die Ausgaben des Systems umzusetzen. So ist im Unterschied zur Fusion der Eingabemodalitäten eine Fission der Ausgabemodalitäten notwendig. Wahlster spricht in diesem Zusammenhang von einer Symmetrie der multi-modalen Dialogführung [Wahlster, 2003].

Abbildung 4.1 zeigt die Umsetzung der Architektur im Embassi-Dialogsystem [Herfet et al., 2001]. Die Medienanalyse umfasst hier die *Eingabekomponenten*  $I_n$  (z. B. einen Spracherkenner) sowie die *Analysekomponenten*  $A_n$ , die die Ergebnisse der Eingabekomponenten filtern und weiterverarbeiten. Weiterhin umfasst die Medienanalyse die *multi-modale Fusionskomponente* MMF, die die Eingabemodalitäten semantisch zusammenführt. Das Ergebnis repräsentiert die Interpretation des Benutzerwunsches durch das System und wird an

## KAPITEL 4. GRUNDLAGEN

den Dialog-Manager weitergegeben. Auf der Dialogebene findet anschließend die Interaktionsverarbeitung statt. Dabei interagiert der *Dialog-Manager*  $D$  mit *Applikationskomponenten*  $P_n$  (z. B. einem Festplattenrecorder), um das von der multi-modalen Fusion übergebene Benutzerziel umzusetzen.

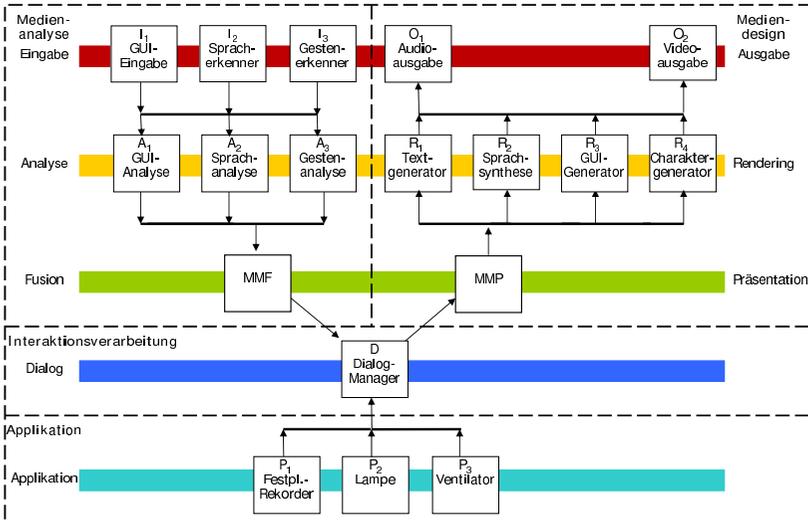


Abbildung 4.1: Beispiel multi-modales Dialogsystem (Vorlage: [Elting et al., 2003]).

Das Ergebnis der Interaktion ist eine Systemausgabe (z. B. „Der Film wurde erfolgreich aufgenommen!“), die durch den Dialog-Manager initialisiert und an die *multi-modale Präsentationskomponente* MMP weitergegeben wird. Die MMP-Komponente interagiert zur Generierung der multi-modalen Präsentationen mit den vorhandenen Ausgabekomponenten  $R_n$ , die als *Rendering-Komponenten* bezeichnet werden.

Ein Beispiel für eine Rendering-Komponente ist eine Textgenerierung, die abstrakte, amodale Konzepte in natürliche Sprache umsetzt. Die Ausgabe erfolgt schließlich durch die *physikalischen Ausgabekomponenten*  $O_n$ , die die Audio- und Videostreams physikalisch auf den Ausgabegeräten ausgeben. Das Mediendesign nach Wahlster umfasst in der Embassi-Architektur somit die Ebenen Präsentation, Rendering und Ausgabe.

Neben den in Abbildung 4.1 gezeigten Komponenten können Sensoren einge-

## 4.1. MULTI-MODALE DIALOGSYSTEME

setzt werden, um weitere Kontextdaten (z. B. die Identität des Benutzers oder den Ort des Benutzers) zu sammeln. Diese Daten werden typischerweise in eine zentrale Datenbank geschrieben, auf die alle Komponenten zugreifen können. Weiterhin ist es sinnvoll, zwischen multi-modaler Fusion, Präsentation und dem Dialog-Manager eine gemeinsam verwaltete Datenbank zur Speicherung der Dialoghistorie zu verwenden. Diese enthält die Historien der multi-modalen Eingaben, der Dialogschritte sowie der multi-modalen Ausgaben des Systems.

Die Datenbank wird von der multi-modalen Fusion verwendet, um Referenzen zu vorhergehenden Dialogen aufzulösen [Elting et al., 2003]. Ein Beispiel hierfür ist die Äußerung „Bitte den Film noch mal zeigen.“. Die Referenz „den Film“ kann nur mit Hilfe der Dialoghistorie aufgelöst werden. Die MMP-Komponente verwendet die Dialoghistorie dazu, um geeignete Übergänge zwischen zwei Präsentationen zu generieren. So könnte z. B. durch Zugriff auf die Ausgabehistorie ein langsames Überblenden zur nächsten Präsentation erfolgen. Weiterhin wird die Ausgabehistorie vom Dialog-Manager dazu verwendet, um Referenzen auf Bildschirminhalte aufzulösen (z. B. „Bitte das Bild auf dem Fernseher maximieren!“).

### 4.1.2 Generierung multi-modaler Präsentationen

Die Generierung multi-modaler Präsentationen durch die MMP-Komponente lässt sich in mehrere Phasen unterteilen, die im Folgenden näher betrachtet werden sollen. In [André & Rist, 1991] findet sich eine Aufstellung dieser Phasen.

Vor der eigentlichen Generierung findet zunächst eine *Auswahl und Strukturierung des Ausgabeinhaltes* statt, der durch das System präsentiert werden soll. Dieser Schritt ist notwendig, falls unterschiedliche Inhalte oder Strukturierungen zur Umsetzung der Ausgabe möglich sind. Dabei kann die Auswahl und Strukturierung manuell durch einen Benutzer geschehen [André et al., 2005] oder automatisiert [André, 1995] erfolgen. In Abbildung 4.1 erfolgt die Inhaltsauswahl durch den Dialog-Manager  $D$  in Kooperation mit den Applikationen  $P_n$ .

Nachdem ein geeigneter Ausgabeinhalt ausgewählt wurde, findet die *Verteilung des Inhaltes* auf eine gegebene Menge von Ausgabekomponenten statt. Dies erfolgt durch die MMP-Komponente in Kooperation mit den Ausgabekomponenten  $R_n$ . Dies können z. B. Komponenten zur Realisierung von statischen Bildern, Animationen oder einer Sprachausgabe sein. Im Allgemeinen kann nicht jede Ausgabekomponente jeden Inhaltstyp verarbeiten. So benötigt z. B. eine Komponente zur Präsentation eines Stadtplans von Heidelberg Angaben über die Geographie von Heidelberg, während eine Komponente zu

Generierung von gesprochener Sprache Zeichenfolgen als Eingabe benötigt. Die Auswahl und Strukturierung des Inhalts sowie die Generierung der Präsentation beeinflussen sich gegenseitig. So kann ein umfangreicher Inhalt bei beschränkten Ausgaberesourcen (z. B. einem Handy-Display) oft nicht dargestellt werden und erfordert eine Wiederholung der Auswahl und der Strukturierung. Andererseits hängt die Auswahl von Ausgabekomponenten zur Generierung der Präsentation vom Ausgabeinhalt ab. Aus diesem Grund müssen beide Prozesse parallel stattfinden. Nachdem der Inhalt auf eine Menge von Ausgabekomponenten verteilt wurde, findet die *modalitätenspezifische Umsetzung* des Inhalts statt. Dies erfolgt durch die Ausgabekomponenten  $R_n$ . Eine Sprachausgabekomponente könnte z. B. eine Zeichenfolge in ein Audiosignal umsetzen, das gesprochene Sprache enthält. Ein GUI könnte dieselbe Zeichenfolge visuell in einem Fenster auf dem Bildschirm erscheinen lassen. Falls kreuzmodale Referenzen zwischen verschiedenen Ausgabeteilen bestehen, beeinflussen sich die modalitätenspezifischen Umsetzungen verschiedener Ausgabekomponenten gegenseitig. Wenn z. B. ein GUI den Text „das Rathaus auf dem Stadtplan“ anzeigt, ist es notwendig, dass auf dem Stadtplan das Rathaus für den Benutzer eindeutig identifizierbar ist. Es können weiterhin Situationen auftreten, in denen die modalitätenspezifische Umsetzung fehlschlägt. Dies kann z. B. der Fall sein, falls zur Darstellung eines Bildes nicht genügend Bildschirmplatz vorhanden ist wie z. B. auf einem Handy-Display. In diesem Fall muss die Verteilung des Inhalts oder die Auswahl und Strukturierung des Inhalts wiederholt werden. Nachdem jede Ausgabekomponente ihren spezifischen Inhalt generiert hat, müssen vor der eigentlichen Ausgabe des Inhalts noch alle Präsentationsteile in ein *multi-modales Layout* integriert werden.

Wenn z. B. eine Ausgabekomponente Bilder anzeigt und auf demselben Gerät eine andere Komponente grafischen Text präsentiert, dann müssen beide Ausgaben koordiniert werden, damit das Bild nicht den Text verdeckt oder umgekehrt. Falls keine Lösung für ein spezielles Layout-Problem existiert (z. B. weil zu wenig Layout-Platz zur Verfügung steht), müssen die Auswahl und Strukturierung des Inhalts, die Verteilung des Inhalts auf die Ausgabekomponenten oder die modalitätenspezifische Umsetzung wiederholt werden.

Die einzelnen Phasen der Präsentationsplanung beeinflussen sich somit gegenseitig. Aus diesem Grund ist es notwendig, bei Fehlschlägen vorhergehende Phasen zu wiederholen bis hin zur Inhaltsauswahl und -strukturierung. Daher muss in Abbildung 4.1 die Kommunikation zwischen dem Dialog-Manager, der MMP-Komponente, den  $R_n$ -Komponenten sowie den  $O_n$ -Komponenten in beiden Richtungen stattfinden.

## 4.2 Planungsansatz

Wie in Kapitel 3 dargestellt wurde, stellt der Planungsansatz eine flexible Methode zur Generierung multi-modaler Präsentationen dar. Im Folgenden wird daher ein kurzer Überblick über die Grundbegriffe des Planens und die damit verbundenen Definitionen gegeben. Dabei orientieren wir uns an den in [Nau et al., 2004] verwendeten Formalismen sowie den in [André et al., 2005] definierten Operatoren.

Ein *Präsentationsplanungsproblem* ist ein Tripel  $P := (O, s_0, g)$  über einer Menge  $S$  von Zuständen mit  $s_0, g \in S$  und einer Menge  $O$  von *Präsentationsoperatoren*  $o \in O$ .  $s_0$  wird als *Startzustand* der Präsentationsplanung bezeichnet.  $g$  wird als *Zielzustand* der Präsentationsplanung bezeichnet. Jeder Präsentationsoperator  $o \in O$  lässt sich wie folgt darstellen:

$$o = (\text{header}(o), \text{constraints}(o), \text{inferiors}(o), \text{spatial}(o), \\ \text{temporal}(o), \text{start}(o), \text{finish}(o))$$

Dabei bezeichnet  $\text{header}(o)$  den Kopf des Operators,  $\text{constraints}(o)$  die Vorbedingungen für die Ausführung des Operators,  $\text{inferiors}(o)$  die Effekte des Operators,  $\text{spatial}(o)$  die Constraints für das Layout der Effekte in  $\text{inferiors}(o)$  und  $\text{temporal}(o)$  die Constraints für die zeitliche Abfolge der Effekte in  $\text{inferiors}(o)$ .  $\text{start}(o)$  gibt den ersten Effekt an, der stattfinden soll.  $\text{finish}(o)$  gibt den letzten Effekt an, der stattfinden soll.

Ein Beispiel für einen Operator  $o \in O$  wird in Abbildung 4.2 gegeben. Der Kopf des Operators enthält den Bezeichner `build-smil-pres` und den Bezeichner `?rc-id` einer Ausgabekomponente sowie die ULR `?im-url` eines Bildes. In den Vorbedingungen für die Anwendung des Operators wird die Existenz einer WAV-Datei mit der URL `?om-url` überprüft.

Dabei symbolisiert das Prädikat `BELP` einen Eintrag in der Wissensbasis des Planers. Falls dies gegeben ist, werden die in  $\text{inferiors}$  beschriebenen Effekte des Operators betrachtet. Zunächst werden Bild- und WAV-Datei mittels eines weiteren Operators zu einer SMIL-Präsentation zusammengeführt (`build-smil-img-with-speech`). Anschließend wird an die Ausgabekomponente `?rc-id` eine Nachricht mit der URL der Präsentation mittels des Operators `send-message` gesendet. Diese Effekte treten jedoch nur ein, wenn die in  $\text{temporal}$  und  $\text{spatial}$  gegebenen Constraints erfüllt sind. Die Layout-Constraints in  $\text{spatial}$  fordern, dass die durch Aktion A1 realisierte Grafik horizontal und vertikal auf dem Bildschirm zentriert sein soll.

Die zeitlichen Constraints in  $\text{temporal}$  fordern, dass zuerst die SMIL-Datei generiert wird (Aktion A1) und anschließend die Nachricht an `?rc-id` ge-

sendet werden soll (Aktion A2). Eine Auflistung aller definierten Constraints findet sich in [André et al., 2000].

```
(define-plan-operator
  :header (A0 (build-smil-pres ?rc-id ?im-url))
  :constraints
    (BELP (wav-file ?om-url))
  :inferiors (
    (A1 (build-smil-img-with-speech ?im-url
      ?rc-id-2 ?om-url))
    (A2 (send-message
      ?rc-id "http://11.0.0.1/pres.smi")))
  :temporal (
    (A1 (m) A2))
  :spatial (
    (centerh A1)
    (centerv A1))
  :start (A1)
  :finish (A2))
```

Abbildung 4.2: Präsentationsoperator.

Ein möglicher Zielzustand  $g$  könnte z. B. die folgende Form haben:

```
g := (build-smil-pres 1421 "http://11.0.0.1/1.gif")
```

Ein möglicher Startzustand  $s_0$  könnte die folgende Form haben:

```
s0 := (BELP (wav-file "http://11.0.0.1/1.wav"))
```

Kann das Ziel  $g$  durch sukzessive Anwendung von Operatoren aus  $O$  bei gegebenem Startzustand  $s_0$  in eine Folge von nicht-zerlegbaren, primitiven Aktionen  $(a_1, \dots, a_k)$  zerlegt werden, die allen geforderten Constraints genügt, so wird diese Folge als *Lösung* des Präsentationsplanungsproblems bezeichnet.

### 4.3 Unimodalitätenmodell nach Bernsen

Im diesem Abschnitt erfolgt eine Einführung in die Grundbegriffe der Modalitätentheorie nach Bernsen [Bernsen, 2001]. Bernsen hat in seiner Arbeit ein Modell der Basisklassen multi-modaler Ausgaben (*Unimodalitäten*) definiert. Er zeigt, dass das Modell vollständig und eindeutig ist. Im folgenden wird ein Überblick über das Modell gegeben.

### 4.3. UNIMODALITÄTENMODELL NACH BERNSEN

Da sich die vorliegende Arbeit auf visuelle Ausgabemedien (vis) und audiotorielle Ausgabemedien (aud) beschränkt, werden die taktilen Ausgabemodalitäten in Bernsens Modell nicht betrachtet. Die resultierenden Ausgabemodalitäten werden in Tabelle 4.1 dargestellt. Bei den Modalitäteneigenschaften unterscheidet Bernsen zwischen linguistischen (li), statischen (sta), dynamischen (dyn), arbiträren (ar) und analogen (an) Unimodalitäten.

Tabelle 4.1: Generische Ausgabeunimodalitätenklassen.

| Unimodalität                                       | Kategorie                 |
|----------------------------------------------------|---------------------------|
| <b>I. Linguistische Modalitäten</b>                |                           |
| 1. Statische analoge visuelle Sprache              | (li, an, -ar, sta, vis)   |
| 2. Dynamische analoge visuelle Sprache             | (li, an, -ar, dyn, vis)   |
| 3. Statische analoge audiotorielle Sprache         | (li, an, -ar, sta, aud)   |
| 4. Dynamische analoge audiotorielle Sprache        | (li, an, -ar, dyn, aud)   |
| 5. Statische nicht-analoge visuelle Sprache        | (li, -an, -ar, sta, vis)  |
| 6. Dynamische nicht-analoge visuelle Sprache       | (li, -an, -ar, dyn, vis)  |
| 7. Statische nicht-analoge audiotorielle Sprache   | (li, -an, -ar, sta, aud)  |
| 8. Dynamische nicht-analoge audiotorielle Sprache  | (li, -an, -ar, dyn, aud)  |
| <b>II. Analoge Modalitäten</b>                     |                           |
| 9. Statische analoge visuelle Modalitäten          | (-li, an, -ar, sta, vis)  |
| 10. Dynamische analoge visuelle Modalitäten        | (-li, an, -ar, dyn, vis)  |
| 11. Statische analoge audiotorielle Modalitäten    | (-li, an, -ar, sta, aud)  |
| 12. Dynamische analoge audiotorielle Modalitäten   | (-li, an, -ar, dyn, aud)  |
| <b>III. Arbiträre Modalitäten</b>                  |                           |
| 13. Arbiträre statische visuelle Modalitäten       | (-li, -an, ar, sta, vis)  |
| 14. Arbiträre dynamische visuelle Modalitäten      | (-li, -an, ar, dyn, vis)  |
| 15. Arbiträre statische audiotorielle Modalitäten  | (-li, -an, ar, sta, aud)  |
| 16. Arbiträre dynamische audiotorielle Modalitäten | (-li, -an, ar, dyn, aud)  |
| <b>IV. Explizite Modalitäten</b>                   |                           |
| 17. Statische visuelle Strukturen                  | (-li, -an, -ar, sta, vis) |
| 18. Dynamische visuelle Strukturen                 | (-li, -an, -ar, dyn, vis) |
| 19. Statische audiotorielle Strukturen             | (-li, -an, -ar, sta, aud) |
| 20. Dynamische audiotorielle Strukturen            | (-li, -an, -ar, dyn, aud) |

Die Klassen eins und zwei in der Tabelle umfassen linguistische Bilder und Animationen. Klasse eins umfasst daher Hieroglyphen. Klasse zwei umfasst die Lippenbewegungen eines animierten Charakters, an denen sich ebenfalls die Phoneme und Wörter des gesprochenen Textes visuell ablesen lassen.

Die Klassen zwei und drei beschreiben audiotorielle, linguistische Bilder. Dies kann z. B. das Geräusch eines Geigerzählers sein, der Radioaktivität durch die

## KAPITEL 4. GRUNDLAGEN

Intensität ihrer Strahlung auf abstrakte Weise auditoriell beschreibt. Statische, auditorielle Ausgaben bedeuten in diesem Zusammenhang eine andauernde Wiederholung der Ausgabe.

Die Klassen fünf und sechs beschreiben statischen und animierten Text. Die Klassen sieben und acht beschreiben Sprachausgaben im auditoriellen Medium. Die Klassen neun und zehn beschreiben statische Bilder sowie Animationen. Die Klassen elf und zwölf beschreiben auditorielle Bilder, wie z. B. den Schrei einer Möwe als Umschreibung der Möwe. Lautmalerei oder -imitation kann eine intuitive Form der Benachrichtigung darstellen.

Im Home-Entertainment-Bereich wird mit „Query by Humming“ die Identifikation von digitalen Musikstücken durch Vorsummen bezeichnet (siehe z. B. [Pauws, 2002]). Ebenso könnten auch kurze lautmalerische Benachrichtigungen als Erinnerung verwendet werden. Ein Beispiel hierfür ist das Geräusch von brutzelndem Fleisch, das als Erinnerung, den Herd auszuschalten, verwendet wird.

Arbiträre Modalitäten kommen gänzlich ohne Semantik aus (Klassen 13-16). Die Klassen 17-20 werden von Bernsen als explizite Strukturen bezeichnet, die weder linguistisch noch analog noch arbiträr sind. In diese Kategorie fallen Strukturierungsmodalitäten, z. B. Tabellen. Im auditoriellen Medium entsprechen die Klassen 19 und 20 Ausgabemodalitäten, die weder linguistisch, noch analog, noch arbiträr sind. Dies entspricht im statischen Fall der Wiederholung von kurzen Jingles. Im dynamischen Fall entspricht dies Musik.

## **5 Modellierung multi-modaler Ausgaben in ubiquitären Informationssystemen**

Die Generierung von multi-modalen Ausgaben in ubiquitären Informationssystemen erfordert eine Erweiterung der Ansätze für multi-modale Dialogsysteme. Die Komponenten eines ubiquitären Informationssystems sind im Unterschied zu den klassischen multi-modalen Dialogsystemen über zahlreiche Geräte verteilt.

Durch die Verwendung mobiler Geräte können dem System weiterhin Ausgabemöglichkeiten dynamisch hinzugefügt oder entzogen werden. Dies erfordert in ubiquitären Informationssystemen insbesondere eine abstrakte Beschreibung von Ausgabediensten (z. B. einer Wiedergabekomponente für MPEG-Videos), Ausgabegeräten (z. B. einem Wandbildschirm in der Küche) sowie Ausgaberesourcen (z. B. der Auflösung eines Bildschirms). Die Modellierung von Ausgabediensten, -geräten und -ressourcen stellt somit die Grundlage für die Kopplung von Ausgabegeräten in ubiquitären Informationssystemen dar.

In diesem Abschnitt werden zunächst die Möglichkeiten und Herausforderungen von multi-modalen Ausgaben mit gekoppelten Ausgabegeräten anhand eines Beispiels erläutert (Abschnitt 5.1). Anschließend wird auf die Architektur von multi-modalen Dialogsystemen, die über mehrere Geräte verteilt sind, eingegangen und die benötigten Modelle zur automatisierten Generierung multi-modaler Ausgaben identifiziert (Abschnitt 5.2).

In Abschnitt 5.3 werden vier Möglichkeiten zur Modellierung von Diensten, Geräten und Ressourcen von Ein-/Ausgabekomponenten gegenübergestellt. Anschließend wird ein allgemeines Modell definiert und anhand von Beispielen erläutert (Abschnitt 5.4). Am Beispiel einer Home-Entertainment-Umgebung werden Ein-/Ausgabekomponenten mittels des Modells beschrieben. Abschnitt 5.5 beschreibt den Einsatz des Modells in der Präsentationsplanung. Das Kapitel schließt mit einer Zusammenfassung (Abschnitt 5.7).

### **5.1 Beispiel**

In diesem Abschnitt sollen die Anwendungsmöglichkeiten und die Herausforderungen einer multi-modalen Ausgabe mit gekoppelten Ausgabegeräten zunächst durch ein Beispiel aus dem Home-Entertainment-Bereich motiviert werden. Typischerweise verfügen Benutzer heutzutage bereits über zahlreiche Ausgabegeräte wie z. B. einen Fernseher, ein Mobiltelefon, eine Stereoanlage und einen PDA. Im Beispiel wird diesen Geräten noch ein weiteres Ausgabegerät, ein elektronischer Bilderrahmen, hinzugefügt. Eine flexible Nutzung

und Kopplung dieser Ausgabegeräte soll durch das folgende Szenario illustriert werden:

*Martin befindet sich in der Küche seiner Wohnung und spült, als ihn sein neues Home-Entertainment-System informieren möchte, dass einer seiner Lieblingsfilme in fünf Minuten startet. Da sich Martin nicht in Sichtweite eines der visuellen Ausgabegeräte befindet, wird eine Sprachausgabe auf die eingeschaltete Stereoanlage geleitet, die über eine genügend große Audioleistung verfügt. Daraufhin schaltet Martin den Fernseher und den daneben stehenden elektronischen Bilderrahmen ein und fragt Informationen zum Film ab, die aus dem elektronischen Programmführer abgerufen werden (Abbildung 5.1, links).*

*Dazu wird auf dem Fernseher ein Bild des Films gezeigt sowie ein animierter Charakter verwendet, um den Filminhalt vorzulesen. Auf dem Bilderrahmen, der nur Standbilder darstellen kann, wird ein weiteres Bild aus dem Film angezeigt. Dabei sind die Layouts an die Auflösungen der zwei Geräte angepasst worden.*



Abbildung 5.1: Präsentationen auf den Gerätekombinationen Fernseher-Bilderrahmen (linke Seite) und PDA (rechte Seite).

*Wenn Martin im Schlafzimmer ist und den PDA verwendet, kann er dieselben Informationen auch mit diesem Gerät abfragen. In diesem Fall erfolgt keine Kopplung mit weiteren Ausgabegeräten, da sich kein weiteres Gerät in der Nähe von Mar-*

## 5.1. BEISPIEL

tin befindet. Das Ergebnis besteht in diesem Fall aus einer an die Ressourcen des PDAs angepassten Ausgabe (Abbildung 5.1, rechts). So wird das Layout auf die geringe Auflösung des PDAs von  $240 \times 320$  Pixeln reduziert. Aufgrund der beschränkten Kennwerte des PDAs werden weiterhin nur Bild- und Audiodateien von reduzierter Qualität verwendet.

Da der PDA tragbar ist, ist es für Martin auch möglich, den PDA mit weiteren Ausgabegeräten wie z. B. dem Fernseher zu kombinieren. So kann auf dem PDA Text dargestellt werden, da er durch die Eingabetasten des PDA leicht scrollbar ist. Bilder können in diesem Fall auf dem Fernseher ausgegeben werden, der über eine größere Auflösung als der PDA verfügt (Abbildung 5.2, links). Ähnlich zum PDA kann Martin auch das Mobiltelefon als Ausgabegerät verwenden, falls der PDA nicht verfügbar ist (Abbildung 5.2, rechts). Im Unterschied zum PDA verfügt das Mobiltelefon jedoch nur über begrenzte Netzwerk- und Ausgabefähigkeiten. Informationen können im MMS-Format<sup>1</sup> dargestellt werden. Im Unterschied zum PDA können hingegen keine komplexen Audiodaten über das WLAN-Funknetzwerk übertragen werden. Aus diesem Grund sind die Filminformationen nur als Text-Bild-Ausgaben verfügbar.

Abends kommt Martins Freundin Melanie nach Hause. Sie verwendet lieber ihren Laptop, um sich über Filme zu informieren. Dabei wird automatisch der Charakter verwendet, den ihr Laptop zur Verfügung stellt. Als Melanie zum Fernseher wechselt, folgt ihr der Charakter automatisch und erscheint auf dem Bildschirm. Melanies Charakter besteht im Unterschied zu Martins Charakter nur aus einem lächelnden Gesicht und kann keine Gesten ausführen. Als der Charakter ein Bild mit den Worten „Das hier ist der Schauspieler Jürgen Vogel.“ präsentiert, leuchtet das Bild kurz auf, da der Charakter keine Geste ausführen kann.

Dieses Szenario verdeutlicht die Herausforderungen, denen ein ubiquitäres Informationssystem mit gekoppelten Ausgabegeräten begegnen muss. Da in dem Beispiel Ausgabegeräte ein- und ausgeschaltet werden können, ist es insbesondere notwendig, dass sich vorhandene Geräte gegenüber dem System identifizieren. Dies betrifft zunächst die Ressourcen der Geräte-Hardware. So ist es notwendig, dass ein PDA seine Auflösung angibt, damit Ausgaben an das Geräte-Layout angepasst werden können.

---

<sup>1</sup>Multimedia Messaging Service, <http://www.3gpp.org>



Abbildung 5.2: Präsentationen auf den Gerätekombinationen PDA-Fernseher (linke Seite) und Mobiltelefon (rechte Seite).

Weiterhin ist es nötig, dass die Ausgabe-Software beschrieben wird, die sich auf den Geräten befindet. So können z. B. auf dem elektronischen Bilderahmen keine Audioausgaben stattfinden, sondern nur Bilder im JPG-Format dargestellt werden. Diese Selbstbeschreibungen ermöglichen es, dass unterschiedliche Gerätetypen wie Mobiltelefone und Fernseher gekoppelt werden können. Diese Anforderung an ubiquitäre Informationssysteme wird als Ad-hoc-Interoperabilität bezeichnet und wird in Abschnitt 5.3 näher betrachtet.

Um die Ad-hoc-Interoperabilität in den Gesamtkontext von verteilten, multi-modalen Dialogsystemen einzuordnen und die im Kontext der Präsentationsgenerierung auftretenden Abhängigkeiten zu identifizieren, wird im folgenden Abschnitt zunächst ein allgemeiner Überblick über die Architektur und die zugrunde liegenden Modelle dieser Systeme gegeben.

## 5.2 Verteilte multi-modale Dialogsysteme

Die in Abschnitt 4.1 vorgestellten multi-modalen Dialogsysteme umfassten nur ein physikalisches Gerät, z. B. einen PC mit Bildschirm und Lautsprechern. Der Vorteil dieses Ansatzes ist, dass alle Komponenten des Systems vollständig bekannt und jederzeit aktiv sind. Diese Voraussetzungen sind jedoch in ubiquitären Informationssystemen nicht mehr gegeben.

## 5.2. VERTEILTE MULTI-MODALE DIALOGSYSTEME

In ubiquitären Informationssystemen sind die Komponenten des Systems über mehrere vernetzte Geräte verteilt. Im Beispiel aus Abschnitt 5.1 befinden sich z. B. Ausgabegeräte sowohl im Wohnzimmer als auch im Schlafzimmer der Wohnung. Weiterhin unterscheiden sich diese Geräte in Bezug auf Software und Hardware. Mobile Geräte wie der PDA im Beispielszenario können den Ort wechseln und jederzeit ausgeschaltet werden. In diesem Abschnitt sollen die Abhängigkeiten, die sich daraus für die automatisierte Generierung multi-modaler Präsentationen ergeben, näher untersucht werden. Dazu wird zunächst die Architektur von verteilten multi-modalen Dialogsystemen im Unterschied zu Systemen auf einem Gerät betrachtet. Anschließend werden die Abhängigkeiten der automatisierten Generierungen multi-modaler Ausgaben in diesen Systemen betrachtet und daraus ein Datenfluss für das Präsentationssystem hergeleitet.

### 5.2.1 Architektur

In verteilten Geräteumgebungen, wie sie in ubiquitären Informationssystemen anzutreffen sind, ergeben sich für multi-modale Dialogsysteme weitere Herausforderungen. Abbildung 5.3 zeigt ein verteiltes Dialogsystem, das sich aus einem PDA, einem Fernseher und einem Server zusammensetzt.

Die Infrastruktur wird von dem Server zur Verfügung gestellt. Auf dem PDA existieren Komponenten zur Spracheingabe sowie zur Sprachanalyse. Weiterhin ist eine Ausgabekomponente zur Wiedergabe von SMIL-Ausgaben [W3C Recommendation, 2005] vorhanden. Auf dem Fernseher ist ebenfalls eine SMIL-Ausgabekomponente vorhanden sowie ein elektronischer Programmführer (Electronic Program Guide, EPG), der aus dem Fernsehsignal Informationen zu laufenden Sendungen abrufen kann. Auf dem Server befinden sich der Dialog-Manager, die Komponenten zur multi-modalen Fusion (MMF) bzw. Präsentation (MMP) sowie eine Textgenerierung und eine Komponente zur Generierung der Sprach- und Grafikdaten des animierten Charakters (vgl. Abbildung 4.1).

Im Unterschied zu multi-modalen Dialogsystemen, die nur ein Gerät umfassen, ergeben sich weitere Herausforderungen. In dem Beispiel wird ein PDA verwendet, der jederzeit vom Benutzer ein- oder ausgeschaltet werden kann. Weiterhin kann der PDA außer Reichweite des Systems sein. Daher müssen multi-modale Präsentationssysteme in verteilten Geräteumgebungen insbesondere in der Lage sein, sich an dynamische Veränderungen in der Geräteumgebung anpassen zu können.

Weiterhin ist die Unterstützung von vielen Interaktionsgeräten wünschenswert. Dies ist insbesondere der Fall, wenn das System von mehreren Benutzern verwendet werden soll und jeder der Benutzer eigene mobile Geräte in

## KAPITEL 5. MODELLIERUNG MULTI-MODALER AUSGABEN

das System einführt. Im Beispiel in Abschnitt 5.1 erfolgt dies durch Melanie, die ihren eigenen Laptop mit dem System verbindet.

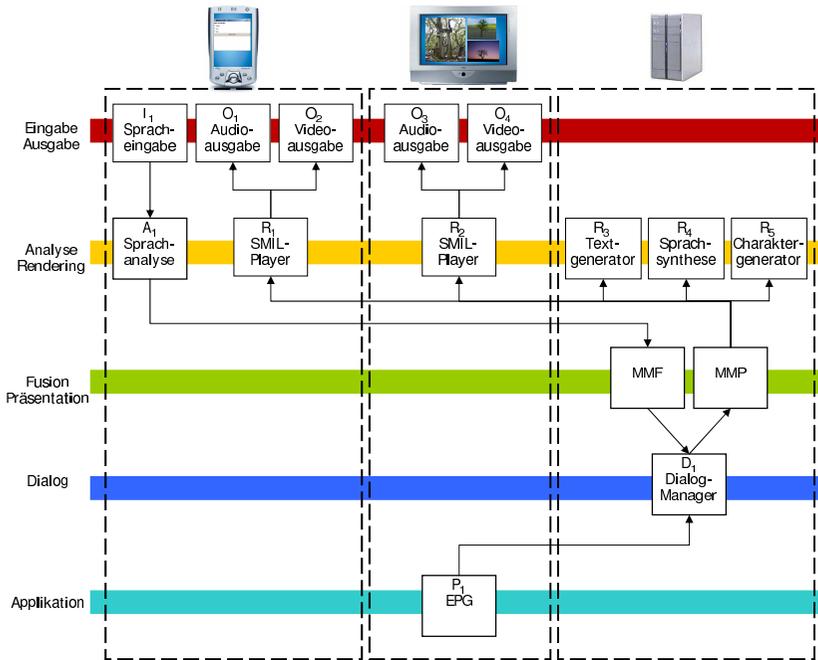


Abbildung 5.3: Beispiel multi-modales Dialogsystem in verteilter Geräteumgebung.

Die multi-modale Ausgabe findet auf mehreren Geräten statt. Insbesondere bedeutet dies, dass das System in der Lage sein muss, Daten von einem Gerät auf das andere zu übertragen. Dies kann z. B. durch einen Webserver erfolgen oder durch spezielle Komponenten zur sukzessiven Übertragung umfangreicher Mediendaten (Streaming). Weiterhin müssen Ausgaben auf mehreren Geräten u. U. synchronisiert werden. So könnte der animierte Charakter auf dem Fernseher z. B. sagen „In diesem Bild sehen Sie Humphrey Bogart“. Sobald der Charakter „diesem Bild“ sagt, erscheint auf dem PDA das Bild des Schauspielers. Hierfür ist insbesondere eine Synchronisation der Systemzeiten zwischen den einzelnen Geräten notwendig. Diese Abhängigkeiten der Präsentationsgenerierung werden im folgenden Abschnitt formalisiert.

## 5.2. VERTEILTE MULTI-MODALE DIALOGSYSTEME

Bei verteilten multi-modalen Dialogsystemen gibt es weitere Herausforderungen, die in dieser Arbeit nicht betrachtet werden. Im Beispiel wird ein Server verwendet, auf dem zentrale Komponenten der Infrastruktur lokalisiert sind. Wird dieser Server vom System getrennt oder abgeschaltet, so kann das Dialogsystem keine Benutzereingaben mehr verarbeiten. Eine mögliche Lösung ist eine physikalisch verteilte Implementierung der Infrastruktur und der zentralen Komponenten des Dialog-Managers sowie der multi-modalen Fusion und Präsentation [Hellenschmidt, 2005; Elting & Hellenschmidt, 2004]. Weiterhin ist es möglich, dass mehrere Geräte mehrere Dialog-Manager einführen. In diesem Fall sind Strategien oder Komponenten zur Konfliktlösung zwischen Dialog-Managern notwendig. Diese verteilte Implementierung der Infrastruktur und des Dialog-Managements sind jedoch nicht Gegenstand dieser Arbeit.

### 5.2.2 Modelle und Abhängigkeiten der Präsentationsgenerierung

Multi-modale Präsentationen werden von verschiedenen Faktoren beeinflusst, die bei der automatisierten Generierung beachtet werden müssen (Abbildung 5.4). In verteilten Geräteumgebungen hängt die Wahl des Ausgabegerätes (und damit der auf dem Gerät möglichen Präsentationen) vom Ort ab, an dem sich das Ausgabegerät befindet [Kray et al., 2003]. Bei visuellen Ausgabegeräten ist weiterhin die Orientierung der Bildschirme relevant.

So können für Präsentationen nur Ausgabegeräte verwendet werden, deren visuellen, auditoriellen oder taktilen Ausgaben vom Benutzer wahrgenommen werden können. Die Wahl des Ausgabegerätes hängt ebenfalls von den Eingabegeräten ab, die der Benutzer verwendet hat. So kann es z. B. sinnvoll sein, eine Präsentation auf einem Ausgabegerät zu generieren, das von dem Eingabegerät aus leicht einsehbar oder wahrnehmbar ist.

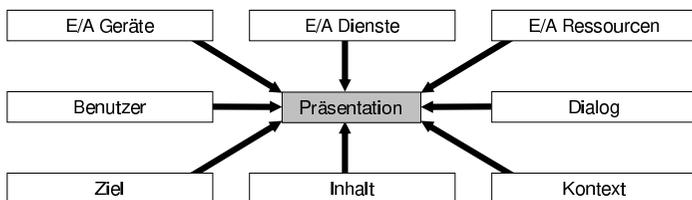


Abbildung 5.4: Abhängigkeiten von multi-modalen Präsentationen in verteilten Geräteumgebungen.

Weiterhin ist es notwendig, eine Präsentation an die auf einem Ausgabegerät vorhandenen Ausgabedienste anzupassen. Mit Ausgabediensten werden die Software-Komponenten bezeichnet, die für die Umsetzung von Ausgaben auf

dem Gerät verantwortlich sind. Auf Ausgabegeräten, auf denen z. B. nur auditorielle Ausgaben möglich sind, müssen Präsentationen so angepasst werden, dass sie ohne visuelle Zusatzausgaben verständlich sind.

Die Eingabedienste eines Gerätes beeinflussen ebenfalls die Präsentation. Stellen Benutzer z. B. eine Anfrage mit der Eingabemodalität Sprache, so ist es i. Allg. sinnvoll, mit einer auditoriellen Ausgabemodalität zu antworten, die in demselben Medium wie die Eingabe erfolgt [Wasinger, 2006]. Darüber hinaus ist es nicht sinnvoll, eine Ausgabe, die den Benutzer zu einer Eingabe auffordert, auf einem Gerät stattfinden zu lassen, das über keinerlei Dienste für eine Benutzereingabe verfügt.

Präsentationen in verteilten Geräteumgebungen müssen analog zu Präsentationen auf einem einzelnen Gerät an die Ressourcen der Ausgabegeräte angepasst werden. So ist es z. B. notwendig, eine Präsentation auf einem PDA an die geringere Auflösung des PDAs anzupassen. Weitere für Präsentationen relevante Geräteeigenschaften sind die Grafik- und Netzwerkennwerte der Ausgabegeräte. So können z. B. auf Geräten mit geringen Netzwerkfähigkeiten nur Bilder oder Animationen von reduzierter Qualität verwendet werden. Die Kennwerte lassen sich mittels geeigneter Software ermitteln. Als Beispiel sei die Software Spb Benchmark<sup>2</sup> genannt, die die Kennwerte von Windows Mobile-basierten PDAs ermittelt.

Ebenso können die Ressourcen der Eingabegeräte die Wahl der Präsentation beeinflussen. So können z. B. für eine Spracheingabe verschiedene Mikrophone unterschiedlicher Qualität vorhanden sein. Eine mögliche Präsentationsstrategie könnte sein, Interaktionen auf dem Ausgabegerät stattfinden zu lassen, auf dem dem Benutzer die qualitativ beste Spracheingabe zur Verfügung steht.

Dienste, Geräte und Ressourcen von Ein-/Ausgabekomponenten sind in verteilten multi-modalen Dialogsystemen von besonderer Bedeutung, falls Ein-/Ausgabegeräte dynamisch hinzugefügt und entfernt werden können. Dies ist der Fall bei Melanies Laptop und Martins PDA im Beispiel in Abschnitt 5.1. In diesem Fall ist eine Beschreibung der Dienste und Ressourcen jedes Ein-/Ausgabegerätes erforderlich, das dem System hinzugefügt wird, um die Präsentationen an die veränderte Geräteumgebung anpassen zu können.

Neben Diensten, Geräten und Ressourcen von Ein-/Ausgabekomponenten ist es nötig, eine Präsentation an den oder die Adressaten anzupassen. Die einfachste Art einer solchen Personalisierung erfolgt durch ein statisches Präferenzenmodell, mit dem der Benutzer das System manuell anpassen kann. Weiterhin ist es möglich, das Verhalten von Benutzern zu beobachten und dar-

---

<sup>2</sup>Spb Benchmark, <http://www.spbsoftwarehouse.com/products/benchmark>

## 5.2. VERTEILTE MULTI-MODALE DIALOGSYSTEME

aus automatisch Rückschlüsse auf deren Präferenzen zu ziehen (siehe z. B. [Bohnenberger et al., 2002]). Der Online-Versandhandel Amazon verwendet dieses Verfahren, um die Kaufempfehlungen auf ihren Webseiten an die bisherigen Käufe und Interessen der Benutzer anzupassen [Linden et al., 2003]. Die Präsentation wird ebenfalls durch den Dialog zwischen System und Benutzer beeinflusst. Falls der Benutzer z. B. fortlaufend dieselben Anfragen stellt, könnte das System folgern, dass der Benutzer Schwierigkeiten bei der Bedienung des Systems hat. Die Generierung von ausführlicheren Präsentationen könnte eine mögliche Reaktion hierauf sein. Weiterhin müssen Präsentationen auch an die Ausgaben des vorherigen Dialogschritts angepasst werden. Dies betrifft insbesondere die Generierung von geeigneten Übergängen zwischen zwei aufeinander folgenden Präsentationen [Zhen et al., 2005].

In einer verteilten Geräteumgebung wird die Präsentation ebenfalls durch die Position und die Orientierung des Benutzers beeinflusst. So wird z. B. in Abschnitt 5.1 eine audiotrielle Ausgabe verwendet, als Martin sich in der Küche befindet. Dieser Einfluss betrifft insbesondere die Ausgabe von geographischen Daten in lokationsbasierten Systemen. Die Ausgabe erfolgt typischerweise in Form von zwei- oder dreidimensionalen Karten, die an den Ort und die Position von Benutzern angepasst sind (siehe z. B. [Coors et al., 2005; Malaka et al., 2004]).

Weiterhin beeinflusst das kommunikative Ziel der Ausgabe die Umsetzung der Präsentation [André, 1995]. So könnte die Präsentation des Satzes „Der Film fängt in fünf Minuten an“ visuell durch eine kurze Meldung erfolgen, wenn der Benutzer nur allgemein informiert werden soll. Soll der Benutzer darauf hingewiesen werden, dass es sich um seinen Lieblingsfilm handelt, so könnte zusätzlich eine Sprachausgabe erfolgen. Der darzustellende Inhalt beeinflusst ebenfalls das Präsentationsformat. Geographische Daten können zum Beispiel sehr gut graphisch durch Landkarten visualisiert werden, während sie sich nur eingeschränkt für eine rein sprachliche Präsentation eignen. Neben Kontextinformationen über den Benutzer und den Dialog sind weitere Typen von Kontextinformationen vorstellbar, die Einfluss auf die Generierung von Präsentationen haben können. So kann bei der Erstellung von Präsentationen z. B. die Helligkeit eines Raumes berücksichtigt werden, um Grafiken in stärkeren oder schwächeren Farbkontrasten darzustellen.

### 5.2.3 Formulierung Präsentationsproblem

Aus der Architektur von verteilten multi-modalen Dialogsystemen (Abschnitt 5.2.1) und den Abhängigkeiten der Präsentationsgenerierung (Abschnitt 5.2.2) ergibt sich ein Datenfluss für die Präsentationsgenerierung. Wie in Abschnitt 5.2.1 beschrieben wurde, ist die MMP-Komponente für die Planung der Prä-

sensation verantwortlich. Im Unterschied zu Abbildung 5.3 wurden in der Abbildung die Kontextdatenbank, die Dialoghistorie sowie Sensoren  $S_n$  berücksichtigt, die Kontextdaten liefern. Auf die Darstellung der Applikationskomponenten  $P_n$  sowie der MMF-Komponente, durch die die multi-modale Fusion erfolgt, wurde in dieser Abbildung verzichtet.

Da Ausgabegeräte (z. B. ein PDA oder ein Fernseher) zu beliebigen Zeitpunkten dem System hinzugefügt oder entzogen werden können (z. B. durch Ein- oder Ausschalten), ist eine Registrierung bzw. Deregistrierung der Ausgabekomponenten bei der MMP-Komponente notwendig. Da Präsentationen sowohl von Eingabe- als auch von Ausgabegeräten beeinflusst werden (vgl. Abschnitt 5.2.2), müssen sich Eingabekomponenten ebenfalls registrieren und deregistrieren.

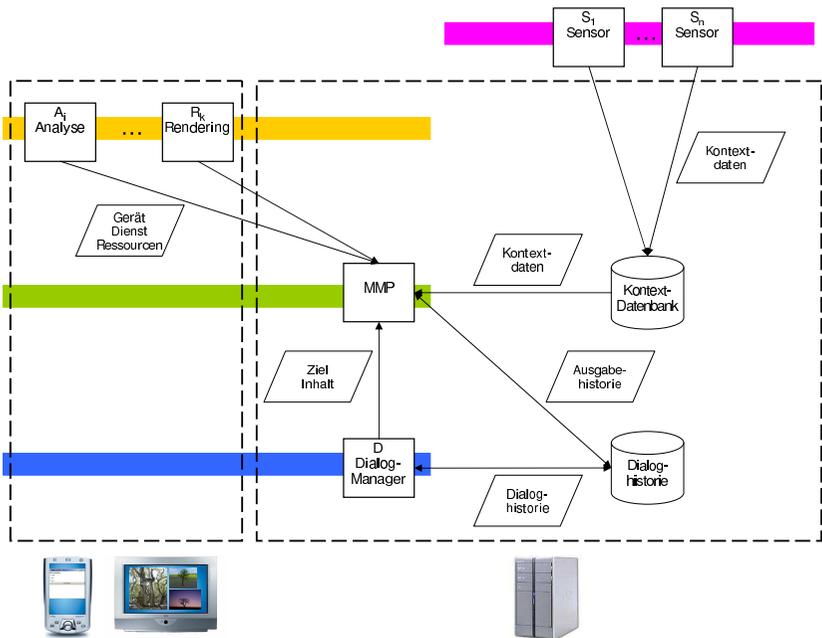


Abbildung 5.5: Datenfluss in einem verteilten multi-modalen Dialogsystem.

Der Einfachheit halber wird die Registrierung auf die Analysekomponenten  $A_n$  sowie die Rendering-Komponenten  $R_n$  beschränkt und auf die Darstellung der Eingabekomponenten  $I_n$  und der physikalischen Ausgabekomponenten

## 5.2. VERTEILTE MULTI-MODALE DIALOGSYSTEME

ten  $O_n$  verzichtet. Dies stellt keine Einschränkung dar, wenn angenommen wird, dass  $R_n$ -Komponenten nur auf  $O_n$ -Komponenten auf demselben Gerät ausgeben, sowie Eingabekomponenten  $I_n$  und die korrespondierende Analysekomponente  $A_n$  als eine Einheit betrachtet werden können.

Die Registrierungsnachrichten der Ein-/Ausgabekomponenten enthalten das Gerät, auf dem die Ein- oder Ausgaben stattfinden sowie die Dienste und die Ressourcenbeschränkungen des Geräts wie in Abschnitt 5.2.2 beschrieben. Diese Informationen werden in der MMP-Komponente gespeichert.

Die Kontextdatenbank wird verwendet, um durch Sensoren  $S_n$  beobachtete Ereignisse festzuhalten. Zum Beispiel kann die Position von Benutzern durch eine Komponente zur Personenlokalisierung in der Datenbank gespeichert werden. Diese Ereignisse werden von der MMP-Komponente abgefragt und können bei der Präsentationsgenerierung verwendet werden.

Weiterhin wird eine Repräsentation von jeder Ausgabe, die von der MMP-Komponente generiert wurde, an die Dialoghistorie übermittelt. Dort steht sie der MMF-Komponente sowie dem Dialog-Manager zur Verfügung. Der Dialog-Manager initialisiert eine neue Ausgabegenerierung durch das Versenden des kommunikativen Zieles der Ausgabe sowie des auszugebenden Inhalts an die MMP-Komponente. Anschließend erfolgt durch die MMP-Komponente die Generierung der Präsentation.

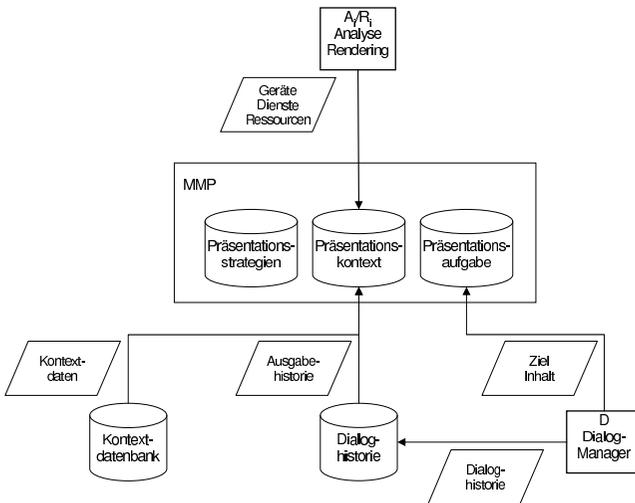


Abbildung 5.6: Präsentationsproblem in verteilten multi-modalen Dialogsystemen.

Aus Abbildung 5.5 ergibt sich der in Abbildung 5.6 dargestellte Aufbau des *Präsentationsproblems* in verteilten multi-modalen Dialogsystemen. Der *Präsentationskontext* speichert alle Daten, die die zurzeit gegebene Ausgabe-situation betreffen. Diese Daten betreffen die Kontextdaten sowie die bisherige Ausgabe-historie. Weiterhin werden die Geräte, Dienste und Ressourcen der zurzeit aktiven Analysekomponenten  $A_n$ , sowie der zurzeit aktiven Rendering-Komponenten  $R_n$  in den Präsentationskontext eingetragen.

Die *Präsentationsaufgabe* wird vom Dialog-Manager  $D$  übermittelt und setzt sich aus dem Ziel der Ausgabe und dem auszugebenden Inhalt zusammen. Die *Präsentationsstrategien* enthalten das Wissen, mit dem das MMP die Präsentation-saufgabe in Abhängigkeit des Präsentationskontextes bearbeitet. Das Lösung des Präsentationsproblems besteht in der Generierung einer Ausgabe, die dem Präsentationskontext und der Präsentation-saufgabe genügt.

### 5.3 Ad-hoc-Interoperabilität von Ausgabegeräten

Interoperabilität ist eine der großen Herausforderungen für ubiquitäre Informationssysteme [Edwards & Grinter, 2001]. Laut [Ranganathan et al., 2005] werden ubiquitäre Informationssysteme daran gemessen, mit wie vielen Personenmonaten Aufwand und geschriebenen Zeilen Quellcode sich neue Komponenten in sie integrieren lassen.

Da ubiquitäre Informationssysteme mit einer großen Anzahl von Komponenten interagieren müssen, ist es nötig, gemeinsame Modelle zu schaffen, die den Komponenten erlauben zu kommunizieren. Die für diese Arbeit relevanten Modelle für verteilte, multi-modale Dialogsysteme wurden bereits in Abbildung 5.4 als Abhängigkeiten der Präsentation-sgenerierung identifiziert. Der Schwerpunkt dieses Abschnitts ist die Definition von Modellen zur Beschreibung der Dienste, Geräte und Ressourcen von Ein-/Ausgabekomponenten.

Dabei stellt das Modell für Ein-/Ausgabedienste das komplexeste Modell dar, da mit ihm eine möglichst große Anzahl von multi-modalen Ausgabekategorien eindeutig beschrieben werden sollen. Ein Modell für Ausgabedienste ist eindeutig, falls unterschiedliche Ausgabekategorien stets durch das Modell auf unterschiedliche Klassen abgebildet werden. Je umfangreicher dieses Modell ist, desto mehr Ausgabegeräte können durch das System automatisch in das verteilte Dialogsystem integriert werden. Ist das Modell eindeutig, so können aufbauend auf dem Modell Präsentation-sstrategien definiert werden, die sämtliche multi-modalen Ausgaben des Modells unterstützen.

Der Schwerpunkt dieser Arbeit liegt hierbei auf der Definition des Modells für Ausgabedienste, das zuerst betrachtet werden soll. Für die Definition dieses Modells existieren drei mögliche Ansätze, die in den folgenden Abschnitten untersucht werden.

## 5.3. AD-HOC-INTEROPERABILITÄT VON AUSGABEGERÄTEN

### 5.3.1 Koordinationsansatz

Der am häufigsten verwendete Ansatz ist der Koordinationsansatz, der eine statische Umsetzung des Modells für Ausgabedienste (Abbildung 5.7) darstellt. Hierbei wird die Formulierung des Präsentationsproblems aus Abschnitt 5.2.3 verwendet (vgl. Abbildung 5.6).

Der Vorteil des Ansatzes ist, dass die Präsentationsstrategien über dem Modell statisch definiert werden können. Wenn das Modell z. B. eine Textgenerierung, eine Sprachausgabe, eine Charakter-Komponente und eine SMIL-Wiedergabekomponente unterstützt, so kann eine Strategie definiert werden, die bei SMIL-Wiedergabekomponenten den Charakter auf dem Gerät darstellt, auf dem die Eingabe erfolgt ist und die auszugebenden Inhalte auf einem anderen Gerät (vgl. Abbildung 5.3). Dies erlaubt eine detaillierte Koordination der Rendering-Komponenten  $R_i$  sowie die Definition einer konsistenten und vorhersagbaren Benutzerschnittstelle.

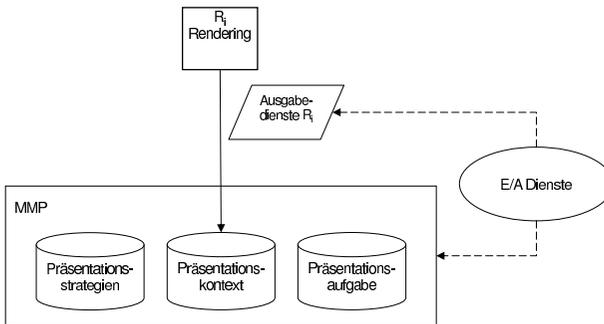


Abbildung 5.7: Statisches Modell für Ausgabedienste.

Der Nachteil dieses Ansatzes ist, dass neben den im Modell enthaltenen Ausgabediensten keine weitere Ausgabedienste unterstützt werden können. Das heißt, falls das Modell einen Ausgabedienst nicht beschreiben kann, so kann er nicht bei der Bearbeitung der Präsentationsaufgabe eingesetzt werden. Die Rendering-Komponente  $R_i$  steht in diesem Fall nicht als Ausgabekomponente zur Verfügung.

Dieses Problem stellt sich nicht, falls das Modell bzgl. aller Kategorien von Ausgabediensten vollständig und eindeutig ist. Dies würde jedoch insbesondere ein allgemeines Modell multi-modaler Ausgaben voraussetzen, das eindeutig und vollständig ist. Aufgrund der Komplexität eines solchen Modells ist es jedoch fraglich, ob sich in der Praxis unter Verwendung des Modells

effiziente Präsentationsstrategien definieren lassen [Bernsen, 2001].

Weiterhin stellt ein statisches Modell eine Herausforderung für den Designer des Modells und der Präsentationsstrategien dar, da sämtliche Abhängigkeiten zwischen den in Abbildung 5.4 dargestellten Faktoren in Betracht gezogen werden müssen. Dies ist aber für komplexe ubiquitäre Informationssysteme nur schwer möglich, denn typischerweise werden durch den Designer des Systems die Abhängigkeiten nicht vollständig erfasst oder umgesetzt [Heider & Kirste, 2005; Edwards & Grinter, 2001].

Da die Verwendung von statischen Modellen jedoch die Realisierung eines komplexen Dialogsystems wesentlich vereinfacht, ist der Koordinationsansatz am häufigsten anzutreffen. Das Gaia System verwendet eine statische Ontologie, um die interaktive Komposition und Ausgabe von Power Point-Präsentationen auf mehreren Bildschirmen anzubieten, falls diese verfügbar sind [Ranganathan et al., 2003]. Das System unterstützt verschiedene Typen von Ein- und Ausgabegeräten, z. B. Touchscreens, Mikrophone, Plasma Bildschirme oder Videowände.

Weiterhin wird die Präsentation von Power Point-Folien als Applikation modelliert. Es ist jedoch nicht möglich, Geräte oder Applikationen in das System einzubinden, die nicht durch diese Ontologie beschrieben werden. Diese können jedoch manuell in die Ontologie gepflegt werden. Die Ontologie modelliert jedoch keine multi-modalen Ausgabedienste.

Das Situated Computing System [Pham et al., 2000] verwendet eine statische Definition von Medientypen und Gerätefähigkeiten in einer Datenbank, um Benutzerschnittstellen und Ausgaben halbautomatisch auf Ausgabegeräte umzuleiten. So kann die HTML-Schnittstelle eines Email-Systems mittels eines PDAs interaktiv zu einem von mehreren verfügbaren PCs umgeleitet werden. Weiterhin kann ein beliebiger PC zur Darstellung von Videodaten verwendet werden. Die Auswahl der Geräte erfolgt automatisch durch das System. Sie kann jedoch durch den Benutzer manuell modifiziert werden. Aufgrund der nicht-hierarchischen Modellierung der Geräte durch wenige Gerätefähigkeiten ist das System jedoch im Unterschied zu Gaia nur schwer erweiterbar.

Im SmartWeb-Projekt wird im Bereich des Semantic Web eine Ontologie verwendet, um die automatische Integration von Webdiensten zu ermöglichen [Oberle et al., 2006]. Dazu werden Ausgabeinhalte auf Basis des MPEG 7-Standards [Martínez, 2004] modelliert, der audiovisuelle Inhalte beschreibt. Die Ontologie ist so aufgebaut, dass eine spätere Erweiterung für weitere Webdienste möglich ist. Ausgabedienste werden jedoch nicht modelliert.

Industriestandards zur Geräteinteroperabilität im Heimbereich wie UPnP [Universal Plug and Play, 2006] oder Havi [HAVi, 2001] stellen ebenfalls statische

### 5.3. AD-HOC-INTEROPERABILITÄT VON AUSGABEGERÄTEN

Modelle für Ein-/Ausgabegeräte zur Verfügung. So werden in UPnP Musikanlagen und Monitore unter dem Konzept „AV-Medien-Renderer“ unterstützt. Selbstbeschreibungen mittels des UPnP-Standards erlauben die konfigurationslose Integration der Komponenten in ein existierendes Netzwerk.

So lassen sich die Komponenten nach Anschluss automatisch durch einen Kontrollpunkt, z. B. einen PC ansteuern. Neben Gerätetypen werden jedoch keine weiteren Ein-/Ausgabefähigkeiten modelliert. So wird z. B. nicht zwischen unterschiedlichen Medientypen wie Filmen, Musikstücken oder Fotos unterschieden. Dies ist ebenso bei dem Java-basierten Havi-Standard der Fall.

#### 5.3.2 Heuristischer Ansatz

Neben dem statischen Koordinationsansatz ist auch eine dynamische Erweiterung des Dienstmodells möglich (Abbildung 5.8). Dies bedeutet, dass ein statisches, domänenunabhängiges Basismodell existiert, das durch jede Ausgabekomponente  $R_i$  domänenspezifisch erweitert werden kann. So könnte das Basismodell z. B. die uni-modalen Basisausgaben von multi-modalen Ausgaben beschreiben, für die Bernsen ein eindeutiges und vollständiges Modell entwickelt hat ([Bernsen, 2001], vgl. Abschnitt 4.3). Diese werden durch die Rendering-Komponente  $R_i$  zu einer Beschreibung der jeweiligen multi-modalen Ausgabe der Komponente (z. B. eines animierten Charakters) erweitert.

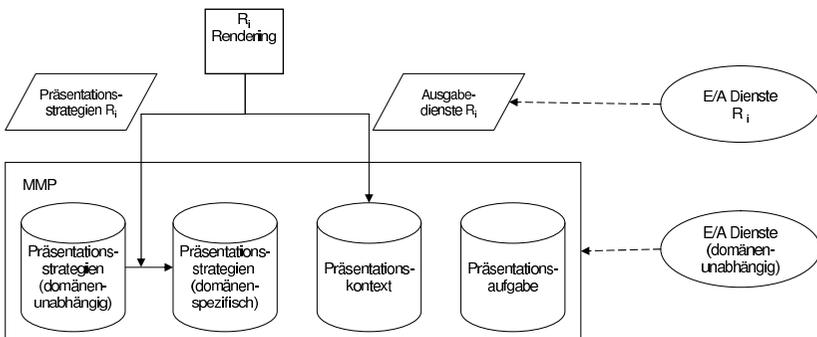


Abbildung 5.8: Dynamisches Modell für Ausgabedienste.

Dies bedeutet insbesondere, dass die Präsentationsstrategien nicht statisch definiert werden können, da die Beschreibung der Ausgabedienste von  $R_i$  nicht bekannt ist. Aus diesem Grund ist es notwendig, dass  $R_i$  die benötigten Prä-

sentationsstrategien liefert. Diese können mit den statischen, domänenunabhängigen Strategien zu einer neuen Menge von domänenspezifischen Strategien vereinigt werden.

Diese Vorgehensweise wird in [Heider & Kirste, 2005] verwendet, um automatisch mittels Lampe und Jalousien die Helligkeit eines Fernsehbildes zu optimieren. Dabei liefern die Lampe und die Jalousien eine Dienstbeschreibung in Form von Planungsoperatoren, die die Präsentationsstrategien definieren. Diese Planungsoperatoren verwenden das statisch definierte Konzept der Raumhelligkeit, um ihre Aktionen mittels eines gemeinsamen Vokabulars beschreiben zu können.

Dieses Modell ist jedoch für Ausgabekomponenten nicht praktikabel, da im Unterschied zu den bei Heider und Kirste verwendeten Diensten Ausgabedienste nicht nur Effekte auf globale Umgebungsvariablen sondern auch auf andere Ausgabedienste haben. So müsste z. B. der visuelle Teil eines animierten Charakters abstrakt beschreiben, wie die Animation der Lippenbewegungen mit dem Audiosignal einer Sprachausgabe synchronisiert wird.

Weiterhin können bei diesem Ansatz konkurrierende Präsentationsstrategien auftreten. So könnte z. B. ein GUI eine Präsentationsstrategie zur Integration eines animierten Charakters liefern. Ebenso könnte ein animierter Charakter eine Strategie zur Integration eines GUI liefern. Die MMP-Komponente müsste entscheiden, welche der beiden Strategien verwendet werden. Dies erfordert allerdings eine tiefgehende Konsistenzprüfung der Strategien.

Heider und Kirste verwenden aus diesem Grund für die intelligente Ansteuerung von Bildschirmen in einem Besprechungsraum einen numerischen Optimierungsansatz anstelle eines Planungsansatzes [Heider & Kirste, 2005], wobei die in Abschnitt 3.2.2 diskutierten Nachteile der numerischen Instabilitäten und der Modellierung der Kostenfunktion in Kauf genommen werden. Das Beispiel von Heider und Kirste zeigt, dass bei einer dynamische Erweiterung des Dienstmodells Heuristiken erforderlich sind. In Folge dessen ist jedoch nur ein reduziertes Maß an Koordination und an Kontrolle der Rendering-Komponenten  $R_i$  möglich.

Der Vorteil des heuristischen Ansatzes gegenüber dem Koordinationsansatz ist, dass theoretisch beliebige Rendering-Komponenten und beliebige Kombinationen von Rendering-Komponenten unterstützt werden können. Der Nachteil des Ansatzes ist jedoch, dass aufgrund der Komplexität multi-modaler Ausgaben in ubiquitären Informationssystemen Heuristiken verwendet werden müssen. Diese Heuristiken bergen die Gefahr die Verständlichkeit und die Konsistenz der multi-modalen Ausgaben zu reduzieren, die wesentliche Bestandteile von Benutzerschnittstellen sind [Norman, 2002].

### 5.3. AD-HOC-INTEROPERABILITÄT VON AUSGABEGERÄTEN

#### 5.3.3 Konfigurierbarer Ansatz

Eine Variante des heuristischen Ansatzes ist der konfigurierbare Ansatz. Dieser Ansatz liegt dem Speakeasy System zugrunde [Edwards et al., 2001]. In diesem System wird Interoperabilität dadurch realisiert, dass sich Komponenten auf Datentypen einigen, die gesucht bzw. benötigt werden (z. B. ein digitales Video-Eingabesignal).

Die Komponenten werden erstmalig durch den Benutzer konfiguriert, indem an einen Browser ein grafisches Installationsprogramm exportiert wird. Dadurch kann durch den Benutzer z. B. ein Drucker konfiguriert werden, ohne dass das System explizite Informationen über Drucker besitzt. Im Speakeasy System wurde ein Szenario realisiert, in dem ein Benutzer mit dem PDA als Fernsteuerung Power Point-Präsentationen auf einem Projektor anzeigen lassen kann.

Der Ansatz ist jedoch nur für einfache Komponentenkonfigurationen praktikabel. Falls Komponenten für viele Kontexte konfiguriert werden müssen, ist es umständlich für den Benutzer, eine Semantik für alle möglichen Zustände einer Komponente anzugeben. Dies ist insbesondere der Fall bei multi-modalen Ausgabediensten, die zusätzlich zum Benutzerkontext von Inhalt der Ausgabe, vom Dialogzustand und von den vorhandenen Ausgabegeräten abhängen. Es wäre sehr umständlich für den Benutzer, Strategien für jede dieser Situationen einzugeben.

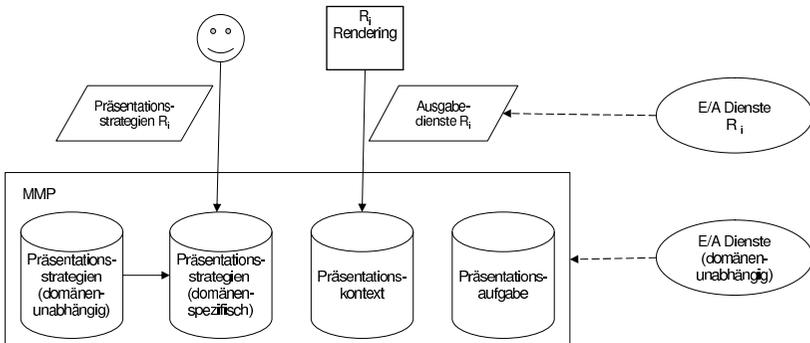


Abbildung 5.9: Konfigurierbares Modell für Ausgabedienste.

Eine Variante dieses Ansatzes verwenden Systeme, die den Benutzer beobachten und daraus das Wissen über die Verwendung von Diensten automa-

tisch ableiten. Dieser Ansatz wird in im MavHome-Projekt [Youngblood et al., 2005] dazu verwendet, um das Licht, die Raumtemperatur, die Luftfeuchtigkeit und einen Ventilator zu kontrollieren. Einen ähnlichen Lernansatz im Bereich Haussteuerung verwendet das Adaptive House-Projekt [Mozer, 2004]. Um die zeitlichen und grafischen Layouts der Ausgabekomponenten zu lernen, müsste der Benutzer jedoch bei jeder Ausgabe das gewünschte grafische oder zeitliche Layout selbst definieren.

Aufgrund der Abhängigkeiten bei der Präsentationsgenerierung (vgl. Abbildung 5.4) wird in Zusammenhang mit dem „Fluch der Hochdimensionalität“ [Bellman, 1961] jedoch eine sehr große Anzahl an Daten benötigt, um ein effektives Lernen zu ermöglichen. Dies bedeutet, dass ein Benutzer bei der Integration einer neuen Ausgabekomponente das Layout sehr oft manuell anpassen muss, bis das System über genügend Daten verfügt, um verlässliche Schlüsse ziehen zu können.

Dies erfordert bei der Definition eines komplexen grafisch-zeitlichen Layouts, das mehrere Geräte umfasst, jedoch umfangreiche Interaktionen. Aus diesen Gründen ist der konfigurierbare Ansatz ebenfalls nicht praktikabel für die Modellierung von Ausgabediensten.

### 5.3.4 Zusammenfassung und Ansatz

Das Ziel dieser Arbeit ist eine Modellierung von Ausgabediensten die eine enge Kopplung von Ausgabegeräten unterstützt. Aus diesem Grund wird der Koordinationsansatz dem heuristischen Ansatz und dem konfigurierbaren Ansatz vorgezogen, die für eine detaillierte Ansteuerung von Rendering-Komponenten nicht geeignet sind. Das Ziel der Modellierung ist es dabei, in dem Modell zwischen domänenunabhängigen und domänenspezifischen Aspekten zu trennen. Der domänenspezifische Teil stellt eine spezifische Erweiterung für spezielle Ausgabedienste dar, z. B. ein GUI zur Auswahl von Fernsehprogrammen.

Der domänenunabhängige Teil des Dienstmodells kann dazu verwendet werden, um eine statische Menge von Präsentationsstrategien zu definieren, die sich auf alle Ausgabekomponenten anwenden lassen. Eine domänenunabhängige Eigenschaft von Ausgabediensten ist z. B. das Medium der Ausgabe, das sich durch die Unterscheidung zwischen visuellen, auditoriellen und taktilen Ausgaben modellieren lässt.

Im Unterschied zum heuristischen Ansatz (Abschnitt 5.3.2) erfolgt die Erweiterung des domänenunabhängigen Modells mit domänenabhängigen Konzepten jedoch nicht dynamisch zur Laufzeit (durch neue, komponentenspezifische Präsentationsstrategien), sondern zur Modellierungszeit der Präsentationsstrategien in der MMP-Komponente. Die so definierte Menge an Präsentations-

## 5.4. MODELLIERUNG MULTI-MODALER EIN-/AUSGABEKOMPONENTEN

strategien ist beim hier gewählten Ansatz im Unterschied zum heuristischen Ansatz statisch. Die Trennung zwischen domänenunabhängigen und domänenabhängigen Konzepten vereinfacht jedoch die Anpassung der Präsentationsstrategien an eine neue Domäne und unterstützt die Portabilität der Software.

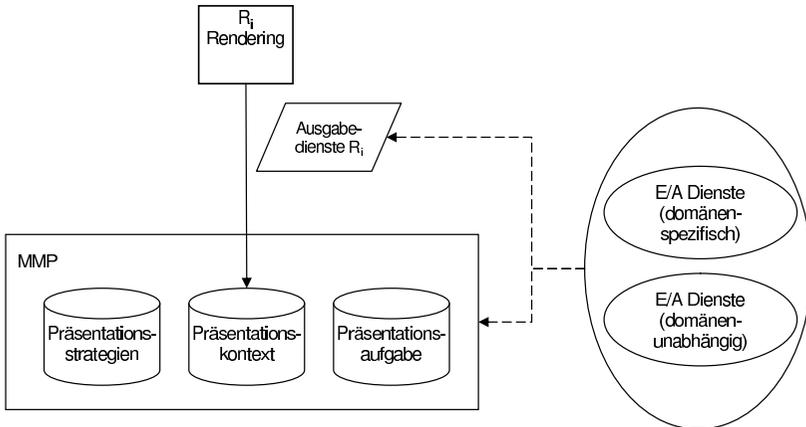


Abbildung 5.10: Statisches Modell für Ausgabedienste mit Trennung nach domänenunabhängigen Eigenschaften.

Der Vorteil des Ansatzes ist somit, dass sich auf der Basis des Modells eine statische Menge von Strategien definieren lässt. Müssen neue Ausgabedienste unterstützt werden, die nicht durch das Modell erfasst werden, so kann in Analogie zu den in SmartWeb [Oberle et al., 2006] oder Gaia [Ranganathan et al., 2003] definierten Ontologien das Modell leicht durch domänenspezifischen Konzepten erweitert werden. Den Präsentationsstrategien können dann neue Strategien, die die Eigenschaften des erweiterten Modells berücksichtigen, hinzugefügt werden.

### 5.4 Modellierung multi-modaler Ein-/Ausgabekomponenten

Den Modellen für Ein-/Ausgabedienste, Ein-/Ausgabegeräte und Ein-/Ausgabebereitschaften kommt bei der Generierung multi-modaler Präsentationen eine besondere Bedeutung zu (Abbildung 5.4). Aus dem Beispiel in Abschnitt 5.1 gehen verschiedene Anforderungen an diese Modelle hervor. Das Präsentationssystem lokalisiert den Benutzer in der Küche und reagiert darauf, in dem das System eine auditorielle Ausgabe anstelle einer visuellen Ausgabe verwendet. Dies erfordert insbesondere, dass die Orte von Ausgabegeräten mo-

delliert werden. Zusätzlich sind die Größe und die Orientierung der Bildschirme relevant, weil sich daraus durch das Präsentationssystem Rückschlüsse auf die Sichtbarkeit von Bildschirmen ziehen lassen (vgl. [Heider & Kirste, 2005]).

Das Präsentationssystem im Beispiel unterscheidet zwischen auditoriellen und visuellen Ausgabediensten. Dies erfordert die Modellierung von Ausgabemedien. Da der elektronische Bilderrahmen nur Standbilder darstellen kann, wird der animierte Charakter auf dem Fernseher dargestellt. Dies bedeutet, dass bei Ausgabediensten die Inhalte modelliert werden müssen, die eine Ausgabekomponente verarbeiten kann. Dies können z. B. JPG-Bilder, SMIL-Texte oder Ontologie-Konzepte sein.

Auf dem PDA wird die Präsentation an die Auflösung sowie die Grafik- und Netzwerkfähigkeiten angepasst, die ebenfalls in dem Modell für Ausgabegeräte erfasst werden müssen. Weiterhin unterstützt das System mehrere animierte Charaktere. Insbesondere passt das Präsentationssystem die Ausgabe an den animierten Charakter von Melanie an, da dieser über keine Gestik verfügt. Aus diesem Grund ist es notwendig, nicht nur multi-modale Ausgabedienste zu modellieren, sondern auch deren uni-modale Basisbestandteile.

### 5.4.1 Beispiele

Im Folgenden werden zwei Beispiele für die Modellierung von Ein-/Ausgabediensten gegeben. Als Ausgabedienst wird hierzu ein animierter Charakter modelliert. Als Eingabedienst werden eine Spracherkennung sowie visuelle Eingaben per Maus exemplarisch modelliert.

Beispiel: Modellierung von Ausgabediensten

Im folgenden Beispiel wird ein animierter Charakter modelliert, der aus einem Kopf und einer Hand besteht, die verschiedene Gesten ausführen kann. Anschließend wird beschrieben, wie sich weitere animierte Charaktere mit anderen Funktionalitäten auf ähnliche Weise durch das Modell beschreiben lassen.

Abbildung 5.11 zeigt die multi-modale Ausgabe des animierten Charakters aus [Rieger & Berner, 2002]. Die multi-modale Ausgabe des Charakters im Beispiel lässt sich mit Hilfe von sechs Unimodalitäten beschreiben: Handgesten, Lippenbewegungen, gesprochene Sprache, Mimik, Blickrichtung und Kopfbewegungen. Jeder dieser Unimodalitäten lassen sich Parameter zuordnen, die für eine Präsentation gesetzt werden müssen. So muss für Ausgaben mittels gesprochener Sprache die Lautstärke der Sprachausgabe gesetzt werden (Parameter *volume*). Für die Integration des animierten Charakters in die Gesamtpräsentation ist es ebenfalls notwendig, Informationen über den aktu-

## 5.4. MODELLIERUNG MULTI-MODALER EIN-/AUSGABEKOMPONENTEN

ellen Status des Charakters zu besitzen. Im Beispiel wird von der Sprachausgabe der Status *is-playing* unterstützt, der angibt, ob die letzte Sprachausgabe noch aktiv ist. Um die Ressourcen-Beschränkungen des Ausgabegerätes in der Präsentationsplanung berücksichtigen zu können, wird ein eindeutiger Bezeichner des Ausgabegerätes verwendet. Im Beispiel sind dies der Bildschirm (*tv1.display*) und die Lautsprecher (*tv1.speakers*) eines Fernsehers. Über diesen Bezeichner lassen sich die Eigenschaften des Ausgabegerätes ermitteln wie z. B. die Position in einem Raum, die aktuelle Auflösung oder die maximale Lautstärke.

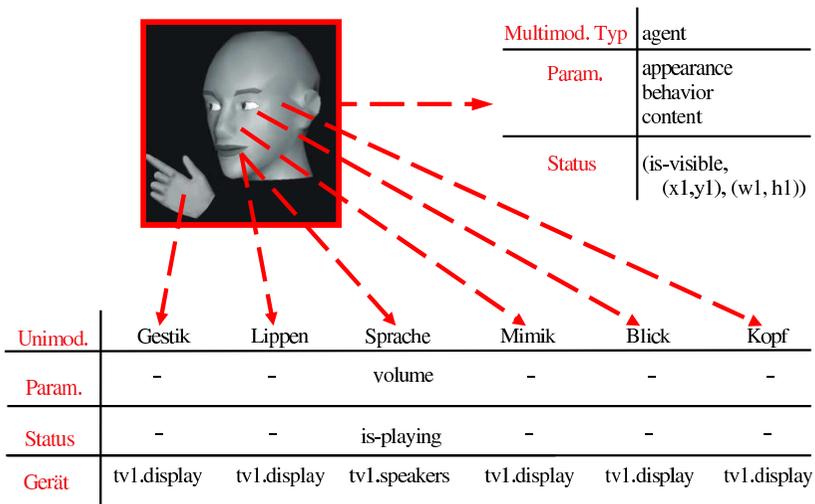


Abbildung 5.11: Modellierung eines animierten Charakters (Bild Charakter: TU Darmstadt).

Um die multi-modale Ausgabe einer Ausgabekomponente als Ganzes charakterisieren zu können, wird der animierte Charakter durch ein weiteres Attribut, den multi-modalen Typ *agent* charakterisiert. Dieser dient zusätzlich der Beschreibung der Basisausgaben durch die Unimodalitäten dazu, die semantische Funktion der Komponente zu beschreiben. So umfasst der Typ *agent* verschiedene Typen von Charakteren, die mit dem Benutzer kommunizieren können. Der Typ *agent* umfasst dabei rein auditive Sprachausgabekomponenten, rein visuelle Charaktere, Charaktere mit und ohne Zeigegesten sowie mit und ohne Modellierung von Körperbewegungen. Aus dem multi-modalen

Typen gehen die multi-modalen Parameter der Ausgabekomponente hervor. Im Gegensatz zu den Parametern der Unimodalitäten konfigurieren die multi-modalen Parameter die gesamte multi-modale Ausgabe. Im Falle des animierten Charakters kann das äußere Erscheinungsbild (Parameter *appearance*) durch einen multi-modalen Parameter konfiguriert werden. Dies erlaubt z. B. die Konfiguration eines männlichen oder weiblichen Erscheinungsbildes. Diese Konfiguration hat Auswirkungen auf die Umsetzung sämtlicher Unimodalitäten und kann somit nicht als uni-modaler Parameter modelliert werden.

Ein weiterer multi-modaler Parameter für Komponenten vom Typ *agent* ist der Parameter *behaviour*, der die abstrakte Verhaltensweise spezifiziert, die der Charakter ausführen soll. Dies kann z. B. die Begrüßung des Benutzers sein, die durch ein Winken mit der Hand oder durch eine kurze sprachliche Begrüßung umgesetzt wird. Dieser multi-modale Parameter hat somit ebenfalls Einfluss auf mehrere Unimodalitäten. Der multi-modale Parameter *content* enthält den Inhalt, den der Charakter ausgeben soll. Dieser kann z. B. direkt als verbalisierter Text im ASCII-Format übergeben werden oder auch als Formulierung amodaler Konzepte aus einer Ontologie.

Neben dem uni-modalen Status *is-playing* der Sprachausgabe kann auch der Status der multi-modalen Ausgabe als Ganzes modelliert werden. Dies ist im Fall von visuellen Ausgabekomponenten des Typs *agent* sinnvoll in Bezug auf die Sichtbarkeit auf der Desktop-Oberfläche (*is-visible*), der Koordinaten des Fensters der Ausgabe auf dem Bildschirm ( $(x1, y1)$ ), sowie der Breite und Höhe des Fensters auf dem Bildschirm ( $w1, h1$ ). Diese Informationen könnten ebenfalls durch die visuellen Unimodalitäten bereitgestellt werden. Jedoch hat es für die Layout-Koordination umständlich, aus der Fenstergröße jeder visuellen Unimodalität die Fenstergröße der gesamten Ausgabekomponente zu berechnen.

Die Modellierung der uni-modalen Basisausgaben von animierten Charakteren ermöglicht es, ähnliche Komponenten mit unterschiedlicher Funktionalität durch dasselbe Modell zu beschreiben. So können z. B. Charaktere ohne Zeigegestik auf ähnliche Weise beschrieben werden, indem bei der Selbstbeschreibung auf die Unimodalität Gestik verzichtet wird. Umgekehrt können animierte Charaktere, die zusätzlich über Körperbewegungen verfügen, modelliert werden, indem eine Unimodalität Körperhaltung hinzugefügt wird.

Eine Komponente zur Sprachausgabe kann ebenfalls als spezielle Instanz eines animierten Charakters modelliert werden, die über keine visuellen Ausgabemöglichkeiten verfügt (hier Gestik, Lippenbewegungen und Blickrichtungsanimation). Dieses Modell ermöglicht auf einfache Weise die Definition von Selbstbeschreibungen animierter Charaktere. Dies ist die Grundlage für

## 5.4. MODELLIERUNG MULTI-MODALER EIN-/AUSGABEKOMPONENTEN

die automatische Integration dieser Komponenten in die Präsentationsplanung eines ubiquitären Informationssystems.

Beispiel: Modellierung von Eingabediensten

Da der Schwerpunkt dieser Arbeit auf der Generierung von Präsentationen liegt, werden Eingabekomponenten im Folgenden nur soweit beschrieben, wie es die Präsentationsgenerierung erfordert. Grundlage des in dieser Arbeit verwendeten Modells für Eingabedienste ist Bernsens Modell für Eingabeunimodalitäten [Bernsen, 2001]. Das Modell ähnelt dem in Abschnitt 4.3 beschriebenen Modell für Ausgabeunimodalitäten und basiert auf den Modalitäteneigenschaften (nicht) linguistisch, (nicht) analog, (nicht) arbiträr, statisch/dynamisch und dem Eingabemedium (visuell, auditoriell oder taktil). Im Unterschied zur Taxonomie für Ausgabeunimodalitäten wird die Eigenschaft statisch/dynamisch nicht betrachtet, da sie für Eingabekomponenten keine Aussagekraft hat.

Somit ist es möglich, die Eingabe mittels einer Spracherkennungskomponente als linguistische, nicht-analoge, nicht-arbiträre, auditorielle Eingabemodalität zu beschreiben. Eine grafische Mauseingabe lässt sich als nicht-linguistische, nicht-analoge, nicht-arbiträre, visuelle Eingabemodalität modellieren. Analog zu den Ausgabegeräten werden den Eingabemodalitäten auch Eingabegeräte zugeordnet. Der Grund hierfür ist, dass der physikalische Standort eines Eingabegerätes Auswirkung auf die Umsetzung einer multi-modalen Ausgabe hat. So wäre es z. B. wünschenswert, dass bei Einsatz eines grafischen Auswahl Fensters als Eingabemöglichkeit das Ergebnis auf demselben Bildschirm dargestellt wird, auf dem auch die Eingabe erfolgt ist.

Darüber hinaus kann es wünschenswert sein, auf einem Display die visuellen Ausgaben mit den visuellen Eingaben in einem einzigen Layout anzuordnen. So könnte z. B. die Ausgabe eines Ergebnisses in einem Fenster rechts neben dem Eingabefenster erfolgen, durch das die Berechnung angefordert wurde. Aus diesem Grund ist es notwendig, Informationen über die Sichtbarkeit, die Bildschirmkoordinaten und die Größe von visuellen Eingabekomponenten zu haben.

### 5.4.2 Vorteile des Ansatzes

Um die Vorteile des in Abschnitt 5.4.1 dargestellten Ansatzes zu verdeutlichen, wird der im Beispiel gewählte Ansatz mit alternativen Modellierungen verglichen. Die einfachste Modellierung von Rendering-Komponenten  $R_i$  stellt die Modellierung jeder Komponente durch einen eindeutigen Bezeichner dar (Abbildung 5.12, links). Der MMP-Komponente sind bei diesem Ansatz lediglich die Bezeichner der zurzeit aktiven Rendering-Komponenten bekannt

(z. B. „render1“). Die Typen oder Ausgabedienste der Komponenten sind der MMP-Komponente nicht bekannt. Weitergehendes Präsentationswissen (z. B., dass die Komponente „render1“ eine Sprachausgabe realisiert) wird den Bezeichnern implizit in der MMP-Komponente zugeordnet.

Der Vorteil dieses Ansatzes ist, dass er auf der Seite der Rendering-Komponenten  $R_i$  keine Modellierung erfordert. Da in der MMP-Komponente die Menge der potentiell vorhandenen Rendering-Komponenten in Form der unterschiedlichen Bezeichner bekannt ist, können die Kommunikationsstrukturen zwischen den Komponenten ebenfalls festgeschrieben werden (z. B. dass die Komponente „render1“ eine Audiodatei liefert, die anschließend an die Rendering-Komponente „render2“ übergeben wird), was die Definition von Präsentationsstrategien erleichtert.

Der Nachteil des Ansatzes ist, dass er die Anzahl der unterstützten Rendering-Komponenten durch die Menge der bekannten Bezeichner festschreibt. Wenn in der MMP-Komponente die Bezeichner „render1“ bis „render10“ bekannt sind, so ist es nicht möglich, eine neue Rendering-Komponente mit dem Bezeichner „render11“ automatisch durch die MMP-Komponente zu unterstützen, da nicht bekannt ist, um was für einen Typ Rendering-Komponente es sich handelt.

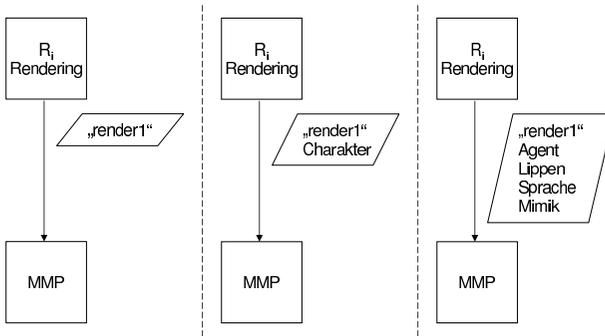


Abbildung 5.12: Drei Ansätze zur Modellierung der Ausgabedienste einer Rendering-Komponente  $R_i$ .

Diese Form der Modellierung bietet sich in abgeschlossenen Umgebungen mit einer eingeschränkten Menge von Rendering-Komponenten an. Aus diesem Grund kommt sie insbesondere bei klassischen multi-modalen Dialogsystemen, die auf ein Gerät beschränkt sind, zum Einsatz [Townsend et al., 1998; Kerpedjiev et al., 1997; André et al., 1993]. Ubiquitäre Informationssysteme

#### 5.4. MODELLIERUNG MULTI-MODALER EIN-/AUSGABEKOMPONENTEN

me verfügen jedoch über eine große Anzahl an Ausgabekomponenten. Mittels mobilen Geräten können weiterhin –wie in Abschnitt 5.1 illustriert– neue Rendering-Komponenten mit unbekanntem Bezeichnern eingeführt werden. Aus diesem Grund ist dieser Ansatz für ubiquitäre Informationssysteme nicht geeignet.

Eine Erweiterung dieses Ansatzes stellt die Modellierung von multi-modalen Typen dar, die die Rendering-Komponente  $R_i$  als Ganzes charakterisieren (Abbildung 5.12, Mitte). So kann z. B. der Typ „Charakter“ verwendet werden, um animierte Charaktere zu modellieren. Der Vorteil dieses Ansatzes ist, dass Rendering-Komponenten unabhängig von ihren Bezeichnern modelliert werden können. Dies ermöglicht in der MMP-Komponente die Definition von Präsentationsstrategien, die unabhängig vom Bezeichner sind und erlaubt die Unterstützung beliebig vieler Rendering-Komponenten der modellierten multi-modalen Typen.

In ubiquitären Informationssystemen trägt dieser Ansatz zur Ausfallsicherheit bei, da z. B. bei einem Ausfall der Komponente „render1“ vom Typ „Charakter“ eine andere Komponente vom Typ „Charakter“ an deren Stelle eingesetzt werden kann. Weiterhin wird durch die detailliertere Modellierung der Rendering-Komponenten das in der MMP-Komponente vorhandene Präsentationswissen im Bezug auf die multi-modalen Typen gekapselt. Somit trägt dieser Ansatz zur Verbesserung der Wissensmodellierung in der MMP-Komponente bei.

Ein Nachteil dieses Ansatzes ist, dass bei der Verwendung ähnlicher Typen von Rendering-Komponenten unterschiedliche Typbezeichner definiert werden müssen. Sollen z. B. neben animierten Charakteren, die über Mimik verfügen, auch animierte Charaktere unterstützt werden, die verschiedene Körperhaltungen unterstützen, so muss ein weiterer Charaktertyp definiert werden.

Dieser Nachteil wird durch den dritten Ansatz (Abbildung 5.12, rechts) behoben. Dieser Ansatz bildet die Grundlage des in Abschnitt 5.4.1 verwendeten Modells. Neben dem multi-modalen Typ werden die Basisausgaben der Rendering-Komponente beschrieben. Dies kann z. B. durch das Unimodalitätenmodell von Bernsen erfolgen (Abschnitt 4.3). Unterschiedliche Typen von animierten Charakteren können so durch denselben multi-modalen Typen modelliert werden. Dies erlaubt ebenfalls eine breitere Definition von multi-modalen Typen. So umfasst in Abschnitt 5.4.1 der multi-modale Typ *agent* nicht nur animierte Charaktere sondern auch Sprachausgaben ohne visuelle Ausgaben.

Der Vorteil des in Abschnitt 5.4.1 beschriebenen Ansatzes besteht somit in einer expliziten Modellierung von Unimodalitäten, die für ubiquitäre Informa-

tionssysteme mit zahlreichen Rendering-Komponenten unerlässlich ist. Die Modellierung erlaubt einerseits ein verbessertes Wissens-Management in der MMP-Komponente, andererseits lässt sich mit dem Ansatz die Ausfallsicherheit von verteilten Systemen verbessern. Die Modellierung der Basisausgaben von Rendering-Komponenten erlaubt es weiterhin, eine große Anzahl unterschiedlicher Varianten desselben multi-modalen Typs zu unterstützen. In den folgenden Abschnitten 5.4.3 und 5.4.4 wird der dritte Ansatz zur Modellierung einer Home-Entertainment Umgebung umgesetzt. In Abschnitt 5.4.5 werden Beispiele für die Modellierung von unterschiedlichen Typen von Rendering-Komponenten gegeben.

### 5.4.3 Domänenunabhängiges Modell

Das Ziel dieses Abschnitts ist die Definition des domänenunabhängigen Modells für Ein-/Ausgabedienste (Abbildung 5.10). Dieses Modell wird ergänzt um Modelle für Ein-/Ausgabegeräte sowie Ein-/Ausgaberesourcen (vgl. Abbildung 5.4). Diese drei Modelle werden unter dem Oberbegriff *domänenunabhängiges Modell für Ein-/Ausgabekomponenten* zusammengefasst. Im folgenden Abschnitt wird dieses Modell dann zu einem domänenspezifischen Modell für eine Home-Entertainment-Umgebung erweitert.

#### Ausgabekomponenten

Die Basis des domänenunabhängigen Modells ist die generische Ebene des Unimodalitätenmodells von Bernsen (Abschnitt 4.3). Bernsen zeigt, dass diese Klassen vollständig und eindeutig sind bzgl. der Beschreibung uni-modaler Ausgaben. Dabei wurden durch Bernsen Unimodalitäten, die für die Erforschung der Mensch-Maschine-Interaktion nicht relevant sind, vernachlässigt. Ein Beispiel hierfür sind Unimodalitäten, die sowohl linguistisch als auch arbiträr sind, d. h. Unimodalitäten, die Sprachen ohne Semantiken beschreiben. Bernsen fügt hinzu, dass diese Klassen noch zu allgemein sind, um im Interaktions-Design praktisch angewendet werden zu können. Aus diesem Grund erweitert Bernsen die generische Ebene exemplarisch in eine atomare und eine sub-atomare Ebene. In diesen Ebenen werden die generischen Unimodalitätenklassen in weitere Teilklassen aufgeteilt, die eine detaillierter Beschreibung von uni-modalen Ausgaben ermöglichen.

Diese Erweiterung ist jedoch nicht für alle Domänen sinnvoll. So würden z. B. alle sechs Unimodalitäten des animierten Charakters aus Abbildung 5.11 unter dem Oberbegriff visuelle Animationen modelliert, wodurch die Selbstbeschreibung von animierten Charakteren wenig Aussagekraft besitzt. Aus diesem Grund beschränken wir uns für das domänenunabhängige Modell auf die generische Ebene des Modells von Bernsen (Tabelle 4.1) und erweitern

## 5.4. MODELLIERUNG MULTI-MODALER EIN-/AUSGABEKOMPONENTEN

diese später im domänen-spezifischen Modell. Das domänen-unabhängige Modell unterscheidet dabei zwischen visuellen und auditoriellen Ausgabemedien. Weitere Ausgabemedien wie z. B. Haptik werden in dieser Arbeit nicht betrachtet.

Tabelle 5.1: Domänen-unabhängiges Modell für Parameter und Status von Ausgabekomponenten.

| Uni-modalität | Parameter                                                                                                                  | Status                                                                                                                             |
|---------------|----------------------------------------------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------|
| Visuell       | $visibility \in \{0, 1\}$<br>$screen-location \in \mathbb{N} \times \mathbb{N}$<br>$size \in \mathbb{N} \times \mathbb{N}$ | $is-visible \in \{0, 1\}$<br>$has-screen-location \in \mathbb{N} \times \mathbb{N}$<br>$has-size \in \mathbb{N} \times \mathbb{N}$ |
| Auditoriell   | $application-volume \in \mathbb{N}$                                                                                        | $has-application-volume \in \{0, 1\}$                                                                                              |
| Dynamisch     | –                                                                                                                          | $is-playing \in \{0, 1\}$                                                                                                          |
| Ausgabe-gerät | Parameter                                                                                                                  | Status                                                                                                                             |
| Visuell       | –                                                                                                                          | $has-resolution \in \mathbb{N} \times \mathbb{N}$<br>$has-graphics-benchmark \in \mathbb{N}$                                       |
| Auditoriell   | –                                                                                                                          | $has-device-volume \in \mathbb{N}$<br>$has-max-device-volume \in \mathbb{N}$<br>$has-min-device-volume \in \mathbb{N}$             |
| Alle          | –                                                                                                                          | $has-id \in \text{String}$<br>$has-networking-benchmark \in \mathbb{N}$                                                            |

Den visuellen und auditoriellen Ausgabe-unimodalitäten lassen sich Mengen von domänen-unabhängigen Parametern zuordnen. In Tabelle 5.1 wird zwischen visuellen, auditoriellen und dynamischen Unimodalitäten getrennt. Neben Parametern und Status für Unimodalitäten lassen sich auch für Ausgabegeräte domänen-unabhängige Definitionen vornehmen. Dies dient der Modellierung der Ressourcenbeschränkungen von Ausgabegeräten.

Diese Beschränkungen müssen bei der Generierung multi-modaler Präsentationen von der MMP-Komponente beachtet werden. Hierbei werden Ausgabegeräte, die visuelle Ausgabe-unimodalitäten unterstützen (Bildschirme), von Ausgabegeräten unterschieden, die auditorielle Ausgabe-unimodalitäten unterstützen (Lautsprecher). Eigenschaften, die alle Typen von Ausgabegeräten betreffen, werden in einer dritten Kategorie betrachtet. Für visuelle Unimodalitäten werden Parameter definiert, die die Sichtbarkeit (*visibility*), die Bildschirmkoordinaten (*screen-location*) und die Größe (*size*) bei einer Ausgabe

bestimmen. Für alle auditoriellen Unimodalitäten wird ein Parameter Lautstärke (*application-volume*) definiert, mit dem sich die Lautstärke der Unimodalität konfigurieren lässt (vgl. Parameter *volume* in Abbildung 5.11). Diese Parameter des domänenunabhängigen Modells werden in dem domänenspezifischen Modell durch weitere Parameterdefinitionen ergänzt.

Das domänenunabhängige Modell enthält weiterhin eine allgemeine Menge von domänenunabhängigen Status, die jeder Ausgabekomponente entsprechend ihrer Unimodalitäten zugeordnet werden. Ausgabekomponenten mit visuellen Unimodalitäten haben die Eigenschaften Sichtbarkeit (*is-visible*), Koordinaten (*has-screen-location*) und Größe (*has-size*). Dabei sind Koordinaten und Größe nur definiert, falls die Sichtbarkeit gegeben ist. Diese Status werden durch die korrespondierenden Parameter *visibility*, *screen-location* und *size* beeinflusst. Ausgabekomponenten mit auditoriellen Unimodalitäten verfügen über einen Status *has-application-volume*, der die durch den Parameter *application-volume* beeinflusste software-seitige Lautstärkeinstellung wiedergibt.

Für Ausgabekomponenten mit dynamischen Unimodalitäten wird der Status *is-playing* modelliert. Dieses Attribut wird gesetzt, falls zurzeit eine dynamische Ausgabe visuell oder auditoriell erfolgt. Diese domänenunabhängigen Parameter und Status werden bei der Definition des domänenspezifischen Modells ergänzt durch domänenspezifische Erweiterungen.

Visuelle Ausgabegeräte haben die Eigenschaft Auflösung (*has-resolution*). Ausgabegeräte verfügen über einen Grafikkennwert (*has-graphics-benchmark*). Grafikkennwerte können dabei auf jedem Gerät durch standardisierte Testsoftware, wie sie z. B. von der SPEC<sup>3</sup> zur Verfügung gestellt wird, ermittelt werden. Kennwerte dienen bei der Präsentationsgenerierung dazu, die Qualität von Ausgabemedien an die Ressourcen eines Ausgabegerätes anzupassen (vgl. Beispiel in Abschnitt 5.1).

Für auditorielle Ausgabegeräte wird der Status Lautstärke modelliert (*has-device-volume*). Während der Parameter *application-volume* von auditoriellen Ausgabekomponenten durch die Applikation selber gesetzt wird (z. B. einer Komponente zum Abspielen von MPEG-Filmen), modelliert der Parameter *device-volume* die physikalische Einstellung der Lautstärke durch das unterliegende Betriebssystem. Zusätzlich werden Status für die maximal bzw. minimal mögliche Lautstärkeinstellung gegeben (*max-device-volume* und *mindevice-volume*), die analog zu *has-resolution* die Ressourcenbeschränkungen des auditoriellen Gerätes wiedergeben. Der Status *has-id* liefert für jedes Ausgabegerät einen eindeutigen Bezeichner in Form einer Zeichenkette

---

<sup>3</sup>Standard Performance Evaluation Corporation, <http://www.spec.org>.

#### 5.4. MODELLIERUNG MULTI-MODALER EIN-/AUSGABEKOMPONENTEN

(String). Weiterhin werden neben den Grafikkennwerten ebenfalls die Netzwerkennwerte für alle Ausgabegeräte festgehalten (*has-networking-benchmark*). Dies ermöglicht die Anpassung der Präsentationen an die Netzwerkfähigkeiten der Ausgabegeräte. Wie im Beispiel in Abschnitt 5.1 anhand von Martins PDA erläutert wurde, können die Medienobjekte der Präsentation im Falle geringer Netzwerkennwerte in ihrer Qualität reduziert werden.

Der physikalische Ort (z. B. das Zimmer, in dem sich ein Ausgabegerät befindet) hat ebenfalls Einfluss auf die Präsentationsgenerierung in einem ubiquitären Informationssystem (vgl. Abschnitt 5.1). Der Detailgrad der Modellierung wird allerdings durch die Domäne der Anwendung bestimmt. So lässt sich z. B. der Geräteort durch die dreidimensionalen Koordinaten, die Orientierung und den Wahrnehmungsbereich des Geräts modellieren. Alternativ dazu kann der Geräteort durch eine Abbildung auf eine zweidimensionalen Ebene [Kray et al., 2003] oder lediglich durch Angabe des Zimmers, in dem sich das Ausgabegerät befindet, modelliert werden.

In einem Home-Entertainment-System kann der Geräteort durch die zweidimensionale Position innerhalb eines Zimmers repräsentiert werden [Brumitt et al., 2000]. In einem elektronischen Touristenführer können zweidimensionale Geokoordinaten eingesetzt werden [Zipf & Malaka, 2000]. Aus diesem Grund wird der physikalische Ort eines Ausgabegerätes nicht im domänenunabhängigen Modell, sondern im domänenspezifischen Modell definiert.

Eine Rendering-Komponente zur Ausgabe statischer Bilder wird somit als visuelle Unimodalität durch die Parameter *visibility*, *screen-location* und *size* modelliert. Weiterhin werden ihr die Status *is-visible*, *has-screen-location* und *has-size* zugeordnet. Die Ressourcenbeschränkungen des PDA-Bildschirms werden durch die Gerätestatus *has-resolution* und *has-graphics-benchmark* modelliert. Wie jedes Ausgabegerät verfügt der PDA über einen eindeutigen Bezeichner (Status *has-id*) sowie über einen Netzwerkennwert (Status *has-networking-benchmark*).

Neben den im Modell festgehaltenen domänenunabhängigen Eigenschaften sind je nach benötigtem Detailgrad und beabsichtigter Verwendung auch Erweiterungen möglich. So lässt sich z. B. bei visuellen Ausgaben der Rand der Grafik als Polygon festhalten. Weiterhin lassen sich die Auflösung, der Kontrast oder die Farben der Grafik festhalten. Bei dynamischen visuellen Ausgaben lässt sich die Bewegung der Animation sowie deren zeitlicher Verlauf modellieren [Bernsen, 2001]. Da die Verwendung dieser Eigenschaften jedoch nicht in jeder Domäne gegeben ist, werden diese Eigenschaften nicht im domänenunabhängigen Modell definiert.

Eingabekomponenten

Wie im Abschnitt 5.11 erläutert wurde, wird neben einem Modell für multi-modale Ausgabedienste und -geräte ebenfalls ein Modell für multi-modale Eingabedienste und -geräte benötigt. Die Definition des domänenunabhängigen Modells für Eingabekomponenten erfolgt analog zu den Ausgabekomponenten und wird in Tabelle 5.2 zusammengefasst.

Tabelle 5.2: Domänenunabhängiges Modell für Parameter und Status von Eingabekomponenten.

| Unimodalität | Parameter                                                                                                                  | Status                                                                                                                             |
|--------------|----------------------------------------------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------|
| Visuell      | $visibility \in \{0, 1\}$<br>$screen-location \in \mathbb{N} \times \mathbb{N}$<br>$size \in \mathbb{N} \times \mathbb{N}$ | $is-visible \in \{0, 1\}$<br>$has-screen-location \in \mathbb{N} \times \mathbb{N}$<br>$has-size \in \mathbb{N} \times \mathbb{N}$ |
| Alle         | –                                                                                                                          | $is-processing-input \in \{0, 1\}$<br>$threw-input-event \in \{0, 1\}$                                                             |
| Eingabegerät | Parameter                                                                                                                  | Status                                                                                                                             |
| Visuell      | –                                                                                                                          | $has-resolution \in \mathbb{N} \times \mathbb{N}$                                                                                  |
| Alle         | –                                                                                                                          | $has-id \in \text{String}$                                                                                                         |

Dabei basiert das Unimodalitätenmodell auf Bernsens Modell für Eingabemodalitäten [Bernsen, 2001]. Als domänenunabhängiger Status werden Informationen über Eingabe-Events verwaltet. Der Status *threw-input-event* ist genau dann gegeben, wenn der Benutzer die jeweilige Modalität verwendet, um eine Eingabe an das System zu formulieren.

Der Status *is-processing-input* gibt an, ob das System zurzeit eine Eingabe des Benutzers verarbeitet. In diesem Fall könnte die Präsentationsgenerierung z. B. reagieren, indem weitere Ausgaben zurückgestellt werden, um den Benutzer nicht zu unterbrechen.

Weiterhin werden bei visuellen Eingabemodalitäten die Sichtbarkeit, die Bildschirmkoordinaten und die Größe modelliert, da diese Konzepte Einfluss auf die Präsentationsgenerierung haben. So könnte z. B. das Fenster einer Ausgabekomponente das Fenster einer Eingabekomponente auf demselben Gerät überdecken. Bei auditoriellen Eingabegeräten (wie z. B. einer Spracherkennung) werden keine Parameter und Status im domänenunabhängigen Modell definiert.

Zwecks einer detaillierten Modellierung von Eingabekomponenten könnte für auditorielle Eingabegeräte ein Parameter *application-input-level*  $\in \mathbb{N}$  eingeführt werden, der dem Parameter *application-volume* bei auditoriellen

## 5.4. MODELLIERUNG MULTI-MODALER EIN-/AUSGABEKOMPONENTEN

Ausgabekomponenten entspricht und den Pegel der auditoriellen Eingabe angibt. Weiterhin könnte das Modell analog zu dem Modell für Ausgabekomponenten um Kennwerte ergänzt werden, die die Leistung des Eingabegerätes beschreiben. Da diese aber für die Präsentationsplanung nicht von direkter Relevanz sind, werden diese Erweiterungen in dieser Arbeit nicht betrachtet.

### 5.4.4 Modellierung einer Home-Entertainment Umgebung

In diesem Abschnitt wird das domänenunabhängige Modell zu einem domänenspezifischen Modell für Ausgabedienste, -geräte und -ressourcen in einem Home-Entertainment-Szenario erweitert. Das System soll Ausgabekomponenten unterstützen, die mittels einer Sprachausgabe und optionalem Grafik-Feedback eine Repräsentation eines Assistenten generieren können, der den Benutzer bei der Verwendung des Systems unterstützen kann.

Tabelle 5.3: Multi-modale Parameter und Status für *agent* und *player*.

| Typ           | Parameter                                                                                                                                             | Status                                                                                                                                                         |
|---------------|-------------------------------------------------------------------------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <i>agent</i>  | <i>visibility</i><br><i>screen-location</i><br><i>size</i><br><i>appearance</i> ∈ String<br><i>behaviour</i> (Tabelle 5.4)<br><i>content</i> ∈ String | <i>is-visible</i><br><i>has-screen-location</i><br><i>has-size</i><br><i>has-appearance</i><br><i>is-playing</i><br><i>has-output-media-list</i> (Tabelle 5.6) |
| <i>player</i> | <i>media-url</i> ∈ URL<br><i>media-format</i>                                                                                                         | –                                                                                                                                                              |

Weiterhin soll das System bei der Medienwiedergabe verschiedene Formate unterstützen, wie z. B. eine Komponente zur Wiedergabe von MPEG-Videos oder eine Komponente zur Wiedergabe von SMIL-Präsentationen. Aus diesem Grund werden für die Taxonomie der Ausgabekomponenten zwei multi-modale Typen definiert, *agent* und *player*. Der Typ *agent* umfasst alle Komponenten, deren Ausgaben einen Assistenten repräsentieren (vgl. Abschnitt 5.4.1). Diese bestehen aus einer Sprachausgabe, die durch beliebige Animationen (z. B. Kopfbewegungen, Lippenbewegungen oder Gesten) ergänzt werden kann.

Der Typ *player* umfasst unabhängig vom Medienformat alle Wiedergabekomponenten. Dies können z. B. MPEG-Filme oder SMIL-Animationen sein. Den multi-modalen Typen *agent* und *player* werden die in Tabelle 5.3 dargestellten multi-modalen Parameter zugeordnet.

### Ausgabekomponenten vom Typ *agent*

Ausgabekomponenten vom Typ *agent* verfügen über die Parameter *visibility*, *screen-location*, *size*, *appearance*, *behaviour* und *content*. Die Parameter *visibility*, *size* und *screen-location* bestimmen dabei bei animierten Charakteren die Sichtbarkeit, die Größe und den Ort, an dem der Charakter auf dem Bildschirm erscheint. Diese Parameter ersetzen die entsprechenden unimodalen Parameter des domänenunabhängigen Modells für visuelle Unimodalitäten (Tabelle 5.1). Der Grund für diese Vorgehensweise ist, dass es bei der Generierung von animierten Charakteren nicht sinnvoll ist, die Größe und Position von Gesten, Lippenbewegungen und Körperbewegungen separat zu definieren. Einfacher ist die Verwendung eines Fensters, das sämtliche visuellen Ausgaben des Charakters umfasst und innerhalb dessen alle Animationen des animierten Charakters stattfinden.

Der Parameter *appearance* dient zu Auswahl des Erscheinungsbildes eines animierten Charakters. So existieren z. B. für animierte Charaktere, die auf dem Microsoft Agent-System aufbauen, die Möglichkeit, zwischen verschiedenen Charakteren (z. B. „Merlin“, dem Zauberer) zu wählen. Im Falle von Sprachausgabekomponenten ohne Grafik-Feedback kann durch *appearance* die zu verwendende Stimme ausgewählt werden. Weiterhin existiert ein multimodaler Parameter *behaviour*, mit dem auf abstrakte Weise das Verhalten des animierten Charakters beschrieben werden kann. Der Parameter *behaviour* steht für eine direkte Ansteuerung der Verhaltensweisen des Charakters, bei der die Generierung und die Darstellung des Charakters kombiniert erfolgen.

Zur Modellierung von *behaviour* sind zwei Ansätze möglich. Animierte Charaktere können mittels einer von den konkreten Verhaltensweisen unabhängigen kommunikativen Absicht modelliert werden (abstrakte Modellierungsebene, vgl. Abschnitt 3.4.2). Beispiele für konkrete Verhaltensweisen sind Mimik oder Gestik. Ebenso können animierte Charaktere auch auf der Ebene konkreter Verhaltensweisen angesteuert werden (physikalische Modellierungsebene). Die Verwendung der abstrakten Modellierung erlaubt die Definition einer Menge von kommunikativen Absichten, die unabhängig vom Erscheinungsbild eines animierten Charakters sind. Sie erfordern jedoch neben der Darstellung der Verhaltensanimationen die Durchführung einer Verhaltensplanung durch die Charakterkomponente, um aus der abstrakten Beschreibung die physikalische Umsetzung abzuleiten.

Die Behaviour Markup Language (BML, [Kopp et al., 2006]) stellt den allgemeinsten Ansatz zur physikalischen Modellierung von animierten Charakteren dar. Weiterhin erlaubt BML eine einfache Integration in das in Abschnitt 5.4.3 verwendete Ausgabeunimodalitätenmodell von Bernsen. Die Ausgabe-

#### 5.4. MODELLIERUNG MULTI-MODALER EIN-/AUSGABEKOMPONENTEN

modalität Gestik kann z. B. durch das Tupel (-li, an, -ar, dyn, vis, gestures) als grafische Animation beschrieben werden (vgl. Abschnitt 4.3). Wenn ein animierter Charakter diese Unimodalität unterstützt, dann muss er auch die das korrespondierende BML-Element `<gesture>` und dessen Parameter unterstützen. Somit könnte der Parameter *behaviour* anhand der Unimodalitäten des Charakters in BML formuliert werden.

In dieser Arbeit wurde BML jedoch nicht verwendet, um Charaktere mittels *behaviour* anzusteuern. Die Definition von BML ist zurzeit noch nicht abgeschlossen. Erste Implementierungen von BML-Renderern existieren zwar [Lee & Marsella, 2006]. Jedoch ist noch keine freie Software zur Umsetzung von BML in die Animationen eines animierten Charakters verfügbar. Die Realisierung einer eigenen, BML-basierten Ausgabekomponente geht jedoch über den Rahmen dieser Arbeit hinaus.

Aus diesem Grund haben wir für die Modellierung des *behaviour*-Parameters in der geplanten Home-Entertainment-Umgebung einen konservativeren Ansatz gewählt. In dieser Arbeit beschränken wir uns auf die in [Rieger, 2003] definierten Verhaltensweisen. Diese werden in Tabelle 5.4 geordnet nach emotionalen Zuständen (z. B. fröhlich oder aufmerksam) und kommunikativem Feedback (z. B. Nicken oder Kopfschütteln) dargestellt. Dieses Modell ist im Gegensatz zu BML der abstrakten Modellierungsebene zuzuordnen und nicht an konkrete Ausgabemodalitäten des Charakters gebunden. Zwar wird nur eine begrenzte Menge an Verhaltensweisen unterstützt. Diese werden jedoch im Hinblick auf die geplante Home-Entertainment-Umgebung als ausreichend betrachtet. Der Parameter *content* in Tabelle 5.3 enthält den Text, der durch den Charakter gesprochen werden soll, im ASCII-Format. Da dieser Text sowohl Einfluss auf die Sprachausgabe als auch auf die Lippenbewegungen des Charakters hat, wird dieser Parameter als multi-modaler Parameter definiert (vgl. Abschnitt 5.4.1).

In Analogie zu den multi-modalen Parametern in Tabelle 5.3 existieren Status *is-visible*, *has-screen-location*, *has-size* und *has-appearance*. Für *behaviour* und *content* wurde kein entsprechender Status definiert, da die Gültigkeit eines solchen Status nur von begrenzter Dauer wäre. Zusätzlich wurde analog zu den dynamischen Unimodalitäten im domänenunabhängigen Modell (Tabelle 5.1) ein Status *is-playing* definiert, der angibt, ob der animierte Charakter zurzeit Animationen darstellt oder Sprache ausgibt.

Für die Umsetzung der multi-modalen Ausgabe von Komponenten des Typs *agent* existieren zwei Möglichkeiten. Die Ausgabekomponente kann die Ausgabe entweder direkt ausgeben oder ein Medienobjekt zurückliefern, das die Ausgabe in kodierter Form enthält (z. B. eine WAV-Datei, ein GIF-Bild oder

ein MPEG-Video). Dieses Medienobjekt kann durch eine *player*-Komponente auf einem verbundenen Rechner ausgegeben werden. Dies wird durch den Status *has-output-media-list* modelliert. Dabei wird definiert, dass Komponenten, die Medienobjekte zurückliefern, grundsätzlich keine Ausgaben selber darstellen. Umgekehrt geben Komponenten, die keine Medienobjekte zurückliefern, die Inhalte direkt aus.

Tabelle 5.4: Verhaltensweisen von animierten Charakteren vom Typ *agent*.

| Emotionale Zustände     |                                                           |
|-------------------------|-----------------------------------------------------------|
| Bezeichner              | Effekt                                                    |
| neutral                 | Neutraler Gesichtsausdruck und Tonfall                    |
| positiv-neutral         | Freundlicher Gesichtsausdruck und Tonfall                 |
| sorry                   | tiefe Stimme, trauriges Gesicht                           |
| regretful               | tiefe Stimme, trauriges Gesicht, Schultern heben/senken   |
| alert                   | Benutzer aufmerksam angucken                              |
| happy                   | Lächeln und fröhlicher Tonfall                            |
| interested              | Benutzer interessiert ansehen                             |
| Kommunikatives Feedback |                                                           |
| Bezeichner              | Effekt                                                    |
| friendly-greeting       | Freundlicher Gesichtsausdruck/Tonfall, winken             |
| friendly-closing        | Freundlicher Gesichtsausdruck/Tonfall, winken, entfernen  |
| friendly-offer          | Freundlicher Gesichtsausdruck/Tonfall, warten auf Eingabe |
| show-agreement          | Nicken                                                    |
| show-rejection          | Kopfschütteln                                             |
| show-alternatives       | Neutraler Gesichtsausdruck/Tonfall, warten auf Eingabe    |
| show-importance         | Stimme lauter, mit Nachdruck sprechen                     |

Beide Ansätze haben Vorteile und Nachteile. Die direkte Ausgabe der zu präsentierenden Inhalte schont Netzwerkressourcen, da die generierten Medienobjekte von anderen Komponenten nicht geladen werden müssen. Der Nachteil dieser Vorgehensweise ist jedoch, dass die Integration mehrerer Komponenten in ein gemeinsames Layout schwieriger ist, da sie viel Kommunikation zwischen den Komponenten erfordert. Wird hingegen durch eine Sprachausgabe eine Audiodatei generiert und als Medienobjekt zur Verfügung gestellt, so kann diese leicht in eine SMIL- oder HTML-Präsentation eingebunden werden.

Dort kann die Ausgabe des Sprachsignals auf einfache Weise mit anderen Medien synchronisiert und in ein gemeinsames Layout integriert werden. Der Nachteil dieser Vorgehensweise ist neben der erhöhten Netzwerkbelastung,

#### 5.4. MODELLIERUNG MULTI-MODALER EIN-/AUSGABEKOMPONENTEN

dass bei animierten Charakteren auch Grafikdateien zur Verfügung gestellt werden müssen und dass diese durch das Präsentationssystem in kohärenter Form entsprechend dem Verhalten des Charakters (Tabelle 5.4) umgesetzt werden müssen. Dies setzt insbesondere bei animierten Charakteren eine Kategorisierung der Grafikdateien voraus.

Beim zweiten Ansatz haben die Netzwerkkennwerte des Ausgabegeräts, auf dem die visuellen oder auditoriellen Medienobjekte dargestellt werden, Einfluss auf das Übertragen der Medienobjekte auf das Gerät. Diese Kennwerte können es erforderlich machen, ein Medienobjekt in seiner Qualität anzupassen, um die Übertragungszeit zu verkürzen.

Tabelle 5.5: Animationen für Komponenten vom Typ *agent*.

| Animation            | Erläuterung                                         |
|----------------------|-----------------------------------------------------|
| <i>restpose</i>      | neutrale Ruhehaltung                                |
| <i>speak</i>         | Sprechanimation                                     |
| <i>gesture-right</i> | Zeigegeste nach rechts                              |
| <i>gesture-left</i>  | Zeigegeste nach links                               |
| <i>look-right</i>    | Blick nach rechts                                   |
| <i>look-left</i>     | Blick nach links                                    |
| <i>look-happy</i>    | Ruheanimation mit fröhlichem Gesichtsausdruck       |
| <i>look-sad</i>      | Ruheanimation mit traurigem Gesichtsausdruck        |
| <i>look-sorry</i>    | Ruheanimation mit entschuldigendem Gesichtsausdruck |
| <i>enter</i>         | Charakter erscheint                                 |
| <i>leave</i>         | Charakter verabschiedet sich                        |

Für das Home-Entertainment-System wird ein Modell von Animationen definiert, die die Animationen von Komponenten vom Typ *agent* kategorisieren und eine Komposition der Animationen durch das Präsentationssystem ermöglichen. In dem Beispielszenario werden die in Tabelle 5.5 dargestellten Animationen verwendet.

Die Kategorien der Animationen basieren auf den Animationen des Ckuckuk Präsentationsplanungssystems [André et al., 2005; André et al., 2000]. Das Modell entspricht einer Teilmenge der Animationen des Microsoft Agent Modells, das ebenfalls in der Modellierungssprache MPML 3.0 [Prendinger et al., 2004] eingesetzt wird. Wie beim Parameter *behaviour* wird hier nur eine begrenzte Menge von Verhaltensweisen definiert. Das Modell ist jedoch leicht durch detailliertere Beschreibungen von Animationen zu erweitern.

Somit wird für Ausgabekomponenten vom Typ *agent* durch *has-output-media-list* eine Liste von Medienobjekten zurückgegeben, die die in Tabelle 5.6

dargestellte Form haben. Bei grafischen Medienobjekten werden Animationen im GIF-Format unterstützt, die ein ressourcenschonendes Format für Animationen darstellen.

Weiterhin wird die Auflösung der Animation festgehalten, um eine Integration in das Gesamt-Layout der Präsentation zu ermöglichen. Um Medienobjekte in Abhängigkeit der Netzwerkennwerte von Ausgabegeräten auszuwählen, wird die Dateigröße der Animation festgehalten. Neben dem Typ der Animation wird der Netzwerkpfad (Uniform Resource Locator, URL) zur Animation angegeben, der einen netzwerkweiten Zugriff auf das Medienobjekt ermöglicht.

Tabelle 5.6: Ausgabemedium für Komponenten vom Typ *agent*.

| <i>output-medium</i> |                                           |                                                 |               |                              |                         |
|----------------------|-------------------------------------------|-------------------------------------------------|---------------|------------------------------|-------------------------|
| visuell              |                                           |                                                 |               |                              |                         |
| Medium<br>vis        | Format<br>GIF                             | Auflösung<br>$\in \mathbb{N} \times \mathbb{N}$ | Größe<br>[kB] | Animationstyp<br>Tabelle 5.5 | URL<br>$\in \text{URL}$ |
| auditoriell          |                                           |                                                 |               |                              |                         |
| Medium<br>aud        | Format<br>$\in \{\text{WAV}, \text{AU}\}$ | Abtastfrequenz<br>[Hz]                          | Größe<br>[kB] | URL<br>$\in \text{URL}$      |                         |

Bei akustischen Medienobjekten werden Dateien im WAV und im AU-Format unterstützt. Um die Qualität unterschiedlicher Medienobjekte unterscheiden zu können, wird die Abtastfrequenz der Audiodatei festgehalten. Analog zu den grafischen Medienobjekten werden ebenso die Dateigröße und die URL verwaltet.

Ausgabekomponenten müssen nicht alle multi-modalen Parameter unterstützen. So können z. B. Ausgabekomponenten vom Typ *agent* existieren, die keine Auswahl des Erscheinungsbildes des Charakters unterstützen. Unterstützt eine Ausgabekomponente vom Typ *agent* den Status *has-output-media-list*, so haben nur die Parameter *appearance* und *size* einen Effekt, da die Komponente in diesem Fall nur Medienobjekte generiert und keine eigene Ausgabe darstellt.

Ausgabekomponenten vom Typ *player*

Ausgabekomponenten vom Typ *player* verfügen über zwei Parameter, die die Netzwerkadresse des Mediums (*media-url*) angeben, das abgespielt werden soll. Weiterhin definiert der Parameter *media-format* das Format des Mediums (vgl. Tabelle 5.3). Dabei enthält *media-url* einen String im URL-

#### 5.4. MODELLIERUNG MULTI-MODALER EIN-/AUSGABEKOMPONENTEN

Format. Für das Home-Entertainment-System werden dazu die Werte „*mpeg*“, „*smil*“, und „*html*“ unterstützt. Da Komponenten vom Typ *player* als Wiedergabekomponenten definiert wurden, geben sie die Ausgabe per Definition direkt aus. Daher wird für diese kein Status *has-output-media-list* definiert.

##### Ausgabeunimodalitäten

Neben der Definition der multi-modalen Parameter und Status erfolgt eine Definition der atomaren Ausgabeunimodalitäten, die durch das System unterstützt werden sollen und von den Komponenten vom Typ *agent* oder *player* eingesetzt werden (Tabelle 5.7). Diese stellen eine Erweiterung der durch das domänenunabhängige Modell definierten generischen Unimodalitätenklassen dar (vgl. Abschnitt 5.4.3).

Tabelle 5.7: Atomare Ausgabeunimodalitäten: Bernsen-Kategorien.

| Unimodalität | Bernsen-Kategorie                                    |
|--------------|------------------------------------------------------|
| Kopf         | ( <i>-li, an, -ar, dyn, vis, head</i> )              |
| Torso        | ( <i>-li, an, -ar, dyn, vis, torso</i> )             |
| Mimik        | ( <i>-li, an, -ar, dyn, vis, facial-expression</i> ) |
| Blick        | ( <i>-li, an, -ar, dyn, vis, gaze</i> )              |
| Körper       | ( <i>-li, an, -ar, dyn, vis, body</i> )              |
| Beine        | ( <i>-li, an, -ar, dyn, vis, legs</i> )              |
| Gestik       | ( <i>-li, an, -ar, dyn, vis, gestures</i> )          |
| Lippen       | ( <i>li, an, -ar, dyn, vis, spoken-text</i> )        |
| Sprache      | ( <i>li, -an, -ar, dyn, aud, spoken-text</i> )       |
| Video        | ( <i>-li, an, -ar, dyn, vis, -</i> )                 |
| Audio        | ( <i>-li, an, -ar, dyn, aud, -</i> )                 |

Auf der generischen Ebene beschreibt die Kategorie (*-li, an, -ar, dyn, vis*) bei Bernsen allgemeine grafische Animationen (Tabelle 4.1). Für animierte Charaktere ergänzen wir diese Klasse durch atomare Unterklassen, die für die Umsetzung der Ausgaben von animierten Charakteren relevant sind: Kopfhaltung (*head*), Torso (*torso*), Mimik (*facial-expression*), Blickrichtung (*gaze*), Körperhaltung/Bewegung (*body*), Beinstellung (*legs*) und Gestik (*gestures*).

Hierbei orientieren wir uns an den in BML definierten Ausgabemodalitäten [Kopp et al., 2006]. Diese werden als nicht-linguistische, visuelle Animationen modelliert. Dabei werden die Lippenbewegungen als linguistische Unimodalität modelliert, da Lippenbewegungen ebenfalls linguistische Informa-

## KAPITEL 5. MODELLIERUNG MULTI-MODALER AUSGABEN

tionen übertragen können. So werden z. B. beim Lippenlesen ausschließlich visuelle Informationen verwendet, um linguistische Informationen auszutauschen. Für die Umsetzung der Sprachausgabe wird die generische Unimodalität (*li*, *-an*, *-ar*, *dyn*, *aud*) analog zu den Lippenbewegungen um eine atomare Klasse *spoken-text* erweitert, um gesprochenen Text zu modellieren.

Tabelle 5.8: Atomare Ausgabeunimodalitäten: Parameter, Status und Geräte.

| Unimodalität | Parameter                                                                         | Status                                                                                     |
|--------------|-----------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------|
| Hände        | <i>ref-window-location</i><br>$\in \mathbb{N} \times \mathbb{N}$                  | –                                                                                          |
| Arme         | –                                                                                 | –                                                                                          |
| Mimik        | –                                                                                 | –                                                                                          |
| Kopf         | –                                                                                 | –                                                                                          |
| Blick        | <i>ref-window-location</i><br>$\in \mathbb{N} \times \mathbb{N}$                  | –                                                                                          |
| Körper       | –                                                                                 | –                                                                                          |
| Bewegung     | <i>start-location</i> ,<br><i>end-location</i> $\in \mathbb{N} \times \mathbb{N}$ | –                                                                                          |
| Lippen       | –                                                                                 | –                                                                                          |
| Sprache      | <i>application-volume</i>                                                         | <i>has-application-volume</i><br><i>is-playing</i>                                         |
| Video        | <i>visibility</i><br><i>screen-location</i><br><i>size</i>                        | <i>is-visible</i><br><i>has-screen-location</i><br><i>has-size</i><br><i>is-playing</i>    |
| Audio        | <i>application-volume</i>                                                         | <i>has-application-volume</i><br><i>is-playing</i>                                         |
| Ausgabegerät | Parameter                                                                         | Status                                                                                     |
| Visuell      | –                                                                                 | <i>has-resolution</i><br><i>has-graphics-benchmark</i>                                     |
| Auditoriell  | –                                                                                 | <i>has-device-volume</i><br><i>has-max-device-volume</i><br><i>has-min-device-volume</i>   |
| Alle         | –                                                                                 | <i>has-id</i><br><i>has-networking-benchmark</i><br><i>has-room-id</i> $\in \text{String}$ |

Für den multi-modalen Typ *player* modellieren wir allgemeine grafische Animationen („Video“) und allgemeine auditive Ausgaben („Audio“), um die

#### 5.4. MODELLIERUNG MULTI-MODALER EIN-/AUSGABEKOMPONENTEN

audiovisuellen Wiedergabefähigkeiten der Komponenten vom Typ *player* repräsentieren zu können. Für die Modellierung dieser Unimodalitäten genügt die generische Ebene des Bernsen-Modells.

Im Folgenden werden die uni-modalen Parameter definiert, die die uni-modalen Parameter im domänenunabhängigen Modell ergänzen (Tabelle 5.8). Analog zu den multi-modalen Parametern ist es nicht notwendig, dass eine Ausgabekomponente alle Parameter unterstützt. Die Unimodalitäten Hände und Blickrichtung verwenden den Parameter *ref-window-location*. Dieser Parameter ermöglicht es, Zeigegesten auf einen Punkt auf dem Bildschirm durchzuführen bzw. den Blick auf diesen Punkt zu lenken.

Bei der Unimodalität Bewegung werden der Startpunkt (*start-location*) und der Endpunkt der Bewegung (*end-location*) übergeben. Die weiteren visuellen Unimodalitäten der animierten Charaktere haben keine Parameter. Die Ausgabe dieser Unimodalitäten wird bestimmt durch die Werte der multi-modalen Parameter von Komponenten des Typs *agent* (Tabelle 5.3).

Bei diesen Unimodalitäten werden keine weiteren Status eingeführt, da die multi-modalen Status des Typs *agent* die multi-modale Ausgabe als Ganzes bereits ausreichend charakterisieren. So ist es für ein Präsentationssystem z. B. nicht relevant, bei welchen Koordinaten sich die Lippen eines animierten Charakters befinden. Relevant ist stattdessen die Position des Fensters, das die gesamten visuellen Ausgaben des Charakters umgibt.

Bei der Unimodalität Sprache wird der Parameter *application-volume* aus dem domänenunabhängigen Modell für dynamische, auditorielle Ausgabekomponenten übernommen (Tabelle 5.1). Der Parameter ist sowohl für Sprachausgabekomponenten, die nur über diese Unimodalität verfügen, als auch für animierte Charaktere, die über zusätzliches Grafik-Feedback verfügen, relevant. Da der Parameter bei Komponenten vom Typ *agent* nur die Unimodalität Sprache betrifft, wird er als uni-modaler Parameter definiert.

Die Status der Unimodalität Sprache werden ebenfalls aus dem domänenunabhängigen Modell übernommen. Die Parameter und Status der Unimodalitäten Audio und Video werden aus dem domänenunabhängigen Modell für dynamische, auditorielle bzw. dynamische, visuelle Ausgabekomponenten übernommen. Da Komponenten vom Typ *player* die Sichtbarkeit, den Ort und die Größe der Ausgabe nicht als multi-modale Parameter modellieren, werden diese durch die Unimodalitäten definiert.

Der Grund hierfür ist, dass es sinnvoll ist, bei Wiedergabekomponenten die Audio- und Video-Ausgaben getrennt zu verwalten. So wird z. B. bei der Wiedergabe von Musik die Videoausgabe nicht verwendet oder bei Filmen ohne Ton keine Audioausgabe. Die Status der visuellen und auditoriellen Ausgabe-

geräte wurden ebenfalls aus dem domänenunabhängigen Modell für Ausgabe-  
komponenten übernommen. Zusätzlich wurde zur Modellierung des Raumes,  
in dem eine Ausgabe stattfindet (z. B. Küche oder Wohnzimmer) ein Status  
*has-room-id* für alle Ausgabegeräte eingeführt, der einen eindeutigen Be-  
zeichner für den Raum liefert, in dem sich das Ausgabegerät befindet. Dieser  
kann auch eine leere Zeichenkette sein, für den Fall, dass der Raum nicht be-  
kannt ist.

Eingabekomponenten

Das Home-Entertainment-System soll zwei Typen von Eingabekomponenten  
unterstützen. Visuelle Eingaben sollen durch GUI erfolgen können, mit denen  
Systemdialoge visuell erfolgen können. Auditorielle Eingaben sollen durch  
Spracherkennungskomponenten erfolgen können. Die Bernsen-Kategorien für  
beide Eingabemodalitäten sind in Tabelle 5.9 dargestellt.

Tabelle 5.9: Atomare Eingabeunimodalitäten: Bernsen-Kategorien.

| Unimodalität      | Bernsen-Kategorie                |
|-------------------|----------------------------------|
| visuelle Eingaben | ( <i>-li, -an, -ar, vis, -</i> ) |
| Spracheingabe     | ( <i>li, -an, -ar, aud, -</i> )  |

Die Parameter und Status der beiden Eingabeunimodalitäten sind in Tabelle  
5.10 dargestellt. Für die Unimodalitäten visuelle Eingabe und Spracherken-  
nung wurden die Parameter und Status aus dem domänenunabhängigen Mo-  
dell (Tabelle 5.2) übernommen.

Tabelle 5.10: Atomare Eingabeunimodalitäten: Parameter, Status und Geräte.

| Unimodalität      | Parameter                | Status                      |
|-------------------|--------------------------|-----------------------------|
| visuelle Eingaben | <i>visibility</i>        | <i>is-visible</i>           |
|                   | <i>screen-location</i>   | <i>has-screen-location</i>  |
|                   | <i>size</i>              | <i>has-size</i>             |
| Spracheingabe     | –                        | <i>threw-input-event</i>    |
|                   |                          | <i>is-processing-input</i>  |
|                   | <i>threw-input-event</i> |                             |
| Eingabegeräte     | Parameter                | Status                      |
| Visuell           | –                        | <i>has-resolution</i>       |
| Alle              | –                        | <i>has-id</i>               |
|                   |                          | <i>has-room-id</i> ∈ String |

## 5.4. MODELLIERUNG MULTI-MODALER EIN-/AUSGABEKOMPONENTEN

Bei der Unimodalität visuelle Eingabe wird der Status *is-processing-input* nicht unterstützt, da dieser für visuelle Eingaben mittels eines GUI keine Aussagekraft besitzt.

Bei den Eingabegeräten werden ebenfalls die Parameter und Status des domänenunabhängigen Modells unterstützt. Analog zu den Ausgabekomponenten werden diese durch einen Status *has-room-id* ergänzt, der einen Bezeichner für den Raum enthält, in dem sich das Ausgabegerät befindet oder eine leere Zeichenfolge enthält, falls der Raum nicht bekannt ist.

### 5.4.5 Beispiele für Selbstbeschreibungen

Im Folgenden werden auf Basis der in Abschnitt 5.4.4 definierten Modelle Beispiele für mögliche Selbstbeschreibungen von Ein-/Ausgabekomponenten gegeben. Dabei werden unterschiedliche Komponenten der Typen *agent* und *player* modelliert, die auf der Basis des Modells automatisch in die Präsentation generiert werden können.



Abbildung 5.13: Microsoft Agent-Charakter Merlin (Nachdruck der Screen Shots mit freundlicher Erlaubnis der Microsoft Corporation).

Zunächst soll ein animierter Microsoft Agent-basierter Charakter als Ausgabekomponente modelliert werden (Bild 5.13). Da die Größe des Charakters nicht veränderlich ist, unterstützt der Charakter den multi-modalen Parameter *size* einer Komponente vom Typ *agent* nicht (Tabelle 5.3). Alle weiteren multi-modalen Parameter einer Komponente vom Typ *agent* werden unterstützt.

Weiterhin unterstützt der Charakter alle Unimodalitäten aus Tabelle 5.8 bis auf die allgemeine Audio- oder Videoausgabe. Der Parameter *ref-window-location* der Unimodalität Blick wird nicht unterstützt, da die Blickrichtung sich an der ausgeführten Gestik orientiert und nicht separat verwendet werden kann. Ansonsten werden durch die Komponente sämtliche in Tabelle 5.8 definierten Parameter unterstützt. Damit ergibt sich für die Komponente die in Tabelle 5.11 dargestellte Selbstbeschreibung.

Tabelle 5.11: Selbstbeschreibung eines Microsoft Agent-Charakters.

| Typ           | Parameter                                                                                                                              | Status                                                                                                           |
|---------------|----------------------------------------------------------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------|
| <i>agent</i>  | <i>visibility</i><br><i>screen-location</i><br><i>appearance</i> ∈ String<br><i>behaviour</i> (Tabelle 5.4)<br><i>content</i> ∈ String | <i>is-visible</i><br><i>has-screen-location</i><br><i>has-size</i><br><i>has-appearance</i><br><i>is-playing</i> |
| Unimodalität  | Parameter                                                                                                                              | Status                                                                                                           |
| Hände         | <i>ref-window-location</i><br>∈ $\mathbb{N} \times \mathbb{N}$                                                                         | –                                                                                                                |
| Arme          | –                                                                                                                                      | –                                                                                                                |
| Mimik         | –                                                                                                                                      | –                                                                                                                |
| Kopf          | –                                                                                                                                      | –                                                                                                                |
| Blick         | –                                                                                                                                      | –                                                                                                                |
| Körper        | –                                                                                                                                      | –                                                                                                                |
| Bewegung      | <i>start-location</i> ,<br><i>end-location</i> ∈ $\mathbb{N} \times \mathbb{N}$                                                        | –                                                                                                                |
| Lippen        | –                                                                                                                                      | –                                                                                                                |
| Sprache       | <i>application-volume</i>                                                                                                              | <i>has-application-volume</i><br><i>is-playing</i>                                                               |
| Ausgabegeräte | Parameter                                                                                                                              | Status                                                                                                           |
| Visuell       | –                                                                                                                                      | <i>has-resolution</i><br><i>has-graphics-benchmark</i>                                                           |
| Auditoriell   | –                                                                                                                                      | <i>has-device-volume</i><br><i>has-max-device-volume</i><br><i>has-min-device-volume</i>                         |
| Alle          | –                                                                                                                                      | <i>has-id</i><br><i>has-networking-benchmark</i><br><i>has-room-id</i> ∈ String                                  |



Abbildung 5.14: Animierter Charakter der TU Darmstadt.

#### 5.4. MODELLIERUNG MULTI-MODALER EIN-/AUSGABEKOMPONENTEN

Dadurch ist es möglich, Komponenten vom Typ Microsoft Agent mit anderen Ausgabekomponenten zu koordinieren, z. B. einem GUI. Dabei kann der Parameter *appearance* die Werte „Genie“, „Merlin“, „Peedy“ und „Robby“ annehmen, die den von Microsoft realisierten Charakteren entsprechen. Da der Microsoft Charakter keine Ausgabemedien zurückliefert, sondern die Inhalte direkt ausgibt, wird der Status *has-output-media-list* nicht unterstützt. Alle weiteren Status aus Tabelle 5.8 werden verwendet.

Tabelle 5.12: Selbstbeschreibung des TU Darmstadt Charakters.

| Typ           | Parameter                                                                                                               | Status                                                                                   |
|---------------|-------------------------------------------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------|
| <i>agent</i>  | <i>visibility</i><br><i>screen-location</i><br><i>size</i><br><i>behaviour</i> (Tabelle 5.4)<br><i>content</i> ∈ String | <i>is-visible</i><br><i>has-screen-location</i><br><i>has-size</i><br><i>is-playing</i>  |
| Unimodalität  | Parameter                                                                                                               | Status                                                                                   |
| Mimik         | –                                                                                                                       | –                                                                                        |
| Kopf          | –                                                                                                                       | –                                                                                        |
| Blick         | <i>ref-window-location</i><br>∈ $\mathbb{N} \times \mathbb{N}$                                                          | –                                                                                        |
| Lippen        | –                                                                                                                       | –                                                                                        |
| Sprache       | <i>application-volume</i>                                                                                               | <i>has-application-volume</i><br><i>is-playing</i>                                       |
| Ausgabegeräte | Parameter                                                                                                               | Status                                                                                   |
| Visuell       | –                                                                                                                       | <i>has-resolution</i><br><i>has-graphics-benchmark</i>                                   |
| Auditoriell   | –                                                                                                                       | <i>has-device-volume</i><br><i>has-max-device-volume</i><br><i>has-min-device-volume</i> |
| Alle          | –                                                                                                                       | <i>has-id</i><br><i>has-networking-benchmark</i><br><i>has-room-id</i> ∈ String          |

Als zweite Komponente wird der animierte Charakter der TU Darmstadt (Abbildung 5.14, [Rieger & Berner, 2002]) modelliert. Da dieser Charakter beliebige Größen annehmen kann, wird der multi-modale Parameter *size* im Gegensatz zum Microsoft Agent-Charakter unterstützt (vgl. Tabelle 5.3). Es werden jedoch keine unterschiedlichen Typen des Charakters modelliert, daher wird der Parameter *appearance* nicht unterstützt.



Abbildung 5.15: DFKI-Charakter Herr Kaufmann.

Der Charakter verfügt über Animationen für keine Arme, Hände oder Körper. Daher unterstützt der Charakter die Unimodalitäten Arme, Mimik, Körper und Bewegung nicht. Mit Hilfe der Unimodalität Blick ist es möglich, den Charakter auf verschiedene Bildschirmkoordinaten schauen zu lassen. Daher wird deren Parameter *ref-window-location* unterstützt. Ansonsten werden alle Parameter aus den Tabellen 5.3 und 5.8 unterstützt.

Tabelle 5.13: Selbstbeschreibung des DFKI-Charakters Herr Kaufmann.

| Typ           | Parameter | Status                                      |
|---------------|-----------|---------------------------------------------|
| <i>agent</i>  | –         | <i>has-output-media-list</i> (Tabelle 5.14) |
| Unimodalität  | Parameter | Status                                      |
| Hände         | –         | –                                           |
| Arme          | –         | –                                           |
| Mimik         | –         | –                                           |
| Kopf          | –         | –                                           |
| Blick         | –         | –                                           |
| Körper        | –         | –                                           |
| Lippen        | –         | –                                           |
| Ausgabegeräte | Parameter | Status                                      |
| –             | –         | –                                           |

Der Charakter Herr Kaufmann des DFKI [André et al., 2005] wird durch eine Menge von GIF Animationen, die jeweils nicht mehr als 100kB groß sind, realisiert. Diese Umsetzung eines animierten Charakters ist daher besonders

#### 5.4. MODELLIERUNG MULTI-MODALER EIN-/AUSGABEKOMPONENTEN

gut geeignet, um von anderen Computern im Netzwerk geladen und dargestellt zu werden. Diese Vorgehensweise wird in der Ckuckuk-Distribution des Präsentationsplanungssystems des DFKI [André et al., 2000] verwendet, das die Animationen des Charakters mittels eines Planungsansatzes mit anderen Medienobjekten in eine Präsentation im SMIL-Format integriert. Eine Selbstbeschreibung des animierten Charakters findet sich in Tabelle 5.13. Die Ausgabemedien der Komponente finden sich in Tabelle 5.14.

Tabelle 5.14: Liste der Ausgabemedien des DFKI-Charakters Herr Kaufmann.

| <i>output-media-list</i> |        |           |               |                      |                                                         |
|--------------------------|--------|-----------|---------------|----------------------|---------------------------------------------------------|
| Medium                   | Format | Auflösung | Größe<br>[kB] | Animations-<br>typ   | URL                                                     |
| vis                      | GIF    | 240 × 175 | 10            | <i>restpose</i>      | <a href="http://1.0.0.1/1.gif">http://1.0.0.1/1.gif</a> |
| vis                      | GIF    | 316 × 230 | 15            | <i>restpose</i>      | <a href="http://1.0.0.1/2.gif">http://1.0.0.1/2.gif</a> |
| vis                      | GIF    | 240 × 175 | 23            | <i>speak</i>         | <a href="http://1.0.0.1/3.gif">http://1.0.0.1/3.gif</a> |
| vis                      | GIF    | 316 × 230 | 34            | <i>speak</i>         | <a href="http://1.0.0.1/4.gif">http://1.0.0.1/4.gif</a> |
| vis                      | GIF    | 240 × 175 | 46            | <i>gesture-right</i> | <a href="http://1.0.0.1/5.gif">http://1.0.0.1/5.gif</a> |
| vis                      | GIF    | 316 × 230 | 58            | <i>gesture-right</i> | <a href="http://1.0.0.1/6.gif">http://1.0.0.1/6.gif</a> |

Der Charakter verfügt über alle Unimodalitäten aus Tabelle 5.8 außer Bewegung sowie allgemeinen Audio- und Videoausgaben. Dabei werden keine Parameter unterstützt, da die Ausgabekomponente Medienobjekte zur Verfügung stellt (vgl. Tabelle 5.6). Dies wird durch die Unterstützung des Status *has-output-media-list* angezeigt.

Tabelle 5.15: Selbstbeschreibung einer Sprachausgabekomponente.

| Typ           | Parameter                 | Status                                     |
|---------------|---------------------------|--------------------------------------------|
| <i>agent</i>  | <i>content</i> ∈ String   | <i>has-output-media-list</i> (Tabelle 5.6) |
| Unimodalität  | Parameter                 | Status                                     |
| Sprache       | <i>application-volume</i> | –                                          |
| Ausgabegeräte | Parameter                 | Status                                     |
| –             | –                         | –                                          |

Es werden die Animationen für die Ruhehaltung, für die Sprechanimation und für eine Zeigegeste nach rechts in Form von animierten GIF-Bildern zur Verfügung gestellt (vgl. Tabelle 5.14). Die Kategorien der Animationen orien-

## KAPITEL 5. MODELLIERUNG MULTI-MODALER AUSGABEN

tieren sich dabei an dem in Tabelle 5.5 definierten Modell. Die Bilder werden in zwei Auflösungen angeboten ( $240 \times 175$  und  $316 \times 230$ ). Die Medienobjekte können durch das Präsentationssystem mit anderen Medien kombiniert werden. So könnte z. B. die Sprechanimation des Charakters mit der Audiodatei eines Sprachsignals kombiniert werden.

Tabelle 5.16: Liste der Ausgabemedien einer Sprachausgabekomponente.

| <i>output-media-list</i> |        |                     |            |                       |
|--------------------------|--------|---------------------|------------|-----------------------|
| Medium                   | Format | Abtastfrequenz [Hz] | Größe [kB] | URL                   |
| aud                      | WAV    | 22.050              | 950        | http://10.0.0.2/1.wav |
| aud                      | AU     | 12.000              | 517        | http://10.0.0.2/1.au  |

Weiterhin könnte neben dem Charakter ein Bild dargestellt werden, auf das der Charakter mittels der Gestenanimation verweist. Dabei könnten alle Medienobjekte in ein gemeinsames Layout mit einem gemeinsamen Hintergrundbild integriert werden.

Tabelle 5.17: Selbstbeschreibung einer Komponente zur SMIL-Wiedergabe.

| Typ           | Parameter                                                  | Status                                                                                   |
|---------------|------------------------------------------------------------|------------------------------------------------------------------------------------------|
| <i>player</i> | <i>media-url</i><br><i>media-format</i> $\in \{smil\}$     | –                                                                                        |
| Unimodalität  | Parameter                                                  | Status                                                                                   |
| Video         | <i>visibility</i><br><i>screen-location</i><br><i>size</i> | <i>is-visible</i><br><i>has-screen-location</i><br><i>has-size</i><br><i>is-playing</i>  |
| Audio         | <i>application-volume</i>                                  | <i>has-application-volume</i><br><i>is-playing</i>                                       |
| Ausgabegeräte | Parameter                                                  | Status                                                                                   |
| Visuell       | –                                                          | <i>has-resolution</i><br><i>has-graphics-benchmark</i>                                   |
| Auditoriell   | –                                                          | <i>has-device-volume</i><br><i>has-max-device-volume</i><br><i>has-min-device-volume</i> |
| Alle          | –                                                          | <i>has-id</i><br><i>has-networking-benchmark</i><br><i>has-room-id</i> $\in$ String      |

## 5.4. MODELLIERUNG MULTI-MODALER EIN-/AUSGABEKOMPONENTEN

Ähnlich wie bei den Medienobjekten des DFKI-Charakters wandelt eine Komponente zur Sprachausgabe Text in eine Audiodatei, die die zum Text passende gesprochene Sprache enthält, um. (Tabellen 5.15 und 5.16). Dazu verwendet die Komponente den in *content* gegebenen Inhalt und wandelt ihn in die in Tabelle 5.16 gegebenen Ausgabeformate um.

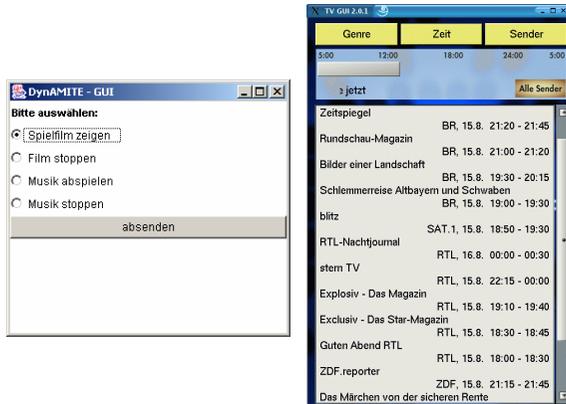


Abbildung 5.16: GUI-Eingabekomponenten aus Dynamite ([Elting & Hellen-schmidt, 2004], links) und Embassi ([Herfet et al., 2001], rechts).

Die Komponente unterstützt daher alle Parameter von Komponenten des Typs *player* aus Tabelle 5.3. Weiterhin werden die Unimodalitäten Audio und Video aus Tabelle 5.8 unterstützt, die die allgemeinen audiovisuellen Ausgaben repräsentieren, die durch die Wiedergabekomponente realisiert werden.

Tabelle 5.18: Selbstbeschreibung einer GUI-Eingabekomponente.

| Unimodalität      | Parameter              | Status                      |
|-------------------|------------------------|-----------------------------|
| visuelle Eingaben | <i>visibility</i>      | <i>is-visible</i>           |
|                   | <i>screen-location</i> | <i>has-screen-location</i>  |
|                   | <i>size</i>            | <i>has-size</i>             |
|                   |                        | <i>throw-input-event</i>    |
| Eingabegeräte     | Parameter              | Status                      |
| Visuell           | –                      | <i>has-resolution</i>       |
|                   | –                      | <i>has-id</i>               |
|                   |                        | <i>has-room-id</i> ∈ String |

Die Größe der erzeugten Dateien variiert dabei von Ausgabe zu Ausgabe. Im Beispiel werden die Audiodateien in zwei Formaten zur Verfügung gestellt: Eine qualitativ höherwertige Datei im WAV-Format mit einer Abtastfrequenz von 22.050 Hz und eine kompaktere AU-Datei mit einer Abtastfrequenz von 12.000 Hz. Als Beispiel für eine Komponente vom Typ *player* wird eine SMIL-Wiedergabekomponente modelliert. Diese Komponente bekommt mittels des Parameters *media-url* eine URL auf eine SMIL-Datei übergeben, die den auszugebenden Inhalt repräsentiert. Dieser wird durch die Komponente geladen und dargestellt.

Tabelle 5.19: Selbstbeschreibung einer Spracherkennungskomponente.

| Unimodalität    | Parameter | Status                                                 |
|-----------------|-----------|--------------------------------------------------------|
| Spracherkennung | –         | <i>is-processing-input</i><br><i>threw-input-event</i> |
| Eingabegeräte   | Parameter | Status                                                 |
| Auditoriell     | –         | <i>has-id</i><br><i>has-room-id</i> ∈ String           |

Als Beispiel für eine visuelle Eingabekomponente werden die Eingabemöglichkeiten eines GUI modelliert (Abbildung 5.16). Dieses GUI unterstützt die visuelle Eingabemodalität aus Tabelle 5.10 mit allen definierten Parametern und Status (Tabelle 5.18). Die Modellierung des GUI beschränkt sich dabei auf die Eigenschaften der Komponente, die für die Präsentationsgenerierung wesentlich sind. So werden durch die Selbstbeschreibung in Tabelle 5.18 sowohl eine Komponente zur Steuerung eines Home-Entertainment-Systems (Abbildung 5.16, links) als auch eine Programmauswahl-GUI (Abbildung 5.16, rechts) modelliert.

Neben visuellen Eingaben mittels eines GUI werden im Home-Entertainment-System ebenfalls auditorielle Eingaben mittels Sprache unterstützt. Eine Komponente zur Spracherkennung wird durch die Unimodalität Spracheingabe (Tabelle 5.10) modelliert. Dabei werden alle Status des Modells unterstützt. Die resultierende Modellierung der Komponente findet sich in Tabelle 5.19.

### 5.5 Präsentationsplanung

Wie in Abschnitt 3.2.2 dargestellt, ist der Planungsansatz als wissensbasierter Ansatz dem numerischen Ansatz überlegen, da er aufgrund der deklarativen Formulierung des Präsentationswissens besser wartbar ist. Weiterhin ist der Planungsansatz aufgrund der dynamischen Layout-Generierung einem



Im Beispiel aus Abschnitt 5.2.1 könnte das Ergebnis der Präsentationsplanung z. B. eine Nachricht im SMIL-Format an die Wiedergabekomponente auf dem PDA sein. Die SMIL-Präsentation ist dabei an die Ausgaberesourcen des PDA angepasst und kombiniert den visuellen Teil eines animierten Charakters mit einer Sprachausgabe. Parallel dazu könnte eine Nachricht im SMIL-Format an die Wiedergabekomponente des Fernsehers gesendet werden, die die Ausgabe zusätzlicher Bilder repräsentiert. Die SMIL-Nachricht an den PDA kann z. B. die folgende Form haben:

```
<smil>
  <head>
    <layout>
      <root-layout width="240" height="320"/>
      <region id="reg1" z-index="1"
        width="316" height="175"
        left="0" top="145"/>
    </layout>
  </head>
  <body>
    <par>
      
      
      <audio begin="0.0s" end="24.0s"
        src="http://1.0.0.1/speech.au"/>
    </par>
  </body>
</smil>
```

In der SMIL-Nachricht wird eine an die Auflösung des PDA angepasste Layout-Region definiert (*reg1*), in der alle Bilder dargestellt werden. Für die ersten 24 Sekunden der Präsentation wird eine Sprechanimation des Charakters dargestellt (*speak.gif*), die parallel zur Audioausgabe einer Sprachdatei (*speech.au*) stattfindet. Anschließend wird eine Ruheposition des Charakters dargestellt (*restpose.gif*).

Um diese Nachrichten automatisch als Lösung des Präsentationsproblems generieren zu können, sind zwei Schritte notwendig. Der erste Schritt besteht in der Lösung des gegebenen Planungsproblems. Eine Lösung des Problems wird als Plan bezeichnet (Abschnitt 4.2), der aus einer Folge von Aktionen ( $a_1, \dots, a_k$ ) besteht. Im Falle der Präsentationsplanung im Beispiel aus Ab-

## 5.5. PRÄSENTATIONSPLANUNG

schnitt 5.2.1 stellen diese eine abstrakte Repräsentation der an den PDA und den Fernseher zu versendenden SMIL-Nachrichten dar. Anschließend wandelt eine Ausführungskomponente den Plan in das endgültige SMIL-Format um und versendet die SMIL-Inhalte an die Komponenten auf dem PDA und dem Fernseher.

Um grafisch-zeitliche Layouts für beliebige Bildschirmauflösungen (im obigen Beispiel für die Auflösung  $240 \times 320$ ) generieren zu können, müssen die Operatoren in der Operatorenmenge  $O$  unabhängig von der Auflösung definiert werden. Um dies zu erreichen, folgen wir dem in Abschnitt 4.2 vorgestellten Ansatz aus [André et al., 2005].

Der Ansatz verwendet Constraints (z. B. „Bild X soll über Bild Y dargestellt werden“), um unabhängig von der Auflösung eines Gerätes ein grafisches Layout zu definieren. Die grafisch-zeitlichen Constraints werden mittels der Allen-Intervallalgebra [Allen, 1983] formuliert. Mittels dieser Algebra können geräteunabhängige Constraints für das grafisch-zeitliche Layout von Präsentationen definiert werden.

Durch die Einführung von Constraints in die Operatoren in  $O$  handelt es sich nicht mehr um ein Planungsproblem im klassischen Sinne [Nau et al., 2004]. André und Kollegen kombinieren daher zur Lösung des Planungsproblems einen klassischen Planer mit einem linearen Constraint-Löser, der konkrete Layout-Koordinaten und feste zeitliche Abläufe berechnet, die den Constraints genügen.

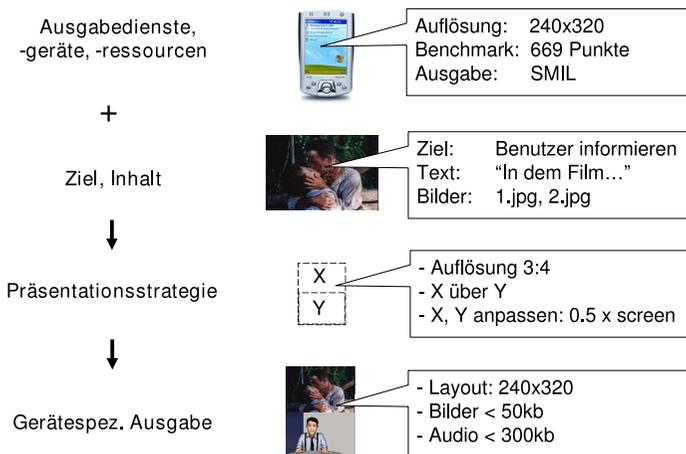


Abbildung 5.18: Ablauf der Präsentationsplanung.

Durch die Kombination von Planer und Constraint-Löser werden konkrete Layout-Koordinaten (z. B. `width="316"`, `height="175"` für die Region `reg1`) sowie feste zeitliche Abläufe (z. B. `begin="24.0s"` und `end="203.0s"` für das Bild `speak.gif`) berechnet, die den Constraints genügen.

Damit ergibt sich im Zusammenhang mit dem Modell für Ausgabedienste, -geräte und -ressourcen der in Abbildung 5.18 dargestellte Ablauf. Ausgabe-geräte und -komponenten senden eine Selbstbeschreibung ihrer Ausgabefähigkeiten an die MMP-Komponente. Diese ist Teil des Startzustands  $s_0$  der Planungskomponente und wird ggf. durch Kontextdaten sowie Ausgabehistorie ergänzt (vgl. Abbildungen 5.5 und 5.6).

Die Präsentationsplanung startet, sobald der Dialog-Manager das Ziel und den Inhalt der Präsentation übermittelt hat. Diese sind Teil des Zielzustands  $g$ . Anschließend löst die MMP-Komponente mittels geräteunabhängiger Präsentationsoperatoren und einem Constraint-Löser das Planungsproblem. Der resultierende Plan wird von einer Ausführungskomponente in ein Format übersetzt, das die Zielkomponente darstellen kann. Anschließend wird der so formatierte Plan an die Zielkomponente versendet. Das Ergebnis ist eine an die Geräteeigenschaften angepasste Präsentation.

### 5.5.2 Vorteile des Ansatzes

In Kombination mit dem Dienst-, Geräte- und Ressourcenmodell für Ein-/Ausgabekomponenten lassen sich verschiedene Präsentationsstrategien definieren, die eine effiziente Präsentationsgenerierung ermöglichen. Die Unterscheidung zwischen den verschiedenen Unimodalitäten einer Komponente (Tabelle 5.8) vereinfacht die Anbindung unterschiedlicher Komponenten mit ähnlichen Eigenschaften. So kann z. B. ein rein visueller Charakter (Tabelle 5.13) mit einer Sprachausgabe (Tabelle 5.15) kombiniert werden. Ansonsten kann er aber genauso behandelt werden wie ein animierter Charakter, der bereits über eine Sprachausgabe verfügt (Tabelle 5.12). Dies erleichtert die Einbindung von neuen Ausgabekomponenten in die Präsentationsgenerierung. Ebenso ist es leicht, neue Komponenten eines bekannten Typs zu unterstützen, z. B. eine weitere SMIL-Wiedergabekomponente.

In der Planung zeigt sich diese erhöhte Ausdruckskraft in einer Reduktion der Planungsoperatoren. Dies soll anhand des Beispielszenarios in Abschnitt 5.1 verdeutlicht werden. Tabelle 5.20 fasst die Ausgabekomponenten des Szenarios zusammen. Dem Szenario wurde ein Multimedia-Server im Keller hinzugefügt, der die Animationen des Charakters zur Verfügung stellt und die Umwandlung von Textdateien in Sprachdateien mittels einer Sprachsynthese durchführt.

## 5.5. PRÄSENTATIONSPLANUNG

Das Szenario umfasst somit acht Ausgabekomponenten. Wird – wie in Abschnitt 5.4.2 illustriert – jede Komponente durch ihren Agentennamen repräsentiert, so sind acht Planungsoperatoren notwendig (vgl. Abbildung 4.2), um die acht unterschiedlichen Typen als Komponenten zu repräsentieren und anzusteuern. Hierbei werden der Einfachheit halber keine Kombinationen von Ausgabekomponenten betrachtet, die weitere Operatoren erfordern. Bei der Modellierung von Komponententypen reduziert sich die Anzahl der Planungsoperatoren auf sechs, da die drei SMIL-Komponenten als ein Komponententyp behandelt werden können.

Tabelle 5.20: Ausgabekomponenten im Beispielszenario.

|                            |              |              |
|----------------------------|--------------|--------------|
| Wohnzimmer                 |              |              |
| Stereoanlage               | Fernseher    | Bilderrahmen |
| MP3                        | SMIL         | SMIL         |
| Mobile Geräte              |              |              |
| PDA                        | Mobiltelefon | Laptop       |
| SMIL                       | MMS          | Charakter 1  |
| Keller                     |              |              |
| Server                     |              |              |
| Charakter 2 Sprachsynthese |              |              |

Da die zwei Charaktere (wie im Beispiel aus Abschnitt 5.1 illustriert) unterschiedliche Fähigkeiten haben, können sie nicht durch denselben Typen behandelt werden. So verfügt Melanies Charakter z. B. nicht über Gesten. Daher wäre allein auf Basis des Typs „Charakter“ unklar, ob bei dem Charakter Gestik eingesetzt werden kann. Aus diesem Grund müssen beide Komponenten durch separate Operatoren modelliert werden.

Bei Verwendung des Modells für Dienst-, Geräte- und Ressourcenmodell für Ein-/Ausgabekomponenten reduziert sich die Anzahl der Planungsoperatoren weiter. Die Wiedergabekomponenten für Medienobjekte von Typ MP3, SMIL oder MMS können als grafisch/akustische Wiedergabekomponenten vom Typ *player* aufgefasst werden (vgl. Abschnitt 5.4.4), die unterschiedliche Inhaltstypen wiedergeben können.

Die zwei Charaktere und die Sprachsynthese können als Komponenten zur Umsetzung eines Benutzerassistenten modelliert werden (Typ *agent*), die sich in den Ausgabeunimodalitäten unterscheiden. Die resultierenden Operatoren sind zwar umfangreicher als die Operatoren bei den zwei vorhergehenden Ansätzen, jedoch ist in ihnen das Präsentationswissen gekapselt und redundante

Definitionen werden vermieden. Damit wird durch Verwendung des Dienst-, Geräte- und Ressourcenmodells die Anzahl der Operatoren von acht auf zwei reduziert. Durch die Kapselung des Wissens in den Operatoren ist es leicht möglich, Strategien zu definieren, die das unerwartete Entfernen von Komponenten vom Typ *agent* oder *player* behandeln. Damit trägt das Modell für Ausgabedienste zu einer fehlertoleranten Realisierung der Systemausgaben von ubiquitären Informationssystemen bei.

Fehlertoleranz ist in ubiquitären Informationssystemen aufgrund des hohen Vernetzungsgrades und der damit verbundenen Gefahr von Instabilitäten von besonderer Wichtigkeit [Edwards & Grinter, 2001]. Typische Fehlerquellen für Komponenten sind Hardware-Fehler, Fehler des Betriebssystems oder logische Fehler in der Komponenten-Software. Diese können zu einer Beeinträchtigung der laufenden Prozesse führen, z. B. dem Verlust der Netzwerkverbindung. Durch die explizite Modellierung von Diensten ist es möglich, im Falle des Versagens von Ausgabekomponenten Ausgaben mittels der verbliebenen Komponenten durchzuführen. Zum Beispiel ist ohne eine Sprachausgabe i. Allg. keine Verwendung des visuellen Teils eines animierten Charakters möglich. Der Inhalt könnte jedoch stattdessen als grafischer Text dargestellt werden.

Weiterhin erleichtert das Modell die Interoperabilität zwischen Komponenten, die auf unterschiedlichen Plattformen realisiert werden. So kann z. B. der auf dem ActiveX-Standard von Microsoft basierende Microsoft Agent-Charakter (Abbildung 5.13) mit Java-basierten Komponenten (z. B. einer grafischen Benutzeroberfläche) koordiniert werden.

Durch die Trennung zwischen domänenspezifischem und domänenunabhängigem Teil wird ebenfalls die Modellierung gänzlich neuer Klassen von Komponenten erleichtert. Somit vereinfacht das Modell die Integration neuer Typen von Applikationen, die Integration neuer Applikationen eines bekannten Typs sowie die Komposition von Applikationen. Dies sind wesentliche Kriterien bei der Realisierung von ubiquitären Informationssystemen [Ranganathan et al., 2005].

### 5.6 Umsetzung des Beispielszenarios

Neben der Modellierung von Ein-/Ausgabediensten und deren Integration in die Präsentationsplanung ergeben sich weitere Herausforderungen für die Umsetzung des in Abschnitt 5.1 dargestellten Beispielszenarios. Da der Schwerpunkt dieser Arbeit jedoch auf der Modellierung der Ein-/Ausgabedienste und deren Integration in die Präsentationsplanung liegt, werden diese Erweiterungen nur skizziert.

## 5.6. UMSETZUNG DES BEISPIELSZENARIOS

### Ausgabedienste und Dialog-Management

Die vorhandenen Rendering-Komponenten und deren Ausgabedienste haben nicht nur Einfluss auf die Form der Ausgabe sondern auch auf die Art des geführten Dialoges. Möchte der Benutzer z. B. einen Krimi im Fernsehen sehen, so könnte eine Ausgabe der zurzeit verfügbaren Krimis auf dem Fernsehbildschirm erfolgen.

Ist jedoch nur eine Sprachausgabekomponente verfügbar, so ist es i. Allg. nicht sinnvoll, dem Benutzer eine lange Liste von Filmtiteln vorzulesen und ihn anschließend zu fragen, welchen Film er sehen möchte. Eine bessere Dialogstrategie könnte darin bestehen, den Benutzer zuerst nach dem gewünschten Genre (z. B. Komödie) zu fragen und dadurch die Auswahl einzuschränken. Für eine begrenzte Anzahl von Filmen können dem Benutzer anschließend die verbleibenden Titel vorgelesen werden.

Die Anpassung der Dialogstrategien erfolgt durch den Dialog-Manager (vgl. Abbildung 5.3). Dies kann dadurch erfolgen, dass der Dialog-Manager analog zur MMP-Komponente die Dialogstrategien in Abhängigkeit des Modells für Ausgabedienste definiert. Mehr Informationen zu der Interaktion zwischen Dialog-Manager und der MMP-Komponente finden sich in [Ludwig, 2004].

### Ausgabedienste und Domänenmodell

Auf der Basis eines wohldefinierten Modells für Ausgabedienste, -geräte und -ressourcen ist es möglich, die Modelle in das Domänenmodell des Dialog-Managers zu integrieren, das die dem Dialogsystem bekannten Konzepte umfasst [Ludwig, 2004]. Der Vorteil dieser Integration wäre, dass sie Benutzern ermöglichen, Anfragen an das Dialogsystem zu stellen, die Ausgabedienste oder -geräte betreffen.

So könnte der Benutzer das System z. B. fragen, welche Bildschirme in der intelligenten Umgebung zur Verfügung stehen. Weiterhin könnte der Benutzer fragen, welche Bilder auf dem Bildschirm vor ihm angezeigt werden. Die Integration des Modells von Ausgabediensten in das Domänenmodell eines Dialogsystems kann somit einen Beitrag zur Selbsterklärung von Dialogsystemen leisten.

### Inkrementelle Planung

Inkrementelle Planung ist ein Planungsansatz, der zum tragen kommt, falls Rendering-Komponenten dem System während der Planungszeit entfernt oder hinzugefügt werden. Die Planungszeit bezeichnet den Zeitraum, in der der die Präsentationsoperatoren  $o$  durch den Planer bereits auf den Startzustand  $s_0$  angewendet werden, aber noch kein Zielzustand  $g$  erreicht wurde (vgl. Kapitel 4.2).

Erfolgt die Entfernung einer Rendering-Komponente während der Planungszeit, so erfolgt bei der MMP-Komponente eine Fehlermeldung bei der Ausführung des generierten Planes (z. B. während des Versendens einer Nachricht an eine bereits entfernte Rendering-Komponente). Der Präsentationsplaner im MMP-Komponente reagiert darauf mit einer Neuplanung, d. h. der gesamte Planungsvorgang wird mittels eines neuen Startzustands  $s'_0$ , der die geänderte Menge an Rendering-Komponenten berücksichtigt, vollständig wiederholt.

Ein inkrementeller Planungsansatz stellt eine effiziente Alternative zu einer vollständigen Neuplanung dar. Eine inkrementelle Planung ermöglicht es, nur die Teile des Planungsvorgangs rückgängig zu machen, die durch die Änderung in der Menge der Rendering-Komponenten beeinflusst werden.

Wird z. B. während des Planungsvorgangs eine Sprachausgabe beendet und dies dem MMP-Modul bekannt gemacht (z. B. durch den Verlust der Netzwerkverbindung), müsste lediglich der letzte Planungsschritt rückgängig gemacht werden, durch den die Ausgabe an die Sprachausgabe initialisiert wurde. Der zu sprechenden Text könnte stattdessen visuell von einer anderen Rendering-Komponente auf einem Bildschirm dargestellt werden. Der Vorteil dieser Vorgehensweise liegt in dem Effizienzgewinn, der sich im Vergleich zu einer vollständigen Neuplanung ergibt. Weitere Informationen zu inkrementeller Ausgabegenerierung finden sich in [Thórisson, 1999; Kilger, 1994].

### Kreuzmodale und geräteübergreifende Referenzen

In Dialogsystemen werden mit Referenzen Verweise auf andere Präsentationssteile bezeichnet. Diese werden in unterschiedlichen Modalitäten i. Allg. unterschiedlich umgesetzt. Eine Referenz innerhalb eines Textes kann z. B. durch das den Ausdruck „Der im vorigen Abschnitt beschriebene Mann“ erfolgen. Eine Referenz innerhalb eines Diagramms kann z. B. durch einen Pfeil erfolgen.

In multi-modalen Dialogsystemen können Referenzen auf modalitätsübergreifend eingesetzt werden [André, 1995]. So kann z. B. mittels eines grafischen Textes die Referenz „der im Bild dargestellte Baum“ auf ein Bild erfolgen. Im Beispielszenario in Abschnitt 5.6 ist der Satz „Das hier ist der Schauspieler Jürgen Vogel“ ebenfalls eine kreuzmodale Referenz des animierten Charakters auf ein Bild.

In ubiquitären Informationssystemen können Referenzen zusätzlich geräteübergreifend erfolgen. So könnte z. B. im obigen Beispiel der Charakter auf einem Laptop dargestellt werden und die Bilder des Schauspielers Jürgen Vogel auf dem Fernseher. Das System könnte die Bilder in diesem Fall mit dem sprachlichen Ausdruck „Auf dem Fernseher siehst du Bilder von Jürgen Vogel“ referenzieren.

## 5.7. ZUSAMMENFASSUNG

Um kreuzmodale und geräteübergreifende Referenzen umzusetzen, sind die folgenden Schritte nötig. Die MMP-Komponente protokolliert erfolgte Ausgaben, die den dargestellten Inhalt beschreiben, in der Ausgabehistorie (vgl. Abbildung 5.5). Dort kann der Dialog-Manager auf die Informationen zugreifen und auf dargestellte Inhalte Bezug nehmen. Falls wie oben beschrieben das Modell für Ausgabedienste und -geräte Teil des Domänenmodells des Dialog-Managers ist, können die Ausgaben in Bezug auf verwendete Modalitäten und Geräte annotiert werden. Dies ermöglicht es dem Dialog-Manager die gewünschten Referenzen kreuzmodal und geräteübergreifend zu generieren.

### 5.7 Zusammenfassung

Zur automatisierten Generierung von Präsentationen in ubiquitären Informationssystemen werden Modelle für die Dienste, Geräte und Ressourcen von Ein-/Ausgabekomponenten benötigt. In dieser Arbeit wurden diese als statische Modelle definiert. Dabei wurde zwischen domänenunabhängigen und domänenspezifischen Modelleigenschaften unterschieden, da ein vollständiges und eindeutiges Modell multi-modaler Ausgabedienste nicht praktikabel ist. Diese Modellierung wurde exemplarisch für eine Home-Entertainment-Domäne durchgeführt. Das Modell kann aufgrund der Trennung zwischen domänenunabhängigen und -spezifischen Eigenschaften jedoch leicht an neue Domänen angepasst werden.

Durch die Modellierung einzelner Unimodalitäten wird die Koordination von unterschiedlichen Ausgabekomponenten desselben Typs (z. B. verschiedener Instanzen von animierten Charakteren) erleichtert. Weiterhin erleichtert das Modell die Koordination von Ausgabekomponenten, die auf unterschiedlichen Plattformen (z. B. Java und ActiveX) implementiert wurden. Durch diesen Ansatz ist eine Verallgemeinerung der Operatoren des Präsentationsplaners in der MMP-Komponente möglich. Im Beispielszenario ließ sich die Anzahl der Planungsoperatoren durch Verwendung des Modells um den Faktor vier im Vergleich zu einer modellfreien Präsentationsplanung reduzieren.

Das Modell ermöglicht dadurch eine flexiblere Ad-hoc-Anbindung neuer Ausgabekomponenten. Dies erhöht ebenfalls die Stabilität des Systems, da das System auf den Ausfall einzelner Ausgabekomponenten (z. B. aufgrund eines Netzwerkfehlers) reagieren und die Ausgabe mittels der verbleibenden Komponenten durchführen kann.



## 6 Implementierung

Das in Abschnitt 5.4.4 definierte Dienst-, Geräte- und Ressourcenmodell für Ein-/Ausgabekomponenten wurde in zwei multi-modalen Dialogsystemen verifiziert. Eine erste Implementierung erfolgte als Teil des Embassi-Systems [Herfet et al., 2001]. Der Schwerpunkt der Implementierung war die Koordination einer Sprachausgabekomponente, eines animierten Charakters und zweier GUIs. Die generierten Präsentationen des Systems waren jeweils auf ein Gerät beschränkt. Die Präsentationsplanung erfolgte durch den in Abschnitt 5.5 vorgestellten Planungsansatz. Jedoch kamen hier keine Constraints zum Einsatz.

Die zweite Implementierung erfolgte innerhalb des Dynamite-Projekts [Elting & Hellenschmidt, 2004] und wird in Abschnitt 6.2 vorgestellt. Die Implementierung realisiert neben einer erweiterten Umsetzung des Modells eine vollständige Umsetzung des in Abschnitt 5.5 vorgestellten Planungsansatzes. Das System umfasst drei Ausgabegeräte, deren Ausgaben gekoppelt werden können und passt Präsentationen an die verfügbare Menge von Ausgabegeräten und -komponenten an.

### 6.1 Prototypimplementierung

Das Embassi-System realisiert eine multi-modale Schnittstelle zu Home-Entertainment-Diensten und -Geräten und war die erste Umgebung, innerhalb der das in dieser Arbeit definierte Dienst-, Geräte- und Ressourcenmodell für Ein-/Ausgabekomponenten umgesetzt wurde. Die Hauptfunktionalität des Systems war eine sprachgesteuerte Suche nach Programminformationen, die mittels eines elektronischen Programmführers aus dem Fernsehsignal entnommen wurden.

#### 6.1.1 Überblick

Die Präsentationsaufgabe in diesem System bestand in der Koordination eines Programmauswahl-GUI und eines animierten Charakters [Rieger & Berner, 2002] auf dem Fernsehbildschirm. Ein zweites GUI konnte auf einem mobilen Gerät aktiviert werden, wodurch das GUI auf dem Fernseher ausgeblendet werden sollte (Abbildung 6.1). Abbildung 6.2 gibt ein Beispiel für eine multi-modale Ausgabe auf dem Fernseher.

Die Architektur des Embassi-Systems wurde bereits in Abschnitt 4.1.1, Abbildung 4.1 beschrieben. Die Aufgabe der MMP-Komponente, die im Rahmen dieser Arbeit realisiert wurde, ist die Layout-Koordination der Textgenerierung, der Sprachsynthese, der Programmauswahl-GUI und des animierten

## KAPITEL 6. IMPLEMENTIERUNG

Charakters. Dabei wurden die Textgenerierung und die Sprachausgabe als eine Konzept-nach-Sprache Einheit aufgefasst. Zusätzlich kann ein weiteres, mobiles GUI auf einem PDA vorhanden sein, mit dem Programminformationen abgefragt werden können, ohne das Fernsehbild zu verdecken. Um die Ausgabe fehlertolerant und die Software wieder verwertbar zu gestalten, wurde für den Charakter, die Sprachausgabe und das GUI auf die in Abschnitt 5.4.4 definierten Modelle zurückgegriffen.



Abbildung 6.1: Mobiles GUI in Embassi (Foto: European Media Laboratory).

## 6.1. PROTOTYPIMPLEMENTIERUNG

Die Umsetzung der MMP-Komponente innerhalb des Projektes fand in Kooperation mit der Sony International (Europe) GmbH statt. In dieser Arbeit werden nur die eigenen Arbeiten beschrieben. Weitere Informationen über die Realisierung der MMP-Komponente finden sich in [Elting et al., 2003].

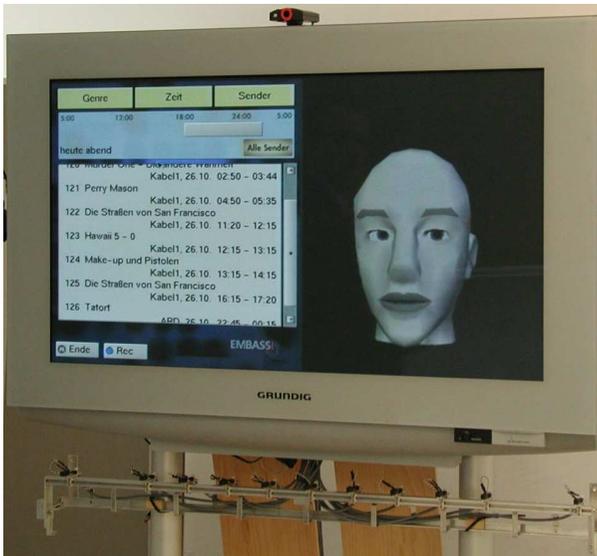


Abbildung 6.2: GUI und animierter Charakter in Embassi (Foto: European Media Laboratory).

### 6.1.2 Modelle

In Embassi kamen drei Modelle zum Einsatz. Neben dem Modell für Ausgabebedienste, -geräte und -ressourcen wurden der Ausgabeinhalt und das Ausgabeziel modelliert (vgl. Abbildung 5.4). Alle drei Modelle gehen in die Präsentationsplanung ein.

Ausgabebedienste, -geräte und -ressourcen

Der animierte Charakter wird mittels der in Abbildung 6.3 dargestellten Selbstbeschreibung im XML-Format modelliert. Die Selbstbeschreibung folgt dem in Tabelle 5.12 definierten Modell. Die Komponente hat den multi-modalen Typ `agent (Element RenderComponent)`. Von dem Modell werden in dieser Implementierung jedoch nur die Unimodalitäten modelliert sowie deren Definition mit Hilfe der Bernsen-Eigenschaften (`Unimodality-Element`).

## KAPITEL 6. IMPLEMENTIERUNG

nte, vgl. Tabelle 5.7). Weiterhin wird der Fernsehbildschirm als visuelles Ausgabegerät modelliert (Element `OutputDevice`).

Der Parameter *behaviour* modelliert das Verhalten des animierten Charakters. Dieser Parameter wird jedoch nicht explizit modelliert, sondern implizit Komponenten vom Typ *agent* mit visuellem Feedback zugeordnet. Dabei wurde das in Tabelle 5.4 definierte Modell für das Verhalten des animierten Charakters verwendet.

Das Ausgabegerät dient dazu, Layout-Konflikte zwischen GUI und animiertem Charakter festzustellen, falls beide auf demselben visuellen Gerät dargestellt werden. Um zwischen einer stationären GUI-Komponente im Wohnzimmer und einer mobilen GUI-Komponente unterscheiden zu können wird weiterhin der Status *has-room-id* unterstützt (Attribut `location`).

In diesem Prototypen wurden im Unterschied zu dem in Abschnitt 5.4.4 definierten Modell keine Eingabekomponenten modelliert. Im Unterschied zur Modellierung der GUI in Tabelle 5.18 wurde das GUI als Ausgabekomponente modelliert, die grafischen Text ausgibt (Abbildung 6.4).

```
- <RenderingAgentProfile>
- <RenderComponent name="543" type="agent">
  <Unimodality name="character-gaze" linguistic="-li"
    analogue="an" arbitrary="-ar" stadyn="dyn"
    medium="vis" subtype="gaze" />
  <Unimodality name="character-head" linguistic="-li"
    analogue="an" arbitrary="-ar" stadyn="dyn"
    medium="vis" subtype="head" />
  <Unimodality name="character-facial-expression"
    linguistic="-li" analogue="an" arbitrary="-ar"
    stadyn="dyn" medium="vis" subtype="facial-
    expression" />
  <Unimodality name="character-lip-movements"
    linguistic="li" analogue="an" arbitrary="-ar"
    stadyn="dyn" medium="vis" subtype="spoken-text" />
</RenderComponent>
<OutputDevice name="tv1-screen" device="tv1"
  medium="vis" location="living-room" maxX="1024"
  maxY="768" />
</RenderingAgentProfile>
```

Abbildung 6.3: Selbstbeschreibung der Charakterkomponente.

Die mobile Variante auf dem PDA hat anstelle des Bezeichners `living-room` den Bezeichner `mobile`. Dies dient dazu, das GUI auf dem Fernseher und das GUI auf dem PDA unterscheiden zu können.

## 6.1. PROTOTYPIMPLEMENTIERUNG

```
- <RenderingAgentProfile>
- <RenderComponent name="95" type="player">
  <Unimodality name="graphic-text" linguistic="li" analogue="-an" arbitrary="-ar"
    stadyn="sta" medium="vis" subtype="written-text" />
  </RenderComponent>
  <OutputDevice name="tv1-screen" device="tv1" medium="vis" location="living-
    room" maxX="1024" maxY="768" />
</RenderingAgentProfile>
```

Abbildung 6.4: Selbstbeschreibung der Programmauswahl-GUI auf dem Fernseher.

Die Sprachausgabe wurde analog zu den Definitionen in den Tabellen 5.7 und 5.15 modelliert als linguistische, nicht-analoge, nicht-arbiträre, audiotrielle Unimodalität vom Typ *agent*. Anhand der Unimodalitäten kann die Sprachausgabe als solche identifiziert werden und mit der Charakterkomponente kombiniert werden.

```
- <RenderingAgentProfile>
- <RenderComponent name="34" type="agent">
  <Unimodality name="speech-output" linguistic="li" analogue="-an"
    arbitrary="-ar" stadyn="dyn" medium="aud" subtype="spoken-text" />
  </RenderComponent>
  <OutputDevice name="tv1-speakers" device="tv1" medium="aud"
    location="living-room" />
</RenderingAgentProfile>
```

Abbildung 6.5: Selbstbeschreibung der Sprachausgabekomponente.

### Ausgabeinhalt und -ziele

Der Inhalt der Ausgabe wird zusammen mit dem Ziel der Ausgabe vom Dialog-Manager an die MMP-Komponente gesendet. Der Inhalt besteht aus Konzepten, die mittels einer Ontologie formalisiert wurden. Nähere Informationen hierzu finden sich in [Ludwig, 2004]. Das Ausgabeziel wird mit Hilfe von Dialogakten modelliert.

Die unterstützten Dialogakte, aus denen der Parameter *behaviour* des animierten Charakters abgeleitet wurde, finden sich in Tabelle 6.1. Diese Abbildung wurde in [Rieger, 2003] definiert und wird durch den Präsentationsplaner ausgeführt. Dabei wird das in Tabelle 5.4 definierte Modell für die Verhaltensweisen des animierten Charakters verwendet.

## KAPITEL 6. IMPLEMENTIERUNG

Tabelle 6.1: Ausgabeziele und Verhaltensweisen des animierten Charakters.

| Dialogakt                 | Status  | Bedeutung           | Verhaltensweise   |
|---------------------------|---------|---------------------|-------------------|
| message-greeting          | -       | Begrüßung           | friendly-greeting |
| message-closing           | -       | Verabschiedung      | friendly-closing  |
| message-inform            | warning | Warnung             | show-importance   |
| message-inform            | busy    | System ausgelastet  | regretful         |
| message-inform            | idle    | System wartet       | regretful         |
| message-inform            | error   | unerwarteter Fehler | sorry             |
| message-inform            | ok      | Erfolg              | positive-neutral  |
| message-inform            | failed  | Fehlschlag          | sorry             |
| message-inform            | offer   | Angebot             | friendly-offer    |
| message-correct           | -       | Korrektur           | neutral           |
| message-command           | -       | Aufforderung        | show-importance   |
| message-accept            | -       | Anfrage akzeptiert  | show-agreement    |
| message-reject            | -       | Anfrage abgelehnt   | show-rejection    |
| message-cancel            | -       | Abbruch             | alert             |
| message-acknowledge       | -       | Anfrage verstanden  | happy             |
| query-input               | -       | Eingabe benötigt    | interested        |
| query-yes-no              | -       | Ja/Nein-Frage       | interested        |
| query-request-acknowledge | -       | Nachfrage           | interested        |
| query-request-repair      | -       | Reparaturanfrage    | interested        |
| query-selection           | -       | Auswahl treffen     | show-alternatives |

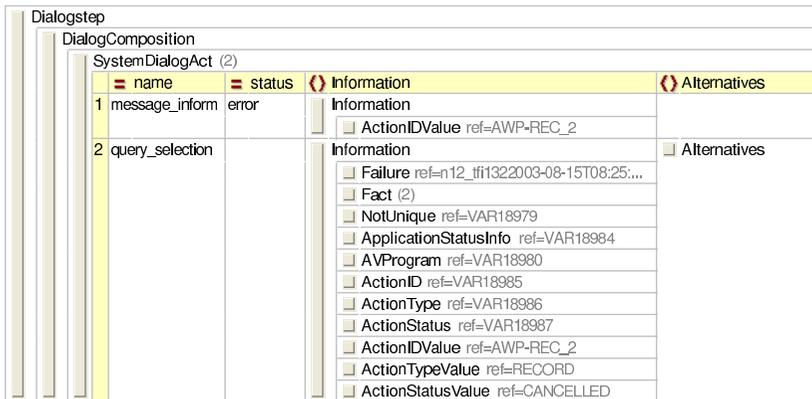


Abbildung 6.6: Ausgabeinhalt mit Dialogakten.

## 6.1. PROTOTYPIMPLEMENTIERUNG

In Abbildung 6.6 findet sich ein Beispiel für einen Ausgabeinhalt. Der Übersicht halber wurden die Daten im XML-Format in reduzierter Form dargestellt. Das Element `SystemDialogAct` enthält die zwei Dialogakte, aus denen die Ausgabe besteht. Der Inhalt behandelt die Fehlermeldung, dass die Eingabe „Ich möchte einen Krimi aufnehmen“ des Benutzers nicht eindeutig bearbeitet werden konnte und der Benutzer daher gebeten wird, aus den verfügbaren Lösungen zu wählen.

### 6.1.3 Präsentationsplanung

Die Präsentationsplanung erfolgt in zwei Planungsschritten. In einem ersten Schritt werden auf Basis der vorhandenen Ausgabedienste der animierte Charakter und die Sprachausgabe zu einer Einheit kombiniert. In einem zweiten Schritt werden aufgrund dieser Kombinationen von Komponenten ein Layout bestimmt und die Parameter mit Werten belegt.

Im Unterschied zum Ansatz aus Abschnitt 5.5 kamen bei der Präsentationsplanung keine Constraints zur Anwendung. Die Implementierung erfolgte unter Verwendung des Ckuckuk-Präsentationsplaners ([André et al., 2000], vgl. Abschnitt 4.2). Mittels des Ansatzes wurden die folgenden Adaptionen realisiert:

1. Ist nur die Sprachausgabekomponente aktiv, so erfolgt bei der Ausgabe aus Abbildung 6.6 die Sprachausgabe „Bitte seien Sie präziser. Welche Sendung soll aufgenommen werden: 'Zeitspiegel' auf Kanal BR, 'Rundschau-Magazin' auf Kanal BR, ...?“
2. Wird zusätzlich das GUI auf dem Fernseher oder dem PDA aktiviert, so erfolgt eine kreuzmodale Referenz auf das GUI: „Bitte seien Sie präziser. Bitte wählen Sie einen Eintrag aus der angezeigten Liste aus.“
3. Wird zusätzlich der animierte Charakter aktiviert, so werden der animierte Charakter und das GUI auf dem Fernseher in ein gemeinsames Layout integriert. Dabei nimmt jede Komponente eine Bildschirmhälfte ein (vgl. Abbildung 6.2). Ist dort kein GUI vorhanden, so nimmt der animierte Charakter den gesamten Bildschirm ein.
4. Wird die Sprachausgabe deaktiviert, so wird auch der animierte Charakter deaktiviert, und das GUI nimmt den gesamten Bildschirm ein.
5. Wird das GUI auf dem PDA gestartet, so wird das GUI auf dem Fernseher ausgeblendet.

Das Layout bestand dabei entweder aus einer senkrechten Bildschirmteilung zwischen animiertem Charakter und des Programmauswahl-GUI (Abbildung

6.2), der Verwendung des gesamten Bildschirmes durch eine der Komponenten oder der Ausblendung beider Komponenten. Mittels dieses Layouts konnte die Sprachausgabekomponente ebenfalls die kreuzmodalen Referenzen auf das GUI generieren. Dabei wurde die kreuzmodale Referenz „angezeigte Liste“ (Punkt 2.) durch eine Textgenerierungskomponente erzeugt. Dies erfolgte, indem die Komponente Informationen über die an der aktuellen Ausgabe beteiligten Rendering-Komponenten und deren Ausgaben (in diesem Fall Programminformationen auf dem GUI) erhalten hat.

### 6.2 Home-Entertainment mit gekoppelten Ausgabegeräten

Aufbauend auf der Prototypimplementierung im Embassi-System wurde die Implementierung im Dynamite-System durch komplexere grafisch-akustische Layouts erweitert. Neben dem PDA und dem Fernseher wurde mit einem elektronischen Bilderrahmen ein weiteres Ausgabegerät eingeführt. Die Modellierung der Ausgabegeräte wurde erweitert und durch die Modellierung von Eingabekomponenten ergänzt.

#### 6.2.1 Überblick

Das Home-Entertainment-System setzt sich aus einem Server und drei Ausgabegeräten zusammen. Die drei Ausgabegeräte können beliebig gekoppelt werden. Ein iPaq hx2750 PDA mit Windows Mobile 2003 ermöglicht es, das System mittels eines GUI zu bedienen sowie Multimediainformationen im SMIL-Format anzuzeigen (Abbildung 6.7).

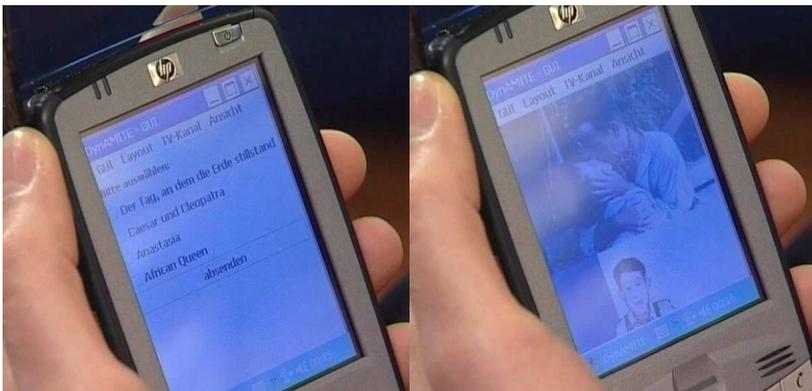


Abbildung 6.7: GUI-Eingabe und animierter Charakter auf PDA (Fotos: European Media Laboratory).

## 6.2. HOME-ENTERTAINMENT MIT GEKOPPELTEN AUSGABEGERÄTEN

Ein Loewe Aconda 9272 Fernseher mit OnlinePlus-Zusatz und eingebettetem Linux-System wird zum Darstellen weiterer Informationen im SMIL-Format verwendet. Als Zusatzbildschirm steht ein 17 Zoll großer elektronischer Bilderrahmen (Digiframe DF-1710) mit eingebettetem Linux-System zur Verfügung, der Bilder und Audiodateien ausgibt (Abbildung 6.8). Auf einem leistungsstarken Server (3,4 GHz Pentium 4 mit 512 MB RAM) werden die ressourcenintensiven Applikationen zur Sprachgenerierung oder Spracherkennung ausgeführt.



Abbildung 6.8: Elektronischer Bilderrahmen und Fernseher mit animiertem Charakter.

Als Applikation wurde ein Filminformationssystem realisiert. Der Benutzer kann zwischen mehreren Filmen wählen, für die Multimediainformationen vorliegen. Die Multimediainformationen beinhalten textuelle Zusammenfassungen und Standbilder der Filme. Das System ist in der Lage, die Ausgabe der multimedialen Filminformationen auf die verfügbaren Ausgabegeräte zu verteilen und an die verwendeten Geräte automatisch anzupassen. Durch das GUI auf dem PDA oder durch Ausschalten kann jedes Gerät zu jeder Zeit deaktiviert werden. Weiterhin hat der Benutzer je nach Präferenz die Möglichkeit, zwischen einem text-basierten Layout und einem charakter-basierten Layout zu wählen.

Neben der GUI auf dem PDA kann der Benutzer das System mittels eines drahtlosen Mikrophons auch per Sprache bedienen. Dabei wird beispielsweise ein Dialog der folgenden Form geführt:

Benutzer: „Kinofilminformationen, bitte!“

## KAPITEL 6. IMPLEMENTIERUNG

System: „Zu welchem Film möchten Sie Kinofilminformationen haben?“

Benutzer: „Anastasia“

Anschließend erfolgt die Präsentation des Filminhalts. Vier Beispiele für die präsentierten Filminhalte finden sich in Anhang A.1. Für den Fall, dass nicht alle Medien durch die erste Präsentation dargestellt werden können, kann sich der Benutzer mittels der Kommandos „weiter“ und „abbrechen“ weitere Bilder zu dem Film anzeigen lassen oder die Präsentation abbrechen. Der Dialog kann ebenfalls per Spracheingabe oder visuell auf dem PDA mittels Stifteingabe auf dem Touchscreen erfolgen (vgl. Abbildung 6.7, linkes Bild).

### 6.2.2 Präsentationen

Mittels des Systems können verschiedene Ausgabegeräte und Ausgabekomponenten gekoppelt werden. Wird nur der PDA verwendet, so werden der animierte Charakter und ein Bild zusammen mit der Sprachausgabe auf dem PDA dargestellt (Abbildung 6.7, rechts). Wird der PDA mit dem Fernseher gekoppelt, so wird der zusätzliche Layout-Platz dazu verwendet, um zwei weitere Bilder darzustellen (Abbildung 6.9).



Abbildung 6.9: PDA und Fernseher mit animiertem Charakter.

## 6.2. HOME-ENTERTAINMENT MIT GEKOPPELTEN AUSGABEGERÄTEN

Anstelle des Charakters und der Sprachausgabe kann auch grafischer Text verwendet werden. In diesem Fall wird auf dem PDA kein drittes Bild dargestellt. Wird der elektronische Bilderrahmen hinzugefügt, so wird durch den Bilderrahmen ein weiteres Bild dargestellt. In diesem Fall wird der Übersichtlichkeit halber auf dem Fernseher nur ein Bild dargestellt. Abbildung 6.10 zeigt die drei Geräte im Text-Layout.

Wird der PDA ausgeschaltet, so wechselt der animierte Charakter auf den Fernseher (Abbildung 6.8). Werden Fernseher und PDA ausgeschaltet, so wird auf dem elektronischen Bilderrahmen ein Bild angezeigt und die Sprachausgabe über dessen Lautsprecher ausgegeben.

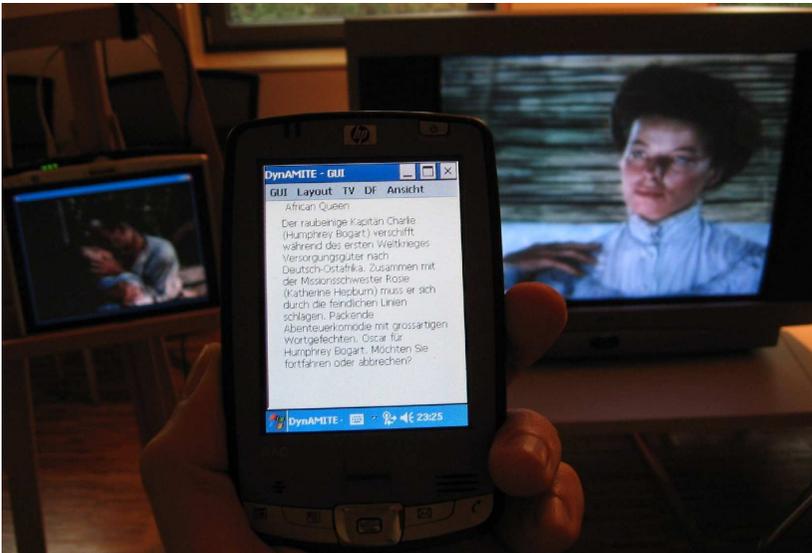


Abbildung 6.10: Bilderrahmen, PDA und Fernseher im Text-Layout.

Neben dem Ausschalten von Geräten reagiert das System auch auf das Ausschalten einzelner Komponenten. Dies kann manuell durch Beenden der Prozesse der Komponente oder aufgrund von Hardware- oder Software-Fehlern geschehen.

Wird die Charakter-Komponente beendet, so wird ein Layout ohne animierten Charakter realisiert. Daher werden Bilder auf dem vollen Bildschirm dargestellt. Werden ebenfalls die beiden SMIL-Komponenten beendet, so wird der

elektronische Bilderrahmen für Bild- und Sprachausgaben verwendet. Wird die Sprachausgabe beendet, so wird grafischer Text dargestellt. Weitere Beispiele zu den vom Home-Entertainment-System generierten Präsentationen finden sich in Anhang A.4.

### 6.2.3 Architektur

Abbildung 6.11 zeigt die Architektur des Home-Entertainment-Systems und realisiert ein verteiltes, multi-modales Dialogsystem mit vier Geräten (vgl. Abbildung 5.3). Auf dem PDA befindet sich eine GUI-Komponente  $I_1$ , mit der der Benutzer zwischen den Filmen wählen kann. Weiterhin befindet sich auf dem PDA eine SMIL-Komponente, die beliebige Nachrichten im SMIL 2.0 Basic Standard darstellen kann.

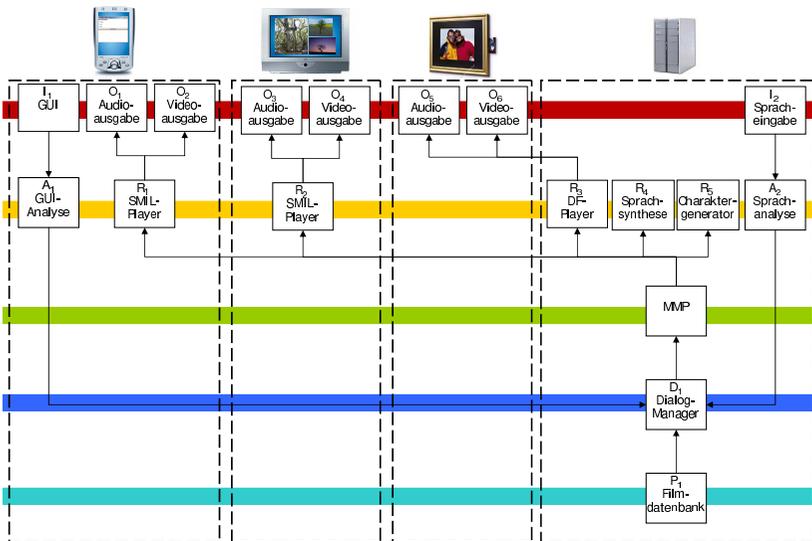


Abbildung 6.11: Architektur des Home-Entertainment-Systems.

Beide Komponenten wurde mittels Java 2 Micro Edition Personal Profile realisiert. Bei der SMIL-Komponente kommt zusätzlich das Java Media Framework sowie die Xsmiles 0.6 Software<sup>1</sup> zum Einsatz. Die Analysekomponente  $A_1$  des GUI beschränkt sich auf die Verarbeitung des durch den Benutzer ausgelösten Eingabeereignisses. Die jeweilige Benutzerwahl wird an den Dialog-

<sup>1</sup>Xsmiles - An Open XML Browser for Exotic Devices, <http://www.xsmiles.org>.

## 6.2. HOME-ENTERTAINMENT MIT GEKOPPELTEN AUSGABEGERÄTEN

Manager  $D_1$  weitergeleitet, der mittels Java SE 1.4 auf dem Server unter Windows XP realisiert wurde.

Mittels des GUI auf dem PDA kann der Benutzer ebenfalls den Fernseher fernsteuern und das aktuelle Programm sowie die Lautstärke ändern. Auf dem Fernseher wurde auf Basis des eingebetteten Linux-Systems und Java SE 1.3.1 eine zur PDA-Implementierung identische SMIL-Komponente  $R_2$  mittels der Xsmiles-Software realisiert. Das Linux-System zeigt die SMIL-Ausgaben auf dem AV-Kanal des Fernsehers. Wird der Kanal durch den Benutzer gewechselt, so deaktiviert sich die Komponente und steht für Ausgaben nicht zur Verfügung.

Auf dem Server befindet sich die Komponente  $I_2$ , die die Spracherkennung und -verarbeitung realisiert. Diese erfolgt mittels IBM Via Voice 10 mittels der Java Speech API-Schnittstelle. Die erkannten Sätze wurden mittels eines Keyword-Parsers analysiert (Komponente  $A_2$ ) und die erkannten Konzepte an den Dialog-Manager zurückgeliefert. Im Unterschied zu Abbildung 5.3 wurde auf die Realisierung einer Komponente zur multi-modalen Fusion (MMF) verzichtet, da diese kein Schwerpunkt dieser Arbeit war.

Für die Sprachsynthese ( $R_4$ ) wurden zwei alternative Implementierungen realisiert. Die erste Implementierung basiert auf der Mbrola Sprachsynthese<sup>2</sup>. Die Nachfolgeversion verwendet die AT&T Natural Voices 1.4 Sprachausgabe unter Verwendung der Java Speech API. Die Eingabe für die Sprachausgabe besteht aus Text, der mittels der Sampa Annotation [Chan et al., 1995] lautschriftlich annotiert wird. Die Sprachsynthesekomponente liefert als Ergebnis eine qualitativ besser wertige Datei im WAV-Format mit einer Abtastfrequenz von 22.050 Hz zurück sowie eine qualitativ schlechter wertige Datei im AU-Format mit 12.000 Hz (vgl. Tabelle 5.16). Die Komponente  $R_5$  realisiert einen animierten Charakter mittels animierter GIF-Bilder (vgl. Tabelle 5.13). Diese Bilder werden mittels des in Tabelle 5.6 definierten Modells für visuelle Medienobjekte annotiert.

Weiterhin befindet sich eine Komponente zur Ansteuerung des elektronischen Bilderrahmens ( $R_3$ ) auf dem Server. Die Komponente befindet sich nicht lokal auf dem Bilderrahmen, da die Kommunikation über ein Netzwerkverzeichnis erfolgt, in das Bild- und Audiodaten geschrieben werden müssen. Die Filmdatenbankapplikation ( $P_1$ ) befindet sich ebenfalls als zentrale Komponente auf dem Server. Hierbei wurde eine frei verfügbare, Microsoft Access-basierte Filmdatenbank<sup>3</sup> verwendet, die 11.030 Filmtitel und deren Inhaltsangaben umfasst. Die Filmdatenbank wurde für eine begrenzte Auswahl von Filmen

---

<sup>2</sup>The Mbrola Project, <http://tcts.fpms.ac.be/synthesis/mbrola>.

<sup>3</sup>Filmdatenbank - Kinofilme online, <http://www.filmdb.de>.

durch Bilder erweitert. Weiterhin wurden Fremdwörter mittels der Sampa Annotation annotiert, um eine korrekte Sprachausgabe zu ermöglichen.

Die MMP-Komponente wurde unter Java SE 1.4 implementiert. Dabei wurde der in [André et al., 2005] vorgestellte Ansatz zur Präsentationsplanung verwendet und mittels der Ckuckuk-Distribution implementiert [André et al., 2000]. Das System stellt eine gemeinsame Arbeit innerhalb des Projekts Dynamite<sup>4</sup> dar. An der Realisierung des Systems waren die European Media Laboratory GmbH, an der die hier beschriebenen Forschungsarbeiten durchgeführt wurden, und das Fraunhofer Institut für grafische Datenverarbeitung beteiligt. Der Schwerpunkt des Fraunhofer Instituts war die Realisierung der Middleware, die nicht Thema dieser Arbeit ist und in [Elting & Hellenschmidt, 2004] dargestellt wird. Das Fraunhofer Institut hat weiterhin Initialversionen der GUI-Komponenten ( $I_1$  und  $A_1$ ) sowie des Dialog-Managers  $D_1$  implementiert, die jedoch im Rahmen dieser Arbeit wesentlich erweitert wurden. Die Implementierung aller weiteren Komponenten sowie die Umsetzung auf den verwendeten Geräten fanden ausschließlich im Rahmen dieser Arbeit statt.

### 6.2.4 Modelle

Das in Abschnitt 5.4.4 beschriebene Modell wurde in Form einer DTD für die Selbstbeschreibungen der Ein-/Ausgabekomponenten des Systems sowie einer darauf aufbauenden Java-Architektur umgesetzt. Die Java-Werkzeuge erlauben das automatische Einlesen der Selbstbeschreibungen aus XML-Dateien sowie das dynamische Setzen der Status und die Rücktransformation nach XML. Eine UML-Darstellung der Architektur findet sich in Anhang A.2. Dabei wurden im Vergleich zu dem in Abschnitt 5.4.3 und Abschnitt 5.4.4 definierten Modell Änderungen vorgenommen. Die Ressourceneinschränkungen des PDAs wurden nicht wie im domänenunabhängigen Modell (Tabelle 5.1) beschrieben mittels Kennwerten modelliert (Status *has-networking-benchmark* und *has-graphics-benchmark*). Stattdessen wurde der Einfachheit halber von der Größe des Bildschirms auf die Grafik- und Netzwerkfähigkeiten geschlossen. Es wurde angenommen, dass eine geringe Bildschirmauflösung ein Gerät mit geringen Grafik- und Netzwerkfähigkeiten impliziert, für das bevorzugt Medien in reduzierter Qualität verwendet werden sollen.

Weiterhin werden die Lautstärke der Applikation (Status *has-application-volume*) und die Lautstärke des Ausgabegeräts (Status *has-device-volume*, *has-min-device-volume* und *has-max-device-volume*) nicht modelliert. Da im Home-Entertainment-System die Benutzerposition als Kontext nicht modelliert wurde, wird weiterhin der Status *has-room-id* nicht modelliert.

---

<sup>4</sup>Dynamite Website, <http://www.dynamite-project.org>.

## 6.2. HOME-ENTERTAINMENT MIT GEKOPPELTEN AUSGABEGERÄTEN

Dienste, Geräte und Ressourcen von Ein-/Ausgabekomponenten

Abbildung 6.12 zeigt die Modellierung der SMIL-Wiedergabekomponente auf dem PDA (vgl. Tabelle 5.17). Der multi-modale Typ *player* wird als Attribut `type` des Elements `RenderComponent` gesetzt. Der Parameter *media-url* wird dem Typ nicht explizit zugeordnet, sondern implizit mit ihm assoziiert.

Dasselbe gilt für die Parameter der Unimodalitäten. Die Parameter werden in der Nachricht an die SMIL-Komponente gesetzt, die das Ergebnis der Präsentationsplanung ist. Die Unimodalitäten Audio und Video werden wie in Tabelle 5.7 dargestellt modelliert. Der multi-modale Parameter *media-format* für SMIL-Präsentationen wird durch das Element `OutputContentType` modelliert. Für die Unimodalitäten Audio und Video werden im Unterschied zu Tabelle 5.17 keine Status *is-visible*, *has-screen-location* und *is-playing* modelliert.

```
- < RenderingAgentProfile >
- < RenderComponent type="player" >
  - < OutputUnimodalityList >
    < OutputUnimodality name="video" linguistic="-li" analogue="an"
      arbitrary="-ar" stadyn="dyn" medium="vis" subtype="none" />
    < OutputUnimodality name="audio" linguistic="-li" analogue="an"
      arbitrary="-ar" stadyn="dyn" medium="aud" subtype="none" />
  </ OutputUnimodalityList >
  - < OutputContentTypeList >
    < OutputContentType id="smil" />
  </ OutputContentTypeList >
</ RenderComponent >
- < OutputDeviceList >
  < OutputDevice name="Dynamite-PDA-2" medium="vis" maxX="240"
    maxY="294" />
  < OutputDevice name="Dynamite-PDA-2" medium="aud" />
</ OutputDeviceList >
</ RenderingAgentProfile >
```

Abbildung 6.12: Selbstbeschreibung der Komponente zur SMIL-Wiedergabe auf dem PDA.

Der Grund für diese Vereinfachung ist, dass die SMIL-Ausgabe stets auf dem gesamten Bildschirm erfolgt und eine laufende Wiedergabe jederzeit durch eine neue Wiedergabe ersetzt werden kann. Ist das SMIL-Fenster durch die GUI-Komponente verdeckt, so wird es im Falle einer Ausgabe automatisch in den Vordergrund des Desktops verschoben (vgl. Abbildung 6.7). Die Mo-

dellierung der Ausgabegeräte findet sich in Element `OutputDevice`. Das Attribut `name` enthält einen eindeutigen Bezeichner (Status *has-id* in Tabelle 5.7).

Das Attribut `medium` modelliert das Medium des Ausgabegeräts, und die Attribute `maxX` und `maxY` modellieren die Auflösung des visuellen Ausgabegeräts (Status *has-resolution*). Die Modellierung der SMIL-Komponente auf dem Fernseher ist identisch mit der Modellierung der Komponente auf dem PDA. Die Auflösung des Fernsehers beträgt im Unterschied zum PDA  $856 \times 600$ .

Der elektronische Bilderrahmen kann statische Bilder anzeigen und Audiodaten wiedergeben. Die Modellierung erfolgt analog zur SMIL-Komponente (Abbildung 6.13). Für die Modellierung der statischen Bildausgaben wurde das Unimodalitätenmodell der Domäne (Tabelle 5.7) um eine Unimodalität für nicht-linguistische, analoge, nicht-arbiträre, statische, visuelle Ausgaben erweitert. Da der Bilderrahmen ein spezielles Präsentationsformat benötigt, wurde hierzu ein separates Medienformat definiert (*df-slide-show*).

```

- <RenderingAgentProfile>
  - <RenderComponent type="player">
    - <OutputUnimodalityList>
      <OutputUnimodality name="audio" linguistic="-li" analogue="an"
        arbitrary="-ar" stadyn="dyn" medium="aud" subtype="none" />
      <OutputUnimodality name="images" linguistic="-li" analogue="an"
        arbitrary="-ar" stadyn="sta" medium="vis" subtype="none" />
    </OutputUnimodalityList>
    - <OutputContentTypeList>
      <OutputContentType id="df-slide-show" />
    </OutputContentTypeList>
  </RenderComponent>
- <OutputDeviceList>
  <OutputDevice name="df-0444b9" medium="vis" maxX="1024"
    maxY="768" />
  <OutputDevice name="df-0444b9" medium="aud" />
</OutputDeviceList>
</RenderingAgentProfile>

```

Abbildung 6.13: Selbstbeschreibung der Wiedergabekomponente des elektronischen Bilderrahmens.

Die Sprachausgabekomponente wird wie im Modell definiert als Komponente vom Typ *agent* modelliert (Abbildung 6.14, vgl. Tabelle 5.15). Der Parame-

## 6.2. HOME-ENTERTAINMENT MIT GEKOPPELTEN AUSGABEGERÄTEN

ter *content* wird durch das Element `OutputContentType` wiedergegeben. Bei der Sprachausgabe ist dies der Typ `dynamite`, durch den reine Textinhalte beschrieben werden. Als Ausgabemedien werden eine WAV-Datei und eine AU-Datei zurückgegeben (Element `OutputMediaList`).

```
- <RenderingAgentProfile>
- <RenderComponent type="agent">
  - <OutputUnimodalityList>
    <OutputUnimodality name="speech-output" linguistic="li"
      analogue="-an" arbitrary="-ar" stadyn="dyn"
      medium="aud" subtype="spoken-text" />
  </OutputUnimodalityList>
  - <OutputContentTypeList>
    <OutputContentType id="dynamite" />
  </OutputContentTypeList>
  - <OutputMediaList>
    <OutputMedium format="wav"
      url="http://11.0.0.1/dynamite/tmp/inText2.wav" />
    <OutputMedium format="au"
      url="http://11.0.0.1/dynamite/tmp/inText2.au" />
  </OutputMediaList>
</RenderComponent>
<OutputDeviceList />
</RenderingAgentProfile>
```

Abbildung 6.14: Selbstbeschreibung der Sprachausgabekomponente.

Dabei hat die WAV-Datei die höhere Qualität (22.050 Hz) und die AU-Datei die niedrigere Qualität (12.000 Hz). Dies wird jedoch im Unterschied zu Tabelle 5.16 durch die Implementierung nicht explizit modelliert, sondern es wird davon ausgegangen, dass die WAV-Datei die höhere Qualität besitzt. Da die Komponente nur Medienobjekte zurückliefert, werden keine Ausgabegeräte modelliert. Der Zugriff auf die Medienobjekte erfolgt mittels eines Webservers, der auf dem Server (mit der IP-Adresse 11.0.0.1) läuft.

Abbildung 6.15 zeigt die Selbstbeschreibung der Charakterkomponente. Wie bei der Sprachausgabekomponente produziert die Charakterkomponente keine Ausgaben, sondern liefert Medienobjekte zurück. Die verwendeten Unimodalitäten sind identisch mit denen im Modell definierten (vgl. Tabellen 5.7 und 5.13).

Bei den Ausgabemedien werden animierte GIF-Bilder in zwei Auflösungen zurückgeliefert (vgl. Tabelle 5.14). Dabei wird auf das in Tabelle 5.5 definierte Animationsmodell zurück gegriffen. Weiterhin wurden die Animationen durch ein Hintergrundbild (`background`) ergänzt. Darüber hinaus wird eine

## KAPITEL 6. IMPLEMENTIERUNG

Sprachanimation und eine Ruheanimation des Charakters verwendet. Wie bei den Medienobjekten der Sprachausgabe wurde auf die Modellierung der Dateigröße verzichtet (vgl. Abbildung 6.14). Stattdessen wird von der Auflösung auf die Qualität und Größe der Bilder geschlossen.

```
- < RenderingAgent Profile >
- < RenderComponent type="agent" >
- < OutputUnimodalityList >
  < OutputUnimodality name="character-head" linguistic="-li" analogue="an"
  arbitrary="-ar" stady="dyn" medium="vis" subtype="head" / >
  < OutputUnimodality name="character-torso" linguistic="-li" analogue="an"
  arbitrary="-ar" stady="dyn" medium="vis" subtype="torso" / >
  < OutputUnimodality name="character-facial-expression" linguistic="-li"
  analogue="an" arbitrary="-ar" stady="dyn" medium="vis" subtype="facial-
  expression" / >
  < OutputUnimodality name="character-gaze" linguistic="-li" analogue="an"
  arbitrary="-ar" stady="dyn" medium="vis" subtype="gaze" / >
  < OutputUnimodality name="character-body" linguistic="-li" analogue="an"
  arbitrary="-ar" stady="dyn" medium="vis" subtype="body" / >
  < OutputUnimodality name="character-gestures" linguistic="-li" analogue="an"
  arbitrary="-ar" stady="dyn" medium="vis" subtype="gestures" / >
  < OutputUnimodality name="character-lip-movements" linguistic="li"
  analogue="an" arbitrary="-ar" stady="dyn" medium="vis" subtype="spoken-
  text" / >
</ OutputUnimodalityList >
< OutputContent TypeList / >
- < OutputMediaList >
  < OutputMedium format="animated-gif"
  url="http:// 11.0.0.1/ dynamite/ tmp/ SpeakContinued-640-480.gif"
  type="speak-continued" res-x="640" res-y="480" / >
  < OutputMedium format="animated-gif"
  url="http:// 11.0.0.1/ dynamite/ tmp/ RestPoseContinued-640-480.gif"
  type="restpose-continued" res-x="640" res-y="480" / >
  < OutputMedium format="animated-gif"
  url="http:// 11.0.0.1/ dynamite/ tmp/ background-640-480.gif"
  type="background" res-x="640" res-y="480" / >
  < OutputMedium format="animated-gif"
  url="http:// 11.0.0.1/ dynamite/ tmp/ SpeakContinued-240-320.gif"
  type="speak-continued" res-x="240" res-y="320" / >
  < OutputMedium format="animated-gif"
  url="http:// 11.0.0.1/ dynamite/ tmp/ RestPoseContinued-240-320.gif"
  type="restpose-continued" res-x="240" res-y="320" / >
  < OutputMedium format="animated-gif"
  url="http:// 11.0.0.1/ dynamite/ tmp/ background-240-320.gif"
  type="background" res-x="240" res-y="320" / >
</ OutputMediaList >
</ RenderComponent >
< OutputDeviceList / >
</ RenderingAgent Profile >
```

Abbildung 6.15: Selbstbeschreibung der Charakterkomponente.

## 6.2. HOME-ENTERTAINMENT MIT GEKOPPELTEN AUSGABEGERÄTEN

Die GUI-Eingabekomponente und die SMIL-Ausgabekomponente auf dem PDA müssen nicht koordiniert werden, da beide Ausgaben in unterschiedlichen Masken auftreten (Abbildung 6.7). Daher wird die GUI-Komponente nicht als Eingabekomponente modelliert. Die Modellierung der Spracheingabekomponente ist in Abbildung 6.16 dargestellt. Die Spracheingabe-Unimodalität wird analog zu den Definitionen aus den Tabellen 5.9 und 5.10 modelliert.

Ebenso wurde der Status *threw-input-event* implementiert (Element `InputEvent`), der angibt, ob die Sprachausgabe im letzten Dialogschritt eine Eingabe verarbeitet hat. Dies kann dazu verwendet werden, im Falle einer Spracheingabe in derselben Modalität mittels einer Sprachausgabe zu antworten. Der Status *is-processing-input* wurde nicht implementiert, da dieser für das realisierte System nicht relevant war.

```
- < InputAgentProfile>
- < InputComponent>
  - < InputUnimodalityList>
    < InputUnimodality name="speech-input" linguistic="li"
      analogue="-an" arbitrary="-ar" medium="aud" />
  </ InputUnimodalityList>
  - < InputStatus>
    < InputEvent />
  </ InputStatus>
</ InputComponent>
- < InputDeviceList>
  < InputDevice name="11.0.0.1" medium="aud" />
</ InputDeviceList>
</ InputAgentProfile>
```

Abbildung 6.16: Selbstbeschreibung der Komponente zur Spracheingabe.

### Ausgabeinhalt und -ziele

Abbildung 6.17 zeigt ein Beispiel für den Inhalt und das Ziel der Ausgabe, die vom Dialog-Manager an die MMP-Komponente gesendet werden (vgl. Abbildung 5.4). Als Ausgabeziele werden die Dialogakte *message-inform* und *query-selection* unterstützt (Element `dialogact`). Den Dialogakten wurde dabei das in [Ludwig, 2004] definierte Modell zugrunde gelegt (vgl. Tabelle 6.1). Dabei stellt *message-inform* eine allgemeine Information über Kinofilme dar. *query-selection* fordert den Benutzer auf, eine Auswahl zu treffen (z. B. bei der Auswahl zwischen verschiedenen Filmtiteln).

Beim Dialogakt *query-selection* wird im Text-Modus der Text zentriert in der Mitte des Bildschirms angezeigt, um zu betonen, dass der Benutzer eine

## KAPITEL 6. IMPLEMENTIERUNG

Auswahl treffen muss. Beim Dialogakt *message-inform* erfolgt eine multimediale Präsentation. Neben dem Ausgabeziel werden der Titel und der Inhalt der Ausgabe als Text übergeben, der mittels der Sampa-Notation lautschriftlich annotiert sein kann (Elemente `title` und `text`). Der Text kann durch beliebig viele Bilder ergänzt werden (Element `images`). Beispiele für die verwendeten Texte und Bilder finden sich in Anhang A.1.

```
- <present>
  <dialogact type="message-inform" />
  <title name="Der Tag, an dem die Erde stillstand" />
  <text>Eine fliegende Untertasse landet in Washington kurz nach Ende des zweiten
  Weltkrieges. Der Ausserirdische Klaatu (Transcription:'ma|k@|} Michael Rennie) und
  sein riesiger Roboter warnen die Menschen vor einem Atomkrieg und fordern sie auf,
  Frieden zu schließen. Um seiner Warnung Nachdruck zu verleihen, läßt er alle
  Maschinen der Erde eine halbe Stunde lang stillstehen. Meilenstein des Science Fiction
  Films.</text>
  - <images>
    <url>http://11.0.0.1/db/media/Der_Tag_an_dem_die_Erde_stillstand/1.jpg</url>
    <url>http://11.0.0.1/db/media/Der_Tag_an_dem_die_Erde_stillstand/2.jpg</url>
    <url>http://11.0.0.1/db/media/Der_Tag_an_dem_die_Erde_stillstand/3.jpg</url>
    <url>http://11.0.0.1/db/media/Der_Tag_an_dem_die_Erde_stillstand/4.jpg</url>
    <url>http://11.0.0.1/db/media/Der_Tag_an_dem_die_Erde_stillstand/5.jpg</url>
    <url>http://11.0.0.1/db/media/Der_Tag_an_dem_die_Erde_stillstand/6.jpg</url>
  </images>
</present>
```

Abbildung 6.17: Ausgabeinhalt und -ziel.

### 6.2.5 Präsentationsplanung

Für die Umsetzung der MMP-Komponente (Abbildung 6.11) mittels des in Abbildung 5.18 illustrierten Ansatzes wurde das Ckuckuk-Präsentationsplanungssystem verwendet [André et al., 2000]. In diesem Abschnitt wird zunächst auf die für das Home-Entertainment eingesetzten Startzustände, Zielzustände und Operatoren für das Ckuckuk-System eingegangen. Anschließend wird beschrieben, wie das System auf Änderungen in der Menge der Ein-/Ausgabekomponenten reagiert. Diese können z. B. durch das Ausschalten oder Hinzufügen eines Ein-/Ausgabegerätes bedingt sein.

#### Startzustand, Zielzustand und Operatoren

Wie in Abbildung 5.6 beschrieben, identifizieren sich die sieben Ein-/Ausgabekomponenten mittels der Selbstbeschreibungen ihrer Dienste, Geräte und Ressourcen. Aus diese Selbstbeschreibungen wird der Startzustand  $s_0$  des Ckuckuk-Planers erzeugt. Abbildung 6.18 zeigt die Selbstbeschreibung der Sprachausgabe im Zustand  $s_0$  (vgl. Abbildung 6.14).

## 6.2. HOME-ENTERTAINMENT MIT GEKOPPELTEN AUSGABEGERÄTEN

```
(render-component rc-id 3)
(render-component-type rc-id 3 rc-type agent)
(output-unimodality uni-id 1 rc-id 3 li-value li an-value -an ar-value -ar
  stady-n-value dyn medium-value aud sub-category spoken-text)
(processes-output-content rc-id 3 oc-type dynamite)
(output-medium rc-id 3 om-format wav om-id 0 om-url
  "http://11.0.0.1/dynamite/tmp/inText2.wav")
(output-medium rc-id 3 om-format au om-id 1 om-url
  "http://11.0.0.1/dynamite/tmp/inText2.au")
```

Abbildung 6.18: Ausgabedienste der Sprachausgabe im Startzustand  $s_0$ .

Das Planungsziel  $g$  besteht aus dem in Abschnitt 6.2.3 beschriebenen Dialogakt. Der Ausgabeinhalt wird als Teil des Startzustands  $s_0$  modelliert. Ein Beispiel für einen vollständigen Startzustand findet sich im Anhang A in Abbildung A.3. Mittels der auf Basis des Modells für Ausgabedienste, -geräte und -ressourcen definierten Operatorenmenge  $O$  ergibt sich für Präsentationen auf PDA, Fernseher und elektronischem Bilderrahmen mit animiertem Charakter der in Abbildung 6.19 dargestellte Planungsbaum (vgl. Präsentationen in Abbildungen 6.9 und 6.10). Der Baum stellt die hierarchische Dekomposition des Ziels  $g$  mittels der an den Knoten des Baumes befindlichen Operatoren dar. Die Operatoren haben die in Abbildung 4.2 gezeigte Struktur. Die vollständigen Definitionen der Operatoren finden sich in Anhang A.3 in den Abbildungen A.4 bis A.10.

Der Operator `BuildSmilPresentationSingleImage` wird verwendet, um die Präsentation eines einzelnen Bildes auf dem Fernseher zu erstellen. Dies geschieht durch die Definition eines grafisch-zeitlichen Layouts mittels der Allen-Intervallalgebra (`PresentStaticImage-640-480`, vgl. Abschnitt 5.5). Mittels `SDeleteImageFromPresentationTask` wird das Bild als ausgegeben vermerkt.

Dies dient dazu, die Ausgabe im nächsten Schritt auf die noch nicht gezeigten Medienobjekte zu beschränken. Die Constraints werden nach der Terminierung des Planungsalgorithmus durch die Aktion `SStartMatsDynamite` in Zahlen umgerechnet. Das Ergebnis ist eine SMIL-Präsentation, die durch Ausführung der Aktion `SAddSmilMessage` an die SMIL-Komponente des Fernsehers versendet wird. Der Operator `BuildSmilPresentation` wird verwendet, um den animierten Charakter auf dem PDA zu generieren. Hierzu wird zunächst das Layout des Charakters, des Filmbildes und der Sprachausgabe auf dem PDA definiert (`PresentStaticImageWithSpeechAnd-`

## KAPITEL 6. IMPLEMENTIERUNG

Character-240-290). Anschließend werden Variablensubstitutionen ermittelt, die den Constraints genügen und eine SMIL-Nachricht generiert (vgl. Abschnitt 5.5). Diese wird an die Ausgabekomponente auf dem PDA versendet.

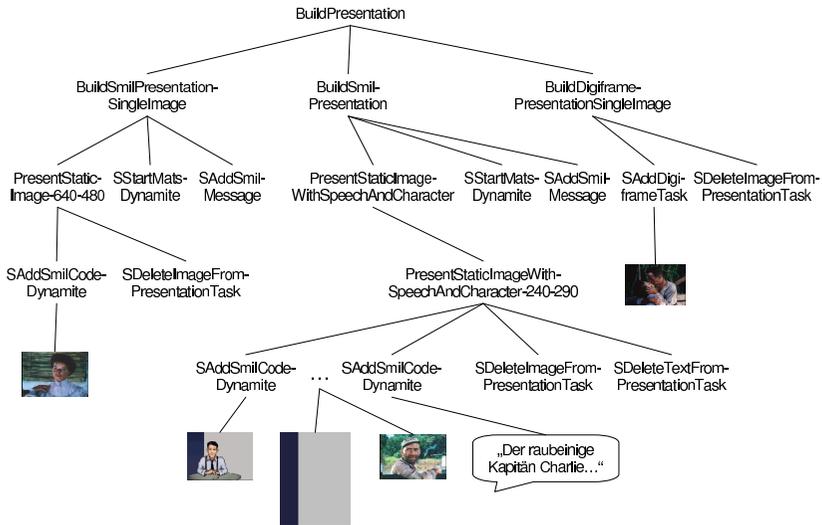


Abbildung 6.19: Planungsbaum für Präsentation mit PDA, Fernseher und Bilderrahmen.

Der Operator `BuildDigiframePresentationSingleImage` erzeugt eine Einzelbildpräsentation im Format des elektronischen Bilderrahmens. Da hier keine Layout-Koordinaten stattfinden müssen, wird der Constraint-Löser nicht verwendet. Um die Filmbilder an die Auflösungen der Geräte anzupassen und in unterschiedliche Layouts integrieren zu können, war es notwendig, Constraints bzgl. der Größe von Bildern zu definieren. Da diese Funktionalität in Ckuckuk nicht vorhanden war, wurde Ckuckuk um Constraints der Form

$$\begin{aligned} (x <= \text{bottom } A1 - \text{top } A1 <= y) \\ (x <= \text{right } A1 - \text{left } A1 <= y) \end{aligned}$$

erweitert. Diese Constraints geben an, innerhalb welcher Grenzen die Größe des durch Aktion `A1` gegebenen Bildes unter Beibehaltung der Seitenverhältnisse verändert werden kann. Eine Einschränkung des Ckuckuk-Planers

## 6.2. HOME-ENTERTAINMENT MIT GEKOPPELTEN AUSGABEGERÄTEN

ist, dass die Überprüfung der Constraints erst nach der Planungszeit durch Ausführung der Aktion `sStartMats` erfolgt. Dies führt dazu, dass durch `Ckuckuk` eine Lösung ausgewählt werden kann, deren Constraints nicht erfüllt sind. Als Folge wird ein Laufzeitfehler angezeigt und die Ausführung des Plans unterbrochen. Eine Modifikation dieser Verhaltensweise war aufgrund der internen Programmstruktur von `Ckuckuk` mit sehr hohem Aufwand verbunden. Weiterhin existierte kein gleichwertiger Planer, der klassisches Planen mit einem Constraint-Löser verbindet.

Der Planer von Scholz [Ding et al., 2006], der den klassischen `Shop2`-Planer verwendet und das Constraint-Lösen in den Planungsalgorithmus integriert, erfüllt die Kriterien. Die Software befindet sich jedoch noch in der Entwicklung. Aus diesem Grund erfolgte die Definition der Operatoren so, dass diese Situationen nicht auftraten.

### Dynamisches Hinzufügen/Entfernen von Ausgabekomponenten

In Abschnitt 6.2.2 wurde illustriert wie das System auf das Hinzufügen oder Entfernen der Ausgabegeräte PDA, elektronischer Bilderrahmen oder Fernseher reagiert. In diesem Abschnitt wird erläutert wie die Adaptionen durch die Präsentationsplanung realisiert wurden.

Eine neue Ausgabe wird durch Versenden von Ausgabeinhalt und -ziel (vgl. Abbildung 6.17) vom Dialog-Manager an die MMP-Komponente eingeleitet. Dies leitet einen neuen Planungsvorgang in der MMP-Komponente ein. Hierzu werden zunächst per Broadcast alle Ein-/Ausgabekomponenten aufgefordert, sich mittels ihrer Selbstbeschreibungen gegenüber der MMP-Komponente zu identifizieren. Dies erfolgt vor jeder Präsentationsplanung, um die Menge der verfügbaren Ein-/Ausgabekomponenten zu ermitteln. Der Grund für diese Vorgehensweise ist, dass Verluste der Netzwerkverbindung (z. B. durch Abziehen eines Netzkabels oder durch Beenden einer WLAN-Verbindung) vom System nicht unmittelbar festgestellt werden können. Diese können erst bei Versenden eines Inhalts beobachtet werden. Eine Alternative Vorgehensweise hierzu sind periodische Überprüfungen der Netzwerkverbindungen in regelmäßigen Intervallen.

Da bei jeder Ausgabe eine Neuplanung erfolgt, resultieren Änderungen in der Menge der Ausgabekomponenten in einem veränderten Startzustand  $s_0$  (vgl. Abbildung 6.18) und damit in einer veränderten Planung. Dies bedeutet, dass wie in Abschnitt 5.6 erläutert keine Änderungen der Menge der Ausgabekomponenten während der Planungszeit (d. h. während des Aufbaus des Planungsbaums aus Abbildung 6.19) berücksichtigt werden. Die Änderungen werden erst während der folgenden Neuplanung berücksichtigt.

### 6.3 Vorteile des gewählten Ansatzes

An den Abschnitten 6.1 und 6.2 beschriebenen Implementierungen zeigen sich die Vorteile der Verwendung des in Kapitel 5.4 beschriebenen Ansatzes. Wie in Abschnitt 5.4.2 erläutert, trägt der Ansatz in beiden Implementierungen zur Ausfallsicherheit bei. Fällt System aus Abschnitt 6.2 das Fernsehen als Ausgabegerät aus, so wird der PDA als alternatives Ausgabegerät verwendet. Fällt der PDA ebenfalls aus, so wird der elektronische Bilderrahmen als alternatives Ausgabegerät angesteuert.

Weiterhin unterstützt das in Java implementierte Modell die Portabilität von Software-Code. So verwenden sowohl die Prototypimplementierung aus Abschnitt 6.1 als auch das Home-Entertainment-System aus Abschnitt 6.2 dieselbe Java-Implementierung des Modells für Ausgabedienste, -geräte und -ressourcen. Diese Software beinhaltet Klassen die aus der XML-Repräsentation der Selbstbeschreibungen (z. B. der Selbstbeschreibung der Sprachausgabekomponente aus Abbildung 6.14) automatisch einen Teil des Startzustandes  $s_0$  des Planers (Abbildung 6.18) erzeugen.

Wird dem System eine zweite Implementierung einer Komponente eines bekannten multi-modalen Typs hinzugefügt (z. B. die AT&T Natural Voices-Sprachausgabe, die wie in Abschnitt 6.2 beschrieben die MBrola-Sprachausgabe ergänzt), so ist hierfür kein zusätzlicher Code erforderlich. Weiterhin kann die Ausdrucksfähigkeit unterschiedlicher Komponenten desselben Typs mittels Bernsens Unimodalitätenmodell detailliert modelliert werden. Somit können unterschiedliche Komponenten wie ein Microsoft Agent-basierter Charakter (vgl. Tabelle 5.11) und der virtuelle Charakter der TU Darmstadt (vgl. Tabelle 5.12) mittels derselben Operatoren beschrieben werden.

### 6.4 Zusammenfassung

Das in Abschnitt 5.4.4 definierte Modell für Ausgabedienste, -geräte und -ressourcen wurde zunächst innerhalb einer Prototypimplementierung im Embassi-System umgesetzt und als Basis der Präsentationsplanung verwendet. In einer Erweiterung der Implementierung innerhalb des Dynamite-Systems wurde ebenfalls der in Abschnitt 5.5 beschriebene Planungsansatz umgesetzt. Das System generiert multimediale Zusammenfassungen von Filminhalten unter Verwendung eines PDAs, eines Fernsehers und eines elektronischen Bilderrahmens, die als Ausgabegeräte gekoppelt werden. Dabei werden die Präsentationen an die jeweils verfügbare Menge von Ausgabegeräten und -komponenten angepasst.

Die Implementierungen zeigen, wie mit Hilfe des Modells Präsentationen mit gekoppelten Ausgabegeräten in dynamischen und heterogenen Geräteumge-

#### 6.4. ZUSAMMENFASSUNG

bungen generiert werden können. Weiterhin wurde gezeigt, dass mit Hilfe des Modells neue Ausgabekomponenten ähnlicher Funktionalität mit geringem Aufwand eingebunden werden können. Das Modell unterstützt ebenfalls die Ausfallsicherheit von ubiquitären Informationssystemen, da defekte Ausgabekomponenten durch alternative Komponenten mit ähnlicher Funktionalität ersetzt werden können.



## 7 Fragebogenstudie

Die in Abschnitt 6.2 beschriebene Implementierung zeigt, wie sich mit Hilfe des in Kapitel 5 entwickelten Ansatzes Präsentationen auf gekoppelten Ausgabegeräten realisieren lassen. In diesem Abschnitt wird betrachtet, wie das implementierte System von Benutzern akzeptiert wird. Dazu erfolgte eine Fragebogenstudie mit 56 Teilnehmern. Die Studie untersuchte, wie sich die vom System generierten Ausgaben in Abhängigkeit der gekoppelten Ausgabegeräte unterscheiden.

### 7.1 Problemstellung und Hypothesen

In diesem Abschnitt werden die Problemstellung und die Hypothesen der Studie beschrieben. Hierzu wird zunächst die Problemstellung erläutert und ein Überblick über Forschungsarbeiten, die die Akzeptanz von Präsentationen auf gekoppelten Ausgabegeräten behandeln, gegeben. Anschließend werden, ausgehend von den existierenden Forschungsarbeiten, Hypothesen über die Akzeptanz des realisierten Home-Entertainment-Systems aufgestellt.

#### 7.1.1 Problemstellung

Das Ziel der Studie war, das in Abschnitt 6.2 beschriebene System in Bezug auf seine Akzeptanz zu untersuchen. Dabei sollten insbesondere die unterschiedlichen Gerätekombinationen des Systems untersucht werden. In der Studie sollten dazu die Kombinationen PDA, Fernseher, Fernseher-PDA, Fernseher-Bilderrahmen und Bilderrahmen untersucht werden. Um den zeitlichen Ablauf der Studie nicht zu komplex zu gestalten, wurde die Kombination Bilderrahmen-PDA nicht untersucht und die verwandte Kombination Fernseher-PDA bevorzugt.

Eine Kopplung aller drei Ausgabegeräte wurde nicht untersucht, da dieser Präsentationstyp zum Zeitpunkt der Studie noch nicht realisiert worden war. Diese Kopplung wurde jedoch in der in Kapitel 8 dargestellten Folgestudie untersucht. Weiterhin sollten die Akzeptanz der Ausgabemöglichkeiten des Systems (animierter Charakter, grafischer Text, Sprachausgabe und Filmbilder) sowie der Eingabemöglichkeiten des Systems (GUI auf dem PDA und Spracheingabe) untersucht werden. Neben der Bewertung einzelner Aspekte des Systems sollte das System auch als Ganzes unabhängig von den Gerätekombinationen bewertet werden.

#### 7.1.2 Akzeptanz von gekoppelten Ausgabegeräten

Bei der Betrachtung gekoppelter Ausgabegeräte finden sich hauptsächlich Studien, die Multi-Monitor-Systeme untersuchen. In einer Feldstudie zur Ver-

## KAPITEL 7. FRAGEBOGENSTUDIE

wendung von Multi-Monitor-Systemen bei IT-Experten [Grudin, 2001] zeigte sich, dass es keine einheitliche Bevorzugung von Konfigurationen für Multi-Monitor Systeme durch die Benutzer zu geben scheint. Der Großteil der 18 Vpn verwendete zwei bis drei Monitore, diese wurden jedoch unterschiedlich eingesetzt. Von den Teilnehmern der Studie verwendeten neun Vpn zwei Monitore, die an denselben PC angeschlossen waren. Drei Teilnehmer setzten einen dritten Monitor ein. Von diesen zwölf Vpn verwendeten sechs Vpn Monitore mit unterschiedlichen Größen (zwischen 15-21 Zoll). Einer verwendete einen Laptop als zweiten Monitor.

Von den fünf verbleibenden Teilnehmern hatten zwei PDAs, die mit dem PC synchronisiert wurden und als Erweiterung des PC-Monitors verwendet wurden. Ein Teilnehmer hatte einen Monitor, der das Bild eines zweiten Monitors in einer geringeren Auflösung zeigte. Ein Teilnehmer verwendete ein Web-Portal, um Daten zwischen zwei separaten PCs mit zwei Monitoren mittels Drag-and-Drop auszutauschen. Ein dritter Teilnehmer benutzte drei virtuelle Desktops auf zwei Monitoren.

Alle Vpn äußerten sich positiv über die Verwendung multipler Monitore. Keiner der Vpn wollte das Multi-Monitor-System durch ein System mit einem einzigen Monitor eintauschen. Die Studie ist ein Hinweis darauf, dass technische Experten effizient mit Multi-Monitor-Systemen arbeiten können und flexibel in Bezug auf die Konfiguration von Multi-Monitor-Systemen sind. Die Akzeptanz von Multi-Monitor-Systemen im Arbeitsumfeld wird durch eine weitere von Microsoft in Auftrag gegebenen Studie unterstützt [Robertson et al., 2005]. In dieser Studie wurden 1.197 Benutzer befragt, die Windows im Arbeitsumfeld verwenden. Die Studie ergab, dass etwa 20% Multi-Monitor-Systeme an ihrem Arbeitsplatz einsetzen.

In der Studie von [Czerwinski et al., 2003] wurde ein 15 Zoll großer Monitor mit einem experimentellen 42 Zoll DSharp-Monitor verglichen. Der 42 Zoll große DSharp-Monitor wurde durch Rückprojektion unter der Verwendung von drei Projektoren erzeugt und entspricht drei nebeneinander gestellten Monitoren. 15 Teilnehmer, die alle Windows-Experten waren, nahmen an der Studie teil. In der Studie wurden die Vpn gebeten, verschiedene typische Büroaufgaben mittels der ihnen zugewiesenen Monitoren zu lösen. In einem Fragebogen wurden den Teilnehmer vier Fragen zur Akzeptanz des Systems gestellt. Bei allen vier Fragen wurde der große Monitor signifikant besser beurteilt.

Eine weitere Studie von Microsoft Research untersuchte die visuelle Trennung von Informationen über mehrere Bildschirme hinweg [Tan & Czerwinski, 2003]. Die Studie wurde mit 24 Benutzern mit fortgeschrittenen Windows-

## 7.1. PROBLEMSTELLUNG UND HYPOTHESEN

Kenntnissen durchgeführt. Es wurden zwei 22 Zoll Monitore und ein Projektionsschirm als Ausgabegeräte gekoppelt und in sechs Kombinationen (darunter ein einzelner Monitor als Kontrollbedingung) zusammengestellt. Die Teilnehmer mussten über mehrere Ausgabegeräte hinweg Aufgaben zu den Themen Korrekturlesen, Email-Benachrichtigungen und Textvergleiche lösen. Entgegen der Hypothesen der Autoren wurden jedoch keine signifikanten Effizienz- oder Akzeptanzunterschiede bzgl. der visuellen Trennung zwischen den beiden Monitoren oder den Monitoren und der Projektionsleinwand festgestellt.

In der Studie von [Oonk et al., 2003] wurden Konfigurationen mit einem bis vier sowie sechs Monitoren im Kontext eines militärischen Kommandozen-trums untersucht. Dabei wurde bei Kopplungen von einem bis drei Monitoren eine horizontale Anordnung gewählt. Bei Kopplungen von vier Monitoren wurde ein Monitor links unten angefügt. Bei Kopplungen von sechs Monitoren wurden zwei Reihen mit je drei Monitoren verwendet. Die Teilnehmer waren 30 Soldaten. Jeder Teilnehmer hat in dem Versuch mit jeder Monitor-Konfiguration gearbeitet. Auf den Monitoren wurden separate Desktops mit Microsoft-Applikationen (z. B. Microsoft Word oder Microsoft Chat) sowie taktische Daten (z. B. eine geographische Karte) dargestellt.

Es wurden zwei Experimente durchgeführt. Im ersten Experiment sollten die Teilnehmer einen Bericht erstellen. Im zweiten Experiment sollten die Teilnehmer sich einen Überblick über die taktische Situation verschaffen. In beiden Experimenten erfolgten parallel Anfragen per Email oder per Chat, die die Teilnehmer schnellstmöglich beantworten mussten. Im ersten Experiment bevorzugte eine signifikante Anzahl von Teilnehmern das System mit vier Monitoren. Im zweiten Experiment ergab sich keine signifikante Präferenz der Teilnehmer.

In einer dieser Arbeit vorhergehenden Studie [Elting et al., 2002] wurden verschiedene Typen von multi-modalen Ausgaben auf einem PDA, einem Fernseher und einem Desktop PC in Bezug auf Ihre Effizienz und Akzeptanz untersucht. Dabei wurden die Kombinationen Text-Bild-Sprache, Bild-Sprache, Text-Bild, Text-Sprache und Text als Präsentationsmöglichkeiten untersucht. Insgesamt wurden die Kombinationen Text-Bild und Text-Bild-Sprache von den Teilnehmern am besten bewertet. Die signifikant unattraktivste Bedingung war Text alleine. Zwischen den drei Gerätegruppen ergaben sich keine signifikanten Unterschiede bzgl. der Attraktivität der Präsentationsbedingungen.

Zahlreiche Studien beobachten positive Effekte von animierten Charakteren in Verbindung mit Informationstechnologie. In [Van Mulken et al., 1998] wurden Präsentationen mit einem animierten Charakter in Bezug auf Ihre Effizienz und Akzeptanz untersucht. Dabei wurden sowohl technische Präsentationen

als auch nicht-technische Präsentationen untersucht. Die Effizienz der Präsentation wurde dabei anhand von Verständnisfragen und Erinnerungsfragen überprüft. Die Akzeptanz wurde mittels eines Fragebogens ermittelt. Die Animationen des Charakters beschränkten sich auf Zeigegeesten und Lippenbewegungen. An der Studie nahmen 30 Personen teil. Die Teilnehmer beurteilten dabei die Präsentationen mit dem animierten Charakter insgesamt unterhaltsamer als die Präsentationen ohne den animierten Charakter. Die Teilnehmer mit den Charakterpräsentationen waren zwar nicht signifikant effizienter als die Teilnehmer mit Präsentationen ohne Charakter, aber sie beurteilten die Tests als weniger schwierig. Dies bezeichnen die Autoren als „Persona Effekt“.

In [Kruppa, 2006] wird ein positiver Effekt eines migrierenden Charakters festgestellt, der von einem Wandbildschirm mit einer Animation verschwindet und auf dem persönlichen PDA mit einer weiteren Animation eingeblendet wird. Der Autor hat die Migration des animierten Charakters mit der Migration eines Symbols bzgl. der Akzeptanz der Benutzer verglichen. Als Kontrollbedingung wurden ebenfalls Präsentationen ohne Migrationsanimation betrachtet. Als Präsentationsinhalt wurden Präsentationen von einer National Geographic DVD verwendet, die auf dem Wandbildschirm und zeitweise auf einem PDA präsentiert wurden. An der Studie nahmen 42 Personen teil. 57% der Teilnehmer bevorzugten dabei die Migrationen des animierten Charakters.

Zusammenfassend lässt sich feststellen, dass gekoppelte Ausgabegeräte lediglich in Bezug auf Bildschirme nähergehend untersucht wurden. Die ausführlichen Studien von Microsoft Research sind weiterhin auf Windows-Anwendungen beschränkt. Systematische Studien zur Akzeptanz der Kopplung von PDAs mit anderen Ausgabegeräten existieren nicht.

### 7.1.3 Hypothesen

Aus den vorhergehenden Untersuchungen lassen sich mehrere Hypothesen ableiten. Die Arbeiten von [Robertson et al., 2005; Czerwinski et al., 2003; Grudin, 2001] weisen auf die Akzeptanz von Multi-Monitor-Systemen durch technische Benutzer hin. Daraus ergibt sich die Hypothese, dass die gekoppelten Ausgabegeräte in Bezug auf die Verwendung mehrerer Bildschirme auch in dieser Studie positiv beurteilt werden. Gemäß den Ergebnissen von [Tan & Czerwinski, 2003] vermuten wir, dass die visuelle Trennung zwischen PDA und Fernseher sowie Bilderrahmen keine wesentliche Rolle bei der Akzeptanzbewertung spielen werden. Weiterhin erwarten wir einen positiven Effekt des animierten Charakters, wie er in den Studien von [Kruppa, 2006; Van Mulken et al., 1998] berichtet wurde. Weiterhin erwarten wir nach den Ergebnissen aus [Elting et al., 2002], dass Präsentationen mit Texten und Bildern ebenfalls sehr positiv bewertet werden.

## 7.2. METHODE

### 7.2 Methode

In diesem Abschnitt wird der Aufbau der Studie beschrieben. Dabei wird auf die verwendeten Variablen, die Teilnehmer und die verwendeten Materialien der Studie eingegangen. Weiterhin werden die Durchführung der Studie und die verwendeten statistischen Methoden beschrieben.

#### 7.2.1 Variablen

Bei der Fragebogenstudie wurden vier Kategorien von abhängigen Variablen verwendet (Tabelle 7.1). Zunächst wurden die Vpn bzgl. der Eingabemodalitäten des Systems befragt. Sie wurden gebeten, die Spracheingabe durch das drahtlose Headset zu bewerten sowie die Stifteingabe auf dem Touchscreen des PDA. Anschließend wurden sie gebeten, die Ausgabemöglichkeiten des Systems zu bewerten. Diese schlossen den animierten Charakter, die Sprachausgabe sowie grafischen Text und Filmbilder ein. Die Bewertung der Ein-/Ausgabemodalitäten erfolgte dabei unabhängig von den Gerätekombinationen.

Tabelle 7.1: Abhängige Variablen der Fragebogenstudie.

| Kategorie                | Variable                                                                                                                                                                                                          |
|--------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Eingabemodalitäten       | Spracheingabe per Headset<br>Stifteingabe per PDA                                                                                                                                                                 |
| Ausgabemodalitäten       | Animierter Charakter<br>Sprachausgabe<br>grafischer Text<br>Filmbilder                                                                                                                                            |
| gekoppelte Ausgabegeräte | PDA<br>TV<br>TV-PDA<br>TV-Bilderrahmen<br>Bilderrahmen                                                                                                                                                            |
| Gesamtsystem             | Bewertung Demonstration<br>Nützlichkeit Kopplung mehrerer Geräte<br>Nützlichkeit Layout-Präferenzen<br>Präferenz Einzelgerät oder Gerätekombination<br>Präferenz System oder PC<br>Ausgabebereitschaft für System |

Anschließend bewerteten die Teilnehmer die unterschiedlichen Kombinationen gekoppelter Ausgabegeräte. Dabei wurden in dieser Studie die Einzelgeräteausgaben auf dem PDA, dem Fernseher und dem Bilderrahmen mit den

## KAPITEL 7. FRAGEBOGENSTUDIE

Kombinationen Fernseher-PDA und Fernseher-Bilderrahmen verglichen. Zuletzt wurden die Vpn gebeten, das Gesamtsystem zu bewerten. Hierbei wurden die Nützlichkeit der Kopplung mehrerer Geräte sowie die Nützlichkeit der Layout-Präferenzen abgefragt. Weiterhin wurden die Vpn gefragt, ob sie Präsentationen auf einem einzelnen Ausgabegerät oder mit gekoppelten Ausgabegeräten vorziehen. Ebenso wurden sie gefragt, ob sie das System gegenüber einem PC mit Internet-Anschluss bei der Ausgabe von Filminformationen vorziehen würden. Zuletzt wurden die Vpn gebeten, anzugeben, wie viel Geld sie in den Kauf eines solchen Systems investieren würden.

Alle abhängigen Variablen bis auf die Präferenz Einzelgerät oder Gerätekombination, die Präferenz System oder PC sowie die Ausgabebereitschaft waren Likert-ähnlich und fünfstufig (siehe z. B. [Bortz & Döring, 2003]). Die Werte der Variablen reichten von völliger Ablehnung der vorangestellten Aussage (z. B. „Überhaupt nicht gefallen“) bis zu völliger Zustimmung (z. B. „Sehr gut gefallen“). Die zwei Präferenzvariablen waren zweistufig. Die Variable Ausgabebereitschaft war fünfstufig. Hier konnten die Vpn zwischen Eurobeträgen von 0 Euro bis 2400 Euro (unterteilt in äquidistante 600 Euro Schritte) wählen.

Als Moderatorvariablen wurden das Alter der Vp und die Stundenzahl pro Woche, die die Vp mit einem PC, einem PDA oder einem elektronischen Bilderrahmen arbeitet, verwendet. Weiterhin wurde festgehalten, welchen Beruf die Vp hat und wie viele Spielfilme die Vp pro Woche im Fernsehen, auf dem Computer oder im Kino anschaut. Die Vp wurde ebenfalls gefragt, ob sie beruflich oder privat auf der CeBit Messe war.

### 7.2.2 Teilnehmer

Der Versuch wurde auf dem Stand des Projektes Dynamite auf der CeBit 2006 zwischen dem 09.03.2006 und dem 15.03.2006 durchgeführt. Interessenten am Stand wurde vom Versuchsleiter zunächst eine Demonstration des Systems gezeigt (Abbildung 7.1). Anschließend wurden sie gefragt, ob sie Interesse hätten, an einer Fragebogenstudie teilzunehmen.

### 7.2.3 Materialien

Das präsentierte System entspricht dem in Abschnitt 6.2 beschriebenen System. Dabei wurde für die Sprachausgabe die Mbrola-Version der Sprachausgabe verwendet. In den folgenden Abschnitten werden die präsentierten Inhalte, die Kopplungen von Ausgabegeräten und der Fragebogen beschrieben.

#### **Präsentationsinhalte**

Jeder Präsentationsinhalt bestand aus einem Titel, einer Inhaltszusammenfassung und beliebig vielen Standbildern aus einem Film. Für die Demonstration wurden vier Filme verwendet („Der Tag, an dem die Erde still stand“, „Caesar

## 7.2. METHODE



Abbildung 7.1: Vp während der Demonstration am Dynamite-Stand (Bild: European Media Laboratory).

und Cleopatra“, „Anastasia“und „African Queen“), denen jeweils zwischen sechs bis zehn Bilder zugeordnet waren. Die präsentierten Inhalte finden sich in Anhang A.1.

### **Gerätekombinationen und Präsentationsmodi**

In dem Versuch wurden die folgenden Gerätekombinationen demonstriert. Bei der Ausgabe der Filminformationen auf dem PDA alleine wurde ein Bild des ausgewählten Filmtitels auf dem PDA dargestellt (vgl. Abbildung 6.7, rechts). Bei der Fernseher-PDA-Demonstration (TV-PDA) wurde der Fernseher verwendet, um zwei zusätzliche Bilder des Filmes anzuzeigen (vgl. Abbildung 6.9). Bei Präsentationen auf dem Fernseher ohne weitere Geräte (TV) wurden der animierte Charakter und ein Filmbild auf dem Fernsehbildschirm dargestellt.

In der Fernseher-Bilderrahmen-Demonstration (TV-Bilderrahmen) wurde diese Ausgabe durch ein weiteres Bild auf dem elektronischen Bilderrahmen ergänzt (vgl. Abbildung 6.8). Bei einer Präsentation auf dem Bilderrahmen alleine wurde auf dem Bilderrahmen ein Bild des ausgewählten Films gezeigt und die Sprachausgabe durch die Lautsprecher des Bilderrahmens ausgegeben.

Das System wurde durch zwei alternative Eingabemodalitäten bedient. Der gewünschte Film konnte sowohl durch Stifteingabe auf dem PDA als auch durch Spracheingabe mittels eines drahtlosen Mikrophons erfolgen. Dabei

hatten die Dialoge stets die folgende Form. Zunächst wurde der Benutzer gefragt, zu welchem Film er Informationen haben möchte. Anschließend wurde durch den Benutzer ein Film ausgewählt, der durch das System präsentiert werden sollte. Nach der Präsentation hatte der Benutzer die Wahl, ob er sich weitere Bilder vom Film anzeigen lassen möchte oder ob er abrechnen und einen neuen Film sehen möchte.

Für die Präsentationen gab es jeweils zwei mögliche Layout-Einstellungen, die mit dem PDA gewählt werden können: Text-Bild und animierter Charakter. Im Text-Bild-Modus wurde anstelle des animierten Charakters auf dem Fernseher oder dem PDA grafischer Text dargestellt (vgl. Abbildung 6.10). In diesem Fall wurde auf dem jeweiligen Bildschirm kein weiteres Bild dargestellt. Neben dem Layout-Modus konnte durch eine Menüleiste auf dem PDA-Bildschirm (vgl. Abbildung 6.7) ebenfalls die Ausgabe von Filminformationen auf dem PDA, dem Fernseher und auf dem elektronischen Bilderrahmen ein- und ausgeschaltet werden. Im Falle des Fernsehers geschah dies durch Zurückschalten vom VGA-Kanal des eingebetteten PCs auf einen TV-Kanal. Dies dient dazu, das Ein- und Ausschalten der Geräte zu simulieren, ohne die Geräte neu booten zu müssen.

### **Fragebogen**

Der verwendete Fragebogen wird in Anhang B wiedergegeben. Die Vp wurde zunächst nach allgemeinen Daten zu ihrer Person gefragt. Sie wurde gebeten, ihr Alter, ihr Geschlecht und ihren Beruf anzugeben sowie ob sie sich beruflich oder privat auf der CeBit aufhält. Außerdem wurden die Vpn gefragt, wie viele Stunden pro Woche sie mit einem Computer, einem Fernseher, einem PDA oder einem elektronischen Bilderrahmen arbeiten. Sie wurden ebenfalls gefragt, wie viele Kinofilme sie sich im Monat anschauen.

Anschließend wurden die Vpn zu den demonstrierten Eingabemöglichkeiten des Systems befragt. Sie wurden gebeten, zu bewerten, wie komfortabel sie die Spracheingabe per Headset und die Stifteingabe auf dem PDA empfanden. Bei der Bewertung der Ausgabemöglichkeiten des Systems sollten die Vpn angeben, wie gut ihnen der animierte Helfer, die Sprachausgabe, der grafische Text und die Filmbilder unabhängig vom konkreten Ausgabegerät gefallen haben.

Danach wurden die Vpn gebeten, die Gerätekombinationen zu bewerten. Hierbei sollten die Vpn die PDA-, TV-, PDA-TV-, TV-Bilderrahmen- und Bilderrahmen-Demonstrationen beurteilen. Abschließend wurden die Vpn gebeten, das Gesamtsystem zu beurteilen. Die Vp wurde zunächst gefragt, wie gut ihr die Demonstration insgesamt gefallen hat. Sie wurde gebeten, die Nützlichkeit der Kombination mehrerer Geräte bei der Ausgabe von Filminformationen so-

## 7.2. METHODE

wie die Nützlichkeit der Wahl zwischen Text-Bild- und Charakter-Modus zu bewerten.

Die Vp wurde ebenfalls gefragt, ob sie bei der Ausgabe von Filminformationen ein Gerät oder mehrere Geräte bevorzugt. Außerdem wurde sie gebeten anzugeben, ob sie das System einem PC mit Internet-Anschluss vorziehen würde und wie viel Geld sie bereit wäre, für das System auszugeben. Am Ende des Fragebogens hatte die Vp noch Gelegenheit, Kommentare, Kritik und Verbesserungswünsche in ein Textfeld einzutragen.

### 7.2.4 Durchführung

Die Demonstration des Systems erfolgte durch den Versuchsleiter, d. h. die Teilnehmer haben nicht selber mit dem System interagiert. Die Demonstration hatte den folgenden Ablauf. Zunächst wurde eine kurze Einleitung durch den Versuchsleiter gegeben. Anschließend wurde die PDA-Demonstration im Charakter-Modus durchgeführt und ein Filminhalt präsentiert (Abbildung 6.7). Danach wurde durch den PDA in den Text-Bild-Modus gewechselt, wodurch der Text des Filmes auf dem PDA erschien (vgl. Fernseher-PDA-Bilderrahmen-Präsentation in Abbildung 6.10). Anschließend wurde durch das PDA-Menü der Fernseher als Ausgabegerät aktiviert, wodurch neben dem Text zwei weitere Bilder auf dem Fernsehbildschirm erschienen. Nach der Präsentation wurde die Ausgabe auf dem PDA deaktiviert und in den Charakter-Modus gewechselt, wodurch der Charakter auf dem Fernseher erschien. Dann wurde das Headset eingeschaltet und der Dialog zur Auswahl eines Filmes per Sprache durchgeführt. Anschließend wurde durch den PDA der elektronische Bilderrahmen aktiviert, wodurch ein weiteres Bild des Filmes dargestellt wurde (Abbildung 6.8).

Am Ende der Demonstration wurde die Ausgabe auf dem Fernseher durch Umschalten des Fernsehprogramms deaktiviert, wodurch die Sprachausgabe des Filminhalts auf den Lautsprechern des Bilderrahmens ausgegeben wurde. Danach wurden die Vpn gebeten, den Fragebogen auszufüllen. Der Versuchsleiter war während dieser Zeit in der Nähe, falls Unklarheiten auftreten sollten. Nachdem die Vp den Fragebogen ausgefüllt hatte, fand ein Abschlussgespräch statt.

### 7.2.5 Statistische Auswertung

Die im Experiment verwendeten Likert-Skalen entsprechen einer Schulnotenskala, bei der nicht nur Kategorien relevant sind (Nominalskala), sondern auch die Rangfolge der Kategorien (Ordinalskala). Da für diese kategorialen Daten die Normalverteilungsbedingung von Varianzanalyse (Anova) und t-Test i. Allg. nicht erfüllt ist, werden Likert-skalierte Daten stattdessen mit

parameterfreien Verfahren ausgewertet [Bortz, 2005]. Die verwendeten Verfahren sind der Mann-Whitney-Test für zwei unabhängige Stichproben über Ordinalskalen, der Wilcoxon-Test für zwei abhängige Stichproben über Ordinalskalen sowie der Friedman-Test für k abhängige Stichproben über Ordinalskalen. Diese Tests entsprechen dem t-Test (für zwei Variablen) bzw. der Varianzanalyse (für k Variablen) bei Intervalldaten. Weiterhin wird der  $\chi^2$ -Test zur Untersuchung der Signifikanz von Häufigkeitsverteilungen (Nominaldaten) eingesetzt. Bei allen Tests wird ein Signifikanzniveau von  $\alpha = 0.05$  verwendet.

Zur besseren grafischen Darstellung wurden die Likert-Skalen durch eine Abbildung auf eine äquidistante Skala über ganzen Zahlen (Werte eins bis fünf) in Intervallskalen überführt. Dabei entsprach der Wert eins einer völligen Ablehnung der korrespondierenden Aussage und der Wert fünf einer völligen Zustimmung. Weiterhin erforderte die Untersuchung spezieller Hypothesen auch die Verwendung von t-Test und Anova auf den intervallskalierten Daten. Dies ist eine gängige Vorgehensweise in der Statistik [Velleman & Wilkinson, 1993]. Für die wichtigsten Daten werden weiterhin im Text die Mittelwerte (MW) und die Standardabweichungen (SD) angegeben, sofern sie nicht aus den Abbildungen hervorgehen.

Für die Visualisierung der Nominaldaten werden Histogramme verwendet, die die gemessenen Häufigkeiten darstellen. Für die Visualisierung der Intervalldaten werden Balkendiagramme verwendet. Dabei bestimmen die Mittelwerte der Messungen die Höhe der Hauptbalken. Die Klammern geben dabei die 95%-Konfidenzintervalle an. Ein 95%-Konfidenzintervall stellt das Intervall dar, in dem sich der wahre Mittelwert der Verteilung mit 95% Wahrscheinlichkeit befindet (siehe z. B. [Bortz & Döring, 2003]) und sind somit ein Maß für die Streuung der Verteilung. Im Unterschied zur Standardabweichung lassen sich durch das Konfidenzintervall bei unabhängigen Variablen direkt Rückschlüsse auf die Signifikanz der Messungen schließen. Zwei unabhängige Stichproben A und B unterscheiden sich genau dann signifikant, wenn sich der Mittelwert von A außerhalb des Konfidenzintervalls von B befindet und umgekehrt. In diesem Fall lassen sich z. B. durch den t-Test signifikante Unterschiede zwischen A und B nachweisen.

### 7.3 Ergebnisse

In diesem Abschnitt finden sich die Ergebnisse der statistischen Signifikanztests. Es wird zunächst eine allgemeine, deskriptive Beschreibung der Stichprobe gegeben. Anschließend werden die Daten in Abhängigkeit verschiedener Faktoren analysiert. Hierbei werden das Alter der Vpn, die Art des Messebesuchs, die technischen Erfahrungen sowie die Anzahl der gesehenen Kino-

### 7.3. ERGEBNISSE

filme betrachtet. Weiterhin werden die Gesamtbewertungen der Eingabe- und Ausgabemodalitäten des Systems untersucht. Abschließend werden die Bewertungen der Gerätekombinationen und des Gesamtsystems sowie die Kommentare der Vpn betrachtet.

#### 7.3.1 Beschreibung der Stichprobe

Es hatten 56 Teilnehmer an der Fragebogenstudie teilgenommen. Der Großteil der Teilnehmer war männlich (50 Teilnehmer). Das Durchschnittsalter der Vpn war 30.1 Jahre. Der jüngste Teilnehmer war 15 Jahre alt. Der älteste Teilnehmer war 72 Jahre alt.

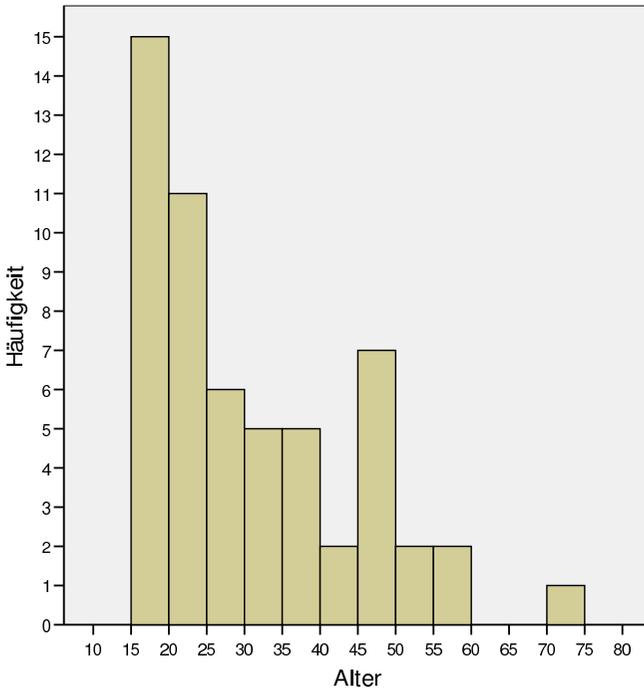


Abbildung 7.2: Histogramm für Altersintervalle von fünf Jahren.

Abbildung 7.2 zeigt die Altersstruktur in einem Histogramm. Von den 56 Vpn besuchten 35 die Messe privat (62.5%), 18 Vpn besuchten die Messe beruflich (32.1%) und drei Vpn besuchten die Messe sowohl privat als auch beruflich (5.4%). Bei der Stundenzahl Computer-Arbeit pro Woche gab keine Vp null Stunden pro Woche (d. h. gar keine Arbeit mit dem Computer) an. Die durch-

schnittliche Stundenzahl betrug 32.3 Stunden pro Woche. Bei der Stundenzahl Fernsehen pro Woche gaben zehn Vpn (5.6%) eine Stundenzahl von null Stunden pro Woche an. Die durchschnittliche Stundenzahl betrug 8.9 Stunden pro Woche. Beim PDA gaben 38 Vpn (67.9%) an, dass sie gar nicht mit dem PDA arbeiten. Die durchschnittliche Stundenzahl der Arbeit mit dem PDA betrug 2.3 Stunden pro Woche.

### 7.3.2 Altersgruppen

In der Studie wurden drei Unterteilungen in jüngere und ältere Vpn untersucht, um mögliche Korrelationen mit den Bewertungen der jeweiligen Gruppen zu untersuchen. Hierzu wurden die Gruppen jünger als 20 Jahre und 20 Jahre oder älter bzw. jünger als 30 Jahre und 30 Jahre oder älter sowie jünger als 40 Jahre und 40 Jahre oder älter untersucht.

#### **Gruppe jünger/älter 20**

15 Vpn waren jünger als 20 Jahre (26.8%, Gruppe jünger 20) und 41 Vpn waren 20 Jahre oder älter (73.2%, Gruppe älter 20). Es wurden mehrere signifikante Unterschiede zwischen den beiden Gruppen gefunden, wobei die Gruppe jünger 20 stets signifikant besser auf der Likert-Skala bewertete. Die Gruppen jünger 20 ( $MW = 4.2$ ,  $SD = .8$ ) und älter 20 ( $MW = 3.6$ ,  $SD = .7$ ) unterschieden sich signifikant bzgl. der Bewertung des Komforts der Spracheingabe (Mann-Whitney,  $p = .010$ ). Weiterhin unterschieden sich beide Gruppen signifikant bzgl. der Bewertung des Komforts der Stifteingabe auf dem PDA (Mann-Whitney,  $p = .021$ ; Gruppe jünger 20:  $MW = 4.4$ ,  $SD = .5$ ; Gruppe älter 20:  $MW = 3.8$ ,  $SD = .9$ ).

Bei den Ausgabemöglichkeiten unterschieden sich beide Gruppen signifikant bzgl. der Bewertung der Sprachausgabe (Mann-Whitney,  $p = .005$ ; Gruppe jünger 20:  $MW = 3.8$ ,  $SD = 1.1$ ; Gruppe älter 20:  $MW = 2.8$ ,  $SD = 1.0$ ). Weiterhin unterschieden sich beide Gruppen signifikant bzgl. der Bewertungen der Ausgaben von grafischem Text (Mann-Whitney,  $p = .013$ ; Gruppe jünger 20:  $MW = 4.3$ ,  $SD = .8$ ; Gruppe älter 20:  $MW = 3.6$ ,  $SD = .8$ ). Bezüglich der Bewertung der einzelnen Gerätekombinationen wurde ein signifikanter Unterschied zwischen den Gruppen bzgl. der Bewertung der Kombination TV-Bilderrahmen gefunden (Mann-Whitney,  $p = .036$ ; Gruppe jünger 20:  $MW = 4.4$ ,  $SD = .6$ ; Gruppe älter 20:  $MW = 3.7$ ,  $SD = 1.3$ ).

#### **Weitere Altersgruppen**

Zwischen den Altersgruppen jünger/älter 30 bzw. jünger/älter 40 ergaben sich deutlich weniger Signifikanzen. Zwischen den Gruppen jünger/älter 30 gab es signifikante Unterschiede bei der Bewertung der Stifteingabe mit dem PDA (Mann-Whitney,  $p = .014$ ; Gruppe jünger 30:  $MW = 4.2$ ,  $SD = .7$ ; Gruppe

### 7.3. ERGEBNISSE

älter 30: MW = 3.6, SD = 1.0). Zwischen den Gruppen jünger/älter 40 gab es signifikante Unterschiede bzgl. der Bewertung der grafischen Textausgabe (Mann-Whitney,  $p = .022$ ; Gruppe jünger 40: MW = 4.0, SD = .8; Gruppe älter 40: MW = 3.3, SD = .8). Analog zu den Gruppen jünger/älter 20 vergab die jüngere Gruppe stets die jeweils besseren Bewertungen.

#### 7.3.3 Art des Messebesuchs

Bei der Art des Messebesuchs wurden die Vpn in zwei Gruppen eingeteilt: beruflicher und privater Messeaufenthalt. Dabei wurden Vpn, die angegeben haben, dass sie sowohl beruflich als auch privat die Messe besucht haben, zu der Gruppe der beruflichen Messeaufenthalte gezählt. Die Gruppe mit beruflichem Messeaufenthalt enthielt 21 Vpn (37.5%). Die Gruppe mit privatem Messeaufenthalt bestand aus 35 Vpn (62.5%). Die Gruppe mit privatem Messeaufenthalt bewertete die grafische Textausgabe signifikant besser als die Gruppe mit beruflichem Messeaufenthalt (Mann-Whitney,  $p = .012$ ; Gruppe privater Aufenthalt: MW = 4.0, SD = .8; Gruppe beruflicher Aufenthalt: MW = 3.4, SD = .9).

#### 7.3.4 Erfahrungen mit Computern, Fernsehern und PDAs

Bezüglich der Verwendung von Computern teilten wir die Vpn in zwei Gruppen ein: Vpn, die bis zu 30 Stunden pro Woche mit einem Computer arbeiten (29 Vpn, 51.8%) und Vpn, die 30 Stunden oder mehr mit einem Computer arbeiten (27 Vpn, 48.2%). Der Grenzwert 30 wurde gewählt, da dieser als Median die Verteilung in zwei homogene Hälften teilte. Die Gruppe, die bis zu 30 Stunden mit einem Computer arbeitete, bewertete die Sprachausgabe signifikant besser als die Gruppe, die mehr als 30 Stunden am Computer arbeitete (Mann-Whitney,  $p = .001$ ; Gruppe weniger als 30 Stunden Computer: MW = 3.6, SD = 1.0; Gruppe mehr als 30 Stunden Computer: MW = 2.5, SD = .9).

Bei der Verwendung von Fernsehern teilten wir die Vpn in zwei Gruppen ein. Die erste Gruppe schaute bis zu zehn Stunden pro Woche Fernsehen (34 Vpn, 60.7%). Die zweite Gruppe schaute zehn Stunden oder mehr Fernsehen (22 Vpn, 39.3%). Es wurde der Grenzwert zehn gewählt, da er die Verteilung in zwei homogene Hälften teilte. Bei dieser Gruppeneinteilung ergaben sich jedoch keine signifikanten Unterschiede bzgl. der Bewertungen der Eingabemöglichkeiten, der Ausgabemöglichkeiten, der Gerätekombinationen oder der Gesamtpräsentation.

Da beim PDA nur eine geringe Anzahl von Vpn überhaupt mit einem PDA gearbeitet hat, wurde die folgende Gruppeneinteilung gewählt: Die erste Gruppe umfasst alle Vpn, die mit einem PDA arbeiten (19 Vpn, 33.9%). Die zweite

## KAPITEL 7. FRAGEBOGENSTUDIE

Gruppe umfasst alle Vpn, die nicht mit einem PDA arbeiten (37 Vpn, 66.1%). Die Gruppe, die nicht mit PDAs arbeitet, bewertete dabei die Gerätekombination TV-Bilderrahmen signifikant besser als die Gruppe, die mit einem PDA arbeitet (Mann-Whitney,  $p = .025$ ; Gruppe ohne PDA: MW = 4.1, SD = 1.0; Gruppe mit PDA: MW = 3.5, SD = 1.1).

### 7.3.5 Kinofilme pro Monat

Bei der Anzahl der Kinofilme pro Monat wurden die Vpn in zwei Gruppen eingeteilt. Dadurch konnten die Bewertungen von Vpn, die häufiger Kinofilme schauen, mit den Bewertungen von Vpn verglichen werden, die sich weniger häufig Kinofilme ansehen. Die erste Gruppe enthielt Vpn, die bis zu drei Kinofilme pro Monat gucken (37 Vpn, 66.1%). Die zweite Gruppe enthielt Vpn, die drei oder mehr Kinofilme pro Monat schauen (19 Vpn, 33.9%). Der Grenzwert von drei Kinofilmen wurde gewählt, da dieser die Verteilung der Bewertungen in zwei homogene Hälften teilte.

Die Gruppe, die bis zu drei Kinofilme pro Monat schaut, bewertete die Ausgaben auf dem Bilderrahmen alleine signifikant besser als die Gruppe, die mehr Kinofilme schaut (Mann-Whitney,  $p = .002$ ; Gruppe bis zu drei Kinofilme: MW = 3.5, SD = 1.2; Gruppe mehr als drei Kinofilme: MW = 2.8, SD = .8). Die Gruppe, die bis zu drei Kinofilme pro Monat schaut, bewertete ebenfalls die Ausgaben auf dem Fernseher signifikant besser (Mann-Whitney,  $p = .007$ ; Gruppe bis zu drei Kinofilme: MW = 4.2, SD = .6; Gruppe mehr als drei Kinofilme: MW = 3.7, SD = .7).

### 7.3.6 Eingabemodalitäten

Bei den Eingabemodalitäten wurde die Stifteingabe auf dem PDA besser bewertet als die Spracheingabe per Headset. Der Unterschied erwies sich jedoch nicht als signifikant (Wilcoxon,  $p = .178$ ; Spracheingabe: MW = 3.76, SD = .751; Stifteingabe: MW = 3.96, SD = .9).

Die Unterschiede zwischen den Eingabemodalitäten wurden ebenfalls in Abhängigkeit der Gruppen jünger/älter 20 betrachtet. Hierzu wurde eine multivariate Varianzanalyse (Pillais PS-Teststatistik, [Bortz, 2005]) verwendet. Zwischen den Gruppen gab es signifikante Unterschiede bei der Bewertung der Eingabemodalitäten (Anova,  $F(2, 51) = 6.9$ ,  $p = .002$ ). Abbildung 7.3 stellt die Unterschiede in einem Balkendiagramm dar. Die Klammern zeigen die 95%-Konfidenzintervalle der Mittelwerte.

### 7.3.7 Ausgabemodalitäten

Zwischen den Ausgabemodalitäten animierter Charakter, Sprachausgabe, grafischer Text und Filmbilder existierten signifikante Unterschiede (Friedman,  $p = .001$ ). Die Mittelwerte werden in Abbildung 7.4 grafisch dargestellt.

### 7.3. ERGEBNISSE

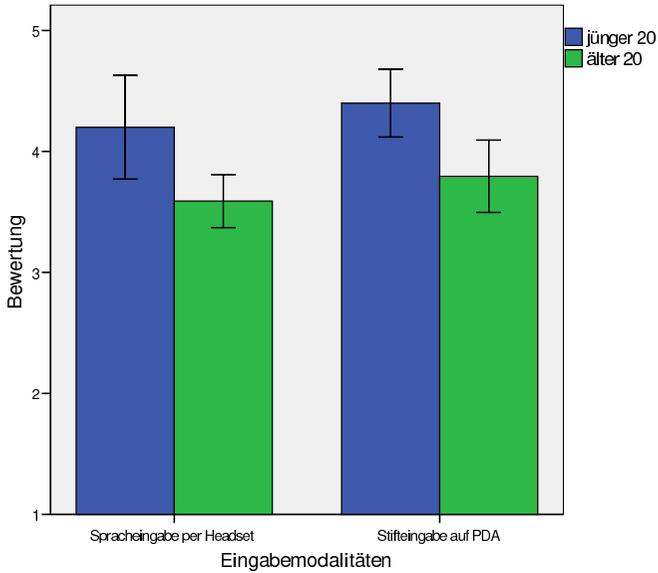


Abbildung 7.3: Bewertungen Eingabemodalitäten für Gruppen jünger/älter 20.

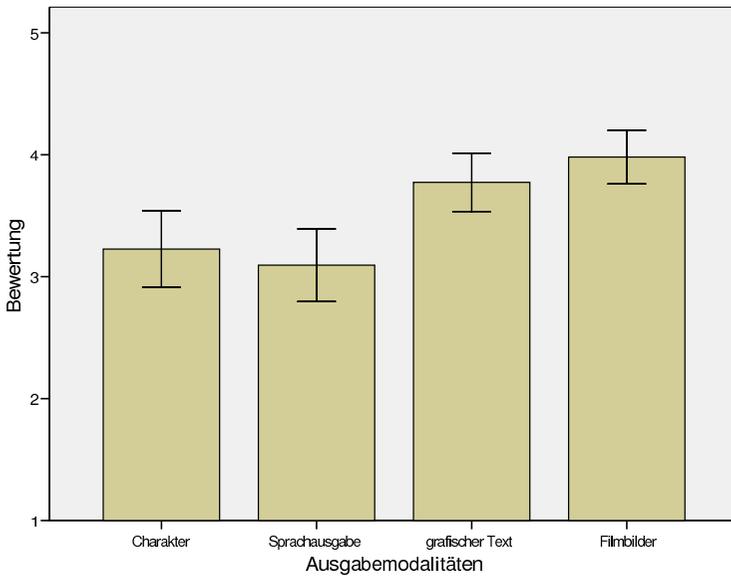


Abbildung 7.4: Bewertungen Ausgabemodalitäten.

## KAPITEL 7. FRAGEBOGENSTUDIE

Um die Unterschiede zwischen den einzelnen Bewertungen genauer zu untersuchen, wurden Wilcoxon-Tests für abhängige Stichproben durchgeführt. Hieraus ergab sich, dass die Bewertungen des grafischen Texts und der Filmbilder jeweils signifikant besser waren als die Bewertungen des animierten Charakters und der Sprachausgabe (Charakter-Text:  $p = .004$ , Charakter-Bilder:  $p = .001$ , Sprachausgabe-Text:  $p = .001$ , Sprachausgabe-Bilder:  $p = .001$ ).

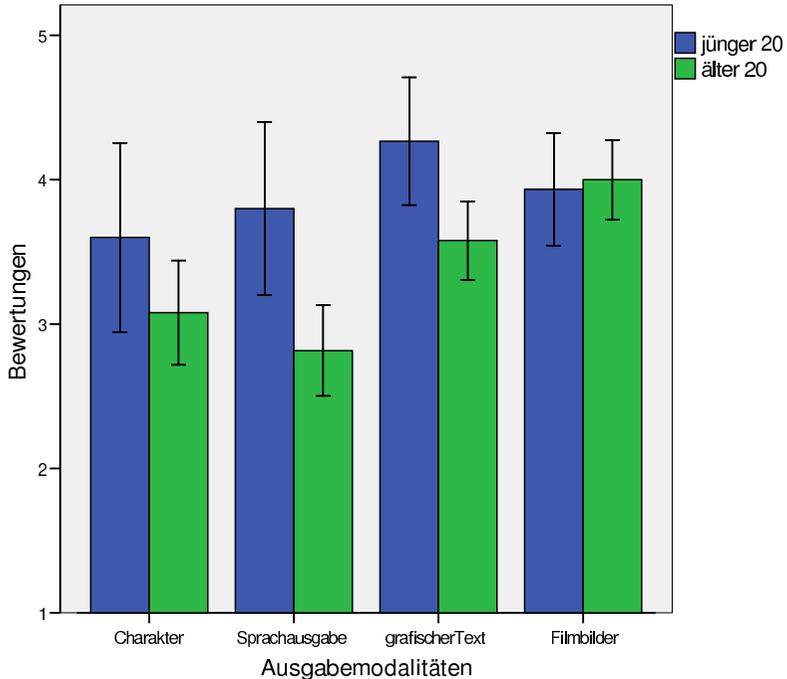


Abbildung 7.5: Bewertungen Ausgabemodalitäten für Gruppen jünger/älter als 20 Jahre.

Die Unterschiede zwischen grafischem Text und Filmbildern ( $p = .083$ ) sowie zwischen dem animierten Charakter und der Sprachausgabe ( $p = .375$ ) waren jedoch nicht signifikant. Die Bewertung der Ausgabemodalitäten wurden analog zu den Eingabemodalitäten mittels einer Varianzanalyse daraufhin überprüft, ob sich die Bewertungen der Gruppen jünger/älter als 20 Jahre unterschieden.

### 7.3. ERGEBNISSE

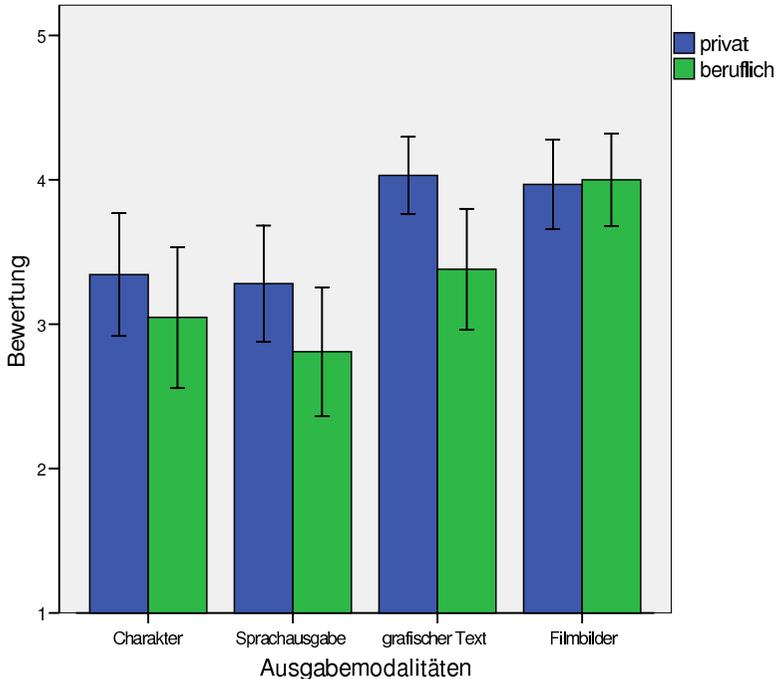


Abbildung 7.6: Bewertungen Ausgabemodalitäten getrennt nach Art des Messebesuchs.

Hierbei wurden signifikante Unterschiede in den Bewertungen der Gruppen festgestellt (Anova,  $F(4, 48) = 4.8, p = .002$ ). Die Mittelwerte werden in Abbildung 7.5 dargestellt. Die Bewertungen der Ausgabemodalitäten wurden ebenfalls signifikant durch die Art des Messeaufenthalts beeinflusst (Anova,  $F(4, 48) = 2.82, p = .036$ ).

Dabei bewerteten Besucher, die ausschließlich privat auf der Messe waren, die Ausgabemodalitäten besser als Besucher, die beruflich auf der Messe waren. Abbildung 7.6 stellt die Mittelwerte grafisch dar. Die Erfahrungen mit Computern hatten einen signifikanten Einfluss auf die Bewertungen der Ausgabemodalitäten.

Vpn, die bis zu 30 Stunden pro Woche mit dem Computer arbeiten, bewerteten die Ausgabemodalitäten signifikant besser als Vpn, die mehr als 30 Stunden pro Woche mit dem Computer arbeiten (Anova,  $F(4, 48) = 4.87, p = .002$ ). Abbildung 7.7 stellt die Unterschiede grafisch dar.

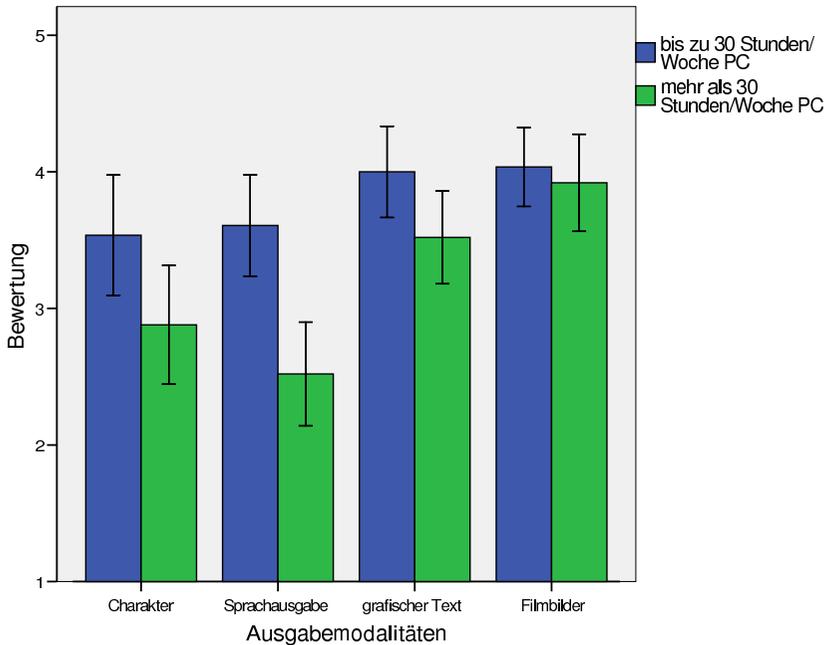


Abbildung 7.7: Bewertungen Ausgabemodalitäten getrennt nach PC-Nutzung.

### 7.3.8 Gerätekombinationen

Zwischen den Bewertungen der fünf Gerätekombinationen wurden signifikante Unterschiede beobachtet (Friedman,  $p = .001$ ). Abbildung 7.8 zeigt die Unterschiede. Mit paarweisen Wilcoxon-Tests wurden die Unterschiede zwischen den einzelnen Gerätekombinationen genauer untersucht. Es zeigte sich, dass die Kombination Bilderrahmen signifikant schlechter als die anderen Gerätekombinationen bewertet wurde (Bilderrahmen-PDA:  $p = .010$ , Bilderrahmen-TV:  $p = .001$ , Bilderrahmen - TV-PDA:  $p = .001$ , Bilderrahmen - TV-Bilderrahmen:  $p = .001$ ). Weiterhin wurde TV-PDA am besten bewertet. Der Unterschied zur zweitbesten Bedingung TV war nicht signifikant (Wilcoxon,  $p = .561$ ). Ebenso war der Unterschied zur drittbesten Bedingung TV-Bilderrahmen nicht signifikant (Wilcoxon,  $p = .149$ ). Der Unterschied zur viertbesten Bedingung PDA war jedoch signifikant (Wilcoxon,  $p = .043$ ).

Die Bewertungen der Gerätekombinationen wurden analog zu den Eingabe- und Ausgabemodalitäten daraufhin überprüft, ob sich signifikante Unterschiede

### 7.3. ERGEBNISSE

de zwischen den definierten Gruppen bzgl. Alter, Nutzung von Computer/Fernseher/PDA, Art des Messebesuchs oder Kinofilmen pro Monat ergaben. Hierbei zeigten sich Unterschiede zwischen den Gruppen, die bis zu drei Kinofilme pro Monat bzw. mehr als drei Kinofilme pro Monat gesehen haben.

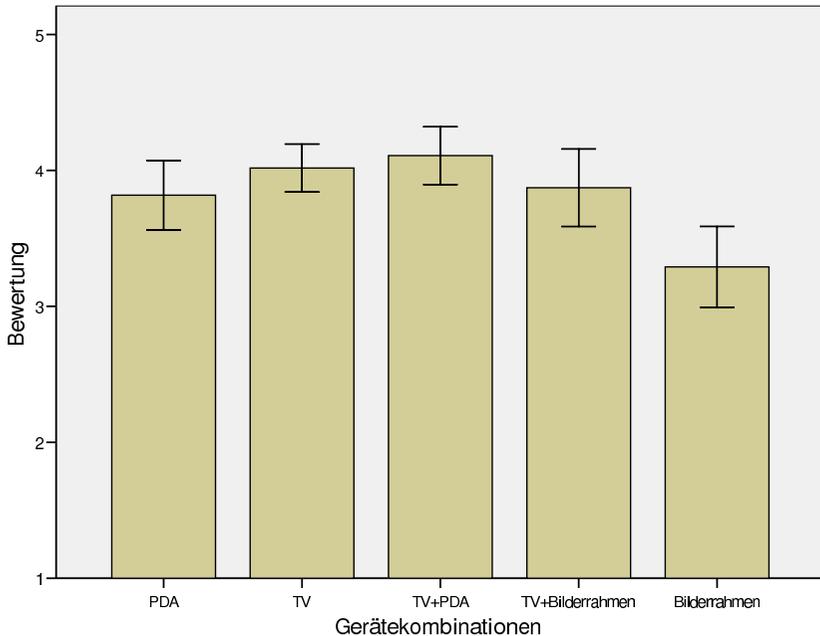


Abbildung 7.8: Bewertungen Gerätekombinationen.

Die Unterschiede zwischen den Bewertungen aller Gerätekombinationen waren zwar nicht signifikant (Anova,  $F(5, 49) = 2.23, p = .066$ ), aber die Unterschiede zwischen den Bewertungen des Fernsehers sowie des Bilderrahmens erwiesen sich, wie in Abschnitt 7.3.5 dargestellt, als signifikant. Abbildung 7.9 zeigt die Mittelwerte der einzelnen Gruppen.

#### 7.3.9 Bewertung des Gesamtsystems

Die Gesamtdemonstration wurde im Mittel mit „recht gut“ (MW=4.2, SD=.6) bewertet. Die gemessenen Häufigkeiten, mit denen eine Bewertungskategorie gewählt wurde, unterschieden sich sehr signifikant ( $\chi^2$ -Test,  $\chi^2(2) = 30.3, p = .001$ , vgl. Abbildung 7.10). Bei der Bewertung der Gesamtdemonstration wurden keine signifikanten Unterschiede bzgl. Alter, der Art des Messe-

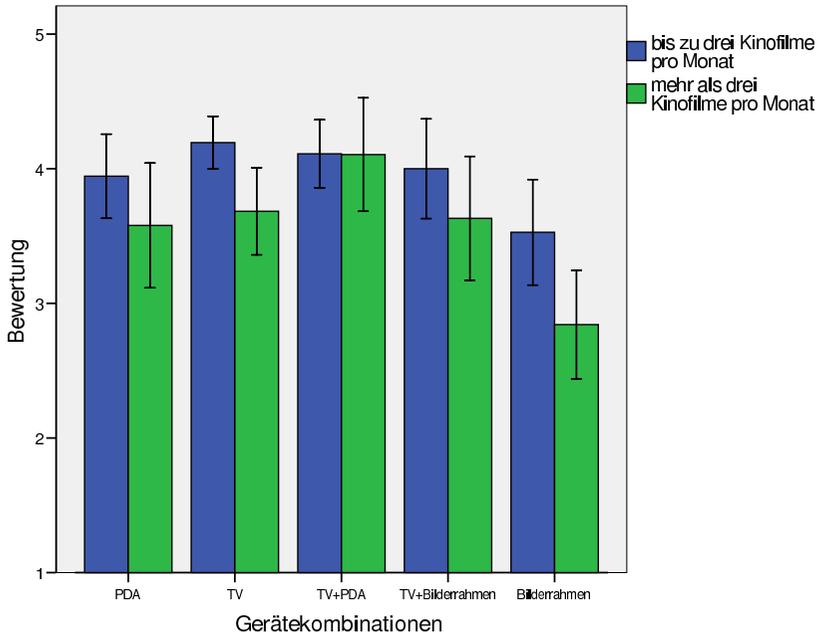


Abbildung 7.9: Bewertungen Gerätekombinationen für Gruppen bis zu drei bzw. mehr als drei Kinofilme pro Monat.

besuchs, der Verwendung von Computern, PDAs und Fernsehern sowie der Anzahl gesehener Kinofilme pro Monat festgestellt.

Die Vpn wurden weiterhin gefragt, wie nützlich sie die Kombination mehrerer Geräte einschätzen. Die Kombination wurde im Mittel mit „recht nützlich“ bewertet ( $MW=3.9$ ,  $SD=1.0$ ). Die Unterschiede zwischen den Häufigkeiten der Bewertungen waren sehr signifikant ( $\chi^2$ -Test,  $\chi^2(3) = 15.0$ ,  $p = .002$ , Abbildung 7.11). Es wurden wie bei der Bewertung der Gesamtpräsentation keine signifikanten Unterschiede zwischen den definierten Gruppen festgestellt. Die Möglichkeit, zwischen Text-Bild Layout ohne animiertem Charakter zu einem Layout mit animiertem Charakter umzuschalten, wurde von den Vpn als „recht nützlich“ bewertet ( $MW = 4.0$ ,  $SD = .9$ ). Die Unterschiede zwischen den Häufigkeiten der Bewertungen waren sehr signifikant ( $\chi^2$ -Test,  $\chi^2(4) = 40.4$ ,  $p = .001$ , Abbildung 7.12). Es wurden keine signifikanten Unterschiede zwischen den definierten Gruppen festgestellt.

### 7.3. ERGEBNISSE

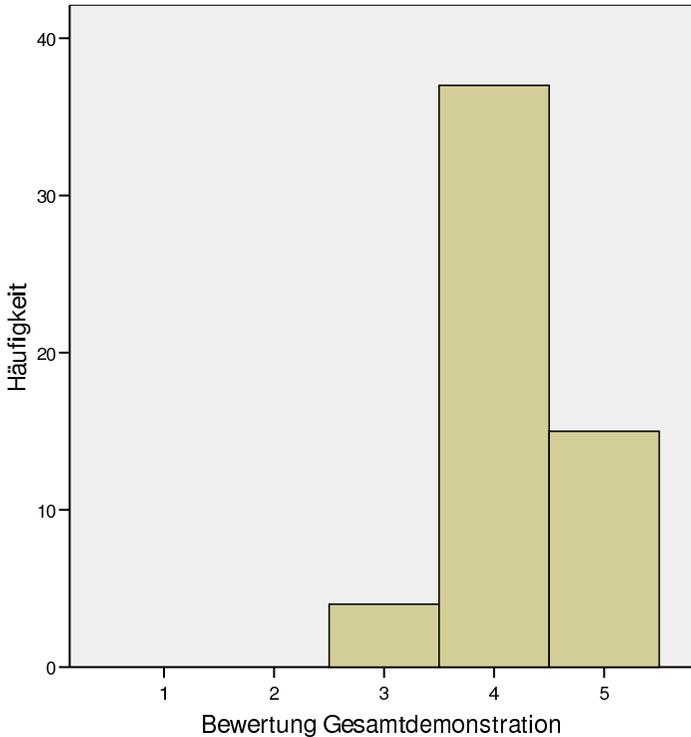


Abbildung 7.10: Histogramm Bewertung der Gesamtdemonstration.

Bei der Frage, ob die Vp mehrere Ausgabegeräte gegenüber einem Gerät bevorzugt, gab es keine signifikante Unterschiede zwischen den Häufigkeiten beider Bewertungen ( $\chi^2$ -Test,  $\chi^2(1) = .7, p = .785$ ). Ebenso ergaben sich keine signifikanten Unterschiede bzgl. der definierten Gruppen. Dabei gaben 26 Vpn an, dass sie ein einzelnes Gerät bevorzugten, während 28 Vpn angaben, dass sie mehrere Geräte bevorzugten. Zwei Vpn machten keine Angaben zu diesem Versuchsteil.

Weiterhin wurden sehr signifikante Unterschiede bei den Beantwortung der Frage gefunden, ob die Vpn das demonstrierte System bei der Suche nach Filminformationen einem PC mit Internet-Anschluss vorziehen würden ( $\chi^2$ -Test,  $\chi^2(2) = 34.9, p = .001$ ). Hierbei bevorzugten 36 Vpn das demonstrierte System, während 16 Vpn einen PC bevorzugten. Eine Vp konnte sich nicht entscheiden. Drei Vpn haben keine Angaben gemacht.

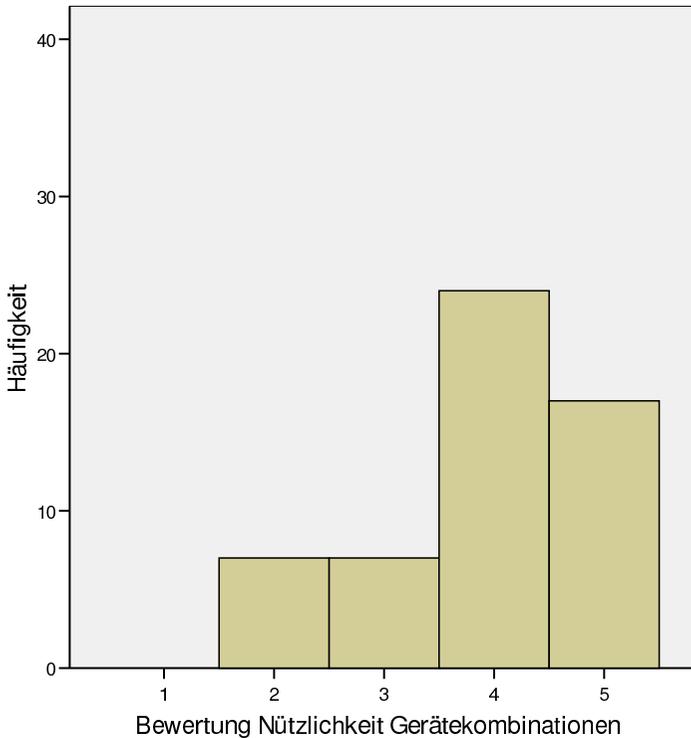


Abbildung 7.11: Histogramm Nützlichkeit der Kombination mehrerer Geräte.

Es gab ebenfalls sehr signifikante Unterschiede zwischen den möglichen Geldbeträgen, die die Vpn in das System investieren würden ( $\chi^2$ -Test,  $\chi^2(4) = 16.7, p = .002$ ). Im Mittel waren die Vpn bereit,  $MW = 962.2$  Euro ( $SD = 689.5$ ) für das System zu bezahlen. Drei Vpn hatten bei der Frage keine Angaben gemacht. Acht der verbleibenden 53 Vpn hatten 0 Euro angegeben, d. h. sie würden das System gar nicht kaufen. Abbildung 7.13 stellt die Häufigkeiten dar.

Teilnehmer, die jünger als 20 Jahre waren, gaben an, dass sie signifikant mehr Geld für das System ausgeben würden als Teilnehmer, die älter als 20 Jahre waren ( $t(51) = 2.49, p = .016$ ; Gruppe jünger 20:  $MW = 1320.0, SD = 823.9$ ; Gruppe älter 20:  $MW = 821.0, SD = 582.2$ ). Weiterhin gaben Vpn, die bis zu 30 Stunden mit einem Computer arbeiten, an, dass sie signifikant mehr Geld für das System ausgaben als Teilnehmer, die mehr als 30 Stunden mit einem Computer arbeiten ( $t(51) = 2.54, p = .014$ ; Gruppe weniger als

### 7.3. ERGEBNISSE

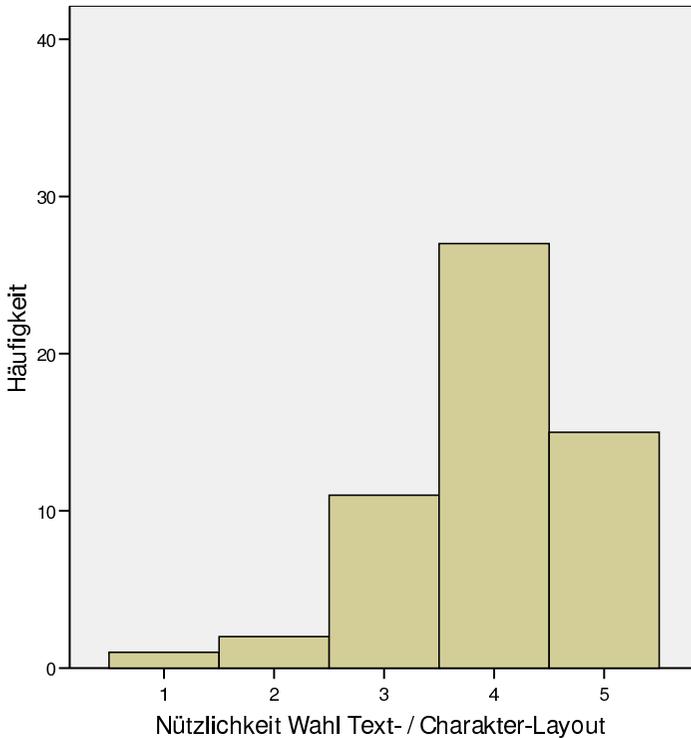


Abbildung 7.12: Histogramm Nützlichkeit der Wahl zwischen Text-/Charakter-Layout.

30 Stunden Computer: MW = 1178.6, SD = 739.0; Gruppe mehr als 30 Stunden Computer: MW = 720.0, SD = 547.7).

#### 7.3.10 Kommentare

Insgesamt 20 der 56 Vpn nutzten den letzten Teil des Fragebogens, um Kommentare abzugeben. Davon lobten sieben Vpn das System oder den Ansatz. Sechs Vpn kritisierten die Sprachausgabe, die besser, flüssiger und individueller gestaltet sein sollte. Drei Vpn kritisierten in ihren Kommentaren den Charakter. Eine Vp kritisierte insbesondere die fehlende Synchronisation zwischen der Sprachanimation des Charakters und der Sprachausgabe. Eine weitere Vp äußerte im Abschlussgespräch, dass die Animationen des Charakters zu simpel waren.

Vier Vpn kritisierten, dass das System zu langsam war. Zwei Vpn merkten

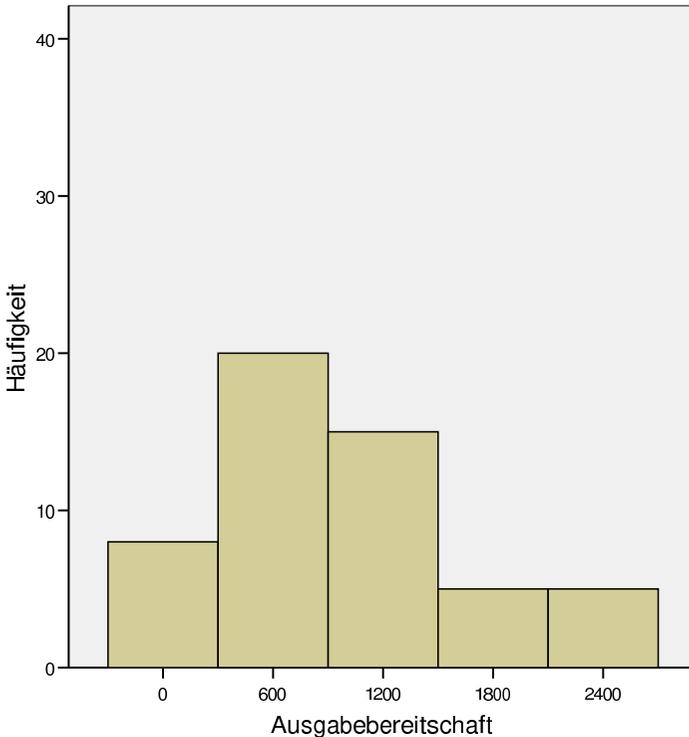


Abbildung 7.13: Ausgabebereitschaft für System in Euro.

an, dass das Layout der Textpräsentation nicht ansprechend war. Zwei Vpn äußerten Kritik an der Spracheingabe. Eine dieser Vpn merkte an, dass sie zu ungenau war. Die zweite Vp kritisierte, dass ein Headset in einem privaten Umfeld nicht akzeptabel ist. Zwei weitere Vpn schlugen eine allgemeine Verbesserung der Benutzeroberfläche des Systems vor.

Drei Vpn schlugen zusätzliche Funktionalitäten für das System vor. Davon schlugen zwei Vpn vor, das Fernsehsignal mit dem System auch auf andere Geräte umzuleiten. Die dritte Vpn schlug vor, den digitalen Bilderrahmen durch einen halb transparenten Spiegel als Ausgabegerät zu ersetzen.

#### 7.4 Diskussion

Die Ergebnisse aus Abschnitt 7.3 zeigen deutlich, dass das Home-Entertainment-System durch verschiedene Benutzergruppen unterschiedlich wahrgenommen wird. Die größten Unterschiede ergaben sich dabei zwischen den

#### 7.4. DISKUSSION

Teilnehmern, die jünger bzw. älter als 20 waren. Die Teilnehmer, die jünger als 20 waren, bewerteten die Eingabemodalitäten und die Ausgabemodalitäten des Systems signifikant besser. Weiterhin waren sie bereit, bis zu 500 Euro mehr für das System auszugeben. Dies lässt sich dadurch erklären, dass ältere Teilnehmer an Ausgaben auf einzelnen Geräten gewöhnt sind und ihnen Präsentationen auf gekoppelten Ausgabegeräten zunächst ungewohnt erscheinen. Jüngere Teilnehmer, die im Zuge der Verbreitung des Internets Anfang der Neunziger Jahre des 20. Jahrhunderts bereits von ihrer Jugend an mit zahlreichen Ausgabegeräten konfrontiert wurden, scheinen offener bzgl. der Kopplung von Ausgabegeräten zu sein.

Weiterhin hat die Studie gezeigt, dass Benutzer mit weniger technischem Wissen das System besser akzeptieren als Benutzer mit mehr technischem Hintergrund. Dies zeigte sich dadurch, dass PDA-Benutzer die Kombination TV-Bilderrahmen signifikant schlechter bewertet haben als Benutzer, die nicht mit einem PDA arbeiten. Weiterhin wurden die Ausgabemodalitäten durch Vpn, die ausschließlich privat auf der CeBit-Messe waren und nicht notwendigerweise einen technischen Beruf hatten, signifikant besser bewertet. Ebenso wurden durch die Gruppe, die bis zu 30 Stunden am Computer arbeitet, die Ausgabemodalitäten signifikant besser bewertet.

Dieses Ergebnis widerlegt die Hypothese aus Abschnitt 7.1.3, dass IT-Fachleute das System besser beurteilen. Mögliche Gründe hierfür können der relativ hohe Automatisierungsgrad des Systems sowie die relativ begrenzte Anzahl an Interaktionen sein, die ein Benutzer durchführen kann. IT-Fachleute sind jedoch von heutigen Programmen (z. B. einem elektronischen Programmführer) ein hohes Maß an Interaktions- und Konfigurationsmöglichkeiten gewöhnt. Dies kann bei IT-Fachleuten zu einer negativen Bewertung des Home-Entertainment-Systems geführt haben.

Dieser Aspekt scheint aber gerade Benutzer mit wenig technischer Erfahrung angesprochen zu haben. Diesen Benutzern gefiel scheinbar der hohe Automatisierungsgrad des Systems. Als mögliche Verbesserung des Systems könnten daher mehr Konfigurations- und Interaktionsmöglichkeiten für fortgeschrittene Benutzer eingeführt werden, während für Benutzer ohne technischen Hintergrund weiterhin ein hoher Automatisierungsgrad verfügbar bleibt.

Die Hypothese, dass das System analog zu den in [Robertson et al., 2005; Czerwinski et al., 2003; Grudin, 2001] untersuchten Multi-Monitor-Systemen positiv beurteilt wird, wurde m. E. bestätigt. Das System wurde insgesamt mit der zweithöchsten Bewertungsstufe („recht gut“) beurteilt. Bei der Bewertung der Gerätekombinationen bestätigte sich die Hypothese, dass wie in [Tan & Czerwinski, 2003] die visuelle Trennung von Medienobjekten durch

Verteilung auf mehrere Ausgabegeräte keinen negativen Effekt auf die Bewertung der Akzeptanz hat. Dies zeigte sich dadurch, dass die Kombination TV-PDA am besten beurteilt wurde. Weiterhin wurde sie signifikant besser als die PDA- sowie die Bilderrahmen-Konfigurationen beurteilt. Sie wurde ebenfalls nicht signifikant schlechter als die Kontrollgruppe (TV-Konfiguration) bewertet. Die Tatsache, dass Vpn, die nur bis zu drei Kinofilme im Monat schauen, die Kombinationen Bilderrahmen und TV-Bilderrahmen signifikant besser beurteilt haben, bedarf jedoch weiterer Untersuchungen.

Ebenso wurde die Hypothese, dass Text-Bild-Präsentationen wie in [Elting et al., 2002] sehr gut akzeptiert werden, bestätigt. Text-Bild-Präsentationen entsprechen dem klassischen Computer-Layout und erlauben dem Benutzer durch die statische Darstellung der Medienobjekte die Möglichkeit der freien Inspektion der Objekte. Die Hypothese, dass der animierte Charakter einen positiven Effekt hat wurde jedoch nicht bestätigt. Der animierte Charakter und die Sprachausgabe wurden von den Ausgabemöglichkeiten des Systems am schlechtesten bewertet. Dies könnte zwei Gründe haben. Zum einen wurde die Sprachausgabe am schlechtesten bewertet, wodurch wahrscheinlich auch der animierte Charakter, dessen Stimme durch die Sprachausgabe generiert wird, negativ bewertet wurde. Weiterhin verwendete der animierte Charakter nur eine eingeschränkte Menge an Animationen (Ruheanimation und Sprachanimation). Dies könnte den Unterhaltungswert des Charakters gesenkt haben. Das Gesamtsystem wurde positiv durch die Benutzer aufgenommen. Dies äußerte sich in den Beurteilungen der Gesamtdemonstration („recht gut gefallen“) und der Bewertung der Nützlichkeit des Kombinierens von Ausgabegeräten („recht nützlich“). Bei der Frage, ob die Vpn bei der Ausgabe von Kinofilminformationen ein oder mehrere Geräte bevorzugen, gab es jedoch keine klare Tendenz. Trotzdem würden sie das System einem PC mit Internetanschluss bei der Ausgabe von Kinofilminformationen vorziehen.

### **7.5 Zusammenfassung**

In der durchgeführten Fragebogenstudie zeigte sich, dass die Kopplung von Ausgabegeräten durch die Teilnehmer der Benutzerstudie gut akzeptiert wurde. Das Gesamtsystem wurde als „recht gut“ bewertet und die Kombination von Ausgabegeräten als „recht nützlich“ empfunden. Weiterhin wurde das System bei der Ausgabe von Filminformationen einem PC mit Internet-Anschluss signifikant vorgezogen. Bezüglich des Alters der Teilnehmer ergaben sich signifikante Unterschiede in den Bewertungen. So wurde das System durch Teilnehmer, die jünger als 20 waren, signifikant besser beurteilt. Dies wurde dadurch erklärt, dass ältere Teilnehmer Ausgaben auf einzelnen Geräten gewöhnt sind und das System daher schlechter beurteilt haben. Ent-

## 7.5. ZUSAMMENFASSUNG

gegen der Erwartungen wurde das System durch Teilnehmer ohne technischen Hintergrund insgesamt besser beurteilt, was durch den hohen Grad der Automatisierung im System erklärt wurde.



## 8 Laborstudie

### 8.1 Problemstellung und Hypothesen

Die in Kapitel 7 beschriebene Fragebogenstudie untersuchte die Akzeptanz des in Abschnitt 6.2 beschriebenen Home-Entertainment-Systems. In diesem Kapitel wird eine Laborstudie beschrieben, die die Effizienz der gekoppelten Ausgabegeräte des Systems untersucht. Darüber hinaus interagieren die Vpn in dieser Studie selbst mit dem System, wodurch eine tiefer gehende Beurteilung der Interaktionen mit gekoppelten Ausgabegeräten ermöglicht werden soll.

Insbesondere stellt sich die Frage, wie die Effizienz einer Präsentation gemessen werden kann. In Übereinstimmung mit anderen Studien messen wir die Effektivität einer Präsentation daran, wie einprägsam sie ist. Dies setzen wir ins Verhältnis dazu, wie lange ein Benutzer benötigt, um sich eine Präsentation anzusehen. Damit erhalten wir ein Maß für die Effizienz des Systems.

#### 8.1.1 Problemstellung

Das Ziel der Studie ist die Untersuchung der Effizienz des implementierten Home-Entertainment-Systems. Hierbei soll die Effizienz von Präsentationen auf gekoppelten Ausgabegeräten mit der Effizienz von Präsentationen auf einem einzelnen Ausgabegerät verglichen werden. Dazu soll als Kontrollgruppe ein einzelner Fernseher als Ausgabegerät und eine Tastatur als Eingabegerät betrachtet werden (Tastatur-TV). Der Fernseher soll weiterhin mit dem elektronischen Bilderrahmen (Tastatur-TV-Bilderrahmen) sowie mit einem PDA (PDA-TV) oder beiden (PDA-TV-Bilderrahmen) gekoppelt werden. In den Gerätekombinationen, die einen PDA enthalten, ersetzt der PDA die Tastatur als Eingabegerät. Damit ergeben sich vier Gruppen, die untereinander bzgl. ihrer Effizienz verglichen werden sollen.

Weiterhin sollen auf jeder Gerätekombination unterschiedliche Modalitätenkombinationen untersucht werden. Hierzu sollen im Experiment die Kombinationen Text-Bild (TB), Bild-Sprache (BS) und Bild-Sprache-animierter Charakter (BSC) betrachtet werden. Neben der Effizienz soll auch die Akzeptanz des Systems untersucht werden und mit den Ergebnissen der vorhergehenden Fragebogenstudie verglichen werden.

#### 8.1.2 Kognitive Last und Effizienz von gekoppelten Ausgabegeräten

In Abschnitt 7.1.2 wurde bereits ein Überblick über die Akzeptanz von gekoppelten Ausgabegeräten gegeben. Diese Darstellung wird in diesem Abschnitt durch die Betrachtung der Effizienz von gekoppelten Ausgabegeräten ergänzt.

## KAPITEL 8. LABORSTUDIE

In der von [Czerwinski et al., 2003] durchgeführten Studie wurde die Effizienz von großformatigen Bildschirmen untersucht. Um die kognitive Last der Vpn bei Verwendung eines 42 Zoll großen Bildschirms mit der eines 15 Zoll großen Bildschirms zu vergleichen, wurden ein Erinnerungsexperiment durchgeführt. Die Vpn wurden zu Beginn des Versuchs gebeten, sich eine Telefonnummer zu merken. Anschließend wurden die Vpn aufgefordert, nach einer Website mittels der AltaVista-Webseite zu suchen und diese aufzurufen. Auf der Webseite sollten die Vpn den Designer der Webseite ermitteln.

Der Name des Designers und die URL der Webseite sollten anschließend in ein Word-Dokument kopiert werden. Weiterhin sollte ein Screenshot der Webseite in ein Power Point-Dokument eingefügt werden. Bei dem Erinnerungsexperiment fanden Czerwinski und Kollegen keine signifikanten Unterschiede zwischen dem kleineren und dem größeren Monitor. Die Interaktionszeiten auf dem kleineren Monitor waren jedoch um 9% größer als auf dem größeren Monitor, was sich als signifikant erwies. Dies wurde mit der größeren Anzahl von Fensterinteraktionen begründet, die auf dem kleineren Monitor nötig waren, um die gestellten Aufgaben zu bearbeiten. Die Studie zeigt, dass sich eine Vergrößerung der Bildschirmfläche positiv auf die Interaktionszeiten auswirken kann. Die Studie ist jedoch auf das Windows-Betriebssystem und dessen Software beschränkt.

In der Studie von [Tan & Czerwinski, 2003] wurde die Kopplung von zwei 22 Zoll Monitoren und einem Projektionsschirm untersucht. Die untersuchten Gerätekombinationen waren ein Monitor, der Projektionsschirm, eine Kombination von beiden sowie zwei gekoppelte Monitore. Jede Vpn hat dabei alle Gerätekombinationen getestet. Die erste Aufgabe bestand darin, in einem Text Rechtschreibfehler zu finden. In der zweiten Aufgabe sollten die Teilnehmer auf eine Instant Messenger-Benachrichtigung möglichst schnell mit dem Drücken einer Taste reagieren. Die dritte Aufgabe bestand aus dem Vergleich zweier Texte, der je nach Gruppe auch über zwei Bildschirme hinweg durchgeführt wurde.

Die Zeiten für die Textvergleiche unterschieden sich nicht signifikant in den verschiedenen Gerätekombinationen. Ebenso unterschieden sich die Reaktionszeiten für die Instant Messenger-Benachrichtigungen nicht signifikant. Bei den Textvergleichen zeigte sich ein einzelner Monitor einer Kopplung von Monitor und Projektor marginal signifikant überlegen. Insgesamt waren die Effekte der Kopplungen der Bildschirme jedoch kleiner als von den Autoren erwartet. Die Autoren schließen, dass Designer anscheinend bei der Erstellung von Multi-Monitor-Präsentationen mehr Freiheiten haben als erwartet. Die Ergebnisse der Studie lassen sich als Hinweis interpretieren, dass Benutzer auch

## 8.1. PROBLEMSTELLUNG UND HYPOTHESEN

bei Verteilung von Informationen auf mehrere, unterschiedliche große Bildschirme effizient mit einem System arbeiten können.

Die in [Oonk et al., 2003] präsentierte Studie untersuchte die Effizienz von einem einzelnen Monitor sowie zwei bis sechs gekoppelten Monitoren in zwei Experimenten. Im ersten Experiment sollte mit Hilfe von Microsoft Office-Anwendungen ein Dokument erstellt werden, das einen militärischen Lagebericht enthält. Während des Experiments wurden den Teilnehmern Anfragen per Email oder per Chat gestellt, auf die die Teilnehmer möglichst schnell reagieren sollten. Hierbei wurde die Antwortzeit protokolliert sowie die Korrektheit der Antwort bewertet. Es zeigten sich signifikante Unterschiede bei der Korrektheit der Chat-Antworten. Diese war in der vier Monitor-Bedingung signifikant höher als in der drei Monitor-Bedingung. Dies korrespondiert ebenfalls mit der signifikanten Präferenz der Teilnehmer für das vier Monitor-System.

In einem zweiten Experiment mit identischem Aufbau lag der Schwerpunkt nicht auf der Erstellung eines Dokuments, sondern auf der Überwachung der taktischen Situation. Wie im ersten Experiment wurden Anfragen per Email und Chat gestellt, deren Antwortzeiten und Korrektheit protokolliert wurden. Im zweiten Experiment zeigten sich jedoch keine signifikanten Effizienzunterschiede.

In der Studie von [Kruppa, 2006] wurde ein animierter Charakter untersucht, der zwischen Präsentationen auf einem PDA und einem öffentlichen Bildschirm durch Migration vermittelt. Dabei haben jeweils drei Vpn zeitgleich eine Präsentation von nicht-technischen Inhalten betrachtet. Die Präsentation begann stets mit einer gemeinsamen audiovisuellen Präsentation auf dem Wandbildschirm und wurde für einen bestimmten Zeitraum durch Zusatzinformationen auf den individuellen PDAs ergänzt, die die Vpn in der Hand hielten.

Als Migrationsmöglichkeiten wurden ein animierter Charakter, ein Symbol oder gar kein Migrationssignal untersucht. Weiterhin wurden zwei Gerätekombinationen untersucht. Die erste Gruppe von Vpn verwendete Präsentationen, die auf den Wandbildschirm beschränkt waren. Die zweite Gruppe verwendete zusätzlich den PDA und migrierende Präsentationen als Stimuli. Die Effizienz der Präsentationen wurde mittels eines Erinnerungsexperiments untersucht. Nach der Präsentation wurden spezifische Fragen zu den präsentierten Textinhalten und Bildern gestellt. Anschließend wurden die Antworten auf ihre Korrektheit überprüft und bewertet. Zusätzlich wurden am Anfang einer Präsentation auf dem PDA für eine kurze Zeit Bilder angezeigt, an die sich die Vpn erinnern sollen.

## KAPITEL 8. LABORSTUDIE

Bei den Präsentationen auf parallelen Bildschirmen zeigte sich ein signifikanter Lerneffekt in der Gerätekombination PDA-Wandbildschirm. Nach der dritten Präsentation war die Erinnerungseffizienz ebenso groß wie bei Präsentationen auf dem Wandbildschirm alleine. Die drei Migrationsmöglichkeiten unterschieden sich weiterhin nicht signifikant bzgl. der Effizienz, mit der die am Anfang einer PDA-Präsentation gezeigten Bilder erinnert wurden. Aus den niedrigen Gesamtbewertungen schließt Kruppa, dass die Bilder zu kurz präsentiert wurden und empfiehlt, als Zeit für einen Fokuswechsel zwischen Geräten mindestens drei Sekunden anzusetzen.

Die Studie von Kruppa demonstriert, dass Benutzer nach kurzer Lernzeit audiovisuelle Präsentationen auf gekoppelten Ausgabegeräten ebenso effizient verarbeiten können wie Präsentationen auf einem einzelnen Gerät. Die Ergebnisse sind daher konform mit den Ergebnissen von [Tan & Czerwinski, 2003].

In [Elting et al., 2002] wurde die Effizienz von multi-modalen Präsentationen auf einem PDA, einem Fernseher und einem Desktop PC untersucht. Auf jedem Gerät wurden die Modalitätenkombinationen Text-Bild-Sprache, Bild-Sprache, Text-Bild, Text-Sprache und Text untersucht. Jede der 45 Vpn wurde einem Gerät zugeordnet. Anschließend wurden den Vpn in allen Kombinationen Sehenswürdigkeiten präsentiert. Jeder Inhalt bestand aus einem Titel, einem Textinhalt und einem Bild. In einem Fragebogen wurde abgefragt, an welche Beiträge sich die Teilnehmer erinnern. Dabei wurde zwischen Titel und Inhalt unterschieden. Weiterhin wurde die Zeit gemessen, die die Teilnehmer zum Betrachten der Information benötigt haben.

Von den drei untersuchten Geräten zeigten sich auf dem PDA die deutlichsten Effekte von kognitiver Last. Die Teilnehmer in dieser Gruppe benötigten signifikant mehr Zeit als die Teilnehmer in den anderen beiden Gruppen. Weiterhin zeigte sich, dass in der Bild-Sprache-Bedingung auf dem PDA signifikant mehr Titel von Sehenswürdigkeiten erinnert wurden als in der Text-Bild-Bedingung auf dem PDA. In den anderen Gerätegruppen trat dieser Effekt nicht auf.

Dieser Effekt wird mit dem Modalitäteneffekt begründet (Abschnitt 3.2.1). So könnte die relativ geringe Erfahrung der Teilnehmer mit dem PDA für kognitive Last gesorgt haben, wodurch der Modalitäteneffekt in der Bild-Sprache-Bedingung beobachtet werden konnte. Weiterhin nahm die Anzahl der erinnerten Titel im Laufe der fünf Bedingungen des Experiments signifikant ab. Das Experiment zeigt, dass der Typ des Interaktionsgeräts Einfluss auf die Effizienz einer Modalitätenkombination haben kann.

Bisher existieren jedoch keine Studien, die die Effekte unterschiedlicher Modalitätenkombinationen mit gekoppelten Ausgabegeräten in Zusammenhang

## 8.1. PROBLEMSTELLUNG UND HYPOTHESEN

gesetzt haben. [Tan & Czerwinski, 2003] untersuchte zwar verschiedene Kombinationen von Ausgabegeräten. Der Schwerpunkt lag hier jedoch auf der Untersuchung typischer Windows-Büroaufgaben. Spezielle Präsentationsformate wurden nicht untersucht. [Kruppa, 2006] untersucht zwar multi-modale Ausgaben auf zwei Gerätekombinationen. Der Schwerpunkt der Studie liegt jedoch auf der Migration von animierten Charakteren. Der in [Elting et al., 2002] beobachtete geräteabhängige Modalitäteneffekt legt jedoch nahe, dass sich die Modalitätenkombinationen auf unterschiedlichen Geräten verschieden auswirken.

### 8.1.3 Hypothesen

Aus den im vorherigen Abschnitt dargestellten Arbeiten lassen sich für die geplante Studie verschiedene Hypothesen ableiten. Analog zu [Tan & Czerwinski, 2003] wird erwartet, dass das Hinzufügen eines Bildschirms keine negativen Einflüsse auf die Erinnerungseffizienz der Vpn hat. Darüber hinaus wird vermutet, dass bei Präsentationen, die sich über mehrere Ausgabeschritte erstrecken, die Erinnerungseffizienz auf gekoppelten Ausgabegeräten höher ist als bei Präsentationen auf einem einzelnen Gerät.

Der Grund hierfür ist, dass bei gekoppelten Ausgabegeräten aufgrund der vergrößerten Bildschirmfläche weniger Ausgabeschritte notwendig sind und die Vpn weniger durch das Wechseln der Präsentationen abgelenkt sind. Weiterhin wird gemäß den Ergebnissen von [Czerwinski et al., 2003] angenommen, dass die Vpn effektiv mit einer größeren Bildschirmfläche interagieren können und dass sich durch Hinzufügen eines weiteren Bildschirms die Interaktionszeiten senken lassen.

Gemäß den Ergebnissen aus [Elting et al., 2002] wird erwartet, dass bei hoher kognitiver Last Sprache-Bild-Präsentationen gegenüber Text-Bild-Präsentationen überlegen sein werden. Hohe kognitive Last wird insbesondere bei Verwendung eines PDAs auftreten, da dieser das am wenigsten gewohnte Interaktionsgerät dargestellt. Weiterhin legen die Ergebnisse von [Van Mulken et al., 1998] nahe, dass animierte Charaktere durch den „Persona Effekt“ die kognitive Last einer Präsentation senken können.

Weiterhin wird erwartet, dass sich die Ergebnisse bzgl. der Akzeptanz von gekoppelten Ausgabegeräten aus Abschnitt 7.1.2 reproduzieren lassen. Daher erwarten wir, dass die gekoppelten Ausgabegeräte von den Benutzern gut akzeptiert werden. Dieses Ergebnis soll insbesondere im Hinblick auf einen Austausch der in der Fragebogenstudie verwendeten Mbrola Sprachausgabe durch eine AT&T Natural Voices 1.4 Sprachausgabe reevaluiert werden.

Auf dieser Grundlage werden die folgenden Hypothesen formuliert:

## KAPITEL 8. LABORSTUDIE

1. Die Erinnerungseffizienz der Vpn wird bei Kopplungen von zwei oder mehr Ausgabegeräten im Home-Entertainment-System höher sein als bei einem einzelnen Ausgabegerät.
2. Die Interaktionszeiten werden bei gekoppelten Ausgabegeräten niedriger sein als bei einem einzelnen Ausgabegerät.
3. Bei Kopplungen, die den PDA als Ausgabegerät enthalten, treten kognitive Lasteffekte auf. Diese Effekte sorgen für Unterschiede in der Erinnerungseffizienz zwischen den Modalitätenkombinationen Bild-Sprache und Text-Bild. Bild-Sprache wird sich dabei als effizienter erweisen.
4. Es wird erwartet, dass sich der animierte Charakter positiv auf die subjektiv empfundene, kognitive Last der Benutzer auswirkt.
5. Nicht jedes Ausgabegerät wird im Laufe einer Präsentation mit gekoppelten Ausgabegeräten gleich häufig betrachtet werden. Geräte, auf denen Text präsentiert wird, werden häufiger betrachtet werden als Geräte, auf denen Bilder präsentiert werden.
6. Wie in der Fragebogenstudie werden die Teilnehmer gekoppelte Ausgabegeräte als gut bewerten.
7. Wie bei der Fragebogenstudie werden sich bei der Akzeptanz der gekoppelten Ausgabegeräte Unterschiede bzgl. des Alters der Vpn ergeben.
8. Die neue AT&T Natural Voices 1.4 Sprachausgabe wird zu besseren Bewertungen der Sprachausgabe und des animierten Charakters als in der Fragebogenstudie führen.

### 8.2 Methode

Zur Untersuchung der Effizienz von Präsentationen auf gekoppelten Ausgabegeräten im Home-Entertainment-System wurden den Vpn Nachrichtenbeiträge präsentiert. Dabei wurden die Versuchspersonen in vier Gruppen unterteilt. Jede dieser Gruppen verwendete eine der vier Gerätekombinationen (Tastatur-TV, Tastatur-TV-Bilderrahmen, PDA-TV oder PDA-TV-Bilderrahmen). Der damit verbundene Aufbau der Studie wird in diesem Abschnitt detailliert beschrieben.

#### 8.2.1 Variablen

Im Folgenden werden die Variablen beschrieben, die im Experiment verwendet wurden. Dazu werden zunächst die abhängigen Variablen beschrieben, deren Werte durch das Experiment gemessen werden sollen. Anschließend

## 8.2. METHODE

werden die verwendeten unabhängigen Variablen dargestellt, die die unterschiedlichen Bedingungen des Experiments repräsentieren. Weiterhin werden die Moderatorvariablen beschrieben, die Einfluss auf die abhängigen Variablen nehmen können.

### Abhängige Variablen

In der Studie wurden die in Tabelle 8.1 dargestellten abhängigen Variablen verwendet. Jeder Nachrichtenbeitrag setzte sich aus einer Folge von mehreren Präsentationen zusammen, deren Anzahl je nach Gerätekombination variierte. Für jede dieser Präsentationen wurde die Dauer in Millisekunden gemessen, die die  $V_p$  zum Betrachten des Nachrichtenbeitrages benötigt. Dabei wurde die Zeit vom Darstellen der Präsentation bis zur Beendigung der Präsentation und der Einleitung der nächsten Präsentation durch den Benutzer gemessen.

Tabelle 8.1: Abhängige Variablen der Laborstudie.

| Kategorie           | Variablen                                                                                                                                                                                                                                      |
|---------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Präsentation        | Dauer der Betrachtung                                                                                                                                                                                                                          |
| Nachrichtenbeiträge | erinnerter Titel<br>erinnerter Inhalt<br>Lebhaftigkeit der Erinnerung<br>Attraktivität Ausgabemodalität<br>Attraktivität multi-modale Präsentation<br>empfundene kognitive Last<br>Interesse<br>Häufigkeit der Verwendung eines Ausgabegerätes |
| Gerätekombination   | Attraktivität<br>Bevorzugte Modalitätenkombination<br>Nützlichkeit von gekoppelten Ausgabegeräten<br>Bevorzugung von gekoppelten Ausgabegeräten                                                                                                |

Weiterhin wurden die erinnerten Titel und Inhalte der Nachrichtenbeiträge abgefragt. Der erinnerte Inhalt der Nachrichtenbeiträge konnte dabei als Freitext verfasst werden. Bei jedem Nachrichtenbeitrag wurde weiterhin die Lebhaftigkeit der Erinnerung abgefragt. Dies erfolgte mittels einer fünfstufigen Likert-ähnlichen Skala („sehr detailliert“ bis „sehr lückenhaft“).

Den Teilnehmern der Studie wurden die Nachrichtenbeiträge in verschiedenen Kombinationen der Ausgabemodalitäten Bilder, Sprachausgabe, animierter Helfer und Text präsentiert. Dabei wurde gemessen, ob die  $V_p$  die verwendeten Ausgabemodalitäten als attraktiv beurteilten. Dies erfolgte wie in

## KAPITEL 8. LABORSTUDIE

der Fragebogenstudie anhand einer fünfstufigen Likert-Skala (von „stimme völlig zu“ bis „stimme überhaupt nicht zu“).

Die Benutzer wurden ebenfalls nach der Attraktivität der gesamten multimodalen Präsentation des Nachrichtenbeitrages gefragt. Um die empfundene kognitive Last der Teilnehmer zu messen, wurden die Vpn gefragt, ob sie sich bei den Präsentationen sehr konzentrieren mussten. Die Messung erfolgte wiederum auf einer fünfstufigen Likert-Skala (von „stimme völlig zu“ bis „stimme überhaupt nicht zu“).

Die Vpn wurden weiterhin gefragt, ob der gezeigte Nachrichtenbeitrag interessant war. Dies wurde ebenfalls mittels einer fünfstufigen Likert-Skala erfasst. Die Benutzer wurden darüber hinaus gefragt, wie häufig sie jedes der verwendeten Ausgabegeräte (Fernseher, PDA oder elektronischer Bilderrahmen) verwendet hatten. Die Vpn wurden ebenfalls gebeten, die gesamte Gerätekombination zu beurteilen. Dazu wurden sie zunächst nach ihrer bevorzugten Modalitätenkombination befragt (BS, BSC oder TB).

Weiterhin wurden den Vpn die Frage gestellt, ob sie die Kopplung von Ausgabegeräten als nützlich empfunden haben. Dies wurde ebenfalls mittels einer fünfstufigen Likert-Skala untersucht. Die Benutzer wurden gefragt, ob sie gekoppelte Ausgabegeräte einem einzelnen Fernseher vorziehen würden. Dabei kam eine fünfstufige Likert-Skala zum Einsatz. Die Teilnehmer hatten weiterhin die Gelegenheit, nach jedem Nachrichtenbeitrag sowie am Ende des Versuchs Freitextkommentare zu verfassen.

### Unabhängige Variablen und Moderatorvariablen

Im Experiment wurden die folgenden unabhängigen Variablen verwendet. Die Effizienz der Präsentationen sollte zunächst anhand der Gerätekombinationen (Tastatur-TV, Tastatur-TV-Bilderrahmen, PDA-TV und PDA-TV-Bilderrahmen) untersucht werden. Weiterhin sollten die drei Präsentationsbedingungen (BS, BSC und TB) untersucht werden.

Die Effizienz und Akzeptanz des Home-Entertainment-Systems wurde in Abhängigkeit der folgenden Moderatorvariablen betrachtet. Zunächst wurde das Alter der Vpn berücksichtigt, da die Fragebogenstudie gezeigt hat, dass das Alter einen starken Einfluss auf die Beurteilung der Akzeptanz des Systems hatte. Ebenso wurde das Geschlecht der Vp berücksichtigt, da Männer und Frauen sich in der Beurteilung von Informationstechnologie unterscheiden können.

Um die Erfahrungen der Vpn mit den Ausgabegeräte einbeziehen zu können, wurden die Vpn nach der Häufigkeit gefragt, mit der sie Fernseher, PDAs, Computer oder elektronische Bilderrahmen bedienen. Die Variablen umfassen

## 8.2. METHODE

ten dabei acht Stufen (null bis sieben Tage pro Woche). Weiterhin wurden die Vpn gefragt, wie häufig sie Nachrichten lesen, hören oder betrachten (null bis sieben Tage pro Woche). Durch diese Variable wurde erfasst, wie vertraut die Vpn mit Nachrichten sind.

### 8.2.2 Teilnehmer

Die Teilnehmer der Studie wurden durch eine Anzeige in einer Heidelberger Tageszeitung sowie durch Aushänge in der Innenstadt über die Studie informiert. Die Teilnehmer erhielten eine Aufwandsentschädigung von 20 Euro.

### 8.2.3 Materialien

Im Folgenden werden die im Experiment verwendeten Materialien beschrieben. Zunächst werden die präsentierten Nachrichtenbeiträge beschrieben. Anschließend werden die verwendeten Gerätekombinationen und Fragebögen dargestellt.

#### Auswahl der zu präsentierenden Inhalte

Bei einem Erinnerungsexperiment ist die Auswahl der zu präsentierenden Inhalte von zentraler Bedeutung. Bei der Fragebogenstudie wurden die in Anhang A.1 beschriebenen Filminhalte präsentiert. Diese eigneten sich jedoch nicht für ein Erinnerungsexperiment.

Dies lag darin begründet, dass bei Filmen große Unterschiede bzgl. des Vorwissens der Benutzers existieren, die die Ergebnisse eines Erinnerungsexperiments verfälschen können. So würde sich z. B. eine Vp, die den Film „African Queen“ bereits vor dem Versuch kannte, wesentlich besser an den Inhalt erinnern als eine Vp, die den Film noch nicht kannte. Weiterhin waren die Beiträge für ein Erinnerungsexperiment nicht komplex genug, da ihr Inhalt nur aus drei bis vier Sätzen bestand, die eine zu geringe Informationsdichte aufwiesen.

Es wurden ebenfalls keine technische Inhalte (z. B. eine Anleitung zur Bedienung eines Funkgeräts) verwendet. Technische Inhalte wären zwar komplex genug für ein Erinnerungsexperiment gewesen, jedoch hätten die Inhalte nicht zu einem Home-Entertainment-System mit Fernseher gepasst, wodurch sich Irritationen bei den Vpn ergeben könnten.

Präsentationen von Nachrichtenbeiträgen stellen einen Kompromiss dar, da sie komplex genug für ein Erinnerungsexperiment sind und ebenfalls einen klaren Bezug zum Home-Entertainment haben. In der Studie wurden hierzu Beiträge des Jahres 2004 aus dem Tagesschau-Archiv im Internet<sup>1</sup> verwendet. Das Jahr 2004 wurde ausgewählt, um zu verhindern, dass die Vpn Vorwissen zu aktuellen politischen Geschehnissen einbringen können.

---

<sup>1</sup>tagesschau.de Archiv, <http://www.tagesschau.de/archiv>.

## KAPITEL 8. LABORSTUDIE

Tabelle 8.2: Nachrichtenbeitrag getrennt nach Segmenten (Quelle: Tagesschau.de).

| Segment | Text                                                                                                                    | Bild                                                                                |
|---------|-------------------------------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------|
| 1       | „Präsident Aristide ist auf Haiti von den Rebellen gestürzt worden und in die zentralafrikanische Republik geflüchtet.“ |    |
| 2       | „Mit Palmenzweigen und Jubelgesängen feierten Tausende ausgelassen den Sturz des Präsidenten.“                          |    |
| 3       | „Wenig später fährt ein Konvoi von 70 Rebellen schwer bewaffnet in Port-au-Prince ein.“                                 |    |
| 4       | „Ihr Führer Guy Philippe traf zu Gesprächen mit der demokratischen Opposition ein.“                                     |    |
| 5       | „Nach dem Sturz Aristides steht Haiti vor dem komplizierten Prozess eines Neuanfangs.“                                  |   |
| 6       | „Opposition und eine internationale Arbeitsgruppe wollen sich über die Bildung einer Übergangsregierung verständigen.“  |  |

Weiterhin wurden hauptsächlich Auslandsbeiträge verwendet. Es wurden keine Beiträge mit bekannten Personen des öffentlichen Lebens verwendet. Insgesamt wurden 24 Beiträge ausgewählt, die jeweils in sechs Sätze und sechs dazugehörige Bilder zerlegt wurden. Eine Kombination von einem Satz und

## 8.2. METHODE

einem Bild wird als Segment bezeichnet. Tabelle 8.2 zeigt einen der Beiträge. Gerätekombinationen

In der Studie wurden die Gerätekombinationen Tastatur-TV, Tastatur-TV-Bilderrahmen, PDA-TV und PDA-TV-Bilderrahmen untersucht. Jede Vp wurde einer dieser Gerätekombinationen zugeteilt. In der Gruppe Tastatur-TV hatten die Vpn nur mittels einer drahtlosen Tastatur und dem Loewe Aconda 9272 Fernseher mit dem System interagiert (Abbildung 8.1).



Abbildung 8.1: Geräte in der Tastatur-TV-Gruppe.

Von der drahtlosen Tastatur wurde lediglich ein Knopf benötigt, um mit der nächsten Ausgabe fortzufahren. Da die Vpn in den PDA-Gruppen aufgrund der eingeschränkten Audioqualität der PDA-Lautsprecher Kopfhörer tragen mussten, wurden auch die Vpn in den Gruppen Tastatur-TV und Tastatur-TV-Bilderrahmen gebeten, Kopfhörer zu tragen. Im Unterschied zur Tastatur-TV-Gruppe wurden in der Tastatur-TV-Bilderrahmen-Gruppe zusätzlich Bilder auf dem elektronischen DF-1710 Bilderrahmen dargestellt (Abbildung 8.2).



Abbildung 8.2: Geräte in der Tastatur-TV-Bilderrahmen-Gruppe.



Abbildung 8.3: Geräte in der PDA-TV-Gruppe.

## 8.2. METHODE

In der PDA-TV-Gruppe wurde die Tastatur durch den HP iPaq hx2750 PDA ersetzt (Abbildung 8.3). Der Fernseher wurde in dieser Gruppe nur zum Anzeigen von Bildern verwendet. An den PDA wurden Kopfhörer angeschlossen, da die Lautsprecher zu leistungsschwach waren, um die Sprachausgabe des Systems verständlich wiederzugeben. Das System wurde mittels eines Stiftes bedient, mit dem die Vpn Eingaben auf dem Touchscreen des PDAs vornehmen konnten. Die Vpn mussten damit den „Weiter“-Knopf auf dem PDA drücken, um mit der nächsten Ausgabe fortfahren zu können. In der PDA-TV-Bilderrahmen-Gruppe wurde die PDA-TV-Gruppe durch den elektronischen Bilderrahmen ergänzt, der eingesetzt wurde, um zusätzliche Bilder anzuzeigen (Abbildung 8.4).



Abbildung 8.4: Geräte in der PDA-TV-Bilderrahmen-Gruppe.

### Präsentationen

Auf jeder Gerätekombination kamen alle drei Modalitätenkombinationen (BS, BSC und TB) zum Einsatz. Jeder Nachrichtenbeitrag setzte sich aus jeweils sechs Segmenten zusammen (vgl. Tabelle 8.2).

## KAPITEL 8. LABORSTUDIE

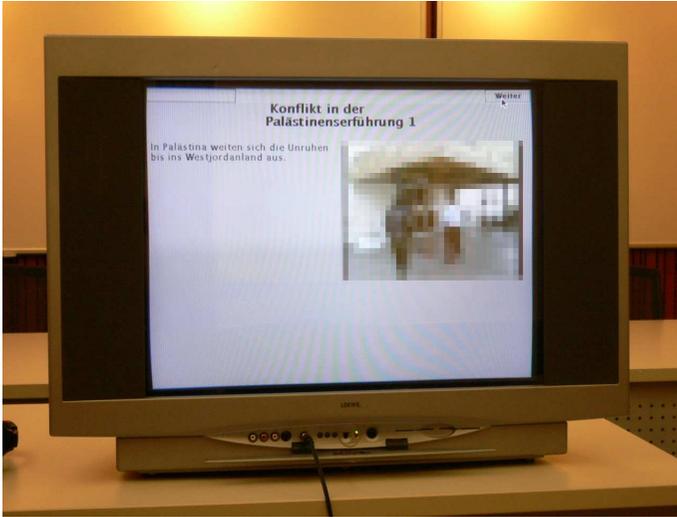


Abbildung 8.5: TB-Bedingung in der Tastatur-TV-Gruppe.



Abbildung 8.6: TB-Bedingung in der Tastatur-TV-Bilderrahmen-Gruppe.

## 8.2. METHODE

Die Anzahl der Segmente, die gleichzeitig bei einer Ausgabe dargestellt wurden, variierte je nach Gerätekombination. Dies lag in dem unterschiedlichen Layout-Platz begründet, der in den vier Gruppen vorhanden ist und der insbesondere die Anzahl der parallel darstellbaren Bilder begrenzt.

In der Tastatur-TV-Gruppe, die über nur einen Bildschirm als Ausgabegerät verfügt, wurde pro Ausgabe ein Segment eines Nachrichtenbeitrages dargestellt. Damit erfolgte die Ausgabe eines Nachrichtenbeitrages in dieser Gruppe in sechs Schritten (vgl. Gruppe Tastatur-TV, Bedingung TB in Abbildung 8.5).

In Gruppen mit zwei Bildschirmen wurden bei jeder Ausgabe zwei Segmente gleichzeitig dargestellt, d. h. die Ausgabe eines Nachrichtenbeitrags erfolgte in drei Schritten (vgl. Gruppe Tastatur-TV-Bilderrahmen, Bedingung TB in Abbildung 8.6). Bei der Ausgabe mittels dreier Bildschirme wurden drei Segmente gleichzeitig dargestellt (vgl. Gruppe PDA-TV-Bilderrahmen, Bedingung BS in Abbildung 8.7). Die Ausgabe eines Nachrichtenbeitrags erfolgte daher in zwei Schritten.

Tabelle 8.3: Anzahl gleichzeitig präsentierter Segmente getrennt nach Gruppe und Bedingung.

| Bedingung | Tastatur-TV | Tastatur-TV-BR | PDA-TV | PDA-TV-BR |
|-----------|-------------|----------------|--------|-----------|
| TB        | 1           | 2              | 1      | 2         |
| BS        | 1           | 2              | 2      | 3         |
| BSC       | 1           | 2              | 2      | 3         |

Ein Spezialfall ist die TB-Bedingung in den PDA-Gruppen (PDA-TV und PDA-TV-Bilderrahmen), in denen aufgrund des beschränkten Bildschirmplatzes auf dem PDA nur Text (ohne Bild) präsentiert wurde, wodurch für Bilder ein Bildschirm weniger zur Auswahl stand (vgl. Gruppe PDA-TV-Bilderrahmen, Bedingung TB in Abbildung 8.8). Dadurch konnten in dieser Bedingung nur zwei Segmente gleichzeitig dargestellt werden.

Tabelle 8.4: Benötigte Anzahl von Ausgabeschritten zur Präsentation eines Beitrages getrennt nach Gruppe und Bedingung.

| Bedingung | Tastatur-TV | Tastatur-TV-BR | PDA-TV | PDA-TV-BR |
|-----------|-------------|----------------|--------|-----------|
| TB        | 6           | 3              | 6      | 3         |
| BS        | 6           | 3              | 3      | 2         |
| BSC       | 6           | 3              | 3      | 2         |

## KAPITEL 8. LABORSTUDIE

Damit ergaben sich in Abhängigkeit der Gruppen und der Präsentationsbedingungen die in Tabelle 8.3 angegebenen Anzahlen von Segmenten, die bei einer Ausgabe gleichzeitig präsentiert werden konnten. Somit ergaben sich die folgenden Anzahlen von Ausgabeschritten, die nötig sind, um einen Nachrichtenbeitrag, der sich aus sechs Segmenten zusammensetzte, vollständig auszugeben (Tabelle 8.4).

In jeder Präsentation existierte ein „Weiter“-Knopf, mit dem der Benutzer zur nächsten Ausgabe gelangen konnte. In den Gruppen ohne PDA erschien dieser auf dem Fernsehbildschirm und konnte mittels eines Knopfes auf der drahtlosen Tastatur gedrückt werden. In den Gruppen mit PDA erschien der Knopf auf dem PDA-Bildschirm und konnte mittels des Eingabestiftes und des Touchscreens bedient werden. Weiterhin wurden sowohl auf dem Fernseher und dem PDA ein Fortschrittsbalken beim Laden einer neuen Ausgabe angezeigt.



Abbildung 8.7: BS-Bedingung in der PDA-TV-Bilderrahmen-Gruppe.

In der TB-Bedingung wurden der Titel des Nachrichtenbeitrags zusammen mit der fortlaufenden Nummer der Ausgabe (z. B. "Konflikt in der Palästinenserführung 1") dargestellt sowie der dazugehörige Text und die dazugehörigen Bilder (vgl. Gruppe Tastatur-TV, Abbildung 8.5).

## 8.2. METHODE



Abbildung 8.8: TB-Bedingung in der PDA-TV-Bilderrahmen-Gruppe.

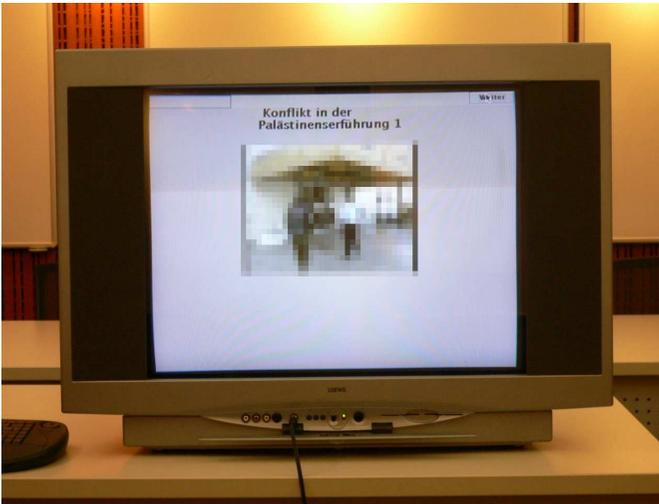


Abbildung 8.9: BS-Bedingung in der Tastatur-TV-Gruppe.

## KAPITEL 8. LABORSTUDIE

In der BS-Bedingung wurden derselbe Titel und dieselben Bilder dargestellt wie in der TB-Bedingung. Der Text wurde jedoch nicht visuell dargestellt, sondern durch eine Sprachausgabe auditoriell wiedergegeben (vgl. Gruppe Tastatur-TV, Abbildung 8.9). Aufgrund der negativen Beurteilung der Sprachausgabe in der Fragebogenstudie wurde die Mbrola Sprachausgabe durch eine AT&T Natural Voices 1.4 Sprachausgabe ersetzt. In der BSC-Bedingung wurde die BS-Bedingung durch den animierten Charakter ergänzt (vgl. Gruppe Tastatur-TV, Abbildung 8.1).

### Fragebögen

Es wurden den Vpn insgesamt drei Typen von elektronischen Fragebögen präsentiert. Mittels des Eingangsfragebogens (Anhang C.1) wurden die deskriptiven Daten der Vpn abgefragt. Ein Zwischenfragebogen (Anhang C.3) wurde verwendet, um die Vpn nach jeder der drei Präsentationsbedingungen (TB, BS und BSC) zu befragen. Dabei wurde insbesondere abgefragt, an wie viele Nachrichtenbeiträge sich die Vpn erinnern. Am Ende der Studie wurden die Vpn in einem Abschlussfragebogen gebeten, die Gerätekombination als Ganzes zu beurteilen (Anhang C.4). Alle Fragebögen wurden an einem Laptop ausgefüllt, der sich neben der jeweils getesteten Gerätekombination befand.

Neben dem Eingangsfragebogen wurden die Vpn zweimal über den Versuchsablauf instruiert. Die erste Instruktion fand am Laptop nach dem Eingangsfragebogen statt (Abbildung C.2). Die zweite Instruktion fand in den Gruppen Tastatur-TV und Tastatur-TV-Bilderrahmen auf dem Fernseher statt (Anhang C.2). In den Gruppen PDA-TV und PDA-TV-Bilderrahmen fand die zweite Instruktion auf dem PDA statt.

Beim Eingangsfragebogen wurden die Vpn gebeten, Alter und Geschlecht anzugeben. Darüber hinaus wurden sie gefragt, an wie vielen Tagen pro Woche sie einen PDA, einen Fernseher, einen Computer oder einen elektronischen Bilderrahmen verwenden. Weiterhin wurden sie gefragt, wie häufig sie pro Woche Nachrichten schauen, lesen oder hören. Im Zwischenfragebogen wurden die Vpn zu der zuletzt gesehenen Präsentationsbedingung (BS, BSC oder TB) befragt.

Im ersten Teil wurden sie gebeten anzugeben, ob ihnen die Textausgabe, die Bilder, die Sprachausgabe und der animierte Helfer gefallen haben. Dabei wurde bei jeder Vp nur nach den Modalitäten gefragt, die die Vp in der zuvor gezeigten Präsentationsbedingung gesehen hat. Die Vpn konnten zusätzlich freie Kommentare zu den gesehenen Präsentationen in ein Textfeld schreiben. Anschließend wurden die Vpn gefragt, wie interessant sie die gesehenen Nachrichtenbeiträge unabhängig von der Art der Präsentation fanden. Abbildung C.6 zeigt den Fragebogen für die BSC-Bedingung.

## 8.2. METHODE

Im zweiten Teil des Zwischenfragebogens (Abbildung C.7) wurden die Vpn gebeten, die Titel der Nachrichten anzugeben, an die sie sich erinnern können. Dabei konnten die Vpn bis zu sechs Titel angeben, da ihnen sechs Nachrichtenbeiträge präsentiert worden sind. Im dritten Teil wurden sie zu jedem erinnerten Nachrichtentitel gebeten, den Inhalt anzugeben, den sie zu diesem Titel noch erinnerten (Abbildung C.8).

Im vierten Teil des Zwischenfragebogens wurden die Vpn gebeten anzugeben, wie häufig sie jedes der Ausgabegeräte (Fernseher, Bilderrahmen oder PDA) bei der letzten Präsentationsbedingung verwendet haben. Dabei wurde nur nach den Geräten gefragt, die in der Gruppe der Vpn auch tatsächlich verwendet wurden. In der Tastatur-TV-Gruppe wurde dieser Teil übersprungen, da hier nur ein Ausgabegerät verwendet wurde. Abbildung C.9 zeigt den Fragebogen für die PDA-TV-Bilderrahmen-Gruppe.

Im Abschlussfragebogen (Abbildung C.10) wurden die Vpn am Ende des Versuchs gebeten, die Gerätekombination als Ganzes zu bewerten. Zunächst wurden die Vpn gefragt, wie gut ihnen die Ausgaben auf der Gerätekombination insgesamt gefallen haben. Anschließend wurde ihnen die Frage gestellt, welche Präsentationsbedingung ihnen am besten gefallen hat (BS, BSC oder TB). In allen Gruppen außer Tastatur-TV wurden die Vpn gefragt, wie nützlich sie die Kombination mehrerer Geräte empfanden und ob sie bei der Ausgabe von Nachrichten die Kombination mehrerer Geräte gegenüber einem Fernseher vorziehen würden. Abschließend konnten die Vpn allgemeine Kommentare zum Versuch in ein Textfeld schreiben.

### 8.2.4 Durchführung

Da die drei Bedingungen BS, BSC und TB nacheinander stattfinden sollten, wurde die Reihenfolge der drei Bedingungen über alle Vpn balanciert, damit keine Bedingung benachteiligt ist. Dies bedeutet, dass –über alle Vpn betrachtet– jede Bedingung gleich oft als erste, zweite oder dritte Bedingung zum Einsatz kam. Im Folgenden wird die resultierende Balancierung des Versuchsaufbaus beschrieben sowie der zeitliche Ablauf des Versuchs.

#### Balancierung der Präsentationsbedingungen

In dem balancierten Versuchsaufbau erschien jede Präsentationsbedingung in jeder Gruppe insgesamt dreimal an jeder der drei möglichen Positionen. Analog zu den Präsentationsbedingungen mussten auch die Nachrichtenbeiträge balanciert werden, damit ein Nachrichtenbeitrag, an den sich die Vpn besonders gut erinnern können, nicht stets als erster, zweiter oder dritter präsentiert wurde. Die Nachrichtenbeiträge wurden in jeder der drei Präsentationsbedingungen in Folgen von jeweils sechs Nachrichtenbeiträgen präsentiert.

## KAPITEL 8. LABORSTUDIE

Tabelle 8.5: Reihenfolge der Präsentationsbedingungen (TB, BS, BSC) und der Folgen von Nachrichtenbeiträgen (1, 2, 3) im Versuch.

| Vp | Zeitpunkt 1 | Zeitpunkt 2 | Zeitpunkt 3 |
|----|-------------|-------------|-------------|
| 1  | TB<br>2     | BS<br>3     | BSC<br>1    |
| 2  | BSC<br>2    | BS<br>1     | TB<br>3     |
| 3  | BS<br>1     | TB<br>3     | BSC<br>2    |
| 4  | BSC<br>1    | TB<br>2     | BS<br>3     |
| 5  | BS<br>3     | BSC<br>1    | TB<br>2     |
| 6  | TB<br>3     | BSC<br>2    | BS<br>1     |
| 7  | TB<br>2     | BS<br>3     | BSC<br>1    |
| 8  | BSC<br>1    | TB<br>2     | BS<br>3     |
| 9  | BS<br>3     | BSC<br>1    | TB<br>2     |

Die resultierende Verteilung der Präsentationsbedingungen BS, BSC und TB sowie der drei Folgen von Nachrichtenbeiträgen (mit Nummern 1, 2, 3) wird in Tabelle 8.5 dargestellt. Dabei kommt jede Präsentationsbedingung und jede Folge von Nachrichtenbeiträgen je dreimal an jedem der drei Zeitpunkte vor. Somit wurden für jede der Gruppen neun Vpn benötigt. Da das Experiment vier Gruppen umfasst (Tastatur-TV, Tastatur-TV-Bilderrahmen, PDA-TV und PDA-TV-Bilderrahmen) wurden für das gesamte Experiment 36 Vpn benötigt.

### Versuchsablauf

Am Anfang erfolgte eine kurze Einführung durch den Versuchsleiter. Anschließend wurden die Vpn gebeten, den Eingangsfragebogen (Abbildung C.1) auszufüllen. Danach erfolgte eine detaillierte Einführung mit Instruktionen zum Versuch (Abbildung C.2). Die Vpn wurden darauf hingewiesen, dass ihr Wissen abgefragt wird und die Zeiten für die Betrachtung einer Ausgabe protokolliert werden. Weiterhin wurden sie darauf hingewiesen, dass es sich bei

## 8.2. METHODE

dem Testsystem um einen Prototypen handelt, der noch keine Serienreife besitzt.

Anschließend wurden die Vpn gebeten, vom Laptop zu den Geräten zu wechseln. Vor den Geräten erfolgte eine weitere Einweisung in den Versuch (Abbildungen C.3, C.4 und C.5), bei der die Vpn den „Weiter“-Button drücken mussten, um zur nächsten Ausgabe zu gelangen. Dies diente dazu, die Vpn mit der Eingabemodalität (drahtlose Tastatur oder PDA-Stift) vertraut zu machen. Anschließend erfolgte ein Testdurchlauf mit einer Testpräsentation, die aus einem Nachrichtenbeitrag bestand, dessen Inhalt für alle Vpn identisch war. Der Testdurchlauf erfolgte in allen Gruppen in der TB-Bedingung.

Nach dem Testdurchlauf verließ der Versuchsleiter den Raum und ging in einen benachbarten Raum, der schalldicht und durch ein Glasfenster vom Testraum getrennt ist. Anschließend erfolgte die Präsentation des ersten Nachrichtenbeitrages, die sich je nach Gruppe aus zwei bis sechs Ausgaben zusammensetzte (vgl. Tabelle 8.4). Insgesamt wurden sechs Nachrichtenbeiträge präsentiert. Um die Nachrichtenbeiträge untereinander zu trennen, erfolgte nach jedem Nachrichtenbeitrag die Meldung „Bitte drücken Sie den 'Weiter'-Button, um mit dem nächsten Beitrag fortzufahren“.

Nach der Präsentation wurde die Vp mittels eines Bildschirmtextes aufgefordert, zurück zum Laptop zu wechseln. Dort erhielt die Vp einen weiteren Fragebogen, der an die vorherige Präsentationsbedingung angepasst war (vgl. Abbildung C.6 für die BSC-Bedingung). Die Vp wurde gebeten, die vorherige Präsentationsbedingung zu bewerten.

In einem weiteren Fragebogen wurde die Vp gebeten, die erinnerten Daten zu reproduzieren. Zunächst wurden dazu die Titel der Nachrichtenbeiträge abgefragt, an die sich die Vp noch erinnert (Abbildung C.7). Anschließend wurde zu jedem angegebenen Nachrichtentitel abgefragt, an welchen Inhalt sich die Vp erinnert (Abbildung C.8). Weiterhin konnte die Vp in einem Textfeld freie Anregungen oder Kritik eintragen.

In allen Gruppen außer Tastatur-TV wurde die Vp anschließend gefragt, wie häufig sie jedes Ausgabegerät in der letzten Präsentationsbedingung verwendet hat (Abbildung C.9). Anschließend wurde die Vp gebeten, zurück zu den Geräten zu wechseln und sich die nächsten sechs Nachrichtenbeiträge anzuschauen.

Dieser Ablauf wurde insgesamt dreimal wiederholt, bis die Vp alle Bedingungen (BS, BSC und TB) durchlaufen hat. Am Ende des Versuchs wurde die Vp gebeten, den Abschlussfragebogen auszufüllen, der zur Bewertung des Gesamtsystems diente (Abbildung C.10). Danach erfolgte eine Abschlussbesprechung mit dem Versuchsleiter.

## Störfaktoren

Während des Experiments traten zwei Typen von Störfaktoren auf. Da der Fernseher und der PDA durch die SMIL-Präsentationen stark ausgelastet waren, erfolgten im Laufe des Experiments gelegentlich Abstürze der verwendeten Java Virtual Machine. Gegen Ende des Versuchsplans fiel bedingt durch einen Hardware-Schaden am Mainboard der elektronische DF-1710 Bilderrahmen irreparabel aus. Der elektronische Bilderrahmen wurde durch einen umkonfigurierten Toshiba M200 Portégé Tablet PC ersetzt. Dieser verfügte jedoch lediglich über einen 12.1 Zoll großen Bildschirm im Gegensatz zum 17 Zoll großen Bildschirm des Bilderrahmens. Der umkonfigurierte Tablet PC wird auf den Abbildungen A.17-A.20 dargestellt. Bei beiden Faktoren wurde festgehalten, bei welchen Vpn und in welchen Präsentationsbedingungen sie auftraten, so dass sie bei der Auswertung berücksichtigt werden konnten.

### 8.2.5 Bewertung der erinnerten Filminhalte

Innerhalb des Zwischenfragebogens wurden die Vpn gebeten, die Filmtitel und die Inhalte der Präsentationen anzugeben, an die sie sich erinnern. Um die Ergebnisse verschiedener Vpn vergleichen zu können, wurden die Erinnerungen jeder Vp für jeden Nachrichtenbeitrag mit einer positiven, ganzen Zahl bewertet. Tabelle 8.6 fasst die maximal möglichen Bewertungen pro Präsentationsbedingung mit sechs Nachrichtenbeiträgen zusammen.

Tabelle 8.6: Maximal mögliche Bewertung der Erinnerungen pro Präsentationsbedingung.

| Kategorie  | max. Bewertung     |
|------------|--------------------|
| Titel      | $6 \times 2 = 12$  |
| Textinhalt | $6 \times 12 = 72$ |
| Bilder     | $6 \times 6 = 36$  |
| $\Sigma$   | 120                |

Für jeden erinnerten Titel wurden bis zu zwei Punkte vergeben. Die volle Punktezahl wurde bei wortwörtlicher Übereinstimmung des erinnerten Titels mit dem Titel des Beitrags vergeben. Bei einem inhaltlich richtig erinnerten Titel, der nicht wortwörtlich mit dem Beitragstitel übereinstimmt, wurde ein Punkt vergeben (z. B. „Geiselnahme in Beslan“ anstelle von „Geiseldrama in Russland“). Bei einem inhaltlich falsch erinnerten Titel wurden keine Punkte vergeben (z. B. „Wahl in Russland“ anstelle von „Geiseldrama in Russland“). Somit konnte eine Vp pro Präsentationsbedingung maximal zwölf Punkte für richtig erinnerte Titel erreichen.

## 8.2. METHODE

Bei der Erinnerung des Textinhalts wurden die erinnerten Daten der Vpn mit den sechs Segmenten jedes Nachrichtenbeitrags verglichen (vgl. Tabelle 8.2). Pro Segment des Nachrichtenbeitrags wurden bis zu zwei Punkte vergeben, d. h. für den Textinhalt eines Nachrichtenbeitrags konnten bis zu zwölf Punkte erreicht werden. Wurde durch die Vp ein Segment korrekt wiedergegeben, so wurde dies mit zwei Punkten bewertet. Das Segment musste dazu vollständig und inhaltlich richtig wiedergegeben werden (z. B. „Nichtrauchen als Normalzustand von Regierung nicht erreicht“ anstelle von „So ist die Regierung immer noch weit entfernt von dem Ziel, Nichtrauchen zum Normalfall zu machen.“).

Fehlten Bestandteile von Segmenten oder waren Teile inhaltlich falsch, so wurde für das Segment ein Punkt vergeben (z. B. „Kinder sind unterernährt“ anstelle von „Betroffen sind vor allem die Kinder, da die nährstoffarme Kost ihr Immunsystem immens geschwächt hat.“). Für vollständig falsche Segmente wurden keine Punkte vergeben. Somit konnten für den Textinhalt der Präsentationen pro Präsentationsbedingung bis zu 72 Punkte erreicht werden.

Neben dem Titel und dem textuellen Inhalt der Nachrichtenbeiträge wurden für die Erinnerung von Bildern Zusatzpunkte vergeben. Bezog sich eine Aussage eindeutig auf ein Bild, so wurde hierfür ein weiterer Punkt vergeben. Somit konnten zusätzlich maximal bis zu sechs Punkte pro Nachrichtenbeitrag bzw. 36 Punkte in der Präsentationsbedingung für die Erinnerung von Bildern erreicht werden. Damit konnten in einer Präsentationsbedingung für die Erinnerung von Titeln, Inhalten und Bildern insgesamt maximal 120 Punkte erreicht werden. Im gesamten Experiment konnten maximal 360 Punkte erreicht werden.

### 8.2.6 Statistische Auswertung

Wie in der Fragebogenstudie wurden hauptsächlich parameterfreie Verfahren zur Evaluierung der Likert-gestufteten Variablen verwendet (vgl. Abschnitt 7.2.5). Hierbei kamen der Mann-Whitney-Test für zwei unabhängige Stichproben über Ordinalskalen, der Wilcoxon-Test für zwei abhängige Stichproben über Ordinalskalen sowie der Friedman-Test für k abhängige Stichproben über Ordinalskalen zum Einsatz.

In dieser Studie wurde ebenfalls der Kruskal-Wallis-Test für k unabhängige Stichproben über Ordinalskalen verwendet. Mit dem Kruskal-Wallis-Test können z. B. die Bewertungen der vier Gerätekombinationen im Abschlussfragebogen auf signifikante Unterschiede untersucht werden. Weiterhin wurde der  $\chi^2$ -Test zur Untersuchung der Signifikanz von Nominaldaten eingesetzt.

Wie in der Fragebogenstudie wurden die Likert-Skalen zur besseren grafischen Darstellung durch eine Abbildung auf eine äquidistante Skala über gan-

zen Zahlen (Werte eins bis fünf) in Intervallskalen überführt. Der Wert eins entsprach einer völligen Ablehnung der vorangestellten Aussage während der Wert fünf einer völligen Zustimmung entsprach. Diese intervallskalierten Daten wurden wie in der Fragebogenstudie ebenfalls mittels t-Test und Anova untersucht. Bei der Anova wurde eine Greenhouse-Geisser-Korrektur durchgeführt, falls die Zirkularitätsannahme verletzt war.

Wurde die Anova mit mehr als einem Faktor durchgeführt (z. B. falls Interaktionen zwischen den Faktoren Gerätekombination und Präsentationsbedingung untersucht werden), so wurden diese Faktoren der Übersicht halber den Ergebnissen hinzugefügt (z. B. Gerätekombination $\times$ Präsentation). Wie in der Fragebogenstudie wurden zusätzlich die Mittelwerte (MW) und die Standardabweichungen (SD) der relevanten Variablen angegeben, sofern diese nicht aus den Abbildungen hervorgingen. Bei allen Tests wurde ein Signifikanzniveau von  $\alpha = 0.05$  verwendet.

Neben Histogrammen wurden Balkendiagramme mit 95%-Konfidenzintervallen für die Mittelwerte verwendet (vgl. Abschnitt 7.2.5). Weiterhin wurden Intervalldaten (z. B. die gemessenen Betrachtungsdauern) mittels Box-Plots visualisiert. Die Visualisierung mittels Box-Plots basiert auf Perzentilen. Das Perzentil  $P_x$  bezeichnet den Messpunkt, der die Verteilung im Verhältnis  $x\%$  zu  $(100 - x)\%$  teilt. Der Median stellt somit  $P_{50}$  dar. Die Boxplots stellen den Median einer Verteilung innerhalb eines Kastens als waagerechten Strich dar. Der Kasten wird unten durch  $P_{25}$  und oben durch  $P_{75}$  begrenzt, d. h. der Kasten umschließt die mittleren 50% der Messpunkte. Aus dem Kasten gehen zwei Balken hervor, die unten durch  $P_5$  und oben durch  $P_{95}$  begrenzt werden. Extremwerte, die sich zwischen 1.5 bis drei Kastenlängen oberhalb der oberen Kante bzw. unterhalb der unteren Kante befinden, werden separat als Extremwerte in Form von Punkten im Diagramm hervorgehoben. Extremwerte werden durch einen Stern als Ausreißer gekennzeichnet, falls sie mehr als drei Kastenlängen vom oberen bzw. unteren Rand des Kastens entfernt sind.

Zur Visualisierung der Abhängigkeiten zwischen mehreren intervallskalierten Faktoren (z. B. die Bewertungen der Erinnerungen in Abhängigkeit vom Alters der Vp) wurden weiterhin Streudiagramme eingesetzt. In diesen Diagrammen werden die Messdaten in Form von Punktwolken dargestellt. Zur Verdeutlichung von Tendenzen in den Punktwolken wurden Loess-Interpolationen der Messpunkte verwendet. Weitere Erläuterungen zu den verwendeten Testverfahren und Diagrammen finden sich in [Bortz, 2005].

### 8.3 Ergebnisse

Dieser Abschnitt umfasst die Ergebnisse der statistischen Signifikanztests. Zunächst erfolgt eine Beschreibung der Stichprobe. Anschließend werden die

### 8.3. ERGEBNISSE

Ergebnisse gliedert nach den relevanten abhängigen Variablen dargestellt. Abschließend erfolgt eine qualitative Analyse der Kommentare der Vpn.

#### 8.3.1 Beschreibung der Stichprobe

An der Untersuchung nahmen 36 Personen teil, darunter 18 Männer. Der Altersdurchschnitt betrug 30.8 Jahre ( $SD = 12.3$ ). Der jüngste Teilnehmer war 12, die älteste Teilnehmerin 67 Jahre alt. Abbildung 8.10 zeigt die Altersstruktur der Studie.

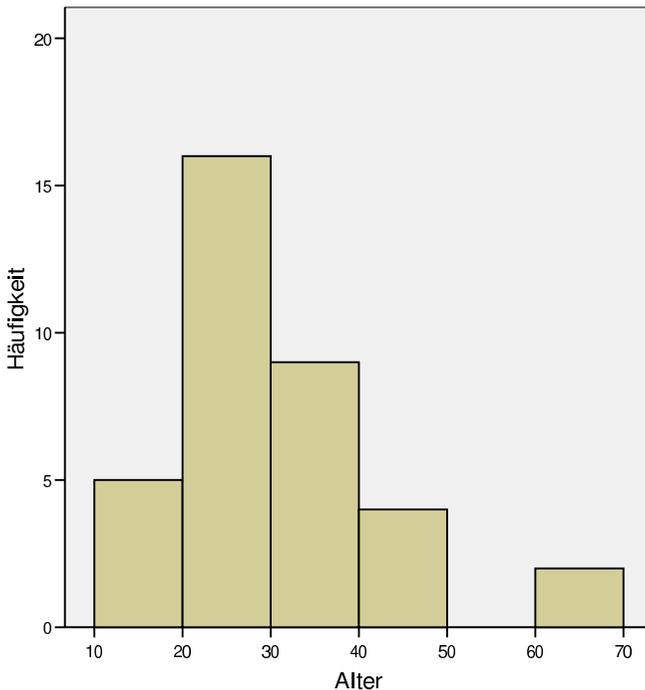


Abbildung 8.10: Histogramm für Altersintervalle von zehn Jahren.

#### 8.3.2 Eingangsfragebogen

Abbildung 8.11 stellt die mittlere Dauer der Nutzung pro Woche von PDAs, Fernsehern, Computern und elektronischen Bilderrahmen dar. Die Balken zeigen die 95%-Konfidenzintervalle. Keine der Vpn hatte angegeben, einen elektronischen Bilderrahmen zu verwenden. Die Unterschiede zwischen den Geräten waren signifikant (Friedman,  $p = .001$ ).

## KAPITEL 8. LABORSTUDIE

Die Häufigkeit der Geräteverwendung wurde auf geschlechtsspezifische Unterschiede untersucht. Es ergaben sich weder beim PDA noch dem Fernseher oder dem Computer signifikante Unterschiede (Mann-Whitney; PDA:  $p = .393$ ; TV:  $p = .077$ ; PC:  $p = .220$ ). Von den 36 Vpn gaben 30 Vpn bei der Nutzung des PDAs einen Wert von null an, d. h. sie nutzen keinen PDA.

Nachrichten wurden von den Vpn durchschnittlich an 5.3 Tagen in der Woche geschaut, gehört oder gelesen ( $SD = 2.1$ ). Auch in der Häufigkeit des Nachrichtenkonsums ergaben sich keine Unterschiede zwischen Männern und Frauen (Mann-Whitney:  $p = .100$ ).

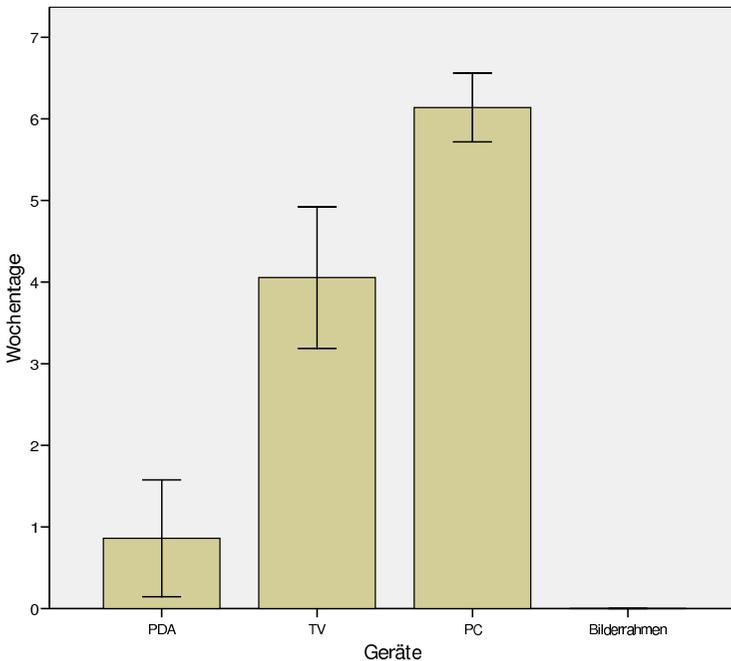


Abbildung 8.11: Erfahrungen mit PDAs, Fernsehern, Computern und elektronischen Bilderrahmen.

### 8.3.3 Erinnerung von Filminhalten

Die Erinnerungen der Filminhalte wurden wie in Abschnitt 8.2.5 beschrieben durch zwei Hilfskräfte bewertet. Neben den Gesamtbewertungen der erinnerten Nachrichtenbeiträge wurden bei der Auswertung die Bewertungen der erinnerten Bilder auch separat betrachtet.

### 8.3. ERGEBNISSE

#### Gesamtbewertung

Die Gesamtbewertung der Beiträge setzte sich aus den Bewertungen der erinnerten Titel, Inhalte und Bilder der Nachrichtenbeiträge zusammen. Durchschnittlich erreichte eine Vp 71.4 (SD = 29.4) von 360 maximal möglichen Punkten im gesamten Experiment. Die minimal erreichte Bewertung betrug 23. Die maximal erreichte Bewertung betrug 150.

Es ergaben sich keine signifikanten Unterschiede bei der Gesamtbewertung der erinnerten Nachrichtenbeiträge zwischen den Gerätekombinationen (Anova,  $F(3, 35) = .08, p = .970$ ). Weiterhin hatte das Alter der Vpn keinen signifikanten Einfluss auf die Gesamtbewertung der erinnerten Beiträge (Anova,  $F(23, 35) = .73, p = .750$ ). Es gab ebenfalls keine signifikanten Unterschiede bzgl. des Geschlechts der Vpn (Anova,  $F(1, 35) = 2.02, p = .164$ ).

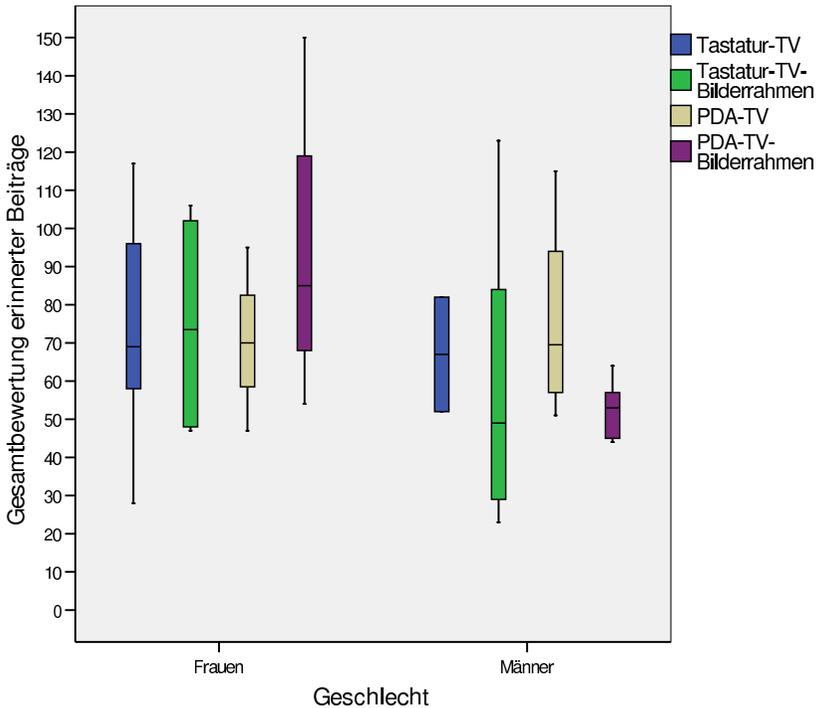


Abbildung 8.12: Gesamtbewertung erinnertes Beiträge getrennt nach Geschlecht und Gerätekombination.

## KAPITEL 8. LABORSTUDIE

Es wurden keine signifikanten Interaktionen zwischen dem Geschlecht und der Gerätekombination bzgl. der Gesamtbewertung der erinnerten Beiträge festgestellt (Anova,  $\text{Geschlecht} \times \text{Gerätekombination}$ ,  $F(3, 28) = .88$ ,  $p = .465$ ). Die Unterschiede werden in Abbildung 8.12 grafisch dargestellt. Die Ergebnisse in der PDA-TV-Bilderrahmen-Gruppe deuten einen Effekt des Faktors Geschlecht an, waren jedoch nicht signifikant (Anova,  $F(1, 7) = 63.90$ ,  $p = .060$ ).

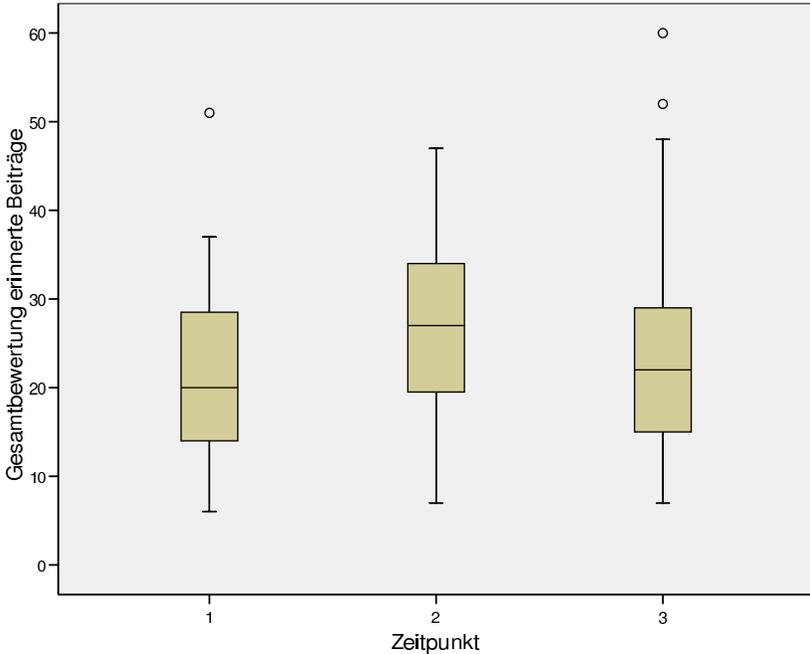


Abbildung 8.13: Gesamtbewertung erinnerten Beiträge getrennt nach Zeitpunkten der Präsentation.

Es zeigten sich signifikante Unterschiede zwischen den drei möglichen Zeitpunkten der Präsentationen (vgl. Tabelle 8.5, Anova,  $F(2, 70) = 7.78$ ,  $p = .001$ ). Abbildung 8.13 stellt die Unterschiede grafisch mittels eines Box-Plots dar. Dabei waren die Unterschiede zwischen den Zeitpunkten eins und zwei signifikant (t-Test; Zeitpunkt 1&2:  $t(35) = -4.77$ ,  $p = .001$ ; Zeitpunkt 1&3:  $t(35) = -1.87$ ,  $p = .070$ ; Zeitpunkt 2&3:  $t(35) = -1.81$ ,  $p = .079$ ). Es

### 8.3. ERGEBNISSE

existierten keine signifikanten Interaktionen zwischen den Faktoren Zeitpunkt der Präsentation und der Gerätekombinationen bzgl. der Gesamtbewertung der erinnerten Beiträge (Anova,  $\text{Zeitpunkt} \times \text{Gerätekombination}$ ,  $F(3, 83) = .81$ ,  $p = .970$ ).

Darüber hinaus ergaben sich keine signifikanten Interaktionen zwischen den Faktoren Zeitpunkt der Präsentation und dem Geschlecht der Vp (Anova,  $\text{Zeitpunkt} \times \text{Geschlecht}$ ,  $F(2, 86) = .28$ ,  $p = .758$ ). Die Interaktion des Faktors Zeitpunkt der Präsentation mit dem Alter der Vp hatte ebenfalls keinen signifikanten Einfluss auf die Gesamtbewertung der erinnerten Beiträge (Anova,  $\text{Zeitpunkt} \times \text{Alter}$ ,  $F(23, 12) = .731$ ,  $p = .750$ ). Neben dem Faktor Zeitpunkt wurde ebenfalls untersucht, ob signifikante Unterschiede bzgl. der Gesamtbewertungen der erinnerten Beiträge zwischen den drei verwendeten Folgen von Nachrichtenbeiträgen existieren (vgl. Tabelle 8.5). Dies war nicht der Fall (Anova,  $F(2, 70) = .576$ ,  $p = .565$ ).

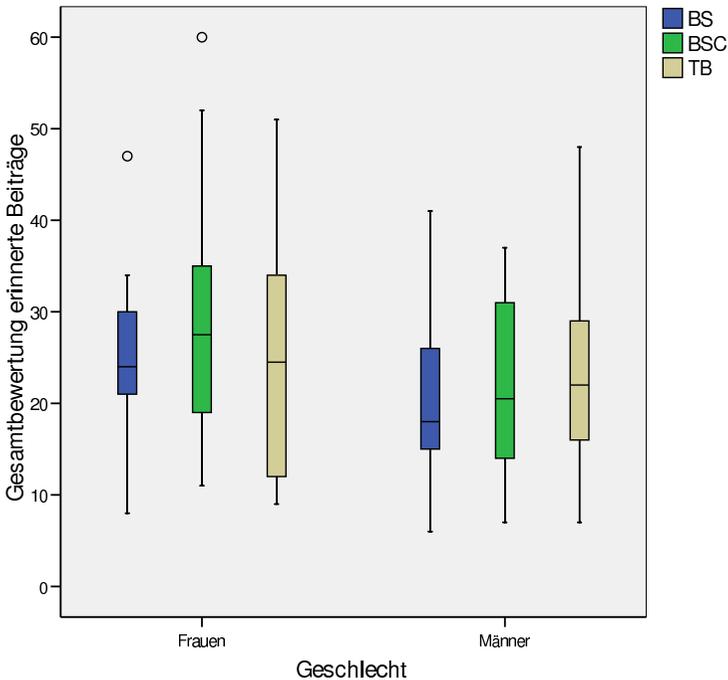


Abbildung 8.14: Gesamtbewertung erinnerten Beiträge getrennt nach Geschlecht und Präsentationsbedingungen.

## KAPITEL 8. LABORSTUDIE

Zwischen den Präsentationsbedingungen BS, BSC und TB existierten keine signifikanten Unterschiede bzgl. der Gesamtbewertungen der erinnerten Beiträge (Anova,  $F(2, 70) = 1.99, p = .144$ ; BS: MW = 22.1, SD = 9.3; BSC: MW = 25.1, SD = 12.3; TB: MW = 24.2, SD = 11.7).

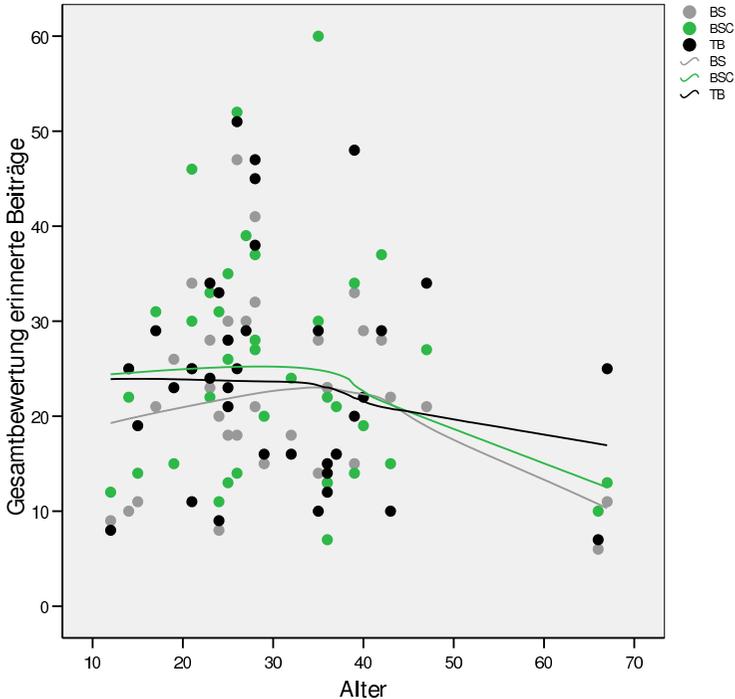


Abbildung 8.15: Gesamtbewertung erinnerten Beiträge getrennt nach Alter und Präsentationsbedingungen.

Weiterhin existierten keine signifikanten Interaktionen zwischen den Präsentationsbedingungen und den vier Gerätekombinationen (Anova, Präsentation  $\times$  Gerätekombination,  $F(6, 64) = .64, p = .635$ ). Es lagen ebenfalls keine signifikanten Interaktionen zwischen der Präsentationsbedingung und dem Geschlecht der Vpn vor (Anova, Präsentation  $\times$  Geschlecht,  $F(2, 68) = 2.23, p = .116$ ). Die Unterschiede zwischen Männern und Frauen in der BSC-Bedingung waren jedoch marginal signifikant (t-Test,  $t(34) = -1.98, p = .056$ ). Abbildung 8.14 stellt die Unterschiede grafisch dar. Die marginale Signifikanz lag in signifikanten Unterschieden zwischen Männern und Frauen

### 8.3. ERGEBNISSE

in der BSC-Bedingung der PDA-TV-Bilderrahmen-Gruppe begründet (t-Test,  $t(7) = -3.49$ ,  $p = .010$ ). Die Interaktionen zwischen dem Alter der Vpn und den Präsentationsbedingungen waren signifikant bzgl. der Gesamtbewertung der Nachrichtenbeiträge (Anova, Präsentation  $\times$  Alter,  $F(46, 24) = 2.26$   $p = .017$ ).

Die Unterschiede werden in Abbildung 8.15 in einem Streudiagramm dargestellt, das die Tendenzen von BS, BSC und TB mittels dreier Interpolationsgeraden (Loess-Interpolation) verdeutlicht. Die Signifikanz scheint daher zu rühren, dass sich BSC bei den unter 25-jährigen als beste Bedingung und BS als schlechteste Bedingung ergab, was sich mit steigendem Alter jedoch nivel-lierte.

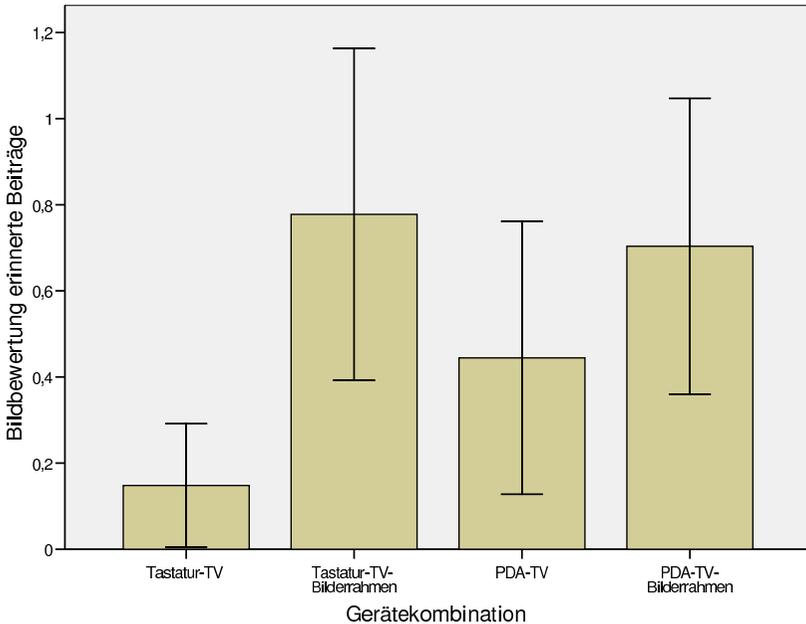


Abbildung 8.16: Bildbewertung erinnerner Beiträge getrennt nach Gerätekombinationen.

#### Bildbewertungen

Insgesamt wurden im Experiment für die drei Präsentationsbedingungen 39 Bildbewertungen mit einem oder mehr Punkten pro Präsentationsbedingung

## KAPITEL 8. LABORSTUDIE

vergeben. Zwölf der 36 Vpn hatten im Versuch eine Bildbewertung von null erhalten. Die maximale Bildbewertung pro Vpn betrug sieben. Im Mittel erreichte jede Vpn eine Bildbewertung von 1.6 Punkten von maximal 108 möglichen Punkten ( $SD = 1.9$ ). Es ergaben sich keine signifikanten Unterschiede bzgl. der Bildbewertungen zwischen den Gerätekombinationen (Anova,  $F(3, 32) = 2.05, p = .127$ ).

Die Unterschiede zwischen Tastatur-TV und den Gerätekombinationen Tastatur-TV-Bilderrahmen sowie PDA-TV-Bilderrahmen waren jedoch signifikant (Anova; Tastatur-TV–Tastatur-TV-Bilderrahmen:  $F(1, 16) = 5.81, p = .028$ ; Tastatur-TV–PDA-TV-Bilderrahmen:  $F(1, 16) = 6.04, p = .026$ ). Abbildung 8.16 stellt die Unterschiede grafisch dar. Die Unterschiede bzgl. der Bildbewertungen zwischen den Präsentationsbedingungen waren nicht signifikant (Anova, Greenhouse-Geisser,  $F(1.56, 54.66) = 2.90, p = .076$ ). Der Unterschied zwischen BSC und TB war jedoch signifikant (t-Test,  $t(35) = -1.54, p = .134$ ). Abbildung 8.17 zeigt die Unterschiede.

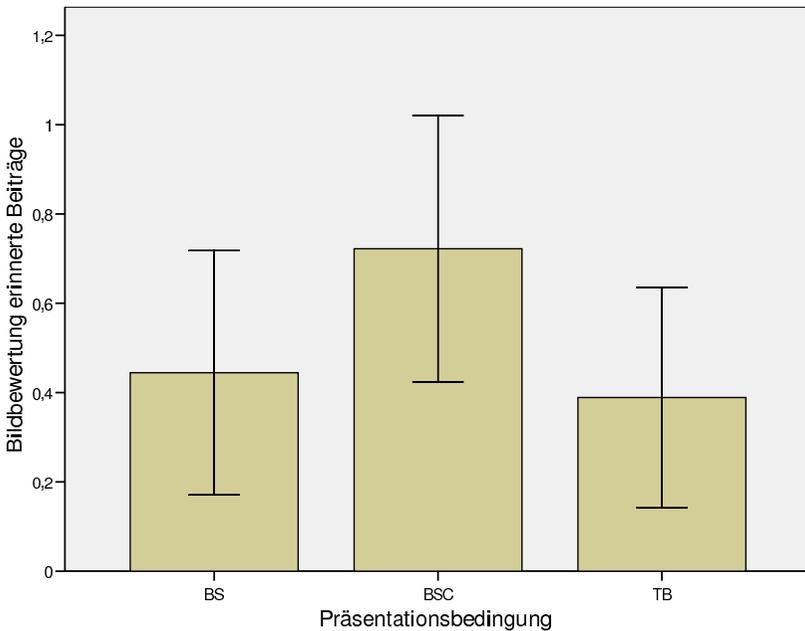


Abbildung 8.17: Bildbewertung erinnerner Beiträge getrennt nach Präsentationsbedingungen.

### 8.3. ERGEBNISSE

#### Lebhaftigkeit

Zu jedem Nachrichtenbeitrag, an den sich die Vpn erinnert haben, wurden sie gebeten anzugeben, wie lebhaft ihre Erinnerung an den Beitrag ist. Dies wurde auf einer Skala von eins („sehr lückenhaft“) bis fünf („sehr detailliert“) erfasst. Nicht erinnerte Beiträge wurden mit null bewertet. Im Mittel ergab sich für einen Beitrag ein Wert von 2.5 (SD = 1.7). Das Alter der Vpn hatte keinen signifikanten Einfluss auf die Bewertungen der Lebhaftigkeit (Anova,  $F(23, 12) = .36, p = .983$ ). Das Geschlecht der Vpn hatte ebenfalls keinen signifikanten Einfluss auf die Bewertungen (Anova,  $F(1, 34) = .99, p = .328$ ). Weiterhin hatte die Gerätekombination keinen signifikanten Einfluss auf die Bewertungen der Lebhaftigkeit der Erinnerungen an die Beiträge (Anova,  $F(3, 32) = .55, p = .654$ ).

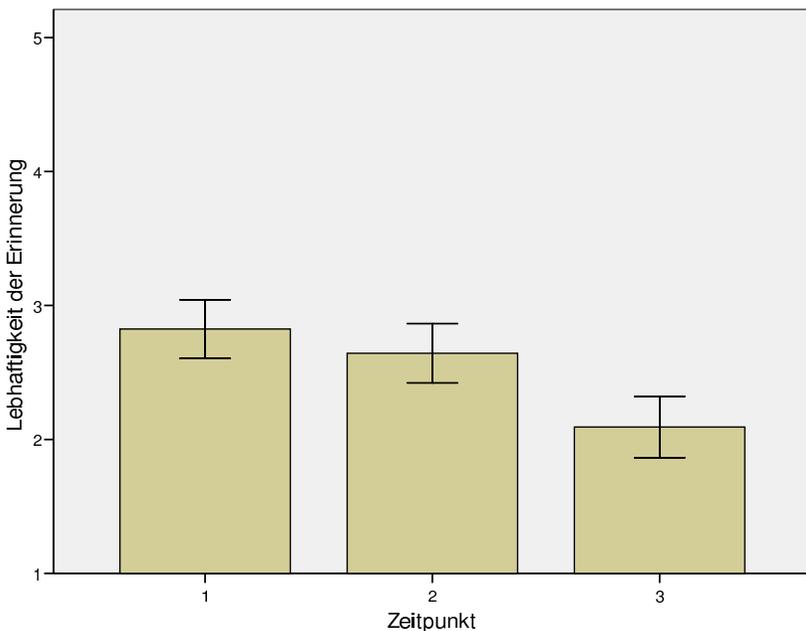


Abbildung 8.18: Lebhaftigkeit der Erinnerungen an Beiträge getrennt nach Zeitpunkten der Präsentation.

Die Lebhaftigkeit der Erinnerungen variierte signifikant zwischen den drei Zeitpunkten einer Präsentation (Anova,  $F(2, 70) = 9.04, p = .001$ ). Die

## KAPITEL 8. LABORSTUDIE

Unterschiede zwischen Zeitpunkt drei und den Zeitpunkten eins und zwei waren signifikant, d. h. die Lebhaftigkeit der Erinnerungen nahm über alle drei Zeitpunkte hinweg signifikant ab (Anova; Zeitpunkt 1&2:  $F(1, 35) = .89$ ,  $p = .352$ ; Zeitpunkt 1&3:  $F(1, 35) = 14.81$ ,  $p = .001$ ; Zeitpunkt 2&3:  $F(1, 35) = 12.87$ ,  $p = .001$ ). Abbildung 8.18 stellt die Unterschiede dar.

Weiterhin existierten signifikante Interaktionen zwischen den drei Präsentationszeitpunkten und den sechs Zeitpunkten, an denen die Beiträge innerhalb einer Präsentationsbedingung präsentiert werden (Anova, Zeitpunkt-Präsentation  $\times$  Zeitpunkt-Beitrag,  $F(10, 350) = 2.35$ ,  $p = .011$ ).

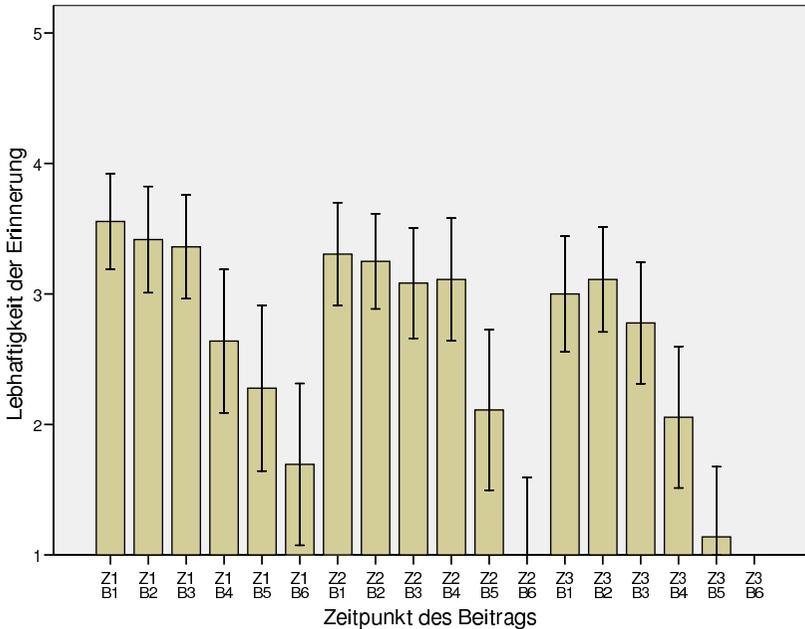


Abbildung 8.19: Lebhaftigkeit der Erinnerungen an Beiträge getrennt nach Zeitpunkten der Präsentation eines Beitrags.

Abbildung 8.19 zeigt die Unterschiede in den drei Präsentationsbedingungen mit je sechs Beiträgen. Bei den Achsenbeschriftungen werden die 18 Zeitpunkte durch Paare  $(Zx, By)$  bezeichnet. Dabei bezeichnet  $x$ ,  $1 \leq x \leq 3$  den Zeitpunkt einer Präsentationsbedingung und  $y$ ,  $0 \leq y \leq 6$  den Zeitpunkt eines Beitrags innerhalb der Präsentationsbedingung  $x$ . Die Abbildung zeigt,

### 8.3. ERGEBNISSE

dass die Signifikanz in einem Abfall der Bewertung der Lebhaftigkeit in jeder Präsentationsbedingung begründet liegt.

Der zeitliche Verlauf wurde darüber hinaus auf signifikante Interaktionen mit den Faktoren Alter, Geschlecht und Gerätekombination untersucht. Es existierten keine signifikanten Interaktionen zwischen den drei Zeitpunkten der Präsentation und dem Alter der Vpn (Anova, Zeitpunkt Präsentation  $\times$  Alter,  $F(2, 46) = 1.50, p = .145$ ). Es existierten ebenfalls keine signifikanten Interaktionen mit dem Geschlecht der Vp (Anova, Zeitpunkt Präsentation  $\times$  Geschlecht,  $F(2, 68) = .67, p = .514$ ). Darüber hinaus waren die Interaktionen zwischen den Zeitpunkten der Präsentation und der Gerätekombinationen nicht signifikant (Anova, Zeitpunkt Präsentation  $\times$  Gerätekombination,  $F(6, 64) = .90, p = .501$ ).

Neben dem zeitlichen Verlauf wurde überprüft, ob die Lebhaftigkeit der Erinnerungen durch die drei Präsentationsbedingungen beeinflusst wird. Es wurden keine signifikanten Unterschiede zwischen den Präsentationsbedingungen festgestellt (Anova,  $F(2, 70) = 2.18, p = .120$ ; BS: MW = 2.6, SD = 1.7; BSC: MW = 2.7, SD = 1.7; TB: MW = 2.3, SD = 1.7). Weiterhin lagen keine signifikanten Interaktionen zwischen der Präsentationsbedingung und dem Alter der Vpn vor (Anova, Präsentation  $\times$  Alter,  $F(46, 24) = .96, p = .558$ ). Die Interaktionen zwischen den Präsentationsbedingungen und dem Geschlecht der Vpn waren ebenfalls nicht signifikant (Anova, Präsentation  $\times$  Geschlecht,  $F(2, 68) = .71, p = .497$ ).

#### 8.3.4 Betrachtungszeit

Zu jeder Präsentation eines Segments eines Nachrichtenbeitrags wurde die Zeit von der Darstellung des Segments bis zum Drücken des „weiter“-Knopfes durch die Vp gemessen. Aufgrund der unterschiedlichen Anzahlen von Ausgabeschritten in den vier Gerätekombinationen (vgl. Tabelle 8.4) wurden die Zeiten zunächst für jeden Nachrichtenbeitrag über alle Ausgabeschritte summiert. Damit ergeben sich insgesamt 18 Messwerte pro Vp.

#### Aufbereitung der Daten

Bei der Analyse der Daten wurden Ausreißer festgestellt. Dabei wird die Definition von Ausreißern aus Abschnitt 8.2.6 verwendet. Abbildung 8.20 stellt die Ausreißer in Sternform dar. Auf die Darstellung des äußersten Ausreißers wurde aus Darstellungsgründen verzichtet (582.5 Sekunden).

Da die Ausreißer die erwarteten Größenordnungen für die Zeiten pro Nachrichtenbeitrag deutlich übersteigen, wurden die fünf betroffenen Extremwerte vernachlässigt. Durch diese Vorgehensweise wurden die fünf betroffenen Einträge (bei insgesamt 684 Zeitmessungen) als fehlende Daten behandelt. Mit

## KAPITEL 8. LABORSTUDIE

Extremwerten betragen die Zeiten pro Nachrichtenbeitrag im Mittel 49.6 Sekunden (SD = 25.8). Ohne Extremwerte betragen sie im Mittel 48.3 Sekunden (SD = 14.1).

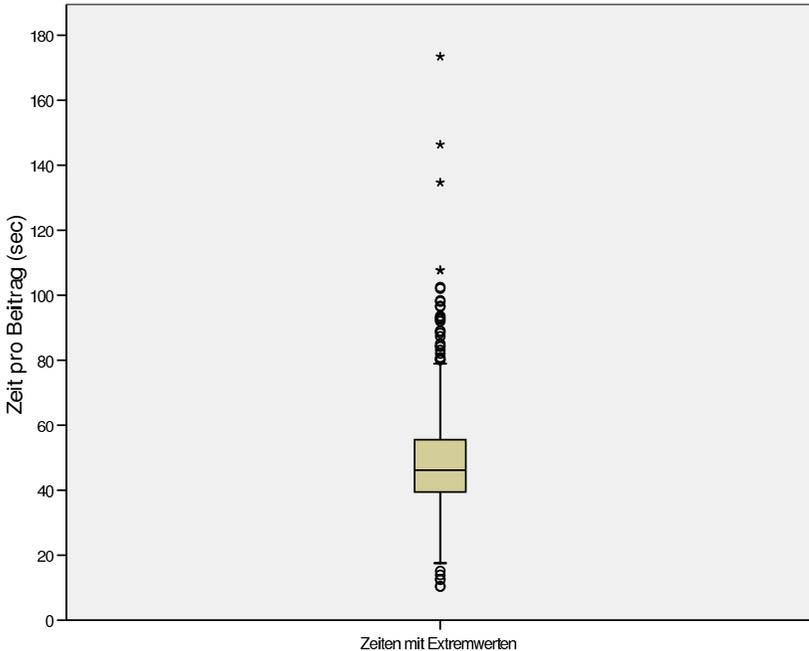


Abbildung 8.20: Zeiten pro Nachrichtenbeitrag mit Extremwerten.

### Gesamtzeit pro Versuchsperson

Es existierten keine signifikanten Unterschiede zwischen den Zeiten pro Nachrichtenbeitrag bzgl. des Alters der Vpn (Anova,  $F(22, 10) = .88, p = .618$ ). Es existierten ebenfalls keine signifikanten Unterschiede zwischen den Zeiten pro Nachrichtenbeitrag bzgl. des Geschlechts der Vpn (Anova,  $F(1, 31) = 1.94, p = .173$ ).

Die Gerätekombination hatte signifikanten Einfluss auf die Zeiten der Nachrichtenbeiträge (Anova,  $F(3, 29) = 4.86, p = .007$ ). Die Unterschiede werden in Abbildung 8.21 grafisch dargestellt. Die Unterschiede zwischen den Gruppen Tastatur-TV und Tastatur-TV-Bilderrahmen, PDA-TV und PDA-TV-Bilderrahmen sowie Tastatur-TV-Bilderrahmen und PDA-TV waren signifi-

### 8.3. ERGEBNISSE

kant (Anova; Tastatur-TV–Tastatur-TV-Bilderrahmen:  $F(1, 14) = 6.29, p = .025$ ; PDA-TV–PDA-TV-Bilderrahmen:  $F(1, 15) = 5.34, p = .036$ ; Tastatur-TV-Bilderrahmen–PDA-TV:  $F(1, 15) = 18.23, p < .001$ ). Die Unterschiede zwischen den Gruppen Tastatur-TV und PDA-TV-Bilderrahmen waren nicht signifikant (Anova,  $F(1, 14) = 1.17, p = .298$ ).

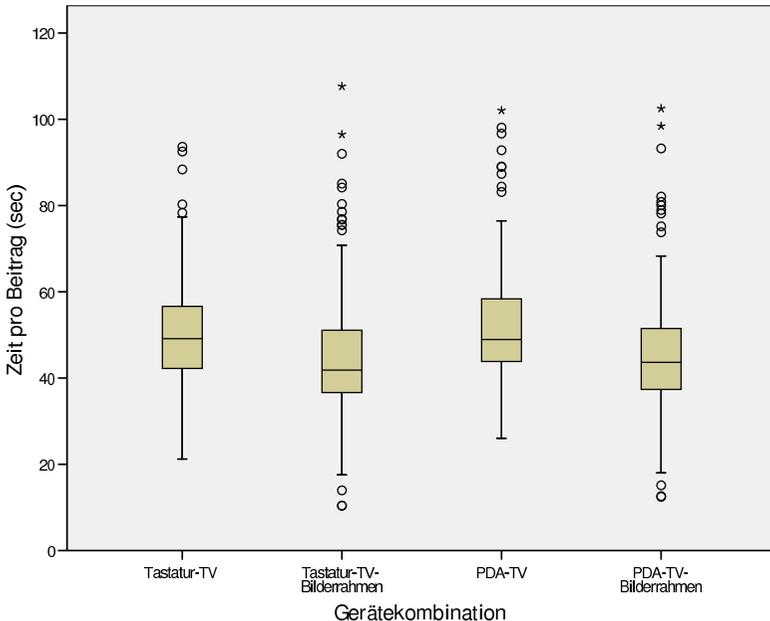


Abbildung 8.21: Zeiten pro Nachrichtenbeitrag getrennt nach Gerätekombinationen.

Weiterhin wurden die Interaktionen zwischen den Faktoren Geschlecht und Gerätekombination bzgl. der Zeiten pro Beitrag untersucht. Diese waren jedoch nicht signifikant (Anova, Geschlecht×Gerätekombination,  $F(3, 25) = 3.48, p = .285$ ).

#### Zeitlicher Verlauf

Neben den Gesamtzeiten pro Vpn wurde ebenfalls der zeitliche Verlauf der Interaktionszeiten betrachtet. Es wurden signifikante Unterschiede zwischen allen 18 Zeitpunkten, an denen Nachrichtenbeiträge im Experiment präsentiert wurden, gefunden (Anova, Greenhouse-Geisser,  $F(5.9, 188.8) = 7.81, p = .001$ ). Diese werden in Abbildung 8.22 dargestellt.

## KAPITEL 8. LABORSTUDIE

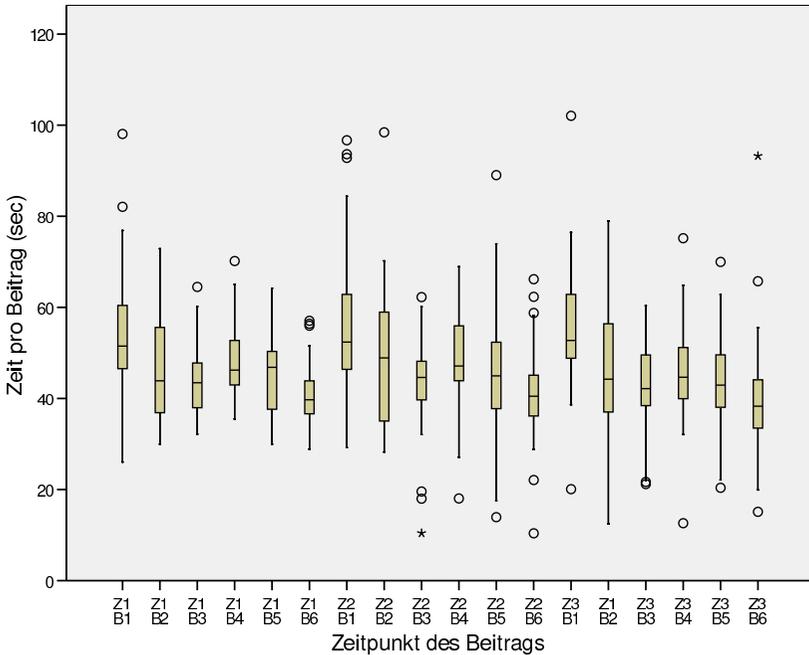


Abbildung 8.22: Zeiten pro Nachrichtenbeitrag getrennt nach Zeitpunkten der Präsentation.

Die Abbildung lässt vermuten, dass die Teilnehmer innerhalb jeder Präsentationsbedingung mit jedem Nachrichtenbeitrag schneller werden. Die Unterschiede zwischen den sechs Zeitpunkten der Beiträge waren signifikant (Anova, Greenhouse-Geisser,  $F(3.3, 105.2) = 65.6, p = .001$ ). Dies bedeutet, dass die Vpn in jeder Bedingung vom ersten Nachrichtenbeitrag bis zum sechsten Nachrichtenbeitrag signifikant schneller wurden. Die Vermutung wurde somit bestätigt.

Es existierten jedoch keine signifikanten Unterschiede zwischen den drei Zeitpunkten der Präsentationsbedingungen (Anova,  $F(2, 64) = .38, p = .688$ ), d. h. zwischen den drei Präsentationsbedingungen traten keine signifikanten Beschleunigungen oder Verzögerungen auf.

Weiterhin existierten zwischen beiden Faktoren keine signifikanten Interaktionen (Anova, Greenhouse-Geisser, Zeitpunkt-Präsentation  $\times$  Zeitpunkt-Beitrag,  $F(4.9, 156.5) = .300, p = .981$ ).

### 8.3. ERGEBNISSE

#### Präsentationsbedingung

Neben dem zeitlichen Verlauf des Versuchs wurden ebenfalls die drei Präsentationsbedingungen BS, BSC und TB untersucht, die bei jeder Vp in unterschiedlicher Reihenfolge präsentiert wurden (vgl. Tabelle 8.5). Hierbei zeigten sich signifikante Unterschiede zwischen den drei Bedingungen (Anova, Greenhouse-Geisser,  $F(1.6, 4.9) = 3.5, p = .047$ ). Abbildung 8.23 stellt die Unterschiede dar. Trotz der optisch nur geringen Abweichung der Mediane ergaben sich die Signifikanzen durch die unterschiedlichen Streuungen der Messwerte.

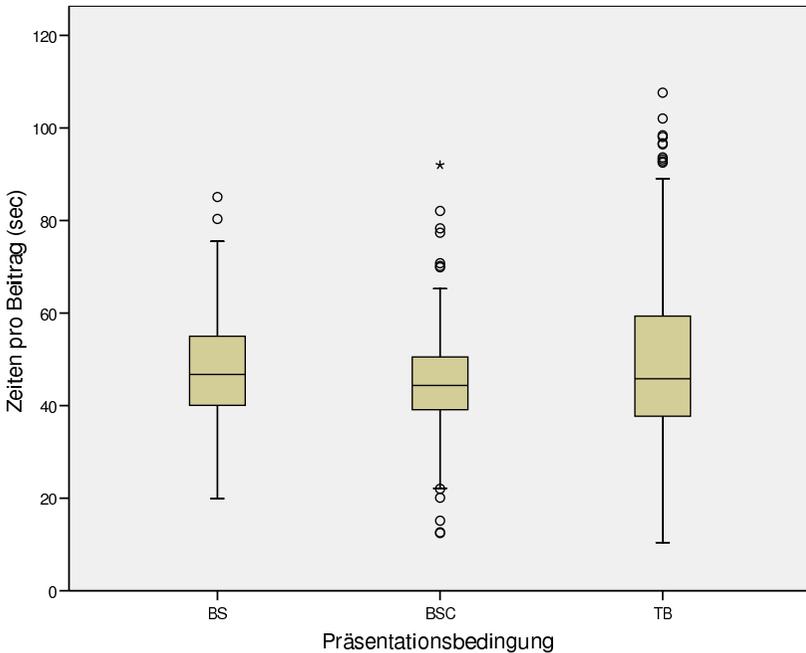


Abbildung 8.23: Zeiten pro Nachrichtenbeitrag getrennt nach Präsentationsbedingungen.

Die Unterschiede zwischen BSC und BS sowie BSC und TB waren signifikant (Anova, Greenhouse-Geisser; BS-BSC:  $F(1, 31) = 5.42, p = .027$ ; BS-TB:  $F(1, 29) = .82, p = .374$ ; BSC-TB:  $F(1, 30) = 4.9, p = .035$ ). Weiterhin existierten signifikante Interaktionen zwischen den drei Präsentationsbe-

## KAPITEL 8. LABORSTUDIE

dingungen und den Gerätekombinationen (Anova, Präsentation×Gerätekombination, Greenhouse-Geisser;  $F(1.6, 4.8) = 3.5, p = .001$ ). Die Unterschiede werden in Abbildung 8.24 grafisch dargestellt. Die Signifikanz ergab sich daher, dass in der PDA-TV-Gruppe die Unterschiede zwischen den Präsentationsbedingungen signifikant waren (Anova,  $F(2, 16) = 27.2, p = .001$ ). In dieser Gruppe waren ebenfalls sämtliche paarweisen Unterschiede zwischen den Präsentationsbedingungen signifikant, d. h. in der BSC-Bedingung waren die Vpn am schnellsten und in der TB-Bedingung am langsamsten (Anova; BS-BSC:  $F(1, 8) = 9.4, p = .015$ ; BS-TB:  $F(1, 8) = 19.2, p = .002$ ; BSC-TB:  $F(1, 8) = 41.7, p = .001$ ).

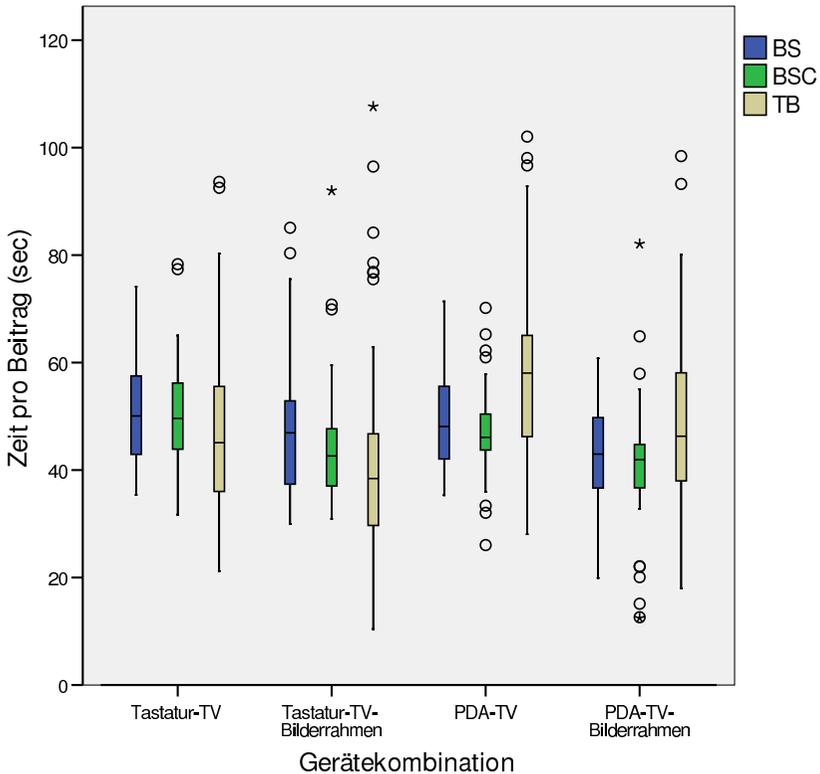


Abbildung 8.24: Zeiten pro Nachrichtenbeitrag getrennt nach Gerätekombinationen und Präsentationsbedingungen.

### 8.3. ERGEBNISSE

#### 8.3.5 Erinnerungseffizienz

Da die Erinnerungsleistung davon abhängt, wie lange eine Präsentation betrachtet wurde, wurden neben den Bewertungen der Erinnerungen auch die Bewertungen pro Minute betrachtet. Dazu wurden zunächst die Zeiten für jede Präsentationsbedingung summiert und in Minuten umgerechnet. Anschließend wurden die Bewertungen der Erinnerungen für jede Präsentationsbedingung durch die so erhaltenen Zeiten dividiert.

Die maximale Erinnerungseffizienz über den drei Präsentationsbedingungen betrug 12.8 Bewertungspunkte pro Minute und Präsentationsbedingung. Die minimale Erinnerungseffizienz betrug 1.7 Bewertungspunkte pro Minute und Präsentationsbedingung. Im Mittel wurden in jeder Präsentationsbedingung 5.0 Bewertungspunkte pro Minute erreicht ( $SD = 2.3$ ).

Gesamteffizienz pro Versuchsperson

Die Unterschiede bzgl. der Erinnerungseffizienz zwischen den vier Gerätekombinationen waren nicht signifikant (Anova,  $F(3, 29) = .21, p = .891$ ; Tastatur-TV:  $MW = 4.8, SD = 2.4$ ; Tastatur-TV-Bilderrahmen:  $MW = 5.0, SD = 2.7$ ; PDA-TV:  $MW = 4.8, SD = 1.5$ ; PDA-TV-Bilderrahmen:  $MW = 5.4, SD = 2.8$ ). Das Alter hatte ebenfalls keinen signifikanten Einfluss auf die Erinnerungseffizienz (Anova,  $F(22, 10) = .49, p = .923$ ).

Weiterhin waren die Unterschiede zwischen Männern und Frauen nicht signifikant (Anova,  $F(1, 31) = .33, p = .572$ ; Frauen:  $MW = 5.2, SD = 2.6$ ; Männer:  $MW = 4.9, SD = 2.1$ ). Es lagen ebenfalls keine signifikanten Interaktionen zwischen dem Geschlecht der Vpn und der Gerätekombination vor (Anova,  $\text{Geschlecht} \times \text{Gerätekombination}, F(3, 25) = .81, p = .500$ ).

Zeitlicher Verlauf

Zwischen den drei Zeitpunkten einer Präsentation ergaben sich signifikante Unterschiede (Anova,  $F(2, 64) = 4.2, p = .020$ ). Dies lag in den signifikanten Unterschieden zwischen den Zeitpunkten eins und zwei begründet (t-Test, Zeitpunkt 1&2:  $t(33) = -3.23, p = .003$ ; Zeitpunkt 1&3:  $t(34) = -1.93, p = .062$ ; Zeitpunkt 2&3:  $t(32) = 1.05, p = .298$ ). Die Unterschiede werden in Abbildung 8.25 dargestellt.

Es existierten keine signifikanten Interaktionen zwischen dem Zeitpunkt der Präsentation und dem Alter (Anova,  $\text{Zeitpunkt} \times \text{Alter}, F(44, 20) = .43, p = .990$ ). Die Interaktionen zwischen den Zeitpunkten und dem Alter der Vpn erwiesen sich ebenfalls als nicht signifikant (Anova,  $\text{Zeitpunkt} \times \text{Geschlecht}, F(2, 62) = .46, p = .632$ ). Es existierten weiterhin keine signifikanten Interaktionen zwischen dem Zeitpunkt der Präsentation und der Gerätekombination (Anova,  $\text{Zeitpunkt} \times \text{Gerätekombination}, F(6, 58) = 1.24, p = .299$ ).

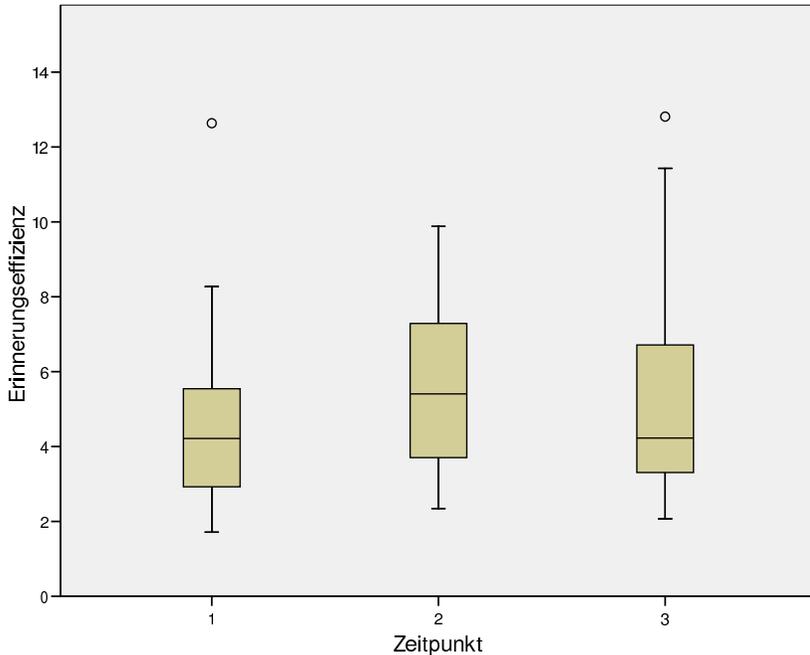


Abbildung 8.25: Erinnerungseffizienz getrennt nach Zeitpunkten der Präsentation.

### Präsentationsbedingung

Neben dem zeitlichen Verlauf der Erinnerungseffizienz wurde ebenfalls die Präsentationsbedingung (BS, BSC oder TB) als Faktor untersucht. Die Unterschiede zwischen den drei Präsentationsbedingungen bzgl. der Erinnerungseffizienz waren nicht signifikant (Anova,  $F(2, 64) = 2.89, p < .063$ ).

Die Unterschiede zwischen den Bedingungen BS und BSC waren jedoch signifikant (t-Test; BS-BSC:  $t(34) = -2.56, p = .015$ ; BS-TB:  $t(32) = -.85, p = .403$ ; BSC-TB:  $t(33) = 1.51, p = .142$ ). Abbildung 8.26 zeigt die Unterschiede. Die Signifikanz zwischen BS und BSC ergibt sich anscheinend durch die größere Streuung von BSC in Richtung höherer Effizienzwerte.

Es existierten keine signifikanten Interaktionen zwischen der Präsentationsbedingung und dem Alter der Vp (Anova, Präsentation  $\times$  Alter,  $F(44, 20) = 1.51, p = .161$ ). Es existierten ebenfalls keine signifikanten Interaktionen zwi-

### 8.3. ERGEBNISSE

schen der Präsentationsbedingung und der Gerätekombination (Anova, Präsentation  $\times$  Gerätekombination,  $F(2, 6) = .81, p = .565$ ).

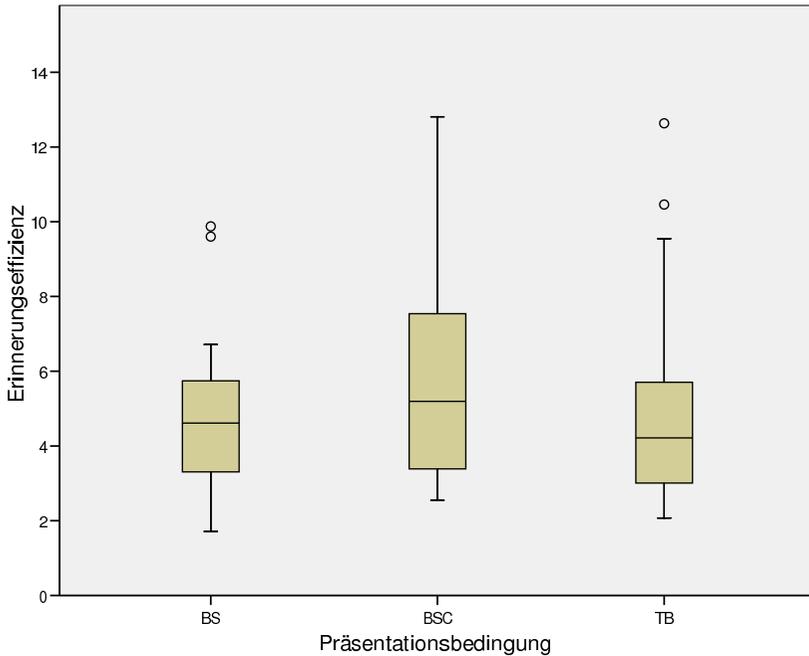


Abbildung 8.26: Erinnerungseffizienz getrennt nach Präsentationsbedingungen.

Es existierten jedoch signifikante Interaktionen zwischen den Faktoren Präsentationsbedingung und Geschlecht bzgl. der Erinnerungseffizienz (Anova, Präsentation  $\times$  Geschlecht,  $F(2, 62) = 4.61, p = .014$ ).

Dies lag darin begründet, dass die Unterschiede zwischen BSC und BS sowie zwischen BSC und TB bei den Frauen signifikant waren (t-Test; BS-BSC:  $t(17) = -2.22, p = .040$ ; BS-TB:  $t(15) = 1.00, p = .333$ ; BSC-TB:  $t(15) = 2.91, p = .011$ ).

Die Unterschiede zwischen BS, BSC und TB waren bei Männern nicht signifikant (t-Test; BS-BSC:  $t(16) = -1.30, p = .211$ ; BS-TB:  $t(16) = -1.85, p = .083$ ; BSC-TB:  $t(17) = -.746, p = .466$ ). Abbildung 8.27 stellt die Unterschiede grafisch dar.

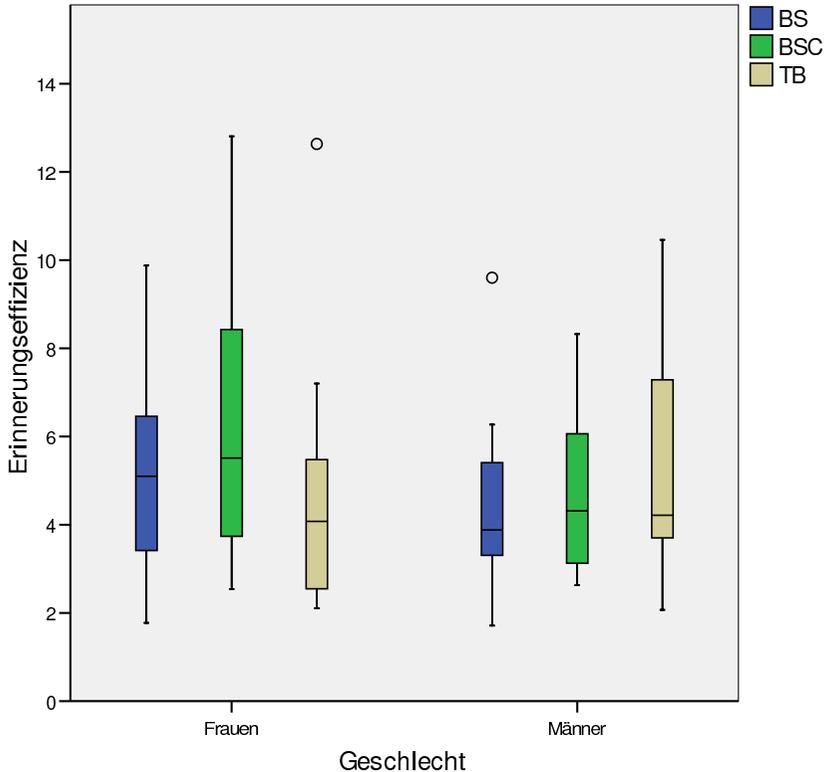


Abbildung 8.27: Erinnerungseffizienz getrennt nach Präsentationsbedingungen und Geschlecht.

### 8.3.6 Attraktivität

Als Teil der Zwischenfragebögen wurden die Benutzer nach der Attraktivität der zuletzt gesehenen Präsentationsbedingung gefragt. Dabei wurden die Benutzer sowohl nach der Attraktivität der einzelnen Ausgabemodalitäten (z. B. Bild und Sprache) als auch nach der Attraktivität des gesamten Präsentationsformats (z. B. BS) gefragt.

#### Ausgabemodalitäten

Nach jeder Präsentationsbedingung wurden die Vpn in den Zwischenfragebögen gebeten, die Ausgabemodalitäten der letzten Präsentationsbedingung zu bewerten. Die Ausgabemodalitäten setzen sich dabei aus Text-, Bild-, Sprach-

### 8.3. ERGEBNISSE

und Charakterausgaben zusammen. Die Vpn wurden nur zu den Ausgabemodalitäten befragt, die in der Präsentationsbedingung verwendet wurden.

In der TB-Bedingung wurden Bilder signifikant besser als Text beurteilt (Wilcoxon,  $p = .037$ ; Text: MW = 3.1, SD = 1.2; Bild: MW = 3.6, SD = .8). In der BS-Bedingung wurden Bilder ebenfalls signifikant besser als Sprache beurteilt (Wilcoxon,  $p = .001$ ; Bild: MW = 3.4, SD = .8; Sprache: MW = 1.9, SD = .8).

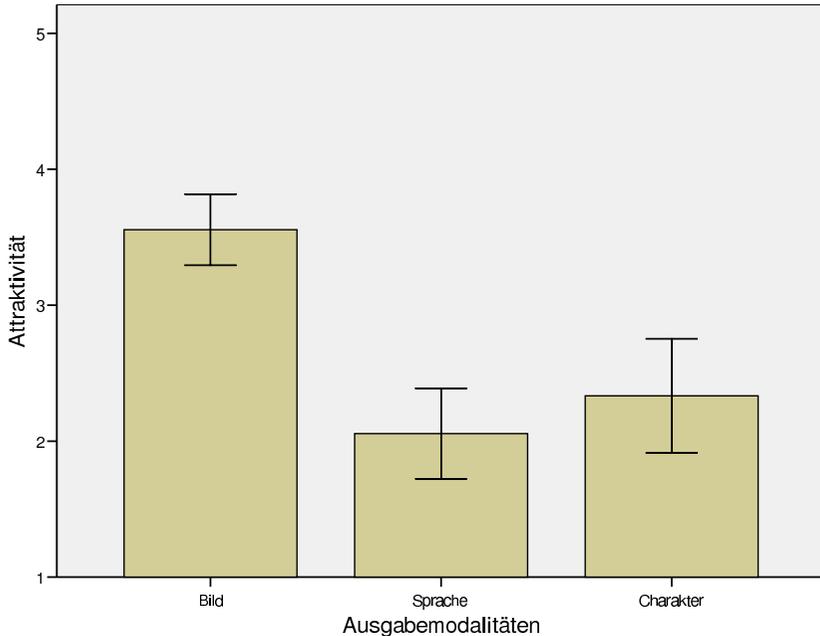


Abbildung 8.28: Attraktivität der Ausgabemodalitäten in der BSC-Bedingung.

Abbildung 8.28 zeigt die Bewertung von Bildern, Sprache und animiertem Charakter in der BSC-Bedingung. Die Unterschiede zwischen den Bewertungen waren signifikant (Friedman,  $p = .001$ ). Dabei wurden die Bilder signifikant besser als die Sprachausgabe oder der animierte Charakter bewertet (Wilcoxon; Bild-Sprache:  $p = .001$ ; Bild-Charakter:  $p = .001$ ; Sprache-Charakter:  $p = .171$ ).

Weiterhin wurde untersucht, ob die Gerätegruppe Einfluss auf die Bewertung der vier Ausgabemodalitäten hatte. Es zeigte sich, dass die Gerätegruppe kei-

## KAPITEL 8. LABORSTUDIE

nen signifikanten Einfluss auf die Bewertungen des Textes, der Sprache oder des Charakters hatte (Anova; Text:  $F(3, 35) = .06, p = .979$ ; Sprache:  $F(3, 71) = 2.0, p = .121$ ; Charakter:  $F(3, 35) = .5, p = .714$ ). Der Einfluss der Gerätegruppe auf die Bewertung der Bilder war jedoch signifikant (Anova,  $F(3, 107) = 5.0, p = .003$ ; Tastatur-TV: MW = 3.8, SD = .6; Tastatur-TV-Bilderrahmen: MW = 3.6, SD = .7; PDA-TV: MW = 3.1, SD = .9; PDA-TV-Bilderrahmen: MW = 3.6, SD = .7).

Die Signifikanz der Anova lag darin begründet, dass in der Gruppe PDA-TV die Bilder signifikant schlechter als in den anderen Gruppen bewertet wurden (t-Test; Tastatur-TV–PDA-TV:  $t(52) = 3.59, p = .001$ ; Tastatur-TV-Bilderrahmen–PDA-TV:  $t(52) = 2.34, p = .023$ ; PDA-TV-Bilderrahmen–PDA-TV:  $t(52) = 2.45, p = .018$ ).

### Präsentationsbedingungen

Neben den Ausgabemodalitäten wurden die Vpn ebenfalls gebeten, in den Zwischenfragebögen die Attraktivität der gesamten Darstellung zu bewerten.

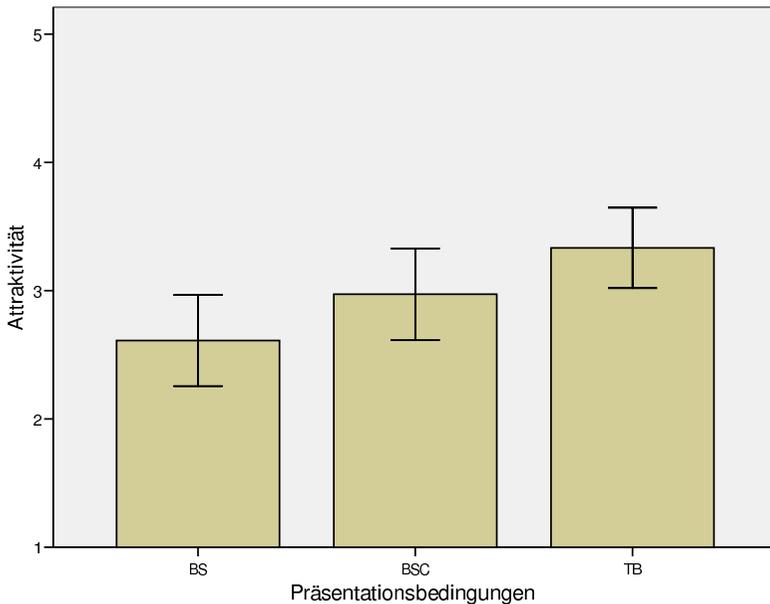


Abbildung 8.29: Attraktivität der Darstellung getrennt nach Präsentationsbedingungen.

### 8.3. ERGEBNISSE

Dabei ergaben sich signifikante Unterschiede zwischen den Bewertungen der drei Präsentationsbedingungen (Anova,  $F(2, 70) = 5.48, p = .006$ ). Die Unterschiede werden in Abbildung 8.29 grafisch dargestellt. Die Unterschiede zwischen BS und TB waren signifikant (t-Test,  $t(70) = -3.10, p = .003$ ).

Weiterhin wurde untersucht, ob die Attraktivität der Darstellung in den vier Gerätegruppen unterschiedlich beurteilt wurde. Es ergaben sich keine signifikanten Unterschiede (Anova,  $F(3, 107) = 1.52, p = .213$ ).

Es wurde ebenfalls untersucht, ob der Zeitpunkt der Präsentation einen Einfluss auf die Bewertung der Darstellung hatte. Wie aus Tabelle 8.5 ersichtlich wird, konnte jede Präsentationsbedingung an erster, zweiter oder dritter Stelle auftreten. Der Zeitpunkt der Präsentation hatte keinen signifikanten Einfluss auf die Bewertung der Attraktivität (Anova,  $F(2, 70) = 0.54, p = .585$ ).

#### 8.3.7 Benötigte Konzentration

In den drei Zwischenfragebögen wurden die Vpn gebeten zu bewerten, in wie weit sie sich konzentrieren mussten, um der Präsentation folgen zu können.

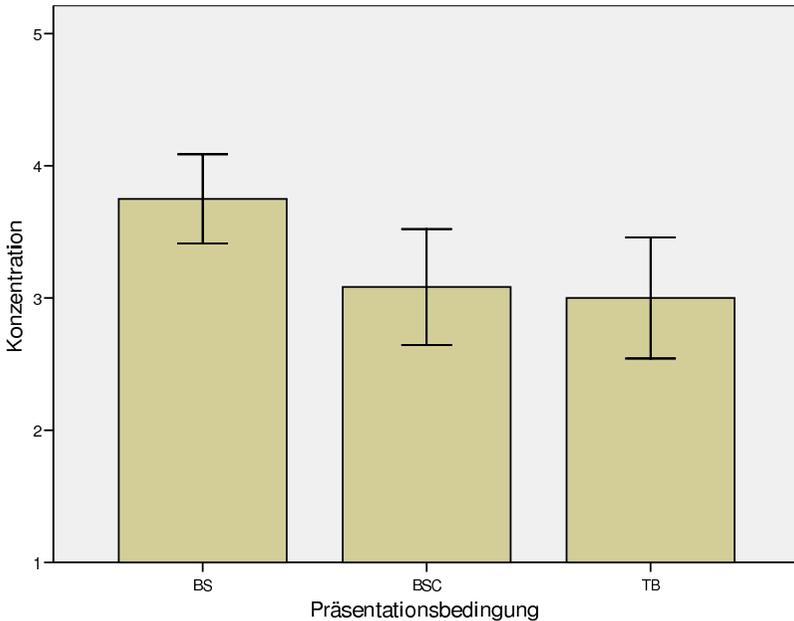


Abbildung 8.30: Benötigte Konzentration getrennt nach Präsentationsbedingungen.

## KAPITEL 8. LABORSTUDIE

Dies sollte Aufschluss über die empfundene kognitive Last der Vpn geben. Dabei zeigten sich signifikante Unterschiede zwischen Frauen und Männern. Frauen gaben eine signifikant höhere Konzentration als Männer an (Anova,  $F(1, 34) = 541.53$ ,  $p = .001$ ; Frauen: MW = 3.6, SD = 1.3; Männer: MW = 3.0, SD = 1.5). Weiterhin zeigten sich signifikante Unterschiede zwischen den drei Präsentationsbedingungen (Anova,  $F(2, 70) = 5.67$ ,  $p = .005$ ). Abbildung 8.30 stellt die Unterschiede grafisch dar. Dabei wurde die empfundene kognitive Last in der BS-Bedingung signifikant höher als in der BSC- oder der TB-Bedingung eingestuft (t-Test; BS-BSC:  $t(70) = 3.44$ ,  $p = .017$ ; BS-TB:  $t(64.4) = 2.68$ ,  $p = .009$ ). Zwischen den Gerätegruppen existierten keine signifikanten Unterschiede bzgl. der benötigten Konzentration der Vpn (Anova,  $F(3, 32) = 0.171$ ,  $p = .915$ ).

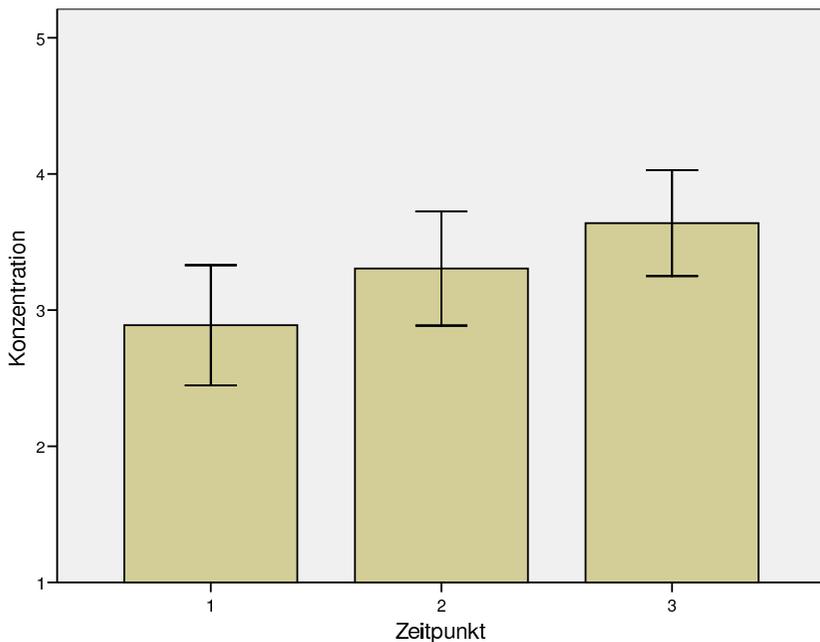


Abbildung 8.31: Benötigte Konzentration getrennt nach Zeitpunkten der Präsentation.

Neben den Präsentationsbedingungen und den Gerätegruppen wurde untersucht, ob sich die Konzentration der Teilnehmer über die drei Zeitpunkte der

### 8.3. ERGEBNISSE

Präsentationsbedingungen änderte. Es zeigten sich signifikante Unterschiede zwischen den drei Zeitpunkten (Anova,  $F(2, 70) = 4.61, p = .013$ ).

Abbildung 8.31 zeigt, dass die benötigte Konzentration über die drei Zeitpunkte des Experiments hinweg anstieg. Dabei waren die Unterschiede zwischen den Zeitpunkten eins und drei signifikant (t-Test; Zeitpunkte 1&2:  $t(35) = -1.93, p = .062$ ; Zeitpunkte 1&3:  $t(35) = -2.93, p = .006$ ; Zeitpunkte 2&3:  $t(35) = -1.25, p = .221$ ).

#### 8.3.8 Häufigkeit der Betrachtung

In den Zwischenfragebögen wurden die Vpn gefragt, wie häufig sie die einzelnen Geräte (PDA, Fernseher oder Bilderrahmen) während der letzten Präsentationsbedingung betrachtet haben.

Dabei wurden die Vpn nur nach den Geräten gefragt, aus denen sich ihre Gerätekombination zusammensetzte. In der Tastatur-TV-Gruppe wurde diese Frage nicht gestellt.

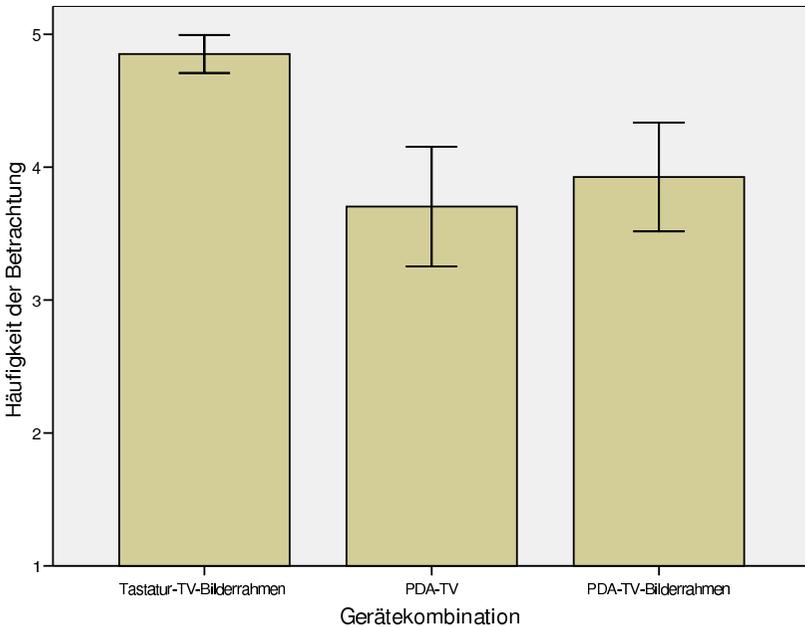


Abbildung 8.32: Häufigkeit der Betrachtung des Fernsehers getrennt nach Gerätekombinationen.

In der Tastatur-TV-Bilderrahmen-Gruppe wurde der Fernseher signifikant häufiger als der Bilderrahmen betrachtet (Wilcoxon,  $p = .001$ ; TV: MW = 4.8, SD = 0.4; Bilderrahmen: MW = 3.8, SD = 1.1). In der PDA-TV-Gruppe wurde der PDA häufiger als der Fernseher verwendet (PDA: MW = 4.3, SD = 0.9; Fernseher: MW = 3.7, SD = 1.1). Die Unterschiede waren jedoch nicht signifikant (Wilcoxon,  $p = .095$ ). In der PDA-TV-Bilderrahmen-Gruppe ergaben sich ebenfalls keine signifikanten Unterschiede bzgl. der Häufigkeit, mit der die Geräte betrachtet wurden (Friedman,  $p = .572$ ; PDA: MW = 4.0, SD = 1.2; Fernseher: MW = 3.9, SD = 1.0; Bilderrahmen: MW = 3.9, SD = 1.1).

Die Häufigkeit der Betrachtung des Fernsehers unterschied sich zwischen den drei Gerätekombinationen signifikant (Kruskal-Wallis,  $p = .001$ ). Dies lag darin begründet, dass der Fernseher in der Tastatur-TV-Bilderrahmen-Gruppe signifikant häufiger als in der PDA-TV- oder der PDA-TV-Bilderrahmen-Gruppe betrachtet wurde (Mann-Whitney; Tastatur-TV-Bilderrahmen–PDA-TV:  $p = .001$ ; Tastatur-TV-Bilderrahmen–PDA-TV-Bilderrahmen:  $p = .001$ ). Die Unterschiede werden in Abbildung 8.32 grafisch dargestellt.

Es ergaben sich darüber hinaus keine signifikanten Unterschiede zwischen den Gruppen PDA-TV und PDA-TV-Bilderrahmen bzgl. der Häufigkeit, mit der der PDA betrachtet wurde (Mann-Whitney,  $p = .511$ ). Ebenso ergaben sich keine signifikanten Unterschiede zwischen den Gruppen Tastatur-TV-Bilderrahmen und PDA-TV-Bilderrahmen bzgl. der Nutzung des Bilderrahmens (Mann-Whitney,  $p = .637$ ).

### 8.3.9 Interesse

Im Zwischenfragebogen wurden die Vpn ebenfalls gefragt, ob sie die sechs präsentierten Nachrichtenbeiträge der letzten Präsentationsbedingung interessant fanden. Der Interessantheitsgrad wurde durch die Vpn im Mittel mit 3.6 bewertet („stimme zu“, SD = .9).

Das Alter der Vpn hatte keinen signifikanten Einfluss auf die Bewertung des Interesses (Anova,  $F(23, 12) = .35$ ,  $p = .986$ ). Die Unterschiede zwischen Männern und Frauen bzgl. des Interesses waren zwar nicht signifikant, deuten aber einen Einfluss des Faktors Geschlecht an (Anova,  $F(1, 34) = 4.03$ ,  $p = .053$ ). Frauen haben den Interessantheitsgrad der Beiträge im Mittel höher als Männer bewertet (Frauen: MW = 3.9, SD = .9; Männer: MW = 3.4, SD = .8).

Es existierten keine signifikanten Unterschiede bzgl. des Interesses zwischen den Gerätegruppen (Anova,  $F(3, 32) = .127$ ,  $p = .943$ ; Tastatur-TV: MW = 3.7, SD = 1.1; Tastatur-TV-Bilderrahmen: MW = 3.7, SD = .8; PDA-TV: MW = 3.5, SD = .8; PDA-TV-Bilderrahmen: MW = 3.7, SD = .9).

### 8.3. ERGEBNISSE

#### Zeitlicher Verlauf

Das Interesse an den Nachrichtenbeiträgen sank signifikant zwischen den drei Präsentationsbedingungen (Anova,  $F(2, 70) = 4.76$ ,  $p = .012$ ). Dabei waren die Unterschiede zwischen der ersten Präsentationsbedingung im Experiment und der zweiten sowie der dritten Präsentationsbedingung signifikant (Wilcoxon; Zeitpunkt 1&2:  $p = .025$ ; Zeitpunkt 1&3:  $p = .015$ ; Zeitpunkt 2&3:  $p = .412$ ). Abbildung 8.33 zeigt die Unterschiede.

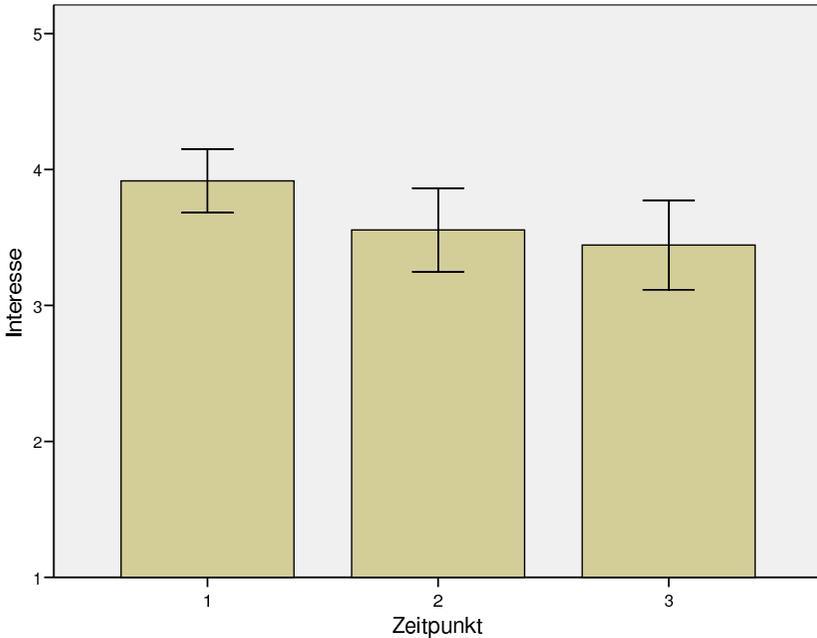


Abbildung 8.33: Interesse am Beitrag getrennt nach Zeitpunkten der Präsentation.

#### Präsentationsbedingungen und Nachrichtenfolgen

Zwischen den drei Präsentationsbedingungen BS, BSC und TB lagen keine signifikanten Unterschiede bzgl. des Interesses an den Nachrichtenbeiträgen vor (Anova,  $F(2, 70) = .375$ ,  $p = .689$ ; BS: MW = 3.6, SD = .9; BSC: MW = 3.7, SD = .8; TB: MW = 3.7, SD = .9). Neben dem Faktor Präsentationsbedingung wurde ebenfalls untersucht, ob signifikante Unterschiede bzgl. des Interesses zwischen den drei verwendeten Folgen von Nachrich-

tenbeiträgen existieren (vgl. Tabelle 8.5). Dies war nicht der Fall (Anova,  $F(2, 70) = 2.81, p = .067$ ). Die Nachrichtenfolge drei wurde jedoch als signifikant weniger interessant als die Nachrichtenfolge zwei beurteilt (Wilcoxon; Folge 1&2:  $p = .766$ ; Folge 1&3:  $p = .054$ ; Folge 2&3:  $p = .034$ ). Weiterhin wurde die Folge drei marginal signifikant weniger interessant als die Nachrichtenfolge eins beurteilt. Abbildung 8.34 stellt die Unterschiede grafisch dar.

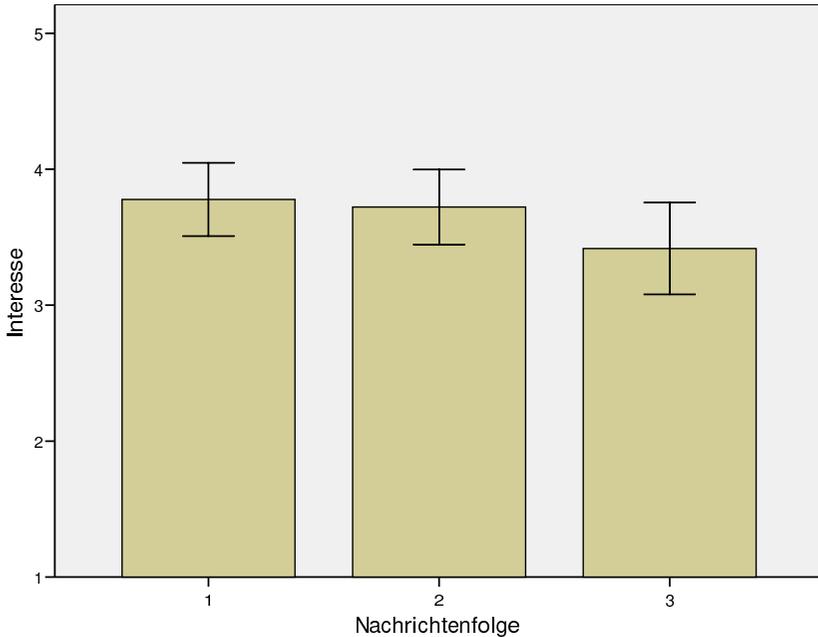


Abbildung 8.34: Interesse am Beitrag getrennt nach Nachrichtenfolgen.

### 8.3.10 Beurteilung des Gesamtsystems

Im Abschlussfragebogen wurden die Vpn gefragt, wie gut ihnen die Gerätekombination gefallen hat. Abbildung 8.35 zeigt die Bewertungen der vier Gerätekombinationen. Die Unterschiede zwischen den vier Gerätekombinationen waren insgesamt nicht signifikant (Kruskal-Wallis,  $p = .127$ ). Der Unterschied zwischen der am schlechtesten bewerteten Kombination Tastatur-TV und der am besten bewerteten Kombination PDA-TV-Bilderrahmen war jedoch signifikant (Mann-Whitney,  $p = .031$ ).

### 8.3. ERGEBNISSE

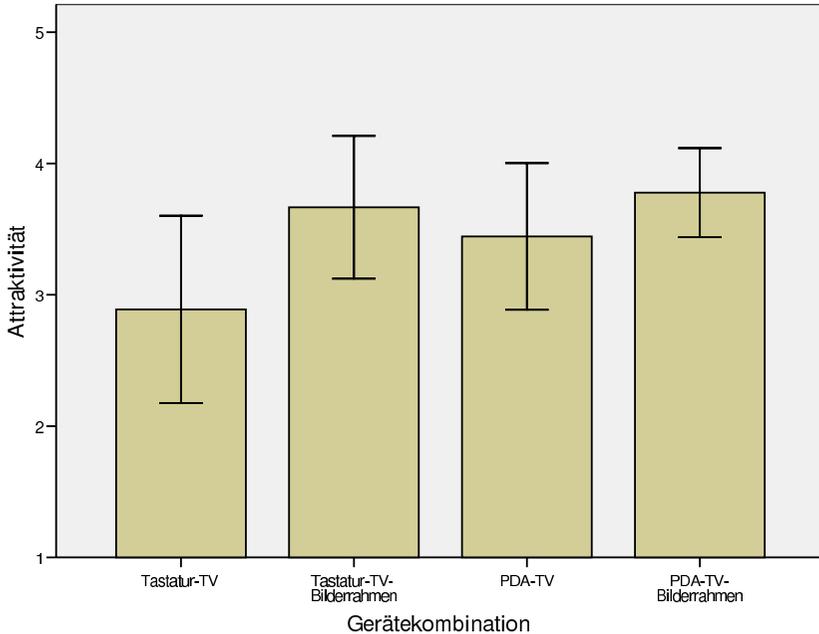


Abbildung 8.35: Attraktivität der Gerätekombinationen.

Im Abschlussfragenbogen wurde ebenfalls nach dem jeweils bevorzugten Präsentationsmodus gefragt. Abbildung 8.36 zeigt ein Histogramm der bevorzugten Präsentationsmodi. Die meisten Vpn haben Text-Bild-Präsentationen bevorzugt. Die Unterschiede zwischen den Präsentationsmodi waren jedoch nicht signifikant ( $\chi^2$ -Test,  $\chi^2(2) = 3.5$ ,  $p = .174$ ). Es existierten weiterhin keine signifikanten Unterschiede bei der Bewertung der bevorzugten Präsentationsmodi zwischen den Gerätegruppen ( $\chi^2$ -Test,  $\chi^2(6) = 2.9$ ,  $p = .821$ ).

In allen Gerätegruppen außer Tastatur-TV wurden die Vpn gefragt, ob sie die Kombination mehrerer Ausgabegeräte als nützlich empfanden. Im Durchschnitt wurde die Aussage mit „stimme nicht zu“ bewertet (MW = 2.4, SD = 1.2). Die Unterschiede zwischen den drei Gerätekombinationen waren nicht signifikant (Kruskal-Wallis,  $p = .465$ ). Die Vpn haben tendenziell abgelehnt, die Gerätekombination einem einzelnen Fernseher gegenüber vorzuziehen. Die Aussage wurde im Durchschnitt mit „stimme nicht zu“ bewertet (MW = 2.2, SD = 1.2). Die Unterschiede zwischen den drei Gerätekombinationen waren nicht signifikant (Kruskal-Wallis,  $p = .760$ ).

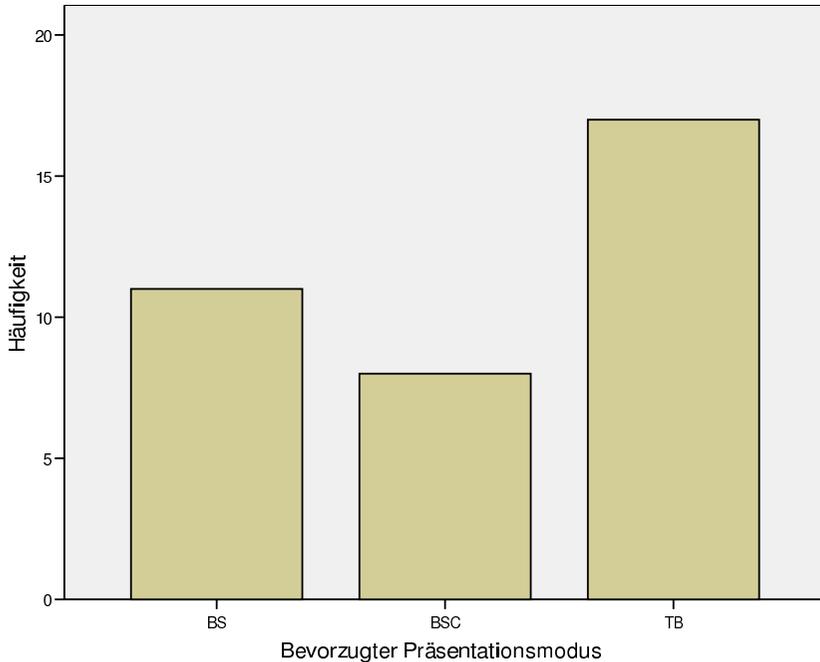


Abbildung 8.36: Histogramm bevorzugte Präsentationsmodi.

### 8.3.11 Störfaktoren

Bei den Störfaktoren wird zwischen Systemfehlern (z. B. dem Neustart des Fernsehers oder des PDAs) und dem Austausch des elektronischen Bilderrahmens durch einen Tablet PC unterschieden. Beide Faktoren werden separat analysiert.

#### Systemfehler

Insgesamt traten bei 24 Vpn während des Versuchs Systemfehler auf. Von diesen 24 Vpn musste bei 13 Vpn der Fernseher im Laufe des Experiments neu gestartet werden. Bei sechs Vpn musste im Laufe des Versuchs ein Neustart des PDAs erfolgen. Bei fünf Vpn musste das Gesamtsystem neu gestartet werden. Jeder dieser Vorfälle wurde als ein Systemfehler gezählt und für jede Präsentationsbedingung jeder Vp notiert. Im Mittel erfolgten bei allen 36 Vpn im gesamten Experiment 1.1 Systemfehler pro Vp ( $SD = 1.2$ ). Die Anzahl der Systemfehler pro Präsentationsbedingung unterschieden sich signifikant

### 8.3. ERGEBNISSE

zwischen den Gerätegruppen (Anova,  $F(3, 32) = 5.1, p = .005$ ). Abbildung 8.37 zeigt die Unterschiede.

Die Unterschiede zwischen der Tastatur-TV-Gruppe und der PDA-TV-Gruppe waren signifikant (Anova,  $F(1, 16) = 8.8, p = .009$ ). Die Unterschiede zwischen der Tastatur-TV-Gruppe und der PDA-TV-Bilderrahmen-Gruppe waren ebenfalls signifikant (Anova,  $F(1, 16) = 7.4, p = .015$ ).

Die Unterschiede zwischen der Tastatur-TV-Gruppe und der Tastatur-TV-Bilderrahmen-Gruppe waren nicht signifikant (Anova,  $F(1, 16) = 2.5, p = .135$ ). Die Unterschiede zwischen der Tastatur-TV-Bilderrahmen-Gruppe und der PDA-TV-Gruppe sowie der PDA-TV-Bilderrahmen-Gruppe waren ebenfalls nicht signifikant (Anova; Tastatur-TV-Bilderrahmen–PDA-TV:  $F(1, 16) = 3.7, p = .072$ ; Tastatur-TV-Bilderrahmen–PDA-TV-Bilderrahmen:  $F(1, 16) = 2.3, p = .147$ ). Dies zeigt, dass die Systemfehler in der Tastatur-TV-Gruppe gehäuft eintraten.

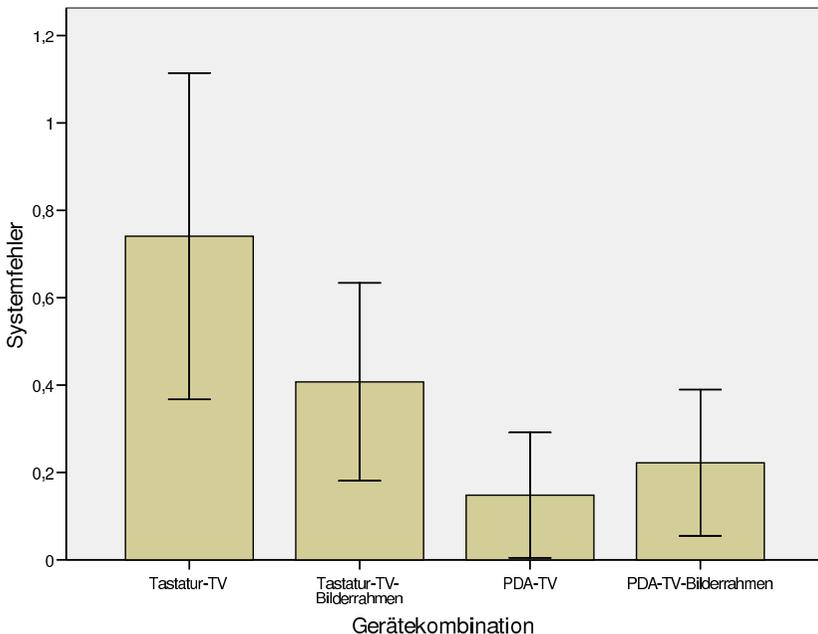


Abbildung 8.37: Systemfehler pro Präsentationsbedingung getrennt nach Gerätekombinationen.

## KAPITEL 8. LABORSTUDIE

Weiterhin wurde der zeitliche Verlauf der Systemfehler untersucht. Hier zeigten sich zwischen den drei Präsentationszeitpunkten keine signifikanten Unterschiede zwischen den Systemstörungen (Anova,  $F(2, 70) = 1.17$ ,  $p = .316$ ; Zeitpunkt 1: MW = .5, SD = .7; Zeitpunkt 2: MW = .4, SD = .6; Zeitpunkt 3: MW = .3, SD = .6). Es wurde ebenfalls untersucht, ob sich die Anzahl der Systemstörungen signifikant zwischen den Präsentationsbedingungen unterschied. Hier zeigten sich signifikante Unterschiede (Anova,  $F(2, 70) = 7.13$ ,  $p = .002$ ).

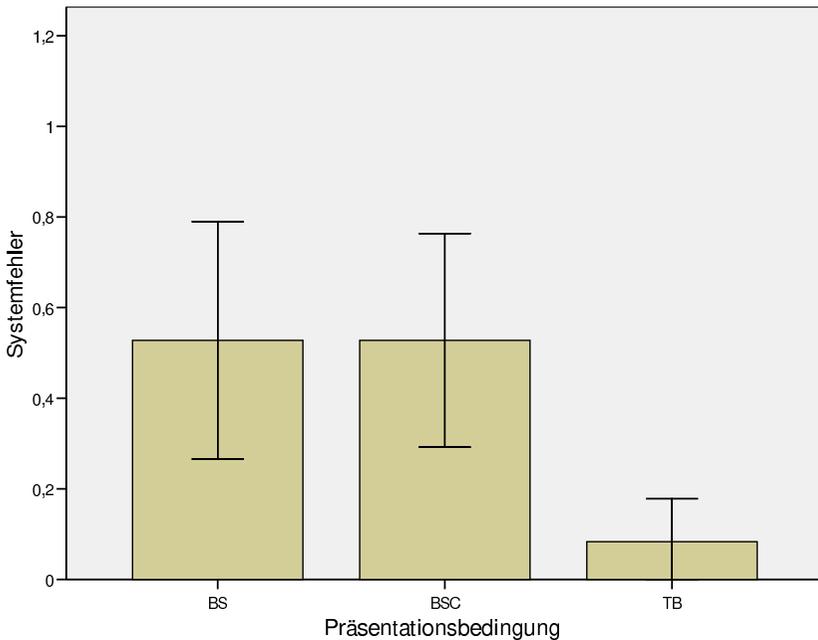


Abbildung 8.38: Systemfehler pro Präsentationsbedingung getrennt nach Präsentationsbedingungen.

Abbildung 8.38 stellt die Unterschiede dar. Dabei war die Anzahl der Systemstörungen in der TB-Bedingung signifikant niedriger als in der der BS- oder der BSC-Bedingung (t-Test; BS-BSC:  $t(35) = 1.0$ ,  $p = 1.0$ ; BS-TB:  $t(35) = 3.04$ ,  $p = .004$ ; BSC-TB:  $t(35) = 3.45$ ,  $p = .001$ ).

Darüber hinaus wurde untersucht, ob die Systemfehler Auswirkungen auf die Gesamtbewertungen der Erinnerungen einer Vp hatten. Dabei wurde nicht die

### 8.3. ERGEBNISSE

Anzahl der Systemfehler untersucht, sondern es wurde untersucht, ob Systemfehler aufgetreten waren oder nicht (unabhängig von deren Anzahl). Der Grund für diese Vorgehensweise ist, dass sonst zuwenig Daten über die einzelnen Häufigkeiten von Systemfehlern vorgelegen hätten. Es existierten keine signifikanten Unterschiede bzgl. der Gesamtbewertung der Erinnerungen zwischen Vpn, bei denen Systemfehler aufgetreten waren und Vpn, bei denen keine Systemfehler aufgetreten waren (Anova,  $F(1, 34) = .01, p = .947$ ).

Weiterhin existierten keine signifikanten Interaktionen zwischen dem Auftreten von Systemfehlern und der Gerätekombination bzgl. der Gesamtbewertungen der erinnerten Nachrichtenbeiträge (Anova, Auftreten Systemfehler  $\times$  Gerätekombination,  $F(3, 28) = .7, p = .551$ ). Die Fehler hatten ebenfalls keinen signifikanten Einfluss auf den zeitlichen Ablauf der Gesamtbewertungen der erinnerten Nachrichtenbeiträge (Anova, Auftreten Systemfehler  $\times$  Zeitpunkt,  $F(1, 34) = 1.25, p = .271$ ). Die Interaktionen der Systemfehler mit den Präsentationsbedingungen waren ebenfalls nicht signifikant (Anova, Auftreten Systemfehler  $\times$  Präsentationsbedingung, Greenhouse-Geisser,  $F(1.7, 57.9) = 2.10, p = .131$ ).

Es wurde ebenfalls untersucht, ob die Systemfehler die Akzeptanz des Systems beeinflusst haben. Die Bewertung der Gerätekombination im Abschlussfragebogen wurde durch das Auftreten von Systemfehlern nicht signifikant beeinflusst (Mann-Whitney,  $p = .830$ ). Das Auftreten von Systemfehlern hatte ebenfalls keine Auswirkung auf die bevorzugten Präsentationsbedingungen der Vpn ( $\chi^2$ -Test,  $\chi^2(2) = 3.91, p = .142$ ). Es existierte ebenfalls kein signifikanter Einfluss des Auftretens der Systemfehler auf die Bewertung der Darstellung in den Präsentationsbedingungen BS, BSC und TB im Zwischenfragebogen (Anova,  $F(2, 68) = .25, p = .776$ ).

#### Austausch elektronischer Bilderrahmen

Von den 18 Vpn in den Gruppen mit Bilderrahmen (Tastatur-TV-Bilderrahmen und PDA-TV-Bilderrahmen) hatten sechs den durch einen Tablet PC simulierten Bilderrahmen verwendet. Von diesen sechs Vpn befanden sich vier Vpn in der Tastatur-TV-Bilderrahmen-Gruppe und zwei Vpn in der PDA-TV-Bilderrahmen-Gruppe. Im Folgenden wird untersucht, ob der simulierte Bilderrahmen Auswirkungen auf die Bewertungen der Erinnerungen oder die Akzeptanz des Systems hatte. Dabei wurden nur die Tastatur-TV-Bilderrahmen-Gruppe und die PDA-TV-Bilderrahmen-Gruppe betrachtet, da nur diese Gerätekombinationen von dem Faktor beeinflusst wurden.

Es wurde untersucht, ob zwischen den Gruppen mit dem echten und dem simulierten Bilderrahmen signifikante Unterschiede bzgl. der Gesamtbewertungen der Erinnerungen bestehen. Dies war nicht der Fall (Anova,  $F(1, 16) =$

.23,  $p = .639$ ; Bilderrahmen:  $MW = 24.4$ ,  $SD = 11.0$ ; Bilderrahmenersatz:  $MW = 21.2$ ,  $SD = 11.7$ ). Derselbe Test wurde auch für die Bewertungen der erinnerten Bilder durchgeführt. Zwischen den beiden Bilderrahmengruppen wurden keine signifikanten Unterschiede gefunden (Anova,  $F(1, 16) = 2.54$ ,  $p = .130$ ; Bilderrahmen:  $MW = .54$ ,  $SD = .9$ ; Bilderrahmenersatz:  $MW = .39$ ,  $SD = .6$ ).

Weiterhin existierten keine signifikanten Interaktionen zwischen den beiden Bilderrahmentypen und den beiden Gerätegruppen mit Bilderrahmen bzgl. der Gesamtbewertung der erinnerten Beiträge (Anova, Bilderrahmentyp $\times$ Gerätekombination,  $F(1, 14) = .01$ ,  $p = .940$ ). Es ergaben sich ebenfalls keine signifikanten Interaktionen zwischen dem Typ des Bilderrahmens und dem zeitlichen Verlauf bzgl. der Gesamtbewertung der erinnerten Beiträge (Anova, Bilderrahmentyp $\times$ Zeitpunkt,  $F(2, 32) = 2.52$ ,  $p = .096$ ). Darüber hinaus existieren keine signifikanten Interaktionen zwischen dem Typ des Bilderrahmens und der Präsentationsbedingung (Anova, Bilderrahmentyp $\times$ Präsentation,  $F(2, 32) = .12$ ,  $p = .891$ ).

Der Typ des Bilderrahmens hatte keinen signifikanten Einfluss auf die Bewertungen der Gerätekombinationen Tastatur-TV-Bilderrahmen und PDA-TV-Bilderrahmen im Abschlussfragebogen (Mann-Whitney,  $p = .750$ ). Der Typ des Bilderrahmens hatte ebenfalls keinen signifikanten Einfluss auf die Wahl der bevorzugten Präsentationsbedingung im Abschlussfragebogen ( $\chi^2$ -Test,  $\chi^2(2) = 1.31$ ,  $p = .519$ ). Weiterhin wurde die Bewertung der Darstellung in den Präsentationsbedingungen nicht signifikant durch den Typ des Bilderrahmens beeinflusst (Anova,  $F(1, 16) = .26$ ,  $p = .614$ ).

### 8.3.12 Kommentare

Während des Versuchs konnten die Vpn sowohl in den drei Zwischenfragebögen als auch im Abschlussfragebogen Kommentare abgeben. Die Kommentare in den Zwischenfragebögen bezogen sich hauptsächlich auf die vorhergehende Präsentationsbedingung. Die Kommentare im Abschlussfragebogen betrafen den Gesamteindruck vom Experiment.

#### Zwischenfragebögen

In den Zwischenfragebögen wurde die Sprachausgabe in der BS- und der BSC-Bedingung am häufigsten kritisiert, wobei die meiste Kritik in der BS-Bedingung geäußert wurde (BS: 23 Vpn; BSC: 17 Vpn). Die Kritik bezog sich auf die verzerrte Ausgabe, die monotone Stimme und die unnatürliche Satzmelodie. Weitere Kritikpunkte betrafen den Sprachfluss und Pausen an unpassenden Stellen. Die Vpn kritisierten ebenfalls die Qualität der dargestellten Bilder (TB: 5 Vpn; BS: 3 Vpn; BSC: 2 Vpn). Die Bilder sollten bewegt, grö-

## 8.4. DISKUSSION

ßer und schärfer sein. In der TB-Bedingung wurden Verbesserungsvorschläge zur Präsentation des Textes gemacht (3 Vpn). Die Vpn regten die Verwendung von Farben und eine größere Schrift an.

### Abschlussfragebogen

In der Tastatur-TV-Gruppe bezog sich die meiste Kritik auf die Sprachausgabe, die als schwer verständlich und mechanisch kritisiert wurde (6 Vpn). Drei Vpn merkten an, dass ihnen die TB-Bedingung am besten gefallen hat. Eine Vp schlug vor, einen realen Nachrichtensprecher anstelle des animierten Charakters zu nehmen. Eine Vp hat den animierten Charakter als ansprechendste Präsentationsbedingung empfunden. Eine Vp hat er irritiert.

In der Tastatur-TV-Bilderrahmen-Gruppe kritisierten drei Vpn die Sprachausgabe. Vier Vpn kritisierten die Verteilung von Informationen auf mehrere Ausgabegeräte. Eine Vp schlug vor, nur ein großes Display anstelle von mehreren kleineren zu verwenden. Eine zweite Vp bemerkte, ein zweiter Bildschirm mache nur dann Sinn, wenn auf ihm zusätzliche Interaktionen möglich seien (z. B. Interaktionen mit einer Landkarte während einer Tagesschauendung). Eine Vp lobte den animierten Charakter.

In der PDA-TV-Gruppe haben fünf Vpn die Qualität der Sprachausgabe kritisiert. Zwei Vpn sagten, dass sie eine Ausgabe auf einem Gerät vorziehen würden. Eine Vp kritisierte die Qualität der Bilder auf dem Fernseher. Eine Vp kritisierte die Qualität der Bilder auf dem PDA, die wegen des spiegelnden Lichtes schlecht zu erkennen gewesen seien. Eine Vp fand es ansprechend, Nachrichten auf Knopfdruck selber steuern zu können.

In der PDA-TV-Bilderrahmen-Gruppe wurde die Sprachausgabe durch vier Vpn kritisiert. Zwei Vpn kritisierten die Bildqualität. Zwei Vpn merkten an, dass die Ausgabe auf drei Geräten verwirrend sei bzw. dass sie einen einzelnen Fernseher bevorzugen würden. Eine weitere Vpn merkte an, dass man auf dem Bilderrahmen anstelle eines dritten Bildes besser ein andere Modalität, z. B. ein Diagramm verwenden sollte. Eine Vpn lobte den Bilderrahmen als angenehm.

### 8.4 Diskussion

Die Analyse der erinnerten Nachrichtenbeiträgen, der Interaktionszeiten und der Akzeptanz zeigte Unterschiede zwischen den Gerätekombinationen und den Präsentationsbedingungen auf. Diese werden im Folgenden nach den relevanten abhängigen Variablen gegliedert und dargestellt.

### Gesamtbewertung erinnertes Nachrichtenbeiträge

Bei der Gesamtbewertung der erinnerten Nachrichtenbeiträge, die die Bewertungen von erinnerten Titeln, Textinhalten und Bildern umfasste (Abschnitt 8.3.3), erreichten die Vpn zwischen 23 und 150 Bewertungspunkte von maximal 360 möglichen Punkten. Dies zeigt, dass die Erinnerungsaufgabe vom Schwierigkeitsgrad her angemessen war. Die Streuung weist jedoch auf individuelle Unterschiede zwischen den Vpn hin.

Die drei verwendeten Folgen von Nachrichtenbeiträgen unterschieden sich wie gewünscht nicht signifikant bzgl. der Gesamtbewertungen der Erinnerungen. Dies bedeutet, dass keine der Folgen signifikant leichter zu erinnern war als die anderen beiden. Dieser Effekt hätte z. B. auftreten können, falls sich eine Folge nur aus Nachrichtenbeiträgen von allgemeinem Interesse zusammengesetzt hätte, während die anderen beiden Folgen nur Beiträge enthalten, die nur für spezielle Gruppen von Vpn interessant sind.

Entgegen den Erwartungen zeigten sich zwischen den vier Gerätekombinationen keine signifikanten Unterschiede bzgl. der Gesamtbewertungen der Erinnerungen. Damit konnte Hypothese 1 aus Abschnitt 8.1.3 bzgl. der Gesamtbewertungen der Erinnerungen nicht bestätigt werden. Bei Präsentationen zwischen gekoppelten Ausgabegeräten ergaben sich anscheinend durch den vergrößerten Bildschirmplatz und die reduzierte Zahl an Interaktionen nur geringe Vorteile für die Vpn. Wie in [Tan & Czerwinski, 2003] führte die Verteilung der Informationen über mehrere Bildschirme jedoch ebenfalls zu keiner signifikanten Verschlechterung der Erinnerungseffizienz.

Zwischen den Präsentationsbedingungen BS, BSC und TB wurden weder in den Gerätekombinationen mit PDA noch in den Gruppen mit Tastatur und Fernseher signifikante Unterschiede bzgl. der Gesamtbewertung der erinnerten Beiträge beobachtet. Damit konnte Hypothese 3 nicht bestätigt werden. Da die Sprachausgabe trotz Austausch der Mbrola-Software durch die AT&T Natural Voices 1.4 Software sehr schlecht beurteilt wurde, kann sich dieser Effekt auch auf die Erinnerungsleistung ausgewirkt haben. Dadurch könnte der Modalitäteneffekt nicht beobachtbar gewesen sein. Dafür spricht ebenfalls, dass die Vpn angegeben haben, dass sie sich in der BS-Bedingung signifikant mehr konzentrieren mussten als in der TB-Bedingung.

Ein weiterer Effekt wurde auf der Gerätekombination PDA-TV-Bilderrahmen beobachtet. Hier zeigten sich marginal signifikante Unterschiede bzgl. des Geschlechts der Vpn. Die Gesamtbewertungen der erinnerten Beiträge der Frauen waren in dieser Gerätekombination in allen Präsentationsbedingungen höher als die der Männer. Diese Unterschiede waren in der BSC-Bedingung dieser Gruppe signifikant (vgl. Abbildung 8.14). Dies kann als Hinweis auf

## 8.4. DISKUSSION

geschlechtsspezifische Unterschiede bei Präsentationen mit gekoppelten Ausgabegeräten gedeutet werden. Bei [Czerwinski et al., 2002] wurde ebenfalls beobachtet, dass Frauen von einer Vergrößerung des Blickfeldes mittels eines größeren Bildschirms mehr profitieren als Männer. Dies wurde beim Vergleich eines 18 Zoll mit einem 36 Zoll großen Bildschirms festgestellt. Jedoch bestand die Aufgabe der Vpn aus einer 3D-Navigation durch einen virtuellen Raum und nicht aus einem Erinnerungsexperiment. Daher sind hier weitere Studien erforderlich, um den hier beobachteten Effekt zu untersuchen.

Es wurden ebenfalls signifikante Interaktionen zwischen den drei Präsentationsbedingungen und dem Alter der Vpn bzgl. der Gesamtbewertung der erinnerten Beiträge festgestellt (Abbildung 8.15). Dies scheint damit zusammenzuhängen, dass Vpn, die jünger als 25 waren, in der BS-Bedingung schlechtere Resultate erzielten. Dieser Zusammenhang bedarf jedoch ebenfalls weiterer Untersuchungen.

Die Bewertungen der Erinnerungen stiegen von der ersten Präsentationsbedingung zur zweiten Präsentationsbedingung im Versuchsablauf signifikant an (vgl. Abbildung 8.13). Dies deutet auf einen Lerneffekt hin, der nach der ersten Präsentationsbedingung und dem ersten Abfragen der erinnerten Nachrichtenbeiträge aufgetreten ist. Dieser setzte sich jedoch in der dritten Präsentationsbedingung nicht fort. Der Lerneffekt ähnelt dem in [Kruppa, 2006] berichteten Lerneffekt bei Präsentationen mit gekoppelten Ausgabegeräten. Jedoch ergaben sich in der hier durchgeführten Studie keine signifikanten Unterschiede zwischen den Gerätekombinationen.

Darüber hinaus zeigte sich, dass die Lebhaftigkeit der Erinnerungen der Vpn im Laufe des Versuchs signifikant abnahm (Abbildung 8.18). Dies korrespondiert mit den Angaben der Teilnehmer über die benötigte Konzentration, die während des Experiments kontinuierlich anstieg (Abbildung 8.30). Aufgrund der ansteigenden Konzentration erfolgte in der dritten Präsentationsbedingung, in der kein Lerneffekt mehr beobachtbar war, jedoch keine Verschlechterung der Gesamtbewertungen der erinnerten Beiträge.

**Bildbewertung erinnerten Nachrichtenbeiträge**

Von den 108 möglichen Punkten für erinnerte Bilder wurden pro Vp nur bis zu sieben erreicht. Dies liegt darin begründet, dass die Teilnehmer in der Instruktion nicht explizit aufgefordert wurden, sich an Bilder zu erinnern. Trotzdem zeigten sich signifikante Unterschiede zwischen den Gerätekombinationen und den Präsentationsbedingungen (vgl. Abbildung 8.16). Hierbei zeigte sich insbesondere ein positiver Einfluss des Bilderrahmens, da in den Gerätekombinationen Tastatur-TV-Bilderrahmen und PDA-TV-Bilderrahmen die meisten Bildpunkte erreicht wurden. Weiterhin waren die Unterschiede

zur Tastatur-TV-Gruppe jeweils signifikant. Dies korrespondiert mit den Bewertungen der Akzeptanz der Gerätekombinationen im Abschlussfragebogen, bei denen die Kombinationen Tastatur-TV-Bilderrahmen und PDA-TV-Bilderrahmen am besten beurteilt wurden. Bezüglich der Bildbewertungen bei den Kombinationen Tastatur-TV-Bilderrahmen und PDA-TV-Bilderrahmen wurde Hypothese 1 im Gegensatz zur Gesamtbewertung der erinnerten Beiträge somit bestätigt. Weiterhin zeigte sich ein positiver Einfluss des animierten Charakters in der BSC-Bedingung, in der signifikant höhere Bildbewertungen als die TB-Bedingung erzielt wurden (Abbildung 8.17). Dies lässt vermuten, dass der animierte Charakter die Aufmerksamkeit der Vpn auf die Bilder gelenkt hat, die auf demselben Gerät wie der Charakter vorhanden waren und dadurch besser erinnert wurden.

### Betrachtungszeiten

Bei den Betrachtungszeiten jeder Präsentation zeigten sich signifikante Unterschiede zwischen den vier Gerätekombinationen (Abbildung 8.21). Hierbei waren die Zeiten in den Gerätekombinationen mit elektronischem Bilderrahmen (Tastatur-TV-Bilderrahmen und PDA-TV-Bilderrahmen) signifikant geringer als bei den Kombinationen ohne Bilderrahmen (Tastatur-TV und PDA-TV). Somit wurde Hypothese 2 für die Tastatur-TV-Bilderrahmen-Gruppe und die PDA-TV-Bilderrahmen-Gruppe bestätigt. Die Überlegenheit der Gruppen mit Bilderrahmen erklärt sich dadurch, dass ein zusätzliches Segment dargestellt werden konnte und dadurch weniger Interaktionen zum Betrachten der Nachrichtenbeiträge nötig waren (vgl. Tabellen 8.3 und 8.4).

In der PDA-TV-Kombination konnten zwar in der BS und der BSC-Bedingung gleich viele Segmente dargestellt werden. In der TB-Bedingung wurde jedoch ein Segment weniger dargestellt, was anscheinend zu einem signifikanten Unterschied zwischen der Tastatur-TV-Bilderrahmen-Gruppe und der PDA-TV-Gruppe bzgl. der Interaktionszeiten führte. Durch die Darstellung eines weiteren Segments in der PDA-TV-Bilderrahmen-Gruppe nahmen die Betrachtungszeiten pro Nachrichtenbeitrag jedoch weiter ab. Die zusätzlichen Ausgabegeräte scheinen im Experiment für keine signifikante Verzögerung des Betrachtens einer Präsentation (z. B. durch Fokuswechsel von einem Gerät zum anderen) gesorgt zu haben.

In jeder Präsentationsbedingung wurden die sechs Beiträge innerhalb einer Bedingung zunehmend schneller betrachtet (Abbildung 8.22). Am Anfang der Präsentation hatten sich die Vpn trotz des Testsegments zu Beginn des Versuchs offenbar erst an das Format der Präsentation gewöhnen müssen. Anschließend setzte ein Lerneffekt ein, so dass die Vpn die Präsentationen schneller betrachten konnten.

## 8.4. DISKUSSION

Weiterhin zeigte sich, dass die Betrachtungszeiten in der BSC-Bedingung signifikant geringer waren als in der BS- oder der TB-Bedingung (vgl. Abbildung 8.23). Da die Vpn weiterhin angaben, sich bei BSC signifikant weniger konzentrieren zu müssen als bei BS (vgl. Abbildung 8.30), scheint der animierte Charakter einen positiven Einfluss gehabt zu haben. Die Ergebnisse lassen vermuten, dass die Vpn im Vergleich zur BS-Bedingung entspannter waren und dadurch die Präsentationen schneller betrachten konnten. Dies lässt sich als Bestätigung des „Persona Effekts“ aus [Van Mulken et al., 1998] interpretieren.

Die Beobachtung, dass BSC ebenfalls schneller als TB war, lässt sich dadurch erklären, dass in der TB-Bedingung in den Gerätekombinationen PDA-TV und PDA-TV-Bilderrahmen aufgrund des komplexeren Layouts mehr Ausgabeschritte notwendig waren als in der BS- oder der BSC-Bedingung (Tabelle 8.4). Die obigen Ergebnisse bestätigen Hypothese 4 in Bezug auf die subjektiv empfundene kognitive Last und die Betrachtungszeiten. Ein positiver Einfluss des animierten Charakters auf die Gesamtbewertung der erinnerten Beiträge konnte jedoch nicht festgestellt werden.

### Erinnerungseffizienz

Die Betrachtung der Erinnerungseffizienz als Quotient der Gesamtbewertung erinnerten Beiträge und der Betrachtungsdauer eines Beitrages hat keine neuen Erkenntnisse erbracht. Der bei den Gesamtbewertungen der erinnerten Beiträge beobachtete Lerneffekt zwischen der ersten und der zweiten Präsentationsbedingung wird auch bei der Erinnerungseffizienz deutlich. Die Effizienz stieg signifikant zwischen beiden Zeitpunkten (vgl. Abbildung 8.25).

Weiterhin bestätigt sich der bei den Betrachtungszeiten pro Beitrag beobachtete „Persona Effekt“. Die BSC-Bedingung erwies sich als signifikant effizienter als die BS-Bedingung (vgl. Abbildung 8.26). Dieser Unterschied erwies sich insbesondere bei Frauen als signifikant (vgl. Abbildung 8.27). Ein analoger Effekt wurde bei der Gesamtbewertung der erinnerten Beiträge in der PDA-TV-Gruppe beobachtet. Beide Ergebnisse deuten darauf hin, dass Frauen von dem animierten Charakter mehr profitierten als Männer.

### Akzeptanz

Wie in der Fragebogenstudie erzielten Gerätekombinationen mit gekoppelten Ausgabegeräten die besten Bewertungen. Im Abschlussfragebogen wurde die Gerätekombination PDA-TV-Bilderrahmen am besten bewertet (vgl. Abbildung 8.35). Dabei wurde die Aussage, dass die Darstellung auf der Gerätekombination ansprechend war, im Mittel mit „stimme zu“ bewertet. Hypothese 6 wurde somit bestätigt. Diese Bewertung erwies sich als signifikant besser

## KAPITEL 8. LABORSTUDIE

als die Bewertung der Tastatur-TV-Gruppe. Im Unterschied zur Fragebogenstudie wurden jedoch keine Abhängigkeiten der Bewertungen vom Alter der Vpn beobachtet. Dies kann verschiedene Gründe haben. Zum einen bekamen die Vpn die Systemfunktionalität nicht nur vorgeführt, sondern konnten sie selber ausprobieren. Weiterhin bestand die Applikation nicht aus einem Filminformationssystem sondern aus einem Nachrichteninformationssystem. Es ist möglich, dass diese Applikation daher keine Altersunterschiede in der Akzeptanz zur Folge hatte. Hypothese 7 konnte somit nicht bestätigt werden.

Die Hypothese, dass die neue AT&T-Sprachausgabe zu besseren Bewertungen von Sprachausgabe und animiertem Charakter führt (Hypothese 8) konnte nicht bestätigt werden. Dies drückte sich in den Bewertungen der Modalitäten Bild, Sprache und Charakter im Zwischenfragebogen aus, bei der die Vpn der Aussage widersprochen haben, dass ihnen die Sprachausgabe oder der animierte Helfer gefallen hat („stimmt nicht“, vgl. Abbildung 8.28 für die BSC-Bedingung).

Bilder wurden hingegen positiv beurteilt („stimmt“). In den Kommentaren wurde jedoch Kritik geäußert, dass die Bilder zu unscharf waren, was durch die begrenzte Auflösung der Bilder bedingt war. Weiterhin wurden Bilder in der PDA-TV-Bedingung signifikant schlechter als in den anderen Gerätekombinationen beurteilt. Die Interpretation dieses Effekts bedarf jedoch weiterer Untersuchungen.

Bei der Sprachausgabe wurden in den Kommentaren insbesondere die monotone Stimme und die unnatürliche Satzmelodie kritisiert. Die schlechte Akzeptanz der Sprachausgabe war vermutlich ebenfalls für die schlechte Akzeptanz des Helfers verantwortlich. Weiterhin können die eingeschränkten Ausdrucksmöglichkeiten des Charakters (Sprech- und Ruheanimationen) ein Grund dafür sein, dass die Teilnehmer keine Vorteile in der Verwendung eines Charakters gesehen und ihn als überflüssig betrachtet haben. Das in [Van Mulken et al., 1998] durchgeführte Experiment verwendete ebenfalls einen animierten Charakter mit beschränkter Ausdruckskraft. Dieser Charakter setzte jedoch im Unterschied zum hier eingesetzten Charakter Zeigegesten ein. Präsentationen mit dem Charakter wurden in dieser Studie als unterhaltsamer im Vergleich zu Präsentationen ohne den Charakter eingestuft.

Eine zweite Möglichkeit ist, dass Vpn den Eindruck hatten, dass der Typ des Charakters nicht zur Applikation eines Nachrichteninformationssystems passte. So erwähnte eine Vpn, dass sie lieber einen realen Nachrichtensprecher als eine „Zeichentrückfigur“ sehen möchte.

Die Vpn fand es unpassend, dass dieser Charakter von Konflikten im nahen Osten berichtet. Trotz der niedrigen Akzeptanz von Sprache und Charakter ist

## 8.4. DISKUSSION

es überraschend, dass BSC eine niedrigere Konzentration, geringere Interaktionszeiten, mehr Bilderinnerungen und eine höhere Erinnerungseffizienz als BS zur Folge hatte.

Diese Bewertungen der Ausgabemodalitäten wurden bei der Bewertung der Präsentationsbedingung im Zwischenfragebogen bestätigt, bei der sich TB als signifikant attraktivste Präsentationsbedingung ergab (vgl. Abbildung 8.29). Weiterhin wurde im Abschlussfragebogen TB als die bevorzugte Präsentationsbedingung angegeben und BSC als die am wenigstens bevorzugte (Abbildung 8.36). Somit bevorzugten die Vpn wie in der Fragebogenstudie und in [Elting et al., 2002] eindeutig die TB-Bedingung.

Wie in der Fragebogenstudie wurden Vpn der Laborstudie, die nicht in der Tastatur-TV-Gruppe waren, gefragt, ob sie die Kombination mehrerer Ausgabegeräte als nützlich empfanden. Die meisten Teilnehmer haben dies im Unterschied zur Fragebogenstudie verneint („stimme nicht zu“). Weiterhin wurden die Teilnehmer gefragt, ob sie bei der Ausgabe von Nachrichten eine Kombination mehrerer Geräte einem einzelnen Fernseher gegenüber vorziehen würden. Im Unterschied zur Fragebogenstudie wurde dies ebenfalls durch die Teilnehmer verneint („stimme nicht zu“).

Die Antworten lassen vermuten, dass Vpn bei einem Nachrichteninformati-onssystem keine Vorteile bei der Kombination mehrerer Geräte sahen. Nachrichten werden im Gegensatz zu multimedialen Filminformationen seit mehr als 50 Jahren auf einem Fernseher audiovisuell dargestellt. Daher ist es möglich, dass die Vpn bei dieser Applikation konservativer urteilten als bei multimedialen Filminformationssystemen, die erst mit dem World Wide Web populär geworden sind.

Weiterhin ist zu berücksichtigen, dass die Teilnehmer in dieser Studie weniger technisch orientiert sind als die Teilnehmer in der Fragebogenstudie. Dies äußert sich z. B. in der Anzahl der Benutzer, die einen PDA verwenden (Fragebogenstudie: 32.1%; Laborstudie: 16.7%).

### Konzentration, Interesse und Betrachtungshäufigkeiten

Frauen haben angegeben, dass sie sich im Experiment signifikant mehr konzentrieren mussten als Männer. Dies lässt vermuten, dass Frauen bei der Beurteilung der eigenen Konzentration selbstkritischer waren als Männer. Weiterhin stieg die benötigte Konzentration der Vpn über die drei Präsentationsbedingungen signifikant an (vgl. Abbildung 8.31). Dies zeigt, dass das Experiment wie beabsichtigt kognitive Lasteffekte hervorgerufen hat.

Eine umgekehrt proportionale Entwicklung zeigt das Interesse an den Nachrichtenbeiträgen, das mit jeder Präsentationsbedingung sinkt (vgl. Abbildung

8.33). Dies reflektiert ebenfalls, dass das Experiment anstrengend für die Vpn war. Das Interesse an der Nachrichtenfolge drei war signifikant geringer als an der Folge zwei. Es war ebenfalls marginal signifikant geringer als an der Folge eins. Eine Analyse der sechs Nachrichtenbeiträge in Folge drei ergab, dass sich in Folge drei keine Deutschland-spezifischen Beiträge im Gegensatz zu den anderen beiden Folgen befanden, was eine mögliche Erklärung für die beobachteten Unterschiede darstellt. Es wurden jedoch im Experiment keine signifikanten Auswirkungen der Unterschiede festgestellt.

Bei den Häufigkeiten, mit denen die einzelnen Geräte betrachtet wurden, zeigte sich in der Tastatur-TV-Bilderrahmen-Kombination wie erwartet, dass der Fernseher, auf dem der Text präsentiert wurde, häufiger betrachtet wurde als der Bilderrahmen. In der PDA-TV-Gruppe und der PDA-TV-Bilderrahmen-Gruppe fanden sich keine signifikanten Unterschiede zwischen den Betrachtungshäufigkeiten zwischen den Geräten. Es zeigte sich jedoch, dass der Fernseher in der Tastatur-TV-Bilderrahmen-Gruppe signifikant häufiger als in der PDA-TV-Gruppe und der PDA-TV-Bilderrahmen-Gruppe betrachtet wurde.

Dies liegt wahrscheinlich darin begründet, dass in der PDA-TV-Kombination und der PDA-TV-Bilderrahmen-Kombination der Text auf dem PDA präsentiert wurde. Somit wurde Hypothese 5 m.E. bestätigt.

### Störfaktoren

Die Systemfehler traten am häufigsten in der Tastatur-TV-Gruppe und den Präsentationsbedingungen mit Sprachausgabe (BS, BSC) auf (Abbildungen 8.37 und 8.38). Dies liegt darin begründet, dass der Fernseher sich als das instabilste Gerät erwiesen hat und dass die BS-Bedingung sowie die BSC-Bedingung die speicheraufwendigsten multimedialen Präsentationen beinhalteten. In der Tastatur-TV-Gruppe fanden diese Präsentationen auf dem Fernseher statt. Es wurden jedoch keine signifikanten Unterschiede zwischen Vpn, bei denen Systemfehler auftraten und Vpn, bei denen keine Systemfehler auftraten, beobachtet.

Dies spiegelt sich auch in den Kommentaren der Vpn wieder, in denen die Systemfehler nicht kritisiert werden. Da den Vpn bereits in der Instruktion gesagt wurde, dass es sich bei dem System um einen Prototypen handelt, bei dem Fehler auftreten können, haben sich die Störungen auf die Akzeptanz vermutlich nicht negativ ausgewirkt. Da die meisten Störungen vom abgetrennten Regieraum aus behoben werden konnten, ergab sich anscheinend ebenfalls keine signifikante Beeinflussung der Bewertungen der erinnerten Beiträge. Der Austausch des elektronischen Bilderrahmens durch einen Tablet PC schien ebenfalls keinen signifikanten Einfluss auf das Experiment gehabt zu haben.

### 8.5 Zusammenfassung

In der durchgeführten Studie zeigte sich in den Kombinationen Tastatur-TV-Bilderrahmen und PDA-TV-Bilderrahmen ein positiver Einfluss des elektronischen Bilderrahmens. In beiden Gruppen erinnerten sich Vpn an signifikant mehr Bilder. Weiterhin haben sich Vpn in diesen Gruppen die präsentierten Nachrichtenbeiträge signifikant schneller angeschaut. Die Gerätekombination PDA-TV-Bilderrahmen wurde ebenfalls als Gerätekombination am besten bewertet und unterschied sich signifikant von den Bewertungen der Tastatur-TV-Gruppe. Weiterhin zeigten sich Hinweise auf einen positiven Effekt des animierten Charakters, der die subjektiv empfundene kognitive Last der Vpn senkte und eine signifikant höhere Erinnerungseffizienz als eine Sprachausgabe ohne grafisches Feedback zur Folge hatte.

Insgesamt wurde das System in der Laborstudie jedoch schlechter als in der Fragebogenstudie beurteilt. Dies liegt vermutlich darin begründet, dass die im Experiment generierten Präsentationen Nachrichteninhalte zum Thema hatten und keine Filminhalte. Weiterhin konnte die neue Sprachausgabe keine Verbesserung der Akzeptanz im Vergleich zur Fragebogenstudie erzielen. Im Experiment traten ebenfalls durch Systemfehler und einen Austausch des elektronischen Bilderrahmens Störfaktoren auf. Diese wurden jedoch als Störvariablen in die Auswertung einbezogen und hatten keine signifikanten Auswirkungen auf das Experiment.

Das Experiment zeigt, dass sich der elektronische Bilderrahmen als zusätzlicher Bildschirm positiv auf die Erinnerung von Bildern und die Interaktionszeiten auswirkt. Der elektronische Bilderrahmen wirkte sich ebenfalls positiv auf die Akzeptanz des Systems aus. Dies zeigt, dass die gekoppelten Ausgabegeräte des Home-Entertainment-Systems das Potential haben, Informationen effizient und für den Benutzer ansprechend darzustellen.



## 9 Wissenschaftlicher Beitrag und Ausblick

Das Thema dieser Arbeit ist die Entwicklung, Umsetzung und Evaluierung eines multimedialen Präsentationssystems für gekoppelte Ausgabegeräte. Die wissenschaftlichen Beiträge dieser Arbeit werden wie folgt dargestellt. Zunächst werden die erbrachten Ergebnisse anhand der Gliederung der Arbeit beschrieben. Anschließend werden die Ergebnisse mit Bezug auf die in Abschnitt 1.2 gestellten Forschungsfragen analysiert. Abschließend erfolgt eine Diskussion der Ansatzpunkte für weitere Forschungsarbeiten.

### 9.1 Zusammenfassung der Ergebnisse

Der wissenschaftliche Beitrag dieser Arbeit setzt sich zusammen aus der Analyse der bisherigen Forschungsarbeiten im Bereich von gekoppelten Ausgabegeräten (Kapitel 3), der Entwicklung eines Ansatzes für die Generierung multimedialer Präsentationen mit gekoppelten Ausgabegeräten (Kapitel 5), der Implementierung des Ansatzes (Kapitel 6) und der Evaluierung der Implementierung in Benutzerstudien (Kapitel 7 und 8). Dabei wurden die folgenden Ergebnisse erzielt:

#### 1. Konzeptuelle Beiträge:

Der Schwerpunkt der konzeptuellen Arbeit war die Definition eines Modells zur automatisierten Generierung von multimedialen Präsentationen auf gekoppelten Ausgabegeräten. Dazu wurde das Problem als Planungsproblem formuliert, wobei auf existierenden Arbeiten zur Präsentationsplanung mit einem Ausgabegerät aufgebaut wurde. Zur Formulierung dieses Problems war insbesondere die Definition eines Dienst-, Geräte- und Ressourcenmodells für Ein-/Ausgabekomponenten notwendig, die ebenfalls in diesem Kapitel erfolgte.

Es wurden vier mögliche Ansätze zur Realisierung von gekoppelten Ausgabegeräten in ubiquitären Informationssystemen identifiziert und diskutiert. Der in dieser Arbeit verwendete Ansatz definiert ein Modell für Ein-/Ausgabedienste, das auf einer Menge von domänenunabhängigen Basiskonzepten aufbaut. Für eine bestimmte Domäne können diese Konzepte gemäß den Anforderungen des Szenarios erweitert und spezifiziert werden. Dies wurde am Beispiel einer Home-Entertainment-Domäne durchgeführt.

Durch das Modell für Ein-/Ausgabedienste können zwar keine völlig unbekanntes Ausgabekomponenten beschrieben werden. Diese können jedoch durch eine Erweiterung des Modells leicht integriert werden.

Dies wird durch die Trennung zwischen domänenunabhängigen und domänenabhängigen Konzepten im Modell unterstützt. Weiterhin können multi-modale Komponententypen definiert werden, um Komponentenklassen (z. B. animierte Charaktere) zu beschreiben. Dies erleichtert ebenfalls die Integration neuer Ausgabekomponenten, was am Beispiel der Modellierung der Home-Entertainment-Domäne demonstriert wurde.

Das definierte Modell für Ausgabedienste erlaubt die bisher umfangreichste und ausdrucksstärkste Modellierung von Ausgabekomponenten in ubiquitären Informationssystemen. Wie in Kapitel 5.5.2 erläutert ermöglicht das Modell eine Verallgemeinerung von Operatoren in der Präsentationsplanung. Im Beispielszenario wurde die Anzahl der Planungsoperatoren um den Faktor vier reduziert. Dadurch lassen sich flexiblere Planungsstrategien formulieren, mit denen sich ausfallsichere Systeme realisieren lassen. Außerdem ermöglichen die auf der Basis des Modells definierten Operatoren die Ad-hoc-Integration einer breiten Menge von Ausgabekomponenten.

### 2. Technische Beiträge:

Der konzeptuell definierte Planungsansatz und das Dienst-, Geräte- und Ressourcenmodell für Ein-/Ausgabekomponenten wurden als Teil von zwei ubiquitären Informationssystemen umgesetzt. Im Embassi-System wurde das Modell für die Koordination einer grafischen Benutzeroberfläche mit einer Sprachausgabe und einem animierten Charakter auf einem Fernseher und einem PDA verwendet.

Innerhalb des Dynamite-Projekts wurde das Modell weiter entwickelt und der definierte Präsentationsplanungsansatz als Teil eines Home-Entertainment-Systems umgesetzt. Mittels dieses Ansatzes konnte das System einen PDA, einen Fernseher und einen elektronischen Bilderrahmen als Ausgabegeräte koppeln und als Teil eines Filminformationssystems einsetzen.

Das System realisiert erstmals eine automatisierte Generierung von Präsentationen mit gekoppelten Ausgabegeräten. Bisherige Systeme zur automatisierten Generierung von Präsentationen sind nicht in der Lage, Präsentationen auf mehrere Ausgabegeräte zu verteilen. Systeme, die gekoppelte Ausgabegeräte unterstützen, verwenden entweder feste Layouts oder sie unterstützen keine multi-modale Ausgabe. Kein bisher existierendes System unterstützt Selbstbeschreibungen von Ausgabekomponenten, die die einzelnen Modalitäten einer Ausgabekomponente beschreiben.

### 3. Empirische Beiträge: Fragebogenstudie

Die vom Home-Entertainment-System generierten Präsentationen auf gekoppelten Ausgabegeräten wurden in einer Fragebogenstudie mit 56 Teilnehmern bzgl. ihrer Akzeptanz untersucht. Dabei wurde das Gesamtsystem als „recht gut“ und die Kopplung von Ausgabegeräten als „recht nützlich“ empfunden. Weiterhin zog die Mehrheit der Teilnehmer das System bei der Ausgabe von Filminformationen einem einzelnen Ausgabegerät signifikant vor.

Es wurde beobachtet, dass sich die Bewertungen des Systems zwischen Teilnehmern, die jünger als 20 sind und Teilnehmern, die älter als 20 sind, signifikant unterscheiden. Ältere Teilnehmer haben das System dabei signifikant schlechter beurteilt. Dies wurde dadurch erklärt, dass ältere Teilnehmer traditionell gewohnt sind, Ausgaben auf einem einzelnen Gerät zu betrachten. Weiterhin wurde das System durch Teilnehmer mit wenig IT-Kenntnissen besser bewertet als durch IT-Fachleute. Die Studie hat ebenfalls gezeigt, dass die vom Home-Entertainment-System generierten Präsentationen mit gekoppelten Ausgabegeräten von den Benutzern gut akzeptiert werden.

Bisherige Akzeptanzstudien zu gekoppelten Ausgabegeräten waren auf Bildschirme als Ausgabegeräte beschränkt. Untersuchungen der Kopplung in Abhängigkeit von Sprachausgabe und animierter Charakter fehlten bisher. Systematische Experimente zur Kopplung von PDAs mit anderen Ausgabegeräten existierten bislang ebenfalls nicht. Diese Zusammenhänge wurden durch die Studie erstmals untersucht.

### 4. Empirische Beiträge: Laborstudie

Neben der Akzeptanz des Home-Entertainment-Systems wurde die Effizienz des Systems in einer Laborstudie mit 36 Teilnehmern untersucht. Zur Untersuchung der Effizienz des Systems wurden die Filminformationen durch Nachrichtenbeiträge ausgetauscht und ein Erinnerungsexperiment durchgeführt. Weiterhin wurden die Zeiten gemessen, die die Teilnehmer brauchten, um sich einen Nachrichtenbeitrag mit dem System anzusehen. Dabei wurden drei Kopplungen von Ausgabegeräten (Fernseher-Bilderrahmen, PDA-Fernseher, PDA-Fernseher-Bilderrahmen) untersucht und mit einem einzelnen Fernseher verglichen.

Es zeigte sich, dass Kopplungen mit dem elektronischen Bilderrahmen zu einer signifikanten Reduktion der Interaktionszeiten führten, ohne dass sich die Erinnerungsleistung der Teilnehmer signifikant verschlechterte. Wie in der Fragebogenstudie haben die Teilnehmer ebenfalls ei-

ne Kopplung von Ausgabegeräten gegenüber einem einzelnen Ausgabegerät signifikant bevorzugt. Die vermuteten Effekte unterschiedlicher Modalitätenkombinationen auf die Erinnerungseffizienz konnten jedoch nicht beobachtet werden. Ein möglicher Grund für das Fehlen dieser Effekte ist die neue Sprachausgabe, die durch die Teilnehmer tendenziell negativ beurteilt wurde.

Durch die Studie wurden erstmals unterschiedliche Modalitätenkombinationen im Zusammenhang mit gekoppelten Ausgabegeräten in Bezug auf Effizienz und Akzeptanz untersucht. Bisherige Untersuchungen zu gekoppelten Ausgabegeräten waren entweder auf Multi-Monitor Systeme beschränkt oder hatten keine unterschiedlichen Kombinationen von Ausgabemodalitäten untersucht.

### 9.2 Bearbeitung der Forschungsfragen

In Abschnitt 1.2 wurden fünf Forschungsfragen gestellt, die durch die Arbeit beantwortet werden sollten. In diesem Kapitel werden die Ergebnisse in Bezug auf diese Fragen diskutiert. Die Arbeit führte zu den folgenden Resultaten:

#### 1. **Welche Eigenschaften von Ausgabegeräten müssen modelliert werden, um sie automatisch in ubiquitäre Informationssysteme einbinden zu können?**

Diese Fragestellung wurde in Kapitel 5 bearbeitet. In Abschnitt 5.4.3 wurde ein Modell definiert, das die Eigenschaften von Ausgabediensten, -geräten und -ressourcen, die unabhängig von einem konkreten Anwendungsszenario sind, erfasst. Eine grundsätzliche Ressourcenbeschränkung von visuellen Ausgabegeräten, die in jeder Domäne auftritt, ist z. B. die Auflösung des Bildschirms. Es wurde motiviert, dass dieses Modell für ein konkretes Anwendungsszenario erweitert und spezifiziert werden muss. Dies wurde in Abschnitt 5.4.4 am Beispiel eines Home-Entertainment-Szenarios durchgeführt. In Kapitel 6 wurde das Modell durch die Implementierung in zwei ubiquitären Informationssystemen verifiziert.

#### 2. **Ist ein Planungsansatz ebenfalls zur Generierung von Präsentationen auf gekoppelten Ausgabegeräten geeignet?**

Die Implementierung des Präsentationsplanungsansatzes in Kapitel 6 hat den Präsentationsplanungsansatz für Kopplungen von einem PDA, einem Fernseher und einem elektronischen Bilderrahmen verifiziert. Dabei wurden multimediale Präsentationen mit Sprachausgabe und animiertem Charakter in Echtzeit durch das Präsentationsplanungssystem

## 9.2. BEARBEITUNG DER FORSCHUNGSFRAGEN

generiert. Dies erfolgte durch Integration des Modells für Ausgabedienste, -geräte und -ressourcen in den Planungsansatz. Die Implementierung zeigt, dass der Planungsansatz im gewählten Home-Entertainment-Szenario gut skaliert.

### 3. **Können multimediale Präsentationen auf gekoppelten Ausgabegeräten mit wenigen Ressourcen effizient umgesetzt werden?**

Eine Teilaufgabenstellung des in Kapitel 6 beschriebenen Home-Entertainment-Systems betraf die Realisierung von SMIL-Präsentationen auf einem iPaq hx2750 PDA und einem Loewe Aconda 9272 Fernseher mit OnlinePlus-Zusatzmodul. Die Herausforderungen beim PDA betrafen insbesondere die geringe Auflösung von  $240 \times 320$ , die beschränkte Netzwerkbandbreite und die geringe Prozessorleistung. Eine Herausforderung beim Fernseher stellte insbesondere die geringe Prozessorleistung dar.

Auf dem PDA wurde eine SMIL-Wiedergabekomponente in Java implementiert. Dieselbe Implementierung wurde auf den Fernseher portiert. Dabei wurde aufgrund der beschränkten Rechenkapazität Java Micro Edition eingesetzt. Die Qualität der Medienobjekte wurde weiterhin an die Prozessorleistung und Bandbreite der Geräte angepasst, was ein schnelleres Laden und einen geringeren Speicherverbrauch beim Darstellen einer SMIL-Präsentation auf dem PDA zu Folge hatte. Die Integration von Constraints zur Bildgröße in die Präsentationsplanung ermöglichte weiterhin eine Anpassung des Layouts an geringe Auflösungen. Das realisierte System demonstriert, dass der gewählte Ansatz sehr gut auf ressourcenbeschränkten Ausgabegeräten skaliert.

In der in Kapitel 8 beschriebenen Evaluierung traten im Laufe von einer Stunde Interaktionszeit durchschnittlich ein bis zwei Systemfehler auf, die insbesondere durch eine zu hohe Speicherauslastung des Fernsehers begründet waren. Es ist jedoch wahrscheinlich, dass sich diese Fehler durch eine nähere Analyse der Speicherverwaltung der Java-Software beheben lassen.

### 4. **Werden von Benutzern des Home-Entertainment-Systems Präsentationen auf gekoppelten Ausgabegeräten gegenüber konventionellen Einzelgeräteausgaben bevorzugt?**

Die zwei Studien, die die Akzeptanz von gekoppelten Ausgabegeräten untersucht haben, haben in diesem Punkt unterschiedliche Ergebnisse geliefert. Die in Kapitel 7 beschriebene Fragebogenstudie ergab, dass die Kombination PDA-Fernseher ebenso gut beurteilt wurde wie

ein einzelner Fernseher. Weiterhin hatten die Teilnehmer der Studie das Home-Entertainment-System gegenüber einem einzelnen PC signifikant vorgezogen. Dabei wurden die gekoppelten Ausgabegeräte durch Teilnehmer, die jünger als 20 waren, besser akzeptiert. Insgesamt wurde das System mit „recht gut“ bewertet.

Bei der in Kapitel 8 beschriebenen Laborstudie wurde durch die Teilnehmer eine Gerätekombination (PDA-Fernseher-Bilderrahmen) gegenüber einem einzelnen Fernseher signifikant vorgezogen. Die Aussage, dass die Darstellung auf der Gerätekombination ansprechend war, wurde durch die Teilnehmer bejaht („stimmt“). Insgesamt wurde das Home-Entertainment-System jedoch schlechter als in der Fragebogenstudie bewertet. Die Frage, ob die Teilnehmer das System einem einzelnen Fernseher vorziehen würden, wurde verneint.

Ein möglicher Grund für die Unterschiede zwischen beiden Studien könnte sein, dass den Teilnehmern in dieser Studie Nachrichtenbeiträge und keine Filminformationen präsentiert wurden. Dies könnte dazu geführt haben, dass den Teilnehmern die Vorteile der Kopplung von Ausgabegeräten nicht bewusst geworden sind und die traditionellen Einzelgeräteausgaben vorgezogen haben.

Beide Studien zeigen, dass die Applikation eine zentrale Rolle bei der Akzeptanz von gekoppelten Ausgabegeräten spielt. Im Bereich von Filminformationssystemen scheinen gekoppelte Ausgabegeräte dabei ein großes Potential aufzuweisen.

### **5. Arbeiten Benutzer mit gekoppelten Ausgabegeräten effizienter als mit Ausgaben auf einem einzelnen Gerät?**

Diese Fragestellung wurde in Kapitel 8 untersucht. Zunächst wurde die Effizienz einer Präsentation mit ihrer Einprägsamkeit assoziiert. Dazu wurde in der Studie abgefragt, an wie viele Informationen sich Benutzer bei Präsentationen mit gekoppelten Ausgabegeräten im Vergleich zu einem einzelnen Ausgabegerät erinnern. Weiterhin wurde die Zeit gemessen, die die Teilnehmer zum Betrachten einer Präsentation benötigt haben.

Dabei zeigte sich, dass in der Fernseher-Bilderrahmen-Kombination und in der PDA-Fernseher-Bilderrahmen-Kombination signifikant mehr Bilder erinnert wurden als in den Kombinationen ohne Bilderrahmen. Weiterhin sanken in beiden Gruppen die Interaktionszeiten signifikant im Vergleich zu den Kombinationen ohne Bilderrahmen. Dies spiegelt die Ergebnisse existierender Forschungsarbeiten zu gekoppelten Monitoren

### 9.3. AUSBLICK

wider und zeigt, dass Benutzer mit gekoppelten Ausgabegeräten effizienter als mit einzelnen Ausgabegeräten interagieren können. Die beobachteten Effekte waren jedoch nicht so groß wie erwartet. Die Steigerung der Erinnerungsleistung konnte nur bzgl. der Bewertungen der erinnerten Bilder beobachtet werden. Bei der Summe der Bewertungen von Nachrichtentitel, Textinhalt und Bildern konnte kein analoger Effekt beobachtet werden. In Bezug auf Bilderinnerungen kann die Frage daher positiv beantwortet werden. Zur grundsätzlichen Beantwortung der Frage sind jedoch weitere Studien erforderlich.

#### 9.3 Ausblick

Im Verlauf der Arbeit ergaben sich Ansätze für mögliche Folgearbeiten. Aus der Integration des Modell für Ausgabedienste, -geräte und -ressourcen in ein ubiquitäres Informationssystem ergeben sich zahlreiche Anknüpfungspunkte für weitere Forschungsarbeiten wie bereits in Kapitel 5.6 illustriert wurde. Hierzu bietet sich insbesondere der Dialog-Manager an. Mögliche Erweiterungen stellen die Integration des Modells in das Domänenwissen des Dialog-Managers sowie die Verwendung einer Ausgabehistorie dar. Die Integration des Modells in das Domänenwissen des Dialog-Managers ermöglicht dem System, über Ausgabegeräte einen Diskurs führen zu können. Die Ausgabehistorie ermöglicht es dem Dialog-Manager, auf eine Repräsentation der zuletzt durch das System umgesetzten Ausgabe zuzugreifen. Auf diese Weise ist es möglich, dass das System Fragen über Teile der aktuellen Ausgabe beantworten kann („Welcher Schauspieler wird auf dem Bild auf dem Fernseher dargestellt?“).

Der Präsentationsplaner bietet weitere Ansätze für mögliche Erweiterungen. Falls Ausgabegeräte oder -komponenten aus dem System entfernt werden, so wird dies bei der in dieser Arbeit realisierten Software durch eine vollständige Neuplanung berücksichtigt, die der nächsten Systemausgabe vorausgeht. Werden Ausgabekomponenten jedoch während der Planung einer multi-modalen Ausgabe entfernt, so erfolgt ein Fehler bei der Ausführung des Plans. Dies liegt darin begründet, dass der Planer noch die Existenz der entfernten Komponente voraussetzt.

Eine Erweiterung des Planers stellt eine inkrementelle Planung dar, die bei einer Änderung in der Menge der Ausgabekomponenten zur Planungszeit nur die Hypothesen neu plant, die durch die Änderung in der Menge der Ausgabekomponenten nicht verifiziert werden können (z. B. dass eine nicht mehr vorhandene Sprachausgabekomponente einen Text akustisch ausgibt).

Eine aktuelle Entwicklung im Bereich von animierten Charakteren ist die Vereinheitlichung der Verhaltensweisen von animierten Charakteren in der

Sprache BML (Behaviour Markup Language) [Kopp et al., 2006]. Durch die in Abschnitt 5.4.4 diskutierte Erweiterung des Modells für Ausgabedienste, -geräte und -ressourcen durch BML könnte die Ausdruckskraft des Modells in Bezug auf animierte Charaktere verbessert werden. Für den erfolgreichen Einsatz dieser Modellierung ist es weiterhin erforderlich, dass die verwendeten animierten Charaktere ihre Verhaltensweisen mittels BML beschreiben.

Im hier realisierten Home-Entertainment-System wurden nur Interaktionen mit einem Benutzer unterstützt. Typischerweise interagieren in einer Home-Entertainment-Umgebung jedoch mehrere Nutzer gleichzeitig mit dem System. In [Kruppa, 2006] findet sich ein System, das die PDAs mehrerer Benutzer mit einem statischen Wandbildschirm in einem Museumskontext koppelt. Hierbei ergibt sich insbesondere die Fragestellung, welche Informationen auf einem privaten Bildschirm und welche Informationen auf einem allgemein sichtbaren Bildschirm darzustellen sind. Ähnliche Präsentationen wie im System von Kruppa könnten auch im Home-Entertainment-System realisiert werden. Eine damit verbundene Studie könnte untersuchen, ob bei mehreren Benutzern die Akzeptanz des Systems steigt, wenn es von mehreren Benutzern mittels geteilter Präsentationen gleichzeitig verwendet werden kann.

Wie in Abschnitt 5.1 illustriert wurde, können Präsentationen mit gekoppelten Ausgabegeräten an den Ort des Benutzers angepasst werden. Dabei können die Ausgaben z. B. an die Entfernung des Benutzers vom Ausgabegerät angepasst werden. So könnte ein größerer Font verwendet werden, falls der Benutzer weiter vom Bildschirm entfernt steht. Bildschirme die vom Blickwinkel des Benutzers aus verdeckt sind, würden nicht verwendet.

Neben den in dieser Arbeit verwendeten Sprach- und Stifteingaben sind bei gekoppelten Ausgabegeräten weitere Interaktionsparadigmen möglich. So diskutiert [Kruppa, 2004] Benutzerschnittstellen für Kopplungen von einem PDA mit einem Wandbildschirm. So könnte z. B. ein PDA als eine Lupe eingesetzt werden, um Teile des Wandbildschirms hervorzuheben. Weiterhin könnte das Home-Entertainment-System durch die Möglichkeit, die Ausgaben auf den Ausgabegeräten direkt manipulieren zu können, erweitert werden. In [Rekimoto, 2000] finden sich mehrere Ansätze, wie durch direkte Manipulation Daten zwischen mobilen und stationären Ausgabegeräten ausgetauscht werden können. Mittels dieser Interaktionstechniken könnten Benutzer im Home-Entertainment-System z. B. Bilder von einem Bildschirm auf einen anderen Bildschirm verschieben.

Zur Verifikation der in der Fragebogenstudie und der Laborstudie gewonnenen Ergebnisse sind Folgestudien notwendig, die die beobachteten Effekte replizieren. Zu den durch die Benutzer vorgeschlagene Verbesserungen, die die

### 9.3. AUSBLICK

Akzeptanz des Home-Entertainment-Systems weiter erhöhen könnten, zählen ein animierter Charakter mit mehr Ausdruckskraft sowie eine Sprachausgabe mit einem natürlicheren Sprachfluss. Weiterhin könnten neben statischen Bildern auch Videos verwendet werden. Die Kommentare der Teilnehmer der Studien lassen vermuten, dass diese Erweiterungen einen positiven Effekt auf die Akzeptanz des Systems haben könnten.



## A Implementierung

In diesem Kapitel werden Beispiele für die Inhalte gegeben, die vom in Abschnitt 6.2 beschriebenen System präsentiert und auf mehrere Geräte verteilt werden (Anhang A.1). Die Inhalte werden von der Filmdatenbankapplikation in Abbildung 6.11 verwaltet. Weiterhin wird ein Überblick über die Architektur der Java-Implementierung des Dienst-, Geräte- und Ressourcenmodells für Ein-/Ausgabekomponenten gegeben (Anhang A.2). Anschließend werden zu dem Beispiel in Abbildung 6.19 ein vollständiger Startzustand  $s_0$  des Planers sowie die Definitionen der verwendeten Operatoren angegeben (Anhang A.3). Abschließend werden Beispiele für Präsentationen gegeben, die durch das in Abschnitt 6.2 beschriebene System generiert werden (Anhang A.4).

### A.1 Präzentierte Inhalte

Dieser Abschnitt enthält vier Beispiele für die durch das Home-Entertainment-System präsentierten Inhalte, die durch die Komponente  $P_1$  in Abbildung 6.11 in einer Filmdatenbank verwaltet werden. Die Bilder wurden den Vorschaubildern der Filme entnommen, die urheberrechtsfrei sind und im Internet Archive verwaltet werden<sup>1</sup>. Die textuellen Zusammenfassungen der vier Filme wurden eigenhändig erstellt.

„African Queen“

„Der raubeinige Kapitän Charlie (Humphrey Bogart) verschifft während des ersten Weltkrieges Versorgungsgüter nach Deutsch-Ostafrika. Zusammen mit der Missionsschwester Rosie (Katherine Hepburn) muss er sich durch die feindlichen Linien schlagen. Packende Abenteuerkomödie mit großartigen Wortgefechten. Oscar für Humphrey Bogart.“



---

<sup>1</sup>Internet Archive, <http://www.archive.org>.



## A.2. DIENST-, GERÄTE- UND RESSOURCENMODELL

„Anastasia“

„Der im Pariser Exil lebende General Bounine (Yul Brunner) will 10 Millionen Pfund erbeuten, indem er ein labiles Mädchen (Ingrid Bergman) als tot geglaubte russische Thronfolgerin Anastasia ausgibt. Während er sie vorbereitet, entsteht in ihm der Verdacht, dass sie tatsächlich Anastasia ist. Filmklassiker. Zweiter Oscar für Ingrid Bergman.“



### A.2 Dienst-, Geräte- und Ressourcenmodell

In diesem Abschnitt wird ein Überblick über die Java-Implementierung des Dienst-, Geräte- und Ressourcenmodells für Ein-/Ausgabekomponenten im Home-Entertainment-System gegeben, das in Abschnitt 6.2 beschrieben wird. Die Implementierung wird in den Abbildungen A.1 und A.2 in Form von UML-Klassendiagrammen dargestellt.

Die Klasse `RenderComponent` implementiert die Definitionen der multi-modalen Typen aus Tabelle 5.3 und der Unimodalitäten aus Tabelle 5.8. Ein Objekt vom Typ `RenderComponent` umfasst einen Bezeichner (`agentName`), eine Liste von unterstützten Inhaltstypen (`outputContentTypes`), sowie eine Liste von Ausgabegeräten (`outputDeviceList`). Ausgabegeräte sind Instanzen der Klasse `OutputDevice` und werden einem übergeordneten Gerät zugeordnet (z. B. einem Fernseher) und einem Ort. Weiterhin besitzen sie die Eigenschaften `Auflösung` und `Medium` (Klasse `OutputDevicePropertyItem`, vgl. Tabelle 5.8). Die Variable `outputContentTypes` modelliert den multi-modalen Parameter *media-format* und enthält eine Liste von Instanzen der Klasse `OutputContentType`. Die Liste der Ausgabemedien (`outputMediaList`) enthält die auditoriellen und visuellen Ausgabemedien von Komponenten vom Typ *agent* (vgl. Tabelle 5.6). Elemente dieser Liste sind Instanzen der Klasse `OutputMedium`. In der Variable `renderComponentType` wird der multi-modale Typ (*agent* oder *player*) der Ausgabekomponente verwaltet.

Die Status der Ausgabekomponenten werden in der Variable `rcStatus` vom Typ `RenderComponentStatus` verwaltet. Bei `rcStatus` wird nicht zwischen multi-modalen und uni-modalen Status unterschieden. Von visuellen Ausgabekomponenten wird die Höhe (`height`), die Breite (`width`), der Ort (`location`) und die Sichtbarkeit (`visibility`) des sie umgebenden Fensters modelliert. Bei dynamischen Ausgabekomponenten wird modelliert, ob zur Zeit eine Ausgabe erfolgt. Dies stellt die Implementierung der Status *has-size*, *has-screen-location*, *is-visible* und *is-playing* dar. Optional können auch Ausgabekomponenten angegeben werden, mit denen ein Synchronisationsprotokoll durchgeführt werden muss `syncPartnerType`. Die Ausgabeunimodalitäten werden in der Variable `unimodalities` verwaltet, die Elemente vom Typ `OutputUnimodality` enthält. Jede Ausgabeunimodalität wird entsprechend der Definitionen aus [Bernsen, 2001] mittels der Klasse `BernsenOutputCategory` beschrieben (vgl. Tabelle 5.7).

Die Modellierung der Eingabekomponenten in der Klasse `InputComponent` (Abbildung A.2) erfolgt analog zu den Ausgabekomponenten. Wie in Abschnitt 5.4.5 erläutert wurde, werden Eingabekomponenten lediglich durch eine Eingabeunimodalität modelliert.

### A.3 Startzustand und Operatoren des Ckuckuk-Planers

In Abbildung A.3 findet sich ein Beispiel für einen Startzustand  $s_0$ , der von dem Ckuckuk-Planer [André et al., 2000] verwendet wird, wie in Abschnitt 6.2.5 beschrieben wurde. Im Beispielzustand sind die Sprachausgabekomponente (vgl. Abbildung 6.14) sowie die Charakterkomponente (vgl. Abbildung 6.15) mit dem System verbunden. Abbildungen A.4 bis A.10 zeigen die Operatoren, die in dem Beispiel aus Abbildung 6.19 zum Einsatz kommen. In diesem Beispiel werden der PDA, der Fernseher und der elektronische Bilderrahmen gekoppelt.

### A.3. STARTZUSTAND UND OPERATOREN DES CKUCKUK-PLANERS

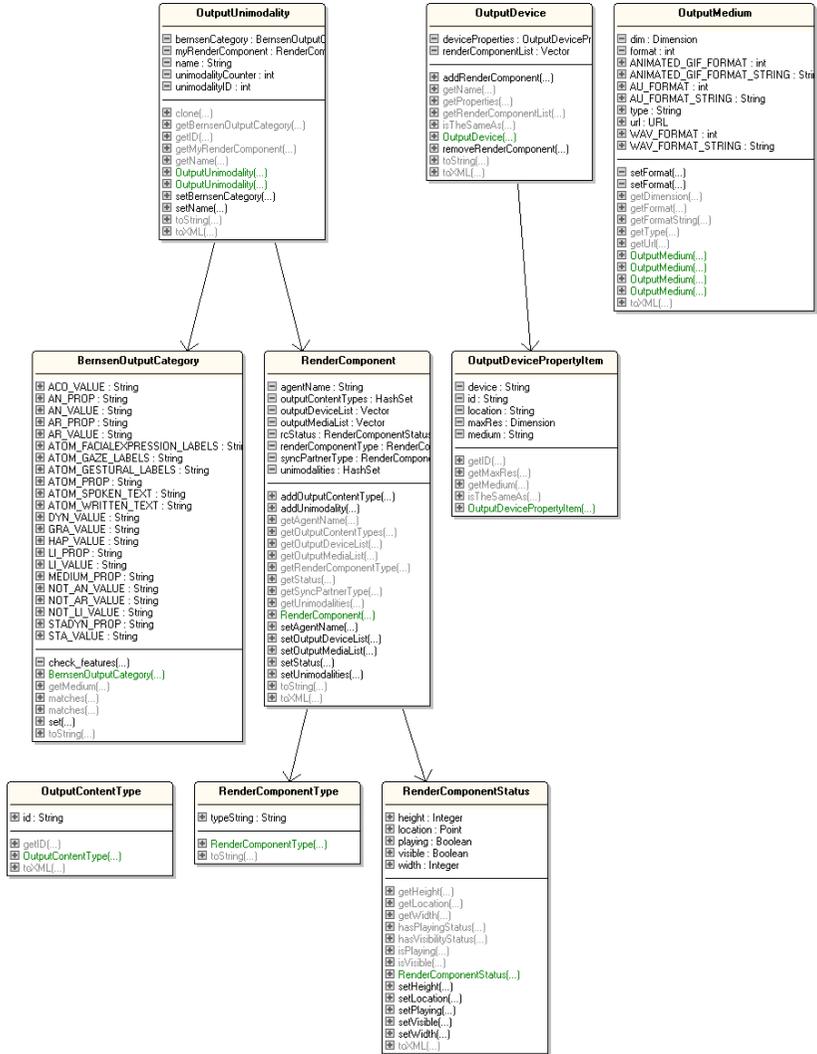


Abbildung A.1: UML-Diagramm Implementierung Ausgabekomponentenmodell.

## ANHANG A. IMPLEMENTIERUNG

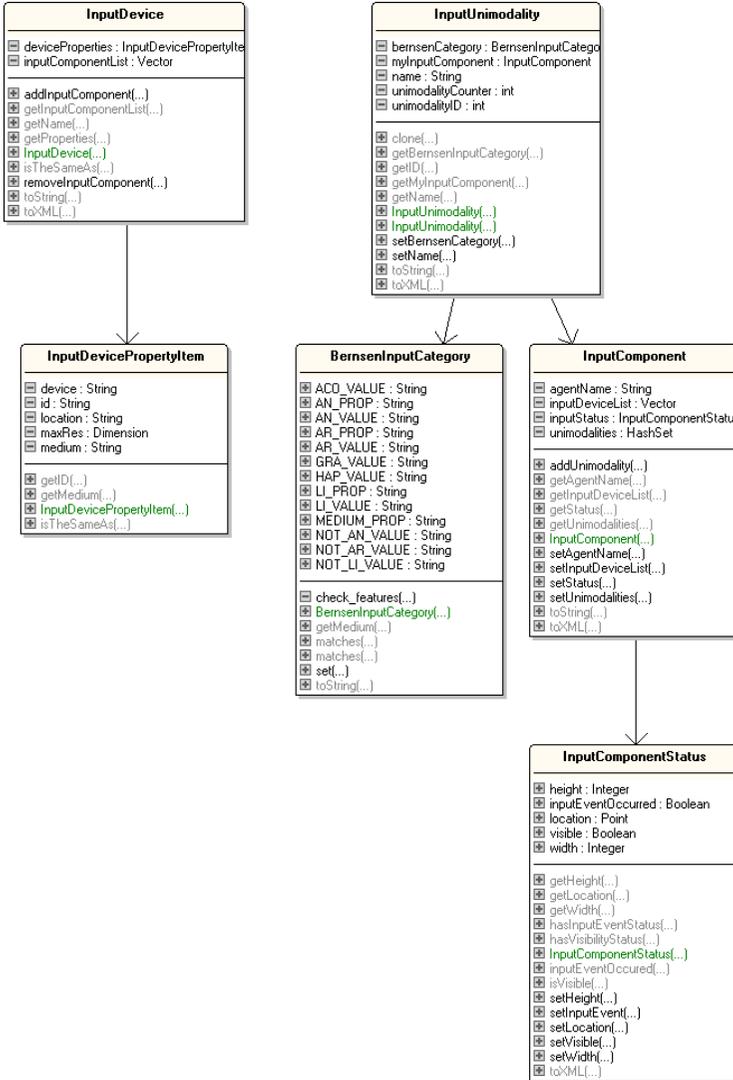


Abbildung A.2: UML-Diagramm Implementierung Eingabekomponentenmodell.

### A.3. STARTZUSTAND UND OPERATOREN DES CKUCKUK-PLANERS

```
((BELP (
  (pres-text "Ich habe Informationen zu Kinofilmen.")
  (pres-title ""))
  (output-preference pref-type layout pref-value automatic)
))

(BELP (
  (render-component rc-id 1147693060392)
  (render-component-type rc-id 1147693060392 rc-type agent)
  (render-component-status rc-id 1147693060392 rc-status stopped)
  (output-unimodality uni-id 1 rc-id 1147693060392 li-value li an-value -an ar-
  value -ar stady-value dyn medium-value aud sub-category spoken-text)
  (processes-output-content rc-id 1147693060392 oc-type dynamite)
  (output-medium rc-id 1147693060392 om-format wav om-id 0 om-url
  "http://11.0.0.1/dynamite/tmp/inText2.wav")
  (output-medium rc-id 1147693060392 om-format au om-id 1 om-url
  "http://11.0.0.1/dynamite/tmp/inText2.au")

  (render-component rc-id 1147693064501)
  (render-component-type rc-id 1147693064501 rc-type agent)
  (output-unimodality uni-id 2 rc-id 1147693064501 li-value -li an-value an ar-
  value -ar stady-value dyn medium-value vis sub-category facial-expression)
  (output-unimodality uni-id 3 rc-id 1147693064501 li-value -li an-value an ar-
  value -ar stady-value dyn medium-value vis sub-category gaze)
  (output-unimodality uni-id 4 rc-id 1147693064501 li-value li an-value an ar-
  value -ar stady-value dyn medium-value vis sub-category spoken-text)
  (output-unimodality uni-id 5 rc-id 1147693064501 li-value -li an-value an ar-
  value -ar stady-value dyn medium-value vis sub-category gestures)
  (output-unimodality uni-id 6 rc-id 1147693064501 li-value -li an-value an ar-
  value -ar stady-value dyn medium-value vis sub-category body)
  (output-unimodality uni-id 7 rc-id 1147693064501 li-value -li an-value an ar-
  value -ar stady-value dyn medium-value vis sub-category head)
  (output-unimodality uni-id 8 rc-id 1147693064501 li-value -li an-value an ar-
  value -ar stady-value dyn medium-value vis sub-category torso)
  (processes-output-content rc-id 1147693064501 oc-type dynamite)
  (output-medium rc-id 1147693064501 om-format animated-gif om-id 0 om-url
  "http://11.0.0.1/dynamite/tmp/SpeakContinued-640-480.gif" om-type speak-
  continued pres-width 640 pres-height 480)
  (output-medium rc-id 1147693064501 om-format animated-gif om-id 1 om-url
  "http://11.0.0.1/dynamite/tmp/RestPoseContinued-640-480.gif" om-type restpose-
  continued pres-width 640 pres-height 480)
  (output-medium rc-id 1147693064501 om-format animated-gif om-id 2 om-url
  "http://11.0.0.1/dynamite/tmp/background-640-480.gif" om-type background pres-
  width 640 pres-height 480)
  (output-medium rc-id 1147693064501 om-format animated-gif om-id 3 om-url
  "http://11.0.0.1/dynamite/tmp/SpeakContinued-240-320.gif" om-type speak-
  continued pres-width 240 pres-height 320)
  (output-medium rc-id 1147693064501 om-format animated-gif om-id 4 om-url
  "http://11.0.0.1/dynamite/tmp/RestPoseContinued-240-320.gif" om-type restpose-
  continued pres-width 240 pres-height 320)
  (output-medium rc-id 1147693064501 om-format animated-gif om-id 5 om-url
  "http://11.0.0.1/dynamite/tmp/background-240-320.gif" om-type background pres-
  width 240 pres-height 320)
  ))
)
```

Abbildung A.3: Beispiel für Startzustand  $s_0$  des Ckuckuk-Planers.

## ANHANG A. IMPLEMENTIERUNG

```

(define-plan-operator
:header (A0 (BuildPresentation pres-id ?pres-id dialog-act ?dialog-act dialog-act-status ?dialog-act-status))
:constraints
  (*and* (
    ;; a digiframe renderer exists
    (BELP(processes-output-content rc-id ?rc-id-3 oc-type df-slide-show))
    ;; a smil renderer exists
    (BELP(processes-output-content rc-id ?rc-id oc-type smil))
    ;; a second smil renderer exists
    (BELP(processes-output-content rc-id ?rc-id-2 oc-type smil))
    ;; and both are not equal
    (*not* ((eval* (equal(?rc-id) (?rc-id-2))))))
    ;; the first smil renderer renders on device ?dev-id
    (BELP(renders-on-device rc-id ?rc-id dev-id ?dev-id))
    ;; with graphics resolution (?dev-max-x, ?dev-max-y)
    (BELP(output-device dev-id ?dev-id dev-medium vis dev-max-x ?dev-max-x dev-max-y ?dev-max-y))
    ;; which is a PDA resolution:
    (*not* ((eval* (> ?dev-max-x 240))))
    (*not* ((eval* (> ?dev-max-y 320))))
    ;; one picture exists
    (BELP (pres-image im-id 0 im-url ?im-url))
    ;; and at least one different picture exists
    (BELP (pres-image im-id 1 im-url ?im-url-2))
    ;; a third picture exists
    (BELP (pres-image im-id 2 im-url ?im-url-3))
    ;; a fourth picture exists
    (BELP (pres-image im-id 3 im-url ?im-url-4))
  ))
:inferiors (
  ;; build single image smil presentation of first picture
  (A1(BuildSmilPresentationSingleImage pres-id ?pres-id dialog-act ?dialog-act dialog-act-status ?dialog-act-status rc-id ?rc-id-2 oc-type smil im-id 1 im-url ?im-url-2))
  ;; build smil presentation of second picture
  (A2(BuildSmilPresentation pres-id ?pres-id dialog-act ?dialog-act dialog-act-status ?dialog-act-status rc-id ?rc-id oc-type smil im-id 0 im-url ?im-url))
  ;; build digiframe presentation of third picture (a single image without speech)
  (A3(BuildDigiframePresentationSingleImage pres-id ?pres-id dialog-act ?dialog-act dialog-act-status ?dialog-act-status rc-id ?rc-id-3 oc-type df-slide-show im-id 2 im-url ?im-url-3))
  ;; send presentation continues
  (A4(SAddRenderingTaskResult
    pres-id      ?pres-id
    content      "presentation_continues"
  ))
)
:start (A1)
:finish (A4)
)

```

Abbildung A.4: Operator `BuildPresentation` des Ckuckuk-Planers.

### A.3. STARTZUSTAND UND OPERATOREN DES CKUCKUK-PLANERS

```

(define-plan-operator
:header (A0 (BuildSmilPresentationSingleImage pres-id ?pres-id dialog-act ?dialog-act
dialog-act-status ?dialog-act-status rc-id ?rc-id oc-type smil im-id ?im-id im-url ?im-url))
:inferiors (
    ;; present all pictures
    (A1 (PresentStaticImage-640-480 im-id ?im-id im-url ?im-url))
    (A2 (SStartMatsDynamite "dynamite-presentation-2.smi" "gen-name"))
    ;; add smil render message
    (A3 (saddsmilmessage
        pres-id      ?pres-id
        rc-id        ?rc-id
        parameter-settings (
            smil-main  "dynamite-presentation-2.smi"
        )
    )
    )
)
)
:temporal (
    ;; "A1 meets A2"
    (A1 (m) A2)
    ;; "A2 meets A3"
    (A2 (m) A3)
)
:spatial (
    ;; center the image
    (centerh A1)
    (centerv A1)
    ;; the picture should be resized to height 480
    (480 <= bottom A1 - top A1 <= 480)
    ;; the picture should be resized to width 640
    (640 <= right A1 - left A1 <= 640)
)
:start (A1)
:finish (A3)
)

```

Abbildung A.5: Operator `BuildSmilPresentationSingleImage` des `Ckuckuk-Planers`.

```

(define-plan-operator
:header (A0 (PresentStaticImage-640-480 im-id ?im-id im-url ?im-url))
:inferiors (
    (
        ;; add the smil code for the picture
        (A1 (SAddSmilCodeDynamite (?im-url)))
        ;; delete the image from the remaining presentation task
        (A2 (SDeleteImageFromPresentationTask im-id ?im-id im-url ?im-url))
    )
)
:temporal (
    (
        ;; present picture for at least 1000 seconds
        (1000 <= duration A1)
        ;; A1 meets A2
        (A1 (m) A2)
    )
)
)

```

Abbildung A.6: Operator `PresentStaticImage-640-480` des `Ckuckuk-Planers`.

## ANHANG A. IMPLEMENTIERUNG

```

(define-plan-operator
:header (A0 (BuildSmilPresentation pres-id ?pres-id dialog-act ?dialog-act dialog-act-status ?dialog-act-status rc-id ?rc-id
oc-type smil im-id ?im-id im-url ?im-url))
:constraints (*and* (
    ;; a presentation text exists
    (BELP (pres-text ?pres-text))
    ;; there is an agent render component
    (BELP (render-component-type rc-id ?rc-id-2 rc-type agent))
    ;; that has a speech unimodality
    (BELP (output-unimodality uni-id ?uni-id rc-id ?rc-id-2 li-value li an-value -an ar-value -ar stady-
value dyn medium-value aud sub-category spoken-text))
    ;; which produced a speech file (format left open)
    (BELP (output-medium rc-id ?rc-id-2 om-format ?om-format om-id ?om-id om-url ?om-url))
    ;; there is an agent render component
    (BELP (render-component-type rc-id ?rc-id-3 rc-type agent))
    ;; which produces a speaking animation
    (BELP (output-medium rc-id ?rc-id-3 om-format animated-gif om-id ?om-id-3 om-url ?om-url-3
om-type speak-continued pres-width ?pres-width pres-height ?pres-height))
    ;; which contains a continuing rest pose animation
    (BELP (output-medium rc-id ?rc-id-3 om-format animated-gif om-id ?om-id-4 om-url ?om-url-4
om-type restpose-continued pres-width ?pres-width pres-height ?pres-height))
    ;; which contains a background picture
    (BELP (output-medium rc-id ?rc-id-3 om-format animated-gif om-id ?om-id-5 om-url ?om-url-5
om-type background pres-width ?pres-width pres-height ?pres-height)))
:inferiors (
    ;; present a picture with speech
    (A1 (PresentStaticImageWithSpeechAndCharacter
    ;; the image to be presented
    im-id ?im-id im-url ?im-url
    ;; the audio file
    rc-id ?rc-id-2 om-id ?om-id om-url ?om-url
    ;; the speaking animation
    rc-id ?rc-id-3 om-id ?om-id-3 om-url ?om-url-3 om-type speak-continued
    ;; the restpose-continued animation
    om-id ?om-id-4 om-url ?om-url-4 om-type restpose-continued
    ;; the background image
    om-id ?om-id-5 om-url ?om-url-5 om-type background
    ;; the smil component which should display the presentation
    smil-rc-id ?rc-id))
    (A2 (SStartMatsDynamite "dynamite-presentation.smi" "gen-name"))
    ;; add smil render message
    (A3 (saddsmilmessage
    pres-id ?pres-id rc-id ?rc-id
    parameter-settings (
    smil-main "dynamite-presentation.smi")
    )))
:temporal (
    ;; "A1 meets A2"
    (A1 (m) A2)
    ;; "A2 meets A3"
    (A2 (m) A3))
:spatial (
    (centerh A1)
    (centerv A1))
:start (A1)
:finish (A3))

```

Abbildung A.7: Operator BuildSmilPresentation des Ckuckuk-Planers.

### A.3. STARTZUSTAND UND OPERATOREN DES CKUCKUK-PLANERS

```

(define-plan-operator
:header (A0 (PresentStaticImageWithSpeechAndCharacter
;; the image to be presented
im-id ?im-id im-url ?im-url
;; the speech file to be presented
rc-id ?rc-id om-id ?om-id
om-url ?om-url
;; the speaking animation to be presented
rc-id ?rc-id-2 om-id ?om-id-2
om-url ?om-url-2
om-type speak-continued
;; the restpose-continued image to be presented
om-id ?om-id-3 om-url ?om-url-3
om-type restpose-continued
;; the background image to be presented
om-id ?om-id-4 om-url ?om-url-4
om-type background
;; the smil component which should display the presentation
smil-rc-id ?smil-rc-id))
:constraints (*and* (
;; the presentation format of the images is 240x320
(BELP(output-medium rc-id ?rc-id-2 om-format animated-gif om-id
?om-id-4 om-url ?om-url-4 om-type background pres-width 240 pres-height 320))
;; use au-files on small devices
(BELP(output-medium rc-id ?rc-id om-format au om-id ?om-id om-url
?om-url))
;; the smil renderer renders on device ?dev-id
(BELP(renders-on-device rc-id ?smil-rc-id dev-id ?dev-id))
;; with graphics resolution (?dev-max-x, ?dev-max-y)
(BELP(output-device dev-id ?dev-id dev-medium vis dev-max-x ?dev-
max-x dev-max-y ?dev-max-y))
;; where ?dev-max-x >= 240
(*eval* (>= ?dev-max-x 240))
;; where ?dev-max-y >= 290
(*eval* (>= ?dev-max-y 290))))))
:inferiors (
;; present a picture with speech for resolution 240x290
(A1 (PresentStaticImageWithSpeechAndCharacter-240-290
;; the image to be presented
im-id ?im-id im-url ?im-url
;; the speech file to be presented
rc-id ?rc-id om-id ?om-id
om-url ?om-url
;; the speaking animation to be presented
rc-id ?rc-id-2 om-id ?om-id-2
om-url ?om-url-2
om-type speak-continued
;; the restpose-continued image to be presented
om-id ?om-id-3 om-url ?om-url-3
om-type restpose-continued
;; the background image to be presented
om-id ?om-id-4 om-url ?om-url-4
om-type background)
)))

```

Abbildung A.8: Operator PresentStaticImageWithSpeechAnd-Character des Ckuckuk-Planers.

## ANHANG A. IMPLEMENTIERUNG

```

(define-plan-operator
:header (A0 (PresentStaticImageWithSpeechAndCharacter-240-290
;; the image to be presented
im-id      ?im-id      im-url      ?im-url
;; the speech file to be presented
rc-id      ?rc-id      om-id      ?om-id
om-url      ?om-url
;; the speaking animation to be presented
rc-id      ?rc-id-2    om-id      ?om-id-2
om-url      ?om-url-2  om-type    speak-continued
;; the restpose-continued image to be presented
om-id      ?om-id-3    om-url      ?om-url-3
om-type     restpose-continued
;; the background image to be presented
om-id      ?om-id-4    om-url      ?om-url-4
om-type     background))
:inferiors (
;; add the smil code for the background image
(A1 (SAddSmilCodeDynamite (?om-url-4)))
;; add the smil code for the picture
(A2 (SAddSmilCodeDynamite (?im-url)))
;; add the smil code for the wav file
(A3 (SAddSmilCodeDynamite (?om-url)))
;; add the smil code for the speaking-animation
(A4 (SAddSmilCodeDynamite (?om-url-2)))
;; add the smil code for the restpose-continued animation
(A5 (SAddSmilCodeDynamite (?om-url-3)))
;; delete the image from the remaining presentation task
(A6 (SDeleteImageFromPresentationTask im-id ?im-id im-url ?im-url))
;; delete the text from the remaining presentation task
(A7 (SDeleteTextFromPresentationTask)))
:temporal (
;; the character is speaking for the time the speech file is playing
(A3 (e) A4)
;; afterwards the restpose animation starts
(A4 (m) A5)
;; which lasts for 1000 seconds
(1000 <= duration A5))
:spatial (
;; the image is at the top of the background image
(0 <= top A1 - top A2 <= 0)
;; the background picture is on the very left
(alignleft A1)
;; the background picture is on the top
(aligntop A1)
;; the character is at the bottom of the background picture
(0 <= bottom A4 - bottom A1 <= 0)
(0 <= bottom A5 - bottom A1 <= 0)
;; the picture should be resized to height 160
(160 <= bottom A2 - top A2 <= 160)
;; the picture should be resized to width 240
(240 <= right A2 - left A2 <= 240)
))

```

Abbildung A.9: Operator `PresentStaticImageWithSpeechAndCharacter-240-290` des Ckuckuk-Planers.

## A.4. BEISPIELPRÄSENTATIONEN

```
(define-plan-operator
:header (A0 (BuildDigiframePresentationSingleImage pres-id ?pres-id dialog-act ?dialog-act dialog-act-status ?dialog-act-status rc-id ?rc-id oc-type df-slide-show im-id ?im-id im-url ?im-url))
:inferiors (
  ;; build digiframe script
  (A1 (SAddDigiframeTask
    pres-id      ?pres-id
    rc-id        ?rc-id
    media-1st (
      im-url ?im-url
    )
  ))
  ;; delete the image from the remaining presentation task
  (A2 (SDeleteImageFromPresentationTask im-id ?im-id im-url ?im-url))
)
:start (A1)
:finish (A2)
)
```

Abbildung A.10: Operator BuildDigiframePresentationSingleImage des Ckuckuk-Planers.

### A.4 Beispielpräsentationen

Dieser Abschnitt enthält Beispiele für Präsentationen, die von dem in Abschnitt 6.2 beschriebenen System generiert werden. Auf den Abbildungen wird anstelle des elektronischen Bilderrahmens (vgl. Abbildung 6.8) ein umfunktionierter Toshiba Portégé M200 Tablet PC verwendet, da der elektronische Bilderrahmen aufgrund eines Hardware-Defekts nicht verwendet werden konnte. Die Funktionalität des Bilderrahmens wurde vollständig auf den Tablet PC migriert. In jeder Präsentation wird der Inhalt des Filmes „African Queen“ präsentiert (vgl. Anhang A.1).



Abbildung A.11: Präsentation auf dem PDA mit Charakter (Bild: European Media Laboratory).

## ANHANG A. IMPLEMENTIERUNG

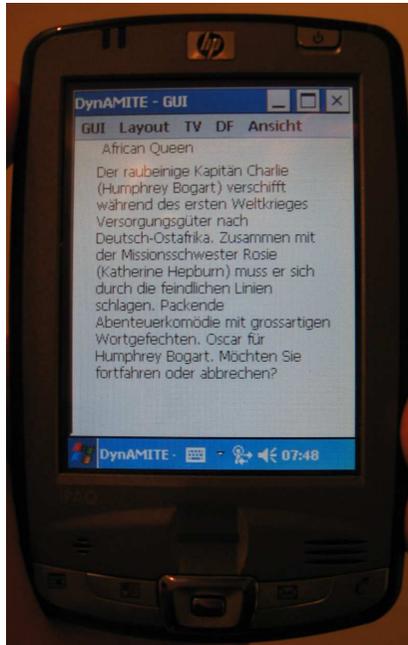


Abbildung A.12: Präsentation auf dem PDA mit visuellem Text.



Abbildung A.13: Präsentation auf dem Fernseher mit Charakter.

#### A.4. BEISPIELPRÄSENTATIONEN

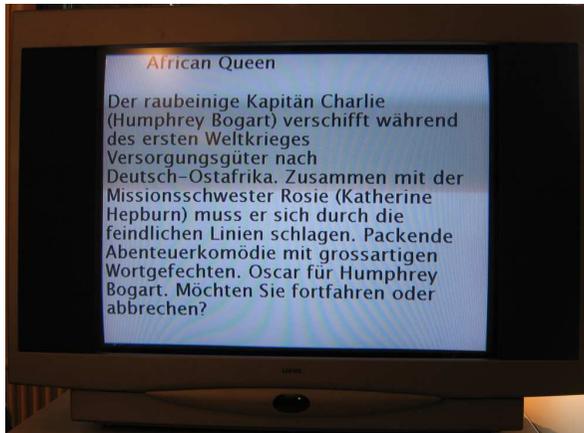


Abbildung A.14: Präsentation auf dem Fernseher mit visuellem Text.



Abbildung A.15: Präsentation auf PDA und Fernseher mit Charakter.

## ANHANG A. IMPLEMENTIERUNG

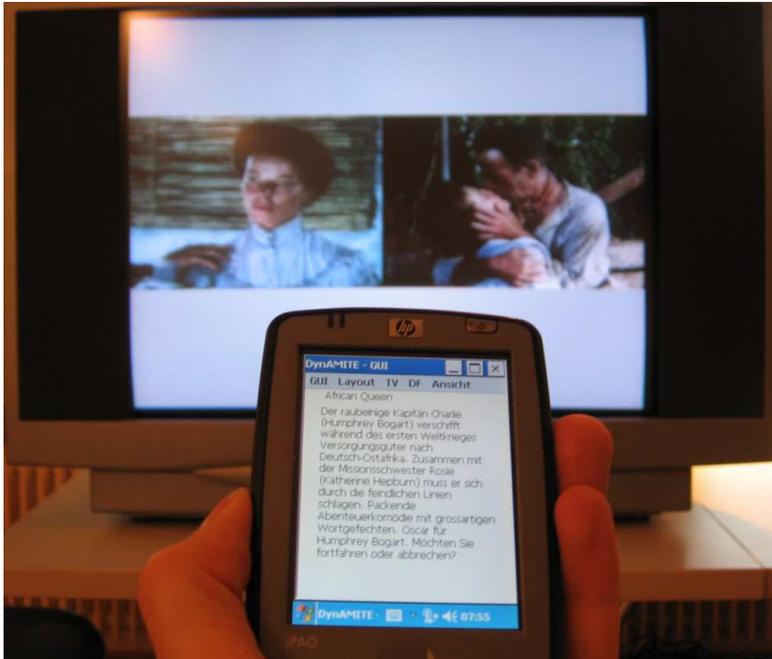


Abbildung A.16: Präsentation auf PDA und Fernseher mit visuellem Text.



Abbildung A.17: Präsentation auf Fernseher und Bilderrahmen mit Charakter.

#### A.4. BEISPIELPRÄSENTATIONEN

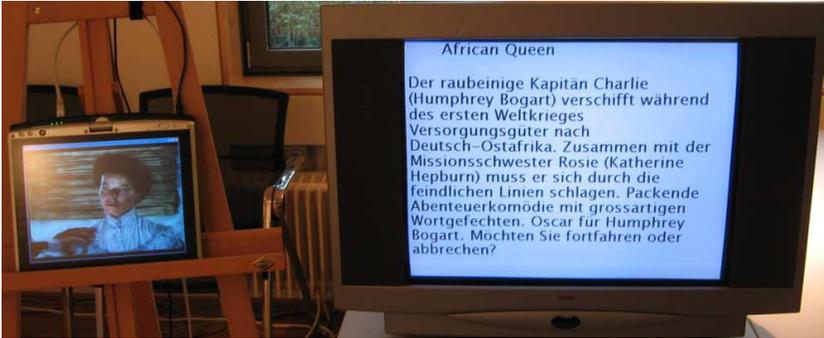


Abbildung A.18: Präsentation auf Fernseher und Bilderrahmen mit visuellem Text.

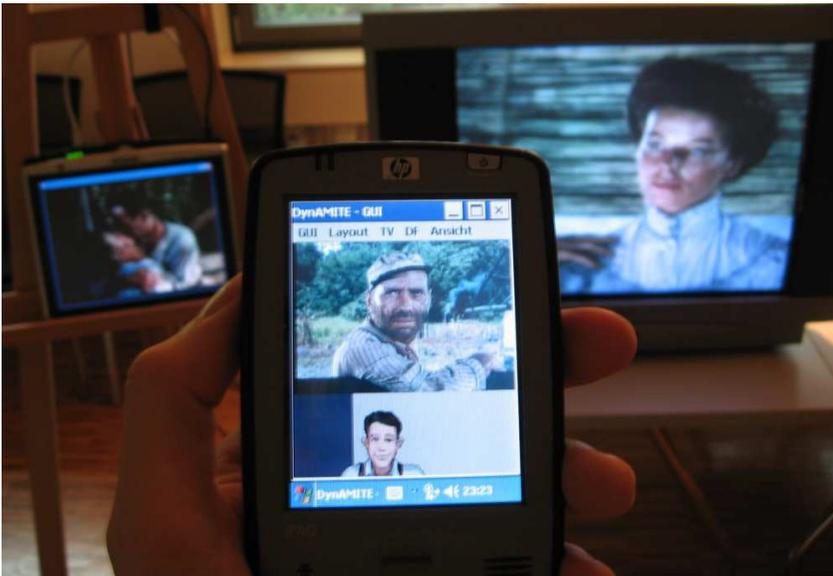


Abbildung A.19: Präsentation auf PDA, Fernseher und Bilderrahmen mit Charakter.

## ANHANG A. IMPLEMENTIERUNG

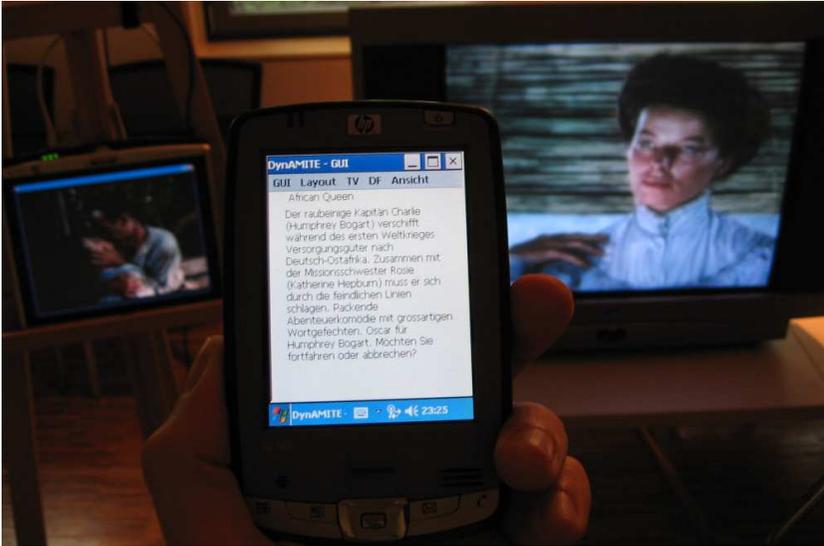


Abbildung A.20: Präsentation auf PDA, Fernseher und Bilderrahmen mit visuellem Text.

## B Materialien Fragebogenstudie

Im Folgenden wird der Inhalt des Fragebogens wiedergegeben, der in der in Kapitel 7 beschriebenen Studie zum Einsatz kam.

### Einleitung

Vielen Dank, dass Sie an unserem Experiment zur Auswertung unseres Heimunterhaltungssystems teilnehmen. Sie helfen uns dadurch, existierende Systeme zu verbessern und neue Systeme zu entwickeln, welche auf die Bedürfnisse der Benutzer zugeschnitten sind. Die im Folgenden von Ihnen erhobenen Daten werden selbstverständlich nur in anonymisierter Form ohne Bezug zu Ihrer Person gespeichert. Es wird durch die Daten keine Möglichkeit bestehen, einen Rückschluss auf Ihre persönlichen Angaben zu ziehen.

Sollten Sie Fragen haben oder Teile der Demonstration noch einmal sehen möchten, so steht Ihnen am Stand ein Film der Demonstration zur Verfügung. Sie können sich ebenfalls jederzeit an einen der Versuchsleiter wenden.

### Allgemeine Angaben

Zunächst möchten wir Sie bitten, einige allgemeine Angaben zu Ihrer Person zu machen.

Bitte geben Sie Ihr Alter an: \_\_\_\_\_ Jahre

Bitte geben Sie Ihr Geschlecht an:

weiblich

männlich

Bitte geben Sie Ihren Beruf an: \_\_\_\_\_

Bitte geben Sie an, ob Sie beruflich oder privat auf der CeBit sind:

beruflich

privat

Bitte geben Sie an, wie viele Stunden pro Woche Sie mit einem der folgenden Geräte arbeiten (0 Stunden pro Woche bedeutet gar nicht):

Computer: \_\_\_\_\_ Stunden pro Woche

Fernseher: \_\_\_\_\_ Stunden pro Woche

Tragbarer Kleinst-Computer (PDA) : \_\_\_\_\_ Stunden pro Woche

Digitaler Bilderrahmen: \_\_\_\_\_ Stunden pro Woche

Bitte geben Sie an, wie viele Kinofilme Sie sich pro Monat im Fernsehen, auf dem Computer oder im Kino ansehen:

\_\_\_\_\_ Kinofilme pro Monat

### Eingabemöglichkeiten

Im Folgenden möchten wir Sie bitten, die Eingabemöglichkeiten des Systems zu beurteilen, das Ihnen gerade demonstriert wurde. Die Eingabemöglichkeiten sind Spracheingabe über das Headset sowie Eingabe per Stift auf dem tragbaren Kleinst-Computer (PDA).

Bitte bewerten Sie, wie komfortabel Sie die Bedienung des Systems per Spracheingabe über das Headset empfanden:

- Sehr komfortabel
- Recht komfortabel
- Weder komfortabel noch kompliziert
- Eher kompliziert
- Sehr kompliziert

Bitte bewerten Sie, wie komfortabel Sie die Bedienung des Systems per Stifteingabe mit dem tragbaren Kleinst-Computer (PDA) empfanden:

- Sehr komfortabel
- Recht komfortabel
- Weder komfortabel noch kompliziert
- Eher kompliziert
- Sehr kompliziert

### Ausgabemöglichkeiten

Im Folgenden möchten wir Sie bitten, die Ausgabemöglichkeiten des Systems zu beurteilen, das Ihnen gerade demonstriert wurde.

Wie gut hat Ihnen (unabhängig vom Gerät) der animierte Helfer gefallen?

- Sehr gut gefallen
- Recht gut gefallen
- Weder gut noch schlecht gefallen
- Eher schlecht gefallen
- Überhaupt nicht gefallen

Wie gut hat Ihnen (unabhängig vom Gerät) die Sprachausgabe gefallen?

- Sehr gut gefallen
- Recht gut gefallen
- Weder gut noch schlecht gefallen
- Eher schlecht gefallen
- Überhaupt nicht gefallen

Wie gut hat Ihnen (unabhängig vom Gerät) die rein grafische Textausgabe gefallen?

- Sehr gut gefallen
- Recht gut gefallen
- Weder gut noch schlecht gefallen
- Eher schlecht gefallen
- Überhaupt nicht gefallen

Wie gut haben Ihnen (unabhängig vom Gerät) die Filmbilder gefallen?

- Sehr gut gefallen
- Recht gut gefallen
- Weder gut noch schlecht gefallen
- Eher schlecht gefallen
- Überhaupt nicht gefallen

Gerätekombinationen

Wir möchten Sie bitten, die unterschiedlichen Kombinationen der drei Geräte der Demonstration (tragbarer Kleinst-Computer, Fernseher oder digitaler Bilderrahmen) bzgl. der Ausgabe von Kinofilm-Informationen zu beurteilen.

Wie gut hat Ihnen die Ausgabe von Kinofilm-Informationen auf dem tragbaren Kleinst-Computer alleine gefallen?

- Sehr gut gefallen
- Recht gut gefallen
- Weder gut noch schlecht gefallen
- Eher schlecht gefallen
- Überhaupt nicht gefallen

Wie gut hat Ihnen die Ausgabe von Kinofilm-Informationen auf dem Fernseher alleine gefallen?

- Sehr gut gefallen
- Recht gut gefallen
- Weder gut noch schlecht gefallen
- Eher schlecht gefallen
- Überhaupt nicht gefallen

Wie gut hat Ihnen die kombinierte Ausgabe von Kinofilm-Informationen auf dem tragbaren Kleinst-Computer und auf dem Fernseher gefallen?

- Sehr gut gefallen
- Recht gut gefallen

## ANHANG B. MATERIALIEN FRAGEBOGENSTUDIE

- Weder gut noch schlecht gefallen
- Eher schlecht gefallen
- Überhaupt nicht gefallen

Wie gut hat Ihnen die kombinierte Ausgabe von Kinofilm-Informationen auf dem Fernseher und auf dem digitalen Bilderrahmen gefallen?

- Sehr gut gefallen
- Recht gut gefallen
- Weder gut noch schlecht gefallen
- Eher schlecht gefallen
- Überhaupt nicht gefallen

Wie gut hat Ihnen die Ausgabe von Kinofilm-Informationen auf dem digitalen Bilderrahmen alleine gefallen?

- Sehr gut gefallen
- Recht gut gefallen
- Weder gut noch schlecht gefallen
- Eher schlecht gefallen
- Überhaupt nicht gefallen

### Gesamtbeurteilung

Abschließend möchten wir Sie bitten, das demonstrierte System als Ganzes zu bewerten.

Wie hat Ihnen die Demonstration insgesamt gefallen?

- Sehr gut gefallen
- Recht gut gefallen
- Weder gut noch schlecht gefallen
- Eher schlecht gefallen
- Überhaupt nicht gefallen

Als wie nützlich empfanden Sie die Kombination mehrerer Geräte bei der Ausgabe von Kinofilm-Informationen?

- Sehr nützlich
- Recht nützlich
- Weder nützlich noch überflüssig
- Eher überflüssig
- Völlig überflüssig

Als wie nützlich empfanden Sie die Wahl zwischen dem Text-Layout und dem animierten Helfer-Layout?

- Sehr nützlich
- Recht nützlich
- Weder nützlich noch überflüssig
- Eher überflüssig
- Völlig überflüssig

Würden Sie bei der Ausgabe von Kinofilm-Informationen Ausgaben auf einem einzelnen Gerät oder Ausgaben auf mehreren Geräten bevorzugen?

- Ich würde Ausgaben auf einem einzelnen Gerät vorziehen
- Ich würde Ausgaben auf mehreren Geräten vorziehen

Würden Sie bei der Suche nach Kinofilm-Informationen das demonstrierte System einem PC mit Internet-Anschluss vorziehen?

- Ja, ich würde es dem PC gegenüber vorziehen
- Nein, ich fände den PC besser

Wie viel Geld wären Sie bereit für das demonstrierte System auszugeben?

- 0 Euro
- 600 Euro
- 1200 Euro
- 1800 Euro
- 2400 Euro

Hier haben Sie noch die Gelegenheit zu allgemeinen Anregungen, Kritik oder Verbesserungswünschen:

---

---

---

---

---

Vielen Dank für Ihre Teilnahme. Falls Sie noch weitere Fragen, Anregungen oder Kritik haben, so sprechen Sie uns bitte an!



## C Materialien Laborstudie

In diesem Kapitel finden sich die in der Laborstudie (Kapitel 8) verwendeten Fragebögen und die Instruktionen zur Verwendung der Gerätekombinationen. Bei den Fragebögen wurden ein Eingangsfragebogen, ein Zwischenfragebogen und ein Abschlussfragebogen verwendet.

### C.1 Eingangsfragebogen

Im Folgenden werden Bilder des Eingangsfragebogens dargestellt, den die Vpn am Anfang der Studie auf einem Laptop ausgefüllt haben.

Vielen Dank, dass Sie an unserem Experiment zur Evaluation eines Nachrichteninformationssystems teilnehmen. Sie helfen uns dadurch, Informationssysteme zu entwickeln, die auf die Bedürfnisse der Benutzer zugeschnitten sind. Weiterhin können wir dadurch auch existierende Informationssysteme verbessern.

Die im Folgenden von Ihnen erhobenen Daten werden selbstverständlich nur in anonymisierter Form ohne Bezug zu Ihrer Person gespeichert. Es wird durch die Daten keine Möglichkeit bestehen, einen Rückschluss auf Ihre persönlichen Angaben oder Leistungen zu ziehen. Wenn Sie mit allen Fragen fertig sind, geht es durch Klick auf den "nächste Seite"-Button weiter.

Bitte geben Sie in das rechte Eingabefeld Ihr Alter ein.

Bitte geben Sie rechts Ihr Geschlecht durch Auswahl des entsprechenden Buttons an.  weiblich  männlich

**Im Folgenden werden Ihnen links mehrere Aussagen präsentiert. Bitte wählen Sie dazu rechts daneben einen zur Aussage passenden Wert durch anklicken mit dem Mauszeiger aus.**

|                                                                                  |                                    |                         |                         |                         |                         |                         |                         |                         |
|----------------------------------------------------------------------------------|------------------------------------|-------------------------|-------------------------|-------------------------|-------------------------|-------------------------|-------------------------|-------------------------|
| Ich verwende einen tragbaren Kleinst-Computer (PDA) an so vielen Tagen die Woche | <input checked="" type="radio"/> 0 | <input type="radio"/> 1 | <input type="radio"/> 2 | <input type="radio"/> 3 | <input type="radio"/> 4 | <input type="radio"/> 5 | <input type="radio"/> 6 | <input type="radio"/> 7 |
| Ich verwende ein Fernsehgerät an so vielen Tagen die Woche                       | <input checked="" type="radio"/> 0 | <input type="radio"/> 1 | <input type="radio"/> 2 | <input type="radio"/> 3 | <input type="radio"/> 4 | <input type="radio"/> 5 | <input type="radio"/> 6 | <input type="radio"/> 7 |
| Ich verwende einen Computer an so vielen Tagen die Woche                         | <input checked="" type="radio"/> 0 | <input type="radio"/> 1 | <input type="radio"/> 2 | <input type="radio"/> 3 | <input type="radio"/> 4 | <input type="radio"/> 5 | <input type="radio"/> 6 | <input type="radio"/> 7 |
| Ich verwende einen elektronischen Bilderrahmen an so vielen Tagen die Woche      | <input checked="" type="radio"/> 0 | <input type="radio"/> 1 | <input type="radio"/> 2 | <input type="radio"/> 3 | <input type="radio"/> 4 | <input type="radio"/> 5 | <input type="radio"/> 6 | <input type="radio"/> 7 |
| Ich lese, schaue oder höre an so vielen Tagen die Woche Nachrichten              | <input checked="" type="radio"/> 0 | <input type="radio"/> 1 | <input type="radio"/> 2 | <input type="radio"/> 3 | <input type="radio"/> 4 | <input type="radio"/> 5 | <input type="radio"/> 6 | <input type="radio"/> 7 |

Abbildung C.1: Eingangsfragebogen, Seite 1: Einleitung und Fragen zu deskriptiven Daten.

Zunächst erfolgte eine kurze textuelle Einleitung in den Versuch. Anschließend wurden den Vpn einige allgemeine Fragen zu ihrer Person gestellt (Abbildung C.1). Anschließend erfolgte eine erste Instruktion der Vpn zum Versuchsablauf (Abbildung C.2).

**Im Folgenden werden Ihnen Nachrichtenbeiträge präsentiert und Sie werden danach aufgefordert, sich an den Titel und den Inhalt des Beitrages zu erinnern.**

**In dieser Untersuchung geht es nicht darum, dass Sie sich an möglichst viele Nachrichtenbeiträge erinnern, sondern dass Sie die Informationen entspannt wahrnehmen (z.B. wie beim Betrachten einer Nachrichtensendung). Versuchen Sie deshalb bitte nicht, die einzelnen Titel und Inhalte auswendig zu lernen. Sollten Sie sich dann nur an eine oder zwei Nachrichtenbeiträge erinnern können, so wäre dies für uns ein realistisches Maß für die Effizienz des Systems. Bitte versuchen Sie nicht, ihre Ergebnisse durch besondere Sorgfalt zu verbessern. Ihre Daten werden auf alle Fälle anonym gespeichert, so dass Sie keine Angst haben müssen, dadurch in ein schlechtes Licht gerückt zu werden. Nochmals vielen Dank für Ihr Interesse, mit dem 'nächste Seite'-Button geht es los!**

Abbildung C.2: Eingangsfragebogen, Seite 2: Instruktionen.

## C.2 Einweisung in die Gerätekombination

Nach dem ersten Teil der Einweisung auf dem Laptop wird die Vp gebeten, zu den Geräten zu wechseln. Hier erfolgt eine weitere Einweisung, die auf drei Bildschirmseiten verteilt ist.

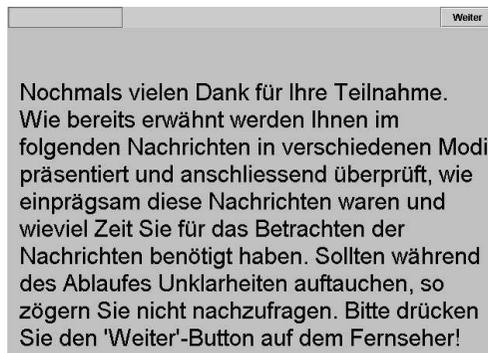


Abbildung C.3: Einleitung auf Fernseher, Seite 1.

### C.3. ZWISCHENFRAGEBOGEN

Abbildungen C.3, C.4 und C.5 zeigen die einleitenden Texte auf dem Fernseher in den Gruppen Tastatur-TV und Tastatur-TV-Bilderrahmen. In den Gruppen PDA-TV und PDA-TV-Bilderrahmen wird derselbe Text auf dem PDA ausgegeben.

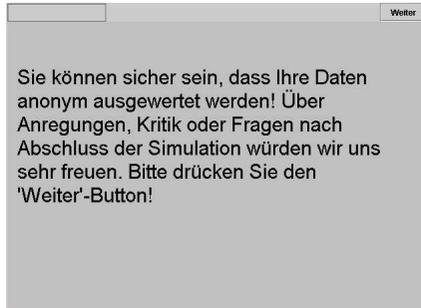


Abbildung C.4: Einleitung auf Fernseher, Seite 2.

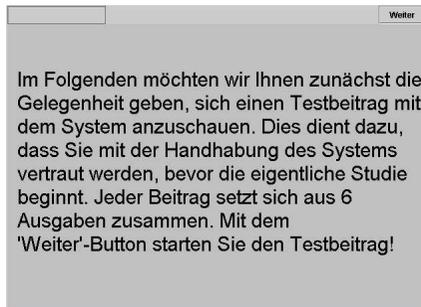


Abbildung C.5: Einleitung auf Fernseher, Seite 3.

### C.3 Zwischenfragebogen

Der Zwischenfragebogen setzte sich aus vier Teilen zusammen. Auf der ersten Seite des Zwischenfragebogens wurden die Vpn gebeten, die gesehene Präsentationsbedingung zu bewerten. Auf der zweiten Seite des Zwischenfragebogens wurden die Vpn gebeten, die Nachrichtentitel anzugeben, an die sie sich noch erinnern können. Anschließend wurden die Vpn gebeten, zu jedem erinnerten Nachrichtentitel anzugeben, an welchen Inhalt der Beiträge sie sich erinnern können. Abschließend wurden die Benutzer gebeten anzugeben, wie häufig sie jedes einzelne Gerät betrachtet haben.

## ANHANG C. MATERIALIEN LABORSTUDIE

**Im Folgenden möchten wir von Ihnen erfahren, wie Sie die Art der Informationsdarbietung fanden. Bitte bewerten Sie dazu die folgenden Aussagen. Klicken Sie dazu bitte mit der Maus auf den Auswahlknopf, der am ehesten Ihrer Meinung entspricht. Wenn Sie alle Fragen beantwortet haben, klicken Sie auf die Taste 'nächste Seite' in der rechten unteren Ecke.**

Die Bilder haben mir gut gefallen.  stimme völlig zu  stimme zu  bin unentschieden  stimme nicht zu  stimme überhaupt nicht zu

Die Sprachausgabe hat mir gut gefallen.  stimme völlig zu  stimme zu  bin unentschieden  stimme nicht zu  stimme überhaupt nicht zu

Der animierte Helfer hat mir gut gefallen.  stimme völlig zu  stimme zu  bin unentschieden  stimme nicht zu  stimme überhaupt nicht zu

Ich musste mich sehr konzentrieren, um den Berichten folgen zu können.  stimme völlig zu  stimme zu  bin unentschieden  stimme nicht zu  stimme überhaupt nicht zu

Ich fand die Darstellung insgesamt ansprechend.  stimme völlig zu  stimme zu  bin unentschieden  stimme nicht zu  stimme überhaupt nicht zu

Im weißen Eingabefeld rechts können Sie Vorschläge machen, was Sie verbessern würden.

**Die folgende Frage bezieht sich nur auf den Inhalt. Versuchen Sie so gut wie möglich von der Art der Darbietung abzusehen.**

Ich fand die gezeigten Nachrichtenbeiträge interessant.  stimme völlig zu  stimme zu  bin unentschieden  stimme nicht zu  stimme überhaupt nicht zu

Abbildung C.6: Zwischenfragebogen, Teil A: Fragen zur Präsentationsbedingung (hier BSC).

**Bitte geben Sie in die folgenden Eingabefelder die Titel der Nachrichtenbeiträge ein (z.B. "Italien wird Weltmeister"), an die Sie sich aus der letzten Präsentationsrunde erinnern. Die Reihenfolge spielt dabei keine Rolle. Bitte achten Sie darauf, dass Sie in jedes Eingabefeld nur einen Titel eintragen. Es macht nichts, wenn einige Felder frei bleiben. Rechts von den jeweiligen Titeln befinden sich Auswahlknöpfe. Geben Sie damit bitte an, wie lebhaft / detailliert Sie sich noch an den Titel, den Sie links eingetragen haben, erinnern können. Betätigen Sie den "nächste Seite"-Button wenn Sie alle Fragen beantwortet haben.**

**Titel des Beitrags**   sehr detailliert  detailliert  weder detailliert noch lückenhaft  lückenhaft  sehr lückenhaft

**Titel des Beitrags**   sehr detailliert  detailliert  weder detailliert noch lückenhaft  lückenhaft  sehr lückenhaft

**Titel des Beitrags**   sehr detailliert  detailliert  weder detailliert noch lückenhaft  lückenhaft  sehr lückenhaft

**Titel des Beitrags**   sehr detailliert  detailliert  weder detailliert noch lückenhaft  lückenhaft  sehr lückenhaft

**Titel des Beitrags**   sehr detailliert  detailliert  weder detailliert noch lückenhaft  lückenhaft  sehr lückenhaft

**Titel des Beitrags**   sehr detailliert  detailliert  weder detailliert noch lückenhaft  lückenhaft  sehr lückenhaft

Abbildung C.7: Zwischenfragebogen, Teil B: Abfrage der erinnerten Nachrichtentitel.

### C.3. ZWISCHENFRAGEBOGEN

Sie haben gerade angegeben, dass Sie sich noch an den Nachrichtenbeitrag mit dem Titel **Unwetter im Mittelmeer** erinnern. Geben Sie bitte im Folgenden ein, an welche Informationen Sie sich aus diesem Nachrichtenbeitrag erinnern. Beginnen Sie für jede Information ein neues Textfeld. Geben Sie bitte nur Informationen an, die in dem gezeigten Nachrichtenbeitrag enthalten waren. Es ist nicht nötig, dass Sie vollständige Sätze formulieren. Ein Beispielfall wäre, dass Sie in das erste Textfeld "Gegner war Frankreich" eintragen, dann zum nächsten Textfeld gehen und dort "Endstand war 6:4" eintragen. Wenn die ersten beiden Textfelder gefüllt sind erscheinen automatisch weitere Textfelder für mögliche weitere Informationen. Wenn Ihnen dann keine Details mehr einfallen geht es mittels des "nächste Seite"-Buttons weiter.

|            |
|------------|
| in Italien |
| in Spanien |
|            |
|            |

Abbildung C.8: Zwischenfragebogen, Teil C: Abfrage der erinnerten Nachrichteninhalte.

Im Folgenden möchten wir von Ihnen erfahren, wie häufig Sie die einzelnen Geräte bei der Betrachtung der Informationen verwendet haben. Bitte wählen Sie dazu die zu den Aussagen passenden Werte mittels der Auswahlknope aus. Anschließend geht es weiter mit dem "nächste Seite"-Button.

|                                                                    |                                        |                                 |                                                    |                                       |                                                 |
|--------------------------------------------------------------------|----------------------------------------|---------------------------------|----------------------------------------------------|---------------------------------------|-------------------------------------------------|
| Ich habe häufig auf dem Fernseher geschaut.                        | <input type="radio"/> stimme völlig zu | <input type="radio"/> stimme zu | <input checked="" type="radio"/> bin unentschieden | <input type="radio"/> stimme nicht zu | <input type="radio"/> stimme überhaupt nicht zu |
| Ich habe häufig auf den tragbaren Kleinst-Computer (PDA) geschaut. | <input type="radio"/> stimme völlig zu | <input type="radio"/> stimme zu | <input checked="" type="radio"/> bin unentschieden | <input type="radio"/> stimme nicht zu | <input type="radio"/> stimme überhaupt nicht zu |
| Ich habe häufig auf den elektronischen Bilderrahmen geschaut.      | <input type="radio"/> stimme völlig zu | <input type="radio"/> stimme zu | <input checked="" type="radio"/> bin unentschieden | <input type="radio"/> stimme nicht zu | <input type="radio"/> stimme überhaupt nicht zu |

Abbildung C.9: Zwischenfragebogen, Teil D: Abfrage der Gerätenutzung.

Die Fragen auf der ersten Seite des Zwischenfragebogens wurden auf die gesehene Präsentationsbedingung (BS, BSC oder TB) abgestimmt. Eine Vp, die z. B. die Bedingung TB gesehen hatte, wurde nur gebeten, Text und Bilder zu bewerten. Die Sprachausgabe oder der animierten Charakter wurden nicht bewertet. Abbildung C.6 zeigt den Fragebogen für die BSC-Bedingung.

Auf der zweiten Seite des Zwischenfragebogens wurden die Titel der Nachrichtenbeiträge abgefragt, an die sich die Vp erinnert (Abbildung C.7). Im dritten Teil des Zwischenfragebogens wurden die Vpn gebeten, zu jedem Titel, den sie im vorherigen Teil angegeben haben, den Inhalt anzugeben, an den sie sich erinnern. Dies erfolgte mittels des in Abbildung C.8 angegebenen Fragebogens.

Im letzten Teil des Zwischenfragebogens wurden die Vpn in Gruppen mit mehr als einem Ausgabegerät (d. h. alle Gruppen außer Tastatur-TV) gefragt, wie häufig sie die einzelnen Geräte verwendet haben. Abbildung C.9 zeigt den Fragebogen für Vpn aus der Gruppe PDA-TV-Bilderrahmen.

#### C.4 Abschlussfragebogen

Im Abschlussfragebogen wurden den Vpn verschiedene Fragen über die jeweils verwendete Gerätekombination gestellt. Abbildung C.10 zeigt den Abschlussfragebogen in der Gruppe PDA-TV-Bilderrahmen.

**Abschließend möchten wir von Ihnen noch erfahren, wie Sie das von Ihnen getestete System als Ganzes beurteilen. Bitte bewerten Sie dazu die folgenden Aussagen. Klicken Sie dazu bitte mit der Maus auf den Auswahlknopf, der am ehesten Ihrer Meinung entspricht. Wenn Sie alle Fragen beantwortet haben, klicken Sie auf die Taste "nächste Seite" in der rechten unteren Ecke, um den Versuch abzuschließen. Vielen Dank!**

Mir haben die Ausgaben auf dem Fernseher, dem Bilderrahmen und dem tragbaren Klein-Computer (PDA) insgesamt gut gefallen.

Der folgende Ausgabebetyp hat mir von den drei gezeigten am besten gefallen:

Ich empfand die Kombination mehrerer Geräte bei der Ausgabe von Nachrichtenbeiträgen als nützlich.

Ich würde bei der Ausgabe von Nachrichten die Kombination mehrerer Geräte einem einzelnen Fernseher gegenüber bevorzugen.

Im weißen Eingabefeld rechts haben Sie Platz für Kommentare oder Verbesserungswünsche zum Versuch.

stimme völlig zu   
  stimme zu   
  bin unentschieden   
  stimme nicht zu   
  stimme überhaupt nicht zu

Text-Bild   
  Bild-Sprache   
  Bild-Sprache-Helfer

stimme völlig zu   
  stimme zu   
  bin unentschieden   
  stimme nicht zu   
  stimme überhaupt nicht zu

stimme völlig zu   
  stimme zu   
  bin unentschieden   
  stimme nicht zu   
  stimme überhaupt nicht zu

Abbildung C.10: Abschlussfragebogen in der PDA-TV-Bilderrahmen-Gruppe.

## Literaturverzeichnis

- Allen, James F. (1983). Maintaining Knowledge about Temporal Intervals. *Communications of the ACM*, 26(11):832–843.
- André, Elisabeth (1995). *Ein plan-basierter Ansatz zur Generierung multimedialer Präsentationen*. Dissertation, Technische Fakultät, Universität des Saarlandes.
- André, Elisabeth, Kristian Concepcion, Inderjeet Mani & Linda Van Guilder (2005). Autobriefer: A System for Authoring Narrated Briefings. In Oliviero Stock & Massimo Zancanaro (Eds.), *Intelligent Multimodal Information Presentation*, pp. 143–158. Dordrecht, Netherlands: Kluwer Academic Publishers.
- André, Elisabeth & Thomas Rist (1991). Generating Coherent Presentations Employing Textual and Visual Material. *AI Review*, 9:147–165.
- André, Elisabeth, Baldes Stephan, Thomas Kleinbauer & Thomas Rist (2000). *CkuCkuCk 1.01 Planning Multimedia Presentations*. <http://www.dfki.de/imedia/miau/software/CkuCkuCk/manual/manual.html>.
- André, Elisabeth (2003). Natural Language in Multimedia/Multimodal Systems. In R. Mitkov (Ed.), *Handbook of Computational Linguistics*, pp. 650–669. Oxford University Press.
- André, Elisabeth, Wolfgang Finkler, Winfried Graf, Thomas Rist, Anne Schauder & Wolfgang Wahlster (1993). WIP: The Automatic Synthesis of Multimodal Presentations. In Mark Maybury (Ed.), *Intelligent Multimedia Interfaces*, pp. 75–93. AAAI Press/The MIT Press.
- Arens, Yigal, Eduard H. Hovy & Mira Vossers (1993). On the Knowledge Underlying Multimedia Presentations. In Mark T. Maybury (Ed.), *Intelligent Multimedia Interfaces*, pp. 280–306. AAAI Press/The MIT Press.
- Baddeley, Alan D. (2001). Is Working Memory Still Working? *American Psychologist*, 56(11):849–864.
- Baldus, Heribert, Markus Baumeister, Huib Eggenhuisen, Andrés Montvay & Wim Stut (2000). WWICE - An Architecture for In-Home Digital Networks. In Klara Nahrstedt & Wu-chi Feng (Eds.), *Multimedia Computing and Networking MMCN00*, pp. 196–203. San Jose, CA.

## LITERATURVERZEICHNIS

- Bandelloni, Renata & Fabio Paternò (2004). Flexible Interface Migration. In *Proc. Ninth International Conference on Intelligent User Interfaces IUI*, pp. 148–155. Island of Madeira, Portugal: ACM Press.
- Becker, Thomas, Helmut Hauptmeier & Katja Helfers (2004). *TV 2010 - Die Digitalisierung des Fernsehens*. Neunkirchen, Germany: Buhl Data Service GmbH.
- Bellman, Richard (1961). *Adaptive Control Processes: A Guided Tour*. Princeton University Press.
- Berners-Lee, Tim & Mark Fischetti (1999). *Weaving the Web: The Original Design and Ultimate Destiny of the World Wide Web by its Inventor*. San Francisco, CA: Harper.
- Bernsen, Niels Ole (2001). Multimodality in Language and Speech systems - from Theory to Design Support Tool. In Björn Granström, David House & Inger Karlsson (Eds.), *Multimodality in Language and Speech Systems*, pp. 93–148. Dordrecht, Netherlands: Kluwer Academic Publishers.
- Bohnenberger, Thorsten, Antony Jameson, Antonio Krüger & Andreas Butz (2002). User Acceptance of a Decision-Theoretic Location-Aware Shopping Guide. In Yolanda Gil & David B. Leake (Eds.), *Proc. 6th International Conference on Intelligent User Interfaces IUI*, pp. 178–179. San Francisco, CA: ACM Press.
- Bolt, Richard A. (1980). Put-That-there: Voice and Gesture at the Graphics Interface. In John C. Thomas, Robert A. Ellis & Harvey Z. Kriloff (Eds.), *Proc. 7th Annual Conference on Computer Graphics and Interactive Techniques SIGGRAPH*, pp. 262–270. Seattle, WA: ACM Press.
- Bortz, Jürgen (2005). *Statistik für Human- und Sozialwissenschaftler*. Berlin, Germany: Springer. Sixth Edition.
- Bortz, Jürgen & Nicola Döring (2003). *Forschungsmethoden und Evaluation für Human- und Sozialwissenschaftler*. Berlin, Germany: Springer. Third Edition.
- Braun, Elmar, Erwin Aitenbichler & Max Mühlhäuser (2004). Talk to your Displays - Position Paper for the WS Ubiquitous Display Environments. In *Workshop on Ubiquitous Display Environments, UbiComp04*. Nottingham, United Kingdom.

## LITERATURVERZEICHNIS

- Braun, Elmar & Max Mühlhäuser (2004). Extending XML UIDLs for Multi-Device Scenarios. In *Workshop Developing User Interfaces with XML: Advances on User Interface Description Languages, AVI 2004*. Gallipoli, Italy.
- Brumitt, Barry, Brian Meyers, John Krumm, Amanda Kern & Steven Shafer (2000). EasyLiving: Technologies for Intelligent Environments. In Hans W. Gellersen & Peter J. Thomas (Eds.), *Proc. Second International Symposium on Handheld and Ubiquitous Computing HUC2k*, pp. 12–29. Bristol, United Kingdom.
- Burger, John D. & Ralph J. Marshall (1993). The Application of Natural Language Models to Intelligent Multimedia. In Mark T. Maybury (Ed.), *Intelligent Multimedia Interfaces*, pp. 174–196. Menlo Park, CA: AAAI Press/The MIT Press.
- Carbonell, Jaime G. (1970). *Mixed-Initiative Man-Computer Instructional Dialogues*. Technical Report 1971, Cambridge, MA: Bolt, Beranek and Newman Inc.
- Cassell, Justine, Hannes Vilhjálmsón & Tim Bickmore (2001). BEAT : the Behavior Expression Animation Toolkit. In Lynn Pocock (Ed.), *Proc. 28th Annual Conference on Computer Graphics and Interactive Techniques SIGGRAPH*, pp. 477 – 486. Los Angeles, CA: ACM Press.
- Chan, Dominic, Adrian Fourcin, Dafydd Gibbon, Björn Granström, Mark Huckvale, George Kokkinakis, Knut Kvale, Lori Lamel, Borge Lindberg, Asunción Moreno, Jiannis Mouropoulos, Franco Senia, Isabel Trancoso, Corin 't Veld & Jerome Zeiliger (1995). EUROM- A Spoken Language Resource for the EU. In José M. Pardo (Ed.), *Proc. Fourth European Conference on Speech Communication and Technology Eurospeech*, pp. 867–870. Madrid, Spain: European Speech Communication Association.
- Channarukul, Songsak, Susan W. McRoy & Syed S. Ali (2004). MULTI-FACE: Multimodal Content Adaptations for Heterogeneous Devices. In Rajeev Sharma, Trevor Darrell, Mary P. Harper, Gianni Lazzari & Matthew Turk (Eds.), *Proc. Int. Conf. on Multimodal Interfaces ICMI*, pp. 319–320. State College, PA: ACM Press.
- Coffman, Kerry G. & Andrew M. Odlyzko (2002). Growth of the Internet. In Ivan P. Kaminow & Tingye Li (Eds.), *Optical Fiber Telecommunications IV B: Systems and Impairments*, pp. 17–56. New York, NY: Academic Press.

## LITERATURVERZEICHNIS

- ComScore Presseerklärung (2002). *Half of U.S. Internet Users Have a Television and PC in the Same Room*. Verfügbar unter <http://www.comscore.com/press/release.asp?id=88>.
- Coors, Volker, Christian Kray, Katri Laakso & Christian Elting (2005). Presenting Route Instructions on Mobile Devices. In Jason Dykes & Alan M. MacEachren (Eds.), *Exploring Geovisualization*, pp. 529–550. Amsterdam, Netherlands: Elsevier.
- Czerwinski, Mary, Greg Smith, Tim Regan, Brian Meyers, George Robertson & Gary Starkweather (2003). Toward Characterizing the Productivity Benefits of Very Large Displays. In *Proc. Interact*, pp. 9–16. Zurich, Switzerland: IOS Press.
- Czerwinski, Mary, Desney Tan & George G. Robertson (2002). Women take a Wider View. In Loren G. Terveen & Dennis R. Wixon (Eds.), *Proc. Conference on Human Factors in Computing Systems CHI*, pp. 195–202. Minneapolis, MN: ACM Press.
- Dahlbäck, Nils, Arne Jönsson & Lars Ahrenberg (1993). Wizard of Oz Studies - Why and How. *Knowledge-Based Systems*, 6(4):258–266.
- DeCarolis, Berardina, Catherine Pelachaud, Isabella Poggi & Mark Steedman (2004). APMML, a Mark-Up Language for Believable Behavior Generation. In Helmut Prendinger & Mitsuru Ishizuka (Eds.), *Life-Like Characters. Tools, Affective Functions and Applications*, pp. 65–85. Berlin, Germany: Springer.
- Ding, Yun, Christian Elting & Ulrich Scholz (2006). Seamless Integration of Output Devices into Intelligent Environments: Infrastructure, Strategies and Implementation. In *Proc. International Conference on Intelligent Environments*, pp. 21–30. Athens, Greece: Institution of Engineering and Technology IET.
- Ding, Yun & Heiner Litz (2006). Creating Multiplatform User Interfaces by Annotation and Adaptation. In Cecile L. Paris & Candace L. Sidner (Eds.), *Proc. Eleventh International Conference on Intelligent User Interfaces IUI*, pp. 270–272. Sidney, Australia: ACM Press.
- Dryer, D. Christopher, Chris Eisbach & Wendy S. Ark (1999). At What Cost Pervasive? A Social Computing View of Mobile Computing Systems. *IBM Systems Journal*, 38(4):652–676. Special issue on Pervasive Computing.

## LITERATURVERZEICHNIS

- Edwards, W. Keith & Rebecca Grinter (2001). At Home with Ubiquitous Computing: Seven Challenges. In Gregory D. Abowd, Barry Brumitt & Steven Shafer (Eds.), *International Conference on Ubiquitous Computing Ubicomp 2001*, pp. 256–272. Atlanta, GA.
- Edwards, W. Keith, Mark W. Newman & Jana Sedivy (2001). *The Case for Recombinant Networking*. Technical Report CSL-01-1, Palo Alto, CA: Xerox Palo Alto Research Center.
- Elting, Christian & Michael Hellenschmidt (2004). Strategies for Self-Organization and Multimodal Output Coordination in Distributed Device Environments. In *Workshop on Artificial Intelligence in Mobile Systems AIMS*, pp. 20–27. Nottingham, United Kingdom.
- Elting, Christian, Stefan Rapp, Gregor Möhler & Michael Strube (2003). Architecture and Implementation of Multimodal Plug and Play. In Sharon L. Oviatt, Trevor Darrell, Mark T. Maybury & Wolfgang Wahlster (Eds.), *Proc. Int. Conf. on Multimodal Interfaces ICMI-PUI03*, p. 93–100. Vancouver, BC, Canada: ACM Press.
- Elting, Christian, Jan Zwickel & Rainer Malaka (2002). Device-Dependant Modality Selection for User Interfaces - An Empirical Study. In Yolanda Gil & David B. Leake (Eds.), *Proc. Seventh International Conference on Intelligent User Interfaces IUI*, pp. 55–62. San Francisco, CA: ACM Press.
- Feiner, Steven K. & Kathleen McKeown (1998). Automating the Generation of Coordinated Multimedia Explanations. In M. Maybury & W. Wahlster (Eds.), *Readings in Intelligent User Interfaces*, pp. 89–98. San Francisco, CA: Morgan Kaufman Publishers.
- Fischermann, Thomas (2004). *Auf den Spuren Big Brothers*. DIE ZEIT.de Dossier, 11.08.2004. [http://www.zeit.de/2004/34/ueberwachung\\_liste](http://www.zeit.de/2004/34/ueberwachung_liste).
- Gajos, Krzysztof, David Christianson, Raphael Hoffmann, Tal Shaked, Kiera Henning, Jing Jing Long & Daniel S. Weld (2005). Fast And Robust Interface Generation for Ubiquitous Applications. In Michael Beigl, Stephen S. Intille, Jun Rekimoto & Hideyuki Tokuda (Eds.), *Proc. Int. Conf. on Ubiquitous Computing UbiComp*, pp. 37–55. Tokyo, Japan: Springer.
- Grudin, Jonathan (2001). Partitioning Digital Worlds: Focal and Peripheral Awareness in Multiple Monitor Use. In Julie A. Jacko, Andrew Sears, Michel Beaudouin-Lafon & Robert J.K. Jacob (Eds.), *Int. Conf. on Human Factors in Computing Systems CHI*, pp. 458–465. Seattle, WA: ACM Press.

- HAVi (2001). *Specification of the Home Audio/Video Interoperability (HAVi) Architecture Version 1.1*. <http://www.havi.org/>.
- Heider, Thomas & Thomas Kirste (2005). Multimodal Appliance Cooperation Based on Explicit Goals: Concepts & Potentials. In Gerard Bailly, James L. Crowley & Gilles Privat (Eds.), *Proc. Joint Conference on Smart Objects and Ambient Intelligence sOc-EUSAI*, pp. 271–276. Grenoble, France: ACM Press.
- Hellenschmidt, Michael (2005). Distributed Implementation of a self-organizing Appliance Middleware. In Gerard Bailly, James L. Crowley & Gilles Privat (Eds.), *Proc. Joint Conference on Smart Objects and Ambient Intelligence sOc-EUSAI*, pp. 201–206. Grenoble, France: ACM Press.
- Herfet, Thorsten, Thomas Kirste & Michael Schnaider (2001). EMBASSI - Multimodal Assistance for Infotainment and Service Infrastructures. In Joaquim Jorge, Rachelle Heller & Richard Guedj (Eds.), *Proc. EC/NSF Workshop Universal on Accessibility of Ubiquitous Computing: Providing for the Elderly*, pp. 41–50. Alcácer do Sal, Portugal: ACM Press.
- Hess, Christopher K., Manuel Román & Roy H. Campbell (2002). Building Applications for Ubiquitous Computing Environments. In Friedemann Mattern & Mahmoud Naghshineh (Eds.), *International Conference on Pervasive Computing 2002*. Zurich, Switzerland.
- Hilbert, David M. & Jonathan Trevor (2004). Personalizing Shared Ubiquitous Devices. *ACM interactions*, 9(3):35–43.
- Hutchings, Douglas Ralph, Greg Smith, Meyers Brian, Mary Czerwinski & George Robertson (2004). Display Space Usage and Window Management Operation Comparisons between Single Monitor and Multiple Monitor Users. In *Proc. Advanced Visual Interfaces*, pp. 32–39. Gallipoli, Italy.
- Jöst, Matthias, Jochen Häußler, Matthias Merdes & Rainer Malaka (2005). Multimodal Interaction for Pedestrians: An Evaluation Study. In Robert St. Amant, John Riedl & Anthony Jameson (Eds.), *Proc. 10th International Conference on Intelligent User Interfaces IUI*, pp. 59–66. San Diego, CA: ACM Press.
- Kerpedjiev, Stephan, Giuseppe Carenini, Steven F. Roth & Johanna D. Moore (1997). Integrating Planning and Task-Based Design for Multimedia Presentation. In *Proc. International Conference on Intelligent User Interfaces IUI97*.

- Kilger, Anne (1994). Using UTAGs for incremental and parallel generation. *Computational Intelligence*, 10(4):591–603.
- Kopp, Stefan, Brigitte Krenn, Stacy Marsella, Andrew N. Marshall, Catherine Pelachaud, Hannes Pirker, Kristinn R. Thórisson & Hannes Vilhjálmsson (2006). Towards a Common Framework for Multimodal Generation in ECAs: The Behavior Markup Language. In Jonathan Gratch, Michael Young, Ruth Aylett, Daniel Ballin & Patrick Olivier (Eds.), *Proc. 6th International Conference on Intelligent Virtual Agents IVA*, pp. 205–217. Marina del Rey, CA: Springer.
- Kopp, Stefan & Ipke Wachsmuth (2004). Synthesizing Multimodal Utterances for Conversational Agents. *Computer Animation and Virtual Worlds*, 15(1):39–52.
- Kray, Christian, Antonio Krüger & Christoph Endres (2003). Some Issues on Presentations in Intelligent Environments. In Panos Markopoulos, Berry Eggen, Emile H. L. Aarts & James L. Crowley (Eds.), *First European Symposium on Ambient Intelligence EUSAI*, pp. 15–26. Veldhoven, Netherlands: Springer.
- Kruppa, Michael (2004). The Better Remote Control - Multiuser Interaction with Public Displays. In Stephen Viller & Peta Wyethh (Eds.), *Workshop on Multi-User and Ubiquitous User Interfaces MU3I*. Madeira, Portugal.
- Kruppa, Michael (2006). *Migrating Characters: Effective User Guidance in Instrumented Environments*. PhD Thesis, Saarland University, Saarbrücken, Germany.
- Lee, Jina & Stacy Marsella (2006). Nonverbal Behavior Generator for Embodied Conversational Agents. In Jonathan Gratch, Michael Young, Ruth Aylett, Daniel Ballin & Patrick Olivier (Eds.), *Proc. 6th International Conference on Intelligent Virtual Agents IVA*, pp. 243–255. Marina del Rey, CA: Springer.
- Linden, Greg, Brent Smith & Jeremy York (2003). Amazon.com Recommendations: Item-to-Item Collaborative Filtering. *IEEE Internet Computing*, 7(1):76–80.
- Ludwig, Bernd (2004). *Ein konfigurierbares Dialogsystem für Mensch-Maschine-Interaktion in gesprochener Sprache*. Dissertation, Universität Erlangen-Nürnberg.

## LITERATURVERZEICHNIS

- Malaka, Rainer, Jochen Häußler & Hidir Aras (2004). SmartKom Mobile - Intelligent Ubiquitous User Interaction. In Jean Vanderdonckt, Nuno Jardim Nunes & Charles Rich (Eds.), *Proc. 9th International Conference on Intelligent User Interfaces IUI*, pp. 310–312. Funchal, Madeira, Portugal: ACM Press.
- Martínez, José M. (2004). *MPEG-7 Overview (version 10)*. <http://www.chiariglione.org/mpeg/standards/mpeg-7/mpeg-7.htm>.
- Massó, José Pascual Molina, Jean Vanderdonckt & Pascual González López (2006). Direct Manipulation of User Interfaces for Migration. In *Proc. Eleventh International Conference on Intelligent User Interfaces IUI*, pp. 140–147. Sidney, Australia: ACM Press.
- Maybury, Mark & Wolfgang Wahlster (Eds.) (1998). *Readings in Intelligent User Interfaces*. San Francisco, CA: Morgan Kaufman Publishers.
- Mayer, Richard E. & Roxana Moreno (1998). A Split-Attention Effect in Multimedia Learning: Evidence for Dual Processing Systems in Working Memory. *Journal of Educational Psychology*, 90(2):312–320.
- Meyer zu Kniendorf, Christoph (2004). Interaktionskonzepte in verteilten und vernetzten Systemen am Beispiel der Unterhaltungselektronik. In C. Stefens, M. Thüring & L. Urbas (Eds.), *Entwerfen und Gestalten. 5. Berliner Werkstatt Mensch-Maschine-Systeme*, Vol. 22, Fortschritt-Berichte VDI 16, pp. 488–506. VDI Verlag GmbH.
- Mollenhoff, Clark R. (1988). *Atanasoff: Forgotten Father of the Computer*, pp. 47–48. Iowa State University Press.
- Moore, Gordon E. (1965). Cramming More Components onto Integrated Circuits. *Electronics Magazine*, 38(8):114–117.
- Mozer, Michael C. (2004). Lessons from an Adaptive House. In Diane Cook & Sajal Das (Eds.), *Smart Environments: Technologies, Protocols, and Applications*, pp. 273–294. Hoboken, NJ: J. Wiley & Sons.
- Myers, Brad A., Jeffrey Nichols, Jacob O. Wobbrock & Robert C. Miller (2004). Taking Handheld Devices to the Next Level. *IEEE Computer*, 37(12):36–43.
- Nau, Dana, Malik Ghallab & Paolo Traverso (2004). *Automated Planning: Theory & Practice*. Morgan Kaufmann Publishers.

## LITERATURVERZEICHNIS

- Neal, Jeannette G. & Stuart C. Shapiro (1991). Intelligent Multi-Media Interface Technology. In Joseph W. Sullivan & Sherman W. Tyler (Eds.), *Intelligent User Interfaces*, pp. 11–43. Reading, MA: ACM Press.
- Nijholt, Anton, Thomas Rist & Kees Tuijnjenbreijer (2004). Lost in Ambient Intelligence? In *Proc. ACM Conference on Computer Human Interaction Extended Abstracts*, pp. 1725–1726. Vienna, Austria: ACM Press.
- Norman, Donald A. (2002). Complexity Versus Difficulty: Where Should the Intelligence Be? In Yolanda Gil & David B. Leake (Eds.), *Proc. Seventh International Conference on Intelligent User Interfaces IUI*, p. 4. San Francisco, CA: ACM Press.
- Oberle, Daniel, Anupriya Ankolekar, Pascal Hitzler, Philipp Cimiano, Michael Sintek, Malte Kiesel, Babak Mougouie, Shankar Vembu, Stephan Baumann, Massimo Romanelli, Paul Buitelaar, Ralf Engel, Daniel Sonntag, Norbert Reithinger, Berenike Loos, Robert Porzel, Hans-Peter Zorn, Vanessa Micelli, Schmidt Christian, Moritz Weiten, Felix Burkhardt & Jianshen Zhou (2006). *DOLCE ergo SUMO: On Foundational and Domain Models in SWIntO (SmartWeb Integrated Ontology)*. Technical Report: AIFB, University of Karlsruhe.
- Obrenovic, Zeljko & Dusan Starcevic (2004). Modeling Multimodal Human-Computer Interaction. *IEEE Computer*, 37(9):65–72.
- Oonk, Heather M., Janel H. Schermerhorn, Dale Glaser & Jeffrey G. Morrison (2003). *Knowledge Desk Limited Objective Experiment*. Technical Report 1896, San Diego, CA: SPAWAR Systems Center.
- Oviatt, Sharon L. (1997). Multimodal Interactive Maps: Designing for Human Performance. *Human-Computer Interaction*, 12(1-2):93–129. Special Issue on Multimodal Interfaces.
- Oviatt, Sharon L. (2003). Multimodal Interfaces. In Julie A. Jacko & Andrew Sears (Eds.), *The Human-Computer Interaction Handbook: Fundamentals, Evolving Technologies and Emerging Applications*, pp. 286–304. Mahwah, NJ: Lawrence Erlbaum Assoc.
- Oviatt, Sharon L., Rachel Coulston & Rebecca Lunsford (2004). When Do We Interact Multimodally? Cognitive Load and Multimodal Communication Patterns. In Rajeev Sharma, Trevor Darrell, Mary P. Trevor, Gianni Lazzari & Matthew Turk (Eds.), *Proc. Sixth International Conference on Multimodal Interfaces ICMI*, pp. 129–136. State College, PA: ACM Press.

## LITERATURVERZEICHNIS

- Oviatt, Sharon L. & Robert van Gent (1996). Error Resolution during Multimodal Human-Computer Interaction. In T. Bunnell & W. Idsardi (Eds.), *Proc. 4th International Conference on Spoken Language Processing ICSLP*, pp. 204–207. Philadelphia, PA: University of Delaware and AI duPont Institute.
- Pauws, Steffen (2002). CubyHum: A Fully Operational Query by Humming System. In *Proc. 3rd International Conference on Music Information Retrieval ISMIR*, pp. 187–196. Paris, France.
- Pham, Thai-Lai, Georg Schneider & Stuart Goose (2000). A Situated Computing Framework for Mobile and Ubiquitous Multimedia Access Using Small Screen and Composite Devices. In Shahram Ghandeharizadeh, Shih-Fu Chang, Stephen Fischer, Joseph A. Konstan & Klara Nahrstedt (Eds.), *Proc. Eighth ACM International Conference on Multimedia*, pp. 323–331. Marina del Rey, CA: ACM Press.
- Pilotta, Joseph J., Don E. Schultz, Gary Drenik & Phil Rist (2004). Simultaneous Media Usage: A Critical Consumer Orientation to Media Planning. *Journal of Consumer Behaviour*, 3(3):285–292.
- Pinhanez, Claudio & Mark Podlaseck (2005). To Frame or Not to Frame: The Role and Design of Frameless Displays in Ubiquitous Applications. In Michael Beigl, Stephen S. Intille, Jun Rekimoto & Hideyuki Tokuda (Eds.), *Proc. Int. Conf. on Ubiquitous Computing UbiComp*, pp. 340–357. Tokyo, Japan: Springer.
- Prendinger, Helmut, Sylvain Descamps & Mitsuru Ishizuka (2004). MPML: A Markup Language for Controlling the Behavior of Life-Like Characters. *Journal of Visual Languages and Computing*, 15(2):183–203.
- Puerta, Angel & Jacob Eisenstein (2002). XML: A Common Representation for Interaction Data. In *Proc. Seventh International Conference on Intelligent User Interfaces IUI*, pp. 214–215. San Francisco, CA: ACM Press.
- Ranganathan, Anand, Jalal Al-Muhtadi, Jacob Biehl, Brian Ziebart, Roy H. Campbell & Brian Bailey (2005). Towards a Pervasive Computing Benchmark. In *Proc. Workshop on Middleware Support for Pervasive Computing PerWare*, pp. 194 – 198. Paris, France: IEEE.
- Ranganathan, Anand, Robert E. McGrath, Roy H. Campbell & Dennis Mickunas (2003). Ontologies in a Pervasive Computing Environment. In *Workshop on Ontologies and Distributed Systems, IJCAI 2003*. Acapulco, Mexico.

## LITERATURVERZEICHNIS

- Reeves, Byron & Clifford Nass (1999). *The Media Equation: How People Treat Computers, Television, and New Media Like Real People and Places*. New York, NY: Cambridge University Press, 1999.
- Reithinger, Norbert, Jan Alexandersson, Tilman Becker, Anselm Blocher, Ralf Engel, Markus Löckelt, Jochen Müller, Norbert Pflieger, Peter Poller, Michael Streit & Valentin Tschernomas (2003). SmartKom - Adaptive and Flexible Multimodal Access to Multiple Applications. In Wolfgang Wahlster, Sharon L. Oviatt, Trevor Darrell & Mark T. Maybury (Eds.), *Proc. Int. Conf. on Multimodal Interfaces ICMI-PUI03*, pp. 101–108. Vancouver, BC, Canada.
- Rekimoto, Jun (2000). Multiple-Computer User Interfaces: "Beyond the Desktop" Direct Manipulation Environments. In Gerd Szwillus & Thea Turner (Eds.), *CHI '00 Extended Abstracts on Human Factors in Computing Systems*, pp. 6–7. The Hague, Netherlands: ACM Press.
- Rieger, Thomas (2003). Avatar Gestures. In Václav Skala (Ed.), *Proc. 11-th International Conference in Central Europe on Computer Graphics, Visualization and Computer Vision WSCG*, pp. 379–386. Plzen, Czech Republic: University of West Bohemia.
- Rieger, Thomas & Uwe Berner (2002). A Scalable Avatar for Conversational User Interfaces. In Noëlle Carbonell & Constantine Stephanidis (Eds.), *7th ERCIM User Interfaces for All Workshop*, pp. 350–359. Paris, France: Springer.
- Robertson, George, Mary Czerwinski, Patrick Baudisch, Brian Meyers, Daniel Robbins, Greg Smith & Desney Tan (2005). The Large-Display User Experience. *IEEE Computer Graphics and Applications*, 25(4):44–51.
- Roth, Steven F., Joseph Mattis & Mesnard Xavier (1991). Graphics and Natural Language as Components of Automatic Explanation. In Sullivan, Joseph W. and Tyler, Sherman W. (Ed.), *Intelligent User Interfaces*, pp. 207–239. ACM Press.
- Stent, Amanda & Hui Guo (2005). A New Data-Driven Approach for Multimedia Presentation Planning. In M. H. Hamza (Ed.), *Proc. IASTED International Conference on Internet and Multimedia Systems and Applications EuroIMSA*, pp. 395–400. Grindelwald, Switzerland: IASTED/ACTA Press.
- Sukaviriya, Noi, Mark Podlaseck, Rick Kjeldsen, Anthony Levas, Gopal Pingali & Claudio Pinhanez (2003). Embedding Interactions in a Retail Store

## LITERATURVERZEICHNIS

- Environment: The Design and Lessons Learned. In Matthias Rauterberg, Marino Menozzi & Janet Wesson (Eds.), *Proc. of the Ninth IFIP International Conference on Human-Computer Interaction INTERACT03*. Zurich, Switzerland: IOS Press.
- Tabbers, Huib K., Rob L. Martens & Jeroen J. G. van Merriënboer (2001). The Modality Effect in Multimedia Instructions. In *Proc. Twenty-Third Annual Conference of the Cognitive Science Society*, pp. 1024–1029. Edinburgh, Scotland: Lawrence Erlbaum Associates Publishers.
- Tan, Desney S. & Mary Czerwinski (2003). Effects of Visual Separation and Physical Discontinuities when Distributing Information across Multiple Displays. In Stephen Viller & Peta Wyeth (Eds.), *Proc. Australian Computer Human Interaction Conference OZCHI*, pp. 184–191. Brisbane, Australia: CHISIG Australia.
- Thórisson, Kristinn K. R. (1999). A Mind Model for Multimodal Communicative Creatures & Humanoids. *International Journal of Applied Artificial Intelligence*, 13(4-5):449–486.
- Tindall-Ford, Sharon, Paul Chandler & John Sweller (1997). When Two Sensory Modes are Better Than One. *Journal of Experimental Psychology: Applied*, 3(4):257–287.
- Towns, Stuart, Charles Callaway & James Lester (1998). Generating Coordinated Natural Language and 3D Animations for Complex Spatial Explanations. In *AAAI Workshop on Representations for Multi-modal Human-Computer Interaction*, pp. 112–119. Madison, WI: AAAI Press.
- Universal Plug and Play (2006). <http://www.upnp.org/>.
- van Dam, Andries (1997). Post-WIMP User Interfaces. *Communications of the ACM*, 40(2):63–67.
- Van Mulken, Susanne, Elisabeth André & Jochen Müller (1998). The Persona Effect: How Substantial Is It? In *British Computer Society HCI Conference*, pp. 53–66. Sheffield, UK: Springer.
- van Ossenbruggen, Jacco, Frank Cornelissen, Joost Geurts, Lloyd Rutledge & Lynda Hardman (2000). *PPP-Persona: Ein objektorientierter Multimedia-Präsentationsagent*. Technical Report, Amsterdam, Netherlands: Centrum voor Wiskunde en Informatica CWI.

## LITERATURVERZEICHNIS

- Vaudry, Christophe & Michel Crampes (2001). AMEBICA - An Adaptive User Interface for Process Control. In Jean Vanderdonckt, Ann Blandford & Alain Derycke (Eds.), *Proc. Interaction Homme Machine - Human Computer Interaction Conference 2001*. Lille, France: Cepaduès-Éditions.
- Velleman, Paul F. & Leland Wilkinson (1993). Nominal, Ordinal, Interval, and Ratio Typologies Are Misleading. *The American Statistician*, 47(1):65–72.
- Vernier, Frederic & Laurence Nigay (2000). A Framework for the Combination and Characterization of Output Modalities. In Björn Granström (Ed.), *Proc. Workshop on Design, Specification and Verification of Interactive Systems DVS-IS 2000*, p. 35–50. Limerick, Ireland: Kluwer Academic Publishers.
- W3C Recommendation (1999). *XSL Transformations (XSLT) Version 1.0*. <http://www.w3c.org>.
- W3C Recommendation (2005). *Synchronized Multimedia Integration Language SMIL 2.1*. <http://www.w3c.org>.
- Wahlster, Wolfgang (1988). Methodologies for Intelligent Systems. In Zbigniew W. Ras & Lorenza Saitta (Eds.), *Proc. 4th International Conference on Spoken Language Processing ICSLP*, pp. 101–111. Amsterdam, North Holland: Elsevier.
- Wahlster, Wolfgang (1998). Intelligent User Interfaces: An Introduction. In Mark Maybury & Wolfgang Wahlster (Eds.), *Readings in Intelligent User Interfaces*, pp. 1–13. San Francisco, CA: Morgan Kaufman Publishers.
- Wahlster, Wolfgang (2003). Towards Symmetric Multimodality: Fusion and Fission of Speech, Gesture, and Facial Expression. In Andreas Günter, Rudolf Kruse & Bernd Neumann (Eds.), *Proc. of the 26th German Conference on Artificial Intelligence KI*, pp. 1–18. Hamburg, Germany: Springer.
- Wahlster, Wolfgang (Ed.) (2006). *SmartKom - Foundations of Multimodal Dialogue Systems*. Berlin, Germany: Springer.
- Wahlster, Wolfgang & Rainer Wasinger (2006). The Anthropomorphized Product Shelf: Symmetric Multimodal Interaction with Instrumented Environments. In José L. Encarnação & Emile H. L. Aarts (Eds.), *True Visions: The Emergence of Ambient Intelligence*, pp. 291–306. Berlin, Germany: Springer.

## LITERATURVERZEICHNIS

- Wasinger, Rainer (2006). *Multimodal Interaction with Mobile Devices: Fusing a Broad Spectrum of Modality Combinations*. PhD Thesis, Saarland University, Saarbrücken, Germany.
- Weiser, Mark (1991). The Computer for the 21st Century. *Scientific American*, 265(3):94–104.
- Weiser, Mark, Rich Gold & John Seely Brown (1999). The Origins of Ubiquitous Computing Research at PARC in the Late 1980s . *IBM Systems Journal*, 38(4):693–696. Special issue on Pervasive Computing.
- Yanagihara, Tadashi, Hiroshi Sakakibara, Ryo Ohsawa, Masao Ideuchi, Naohiko Kohtake, Masayuki Iwai, Kazunori Takashio & Hideyuki Tokuda (2005). A Self Configurable Topology-Aware Network For Smart Materials. In *Proc. 5th International Workshop on Smart Appliances and Wearable Computing IWSAWC*, pp. 469–474. Columbus, OH: Computer Society.
- Youngblood, Michael G., Edwin O. Heierman, Lawrence B. Holder & Diane J. Cook (2005). Automation Intelligence for the Smart Environment. In Leslie Pack Kaelbling & Alessandro Saffiotti (Eds.), *Proc. Int. Joint Conference on Artificial Intelligence IJCAI*, pp. 1513–1516. Edinburgh, Scotland: ACM Press.
- Zancanaro, Massimo, Cesare Rocchi & Oliviero Stock (2003). Automatic Video Composition. In Andreas Butz, Antonio Krüger & Patrick Olivier (Eds.), *Proc. Smart Graphics, Third International Symposium*, pp. 192–201. Heidelberg, Germany: Springer.
- Zhen, Wen, Michelle X. Zhou & Vikram Aggarwal (2005). An Optimization-based Approach to Dynamic Visual Context Management. In John T. Stasko & Matthew O. Ward (Eds.), *Proc. IEEE Symposium on Information Visualization InfoVis*, pp. 187–194. Minneapolis, MN: IEEE Press.
- Zhou, Michelle X. & Steven K. Feiner (2001). IMPROVISE: Automated Generation of Animated Graphics for Coordinated Multimedia Presentations. In Harry Bunt & Robert-Jan Beun (Eds.), *Cooperative Multimodal Communication*, pp. 43–63. Berlin, Germany: Springer.
- Zipf, Alexander & Rainer Malaka (2000). Deep Map: Challenging IT Research in the Framework of a Tourist Information System. In Daniel R. Fesenmaier, Stefan Klein & Dimitrios Buhalis (Eds.), *Proc. International Conference on Information and Communication Technologies in Tourism ENTER*, pp. 15–27. Barcelone, Spain: Springer.

## Sachwortregister

### A

- Ad-hoc-Interoperabilität . . . . . 86–93  
  heuristischer Ansatz . . . . . 89–90  
  konfigurierbarer Ansatz . . . . . 91–92  
  Koordinationsansatz . . . . . 87–89  
*agent* . . . . . 95, 96, 105–120, 127, 135, 136,  
  148, 150  
  multi-modale Parameter . . . . . 105  
Analysekomponente . . . . . 67, 84–85, 144  
Animierter Charakter . . . . . 62, 76, 142, 159,  
  187  
  Animationen . . . . . 109  
  DFKI-Charakter . . . . . 118  
  Selbstbeschreibung . . . . . 115–120, 135,  
  149, 150  
  Verhaltensweisen . . . . . 106, 137  
Anova . . . . . *siehe* Varianzanalyse  
*appearance* . . . . . 96, 106, 110, 117  
*application-volume* . . . . . 101, 102, 104,  
  112–120  
Applikationskomponente . 68, 69, 84, 145  
AT&T Natural Voices 145, 191, 204, 246  
auditoriell . . . . . *siehe* Modalität  
Ausgabeinhalt . . . . . 83, 137, 151, 164  
Ausgabekomponente . 46, 68, 70, 84, 145  
  physikalisch . . . . . 68, 85  
Ausgabeziel . . . . . 68, 83, 137, 151  
Ausreißer . . . . . 210

### B

- Balkendiagramm . . . . . 168, 210  
*behaviour* . . . . . 96, 105–107, 109, 116, 117,  
  136, 137  
Bernsen-Unimodalitätenmodell . . . . *siehe*  
  Modalitätentheorie  
BML . . . . . 64, 106, 111  
Box-Plot . . . . . 210

### C

- $\chi^2$ -Test . . . . . 168, 209  
Ckuckuk . . . . . 139, 146, 152  
*content* . . . . . 96, 106, 107, 121, 149

### D

- Dialog-Manager . 68, 70, 79, 81, 85, 126,  
  137, 145, 151  
Dialogakt . . . . . 137  
Domänenspezifisches Modell . . . . . *siehe*  
  Ein-/Ausgabediens  
Domänenunabhängiges Modell . . . . *siehe*  
  Ein-/Ausgabediens  
Dynamite . . . . . 140, 156, 164

### E

- Ein-/Ausgabediens . . . . . 81, 126, 135, 147  
  Ausgabeunimodalität 111, 135, 147  
  domänenspezifisch . . . . . 105  
  domänenunabhängig . . . . . 100  
  Eingabeunimodalität . . . . . 114  
  Medienobjekt . . . . . 107, 149  
  MMP-Komponente . . . . . 152  
  Modellierung . . . . . 86, 93, 97–100  
  multi-modale Parameter und Status  
  105  
  multi-modaler Typ . . . . . 105  
Ein-/Ausgabegerät . . . . . 46  
Eingabekomponente . 46, 67, 84, 97, 104,  
  114, 122, 136, 140, 144  
Elektronischer Bilderrahmen . . . . . 76, 141,  
  159, 187, 208  
  Selbstbeschreibung . . . . . 148  
Embassi . . . . . 40, 133  
Extremwert . . . . . 210

### F

- Fernseher . . . . . 75, 133, 141, 159, 187, 208  
Filmbilder . . . . . 159  
Filminformationssystem 51, 76, 141, 164  
Fragebogenstudie . . . . . 159  
  abhängige Variable . . . . . 163  
  Akzeptanz . . . . . 159  
  Charakter-Layout . . . . . 166  
  Fragebogen . . . . . 166  
  Gerätekombination . . . . . 163  
  Gruppe beruflicher/privater Messe-  
  aufenthalt . . . . . 171

## SACHWORTREGISTER

- Gruppe jünger/älter 20 . . . . . 170  
 Gruppe jünger/älter 30 . . . . . 170  
 Gruppe jünger/älter 40 . . . . . 170  
 Gruppe ohne/mit PDA . . . . . 172  
 Gruppe weniger/mehr als 10 Stunden  
 Fernseher . . . . . 171  
 Gruppe weniger/mehr als 30 Stunden  
 Computer . . . . . 171  
 Gruppe weniger/mehr als drei Kino-  
 filme . . . . . 172  
 Hypothese . . . . . 162, 183  
 Moderatorvariable . . . . . 164  
 PDA . . . . . 165  
 Text-Bild-Layout . . . . . 166  
 TV . . . . . 165  
 TV-Bilderrahmen . . . . . 165  
 TV-PDA . . . . . 165  
 Versuchsleiter . . . . . 164  
 Friedman-Test . . . . . 168, 209
- G**
- Gekoppelte Ausgabegeräte . . 41, 75, 140,  
 159, 187  
 Bildschirme . . . . . 41, 52  
 empirische Untersuchungen . . . . 52  
 technische Umsetzung . . . . . 56  
 Grafische Benutzeroberfläche . *siehe* GUI  
 Grafischer Text . . . . . 159  
 Greenhouse-Geisser . . . . . 210  
 GUI . . . . . 122, 133, 140, 159  
 Selbstbeschreibung . . . . . 121, 136
- H**
- has-application-volume* 102, 104, 113,  
 120, 146  
*has-device-volume* . 102, 112–120, 146  
*has-graphics-benchmark* . . . . . 102,  
 112–120  
*has-id* . . . . . 101, 104, 112–122, 148  
*has-networking-benchmark* . . . . 101,  
 112–120  
*has-output-media-list* . . 105, 108–110,  
 117, 119  
*has-resolution* 102, 104, 112–117, 120,  
 121, 148  
*has-room-id* . . . 101, 114–117, 120–122,  
 136, 146  
*has-screen-location* 101–107, 112–117,  
 120, 121, 147, 268  
*has-size* . . 101–107, 112–116, 120, 121,  
 268  
 Histogramm . . . . . 168, 210  
 Home-Entertainment  
 System . . . . . 140, 159, 187  
 Umgebung . . . . . 76, 105
- I**
- Intervall-Skala . . . . . 210  
 Intervallskala . . . . . 168  
*is-playing* . . . 95, 96, 101, 105, 112, 116,  
 120, 147, 268  
*is-processing-input* 104, 115, 122, 151  
*is-visible* . . 96, 101–107, 112–121, 147,  
 268
- K**
- Kognitive Last . . . . . 46, 187, 194  
 Konfidenzintervall . . . . . 168, 210  
 Kruskal-Wallis-Test . . . . . 209
- L**
- Laborstudie . . . . . 187  
 abhängige Variable . . . . . 193  
 Akzeptanz . . . . . 187  
 Ausgabeschritt . . . . . 202  
 Balancierung . . . . . 205  
 Bewertung erinnerte Inhalte . . . 208  
 BS . . . . . 187, 204  
 BSC . . . . . 187, 204  
 Dauer der Betrachtung . . . . 193, 221  
 Effizienz . . . . . 187, 227  
 Fragebogen . . . . . 204  
 Gerätekombination . . . . . 187, 197  
 Hypothese . . . . . 191, 246  
 Moderatorvariable . . . . . 194  
 Nachrichtenbeitrag . . . . . 195  
 Nachrichtenfolge . . . . . 205

## SACHWORTREGISTER

- PDA-TV ..... 187  
 PDA-TV-Bilderrahmen ..... 187  
 Präsentation ..... 199  
 Segment ..... 197, 202  
 Störfaktor ..... 208, 240  
 Tastatur-TV ..... 187  
 Tastatur-TV-Bilderrahmen ..... 187  
 TB ..... 187, 202  
 unabhängige Variable ..... 194  
 Versuchsleiter ..... 206  
 Zeitpunkt der Präsentation ..... 205  
 Likert-ähnlich ..... 164, 193, 209  
 Loess-Interpolation ..... 210, 217
- M**
- Mann-Whitney-Test ..... 168, 209  
 Mbrola ..... 145, 164, 191, 204, 246  
*media-format* ..... 105, 120, 147  
*media-url* ..... 105, 120, 122, 147  
 Median ..... 210  
 Microsoft Agent ..... 106, 115, 128  
 Mittelwert ..... 168, 210  
 MMF-Komponente . *siehe* Multi-modales Dialogsystem  
 MMP-Komponente . *siehe* Multi-modales Dialogsystem  
 Modalität ..... 45–46, 61–64, 72–74  
   arbiträr ..... 46, 72–74  
   auditoriell ..... 45–46, 61, 72–74  
   Code ..... 45–46  
   linguistisch ..... 46, 72–74  
   Medium ..... 45–46, 72–74  
   statisch/dynamisch ..... 46, 72–74  
   taktil ..... 45–46, 61, 72–74  
   visuell ..... 45–46, 61, 72–74  
 Modalitäteneffekt ..... 47, 190  
 Modalitätentheorie ..... 61–64  
   Unimodalität . 61, 94, 97, 100, 104, 105, 135  
   Unimodalitätenmodell ..... 61–62, 72–74, 89, 100, 104, 106, 111, 112, 114, 136, 137, 147–149, 151, 268  
 Multi-modale Ausgabe ..... 38
- Abhängigkeiten ..... 81  
 auf einem Gerät ..... 46  
 empirische Untersuchungen . 46, 52  
 Generierung ..... 69  
 in verteilter Geräteumgebung ... 51, 79, 81, 83, 86  
 Modellierung ..... 60  
 numerischer Ansatz ..... 50  
 technische Umsetzung ..... 48, 56  
 template-basierter Ansatz ..... 48  
 wissensbasierter Ansatz ..... 48  
 Multi-modale Eingabe .... 38, 67, 69, 79  
 Multi-modale Fission *siehe* Multi-modale Ausgabe  
 Multi-modale Fusion *siehe* Multi-modale Eingabe  
 Multi-modale Präsentation ..... *siehe* Multi-modale Ausgabe  
 Multi-modales Dialogsystem ..... 36–40  
   auf einem Gerät ..... 67  
   Datenfluss ..... 83  
   in verteilter Geräteumgebung ... 79  
   MMF-Komponente ..... 67, 79, 84  
   MMP-Komponente . 68, 69, 79, 84, 133, 146  
 Multi-Monitor-System . *siehe* Gekoppelte Ausgabegeräte  
 MW ..... *siehe* Mittelwert
- N**
- Nominalskala ..... 167, 209  
 Normalverteilungsbedingung ..... 167
- O**
- Ordinalskala ..... 167, 209
- P**
- Parameterfreie Verfahren ..... 168, 209  
 PDA ..... 75, 133, 140, 159, 187, 208  
 Perzentil ..... 210  
*player* 105, 108, 110–113, 115–122, 127, 147, 148  
   multi-modale Parameter ..... 105  
 Präsentation

## SACHWORTREGISTER

- Aufgabe . . . . . 86–92, 123  
Historie . . . . . 86–92, 123  
Kontext . . . . . 86–92, 123  
Planung *siehe* Präsentationsplanung  
Problem . . . . . 86–92  
Präsentationsplanung . 71, 122–128, 139,  
152–157  
Ablauf . . . . . 126  
Allen-Intervallalgebra . . . . . 125  
Constraints . . . . . 71, 125  
inkrementell . . . . . 129–130, 155  
kreuzmodale Referenzen . 130–131,  
139–140  
Operator . . . . . 71, 123, 152–157  
Planungsansatz . . . . . 71, 122–128  
Planungsbaum . . . . . 153  
Planungsproblem . . . . . 71, 124  
Prototypimplementierung . . . . . 133
- R**  
*ref-window-location* . . . . . 112, 113,  
115–118  
Rendering-Komponente . . . . . 68, 84–93,  
97–100, 135, 153, 267
- S**  
*screen-location* . . . . . 101, 102, 104–107,  
112–121  
SD . . . . . *siehe* Standardabweichung  
Signifikanzniveau . . . . . 168, 210  
*size* . . . . . 102, 105, 110, 112–117, 120, 121  
SMIL . . . . . 48, 49, 60, 124, 140, 141, 208  
SMIL-Ausgabekomponente . . . . . 122  
Selbstbeschreibung . . . . . 120, 147  
Split-Attention-Effekt . . . . . 47  
Sprachausgabekomponente 121, 142, 159,  
191  
Selbstbeschreibung . . 137, 148, 152  
Spracherkennungskomponente . 122, 145,  
163  
Selbstbeschreibung . . . . . 122, 151  
Sprachsynthese . . . . . *siehe*  
Sprachausgabekomponente  
Standardabweichung . . . . . 168, 210
- Streudiagramm . . . . . 210, 217
- T**  
t-Test . . . . . 167, 210  
Teilnehmer . . . . . 164, 195  
Textgenerierung . . . . . 48, 68, 79, 133  
*throw-input-event* . 104, 114, 121, 122,  
151  
TU Darmstadt Charakter . . . . . 117, 135
- U**  
Ubiquitäres Informationssystem . . . . . 33  
empirische Untersuchungen . . . . . 54  
Home-Entertainment . . . . . 55  
Unimodalitätenmodell . . . . . *siehe*  
Modalitätentheorie
- V**  
Varianzanalyse . . . . . 167, 210  
*visibility* . 102, 105, 112, 114, 116, 117,  
120, 121  
visuell . . . . . *siehe* Modalität  
*volume* . . . . . 94, 102  
Vp . . . . . *siehe* Teilnehmer
- W**  
Wilcoxon-Test . . . . . 168, 209
- X**  
Xsmiles . . . . . 49, 144

# Lebenslauf



## Persönliche Daten

geboren am: 29.04.1973  
Geburtsort: Herne  
wohnhaft in: Heidelberg  
Familienstand: in fester Partnerschaft  
Staatsangehörigkeit: Deutsch

## Ausbildung

|             |                                                                                                                                                                                                 |
|-------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 09/79-08/83 | Grundschule in Bochum                                                                                                                                                                           |
| 09/83-08/92 | Hildegardis Gymnasium in Bochum<br>Abschluss: Abitur                                                                                                                                            |
| 09/92-11/93 | Zivildienst Augusta Krankenhaus, Bochum                                                                                                                                                         |
| 10/93-09/99 | Studium der Informatik an der RWTH Aachen<br>- Nebenfach: Physik<br>- Abschluss: Dipl.-Informatiker<br>- Dipl. Arbeit: „Wortgrenzenmodellierung und Suche in der automatischen Spracherkennung“ |

## Berufspraxis

|             |                                                                                                                                                                                            |
|-------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 02/00-09/06 | Firma European Media Laboratory GmbH, Heidelberg<br>- Position: wissenschaftlicher Mitarbeiter<br>- Bereiche: adaptive Benutzerschnittstellen, eingebettete Systeme, Multimedia, Usability |
| 10/06-12/06 | Stipendium Klaus Tschira Stiftung gGmbH, Heidelberg zur Fertigstellung der Promotion                                                                                                       |
| 02/07-05/07 | Firma Pentamino GmbH, Heidelberg<br>- Position: IT-Consultant<br>- Bereich: Marktdatensysteme                                                                                              |
| seit 06/07  | Max-Planck Institut für Astronomie, Heidelberg<br>- Position: wissenschaftlicher Mitarbeiter (Postdoc)<br>- Bereich: Automatisierte Klassifikation von astrophysikalischen Objekten        |

Heidelberg, den 14. Mai 2008