Darstellung von Emotionen als Repräsentation von Dialogerfolg

Diplomarbeit am Institut Interactive Systems Labs
Prof. Dr. Alex Waibel
Fakultät für Informatik
Universität Karlsruhe (TH)

von

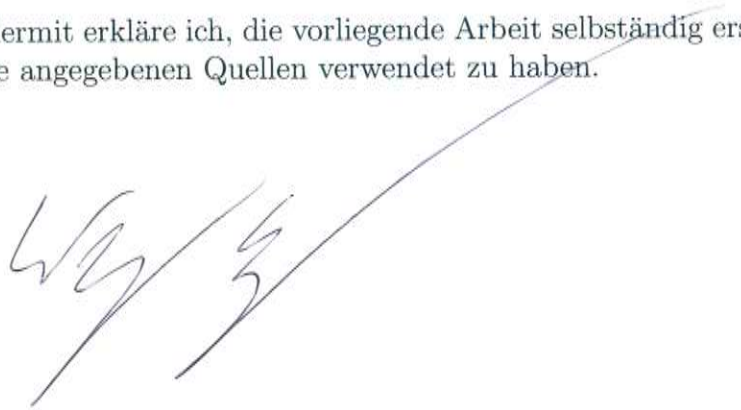
cand. inform.
Wenzel Svojanovsky

Betreuer:
Prof. Dr. Alex Waibel
Dipl.-Inform. Hartwig Holzapfel

Tag der Anmeldung: 1. Juni 2006
Tag der Abgabe: 30. November 2006

Erklärung

Hiermit erkläre ich, die vorliegende Arbeit selbständig erstellt und keine anderen als die angegebenen Quellen verwendet zu haben.



Karlsruhe, 30.11.2006
Wenzel Svojanovsky

Kurzfassung

Die Erforschung und Entwicklung humanoider Roboter machte in den letzten Jahren große Fortschritte. Um sich besser in den menschlichen Alltag integrieren zu können und um einen Eindruck von Lebendigkeit zu erwecken, werden diese Roboter mit künstlichen Emotionen versehen. Studien im Bereich des *"Affective Computing"* zeigten, dass emotional handelnde Maschinen auf den Menschen angenehmer wirken und somit die *Effizienz der Mensch-Maschine Kommunikation steigern* kann.

In dieser Arbeit wird untersucht, ob sich diese Effizienzsteigerung auch auf das *Lernverhalten des Menschen* auswirkt. In einem multimodalen Dialog kann die Darstellung von Emotionen als Repräsentation von Dialogerfolgen verwendet werden. Hierfür wird ein Roboter um die Fähigkeit Emotionen ausdrücken zu können erweitert. Der Benutzer erhält durch dieses *emotionale Feedback* einen Eindruck davon, wie gut er das System bedient und kann sein Verhalten entsprechend anpassen.

Es wird der Entwurf eines emotional agierenden Avatar beschrieben. Dieser stellt eine Erweiterung zu einem bestehenden Roboter dar, der ebenfalls vorgestellt wird. *EmoMaps*, ein in diesem Zusammenhang entwickeltes Emotionssystem, bestimmen die Emotion und wirken sich auf die gesprochene Sprache und die Mimik des Avatar aus.

In einer Benutzerstudie werden abschließend Probanden in einem Barkeeper-Szenario mit diesem Roboter kommunizieren. Dabei sollen sie die Benutzung des Systems intuitiv erlernen. Die Auswirkung des emotionalen Feedbacks auf die Leistung des Probanden wird analysiert und Rückschlüsse auf dessen Lernverhalten gezogen.

Danksagung

Als erstes möchte ich mich bei Hartwig Holzapfel für die erstklassige Betreuung bedanken. Ohne seine Geduld und seine wertvollen Ratschläge hätte diese Arbeit nicht so umfangreich und interessant werden können. Ich danke ihm auch für die humorvollen Gespräche, die immer wieder aufs Neue zur Motivationen beitrugen.

Natürlich gilt mein Dank auch Professor Alexander Waibel, durch den diese Arbeit erst ermöglicht wurde.

Ich möchte mich auch bei Satoshi Nakamura und Nick Campbell bedanken, die mich zuvor auf meinem akademischen Werdegang begleiteten. In diesem Zusammenhang möchte ich auch Rainer Gruhn hervorheben, von dem ich viel lernen konnte.

Danke auch an alle Mitarbeiter des Lehrstuhls für ihre Unterstützung in jeglicher Hinsicht.

Meiner Familie, vor allem meinen Eltern, möchte ich ebenfalls an dieser Stelle danken. Sie ermöglichten mir erst das Studium und gaben mir auch sonst Rückhalt.

Folgende Personen begleiteten mich über die letzten Monate und unterstützen mich auf fachliche aber auch auf menschliche Art und Weise: Adrian, Alex, Andrea, Andreas, Anni, Annika, Armin, Benjamin, Bernhard, Boris, Carina, Carmen, Christian, Christoph, Clara, Clauia, Cornelia, Dagmar, Dennis, Dieter, Eva, Evelyn, Felipe, Frank, Franziska, Heiko, Henni, Holger, Ikarus, Jan, Johanna, Johannes, Julia, Kachun, Kai, Kirsten, Ko, Lena, Marcel, Markus, Martin, Michael, Nico, Nicole, Sebastian, Simon, Stefan, Thomas, Till, Timea und Tobias.

Die Zeit während der Diplomarbeit ist sehr anstrengend und kann in bestimmten Phasen sehr arbeits- und zeitintensiv sein. Deswegen möchte ich hier ein großes Dankeschön an meine Freundin Anja Gohl aussprechen. Sie hatte sehr viel Geduld mit mir und schaffte es stets mich auch in schwierigen Phasen zu motivieren.
Danke!

Zu guter Letzt will ich meinem Computer Annika danken, die mir diese lange Zeit über treu zur Seite stand. Sie funktionierte Tag und Nacht einwandfrei und ließ mich auch in der heißen Endphase dieser Arbeit nicht im Stich. (Siehe hierfür *The Media Equation* [ReNa03])

Inhaltsverzeichnis

Abbildungsverzeichnis

Tabellenverzeichnis

1. Einleitung

1.1 Motivation

Die Besucher der E3 2006[1] waren hellauf begeistert. Endlich wurde eine Innovation präsentiert; eine Innovation worauf die Spielewelt seit Jahren wartete. Etliche Menschen wurden in den Bann einer virtuellen Figur gezogen, die der französisch-amerikanischen Schauspielerin Aurelie Bancilhon nachempfunden ist.

Abbildung 1.1: Bilder aus dem E3-Trailer zu Heavy Rain von Quantic Dream

Der Chefredakteur des Online-Magazins *www.4Players.de* Jörg Luibl [JLui06] schrieb in seinem Bericht über die Spielemesse:

> [...] Kein anderer E3-Trailer konnte mich so faszinieren wie diese weibliche Charakterstudie. Diese letzten Sekunden mit Tränen in den Augen und dem Revolver in der Hand waren das Beste, was ich bisher aus Los Angeles gesehen habe [...]

Die Rede ist vom Trailer zum Videospiel *Heavy Rain* von Quantic Dream. Im Gegensatz zu den meisten anderen Trailern der E3 werden hier keine schmierigen Blutlachen, realistische Schatteneffekte oder atemberaubende Physiksimulationen gezeigt. Hier wird ein virtuelles Casting vorgespielt, in dem eine Figur namens Mary

[1] Electronic Entertainment Expo in Los Angeles

Smith eine emotionale Szene eindrucksvoll nachspielt. Abb. 1.1 und Anhang D zeigen Bilder von Mary in unterschiedlichen Posen mit jeweils verschiedenen Emotionen. In einem Artikel über die 10 wichtigsten Erkenntnisse dieser Messe [MOer06] im selben Magazin wurde unter der Überschrift *Der Wunsch nach emotionaler Tiefe* folgender Text veröffentlicht:

> [...] Es scheint nicht nur den Wunsch nach kreativen Ideen, sondern auch den nach mehr Emotionalität zu geben: Die Charakterstudie Heavy Rain von Quantic Dream [...] konnte sich von null auf hundert in die Herzen der Trailerfans spielen. [...] Dieser Film bewegte und inspirierte, weckte bei vielen die Hoffnung auf erwachsenere, erzählerisch anspruchsvollere und nachhaltig beeindruckendere Spiele. [...]

Dieses Spiel löste in der Spielewelt eine wahre Euphorie aus. Endlich würde die Technik aktueller Rechner nicht nur zur Darstellung immer realistischerer Physiksimulationen, Licht- und Schatteneffekte und detaillierte Animationen benutzt werden. Es wird in die virtuellen Welten Leben eingehaucht. Durch die realistischere Darstellung von Emotionen erhofft sich die Spielergemeinde mehr spielerische Tiefe. Virtuelle Charaktere, die überzeugend emotional handeln, lassen eine stimmigere Atmosphäre aufkommen. Das Spiel wird faszinierender, der Spieler versetzt sich tiefer in das Geschehen, die Virtualität wirkt realistischer und plastischer. Dadurch eröffnet sich dem Spieler eine völlig neue Spielerfahrung.
Aber nicht nur die Welt der Videospiele horchte auf, der englischsprachige Wikipediaartikel [Wiki06b] zu Moris *Uncanny Valley* (siehe Seite 3) wurde aktualisiert, so dass nun auch das Spiel Heavy Rain dort erwähnt wird.

Was die Spieleindustrie erst jetzt für sich entdeckt hat, wird in der Forschung der Mensch-Maschine Interaktion schon länger untersucht. Bereits 1995 veröffentlichte Picard das Paper *Affective Computing* [Pica95] in dem dargelegt wird, dass Emotionen ein wichtiger Bestandteil der menschlichen Wahrnehmung darstellen. Um menschlichere Maschinen zu schaffen, die sich im Alltag problemlos integrieren können, sind Emotionen demnach unausweichlich.

Dass Menschen euphorisch auf virtuelle Emotionen reagieren, zeigt dass diese Emotionen einen gewissen Einfluss auf sie haben. Es ist daher interessant zu untersuchen, wie man diese Emotionen in der Mensch-Maschine Kommunikation einsetzen kann um bessere Systeme zu schaffen.
Das Thema „Emotionen" ist sehr schwierig, da sehr viele individuelle und psychische Faktoren des Menschen mit einfließen. Es fehlt an objektiven Bewertungskriterien, um unterschiedliche Ergebnisse besser vergleichen zu können. Um aber einen Roboter zu bauen, der sich in die menschliche Welt integriert, darf man keine Scheu vor diesem Thema haben.
In dieser Arbeit soll ein emotionaler Roboter gebaut werden, der die Emotionalität für effizientere Dialoge zwischen Mensch und Maschine einsetzt.

In den folgenden Kapiteln wird der Bau und Einsatz dieses Systems beschrieben. Hierbei wird zuerst auf allgemeine Aspekte emotionaler Avatare eingegangen. Danach werden die Grundlagen und der aktuelle Stand der Forschung dargestellt. Es folgt eine technische Vorstellung des Systems. Anschließend werden die für eine Benutzerstudie notwendigen Experimente erklärt, deren Ergebnisse anschließend diskutiert werden.

1.2 Herausforderung

Der in 1.1 vorgestellte Trailer zeigt eine emotional agierende Frau. Hier soll aber das Wort „emotional" nicht im negativen Sinne interpretiert werden, was den Verlust von Rationalität beinhaltet. Gemeint ist vielmehr die Fähigkeit Emotionen ausdrücken zu können. Dieser Punkt birgt einige Herausforderungen.

Ist man nun in der Lage, diese Fähigkeit auf Roboter zu übertragen, so begibt man sich in einen Forschungsbereich, der als *Soziale Robotik* bezeichnet wird. Einen kurzen Überblick über den aktuellen Stand der Forschung und deren Gebiete ist in der Ausgabe 2/2006 des c't Magazins im Artikel *Der Roboter als Menschenfreund* von Jutta Weber nachzulesen [Webe06]. Das Ziel dieses Bereichs ist es freundliche Maschinen zu entwickeln, die als *verständige und glaubhafte Interaktionspartner* mit dem Menschen natürlich kommunizieren.
Mit dem sehr hoch entwickelten Roboter Repliee Q2 konnte der Wissenschaftler Hiroshi Ishiguro bereits interessante Ergebnisse erzielen. Dieser Android ist äußerlich einer japanischen Nachrichtensprecherin nachempfunden und vermittelt durch Mikrobewegungen wie Lidschlag oder Atmung einen natürlich wirkenden Eindruck. Dem BBC sagte Ishiguro [WhIs05]:

> We have found out that people forget she is an android while interacting with her. (dt.: Wir fanden heraus, dass Menschen vergessen, dass sie ein Android ist, während sie mit ihr interagieren.)

Beim Entwurf eines humanoiden Roboters tritt ein großes Problem auf, das als *Uncanny Valley* (dt.: Unheimliches Tal) bekannt ist. Mit dem Ziel einen akzeptierten und nicht abschreckenden Roboter zu entwickeln, muss diese Hürde erst überwunden werden.

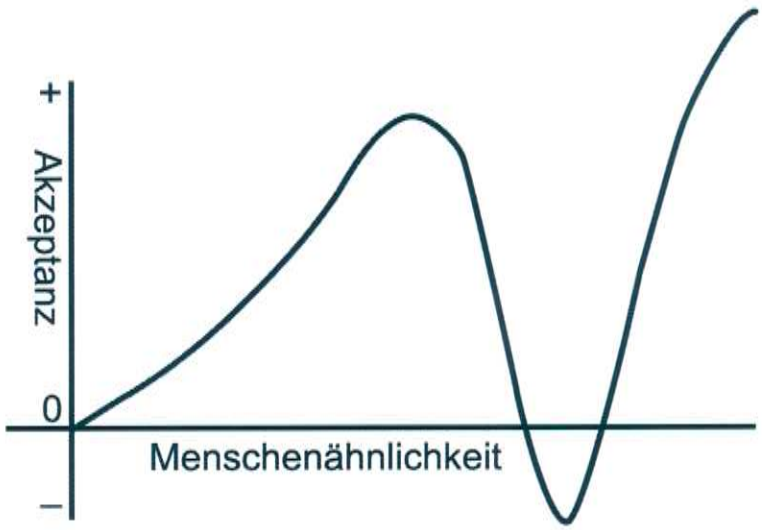


Abbildung 1.2: Das Unheimliche Tal von Mori: Die Akzeptanz eines Roboters steigt nicht linear mit der Ähnlichkeit zum Menschen

Die Entdeckung dieses unheimlichen Tals geht auf Masahiro Mori [MMor70] zurück. Es besagt, dass Menschen unterschiedlich auf Roboter reagieren. Die Akzeptanz eines Roboters verhält sich nicht linear zur Ähnlichkeit zum Menschen. Ab einem gewissen Punkt nimmt der positive Eindruck wieder stark ab und geht sogar ins Negative

über. Dieser Punkt der Verstörtheit tritt z.B. bei künstlichen Körperteilen (Prothese) oder leblos wirkenden Gestalten (Zombie) auf. Erst nach diesem Punkt nimmt die Vertrautheit und Akzeptanz des Wesens wieder zu. Abb. 1.2 zeigt schematisch dieses Verhalten der menschlichen Wahrnehmung.

Abbildung 1.3: Das unheimliche Tal angewandt auf Beispielroboter

In Abb. 1.3 wird das Schema des unheimlichen Tals auf konkrete Beispiele angewandt; links ein Bild von Robovi, rechts ein menschliches Kind.

Die Akzeptanz der Maschine muss sich aber nicht zwingend aus Nachahmungen der Realität ergeben. Joseph Bates beschreibt in *The Role of Emotion in Believable Agents* [Bate94] verschiedene Wege einen glaubwürdigen Roboter zu gestalten. In diesem Fall wird aber „glaubwürdig" als „lebendig wirkend" angesehen.
In Animationsfilmen, wie beispielsweise Disney-Filmen, sind es selten Menschen, die die Hauptrollen spielen. Trotzdem wird durch die Handlung, aber vor allem auch durch die Animationstechnik, Leben in die Charaktere eingehaucht. Dieser Eindruck bringt Menschen zum Lachen oder sogar zum Weinen. Obwohl den Figuren äußerlich das menschliche Aussehen fehlt, weist man ihnen dennoch menschliche Eigenschaften zu. Man ist bereit sich mit den Geschöpfen zu identifizieren oder sie auch zu Gefährten oder Haustieren zu machen, mit denen man einen Teil seines Lebens verbringen würde. Genau das macht *Glaubwürdigkeit* aus. Wie dieser Eindruck durch Zeichentechnik erzeugt werden kann, beschreiben zwei Disney-Zeichner in ihrem Buch *The Illusion of Life* [JoTh81].
Johnston formuliert drei Regeln, die zu mehr Glaubwürdigkeit beitragen:

- Der emotionale Zustand des Charakters muss klar definiert sein.
- Gedankenprozesse offenbaren die Gefühle.
- Gelegentliche Akzentuierungen tragen zum Verständnis bei. Der Zuschauer braucht aber auch Zeit um diese Momente auskosten zu können.

Anhand dieser Regeln entwarf Bates glaubwürdige Avatare (die Woggles), die lediglich aus einer verzerrbaren Kugel mit Augen bestehen. In einer Studie wurde dann die Glaubwürdigkeit dieser Avatare untersucht. Durch einen Programmierfehler tickte ein Woggle gelegentlich aus, was die Benutzer als Teil seiner Persönlichkeit interpretierten und diesen auch als am glaubwürdigsten einstuften.
Es kann demnach auch ohne Ähnlichkeit zum Menschen der Eindruck von Lebendigkeit glaubwürdig vermittelt werden. Diese Mechanismen bezeichnet Bates als *alternate Artificial Intelligence.*

Clifford Nass belegte in einer Untersuchung zur Mensch-Computer Interaktion [NaSu94], dass Menschen durchaus dazu bereit sind Gegenständen Emotionen zuzuordnen. Oft gehen Menschen mit Gegenständen wie Computern emotional um, obwohl sie wissen, dass diese keine Gefühle besitzen. Vermittelt dieser Gegenstand zusätzlich Gefühle, so wirkt sich das auf die Kommunikation zwischen Mensch und Computer aus.

Es ist mittlerweile bekannt, dass Roboter durchaus natürlich wirken können, diese Natürlichkeit aber auch durch andere Methoden vermittelt werden kann. Außerdem reagieren Menschen auch anders auf emotionale Maschinen als auf neutrale. Vereint man diese Techniken, so kommt man Androiden, die mit Menschen interagieren und sich in deren Alltag auf natürliche Weise einfügen, schon sehr nahe.
Sollen diese Maschinen aber nicht unangenehm auffallen, so müssen sie auch über soziale Kompetenzen verfügen. Der Begriff der *Emotionalen Intelligenz* geht auf Salovey und Mayer zurück, wurde aber erst durch das gleichnamige Buch von Daniel Goleman [Gole95] populär.
Emotionale Intelligenz wird hier als übergeordnete Fähigkeit angesehen, von der es abhängt wie gut ein Mensch seine sonstigen Fähigkeiten, darunter auch den Verstand, zu nutzen versteht. Sie besteht dabei aus fünf Teilkonstrukten:

Selbstbewusstsein: Fähigkeit eines Menschen, seine Stimmungen, Gefühle und Bedürfnisse zu akzeptieren und zu verstehen und deren Wirkung auf andere einzuschätzen.

Selbstmotivation: Begeisterungsfähigkeit für die Arbeit; die Fähigkeit sich selbst unabhängig von finanziellen Anreizen oder Status motivieren zu können.

Selbststeuerung: Planvolles Handeln in Bezug auf Zeit und Ressourcen.

Soziale Kompetenz: Fähigkeit, Kontakte zu knüpfen, tragfähige Beziehungen aufzubauen und zu halten.

Empathie: Fähigkeit, emotionale Befindlichkeiten anderer Menschen zu verstehen und angemessen darauf zu reagieren.

In Bezug auf den Forschungsbereich *Affective Computing* ist die eben vorgestellte *Empathie* hervorzuheben. Denn dies ist genau der Punkt, den es weiter zu erforschen gilt. Sowohl in der Emotionserkennung als auch in deren Verständnis und Verarbeitung ist noch ein weiter Weg zu gehen.

Brian Duffy warnt in seinem Bericht *Why Humanoids?* [DuJo05] vor einer voreiligen Euphorie. Die Erwartungen des Benutzers hängen stark vom äußeren Eindruck des

Systems ab. Wirkt ein humanoider Roboter sehr natürlich, so weckt dies auch Erwartungen an eine hohe Intelligenz und Mobilität. Dass diese Erwartungen zum gegenwärtigen Zeitpunkt nicht erfüllt werden können, beweist beispielsweise der Android *Repliee Q2*, der trotz sehr natürlich wirkender menschlicher Gestalt nicht stehen oder gehen kann. Dies enttäuscht und der Roboter verliert seine Glaubwürdigkeit.

1.3 Zielsetzung

Die Fähigkeit Emotionen ausdrücken zu können macht eine Maschine zu einem effizienteren Gesprächspartner. Emotionen helfen dabei mögliche Reaktionen des Partners berechenbarer zu machen und im Gespräch darauf einzugehen. Dazu gewähren sie einen kleinen Einblick in den Denk- und Fühlprozess des Gegenübers. Die Kommunikation wird dadurch vereinfacht.

Ein ausgereifter humanoider Roboter mit der Fähigkeit Emotionen auszudrücken sollte demnach die Effizienz der Kommunikation steigern. Bedingung ist allerdings, dass dieser als vollwertiger Partner akzeptiert wird.

Nun kann man vermuten, dass sich diese Effizienzsteigerung auch auf das Lernverhalten des Menschen auswirkt. Häufig wiederholen Menschen bei Interaktionen mit einer Maschine die gleichen Fehler wieder und wieder. Die Maschine gibt eine Fehlermeldung aus, die sich bei jedem gleichen Fehler aufs Neue wiederholt. Würde die Maschine nun beim wiederholten Male anders reagieren, könnte dies den Benutzer veranlassen seine Eingabe zu ändern.
Dieses Konzept ist mit einem emotionalen System umsetzbar. Eine Fehlermeldung würde Einfluss auf die aktuelle Emotion des Roboters ausüben. Die Reaktion bei einem Fehler könnte demnach von Emotion zu Emotion unterschiedlich sein. In einem intuitiv erlernbaren System kann es also durchaus sein, dass Emotionen das Erlernen beschleunigen. Durch emotionales Feedback wird der Benutzer ständig über Erfolg und Misserfolg in Kenntnis gesetzt. Er kann also nachvollziehen ob er vom System verstanden wurde oder ob die letzte Eingabe fehlerhaft war.

Hieraus ergibt sich die zentrale Fragestellung dieser Arbeit: *Haben Emotionen Einfluss auf die Erlernbarkeit eines Systems und inwieweit können Emotionen den Benutzer in einem Dialog führen?*

Um dieser Frage auf den Grund zu gehen, wird im Zuge dieser Arbeit ein emotionaler Roboter gebaut werden. Der Benutzer wird in einem Barkeeper-Szenario einen Dialog mit dem Roboter durchführen. Dabei soll er diesen Dialog während der Benutzung erlernen. Durch die Experimente wird das Lernverhalten untersucht werden. Die Dialogentscheidungen sollen dabei unabhängig von der Emotion des Systems sein. Die Emotionen selbst dienen als Repräsentation von Dialogerfolgen und haben Einfluss auf die Art und Weise wie sich der Roboter äußert. Bei häufigen Misserfolgen wird das System die Fragen um mögliche Antworten erweitern um dem Benutzer „auf die Sprünge zu helfen“. Wird das System dagegen gut bedient, sollen die Fragen kürzer ausfallen, um Zeit zu sparen.

1.4 Vorbemerkungen

URLs:

Alle in dieser Arbeit verwendeten *URLs* sind am 24. November 2006 um 18:00 (MEZ) überprüft worden. Eine inhaltliche Korrektheit dieser wird nicht garantiert. Daher werden diese Artikel als Zusatzinformation angesehen und dienen nicht als Grundlage für wissenschaftliche Argumentationen.

Übersetzungen:

Es wird versucht *alle Wörter und Zitate ins Deutsche zu übersetzen.* Dies ist allerdings nicht immer möglich. Zudem stellen manche englische Begriffe auch Eigennamen dar. In solchen Fällen gibt es keine Übersetzung.

Die Wörter „affektiv" und „emotional":

Das Wort „affektiv" bezieht sich auf den Begriff „Affective Computing" und wird in dieser Arbeit auch als „emotional" verwendet. Dabei ist mit „emotional" die Fähigkeit Emotionen auszudrücken gemeint und nicht das menschliche Verhalten, das einen Verlust von Rationalität bedeutet.

Affekte:

Das Wort „Affekt" bedeutet in dieser Arbeit eine *kurze spontane Emotionsäußerungen.* Dabei ist nur gering Bezug auf „affektiv" zu nehmen, was in diesem Zusammehang die Fähigkeit Emotionen ausdrücken zu können bedeutet.

Die Begriffe „Android" und „humanoider Roboter":

Die beiden Begriffe „Android" und „humanoider Roboter" werden in dieser Arbeit als gleichbedeutend angesehen.

„Proband", „Teilnehmer" und „Benutzer":

Die „Teilnehmer" an der, in dieser Arbeit durchgeführten Studie werden auch als „Probanden" bezeichnet. Beide Begriffe sind demnach gleichbedeutend. „Benutzer" sind Personen, die auch außerhalb dieser Studie mit dem Roboter kommunizieren, wie es in einer späteren praktischen Anwendung sein kann.

Turn:

Ein *Turn* ist genau ein Schritt im Dialog. Also die Zeitspanne von der Frage des Avatar bis zur Reaktion auf die Antwort des Benutzers (siehe Abschnitt 5.3.1).

2. Grundlagen

2.1 Emotionen

Das dritte Hauptwerk von Immanuel Kant aus dem Jahre 1790 trägt den Titel *Kritik an der Vernunft*. Hier formuliert er, dass das Gefühl der Lust und der Unlust das Mittelglied zwischen Erkenntnisvermögen und Begehrungsvermögen ist. Er sieht dabei Emotionen als Feinde der Vernunft an.
Das menschliche Leben funktioniert jedoch nicht ohne Emotionen. In der Entscheidungsfindung spielen sie eine wichtige Rolle. Der Satz „Entschuldige bitte, ich habe nicht nachgedacht!" klingt jedem vertraut, einen ähnlichen Satz „Entschuldige bitte, ich habe nicht gefühlt!" hört man jedoch selten. Laut Izard [Izar93] sind Emotionen sowohl Motivation als auch führende Kraft im menschlichen Empfinden und der Wahrnehmung.

Rosalind W. Picard verfasste einen Überblick über die verschiedenen Bereiche der Emotionsforschung und deren mögliche Auswirkungen auf zukünftige Computersysteme. Ihr Werk *Affective Computing* [Pica95] gab diesem Forschungsbereich seinen Namen. Sie definiert diesen Begriff so.

> I call „affective computing," computing that relates to, arises from, or deliberately influences emotions.
> (dt.: Ich bezeichne „affective computing" als *Computing*, das sich auf Emotionen bezieht, daraus entsteht oder sie vorsätzlich beeinflusst.)

Bevor man Emotionen aber in Computern einsetzen kann, sollte man sich vorher klar machen, was Emotionen sind und welche es gibt.
Den Versuch Emotionen zu erklären und zu analysieren wurde laut Dieter Ulich [UlMa03] bereits von den alten Indern ca. 800 v.Chr. unternommen und konzentrierte sich auf die Formulierung „Alles Leben ist Leiden!".

Im Folgenden sollen nun Methoden der Modellierung von Emotionen vorgestellt werden und wie sich Emotionen auf den Mensch auswirken.

2.1.1 Auswirkung von Emotionen

Emotionen haben einen enormen Einfluss auf das Verhalten des Menschen. Dieses wurde in Abschnitt 1.2 ausführlich dargestellt. Die Mimik gibt dabei Hinweise über die aktuelle Emotion eines Charakters. Die Stimme und Sprache werden allerdings ebenfalls beeinflusst. Dies äußert sich z.B. durch Lautstärke, Sprachtempo und Intonation. Der Einfluss dieser Merkmale wird als *Prosodie* bezeichnet. Prosodie definiert Hiroya Fujisaki [FuCa96] wie folgt:

> Prosody is a systematic organization of various linguistic units into an utterance or a coherent group of utterances in the process of speech, and serves to convey not only linguistic information, but also paralinguistic and non-linguistic information.
> (dt.: Prosodie ist eine systematische Organisation verschiedener linguistischer Elemente in einer Äußerung oder zusammenhängenden Äußerungen im Sprachprozess. Sie dient nicht nur der Vermittlung linguistischer Information, sondern auch paralinguistischer und nicht-linguistischer Information.)

Mehr über Prosodie ist auch im Buch *Prosody and speech recognition* [Waib88] nachzulesen.

Bei Äußerungen, deren Bedeutung aus dem Wortlaut selbst nicht hervorgeht, wie auch bei Ironie, spielt Prosodie eine sehr wichtige Rolle. So kann anhand prosodischer Merkmale eine Bedeutung einer nicht eindeutigen Äußerung zugewiesen werden [SvGN04].

Prosodie wird laut Halliday [Hall70] durch drei Faktoren beeinflusst.

- Die emotionale Einstellung des Sprechers zum Hörer.
- Die emotionale Einstellung des Sprechers zum Inhalt des Gesagten.
- Der emotionale Zustand des Sprechers.

All diese Faktoren haben auch Einfluss auf die automatische Spracherkennung. Thomas Polzin [PoWa98] untersuchte dieses Phänomen und stellte fest, dass die Erkennungsrate (word accuracy) bei einem emotional gesprochenen Korpus etwa um 15% sinkt. Beachtet man allerdings Prosodie bei der Erkennung, so steigt die Erkennungsrate um 7%.
Außerdem konnte eine Emotions-Erkennungsrate von 70% erreicht werden, das in etwa der menschlichen Leistung entspricht.

2.1.2 Basisemotionen

Robert Plutchik [Plut80] behauptet, dass Emotionen eine biologische Funktion haben. Die Evolution hätte sie sonst abgeschafft. Sie stellen komplexe Ketten von Reaktionen mit stabilisierenden Rückmeldeschleifen dar, die eine gewisse Art von Homöostase des Verhaltens herstellen.
Plutchik zählt dabei acht *Basisemotionen* auf, die eine biologische Funktion haben

Emotion	Funktion
Furcht	Schutz
Ärger	Zerstörung
Freude	Fortpflanzen
Traurigkeit	Reintegration
Vertrauen, Akzeptieren	Einverleiben
Ekel	Zurückweisen
Erwartungen	Erkunden
Überraschung	Orientierung

Tabelle 2.1: Plutchik: 8 Basisemotionen und ihre Funktionen

und somit sowohl psychologisch als auch biologisch grundlegend sind. Diese Emotionen sind Tabelle 2.1 zu entnehmen.

Neben diesen Basisemotionen (primären Emotionen) gibt es auch sekundäre Emotionen die eine Kombination verschiedener primären Emotionen sind. Sie werden auch als *Dyaden* bezeichnet.
Die hier vorgestellten Basisemotionen werden oft als diskretes Emotionsmodell eingesetzt. Der Grund hierfür ist, dass durch die geringe Anzahl an Zuständen eine Implementierung sehr leicht ist. Zudem werden die Emotionen gleich mit einer Funktion gleichgesetzt. Das Verhalten einer Maschine ist somit direkt aus der Emotion ableitbar.

Paul Ekman [Ekma99] greift dieses Konzept auf und ordnet Basisemotionen drei grundlegende Bedeutungen zu:

- Die Emotionen unterscheiden sich durch wichtige Eigenschaften.
- Die Emotionen haben Einfluss auf lebenswichtige Aufgaben.
- „Basis“ steht für ein grundlegendes Element, aus dem sich neue kombinieren lassen.

2.1.3 OCC-Modell

Ein weiteres diskretes Modell ist das *OCC-Modell*, das von Ortony, Clore und Collins [OrCC88] entwickelt wurde. Es besteht aus 22 Emotionen, die nach Bedingungen und kognitiven Auswirkungen kategorisiert sind. Diese Kategorien sind abstrakte Klassen, die Zustände mit ähnlichen Eigenschaften gruppiert. Jeder Zustand wird durch Situation und Reaktion definiert. Tabelle 2.2 listet die Gruppen im OCC-Modell auf und zeigt die zu einer Gruppe gehörenden Emotionen.
Das OCC-Modell wird häufig in der Forschung eingesetzt, so basieren z.B. die in Abschnitt 1.2 vorgestellten Woggles von Bates [Bate94] auf diesem Modell. Es gibt allerdings auch einige Kritikpunkte. Christoph Bartneck [Bart02] bringt verschiedene Punkte auf. Ein Problem ist, dass durch die Komplexität des Modells ein enormer Aufwand bei der Implementierung notwendig ist. Ohne ein ausreichendes Kontext- oder Weltmodell ist diesem nicht zufriedenstellend lösbar. Die Theorie hinter dem Modell deckt sehr viel ab. Darüber, wie die Reaktionen dann tatsächlich aussehen, gibt es jedoch keine Hinweise. Ein ebenfalls großer Kritikpunkt ist die fehlende History-Funktion. Zustandsübergange sind absolut unabhängig vom aktuellen Zustand.

Gruppe	Beschreibung	Name und Emotionstyp
Well-Being	appraisal of a situation as an event	**joy:** pleased abount an event **distress:** displeased about an event
Fortunes-of-Others	presumed value of a situation as an event affecting another	**happy-for:** pleased about an event desirable for another **gloating:** pleased about an event undesirable for another **resentment:** displeased about an event desirable for another **sorry-for:** displeased about an event undesirable for another
Prospect-based	appraisal of a situation as a prospective event	**hope:** pleased about a prospective desirable event **fear:** unpleased about a prospective undesirable event
Confirmation	appraisal of a situation as confirming or disconfirming an expectation	**satisfaction:** pleased about a confirmed desirable event **relief:** pleased about a disconfirmed undesirable event **fears-confirmed:** displeased about a confirmed undesirable event **disappointment:** displeased about a disconfirmed desirable event
Attribution	appraisal of a situation as an accountable act of some agent	**pride:** approving of one's own act **admiration:** approving of another's act **shame:** disapproving of one's own act **reproach:** disapproving of another's act
Attraction	appraisal of a situation as containing an attractive or unattractive object	**liking:** finding an object appealing **disliking:** finding an object unappealing
Well-Being / Attribution	compound emotions	**gratitude:** admiration + joy **anger:** reproach + distress **gratification:** pride + joy **remorde:** shame + distress
Attraction / Attribution	compound emotion extension	**love:** admiration + liking **hate:** reproach + disliking

Tabelle 2.2: OCC-Modell

2.1.4 Arousal-Valence Modell

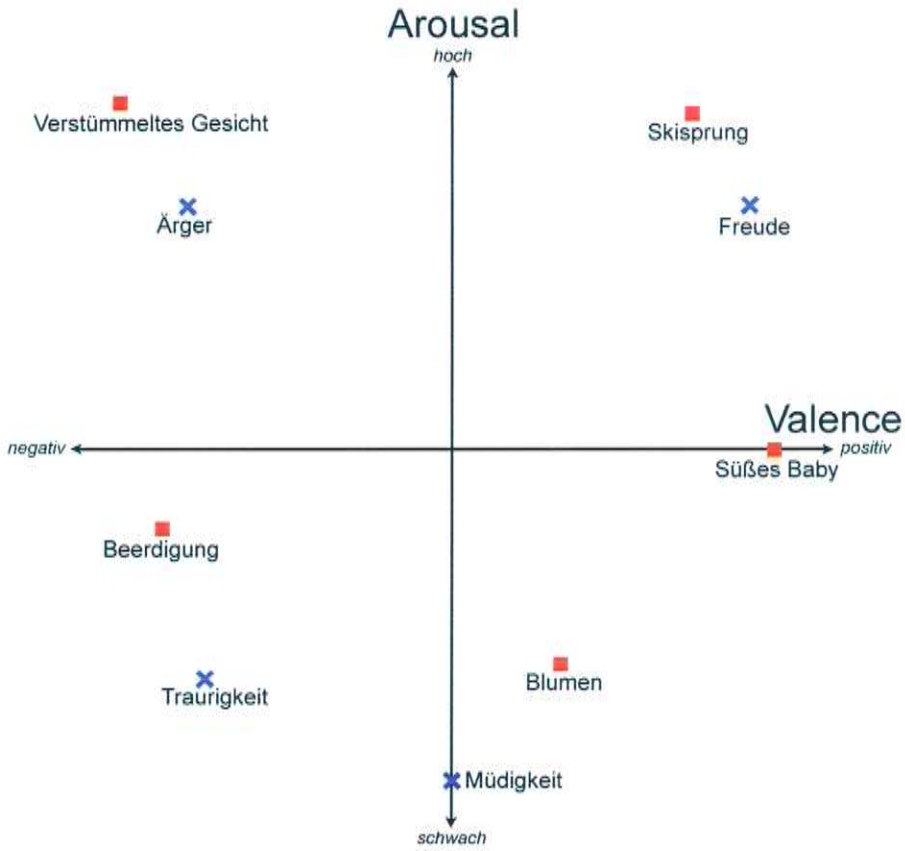


Abbildung 2.1: Arousal-Valence Modell

In Untersuchungen von Emotionen in der menschlichen Sprache entdeckte Bernd Tischer [Tisc93] drei Hauptmerkmale, die sich durch Emotionen verändern: *Intensität, Sprachtempo* und *Grundfrequenz*. Da diese Veränderungen in verschiedenen Graden auftreten, reicht eine diskrete Beschreibung nicht aus. Er schlägt daher ein kontinuierliches Emotionsmodell basierend auf den Dimensionen *Aktivität/Erregung, Potenz/Stärke* und *Valenz/Wohlbefinden* vor. Die oben genannten Sprachmerkmale ordnet Tischer bei Erregung ein.

Aktivität/Erregung: Der Erregungsgrad einer Person kann hauptsächlich durch die drei Merkmale Lautstärke, Sprachtempo und Grundfrequenz gemessen werden. Eine erregte Person spricht relativ zum gemäßigten oder mittleren Aktivitätsniveau bei vergleichsweiser hoher Grundfrequenz relativ schnell und lautstark. Diese Erregung wurde intensiv von Klaus Scherer [Sche86] untersucht.

Potenz/Stärke: Die Ergebnisse der Studien über Stärke in der vokalen Sprache gehen stark ausseinander. Es lässt sich nicht klar beschreiben, welche Merkmale sich durch die Stärke verändern. Man kann jedoch annehmen, dass die Potenzdimension bei der vokalen Kommunikation von Gefühlen nicht orthogonal zur Aktivitätsdimension steht. Es gibt zu viele akustische Merkmale, deren Ausprägung mit der Aktivitäts- und Potenzdimension korrespondieren.

Valenz/Wohlbefinden: Valenz wird auch häufig als „Wertigkeit" übersetzt und spiegelt sich bei positiven Werten in einer großen Grundfrequenzvariation wieder. Die Merkmalsausprägungen sind aber auch hier nicht immer eindeutig. Valenz unterscheidet sich im Vergleich zu den anderen Gefühlsdimensionen im Frequenzverlauf (eher fallend bei positiver Valenz), unspezifischer Tonhöhe und Wärme. Bei positiver Valenz werden außerdem leise Passagen mit eher hoher und laute Passagen mit tiefer Grundfrequenz gesprochen.

Peter Lang untersuchte ebenfalls diese Dimensionen und beschreibt dies in seinem Artikel *Emotion, Attention, and the Startle Reflex* [LaBC90]. Dort wird das *Arousal-Valence Modell* beschrieben, das auf zwei der drei oben genannten Dimensionen zurückgreift. Eine schematische Darstellung dieses Modells ist in Abb. 2.1 zu sehen. Rote Quadrate zeigen die Einordnung von Bildern, blaue Kreuze die Einordnung emotionaler Zustände.

Das Arousal-Valance Modell ist hier von besonderer Bedeutung, da es zu dem für diese Untersuchung entwickelten Emotionsmodell inspirierte (siehe Abschnitt 3.2.1.1).

2.2 Avatare und Roboter

In den letzten Jahren gab es große Fortschritte im Bereich der Robotik.

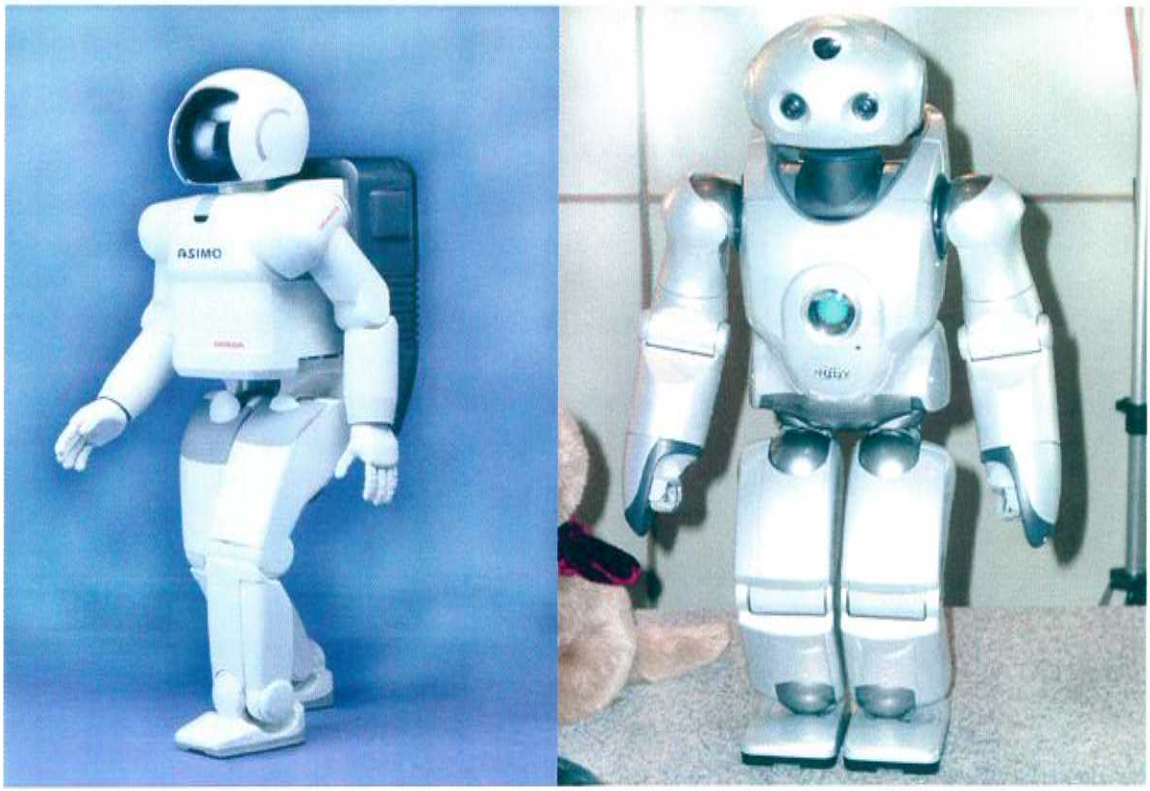


Abbildung 2.2: links: ASIMO von Honda; rechts: QRIO von Sony

Abb. 2.2 zeigt zwei bekannte Vertreter aus dem Bereich der humanoiden Roboter. Links ist *ASIMO* von Honda [Hond03] zu sehen. Dieser Roboter ist 120 cm hoch und ist dank seines weiten Fortschritts auf diesem Gebiet in die *Robot Hall of Fame* [Morr03] aufgenommen worden. Der 52 kg schwere Roboter erreicht inzwischen eine

Geschwindigkeit von 6 km/h und ist der Lage Gegenstände zu greifen und wieder loszulassen. Bis 2010 soll ASIMO in Bürogebäuden Botendienste erledigen können [Kö05].
Rechts daneben ist das wesentlich kleinere Exemplar *QRIO* von Sony zu sehen. Dieser eher als Spielzeug anzusehende Roboter verfügt wie ASIMO bereits über die wesentlichen Fähigkeiten, die ein humanoider Roboter benötigt [Robo06]. Die Forschungen an QRIO und seinem populäreren Vorgänger AIBO wurden 2006 eingestellt [jk06]. Die dort angewandte Technologie soll aber in zukünftigen Projekten verwendet werden.

Hält Honda das Versprechen bis 2010 ASIMO im Alltag voll einsetzbar zu machen, werden sich einige Zeit später viele Personen ein Leben ohne humanoide Roboter nicht mehr vorstellen können. Was den kurz vorgestellten Modellen noch fehlt ist ein menschliches Gesicht. Zwar können Emotionen auch über Körperhaltung vermittelt werden, Eindeutigkeit ist jedoch schwierig, da diese Faktoren durchaus von der Kultur abhängig sind. In der deutschen Einleitung des Buches *Codierung von Emotionen im Mittelalter* [JaKa03] schreibt die Autorin Ingrid Kasten:

> [...]Modellierung von Gefühlen und die Formen des verbalen wie nonverbalen Ausdrucks von Emotionen sind kulturell bedingt und können einen hohen Grad an Ritualisierung aufweisen. [...] Die Codierung von Emotionen ist historisch bedingt und kulturell variabel. Wo sind aber Grenzen zwischen Universalien und Variablen? [...]

Ein bekanntes Beispiel für diese kulturelle Abhängigkeit des Ausdrucks sind gehobenen Arme. Während in arabischen Kulturen erhobene Arme meist ein Ausdruck von Trauer oder Empörung ist, so ist dies in der westlichen Kultur eher ein Zeichen der Freude oder Erfolg.
Genau aus diesem Grund ist ein Gesicht notwendig, durch das Emotionen eindeutig vermittelt werden können.

Im Folgenden werden vier Systeme zum Ausdruck von Gesichtsmimik vorgestellt. Die für diese Untersuchung verwendete Technik basiert auf dem virtuellen Avatar Baldi, der in Abschnitt 3.2.2 genauer vorgestellt wird. Als Roboter dient eine Teilplattform des ARMAR3-Systems. Dieses wird in Abschnitt 3.1.1 beschrieben.

2.2.1 Kismet

Der wahrscheinlich bekannteste in Hardware implementierte Avatar ist der Roboter *Kismet* (Abb. 2.3), der 1997 am M.I.T. von Cynthia Breazeal [Brea00] entwickelt wurde. Dabei wurde das Ziel verfolgt, menschliche Emotionen darzustellen.
Ähnlich wie bei einem Kleinkind soll hier eine Person in der Lage sein dem Roboter Dinge zu zeigen. Auf interessante Objekte reagiert Kismet emotional.

Zur Darstellung der Emotionen verfügt Kismet über 18 Freiheitsgrade. Jedes Ohr, die Augenlider und Augenbrauen, Augen, Ober- und Unterlippe sowie der Hals können separat gesteuert werden. Kismet ist damit auch in der Lage Objekte mit den Augen zu fixieren. Abb. 2.4 zeigt Kismet mit verschiedenen Emotionen.

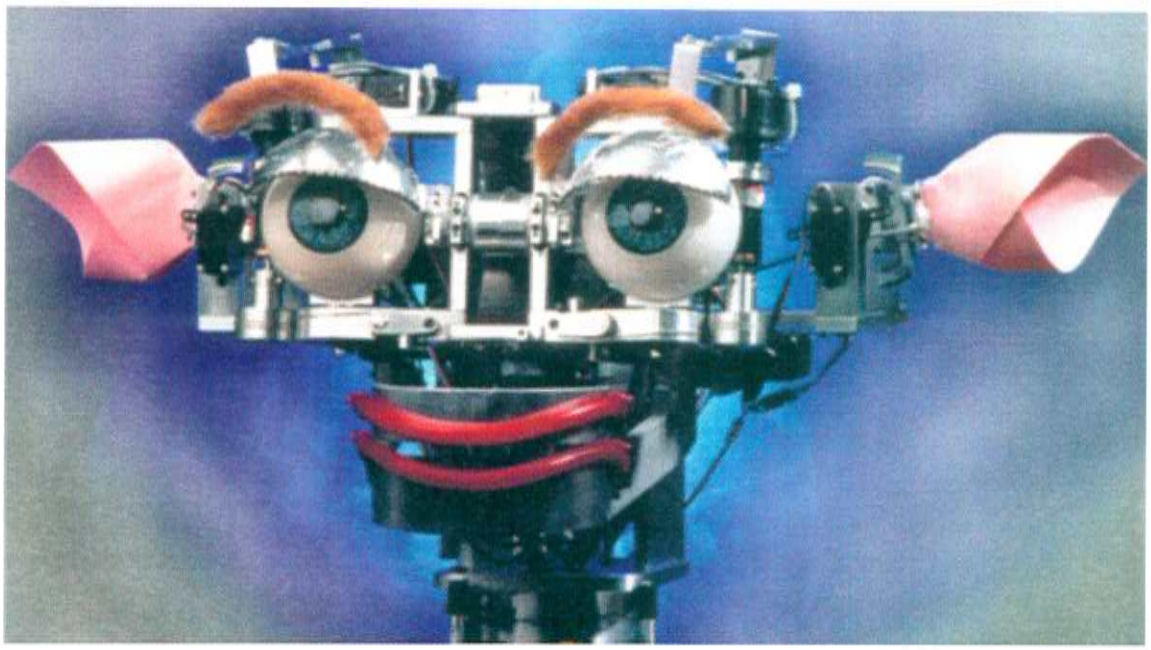


Abbildung 2.3: Roboter Kismet vom M.I.T.

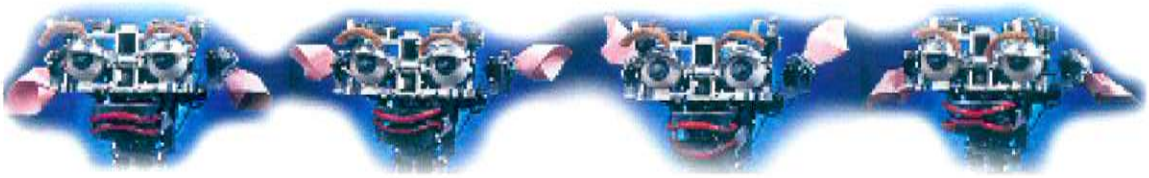


Abbildung 2.4: Kismet: Ausdruck verschiedener Emotionen

2.2.2 K-Bot

K-bot ist aktuell die am weitesten entwickelte Form eines künstlichen Gesichts. Dieser Roboter entstand 2003 an der Universität von Texas und wurde von David Hanson [Hans03] entwickelt. Hanson arbeitete davor bei der Firma Disney als Berater.

Wie man von Abb. 2.5 erkennen kann, ist K-bot Kismet in einigen Dingen überlegen. Die Haut besteht aus einem aufgeschäumten Elastomer, das eigens für diesen Zweck entwickelt wurde und als *F'ubber* bezeichnet wird. Diese Oberfläche gibt dem Roboter ein natürlich wirkendes weibliches Gesicht. Durch 24 mechanische Muskeln, die die künstliche flexible Haut verziehen, kann nahezu jede menschliche Mimik nachgeahmt werden und somit auch Emotion ausgedrückt werden.

Das Gewicht von K-bot beträgt etwa 2 Kilogramm und die Matrialkosten belaufen sich auf 400$. Es ist demnach davon auszugehen, dass dieses System bald eine Verwendung bei humanoiden Robotern finden wird und damit ein menschlicherer Eindruck vermittelt werden kann.

2.2.3 Repliee Q2

Der zur Zeit am weitesten entwickelte humanoide Roboter ist *Repliee Q2* von Hiroshi Ishiguro [MMMI06]. Diese Maschine ist einer bekannten japanischen Nachrichtensprecherin nachempfunden. Abb. 2.6 zeigt zwei Bilder dieses Roboters. Links in einer sitzenden Pose, rechts das Vorbild mit ihrer mechanischen Nachahmung.

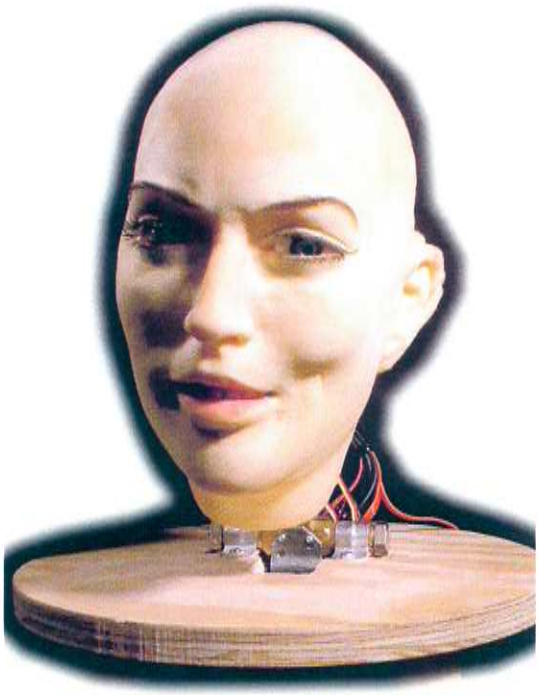


Abbildung 2.5: K-bot

Repliee Q2 verfügt insgesamt über 42 Freiheitsgrade im oberen Teil des Körpers und ist teilweise mit einer Silikonhaut überzogen. Das Gesicht kann mit 13 Freiheitsgraden manipuliert werden. Die Lippen bewegen sich dabei synchron zur Sprache und Emotionen können durch Mimik vermittelt werden.
Was den Ausdruck von Emotionen angeht, kann Q2 nicht mit K-bot mithalten, bei dem fast doppelt so viele Freiheitsgrade für die Mimik verantwortlich sind. Zudem ist K-bot mit einem anderen, der menschlichen Haut ähnlicher wirkenden Material überzogen.

Interessant an Repliee Q2 ist, dass durch Mikrobewegungen eine höhere Natürlichkeit vermittelt wird. Dank einer simulierten Atmung, Lidschlag und Augenrollen wurde in einem Turing Test experimentell nachgewiesen, dass Menschen erst nach einigen Sekunden merken, dass sie es mit einem mechanischem Wesen zu tun haben. Beim Deaktivieren dieser Bewegungen ging dieser Eindruck der Natürlichkeit früher verloren.

Der BBC gab Hiroshi Ishiguro folgende Auskunft [MMMI06]:

> An android could get away with it for a short time, 5-10 seconds. However, if we carefully select the situation, we could extend that, to perhaps 10 minutes.
> (dt.: Ein Android kommt für eine kurze Zeit von 5-10 Sekunden damit durch. Wenn wir allerdings die Situation sehr sorgfältig auswählen, können wir das vermutlich auf 10 Minuten ausdehnen.)

2.2.4 Grace

Der Roboter *Grace* basiert auf dem Versuch ein System zu bauen, das den Anforderungen *AAAI-2002 Mobile Robot Competition* gewachsen ist [Simm03]. Sie entstand

Abbildung 2.6: links: Repliee Q2; rechts: weibliches Vorbild mit ihrem Imitat

aus der Zusammenarbeit der Carnegie Mellon University, des Naval Research Laboratory, Metrica Inc. der Northwestern University und des Swarthmore Colleges heraus.

Abbildung 2.7: Roboter Grace mit animierten Gesicht

Wie auf Abb. 2.7 zu erkennen ist, ist Grace dem Roboter in dieser Untersuchung in vielen Dingen ähnlich. Gesichtsausdrücke werden hier über eine Animation dargestellt um die Forderung nach *friendly human-robot interaction (dt.: freundliche Mensch-Roboter Interaktion)* zu erfüllen. Der Avatar basiert auf einer Standardimplementierung aus dem Buch *Computer Facial Animation* von Parke und Waters [PaWa96] und kann sowohl seine Lippen synchron zur erzeugten Sprache bewegen, als auch Emotionen ausdrücken. Die Sprache wir dabei durch das Sprachsynthesesystem Festival erzeugt.

2.2.5 Albert

Der *Serviceroboter Albert* wurde an der Universität Karlsruhe entwickelt. Er verfügt über einen virtuellen Avatar, der dazu dient innere Zustände des Systems zu visualisieren. Auf diese Weise bekommt der Benutzer einen Einblick in das System und kann seine Erwartungen entsprechend anpassen.

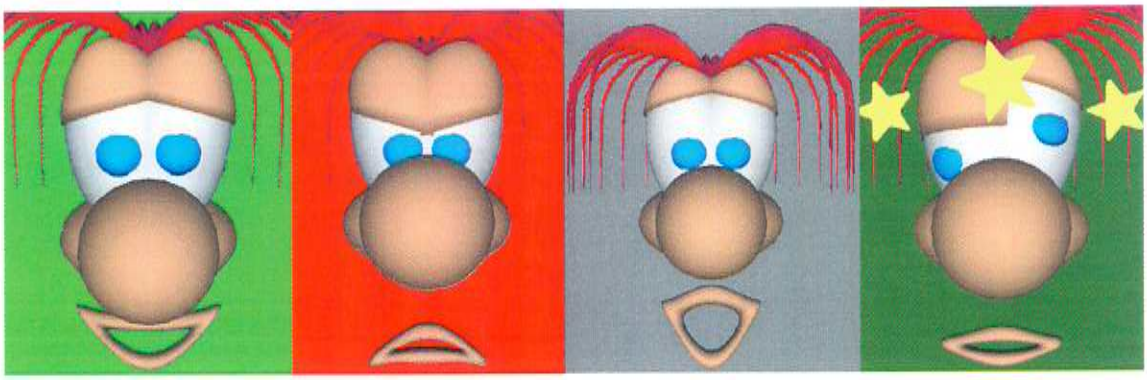


Abbildung 2.8: Avatar des Serviceroboter Albert bei der Visualisierung verschiedener Zustände

Jan Minar beschreibt in seiner Diplomarbeit [MiKD04] die Entstehung dieser Visualisierung.
Der Roboter soll die Zusände *abgeschaltet, wartend, suchend, beschäftigt, Aktion erfolgreich, Aktion fehlgeschlagen, Fehler, fragend* und *neuen Sachverhalt gelernt* darstellen können. Um diese zu untermalen wird zusätzlich zu Effekten gegriffen, wie sie in der Comicsprache üblich sind (z.B. Sterne bei einem Fehler). Abb. 2.8 zeigt vier der insgesamt neun Zustände.

Die Visualisierung von Albert ist vorerst abgeschlossen. Eine Weiterentwicklung wäre aus vielerlei Hinsicht interessant, findet aber zum gegenwärtigen Zeitpunkt nicht statt. Alberts Fähigkeiten sind im Moment noch sehr beschränkt, da das Hauptziel der Entwicklung die Darstellung der Zustände war. Eine freie Konfiguration und Steuerung der Animation ist bisher nicht möglich. Somit ist auch eine synchrone Mundbewegung beim Sprechen nicht möglich.

3. Aufbau

Dieser Untersuchung liegt ein Prototyp von Thomas Prommer zu Grunde. Eine ausführliche englische Beschreibung ist der Diplomarbeit [TPro06] zu entnehmen. In diesem Kapitel wird es daher nur eine kurze Darstellung der bereits vorhandenen Komponenten geben. Auf die Erweiterungen, die für diese Experimente notwendig waren, wird in 3.2 detaillierter eingegangen.

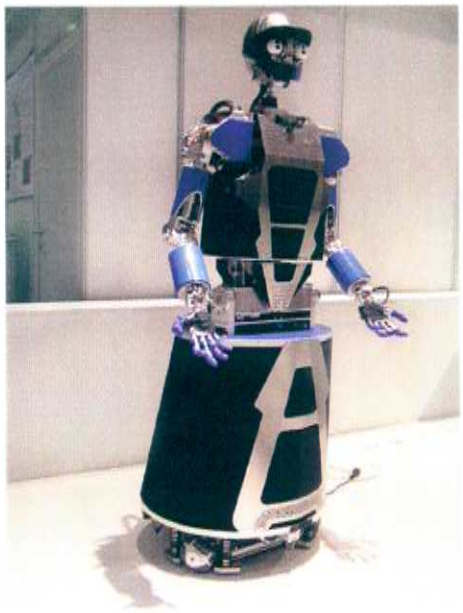


Abbildung 3.1: ARMAR 3 (SFB588)

3.1 Anwendung

Dieser Abschnitt stellt die bereits im Prototypen vorhandenen Systeme vor. Zunächst wird der Hintergrund der Entstehung erläutert, anschließend das Szenario, für das der Dialog entworfen wurde, sowie den Dialogmanager selbst. Zu guter Letzt wird ein Blick auf die Gestikerkennung geworfen.

3.1.1 Hintergrund

Dieser Dialog ist aus dem Ziel heraus entstanden einen humanoiden multimodalen Roboter zu bauen, der im alltäglichen Leben kooperativ dem Menschen zur Seite stehen kann. Hierfür hat die *Deutsche Forschungsgemeinschaft (DFG)* im Jahr 2001 den Sonderforschungsbereich *SFB588 „Humanoide Roboter – Lernende und kooperierende multimodale Roboter"* an der Universität Karlsruhe ins Leben gerufen [Robo01]. Besondere Erwähnung verdient der Roboter ARMAR3 [TAsf01], der über 25 mechanische Freiheitsgrade verfügt. Abb. 3.1 zeigt ein Bild dieses Roboters. Er besitzt zwei anthropomorphe Arme und ist auf einer autonomen Bewegungsplattform montiert. Außerdem ist ein beweglicher Kopf mit Stereokamera und anderen Instrumenten, die multimodale Interaktion ermöglichen, vorhanden.
Die Experimente laufen jedoch nur auf einer Teilplattform von ARMAR ab. Da nur das Verhalten der Benutzer in einem beschränkten Szenario untersucht wird, ist eine vollständige Ausführung nicht zwingend notwendig.

3.1.2 Szenario

Das Experiment spielt sich in einem Barkeeper-Szenario ab. Der Roboter übernimmt dabei die Rolle des Barkeepers. Er ist fest hinter einem Tisch installiert und kann von dort auf verschiedene Objekte mit der Kamera zeigen. Diese Objekte unterscheiden sich sowohl in der Farbe, als auch in ihrer Form. Auf dem Tisch wurden insgesamt 20 Tassen, Teller oder Flaschen in den Farben Blau, Gelb und Rot platziert. Abb. 3.2 zeigt den Aufbau des Szenarios.

Abbildung 3.2: Experiment Aufbau

Der Proband hat die Aufgabe dem Barkeeper zu vermitteln, welches Objekt er gerne serviert bekommen will. Hierfür stellt das System Fragen, die der Benutzer beantwortet. Er muss dabei auf Typ und Farbe eingehen. Als zusätzliche Modalität steht

eine Gestikerkennung zur Verfügung, mit der Proband durch Zeigen auf die Position des Objektes hinweisen kann. Auf diese Weise soll der Gegenstand eindeutig identifiziert werden. Um das ausgewählte Objekt zu verifizieren, sind die Gegenstände mit Nummern versehen.
Der bestmögliche Dialog könnte wie folgt aussehen:

System:	*„Hello, my name is Robbi. What item do you want me to serve you?“*
Proband:	*„I want the blue plate.“*
System:	*„Do you ask for a blue plate?“*
Proband:	*„Yes.“*
System:	*„Could you please point towards the item?“*
Proband:	*(Zeigt auf das Objekt)*
System:	*(Zeigt mit Kamera auf Objekt) „Is the item number 26 I am pointing at the one you wish for?“*
Proband:	*„Yes“*
System:	*„Thank you for your patience. I will serve you item number 26 now.“*

3.1.3 Dialogmanager TAPAS

Um den Ablauf des Dialogs zu steuern, wird der *Dialogmanager TAPAS* benutzt. TAPAS verwendet die Algorithmen des ARIADNE Projekts von Matthias Deneke [Dene05]. Dieses Tool ist sowohl von Sprache als auch von der Domäne unabhängig und eignet sich für Rapid Prototyping Anwendungen. Eine besondere Stärke des TAPAS-Systems liegt in Schnittstelle für multimodale Dialoge. Als Kommunikationsschnittstelle zwischen den einzelnen Komponenten wird die Middleware *one4all* verwendet. Hartwig Holzapfel beschreibt in *Towards Development of Multilingual Spoken Dialogue Systems* [Holz05] diesen Dialogmanager genauer.

3.1.4 3D-Gestikerkennung

Der Roboter verfügt über eine 3D-Gestikerkennung die von Kai Nickel [NiSW03] entwickelt wurde. Die Erkennung basiert auf einem Stereokamera-Signal, aus dem die Position von Hand und Kopf kontinuierlich extrahiert wird. Die Lokalisierung dieser Gliedmaßen funktioniert anhand Hautfarben-Cluster. Nach einer räumlichen Schätzung und mit Hilfe eines trainierten Modells kann die Pose des Benutzers erkannt werden. Abb. 3.3 zeigt links die reale Szene und rechts, wie sie im Computermodell umgesetzt wird.

3.2 Erweiterungen

Ein besonderes Augenmerk richtet sich in dieser Untersuchung auf die Auswirkung von Emotionen und das Lernverhalten der Benutzer. Um diese Emotionen zu vermitteln, wurde das System um einen virtuellen Avatar erweitert, der seine Emotionen auf einem Monitor zum Ausdruck bringt. Die Emotionen werden von einem Benutzer mittels einer Wizard-of-Oz Anwendung gesteuert. Genauere Informationen über den Ablauf während der Experimente sind Abschnitt 4.1.2 zu entnehmen.
In den folgenden Abschnitten werden die einzelnen Komponenten dieses Avatar vorgestellt.

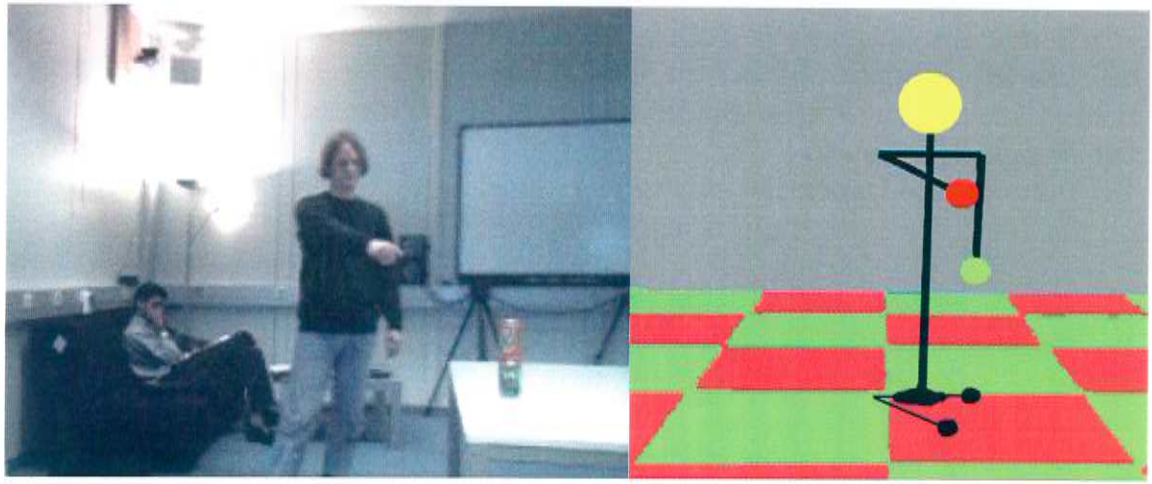


Abbildung 3.3: 3D-Gestikerkennung. links: reale Szene; rechts: Umsetzung in Computermodell

3.2.1 Emotionsmodell EmoMap

Für diese Untersuchung wurde ein neues Emotionsmodell entwickelt, das in den folgenden Abschnitten kurz vorgestellt wird.

3.2.1.1 Grundidee

In dieser Untersuchung sind Emotionen der grundlegende Pfeiler. Aus diesem Grund wurde ein neues System zur Darstellung und Repräsentation von Emotionen entwickelt. Die Idee ist an der Universität Karlsruhe entstanden und basiert auf dem kontinuierlichen *Arousal-Valence* Modell [Lang95] und der dreidimensionalen Visualisierung der Allgemeinen Relativitätstheorie (z.B. unter Wikipedia [Wiki06a]) von Albert Einstein. Abb. 3.4 zeigt ein Beispiel dieser Kombination. Dieses Konzept wird als *EmoMap* bezeichnet und soll im Rahmen dieser Arbeit vorgestellt werden.

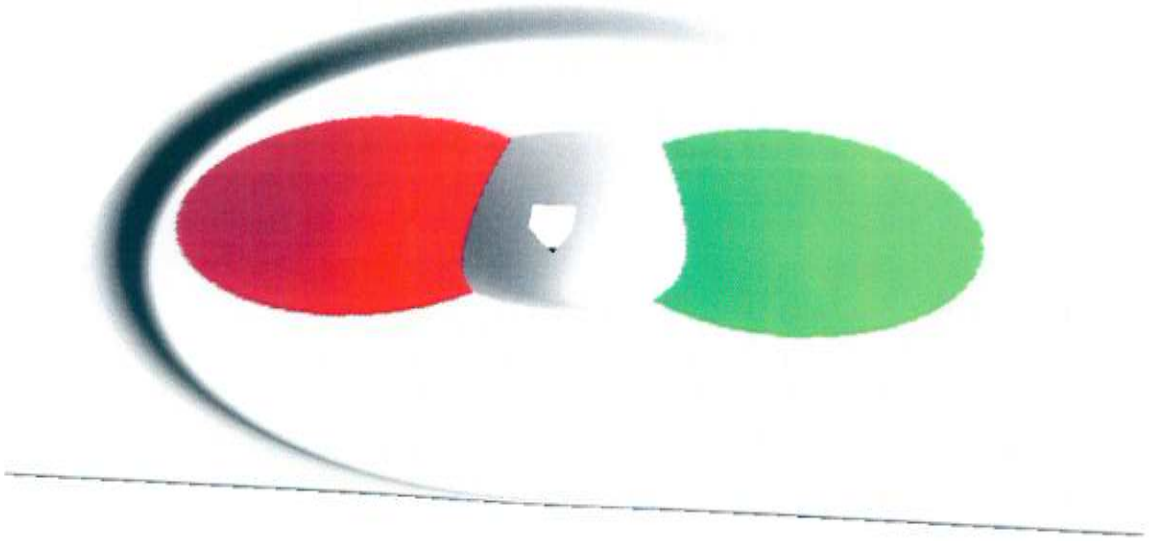


Abbildung 3.4: 3D-Darstellung einer EmoMap

Auf einer Ebene können verschiedene Emotionen eingetragen werden, die durch eine Farbe repräsentiert werden. Die Emotionen wirken wie ein Gewicht, das die Ebene

nach unten drückt. Durch die somit entstandene Krümmung bildet sich ein Potential, innerhalb dem ein stabiler Zustand herrscht. Ein Zeiger repräsentiert nun die aktuelle Emotion. Durch Ereignisse kann dieser Zeiger einen *Impuls* erhalten, der ihn von seiner aktuellen Position verdrängt. Ist der Zeiger mitten in einem Potential hat dieser erst eine Steigung zu überwinden, um den aktuellen Zustand zu verlassen. Es ist also ein gewisser Aufwand erforderlich, um den Zeiger die Steigung überwinden zu lassen.
Zusätzlich kommt ein zeitlicher Faktor ins Spiel, der ebenfalls die aktuelle Geschwindigkeit und Richtung des Zeigers beeinflusst. So wirken über die Zeit *Gravitationskräfte*, die den Zeiger wieder zurück in die Vertiefung zwingen.
Um einen Zustand zu verlassen, ist also entweder ein größerer Impuls notwendig. Passiert über eine längere Zeit nichts, stabilisiert sich der Zeiger in einem Potential.

Die Idee ist also mit Planetenbahnen vergleichbar. Der Zeiger ist ein Trabant, der um einen Planeten kreist. Dieser kann durch äußere Auswirkungen aus dieser Bahn gedrängt werden und kreist danach, falls die Kraft groß genug war, stabil um einen anderen Planten. Allerdings werden die Emotionen in diesem Modell als lokal invariant betrachtet.

Mit dieser Technologie ist es auf einem einfachen Weg möglich, einen Avatar mit verschiedenen Emotionen zu implementieren. Erwünschte Emotionen wie Freude würden demnach häufiger auf der EmoMap verteilt sein, während eher negative Emotionen auf Erhöhungen liegen würden, so dass diese Emotion automatisch wieder verlassen wird.

3.2.1.2 Implementierung

Eine wichtige Entwurfsentscheidung war eine Client-Server-Architektur. Der Client kann leicht in den Avatar eingebunden werden während der Server auf einer anderen Plattform laufen kann.
Zum Zeitpunkt des Entwurfs war bereits abzusehen, dass sowohl Linux- als auch Windows-Systeme zum Einsatz kommen. Umso wichtiger war eine plattformunabhängige Implementierung. Deswegen fiel die Entscheidung auf Java. Java bietet neben einer reibungslosen Plattformunabhängigkeit viele vorgefertigte Klassen, die die Kommunikation über Sockets vereinfachen. Zudem existiert mit Java3D eine Schnittstelle zu 3D-Grafik Bibliotheken, die eine Visualisierung der EmoMaps ermöglichen.

Die Konfiguration einer EmoMap findet in einer XML-Datei statt. Die Emotionen werden durch Gauß-Glocken erstellt, welche sich gut für glattere Übergänge an den Rändern eignen. Sie lassen sich zudem durch wenige Parameter gestalten. Neben den bekannten Argumenten wie Varianz, Höhe und Mittelpunkt, kann ein Wirkungsbereich und ein Radius eingestellt werden. Durch einen beschränkten Wirkungsbereich wird verhindert, dass besonders ausschlaggebende Emotionen auf die komplette EmoMap Einfluss nehmen. Mit dem Radius lässt sich eine Ebene um den Mittelpunkt erzeugen, um breitere Potentiale besser gestalten zu können.
Neben den mathematischen Einstellungen lässt sich noch die Farbe und Emotion angeben. Ersteres spielt allerdings nur bei der Visualisierung eine Rolle.
Die EmoMap selbst verfügt über einige Attribute, die global gesetzt werden. So lässt sich die Reibung an der Oberfläche (*friction*) und die Richtung und Stärke des Gravitationsvektors bestimmen.

In der aktuellen Implementierung sind Anfragen zu Position und Emotion an den Server möglich. Außerdem lässt sich eine direkte Versetzung des Zeigers, sowie die Angabe eines Impulses durchführen.

Die Weiterentwicklung wurde vorerst eingestellt, da die Entscheidung für ein sehr simples Emotionsmodell getroffen wurde (Begründung siehe 3.2.1.4). Das umzusetzende Emotionsmodell ist jedoch mit der jetzigen Implementierung problemlos möglich. Dies wird in Abschnitt 4.1.2.1 beschrieben.

Folgende Ideen und Funktionen sind geplant und sollen bei einer Weiterentwicklung des Projekts implementiert werden:

- Das Importieren von Graustufenbildern (*Heightmaps*), die den Entwurf einer EmoMap in einem Zeichenprogramm ermöglichen.
- Mehr Einstellungsmöglichkeiten bei den Gauß-Glocken; Verzerrung und Rotation.
- Verschiedene Kombinationsarten der Gauß-Glocken (additiv, subtraktiv, Minimum, ...) um glattere EmoMaps zu erzeugen.
- Prioritäten bei Emotionen, damit bei einem Schnitt von zwei Gauß-Glocken eine genauere Entscheidung gefällt werden kann.
- Inverse Emotionen um Erhebungen in der EmoMap zu erzeugen.
- Online-Manipulationen der Zustände oder Direktmanipulationen an der EmoMap für dynamische Anpassungen.
- Scriptsprache um das Setzen von mehreren Emotionen zu vereinfachen oder verschiedene Bedingungen zu formulieren, die bei dynamischen Veränderungen ausschlaggebend sind.
- Magnete, die statt einen Impuls in eine vom Zeiger unabhängige Richtung, einen Punkt angeben, auf den sich der Zeiger hinbewegen soll. Der Zeiger näherst sich also stückweise dem Ziel (Emotion), unabhängig von dessen aktueller Position.

3.2.1.3 Theoretische Möglichkeiten

Die EmoMaps sollen eine einfache Modellierung eines Emotionssystems ermöglichen, das zudem flexibel ist. Wie bereits in 3.2.1.1 erwähnt handelt es sich hierbei um eine direkte Anwendung des kontinuierlichen *Arousal-Valence* Modells. Eine direkte Modellierung ist hier möglich.
Aber nicht nur diese Theorie lässt sich damit in die Praxis umsetzen. Der Zeiger könnte auf mehr als nur *einen* Punkt zeigen, also eine Fläche unter sich abfragen. Man könnte so sämtliche Emotionen, die in dieser Fläche enthalten sind, messen und einen kontinuierlichen Zustandsübergang darstellen. Somit wäre eine Konstellation wie z. B. 33% Emotion 1, 15% Emotion 2 und 52% Emotion 3 möglich. Fließen Emotionen in die Entscheidung des Systems mit ein, so kann eine Gewichtung der einzelnen Faktoren erzielt werden und harte Übergänge werden vermieden.
Das alles lässt sich auch als Fuzzy-Logik beschreiben, die dem menschlichen Wissen

nachempfunden ist. Diese unscharfen Zustände können helfen menschliche Charakteristiken oder Psychen nachzubilden. Mehrere parallel laufende EmoMaps können dabei verschiedene Charakteristiken aufweisen, die je nach Situation unterschiedlich gewichtet werden. Auf diese Weise lässt sich ein Avatar schaffen, der je nach Situation ein unterschiedliches emotionales Verhalten aufweist.

Die hier vorliegende Implementierung lehnt sich an das *Arousal-Valence* Modell an, und ist demnach nur zweidimensional. Die Grundidee lässt sich aber um beliebig viele Dimensionen erweitern. Eine Visualisierung wird dadurch jedoch schwieriger.

Durch eine Manipulation des Gravitationsvektors ist die Simulation eines natürlichen Gefälles möglich. Wäre z. B. die rechte Seite eher mit positiven und wünschenswerten Emotionen bestückt, so wäre eine Konfiguration möglich, in der Zeiger automatisch eine Tendenz nach rechts besitzt und diese Zustände demnach leichter und häufiger erreicht werden.

Weitere Möglichkeiten wären das Verändern der EmoMaps während der Laufzeit, entweder durch direkte Manipulation an der EmoMap oder durch das Verschieben und reorganisieren von Emotionen. Hiermit sollte es möglich sein ein sich dynamisch an den Benutzer anpassenden Avatar zu schaffen. Durch verschiedene Lernverfahren würden sich die Emotionen auf den EmoMaps reorganisieren und nach einer Zeit eine optimierte Konfiguration erzielen, die an den Benutzer und seine Umgebung angepasst ist.

3.2.1.4 Modell

Die Modellierung einer authentischen Psyche ist sehr anspruchsvoll und benötigt viel Erfahrung. Mit den EmoMaps steht aber eine Technologie zur Verfügung, die in der Lage ist, komplexe Vorgänge zu modellieren.
In diesem Experiment sollen jedoch lediglich die Reaktionen von Emotionen auf das Lernverhalten der Benutzer untersucht werden. Hätte der emotionale Avatar eine sehr komplexe Psyche, wäre der Zustandsraum zu umfangreich. Es ist also zu erwarten, dass in der Praxis viele Probanden bestimmte Zustände gar nicht erst erreichen würden. Demnach wäre eine sehr große Menge an Probanden nötig, um aussagekräftige Ergebnisse zu erlangen.
Sollten die Ergebnisse zudem nicht den Erwartungen entsprechen, kann man keine allgemeine Aussage über Emotionen treffen. Im Gegenteil, denn es könnte durchaus sein, dass Emotionen die erwarteten Vorteile bringen, die Komplexität des Modells allerdings Nachteile mit sich bringt und somit das Ergebnis verfälschen kann.

Aus den oben genannten Gründen fiel die Entscheidung auf ein sehr einfaches Modell, bestehend auf 5 Emotionen. Zwei Affekte (positiv und negativ) sowie drei Basisemotionen (Ärger, Neutral und Freude). Eine ausführliche Beschreibung der Auswirkungen dieser Emotionen erfolgt in Abschnitt 4.1.2.

Da es sich bei den Emotionen um wenige, diskrete Zustände handelt, wäre eine Modellierung als endlicher Automat ebenfalls denkbar. Die Umsetzung dieses Modells in eine EmoMap wird auf Seite 38 beschrieben.

3.2.2 Der Avatar Baldi

Als animierter Avatar wird *Baldi* [Tool06] verwendet. Baldi basiert auf dem *CSLU-Toolkit*, welches bereits häufig in Forschung und Lehre Anwendung fand. Mit Hilfe der in Baldi integrierten realistischen Zungen- und Gesichtsanimationen wurden in der Spracherziehung von taubstummen Menschen bemerkenswerte Erfolge erzielt [Ston99]. Der *Rapid Application Developer (RAD)*, der im CSLU-Toolkit enthalten ist, ermöglicht es auch unerfahrenen Benutzern die Gestaltung von einfachen Dialogen [McTe99].
Eine genaue Auflistung der Möglichkeiten von Baldi stellt Ron Cole [Cole99] vor.

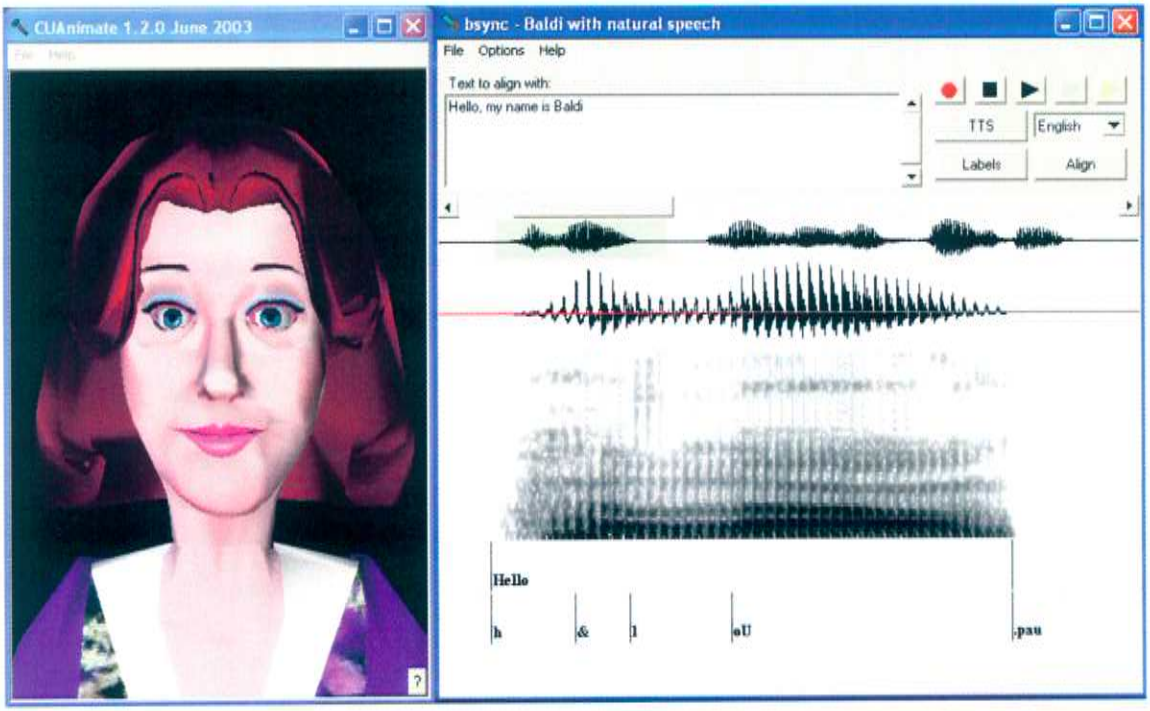


Abbildung 3.5: BaldiSync aus den CSLU-Toolkit

Abb. 3.5 zeigt das Tool *BaldiSync* das dem CSLU-Toolkit beiliegt. Es erlaubt das interaktive Erstellen und Abspielen von Sprachanimationen mit Baldi. Außerdem wird neben dem PCM-Signal noch das Spektrum samt Transcript angezeigt.

Baldi ermöglicht die Kommunikation von Emotionen über alle Modalitäten, die das System zur Verfügung hat. Die Gesichtszüge lassen sich durch sieben Parameter einstellen. Alle benötigten Emotionen lassen sich somit über die Mimik ausdrücken, wie in Abb. 3.6. zu erkennen ist. Dass Baldi dreidimensional animiert ist eröffnet zusätzlich die Möglichkeit die Position zum Betrachter auszurichten. Ablehnung kann somit über eine weitere Entfernung zum Probanden ausgedrückt werden.

Die Sprachsynthese von Baldi ist eine Erweiterung zum TTS-System Festival. Diese ist jedoch in das CSLU-Toolkit fest einkompiliert. Der Patch, der eine deutsche Sprachsynthese [Mö06] erlaubt, ist hier also nicht anwendbar. Eine emotionale Sprache ist auf diesen Weg nicht zu erreichen.
Baldi ist allerdings über eine spezielle Tcl-Shell steuerbar. Auf diese Weise lässt sich die Sprache mit anderen System erzeugen und in das Animationssystem einspeisen. Dieser Vorgang wird in 3.2.4 genauer erklärt.

Abbildung 3.6: Verschiedene Gesichtsausdrücke von Baldi

Ein weiterer Vorteil von Baldi liegt in seiner Geschwindigkeit. Die Animation wird durch sog. Bones und nicht durch ein Muskelsystem erzeugt, was den Rechenaufwand erheblich senkt. Das CSLU-Toolkit läuft bisher nur auf Windows-Plattformen, wodurch jedoch eine ausreichende 3D-Grafikkartenunterstützung garantiert wird.

3.2.3 Sprachsynthese OpenMary

Die Auswahl der Sprachsynthese (TTS[1]) stellte sich als eine der größten Herausforderungen bei der Entwicklung des Prototypen heraus. Das in das CSLU-Toolkit integrierte Festival eignet sich nicht, da es nicht die Möglichkeiten der emotionalen Sprachsynthese besitzt.
Um einen reibungslosen Ablauf mit Baldi zu garantieren, sollte die stand-alone Version von Festival [Alan97] als TTS eingesetzt werden. Festival ist frei verfügbar [Alan06] und via Patch auch zu deutscher Sprachsynthese fähig [Mö01] [Mö06]. Die deutsche Version macht sich MBROLA [Lab96] zu Nutze, das den Vorteil mit sich bringt, dass MBROLA von vielen Wissenschaftlern unterstützt wird und daher für eine große Auswahl an zusätzlichen Werkzeugen gibt. So auch *Emofilt* [Burk03], das aus einer Phonemisation durch Prosodie-Transformation den Eindruck emotionaler Sprache erweckt.
Dieses Programm schien für die Experimente geeignet zu sein. Es musste jedoch ein Patch mit der Funktion geschrieben werden, der die aus Text erzeugte Phonemisation nicht an MBROLA, sondern an Emofilt weiterleitet. Aus der transformierten Phonemkette wird dann mit MBROLA emotionale Sprache erzeugt.

Im Februar 2006 wurde allerdings *OpenMary* des *Deutschen Forschungszentrums für Künstliche Intelligenz GmbH (DFKI)* veröffentlicht [Schr06]. Ein Programm, das allen Anforderungen dieser Untersuchung genügte.

OpenMary wurde von Marc Schröder [ScTr03] entwickelt und ist durch seine Implementierung in Java auch plattformunabhängig. Zudem stellt es eine Client-Server-Architektur zu Verfügung, die das Integrieren der Sprachsynthese in den Avatar erheblich vereinfacht. Besonders hervorzuheben ist allerdings das Programm *EmoSpeak*, welches in Abb. 3.7 auf der linken Seite zu sehen ist.

[1] Text to Speech

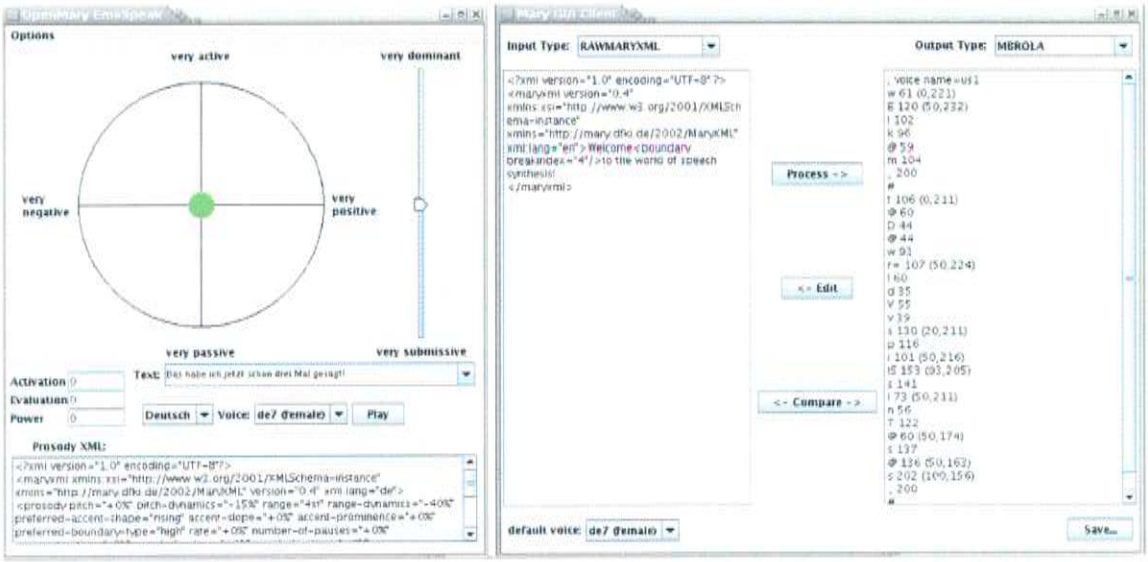


Abbildung 3.7: EmoSpeak (links) und Maryclient (rechts) aus OpenMary

Wie zu erkennen ist, ermöglicht EmoSpeak eine kontinuierliche Einstellung der Emotion. Dabei können die Parameter *Activation*, *Evaluation* und *Power* einzeln bestimmt werden. Hier ist also eine Übertragung des Arousal-Valence Modells auf EmoSpeak möglich. Außerdem könnten die zu wählenden Parameter direkt aus den EmoMaps gelesen werden. Ein kontinuierlicher Emotionsübergang in der Sprache (und auch in der Mimik) sind somit möglich.
Auf der rechten Seite von Abb. 3.7 ist ein Bild der Anwendung *Maryclient* abgebildet, welches eine GUI zum beiliegenden Client darstellt. Sämtliche Anfragen lassen sich hier eingeben und überprüfen, was die Entwicklung eines separaten Clients wesentlich vereinfacht.
OpenMary ist in der Lage deutsche, englische und tibetanische Sprache emotional zu erzeugen.

Beim Experimententwurf fiel die Wahl auf ein sehr einfaches Emotionssystem, mit diskreten Zuständen (siehe 4.1.2.1). Das Potenzial dieser Technologie konnte demnach nicht voll ausgeschöpft werden.

3.2.4 Baldi-Mary Wrapper BMM

Das Herzstück des Avatar ist der Wrapper *Baldi marries Mary (BMM)*, der während dieser Untersuchung entwickelt wurde. Da Baldi bisher nur auf Windows-Plattformen läuft, der Rest der Experimente aber auf Linux, muss BMM plattformunabhängig sein. Daher wurde auch hier Java für die Entwicklung verwendet.
BMM stellt eine Schnittstelle zwischen den EmoMaps, Baldi und OpenMary zu Verfügung. Es handelt sich hierbei größtenteils um statische Methoden, die primitive Funktionen wie Anfragen an den EmoMap-Server erfüllen.
BMM enthält zudem einen minimalen Client zu OpenMary, so dass das TTS-System nicht auf dem Client-Rechner installiert sein muss. Ein EmoMap-Client steht ebenfalls mit BMM zu Verfügung. Um emotionale Sprache zu erzeugen, kann also eine Anfrage nach der aktuellen Position oder Emotion an den EmoMap-Server erfolgen und die Antwort wird zusammen mit dem zu synchronisierenden Text an den OpenMary-Server geschickt. BMM erhält je nach Anfrage wahlweise eine Phonemisation oder eine Wave-Datei.

Durch einen Bug in älteren OpenMary-Versionen ist der Header der Wave-Datei defekt. Hierfür wurde ein Patch entwickelt, der einen neuen Header erstellt. Ab Version 3.0.2 ist dieser Fehler aber auch direkt bei OpenMary behoben.

Komplizierter war es mit Baldi, da hier ein kompletter Server entwickelt werden musste. Baldi lässt sich von einer speziellen Tcl-Shell, der *CSLUsh*, steuern. Der Server wurde demnach in Tcl programmiert und ruft Baldi-eigene Prozesse und Funktionen auf.
Die Audio-Datei kann direkt vom Audio-System geladen werden. Für die Lippensynchronisation ist ein anderer Teil verantwortlich, an den spezielle Phonem- und Zeitinformationen gegeben werden müssen. Diese Informationen werden aus dem MBROLA-Format, wie es OpenMary zurückgeben kann, extrahiert und konvertiert.
Es wurde das BMU-Dateiformat (Baldi Mary Utterances) entwickelt, das sowohl die Phonemisation als auch das Audio-Signal enthält. Dies ermöglicht eine Vorberechnung,die durch BMM abgespielt werden kann. In diesem Fall sind EmaMap und OpenMary-Server nicht notwendig.
Neben OpenMary ist aber auch das interne TTS-System des CSLU-Toolkits verwendbar. Auf diese Weise lässt sich jedoch nur englische und spanische Sprache erzeugen, die allerdings keine Emotionen enthält. Im Gegenzug ist eine zeitliche Beschleunigung feststellbar. Das Erzeugen emotionaler Sprache und das Übertragen an den Baldi-Server kostet verständlicherweise Zeit und bildet einen Flaschenhals. Hierdurch ergibt sich eine spürbare Latenz, die abhängig von Emotion und Länge der Aussage ist. In kritischen Situationen sollte daher das interne TTS-System oder eine BMU-Datei (siehe 3.2.6) verwendet werden.

Neben der Sprachsynthese, lässt sich natürlich auch Baldis Gesicht steuern. Die Mimik, die Kamera oder das Erscheinungsbild (Mann, Frau, Junge, Mädchen, Affe, Alien) lassen sich einstellen, was für einen glaubhaft agierenden Avatar unabdingbar ist.
Zudem lässt sich auch Tcl-Code ausführen und somit auch bisher noch nicht implementierte Funktionen ausführen. Dies ist allerdings ein Sicherheitsrisiko und sollte auf ungesicherten System abgeschaltet werden.

BMM wurde bereits zu Lehrzwecken im Zuge eines Praktikums verwendet.

3.2.4.1 Ausblick

Eine Offenlegung von BMM ist wünschenswert, die endgültige Entscheidung ist allerdings zum gegenwärtigen Zeitpunkt noch nicht getroffen worden. Es besteht allerdings Bedarf an der Fähigkeit Baldi Deutsch sprechen zu lassen [Tool06], das dazu emotional ausgesprochen wird. Dieses Projekt befindet sich noch in einem frühen Stadium der Entwicklung und müsste für eine Offenlegung erst weiter verallgemeinert werden.

In Zukunft könnte dann eine Unterstützung von anderen TTS-Systemen, wie z.B. *Cepstral*, unterstützt werden, welches zwar kommerziell ist, allerdings auch bessere Ergebnisse liefert.
Ein Einsatzgebiet für BMM im Alltag wäre beispielsweise eine Erweiterung eines *IRC- oder ICQ-Clients* sein, die das Geschriebene in emotionale Sprache verwandelt und passender Mimik abspielt. Die Emotion könnt dabei über eine Markup-Sprache gesteuert werden oder aus Emoticons abgelesen werden.

Um allerdings reibungslos zu funktionieren muss der Flaschenhals bei der Übertragung beseitigt werden. Hier gilt es die interne Struktur des Baldi-Audio-Systems zu untersuchen und festzustellen, ob eventuell eine direkte Einspeisung des OpenMary-Server möglich wäre.

Neben den zusätzlichen Funktionen könnte man den Entwurf von BMM um Objekte erweitern. Einzelne Programme instanzieren dann BMM-Objekte, die automatisch die Verbindungen zwischen den einzelnen Komponenten verwalten und für verschiedene Aufgaben zuständig sind. So sind Audio-, MBROLA- oder Code-Objekte denkbar, die auf ihre Aufgaben spezialisiert sind.

3.2.5 Avatar und Wizard-of-Oz

Der Avatar wird durch Baldi visualisiert. In der Implementierung des Avatar, die in Java erfolgte, ist eine Schnittstelle zu *BMM* und *one4all*, einer Kommunikationsplattform, integriert. Zusätzlich besteht eine direkte Verbindung zu einer *Wizard-of-Oz* Anwendung, von der aus die Emotionen, einzelne Gestiken oder auch Animationen samt Äußerung abgespielt werden können.

3.2.6 Überblick

Es soll nun ein schematischer Überblick über das System und seine Funktionsweise gegeben werden.

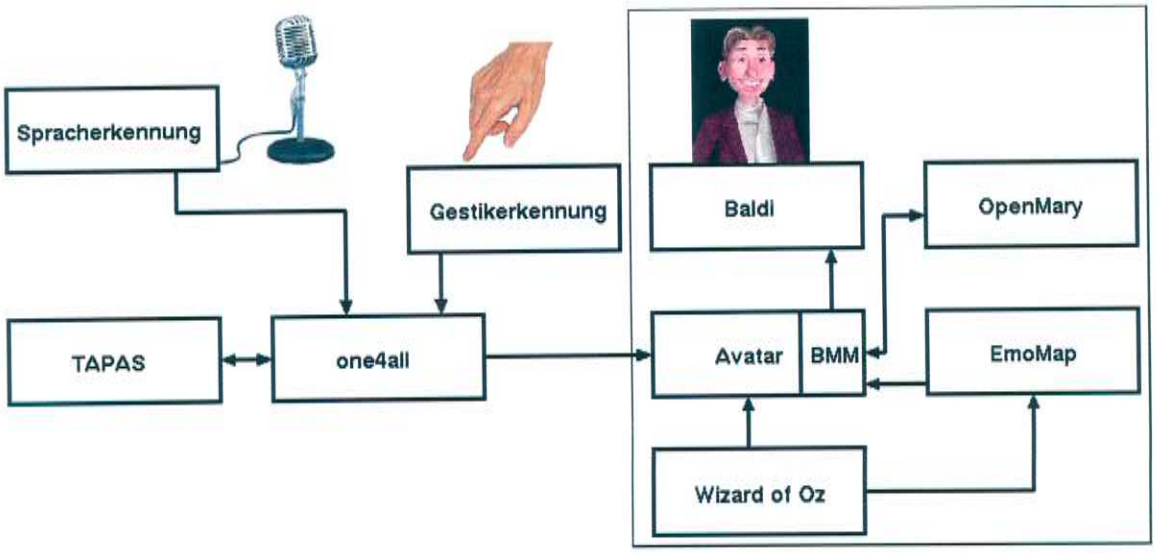


Abbildung 3.8: Schematische Darstellung des Dialogsystems

Abb. 3.8 zeigt dieses Schema. Der umrahmte Teil rechts im Bild stellt die Erweiterungen des Systems dar.

Der Proband kann mit Zeigegestik und Sprache eine Eingabe durchführen. Die Gestikerkennung bzw. der Spracherkenner schicken diese Eingabe über die *one4all*-Kommunikationsplattform an TAPAS, den Dialogmanager. Dieser entscheidet nun, was als nächstes geschieht, z.B. ob eine Sprachausgabe durch den Avatar erfolgt. Tapas schickt über one4all eine Nachricht an den Avatar. Der Avatar ermittelt nun seine aktuelle Emotion anhand der EmoMap. Diese Anfrage erfolgt über BMM. Der

zu sprechende Text samt Emotion wird an OpenMary gesendet, das eine BMU[2]-Datei erzeugt, die an Baldi weitergeleitet wird. Baldi berechnet nun aus der BMU-Datei eine Animation und spielt diese ab. Der Proband sieht diese Animation und wird entsprechend darauf reagieren. Diese Reaktion erfolgt wieder durch Gestik oder Spracher und ein neuer *Turn* im Dialog beginnt.

Durch die Wizard-of-Oz Anwendung ist es möglich direkt Einfluss auf den Avatar oder die EmoMaps zu nehmen. So kann ein externer Beobachter die Reaktionen des Avatar steuern.

[2]Baldi-Mary Utterance

4. Experimente

In diesem Kapitel wird die Konfiguration des Prototyps dargelegt. Hierbei wird auch kurz auf die Hardware eingegangen. Viel entscheidender sind jedoch die einzelnen Emotionen und das Emotionsmodell, die für die Experimente gewählt wurden.
Es folgt eine detaillierte Beschreibung des Ablaufs, den jeder Proband durchläuft. Anschließend werden die Testgruppen und der Fragebogen vorgestellt. Das Kapitel endet mit einem kurzen Bericht über die Vorstudie und die daraus gewonnenen Erkenntnisse.

4.1 Konfiguration

Dieser Abschnitt erläutert die ausschlaggebenden Komponenten der Experimente. Diese sind einzelne Hardware-Aspekte, sowie die Konfiguration des Dialogmanagers und der Entwurf des Emotionsmodells.

4.1.1 Hardware

Der Dialog läuft vollständig in Englisch ab. Da es sich bei den Probanden um Studenten einer Technischen Hochschule handelt, ist davon auszugehen, dass sie über ausreichende Englischkenntnisse verfügen. Der Dialog ist nicht sehr komplex, so dass auch die Probanden, die über einen längeren Zeitraum nicht mehr Englisch gesprochen hatten, diese Aufgabe bewältigen können.
Der Hauptgrund für diese Entscheidung liegt darin, dass eine englische Konfiguration für diesen Prototyp existiert. Der bestehende Dialog wurde also bereits in vorangehenden Experimenten getestet. Eine Übersetzung hätte zu einer zusätzlichen Fehlerquelle führen können.
Ein weiterer Grund ist die leichte Internationalisierung des Experiments. Es wäre eine kooperierende Benutzerstudie zwischen der Universität Karlsruhe (TH) und der Carnegie Mellon University denkbar.

Die Experimente wurden mit einer Videokamera aufgezeichnet. Dabei handelt es sich um das Modell *Sony CDR-HC90E*. Abb. 4.5 zeigt ein Bild aus den Aufnahmen. Die entstanden Videos haben keine Auswirkungen auf das Experiment selbst, erleichtern

jedoch die spätere Auswertung erheblich. Zudem geben die Videos noch interessante Hinweise auf verschiedene Gestiken und andere Verhaltensformen der Probanden

Für die Sprachaufnahmen wurde ein *Countryman close-talking Mikrofon* verwendet. Die Daten wurden über eine *Sennheiser Funkverbindung* an den Aufnahmerechner übertragen. Der Proband konnte sich demnach uneingeschränkt vor dem System bewegen.

Das System verfügt über zwei Entscheidungs-Strategien ([TPro06], Seite 83); der *Baseline-Strategie* und der *Learned-Strategie*. Da aber eventuelle Veränderungen im System vermieden werden sollten, um möglichst vergleichbare Ergebnisse zu erzielen, wird in den Experimenten die regelbasierte Baseline-Strategie verwendet.

4.1.2 Emotionen

In Abschnitt 3.2.1.4 wurden bereits Gründe für ein sehr einfaches diskretes Emotionsmodell dargelegt. Daher wurden nur drei Basisemotionen ausgewählt, die deutlich unterscheidbar sind: Ärger, Neutralität und Freude.

Abbildung 4.1: Visualisierung der emotionalen Zustände des Avatar

Die Emotionen haben im Dialog Auswirkungen auf die Wortwahl, den Tonfall, die Mimik und die Position des Avatar.

Ärger: Das Gesicht bildet die Mimik einer verärgerten Person nach (Abb. 3.6 links). Baldi verschwindet weiter in den Hintergrund und dreht sich weg, wie auf Abb. 4.1 oben links zu erkennen ist. Die Stimme wird aggressiver (*Activation 45, Evaluation -85*) und die Wortwahl zwanglos. Außerdem werden bei Fragen mehrere Antwortmöglichkeiten am Ende des Satzes formuliert. Ein Beispiel hierfür ist: *„Again! What type is the item? Cup, plate, or bottle!“*
Auf diese Weise soll der Benutzer aufgefordert werden seine Antwort anders zu formulieren und eine der möglichen Antworten zu verwenden.

Neutral: Dies soll einen Butler nachahmen. Die Stimme ist neutral (*Activation 0, Evaluation 0*), das Gesicht freundlich (Abb. 3.6 Mitte). Der Avatar hat noch einen gewissen Abstand zum Benutzer und schaut diesem nicht direkt ins Gesicht, sondern leicht vorbei. Die Wortwahl ist sehr höflich. Der obige Beispielsatz lautet bei einem neutralen Avatar so: *„Sorry I have to ask you again, but what type is the item?"*. Die Wortwahl stimmt mit den vorangegangen Experimenten überein.

Freude: Hier drückt das Gesicht Freude aus (Abb. 3.6 rechts) und kommt dem Benutzer sehr nah, wie auf Abb. 4.1 oben rechts zu sehen ist. Die Stimme wird etwas höher und aufgeregter, bleibt aber freundlich (*Activation 40, Evaluation 90*). Bei der Wortwahl fasst sich der Avatar kurz und drückt sich eher persönlich aus. *„Sorry. What type is it?"* lautet der Beispielsatz bei einem freundlichen Avatar.
Auf diese Weise soll eine höhere Effizienz erreicht werden. Da der Avatar freundlich ist, geht er davon aus, dass der Benutzer mit der Bedienung vertraut ist. Andernfalls wäre dieser Zustand auch nicht erreicht worden. Durch schnelle kurze Fragen wird der Benutzer weniger belästigt und kann ebenfalls schneller antworten. Die Effizienz steigt somit.

Das Wechseln der Position von Baldi zur Kamera geschieht fließend, so dass der Benutzer nicht durch abrupte Bewegungen irritiert wird.

Auf das Verhalten des Dialogmanagers haben die Emotion keine Auswirkung. Eine Entscheidung im Dialog wird ausschließlich anhand der gegebenen Regeln gefällt. Die möglichen Eingaben belaufen sich auf Sprache und Gestik.

Neben diesen Emotionen verfügt der Avatar noch über zwei Affekte, einer positiven und negativen Überraschung. Diese kommen nach jeder Aussage des Benutzers zum Zuge und sollen Aufschluss darüber geben, ob das Gesagte auch verstanden wurde. Diese Bewertung erfolgt über einen externen Beobachter.
Über eine *Wizard-of-Oz* Anwendung steuert der Beobachter die Affekte und Emotionen des Avatar. Er sieht dabei die Hypothese des Spracherkenners und vergleicht sie mit dem Gesprochenen. Zudem muss er überprüfen, ob der Dialogmanager die Eingabe auch korrekt verarbeitet hat. Sind alle Faktoren stimmig (es ist keine 100%ige Spracherkennung nötig, so lange bestimmte Schlüsselworte verstanden wurden) gilt dies als positiver Turn. Der Avatar drückt dann für drei Sekunden seinen positiven Affekt (Abb. 4.1 unten rechts) aus.
Treten allerdings Fehler auf, z.B. wenn etwas falsch verstanden wurde, dann wird dieser Turn negativ gewertet und der Avatar zeigt für drei Sekunden den negativen Affekt (Abb. 4.1 unten links).
Der Proband kann bei der Bewertung nicht auf den Monitor des Beobachters sehen, welcher leicht abseits des Systems sitzt. Siehe hierfür Abb. 3.2 auf Seite 22, welches den Aufbau aus einer anderen Perspektive zeigt. Würde der Benutzer auf diesen Bildschirm sehen können, würde dies wahrscheinlich zu Irritationen führen, da er dort seinen Erfolg sehen könnte und seine Aufmerksamkeit nicht auf den Avatar richten würde.

Dieser Beobachter kann in Zukunft durch ein System ersetzt werden, das die Frage und die Antwort des Benutzers analysiert. Wird eine Frage zum Typ des Objekts gestellt und der Benutzer antwortet mit „Yes", so kommt diese Antwort unerwartet

und es gibt keinen Informationsgewinn. Hier wäre demnach eine schlechte Bewertung zu geben. Entspricht die Antwort allerdings den Erwartungen oder übersteigt sie diese sogar (im Falle, dass statt „bottle“ „red bottle“ gesagt wird), wird eine positive Bewertung vergeben.
Fehler in der Spracherkennung sind natürlich auf diese Weise nicht auszuschließen. Die Wahrscheinlichkeit der Hypothese (also wie sicher der Satz erkannt wurde) kann aber ebenfalls in die Bewertung mit einfließen. Dies ist allerdings weiter zu untersuchen.
Durch einen menschlichen Beobachter wird hier ein Goldstandard (Ideal) für die Bewertung gesetzt, der in dieser Form nicht automatisiert werden kann. Dies ist dennoch sinnvoll, da bei einer unsauberen Bewertung keine Aussage über das System selbst gemacht werden kann. Sollte das Experiment scheitern, so könnte ein Grund dafür diese Fehlbewertungen sein. Ein weiterer Störfaktor wird auf diese Weise beseitigt. Außerdem ist mit sinkender Fehlerrate bei der Spracherkennung zu rechnen, da sich diese durch die Benutzung immer weiter adaptiert.

4.1.2.1 EmoMap: Umsetzung des Emotionsmodell

Die Wahl fiel auf ein sehr einfaches Verhaltensmodell. Werden zwei Turns in Folge schlecht bewertet, verändert sich die Emotion des Avatar ins Negative. Bei einer positiven Bewertung passiert das Äquivalente in die andere Richtung. Abb. 4.2 zeigt dieses einfache Modell als endlichen Automaten.

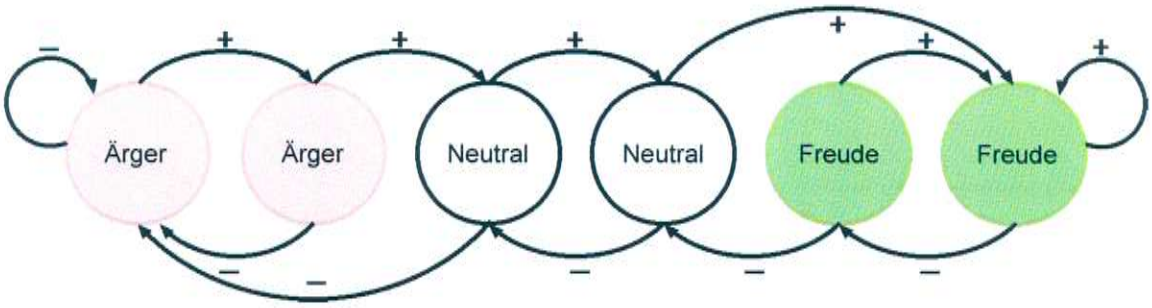


Abbildung 4.2: Darstellung des Emotionsmodell als endlicher Automat

Beachtet man den Automaten genauer, fällt auf, dass eine zweidimensionale Umsetzung nicht notwendig ist. Es gibt nur eine Achse, die der Positivität. Gäbe es nicht der Doppelsprung in die seitlichen Zustände, wäre der Automat leicht als EmoMap darstellbar. Jeder Zustand wäre ein Vertiefung, die in gleichen Abständen nebeneinander angeordnet wären. Die Impulse sollten dann so eingestellt sein, dass der Zeiger genau soviel Energie erhält, dass er nach einem Schub in den nächsten Zustand gleitet. Man muss dabei beachten, dass es während der Bewegung zu einem Energieverlust durch Reibung kommt. Würde keine Reibung existieren, wäre solch ein Schub nicht machbar. Die Energie würde ausreichen, um über alle Steigungen zu kommen, da beim Abstieg ins Tal die gleiche Energie wieder aufgenommen wird.
Der Doppelsprung lässt sich folgendermaßen verwirklichen: Der seitliche Zustand sitzt tiefer als die anderen. Der Übergang vom vorgehenden Zustand weist nur eine sehr leichte Steigung auf. Abb. 4.3 zeigt eine schematische Seitenansicht dieser Umsetzung.

Beim Abstieg vom z.B. rechten neutralen Zustand zu Freude muss die Energie ausreichen, um das erste Freude-Tal vollständig zu überwinden. Da hier die Steigung

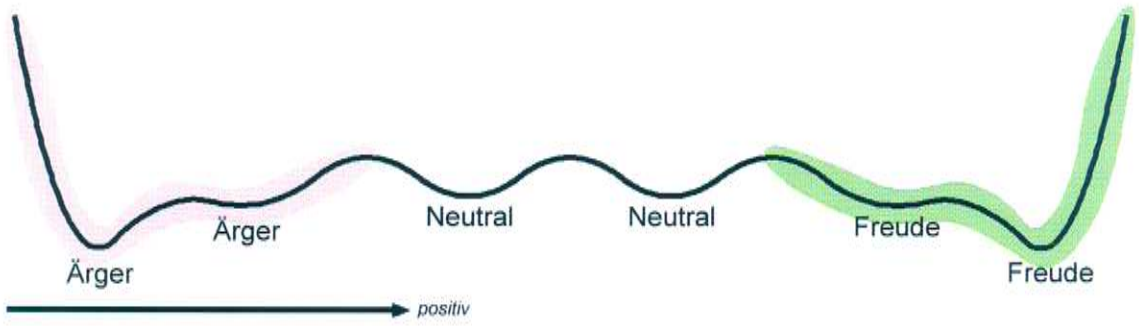


Abbildung 4.3: Schematische Seitenansicht des Emotionsmodell als EmoMap

aber in das zweite Tal kein nennenswertes Hindernis darstellt, nutzt der Zeiger seine Energie, um gleich in das tiefere Tal zu gleiten. Die Steigung in die andere Richtung muss so gewählt werden, dass der Zustand verlassen wird, sich aber im benachbarten Potential wieder stabilisiert.
Die unendlichen Potentiale am Rand verdeutlichen die Reflexivität dieses Zustands. In diese Richtung kann er nicht verlassen werden.
Abb. 4.4 zeigt eine konkrete Umsetzung dieser Theorie in eine EmoMap.

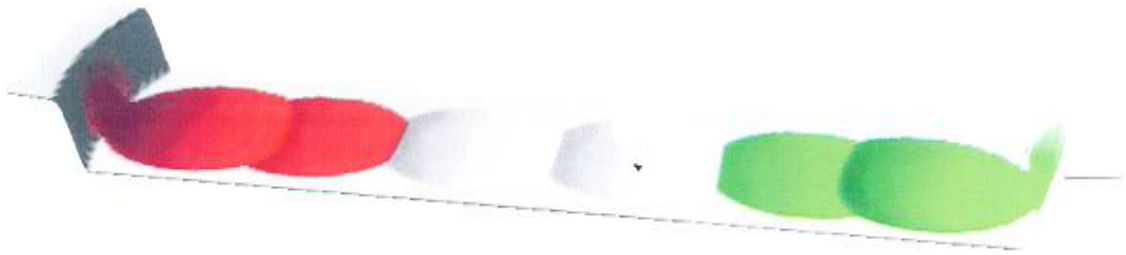


Abbildung 4.4: 3D-Visualisierung der umgesetzten EmoMap

Es wurde hiermit gezeigt, dass sich bestimmte endliche Automaten in eine EmoMap umsetzen lassen. Liegt den Zuständen eine (mehrdimensionale) Ordnung zu Grunde, lassen sie sich auf der EmoMap anordnen. Über einen Impuls oder eine Gravitationsquelle kann dann zwischen den Zuständen gewechselt werde.

4.2 Probanden

Da die Probanden nicht bezahlt werden können, musste die Arbeit auf freiwilliger Basis erfolgen. Hierfür wurden Personen aus dem Freundeskreis oder Kommilitonen angesprochen. Da diese Probanden den Umgang mit dem System erlernen sollten, war es wichtig, dass sie vorher keine Erfahrung mit dem System gesammelt hatten. Keiner wurde vorher genauer unterrichtet, um was es in den Experimenten geht und wie sich der Avatar verhält. Somit konnten alle Experimente unter den gleichen Bedingungen laufen.

Insgesamt wurden 24 Probanden beauftragt, einer wurde bereits bei einem frühen Teststadium herangezogen, um die Funktionsweise des Systems zu testen. Dieser hat auch keinen Fragebogen ausgefüllt. Eine Evaluierung dieser Person ist daher ausgeschlossen und wird in den Ergebnissen auch nicht berücksichtigt.

Es gibt *zwei Testgruppen*, eine beginnt mit einem emotionalen System und hat im zweiten Teil des Experiments den neutralen Dialog; bei der zweiten Gruppe ist dies umgekehrt. Der Grund für die Teilung liegt darin, dass eine Verbesserung der Ergebnisse im zweiten Teil des Experiments zu erwarten ist. Der Proband hat in den ersten drei Dialogen bereits Erfahrung gesammelt und wird diese in den Folgenden einsetzen. Würde hier keine Trennung geschehen, wären die Ergebnisse des ersten und zweiten Teils nicht mehr objektiv vergleichbar.

Um die Auswahl der Gruppen willkürlich zu gestalten, wurde jeder zweite Proband (jede gerade Nummer) in die Gruppe mit dem emotionalen Start gesetzt. Diese Willkür kam mit der Reihenfolge der Probanden, die nicht zu beeinflussen war. Auf diese Weise sollten beide Gruppen hinreichend gut durchmischt sein.

4.3 Fragebogen

Die Fragebögen sind dem Anhang A zu entnehmen.

Der Fragebogen besteht aus fünf Teilen. Der erste Teil ist eine rechtliche Zustimmung, dass während des Experiments Daten aufgenommen werden und diese auch im Zuge dieser Diplomarbeit verarbeitet werden dürfen.

Im zweiten Teil werden persönliche Angaben zum Probanden gemacht. Neben Alter, Geschlecht und Beruf werden noch Englischkenntnisse und technische Erfahrungen, speziell im Bereich der mulitmodalen Schnittstellen und Robotik geprüft.

Die Reihenfolge vom dritten und vierten Teil werden je nach Testgruppe vertauscht. Hier wird der dritte Teil als der Fragebogen zum emotionalen Dialog und entsprechend der vierte zum neutralen Dialog betrachtet.

Der dritte Teil überprüft erst, ob überhaupt Veränderungen im Verhalten des Avatar aufgefallen sind, und wenn ja, welche. Es soll überprüft werden, welche Emotionen erkannt wurden und wie der Benutzer diese bezeichnen würde, also ob sie *eindeutig erkannt* wurden.
Dazu kommen Fragen, wie nützlich diese Emotionen empfunden wurden oder ob das Gefühl vermittelt wurde, dass das System dadurch menschlicher oder sympathischer wirkt. Hierbei geht um eine persönliche Selbsteinschätzung.
Am Ende steht Platz für Kritik oder Anmerkungen zu Verfügung. Hier können besondere Aspekte hervorgehoben werden.

Der vierte Teil behandelt den neutralen Dialog und ist der kleinste Teil des Fragebogens. Neben der Selbsteinschätzung und der Kritik wird noch gefragt, ob der Avatar eher einen positiven Beitrag lieferte oder eher als nutzlos betrachtet wurde, da sich dieser während des Dialogs nicht ändert.

Im letzten Teil werden beide Dialoge verglichen. Es wird gefragt, welches System die eigenen Erwartungen besser erfüllen konnte und welches leichter zu bedienen war. Die wichtigste Frage ist jedoch, wie die Probanden *den eigenen Lernerfolg bezüglich des emotionalen Systems einschätzen.* Diese Frage ist im Vergleich mit den tatsächlichen Ergebnissen besonders interessant, da sie die zentrale Fragestellung dieser Diplomarbeit repräsentiert.

4.4 Verlauf der Studie

In diesem Abschnitt wird der Testlauf und die Vorstudie beschrieben. Diese Schritte sind notwendig, um auf Eventualitäten während der Benutzerstudie vorbereitet zu sein. Gerade eine Vorstudie gibt Aufschluss über einige Merkmale des Experiments, die bei dem Entwurf nicht oder nur kaum beachtet wurden.

Insgesamt nahmen 24 Personen an dem Experiment teil. Zwölf waren in die finale Benutzerstudie involviert, eine für den Testlauf und die restlichen elf für die Vorstudie. Dabei haben sowohl die Probanden aus der Vorstudie, als auch aus den abschließenden Experimenten den gleichen Fragebogen ausgefüllt.

4.4.1 Der erste Testlauf

Bevor die Experimente starteten, wurde der Prototyp mit einer Testperson, die vorher keine Erfahrung mit dem System hatte, getestet. Es wurden mehrere Dialoge mit verschiedenen Zielen durchprobiert. Der Proband musste keinen Fragebogen ausfüllen, da der Versuch nicht unter den geplanten Versuchsbedingungen ablief. Es sollte lediglich die Funktionsweise des Systems getestet werden.
Im Großen und Ganzen wurde festgestellt, dass der Prototyp funktionstüchtig ist und die Experimente beginnen können. Lediglich kleine Fehler bei der Objektnummerierung und bei der Aussprache von Zahlen in der Sprachsynthese (der Punkt am Satzende wurde mit ausgesprochen) wurden festgestellt und sofort korrigiert.
Zudem war dieser Test auch ein kurzes Training für den Beobachter, der so nochmal üben konnte den gesamtem Dialog im Überblick zu halten und zu den passenden Situationen die Wizard-of-Oz Anwendung zu bedienen.

4.4.2 Vorstudie

An der Vorstudie nahmen elf Personen teil, sechs davon begannen mit einem neutralen Dialog. Alle Teilnehmer füllten auch den für das Experiment vorgesehenen Fragebogen aus. Da es an diesem keine Beanstandungen gab, wurde er auch nicht für die Benutzerstudie abgeändert. Die einzige Änderung war, dass nach wenigen Durchlauf bereits beschlossen wurde den Dialog nach dem ersten Teil zu unterbrechen und den entsprechenden Fragebogen ausfüllen zu lassen. Außerdem wird vor dem Start dem Probanden etwas Zeit gegeben, die vom Dialog unabhängigen Fragen (zur eigenen Person, Erfahrungen, ...) auszufüllen.

Es wurde außerdem beobachtet, dass Probanden, die besonders geschickt mit dem System umgingen sehr schnell mit der positiven Emotion konfrontiert wurden. Diese Emotion wurde dann nur noch selten verlassen und wenn, dann war sie wieder neutral. Der Proband konnte die negative Emotion daher gar nicht wahrnehmen. Äquivalent ging es mit Teilnehmern, die große Probleme mit dem System hatten. Daher wurde beschlossen, den dritten Durchlauf des emotionalen Dialogteils in der noch fehlenden Emotion zu starten, so dass der Proband wenigstens einen kurzen Eindruck von dieser Emotion erhält. Der rein theoretische Fall, dass trotz Emotion das System stets neutral bleibt (schlechte und gute Turn immer abwechselnd) trat weder in der Vorstudie noch in der Benutzerstudie ein.
Demnach scheinen drei Emotionen für diesen kurzen Dialog absolut ausreichend zu sein. Um mehr Emotionen verwenden zu können, müsste man die Anzahl der

Durchläufe erhöhen. Dabei würde aber der Proband zu viel lernen, so dass der jeweils andere Teil des Dialogs nicht mehr betrachtet werden könnte. Der Proband hätte bis dahin zu viel Erfahrung gesammelt.

Ein eher technischer Mangel wurde erst in der Vorstudie entdeckt. Antwortete der Proband zweimal oder bereits während die Animation zur Frage abgespielt wurde, trat eine Überlappung (*Margin*) auf. Während der Dialogmanager bereits eine neue Frage abspielen wollte, wurde die laufende Animation des Avatar nicht unterbrochen. Es kam zu Staus und der Dialog konnte nicht mehr sauber beendet werden.
Das System wurde dahingehend verändert, dass vor dem Abspielen die aktuell laufende Animation abgebrochen wird. Dies führte in manchen Situationen zu abrupten Abbrüchen mitten in der Frage. Allerdings wird auf diese Weise auch vermittelt, dass das System etwas verstanden hat und weitere Fragen stellen wird. Außerdem wird auf diese Weise die passende Antwort zur passenden Frage geben, was für einen Dialog unabdingbar ist.
Diese Veränderung wurde bereits nach fünf Probanden eingeführt.

Eine sehr interessante Beobachtung war, dass die Objekte *verschiedene Schwierigkeitsgrade* haben. Blaue Gegenstände sind etwas schwieriger, da „blue" vom Spracherkenner auch gerne als „no" verstanden wird. Da „no" ein wichtiges Wort im Dialog ist, kam es hier zu einigen Verwechslungen. Ähnliche Probleme samt Konsequenzen traten bei „yellow" und „red" wesentlich seltener auf. Sowohl in der Vorstudie als auch in der Benutzerstudie wurde die Aufstellung aus den Vorexperimenten gewählt [TPro06]. Die Anzahl gleicher Objekttypen variiert in dieser Aufstellung. Die blaue Tasse und der gelbe Teller sind jeweils einmalig. Ist vom System Farbe und Typ verstanden worden, ist keine Verwechslung mehr möglich, z.B.:

System: *„Hello, my name is Robbi. What item do you want me to serve you?"*
Proband: *„I want the yellow plate."*
System: *„Do you ask for a yellow plate?"*
Proband: *„Yes."*
System: *„Thank you for your patience. I will serve you item number 4 now."*

Diese beiden Objekte wären also in mindestens zwei Turns erreichbar, während für alle anderen Gegenstände mit mindestens vier Turns (Aufforderung zur Zeigegestik und darauf folgende Bestätigung) zu rechnen ist.
Aus diesem Grund wurde für die Benutzerstudie eine feste Objektreihenfolge bestimmt, was ein Unterschied zur vorangegangenen Untersuchungen ist. Dort durften sich die Probanden einen Gegenstand vorher aussuchen. Da die Ergebnisse zwischen beiden Dialogteilen aber vergleichbar sein sollten, wurde pro Teil ein einzigartiger Gegenstand (gelber Teller und blaue Tasse), ein blauer Teller (dreimaliges Vorkommen auf dem Tisch) und eine rote Flasche (vier mal auf dem Tisch, dazu sehr nahe beieinander stehend) ausgewählt. Die Objekte haben so einen steigenden Schwierigkeitsgrad. Außerdem sind auf diese Weise alle Farben und Typen enthalten.

Die Sprecheradaption vor dem Dialog war anfangs zu kurz. Während die ersten Probanden nur wenige einzelne Worte sagen mussten und gelegentlich sogar Deutsch ins Mikrofon sprachen, wurde dies bei der Benutzerstudie geändert. Während des Dialogs durfte kein Wort Deutsch gesprochen werden. Dazu wurde mehrere ganze

Sätze für die Adaption verwendet. Die Sätze stammen dabei aus dem Dialog selbst, um sich an das System ein wenig zu gewöhnen. Beispiel für einen Satz wäre: „Give me the blue cup." oder „I would like to get a plate". Dabei wurden bewusst verschiedene Sätze gewählt, um dem Benutzer zu vermitteln, dass er mit der Formulierung auch etwas experimentieren kann und soll.

Der Beobachter wurde angehalten keinen Einfluss auf den Ablauf zu nehmen und keinerlei Fragen des Probanden zu beantworten. Diese Regelung wurde für die Benutzerstudie allerdings gelockert. Sollte der Benutzer Probleme mit der englischen Sprache haben und ein Wort z.B. falsch aussprechen oder es einfach nicht wissen, so hat das nichts mit dem Dialog selbst zu tun und es wäre kontraproduktiv ihn falsch weitermachen zu lassen. So würden die Ergebnisse nur verfälscht werden. Außerdem kann der Dialog nicht reibungslos weitergehen, wenn der Proband die Aussprache nicht korrekt beherrscht. Sollte dieses Problem also mehrfach auftreten, dann wurde kurz nachgeholfen. Ein typischen Beispiel aus der Spracherkennung ist, dass das Wort „blue" oft als „no" verstanden wurde. Wurde also mehrfach nach der Farbe gefragt und das System verstand nur „no" wurde der Benutzer darauf hingewiesen, das Wort in einem Satz zu formulieren, wie „It is a blue cup.". Auf diese Weise wurde es dann meistens verstanden.

4.4.3 Ablauf der Benutzerstudie

Die Aufstellung des Experimente ist auf Abb. 4.5 zu sehen. Der Proband steht während des Dialogs vor dem Tisch mit den Objekten und sieht frontal den Roboter an.

Abbildung 4.5: Kameraaufnahme des Experiments

Bevor der Dialog allerdings beginnen kann, muss der Proband sich mit dem System vertraut machen. Hierbei werden ihm die einzelnen Komponenten des Systems vorgestellt. Der Avatar begrüßt den Benutzer durch die Lautsprecher, bleibt dabei aber neutral, um nichts vorweg zu nehmen.
Das Ziel der einzelnen Dialoge wird erklärt. Der Benutzer soll dem System ein bestimmtes Objekt beschreiben. Hat er dabei Erfolg, erhält er ein Hanuta. Bei einem Misserfolg erhält er nichts. Diese Motivation soll gewollte Misserfolge durch gezielte Verwirrung des Systems verhindern.
Insgesamt werden sechs Dialoge durchgeführt. Drei sind emotional und richten sich nach dem auf Seite 36 vorgestellten Emotionsmodell. In den anderen drei Dialogen bleibt der Avatar stets neutral.
Die Reihenfolge der Objekte wird dabei vorgegeben, da einzelne im Schwierigkeitsgrad variieren. Der gelbe Teller ist z.B. nur einmal auf dem Tisch vertreten, die Zeigegestik fällt dadurch weg. Der Dialog wird kürzer und leichter. Andere Objekte sind schwieriger, da sie häufiger auf dem Tisch vorkommen und außerdem eng beieinander stehen. Die Zeigegestik wird dadurch fehleranfällig. Der Dialog verlängert und erschwert sich dadurch. Pro Experimentteil (emotional und neutral) wurde ein leichter und zwei schwierigere Gegenstände bestimmt. Auf diese Weise kann die Leistung der beiden Teile verglichen werden.
Vor der Durchführung wird der Fragebogen ausgehändigt und auf jede Frage kurz eingegangen. Der Proband bekommt so einen Eindruck von dem, was er während des Experiments beachten muss. Die Fragen direkt zur Person werden auch gleich zu diesem Zeitpunkt ausgefüllt um keine Ablenkung für die späteren Fragen nach dem Dialog zu haben.
Anschließend wird der Spracherkenner mit kurzen Sätzen trainiert. Darunter sind hauptsächlich Sätze, die auch so im Dialog stattfinden können. Beispiele sind „Give me the red cup!“, „I would like to have the blue plate.“ oder „I want the yellow bottle!“ Auf diese Weise werden dem Benutzer Formulierungen vorgetragen, die er im Dialog gebrauchen kann. Dass diese Sätze den Dialog steuern, wird dabei nicht explizit erwähnt.
Nach dem Training wird der Proband angehalten keine Fragen an den Beobachter zu stellen. Das System würde eine Interpretation versuchen, was zu unvorhersehbaren Ergebnissen führen würde. Außerdem würde der Spracherkenner die Adaption verfälschen, da die Frage wahrscheinlich auf Deutsch gestellt werden würde. Der Beobachter wird nur so viel Hilfestellung wie nötig geben. Lediglich bei falscher Aussprache oder Problemen mit der Spracherkennung wird er einspringen. Fragen wie „Was hat er gerade gesagt?“ bleiben unbeantwortet, um keinem Probanden einen Vorteil zu verschaffen.
Nach den ersten drei Dialogen bekommt der Proband den Fragebogen zurück, auf dem er nun den entsprechenden Teil ausfüllt. Der Proband kann in dieser Pause versuchen kurze dringende Fragen klären.
Nachdem alle Fragen ausgefüllt sind, erfolgen die nächsten drei Dialoge, für die dasselbe gilt wie in den Dialogen davor. Sind diese vorüber, wird wieder der Fragebogen ausgehändigt und die letzten Teile ausgefüllt. In diesem Schritt kann der Beobachter nochmals zu Rate gezogen werden, wobei er nur die Fragen nochmal erklären kann, jedoch keine Antwort geben darf.
Das Experiment endet mit dem vollständig ausgefüllten Fragebogen. Insgesamt sind etwa 45–60 Minuten für das Experiment zu veranschlagen. Vor dem Beginn gibt es meist noch einige Fragen zu klären, der Spracherkenner muss adaptiert werden

und das Ausfüllen des Fragebogens kostet auch seine Zeit. Die zwei mal drei Dialoge, das eigentliche Herzstück des Experiments, beanspruchen insgesamt etwa 10–15 Minuten.

Beim emotionalen Dialog ist zu beachten, dass die Emotionen durch die Wizard-of-Oz Anwendung gesteuert werden. Da es wichtig ist, dass jeder Proband auch jede Emotion wenigstens einmal gesehen hat, wird eventuell beim dritten Durchlauf des emotionalen Dialogs der Initialzustand auf die fehlende Emotion gesetzt. Bei Probanden, die z.B. Probleme mit dem System haben, würde die positive Emotion nicht auftreten. Der dritte Dialog ist dann mit Freude zu starten um wenigstens einen kurzen Eindruck dieser Emotion vermitteln zu können. Analog wird so mit Probanden verfahren, die „zu gut“ mit dem System umgehen.

5. Ergebnisse und Analyse

In diesem Kapitel wird das Ergebnis der Untersuchung besprochen. Zuerst werden die Probanden statistisch erfasst und deren Fragebögen ausgewertet. Hierbei wird jeder Teil einzeln betrachtet. Am Ende wird auf die in den Fragebögen genannte Kritik eingegangen. Anschließend werden die praktischen Ergebnisse der Experimente untersucht und verglichen. Zum Abschluss erfolgt eine Diskussion dieser Ergebnisse und den daraus gewonnenen Erkenntnissen.

5.1 Erfassung der Probanden

Alle Probanden sind Studierende an einer Karlsruher Hochschule. Sie nahmen freiwillig teil und wurden für ihre Mühen nicht finanziell entlohnt. Es konnten so 24 Probanden für die Untersuchung gewonnen werden. Einer diente als Testperson für die ersten Probeläufe des Systems und wurde anschließend auch nicht befragt. Die verbleibenden 23 Personen wurden Teilnehmer an der Studie und füllten dementsprechend einen Fragebogen aus. Ihr Alter liegt zwischen 20 und 32 Jahren, wobei 16 davon zwischen 22 und 26 Jahren alt sind. Tabelle 5.1 liefert einen kurzen Überblick über die Eigenschaften der Probanden.

Wie zu entnehmen ist, verfügen alle Teilnehmer über ausreichende Englischkenntnisse. Etwas mehr als die Hälfte sogar über fortgeschrittene Kenntnisse.
Knapp ein Drittel der 23 Teilnehmer ist weiblich. In der Benutzerstudie, an der nur zwölf Probanden beteiligt waren, ist es sogar die Hälfte.
Im Fragebogen wurden auch Computerkenntnisse abgefragt. Diese finden hier keine weitere Erwähnung, da sie sich mit den Erwartungen decken. Informatiker gaben hier in der Regel gut bis sehr gute Kenntnisse an. Niemand hatte bisher keine oder nur wenig Erfahrung mit Computern.
Mit dem System war kein Proband bisher vertraut und nur wenige mit Spracherkennung oder Robotik.

5.2 Auswertung der Fragebögen

Da weder am Fragebogen noch am Emotionsmodell eine Änderung nach der Vorstudie vorgenommen wurde, werden in diesem Abschnitt alle 23 ausgefüllten Fragebögen

	abs. Anzahl	in Prozent
Anzahl Teilnehmer	23	-
männl. Teilnehmer	16	69,6%
allg. Hochschulreife	23	100,0%
Studienrichtung		
Informatik	7	30,4%
Informationswirtschaft	3	13,0%
Maschinenbau	5	21,7%
sonst. Ingenieur	5	21,7%
sonst. nicht Ingenieur	3	13,0%
Englischkenntnisse		
mind. 7 Jahre Schule	23	100,0%
erweiterte Englischkenntnisse	12	52,2%

Tabelle 5.1: Überblick über Probanden

ausgewertet. Die Änderungen am System waren rein technischer Natur, so dass von einer Konsistenz in der Befragung auszugehen ist.

5.2.1 Wie wurden die Emotionen erkannt

Die Probanden sollten angeben, welche Emotionen sie erkannten und wie sie diese bezeichnen würden. Auf diese Weise sollte festgestellt werden, ob die Emotionen eindeutig interpretiert werden konnten. Während der Entwicklung wurden die Emotionen als *Neutral*, *Freude* und *Ärger* bezeichnet.
Bei der Angabe der Emotionen gaben nur drei Probanden an, dass das System im emotionalen Dialog auch neutral sein kann. Dass nur so wenige dies angaben, könnte daran liegen, dass *Neutral* der *Normalzustand* ist. Emotionen sind lediglich eine Abweichung von diesem Zustand.
14 Teilnehmer wählten für die Bezeichnung der positiven Emotionen die Wörter „Freude", „freundlich" oder „Freundlichkeit". Andere Bezeichnungen, die aus den Fragebögen entnommen wurden, sind: „flirten", „lächeln", „fröhlich", „glücklich", „lachen" und „nett". Insgesamt gaben 20 Probanden eine positive Emotion an. Ein hoher Anteil nahm also diese Emotion wahr und empfand sie als positiv.
Die Benennung der negativen Emotion fiel dagegen anders aus. So gaben nur 18 Probanden eine negative Emotion an. Sechs davon wählten das Wort „Ärger" oder „verärgert". In diesem Kontext wurden auch „unfreundlich", „Wut", „böse", „nachdenklich", „Verlegenheit", „ratlos", „sauer", „Ablehnung", „Ignoranz", „Frustration", „Desinteresse" und „beleidigt" genannt. Hieraus kann man schließen, dass das eindeutige Vermitteln negativer Emotionen schwieriger ist als das der positiven. 78,3% der Teilnehmer verstanden die Negativität trotzdem. Es ist demnach schwer eine eindeutige Bezeichnung für diese Emotion zu finden.
Dass das System Erfolge durch Affekte ausdrückt, schien von nur vier Probanden bewusst bemerkt worden zu sein. Diese wurden von ihnen als „Unverständnis", „Enttäuschung", „Überraschung" und „begierig" bezeichnet.

Emotionen drücken sich durch vier Faktoren aus. Die *Mimik* passt sich der Emotion an, sowie die *Position* des Avatar zum Benutzer. Weitere Änderungen sind die *Stimmlage* und die *Wortwahl*. Eine genaue Erklärung ist in Anschnitt 4.1.2 zu finden.

Ein Proband aus der Vorstudie kam mit dem System nicht klar. Dieser stellte keinerlei Emotion fest, da er mit dem Dialog selbst überfordert war. Alle anderen Probanden konnten erkennen, dass sich die Mimik des Avatar änderte. Lediglich 14 gaben zusätzlich an, dass die Position des Avatar variierte. Den Wechsel der Stimmlage stellten nur noch vier Probanden fest. Ein Einziger war in der Lage bewusst heraus zu hören, dass sich die Wortwahl mit den Emotionen anpasste.
Der Grund für diese Erkennungsrate mag darin liegen, dass bei einem animierten Gesicht eine Änderung der Mimik durchaus erwartet wird. Eher unerwartet ist die Fähigkeit, dass eine synthetische Stimme Emotionen ausdrücken kann. Daher wurde auch weniger darauf geachtet. Auch die Änderung der Wortwahl ist für Benutzer eher ungewohnt und verlangt bewusstes Hinhören.

Dass es Probleme mit der Eindeutigkeit der negativen Emotionen gab, bemerkte auch ein Proband, der im Fragebogen explizit angab, dass die *positive Emotion eindeutig* vermittelt wurde, die *negative Emotion* jedoch nur *einigermaßen*. Dies deckt sich mit der Beobachtung aus den Emotionsbezeichnungen.
Diese Angabe wurde bei der Frage *„Wie gut konnten die Emotionen vermittelt werden"* gemacht. Hier gaben insgesamt 17 Probanden (oben genannten mit eingeschlossen) an, dass die Emotionen *recht gut* oder *eindeutig* erkannt wurden. Dies entspricht 73,9%.

Interessanter ist die Wirkung der Emotionen auf den Probanden. Hierüber sollten drei Fragen Aufschluss geben.
Die Frage *„Wirkte das Verhalten durch Emotionen natürlicher?"* soll die Meinung des Teilnehmers ermitteln inwieweit sich die Emotionen positiv auf die Natürlichkeit (und damit Integrationsfähigkeit in den Alltag) ausgewirkt hat. Dies sollte fern von der persönlichen Sympathie zu diesem System beantwortet werden. Fünf gaben dabei an, dass der Roboter *überhaupt nicht* natürlicher wirkte. 15 Probanden, und damit knapp zwei Drittel, sind der Meinung, dass Emotionen *einigermaßen* zur Natürlichkeit beitrugen. Drei Teilnehmer empfanden das System als *sehr natürlich*.
Die Sympathie des Systems soll ein subjektives, positives Gefühl beim Benutzer wecken. Dies soll die Motivation, das System nutzen zu wollen, steigern und auch das Lernen fördern. Sollen sich Roboter später in den Alltag integrieren, müssen sie auch sympathisch auf ihre Umwelt wirken. Das emotionale System weckte bei 15 Probanden Sympathie. Zwei waren durch die Emotionen dem Roboter eher abgeneigt.
Dass die Emotionen bei der Benutzung geholfen haben, gaben nur knapp ein Drittel (8 Probanden) an. Die restlichen 15 waren vom Gegenteil überzeugt. Einer empfand Emotionen sogar hinderlich bei der Benutzung.

In der persönlichen Einschätzung geben 20 Teilnehmer an, dass sie ihr Verhalten dem Roboter angepasst hatten. Elf davon stuften dabei ihre Anpassung auf *ein bisschen* ein. Drei gaben an sich überhaupt nicht angepasst zu haben.

5.2.2 Wie wirkte der neutrale Dialog

Der Fragebogen zum neutralen Dialog war sehr kurz. Interessant ist allerdings die Frage, wie der Roboter samt Avatar auf den Probanden wirkte. Hier gaben fünf an, dass der neutrale Avatar den Roboter *befremdlich* wirken ließ. Dagegen meinten drei, dass der Avatar den Roboter trotz Neutralität sympathischer wirken ließ.

Als *störend* wurde der animierte Kopf nur von einem Probanden empfunden. Dieser gab aber in der vorigen Fragen an, dass der Roboter auf sie neutral wirkte.

Als *lächerlich* empfand ihn ebenfalls nur ein Teilnehmer. Dieser meinte aber, dass der Avatar durchaus eine sympathische Wirkung auf ihn hatte. Dieser Angabe ist wohl zu entnehmen, dass sich diese Person über das System amüsierte.
Über die Hälfte (13 Probanden) stuften den animierten Kopf als *nutzlos* ein. Hier wurde ein hoher Anteil erwartet, da der Avatar im neutralen Dialog keine relevante Funktion hat.
Eher überraschend ist daher, dass etwa ein Drittel (8 Probanden, 34,8%) den Avatar als *hilfreich* einordneten. Dies kann an der Mundbewegung liegen. Durch eine Bewegung beim Sprechen wirkt das System etwas natürlicher als bewegungslose Lautsprecher. Außerdem wird dem Benutzer das Gefühl vermittelt, dass der Roboter aktiv mit ihm kommuniziert. Interessant wäre hier gewesen, die Mundbewegung zu deaktivieren. Dann wäre zu erwarten gewesen, dass mehr Probanden den neutralen Avatar als störend eingestuft hätten, da bestimmte Erwartungen nicht erfüllt worden wären.

5.2.3 Vergleich

Der vergleichende Fragebogen sollte Aufschluss über den Erfolg des Experiments geben. Die Angaben sind jedoch sehr durchmischt und zeigen in diesem Fall kein klares Ergebnis.

Beide Systeme liegen bei der Erfüllung der Erwartungen der Probanden etwa gleichauf. Zehn meinten, dass keines der beiden Systeme die Erwartungen besser erfüllen konnte. Sechs gaben das emotionale und sieben das neutrale System bei der besseren Erfüllung an.
Auf zwölf Probanden wirkte das emotionale System sympathischer als das neutrale System. Das Gegenteil empfanden dagegen nur sechs.
Dieser Trend setzt sich in umgekehrter Reihenfolge bei der Frage nach der leichteren Bedienung fort. Hier schneidet das neutrale System mit zehn Stimmen besser ab als das emotionale System, das nur sieben Teilnehmer als leichter zu bedienen empfanden.

Die interessanteste Frage ist die des Lernerfolges. Auf Grund der persönlichen Einschätzung sollten die Probanden angeben, ob sie die Bedienung des emotionalen Systems schneller oder besser erlernten als die des neutralen Systems. Hier gaben zehn *wenig* oder *wesentlich bessere Lernerfolge* durch Emotionen an. Acht hielten beide Systeme für *gleich* gut erlernbar. Nur fünf Probanden, nicht einmal einem Viertel der Teilnehmer, empfanden das emotionale System als schwieriger erlernbar.

Wieder etwa gleichauf liegen die Systeme bei der Einschätzung zur späteren Integration in den menschlichen Alltag. Zwölf Probanden stuften das emotionale System als alltagstauglicher ein, zehn dagegen das neutrale. Ein Proband hielt keines der beiden Systeme für den Alltag geeignet.

5.2.4 Kritik

Auf den Fragebögen hatten die Probanden genug Platz konstruktive Kritik zu üben. Die Kritik wird hier in verschiedene Kategorien aufgeteilt: *technische Aspekte*, Besonderheiten des *Dialogsystems*, Kritik an der *Emotionalität* und Eigenheiten des *Avatar*.
Folgende Abschnitte fassen die genannte Kritik zusammen. Wie man bereits der bisherigen Auswertung entnehmen kann, gehen die Meinungen zu den Systemen auseinander. So werden sich einige Kritikpunkte auch widersprechen.

5.2.4.1 Technische Aspekte

Da es sich beim Roboter um einen Prototypen handelt, ist selbstverständlich noch das ein oder andere technische Problem vorhanden. Dies soll der Wissenschaft kein Hindernis sein, solange er funktionstüchtig ist. Es ist unmöglich alle technischen Eigenheiten in einem so komplexen System auszuschließen.
Hier werden nun die technischen Aspekte der Kritik erhoben, so dass diese bei weiterführenden Untersuchungen berücksichtigt werden können.

Der am häufigsten genannte Kritikpunkt war die *Sprachsynthese*. 13 Teilnehmer fanden diese zu undeutlich. Die Sprachsynthese ist ein sehr wichtiger Punkt der Schnittstelle zwischen Mensch und Maschine. Ohne sie ist das System nicht in der Lage eine Frage zu formulieren. Diesem Kritikpunkt sollte daher eine besondere Bedeutung zukommen. Um so bedauerlicher ist es, dass die Hälfte der Teilnehmer nicht zufrieden waren.
Da es für diese Untersuchung aber wichtig war, Emotionen mit in die Aussprache zu bringen, war die Auswahl an TTS-Systemen eingeschränkt. Ein kommerzielles System bietet eventuell eine deutlichere Synthese mit Emotionen. Als Alternative wäre auch Deutsch als Sprache zu nehmen. Die Sprachqualität würde dadurch nicht besser, allerdings ist die Hoffnung auf ein besseres Verständnis berechtigt.

In der Vorstudie hatten noch einige Probanden Probleme mit der *Spracherkennung*. Fünf gaben an, dass diese leicht überfordert war. In der Benutzerstudie wurde daher eine ausführlichere Sprecheradaption durchgeführt. Die Spracherkennung wurde nur noch von einem Probanden bemängelt. Diese merkte an, dass sie zu laut sprechen musste, um verstanden zu werden. Das Problem hierbei kann an einer ungenaueren Einstellung des Mikrofons gelegen haben. Da andere aus der Benutzerstudie diesen Punkt nicht anbrachten, wurde diese Kritik nicht weiter verfolgt.

Die *Bewegung der Kamera* wurde zweimal erwähnt. Einmal lobend, da sie bei der Objektbestätigung eine gute Unterstützung lieferte und man das abgefragte Objekt im Voraus erahnen konnte. Eine Rüge ging an die Lautstärke der Bewegung. Diese hat zu Verständnisproblemen geführt, da sie die Sprachsynthese übertönte.

In Bezug auf die Kamera wurde auch die *Gestikerkennung* genannt. Dass diese überhaupt existiert wurde von einem Probanden positiv hervorgehoben. Ein anderer gab an, dass diese nicht funktionierte.
Wie bei der Spracherkennung auch, wurde dieser Punkt nicht weiter untersucht, da auch hier keine weiteren Teilnehmer diese Kritik äußerten. Zwar hatten einige Probanden mit der Gestikerkennung Probleme, gaben dies aber nicht explizit in ihrer Kritik an. Die meisten waren eher überrascht, dass diese überhaupt existiert.

In diesem Zusammenhang soll erwähnt werden, dass zwei Probanden angaben, dass das *System besser als erwartet funktionierte*. Einer meinte, dass er seine Zweifel hatte, als er das System sah. Da er aber das Dialogziel mehrfach erreichte, erwiesen sich diese Zweifel als unbegründet.

5.2.4.2 Dialog

Zum Szenario, dem Dialog und seiner Konfiguration gab es natürlich auch Kritik. Diese ist in diesem Bereich jedoch etwas weniger fundiert. Kommentare zu technischen oder emotionalen Aspekten kamen häufig von mehreren unabhängigen Probanden. Bei den in diesem Abschnitt vorgestellten Kritikpunkten handelt es sich dagegen um vereinzelte Meinungen einiger Probanden.

Positiv wurde von zwei Teilnehmer angemerkt, dass das System sowohl mit *kompakten Kommandos* als auch mit ganzen Sätzen umgehen konnte. So kann ein erfahrener Benutzer das System effizient steuern, während ein Anfänger trotzdem zurecht kommt.

Die Methode das *Margin* zu umgehen (siehe hierfür Abschnitt 4.4.2), welche während der Vorstudie implementiert wurde, wurde von einem Probanden als sehr störend empfunden. Diese Selbstunterbrechung des Avatar hatte ihn abgelenkt.

In Bezug auf die *Rückfragen* des Dialogmanagers, gingen die Meinungen auseinander. Zwei Teilnehmer gaben an, dass sie die häufigen Rückfragen als nützlich empfanden. Vier meinten dagegen, dass diese eher störten und den Dialogverlauf verlangsamten. Die Aussage der letzteren Gruppe ist richtig, allerdings sind solche Rückfragen notwendig, da durch sie eine Entscheidung noch einmal bestätigt wird. Besonders für unerfahrene Benutzer ist diese Funktion wichtig. Bei kritischen Handlungen muss der Roboter daher sicher gehen, dass er auch das Richtige tut. Besonders für unerfahrene Benutzer ist diese Funktion wichtig.
Ein Proband bemängelte den Neuanfang nach einer Verneinung einer Rückfrage. Hatte sich das System für einen Objekttyp und Farbe entschieden, so wollte es diese bestätigt haben. Folgte auf diese Frage ein „No“, musste die komplette Prozedur von Neuem beginnen.
Ein wichtiger Kritikpunkt bezieht sich auch auf die Objektbestätigungen. Sollte hier das System „Yes“ verstehen und ein falsches Objekt bestätigt haben, so ließ sich dieser Fehler nicht mehr korrigieren. Dies kann in kritischen Situationen fatale Folgen haben. Dieser Mangel wurde von zwei Probanden genannt.
Diese genannten Probleme ließen sich durch eine Neustart- oder Zurück-Funktion lösen.

Ein Proband bemängelte die Wartezeit zwischen den Turns. Diese war zu kurz eingestellt. Die meisten Probanden warteten, bis die Frage zu Ende gestellt war. Bei längeren Fragen wie bei der Objektbestätigung oder einem verärgerten Avatar, blieb dem Benutzer nicht mehr viel Zeit zum antworten. Wartete der Proband zu lange, wurde seine Antwort schon auf die nächste Frage bezogen. Dies kann bei Antworten wie „Yes“ und “No“ große Auswirkungen haben.
Interessanterweise hatten mehrere Teilnehmer Probleme dieser Art, gaben dies aber nicht an. Sie konnten ihre Schwierigkeiten wahrscheinlich nicht auf dieses Phänomen beziehen.

Ein Proband erwartete ein komplexeres Szenario. Ein anderer meinte, dass der Ablauf zu eintönig gewesen sei und sich keine Langzeitmotivation einstellte. In weiterführenden Studien sollte tatsächlich ein komplexeres Szenario gewählt werden, da dies schwieriger zu erlernen ist und daher unvorhersehbarer. Die Auswirkung der Emotionen würde sich dort wahrscheinlich prägnanter zeigen und emotionale Benutzerführung ließe sich besser untersuchen.

Zwei Probanden gaben an, dass das System für Anfänger ungeeignet war. Sie hatten sich etwas wie Testdurchläufe oder ein Tutorial gewünscht. Dies ist für eine zukünftige Anwendung sicher sinnvoll, aber in dieser Untersuchung ging es gerade darum, wie die Benutzung des Systems erlernt wird und ob Emotionen zu einem intuitiven Verständnis beitragen.

5.2.4.3 Emotionalität

In diesem Bereichen gingen die Meinungen auseinander und widersprachen sich teilweise. Einige Probanden empfanden Emotionen als positiv und äußerten dies entsprechend auch in den Fragebögen. Andere, die künstlichen Emotionen eher kritisch gegenüber standen, drückten dies ebenfalls in den Fragebögen aus.

Sieben Probanden gaben in der Kritik an, dass Emotionen die *Langzeitmotivation* steigern würden. Bei dem Roboter handelte es sich aber nicht um ein Spielzeug, das den Besitzer möglichst lange faszinieren soll. Es ist aber möglich, dass sich diese Motivation auch positiv auf das Lernverhalten eines Benutzers auswirkt, da er bei Fehlern nicht die Lust verliert. Menschen gehen positiver an das System heran, und wenn sie Spaß daran haben das System zu bedienen, dann wollen sie auch mehr darüber lernen und testen es weiter aus.

Ein wichtiger Punkt, und auch wesentlicher Bestandteil der Motivation zum Bau dieses Systems, ist eine zu erwartende *Effizienzsteigerung* bei der Bedienung. Durch das emotionale Feedback erfährt der Benutzer direkt nach seinem Kommando, ob der Befehl verstanden wurde oder nicht. Auf diese Weise kann er besser auf das System reagieren. Dieser Aspekt wurde von sechs Probanden angegeben.

In Bezug auf die nicht zufrieden stellenden Sprachsynthese ist ein Punkt hervorzuheben: Fünf Probanden äußerten die positive Kritik das System durch die Emotionen viel leichter verstanden zu haben, da sie durch das Feedback bereits ahnen konnten, was das System als nächstes fragen wird. Zwei weitere gaben an, dass Emotionen zu einer besseren Interaktion führen. So meinen insgesamt sieben Probanden, dass *Emotionen die Kommunikation zwischen Mensch und Maschine verbessern.*

Drei Teilnehmer erwähnten in diesem Bereich noch einmal, dass das System durch Emotionen viel *menschlicher* wirkte. Dies ist diesen Probanden wohl besonders aufgefallen, sonst hätten sie es nicht expliziz erwähnt.

Einem Probanden ist aufgefallen, dass durch die Emotionalität der Benutzer mehr über die Funktionsweise des Systems erfährt. Würde man diesen Aspekt weiter ausbauen, so könnte man das *Erlernen des Systems* weiter vereinfachen.
Dieser Punkt wird indirekt durch eine Anmerkung unterstützt, die fünf weitere Teilnehmer anbrachten. Diese meinten, dass *Emotionen unnötig seien, sobald das System verstanden worden ist.* Diese verspürten demnach keine Langzeitmotivation oder Freude daran das System zu bedienen. Laut diesen Probanden sind Emotionen anfangs nützlich, verlieren aber, mit zunehmender Beherrschung des Systems, an Bedeutung.

Drei Probanden gaben dagegen an, dass sie durch Emotionen eher *abgelenkt* wurden. Wie diese Ablenkung tatsächlich aussah, wurde nicht formuliert.

Ein sehr interessanter Kritikpunkt wurde von zwei Probanden geäußert. *Emotionen lassen ein komplexeres System (höhere KI) erwarten.* Da das System menschlicher wirkt, werden ihm menschliche Eigenschaften und Fähigkeiten zugeschrieben. Da in beiden Teilen der selbe Dialogmanager samt Entscheidungsstrategie verwendet wurde, konnte diese Erwartung nicht erfüllt werden. Dieser Kritikpunkt wird auch in *Why Humanoids?* [DuJo05] genannt.

Zum neutralen Dialog äußerten sich zwei Probanden:
Einer sagte, dass ihm der Sprung vom emotionalen auf den neutralen Dialog sehr

schwer fiel. Das System hätte sehr viel von seiner Sympathie eingebüßt und er hätte sich dadurch leichter aufgeregt. Emotionen wirkten auf ihn daher eher besänftigend. Das genaue Gegenteil wurde aber auch von einem Teilnehmer behauptet. Dieser meinte, dass ihm die ständige Neutralität des Avatar geholfen habe ruhig zu bleiben. Die Emotionen wirkten hektisch und verleiteten ihn zur Aufregung.

5.2.4.4 Avatar

Der Grund für einen separaten Abschnitt für die Kritik am Avatar liegt darin, dass die hier geäußerten Punkte nichts mit den Emotionen im System an sich zu tun haben. Hier geht es um Entwurfsentscheidungen, die allein den Avatar betreffen. Die Kritik ist also nicht auf Emotionalität im Allgemeinen übertragbar.

Die *Sprachsynthese* würde von vielem Probanden beanstandet (siehe Seite 51). Zwei Probanden gaben dabei an, dass die Sprache erst bei einem freundlichen Avatar *zu schnell und zu hoch* war. Eine Änderungen der Stimmlage bemerkten dabei nur vier der Teilnehmer. In einem neuen Prototypen wird eine andere Sprachsynthese zum Einsatz kommen. Dies wird dieses Problem lösen.

Auch bei der Darstellung der Emotionen, gingen die Meinungen auseinander. Drei Probanden sagten, dass die Emotionen gut dargestellt wurden. Zwei meinten dagegen, dass diese übertrieben waren. Drei weitere gaben an, dass die Emotionen nicht natürlich genug wirkten.
Das Problem hierbei ist, dass es eine *eindeutige* Darstellung der Emotionen sein sollte. Hätte sich die Mimik des Avatar nur in feinen Nuancen geändert, wären die Emotionen womöglich nur von wenigen Probanden erkannt worden. Dass die Aussage der Mimik gut vermittelt wurde, belegen die Bezeichnungen der Emotionen, die auf Seite 48 diskutiert wurden.

Der Ausdruck von Emotion über die *Position des Avatar* schien ungeeignet zu sein. Insgesamt sechs Probanden bemängelten diese Darstellung von Ärger. Die große Entfernung des Avatar zum Probanden wirkte eher verwirrend und ablenkend. Ein Proband dachte sogar, dass das System ihm jetzt nicht mehr zuhören würde. Er behandelte den Roboter dann wie ein Kleinkind, da er sich wie eines verhielt. Da der Proband sein Verhalten sehr veränderte, konnte das System ihn nicht mehr verstehen. Dieses Unverständnis wurde dann vom Probanden als „Sturheit" interpretiert. Diese Person gab an, dass sie einen Butler erwartet hatte und diese Erwartung in keinster Weise erfüllt wurde.
Sehr unerwartet war die Kritik bezüglich der Position beim Ausdruck der Freude. Eine Proband meinte, dass er sich vom Avatar bedroht fühlte, da er ihm zu nahe kam. Dieses Phänomen lässt sich durch die Verletzung des *persönlichen Freiraums* leicht erklären. Der Begriff „personal space" geht auf David Katz [Katz37] zurück und wird von Robert Sommer [Somm69] wie folgt definiert.

> [...] an area with invisible boundaries surrounding a persons body into which intruders may not come.
> (dt.: [...] ein Bereich mit unsichtbaren Grenzen, der den Körper einer Person umgibt und ein Eindringling nicht betreten darf.)

Auch wenn sich der Roboter nicht dem Probanden physisch näherte, so änderte sich jedoch die Perspektive auf den Avatar. Der obere Rand der virtuellen Haare wurde

dabei vom Monitor abgeschnitten. Laut Akira Toriyama [Tori01] nimmt dieser Effekt dem Bild den Bewegungsspielraum. Der Avatar wirkte demnach viel näher als es es tatsächlich war und drang zumindest virtuell in den persönlichen Freiraum ein. Dieses Phänomen kann das Gefühl von Bedrohung vermitteln.
Auch wenn dieser Kritikpunkt nur von einem Probanden angebracht wurde, ist dieser nicht zu unterschätzen. Humanoide Roboter sollen in Zukunft mit dem Menschen auf einer persönlichen Ebene kommunizieren. Ein Einsatz im Bereich der Pflege wird ebenfalls angestrebt. Roboter werden dem Menschen sehr nahe kommen und in manchen Situationen dessen persönlichen Freiraum verletzen. Dieser Punkt muss beim Entwurf von Robotern beachtet werden.

Der Avatar verfügt über *Leerlaufanimationen*, die aus einem gelegentlichen leichten Kopfneigen oder Lidschlag bestehen. Die Meinungen zu diesen Bewegungen waren allerdings zweigeteilt. Die meisten Teilnehmer erwähnten diese Animationen nicht. Ein Proband beschrieb diese Bewegungen als „nervig", ein anderer hebte sie positiv hervor und meinte, sie trage zur Natürlichkeit bei.
Diese Bewegungen werden von Ishiguro [MMMI06] auch als Mikrobewegungen bezeichnet und erwecken den Eindruck von Lebendigkeit (siehe Abschnitt 2.2.3).

In diesem Zusammenhang soll erwähnt werden, dass sich drei Probanden *mehr Einstellungsmöglichkeiten* beim System wünschten. Das Gesicht des Avatar, eine Stimme, ... könnten darin eingestellt werden. Auch die eben genannten Leerlaufanimationen wären hier abschaltbar. Diese Individualisierung ist für eine spätere Nutzung des Systems durchaus sinnvoll. Ein individuell angepasster Avatar sollte zu einer Steigerung der Langzeitmotivation beitragen. Es war wichtig, dass alle Dialoge unter den gleichen Bedingungen durchgeführt wurden, um am Ende vergleichbare Ergebnisse zu haben. Eine freie Wahl des Avatar würde eine zusätzliche Quelle der Ablenkung bieten. Zudem würde dann Stimme und Avatar eventuell nicht mehr zusammenpassen, was ebenfalls eine verwirrende Wirkung haben kann.

Zwei Probanden wünschten sich eine Stimme mit einer differenzierteren Sprachausgabe. Die Kommunikation könnte laut diesen Probanden durch Floskeln lebendiger gestaltet werden. Interessant ist, dass diese beiden nicht bemerkten, dass es bereits beim emotionalen Dialog differenziertere Formulierungen gab. Durch kurze Äußerungen wie Seufzern würde der Avatar zwar an Natürlichkeit gewinnen, wären aber im Dialog hinderlich, da diese fehlinterpretiert werden könnten.

5.2.5 Überblick

Dieser Abschnitt schafft einen kurzen Überblick über die Fragebögen. Die hier genannten Werte beziehen sich auf die Benutzerstudie, da diese im kommenden Abschnitt genauer ausgewertet wird. Es werden nur die Fragebögen der zwölf Probanden aus der Benutzerstudie betrachtet.

Eine tabellarische Auswertung der Fragebögen ist in Anhang B zu finden. Dort sind alle Fragen erfasst, die durch Ankreuzen beantwortet wurden. Die rechte Seite enthält die Werte allein aus der Benutzerstudie. Auf der linken Seite sind die Werte aller Teilnehmer eingetragen. Wie zu erkennen ist, sind in der Benutzerstudie die Hälfte der Probanden weiblich. Insgesamt fallen beide Betrachtungen bei den meisten Punkten sehr ähnlich aus. Lediglich ein Wert sticht in diesem Vergleich deutlich heraus. Bei den wahrgenommen Modalitäten der Emotion wurde die Änderung der

Position von 10 Probanden (83,3%) bemerkt. Im Gesamtüberblick waren dies nur 60,9%. Die Erklärung hierfür ist, dass in der Benutzerstudie das System etwas genauer vorgestellt wurde. Es wurde zwar nicht direkt gesagt, dass der Avatar seine Position verändert, dennoch wurde beim Durchsprechen des Fragebogens erwähnt, dass die Probanden auf den Avatar achten sollen.

Im Vergleich schneiden beide Systeme etwa gleich gut ab. Dies wird besonders durch die abschließenden Fragen deutlich. Jeweils zwei Probanden sagten, dass das jeweilige System ihre Erwartungen besser erfüllte. Bezüglich des Lernerfolges gab die Hälfte der Probenden leichten bis wesentlichen Lernerfolg durch Emotionen an. Das emotionale System wirkte sympathischer, das neutrale ließ sich bedienen.

Betrachtet man nur die Benutzerstudie, so gaben jeweils 58,3% bei der Beherrschung des System *recht gut* oder *super*. Bei genauerer Betrachtung schneidet das emotionale System leicht besser ab. Diesen Vorsprung gibt es auch bei der Bewertung durch alle 23 Teilnehmer.

91,3% der Probanden gaben an, dass die Emotionen gut vermittelt wurden. Die gleiche Menge meinte auch, dass der Roboter dadurch natürlicher wirkte und dass sie ihr Verhalten dem Roboter angepasst hatten. Für 58,3% gewann der Roboter an Sympathie durch die Emotionen. Etwas weniger als die Hälfte gab an, dass sie bei der Benutzung durch die Emotionen unterstützt wurden.

Bezüglich des neutralen Dialogs fand nur ein Proband, dass der Roboter natürlicher wirkte. Ein Drittel empfand den animierten Kopf hilfreich.

Dieses Ergebnis beweist, dass die Konzepte hinter *Affective Computing* richtig angewandt wurden. Die angebrachte Kritik zeigt aber auch, dass es noch einige Punkte gibt, in denen das System noch weiter verbessert werden kann.

5.3 Auswertung der Dialogerfolge

Die Auswertung der Fragebögen hat ergeben, dass beide Systeme etwa gleich abschneiden. In diesem Abschnitt wird die Auswirkung der Emotionen auf den Dialog untersucht. Die Ergebnisse sind dabei nicht von den persönlichen Einschätzungen der Probanden abhängig, sondern sind direkt aus der praktischen Ausführung der Experimente entnommen.

5.3.1 Turns

Eine wichtige Grundlage dieser Bewertung sind die sog. Dialogturns. Ein *Turn* ist dabei genau ein Dialogschritt, also *ein Frage, eine Antwort und die darauf folgende Reaktion.*

Aber was war, wenn auf eine Frage keine Antwort kam? Dies kam meist gegen Ende bei der Objektbestätigung vor, da hier einige Probanden auf das richtige Objekt gewartet hatten oder zu lange für ihre Antwort brauchten. Dies zählte ebenfalls als ein Turn. In dieser Situation arbeitete der Dialog eine Liste von möglichen Objekten ab. Wurde keine Antwort gegeben, wurde einfach nach dem nächsten Objekt gefragt. Wurde auf andere Fragen keine Antwort gegeben, so wurde diese wiederholt. Auch dies zählte als ein vollständiger Turn.

Ein viel größeres Problem stellte das *Margin* dar. Dies konnte bewirken, dass der Roboter bereits eine neue Frage anfing, während die andere noch lief. Dieser Effekt trat auf, wenn Probanden zwei statt einer Antwort gegeben hatten, sich also korrigierten oder eine Pause bei der Äußerung machten, aber eventuell bei einem Seufzen oder Schmatzen. Wartete der Benutzer nicht, bis die Frage zu Ende formuliert war, konnte dies zu einer unkontrollierten Eskalation des Dialoges ausarten.
Da die Probanden im Voraus nicht zu viel über das System wissen durften, wurden sie auch nicht über dieses Phänomen aufgeklärt. In einer derartigen Situation unterbrach sich der Avatar selbst, um das Ende diese Eskalation zu beschleunigen. Das Werten einer Unterbrechung als vollständigen Turn wäre aber unangebracht gewesen und hätte die Ergebnisse verfälscht. Ein Seufzer, der das eventuell auslöste, sollte die das Ergebnis eines Probanden komplett ruinieren; Vor allem da einige Probanden trotz dieser Situation einen Erfolg erzielten.
Es wurde entschieden, dass diese Situation mit maximal drei Turns berechnet wird. Sobald das System sich einmal oder zweimal unterbricht, sind es zwei Turns, bei drei oder mehr Unterbrechungen werden drei Turns gezählt.

5.3.2 Bewertung und Ergebnis

Eine objektive Bewertung der Ergebnisse stellte eine Herausforderung dar. Der naive Ansatz $\frac{Dialogerfolge}{Turns}$ enthält die Idee, dass eine höhere Anzahl an Turns auch zu einer niedrigeren Bewertung führt.
Eine Analyse der Videoaufnahmen und der Log-Dateien der Experimente ergab, dass dieser Ansatz der Praxis nicht gerecht geworden wäde. Ein Proband, der einen Dialogdurchlauf sehr erfolgreich und einen Durchlauf mäßig erfolgreich hinter sich brachte, im dritten Durchlauf aber Probleme hatte und die Anzahl der Turns in die Höhe trieb, schnitt dann schlechter ab, als einer mit nur einem Erfolg, aber insgesamt weniger Turns.
Zudem wurde festgestellt, dass gerade am Ende, bei der Objektbestätigung, häufig Fehler auftreten. Das System wartete nicht lange genug auf die Antwort des Probanden (siehe Seite 52). Sagte dieser etwas zu spät „Yes“, so entschied sich der Dialogmanager für das falsche Objekt. Bei der Verneinung sah es ähnlich aus. Ein verspätetes „No“ konnte dann das richtige Objekt ausschließen.
Aus diesem Grund wurde ein Punktesystem entworfen, um diese Eventualitäten berücksichtigen zu können. Wurd das Ziel erreicht und genau der gewünschte Gegenstand identifiziert, dann ggab es hierfür drei Punkte. Wurde ein Gegenstand vom gleichen Typ (sowohl Farbe als auch Form) identifiziert, so wurde der erste Teil des Dialogdurchlaufs erfolgreich abgeschlossen, lediglich die abschließende Objektbestätigung wurde nicht bewältigt. Diese Situation wird als *Teilerfolg* bezeichnet und gibt einen Punkt. Wurde das Dialogziel sehr schnell erreicht, so gibt es hierfür einen Bonus. Für einen perfekten Durchlauf gab es zwei Extrapunkte, bei nur wenigen Fehlern immer noch einen.
Für ein gescheitertes Dialogziel (weder Erfolg noch Teilerfolg) gab es keine Punkte, unabhängig von der Anzahl der Turns.
Tabelle 5.2 zeigt die genaue Punktvergabe.

Die unterschiedliche Verteilung bei den Objekten kam daher, dass der Objekttyp von Objekt1 einmalig im Szenario vorkam. Eine Verwechslung, und damit auch ein Teilerfolg, war nicht möglich. Da eine Objektbestätigung demnach auch nicht notwendig war, ist die Anzahl der Turns für einen perfekten Durchlauf geringer.

	Turns	Punkte
Objekt1		
Erfolg	2	5
	3	4
	> 3	3
Objekt2 \Objekt 3		
Erfolg	4	5
	5–6	4
	> 6	3
Teilerfolg	4	3
	5–6	2
	> 6	1

Tabelle 5.2: Punktvergabe abhängig von Erfolg und Turns

In beiden Experimentteilen unterschieden sich zwar die Objekte, allerdings sind die Schwierigkeitsgrade ähnlich gewählt worden. Somit konnte die Bewertung auf beide Teile angewandt werden und die Ergebnisse wurden vergleichbar.

Im optimalen Fall waren also 15 Punkte in 10 Turns erreichbar. Der oben genannte naive Ansatz wurde auf dieses System angewendet und auf 100 normiert. Der sich ergebende Wert wird als *Score* bezeichnet und berechnet sich durch die Formel 5.1.

$$Score = \frac{Punkte1 + Punkte2 + Punkte3}{Turn1 + Turn2 + Turn3} * \frac{10}{15} * 100 \tag{5.1}$$

Mit diesem Score wurden nun alle Dialoge der Benutzerstudie ausgewertet. Das Ergebnis ist der Tabelle 5.3 zu entnehmen. Hier wird nur der Durchschnitt der Scores gezeigt, eine genaue Auflistung liegt in den Tabellen in Anhang C vor.
Teil 1 bzw. *Teil 2* geben das Ergebnis aus den jeweiligen Experimentteilen wieder. *Gruppe 1* benutzte im ersten Teil den *emotionalen Dialog* und im zweiten den *neutralen. Gruppe 2* begann entsprechend mit dem *neutralen Dialog.* In der Spalte ∅ *Gesamt* wurde der Gesamtscore ausgewertet, der alle sechs Dialogdurchläufe umfasst. Dagegen listet Spalte ∅ *Summe 1+2* die Summe der beiden ersten Scores auf.

	∅ Teil 1	∅ Teil 2	∅ Gesamt	∅ Summe 1+2
alle	29,9	28,2	28,2	58,1
Gruppe 1	27,2	30,2	28,3	57,4
Gruppe 2	32,6	26,2	28,2	58,8

Tabelle 5.3: Durchschnittswerte nach Dialogauswertung (Gruppe 1: Start emotional; Gruppe 2: Start neutral)

Insgesamt haben beide Gruppen nahezu gleich abgeschnitten. Der Score des jeweils emotionalen Teils fällt aber in beiden Gruppen schlechter aus. Dieses Ergebnis widerspricht den Erwartungen und soll nun die Grundlage der kommenden Diskussion 5.4 sein.

5.4 Diskussion

Der Prototyp wurde mit dem Gedanken entwickelt, durch eine emotionale Benutzerführung das Erlernen des Dialoges zu vereinfachen und gleichzeitig die Effizienz zu verbessern. Wie Tabelle 5.3 aber zeigt, konnte diese Erwartung nicht erfüllt werden. Das Gegenteil ist eingetreten. In diesem Abschnitt sollen die Ursachen dieser Resultate diskutiert und nach möglichen Lösungen gesucht werden.

Es muss hervorgehoben werden, dass beide Gruppen nahezu gleich abgeschnitten haben. Man kann also davon ausgehen, dass beide Gruppen unter gleichen Bedingungen die Experimente durchführten und die Einteilung in die Gruppen willkürlich genug war.

In der Kritik wurde geäußert, dass Emotionen das Erlernen vereinfachen würden und später bei der eigentlichen Verwendung des Systems unnötig seien. Demnach wäre davon auszugehen, dass Gruppe 1, die mit den emotionalen Dialogen begonnen hat, besser abschneidet als Gruppe 2. Dies ist zwar der Fall, der Unterschied ist aber so verschwindend gering, dass er keine statistische Relevanz darstellt.

Bei Gruppe 1 fällt der zweite Teil besser aus, als der erste. Eine mögliche Interpretation wäre, dass im ersten Teil das System noch kennen gelernt und dieses Wissen später angewandt wurde. Dies würde auch dem Gedanken der oben genannten Kritik entsprechen.
Diese Interpretation ist jedoch falsch. Betrachtet man Gruppe 2, so müsste diese nach dem Einlernen im zweiten Teil auch besser abschneiden. Hier ist aber das genaue Gegenteil der Fall, es tritt eine deutliche Verschlechterung ein. Würden Emotionen den Dialogfluss verbessern, so müsste hier der zweite Teil besser ablaufen.

Interessant ist ebenfalls der kreuzweise Vergleich. Vergleicht man von beiden Gruppen jeweils den neutralen Teil, so schneiden beide etwa gleich gut ab. Gruppe 2 ist ein wenig besser, was überraschend ist, denn schließlich hatte Gruppe 1 zu diesem Zeitpunkt des Experimentes bereits etwas Erfahrung.
Stellt man aber nun die Werte der emotionalen Dialoge gegenüber, so schneidet diesmal Gruppe 2 etwas schlechter ab.
Insgesamt haben also beide Gruppen in den jeweiligen Teilen etwa gleich abgeschnitten. Im zweiten Teil schnitten beide Gruppen im Vergleich schlechter ab. Dies sollte aber nicht am Schwierigkeitsgrad dieses Teils liegen. Wäre dieser erheblich höher, würden die Ergebnisse der beiden Gruppen nicht so nah beieinander liegen. Die Ursache ist wahrscheinlich bei dem Verhalten der Probanden selbst zu finden. Durch deren Erfahrungen aus dem ersten Teil gehen diese anders mit dem System um. Dies zu untersuchen sprengt jedoch den Bereich dieser Arbeit, in der nur technische Aspekte des Prototyps und deren Auswirkungen auf den Dialogverlauf betrachtet werden.

Die Folgerung aus diesen Ergebnissen ist, dass die geplante emotionale Benutzerführung keine Verbesserungen, sondern eine Verschlechterung mit sich bringt.
Eine mögliche Ursache ist ein zu undifferenziertes Szenario. Die Probanden wurden immer mit den gleichen Fragen konfrontiert. Durch ein komplexeres Szenario mit einer tieferen Objektstruktur und weiteren Modalitäten (z.B. Handschrifterkennung beim Unterzeichnen oder Emotionserkennung aus der Sprache) wäre hiermit Abhilfe zu schaffen. Der Dialog könnte so immer einen unterschiedlichen Verlauf nehmen, wodurch die Probanden ständig neuen Fragen ausgesetzt wären. Dies würde eine

objektive Messung und den anschließenden Vergleich der Ergebnisse erschweren, allerdings könnten sich auch auf diese Weise die Fähigkeiten der emotionalen Benutzerführung hervortun.
In der Kritik wurde formuliert, dass Emotionen ablenken würden. Da beide Gruppen im emotionalen Teil schlechter abschnitten, wurden sie wohl abgelenkt. Diese Behauptung ist also nachvollziehbar. Interessant wäre eine Untersuchung, in denen der Avatar nicht übertrieben reagiert, sondern Emotionen mit feinen Nuancen in der Mimik ausdrückt. Dies könnte diese Ablenkung unterbinden.

Um die Analyse zu festigen wird nun ein weiteres Messkriterium betrachtet, die Standardabweichung (Varianz). Die hierfür verwendeten Werte berechnen sich aus den Tabellen C.1 und C.2. Die Standardabweichung bei Gruppe 1 liegt beim emotionalen Dialog bei 18,0 und beim neutralen bei 13,4. Gruppe 2 hat eine Standardabweichung von 14,2 beim emotionalen und 13,6 beim neutralen Dialog. Die Werte im neutralen Dialog sind nahezu identisch und jeweils niedriger als beim emotionalen Dialog.
Wie sind nun diese Ergebnisse zu deuten? Die Ergebnisse beim emotionalen Dialog sind breiter gestreut. Beide Gruppen weisen unabhängig voneinander eine fast identische Standardabweichung beim neutralen Dialog auf, sowie einen sehr ähnlichen durchschnittlichen Score.
Da es beim neutralen Teil nur eine geringe Schwankung in den Werten gibt, kann man erwarten, dass sich diese Werte auch bei einer größeren Testgruppe nicht stark verändern würden. Dies müsste man aber durch eine Studie belegen. Interessant ist, dass die erfahrenere Gruppe 1 nicht besser abgeschnitten hat als Gruppe 2, die dieses System zum ersten mal benutzte. Diese Leistung scheint also vom Training unabhängig zu sein. 0

Beim emotionalen System sieht es etwas anders aus. Die höhere Standardabweichung lässt darauf schließen, dass einige Probanden sehr gut abgeschnitten haben, andere aber eher unerfolgreich waren. Da etwa die Hälfte der Probanden angaben, dass sie das neutrale System besser bedienen konnten, kann man davon ausgehen, dass das emotionale System keine universale Benutzerführung liefert. Tabelle 5.4 zeigt eine Aufstellung aller Experimentteile nach dem Score sortiert. Betrachtet man nun die oberen Plätze, so ist dort zu sehen, dass unter den ersten sechs Plätzen sowohl drei emotionale als auch drei neutrale Dialoge liegen. In diesen Teilen wurde jeweils ein Score über 40,0 erzielt. Anders sieht es im Mittelfeld aus. Einen Score zwischen 20,0 und 40,0 erreichten neun Probanden. Die Verteilung von emotional und neutral verteilt sich hier aber anders. In diesem Bereich liegt siebenmal ein neutraler Dialog und entsprechend nur zweimal ein emotionaler. Das genau umgekehrte Verhältnis weisen die unteren neun Plätze auf.

Daraus kann man schließen, dass es für Probanden, die das System gut beherrschten, keinen Unterschied machte, ob der Dialog emotional oder neutral war. Bei den anderen Probanden war allerdings ein Unterschied feststellbar. Diese schienen durch die Emotionen irritiert worden zu sein und schnitten entsprechend schlechter ab. Dies erklärt auch das bessere Abschneiden und die geringerer Standardabweichung beim neutralen Dialog.
Abb. 5.1 stellt diese Hypothese schematisch dar. Die grüne Linie repräsentiert die Scoreverteilung bei einem emotionalen Dialog, die magentarote Linie bei einem neutralen Dialog. Wie zu erkennen ist, sind bei hohen Scores die Verteilungen ähnlich, unterscheiden sich aber bei niedrigeren Scores. Die türkise Linie zeigt eine angestreb-

	ID	Score	Teil	Dialog
1	14CK	57,8	1	emotional
2	17CH	53,3	1	neutral
3	22MV	50,0	2	neutral
4	21TP	45,8	2	emotional
5	24CK	44,4	2	neutral
6	19ST	43,1	2	emotional
7	13DC	38,1	1	neutral
8	19ST	35,6	1	neutral
9	22MV	35,3	1	emotional
10	20HS	29,6	1	emotional
11	23JS	29,6	1	neutral
12	21TP	26,7	1	neutral
13	12CM	24,6	2	neutral
14	16FK	22,2	2	neutral
15	20HA	21,1	2	neutral
16	13DC	19,1	2	emotional
17	18AP	19,1	2	emotional
18	16FK	18,5	1	emotional
19	17CH	17,4	2	emotional
20	15AU	15,9	2	emotional
21	23JS	15,9	2	emotional
22	18AP	13,9	1	emotional
23	15AU	12,4	1	neutral
24	12CM	8,0	1	emotional

Tabelle 5.4: Platzierung der einzelnen Experimentteile nach Score sortiert

te Verteilung, die vielleicht durch eine besser abgestimmte emotionale Benutzerführung erreicht werden kann.

Diese Theorie wird durch eine Betrachtung von Tabelle 5.5 untermauert. Hier wird der *Gesamtscore* ausgewertet, also der Wert, der sich ergibt, wenn man den Score auf alle Turns und alle Punkte (alle 6 Durchläufe) anwendet. Auch hier tritt das eben geschilderte Phänomen auf. Die oberen vier Plätze werden von zwei Probanden belegt, die mit dem emotionalen und zwei mit dem neutralen Dialog begonnen hatten. Auf den mittleren vier Plätzen befinden sich drei mit einem neutralen Beginn und entsprechend auf den letzten vier Plätzen drei mit einem emotionalen Dialog als ersten Experimentteil. Da hier jeweils die Gesamtleistung des Probanden bewertet wird, ist diese Beobachtung nicht unbedingt zu erwarten, deckt sich aber Aber auch hier lässt sich das Ergebnis so interpretieren, dass die Probanden mit einem emotionalen Start eher irritiert wurden und ein falsches Bild von dem System bekamen. Daher haben sie auch insgesamt schlechter abgeschnitten. Die Probanden auf den oberen Plätzen kamen mit dem System so gut zurecht, dass diese Verwirrung nicht eintrat.

Die Tendenz, die bei der Emotionalität der Dialoge zu finden ist, lässt sich nicht auf die Reihenfolge der Experimente übertragen. Bei der Betrachtung dieses Ergebnisses in Tabelle 5.4 ist kein Muster bezüglich des Experimentteils zu erkennen. Hier wirkt die Verteilung willkürlich. Es scheint also keinen großen Unterschied zu machen, ob

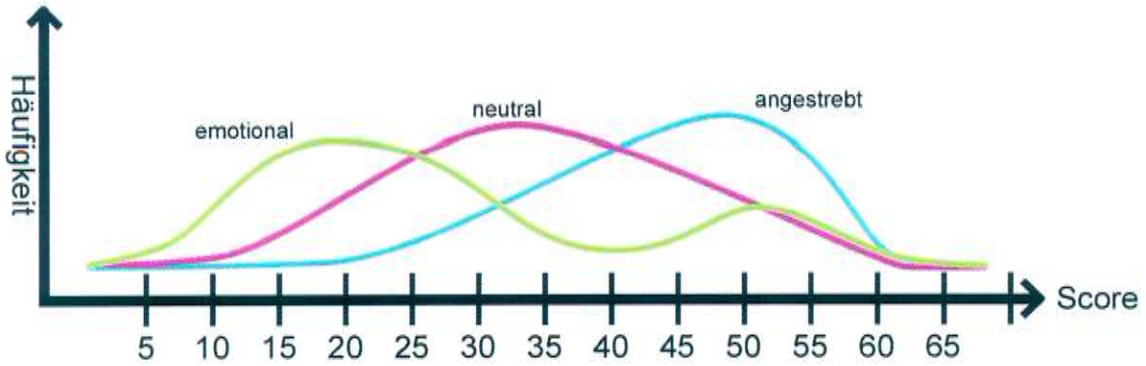


Abbildung 5.1: Schematische Darstellung der erwarteten Scoreverteilung. grün: emotionaler Dialog; magentarot: neutraler Dialog; türkis: angestrebte Verteilung

	ID	1. Teil	Score
1	14CK	emotional	51,1
2	22MV	emotional	41,4
3	19ST	neutral	39,6
4	21TP	neutral	35,2
5	17CH	neutral	31,6
6	13DC	neutral	26,7
7	20HS	emotional	25,2
8	23JS	neutral	22,2
9	16FK	emotional	20,5
10	18AP	emotional	16,3
11	12CM	emotional	15,2
12	15AU	neutral	13,9

Tabelle 5.5: Platzierung der Experimente nach Gesamtscore sortiert

der Proband zuerst den neutralen oder emotionalen Dialog durchführt.
Dies ist wieder ein Hinweis darauf, dass das Szenario nicht komplex genug gewählt wurde. Mit einer höheren Komplexität sollte zu erwarten sein, dass die zweiten Experimentteile besser abschneiden, da dort die Probanden von der Erfahrung aus dem ersten Teil profitieren.

Der berechnete Score enthält sowohl die *Effizienz* als auch das Erreichen des Dialogziels. Die Absicht beim Entwurfs des System lag darin, durch eine emotionale Benutzerführung ein besseres Erlernen und eine höhere Effizienz zu erreichen. Dass Ersteres nicht erfolgreich war, wurde hier bereits diskutiert. Im Folgenden soll nun die Effizienz betrachtet werden. Hierfür ist die *Anzahl der Turns* ausschlaggebend. Geht man davon aus, dass eine bessere Technik den Erfolg des Dialoges sichert, so ist nur noch diese Zahl interessant. Es könnte ja durchaus sein, dass der emotionale Dialog zu einem schnelleren Ende, jedoch mit Misserfolg führte.

Auch hier belegen die Zahlen ein ähnliches Phänomen wie bei der Betrachtung des Scores. Gruppe 1 brauchte insgesamt im Durchschnitt 37,3 Turns und erzielte 14,7 Punkte. Gruppe 2 benötigte hier durchschnittlich 38,0 Turns und erreichte 15,5 Punkte. Auch hier liegen die Zahlen sehr nahe beieinander. Dies ist bei einem fast identischen Gesamtscore auch zu erwarten. Beide Gruppen haben also fast die gleiche Anzahl an Turns für das Experiment gebraucht.

Beim Betrachten der einzelnen Teile benötigte Gruppe 1 für den ersten Teil durchschnittlich 19,5 Turns und erzielte 7,2 Punkte. Im zweiten Teil, der aus dem neutralen Dialog bestand, wurden nur 17,8 Turns gebraucht und 7,5 Punkte erzielt. Beide Werte haben sich also der neutralen Versuchsumgebung verbessert. Bei Gruppe 2 sieht es ähnlich aus. Im Durchschnitt benötigten die Probanden im ersten, neutralen Teil 18,2 Turns und erzielten 8,2 Punkte. Im zweiten Teil verschlechterten sich beide Werte auf 19,8 Turns und 7,3 Punkte.
Beide Gruppen benötigten für den emotionalen Teil etwa 1,7 Turns weniger, dazu erreichten beide Gruppen im neutralen Teil eine höhere Punktzahl. Der *neutrale Dialog erzielt demnach eine höhere Effizienz.*

Es bleibt offen, warum bei der Frage nach der Beherrschung des Systems das emotionale System leicht besser abgeschnitten hat, in der Praxis aber das Gegenteil eingetreten ist.
Folgende drei Fälle sind bei der Befragung möglich.

- Der Proband stufte die Beherrschung *beider System gleich* ein. Dies ist fünfmal geschehen.
- Der Proband stufte sich beim *emotionalen System besser* ein. Dieser Fall ist viermal eingetreten.
- Der Proband stufte sich beim *neutralen System besser* ein. Diesen Fall gab es dreimal.

Nur die Hälfte der zwölf Probanden hatten mit ihrer Einschätzung Recht. Davon haben zwei angeben, dass sie den neutralen Dialog besser beherrschten und vier meisterten entsprechend den emotionalen Experimentteil besser.
Bei fünf Probanden trat ein interessantes Phänomen auf. Diese schienen von einem *falschen Eindruck des emotionalen Systems* getäuscht worden zu sein. Nicht nur, dass sie sich bei der Einschätzung irrten, sie stuften sich beim emotionalen System besser ein, als sie tatsächlich waren. So dachten vier Probanden, dass sie etwa gleich gut mit beiden Systemen umgegangen seien, haben aber in Wirklichkeit mit dem emotionalen System ihre Schwierigkeiten gehabt. Lediglich bei einem Probanden ist dieser Effekt nicht aufgetreten. Dieser stufte sich bei beiden System gleich ein, schnitt aber im emotionalen Teil besser ab.

Dies erweckt den Eindruck, dass *Emotionen über Misserfolge hinweg täuschen.* Dieses Phänomen ist bereits Clifford Nass [NaBr05] aufgefallen. Dem Probanden war demnach nicht richtig bewusst, dass er gerade ineffizient mit dem System umging. Durch diese Manipulation durch den Avatar kann die Langzeitmotivation steigen indem Frust bei der Bedienung verhindert wird. Allerdings verhindert diese Täuschung auch, dass die Effizienz weiter gesteigert wird, da der Proband nicht die Notwendigkeit sieht sich weiter und detaillierter mit dem System zu beschäftigen.

5.5 Schlussfolgerung

Insgesamt ergibt sich aus dem Experiment, dass die Erweiterung des Dialogsystems technisch funktioniert, auch wenn sie noch nicht ganz ausgereift ist.

Die Emotionen konnten zufriedenstellend vermittelt werden. Diese Vermittlung geschah hauptsächlich durch visuelle Faktoren. Die von der Emotion abhängigen Änderungen in der Sprache wurden nur von wenigen bemerkt.

Das gewünschte Ziel der Effizienzsteigerung durch Emotionen konnte nicht erreicht werden. Beide Testgruppen schnitten beim emotionalen Dialog durchschnittlich schlechter ab. Im Vergleich benötigten die Benutzer mit dem emotionalen System etwa 9% mehr Turns als mit dem neutralen System. Dabei konnten im emotionalen Teil auch keine Verbesserung bezüglich der Dialogerfolge erzielt werden. Betrachtet man den Score, so schneidet Gruppe 1 im emotionalen Teil um 11,0% und Gruppe 2 um 22,9% schlechter ab.

Die Verärgerung des Avatar wurde etwas zu übertrieben dargestellt. Die Idee, die Antworten durch die Rückfrage des Systems vorzugeben, ging jedoch auf. Auch wenn viele durch diese Emotion verwirrt wurden, übernahmen einige das Antwortschema. Es stellte sich aber heraus, dass dieses Schema zu knapp ist. Hier sollte noch an einer geeigneten Formulierung gearbeitet werden. Andere änderten ihr Verhalten, da sie nun einen störrischen Roboter erwarteten, was nicht der Fall war. Die Erwartungen des Benutzers konnten in diesem Fall auch nicht erfüllt werden.

Die spontanen Affekte des Avatar wurden positiv wahrgenommen. Viele Benutzer konnten sich somit auf die nächste Frage vorbereiten. Dies entspricht den Erwartungen beim Entwurf. Durch kurze verbale Äußerungen wie „Super!“ oder „Mist!“ könnten diese Affekte weiter untermalt werden, so dass sie auch von denen wahrgenommen werden, die sie nicht bemerkt hatten.

Ein großer Kritikpunkt ist die Sprachsynthese gewesen. Da viele Benutzer dies nicht gewöhnt sind, kam es hier gelegentlich zu Kommunikationsproblemen. Eine Übersetzung des Dialoges auf Deutsch könnte hier bereits helfen. Man sollte aber in zukünftigen Experimenten mehr auf die Sprachqualität achten. Ein kommerzielles System hätte wahrscheinlich besser funktioniert.

Bei der Befragung schnitten beide Systeme sehr unterschiedlich ab. Dies ist ein Hinweis darauf, dass der Wunsch nach Emotionen im Dialog individuell ist. Mehr Emotion bedeutet demnach nicht unbedingt mehr Sympathie. Auch bei den Angaben zu den Lernerfolgen gingen die Meinungen auseinander. Eine universelle Benutzerführung ist demnach nicht zu erwarten. Allerdings könnte man diese Theorie auf verschiedene Benutzergruppen anwenden. Das System passt sich dann seiner Zielgruppe an. Ingenieure haben an einen Roboter andere Erwartungen als andere Berufsgruppen.

Im Vergleich der einzelnen Probanden hat sich ergeben, dass denen, die das System sehr gut beherrschten, die Emotionalität egal ist. In diesem Bereich schnitten beide Systeme etwa gleich ab. Die Dialoge, die aber sehr schlecht liefen, waren meist emotional, die mit mittlerem Score waren meist neutral. Also gerade die Probanden, die mit dem System nicht so gut zurecht kamen, wurden von den Emotionen abgelenkt und nicht geführt.

Interessant ist, dass bei der Bewertung des Systems durch den Benutzer das emotionale System leicht besser abschnitt, als das neutrale System. Emotionen scheinen also den Eindruck von Erfolg zu vermitteln. Dieser wird anscheinend bewusster oder intensiver erlebt. Es wird über den Misserfolg hinweggetäuscht, was sich auch auf die Motivation des Benutzer auswirkt.

Bei einzelnen Probanden konnten Emotionen interessante Reaktionen hervorbringen. Ein Proband fühlte sich vom freundlichen Avatar bedroht, da er ihm zu nahe kam und seinen persönlichen Freiraum verletzte. Ein anderer Proband änderte sogar seine Art und Weise mit dem System zu sprechen, als sich der verärgerte Avatar zurückzog. Er erwartete einen sturen Roboter und änderten sein Vorgehen. Ein Proband gab an, dass ihn die Emotionen beruhigten, ein anderer meinte, dass der neutrale Avatar eher beruhigend wirkt.

Das Szenario wurde schnell überblickt und die Funktionsweise des Dialogs von den meisten verstanden. Die Lernerfolge sind daher nicht gut messbar. Ein komplexeres System würde hier mehr Aufschluss geben. Dazu wäre es interessant eine weitere Testgruppe zu haben, denen vor dem Experiment der Dialog detailliert erklärt wird. Der Vergleich dieser Ergebnisse würde weitere Erkenntnisse für eine gezielte Benutzerführung ergeben.

6. Zusammenfassung und Ausblick

6.1 Zusammenfassung

In Bezug auf die Forschung im Bereich *Affective Computing* wurde der Einfluss von Emotionen auf die Mensch-Maschine Kommunikation untersucht. Ein Prototyp des Roboters *ARMAR 3* konnte um einen virtuellen Avatar erweitert werden, der in der Lage, ist Emotionen sowohl durch Mimik als auch Sprache auszudrücken.

Im Zuge dieser Arbeit wurde das Emotionssystem *EmoMaps* entwickelt, das an das Arousal-Valence Modell angelehnt ist. Diese Technologie ermöglicht einen flexiblen und schnellen Entwurf einfacher Emotionsmodelle. Für die grafische Darstellung wurde das System *Baldi* verwendet, welches bereits häufig in wissenschaftlichen Untersuchungen zum Einsatz kam. Die Sprache wurde durch das freie TTS-System *OpenMary* erzeugt, das in der Lage ist, emotionale Sprache sowohl in Englisch als auch Deutsch zu erstellen.
Den Austausch an Information zwischen diesen Komponenten ermöglichte der Wrapper *BMM*, der für diese Anwendung speziell entwickelt wurde. BMM wurde in Java implementiert und kann über den Dialogmanager *TAPAS* angesteuert werden.

In einer abschließenden Benutzerstudie wurde untersucht, ob Emotionen als Darstellung von Dialogerfolgen Einfluss auf die Effizienz der Benutzung haben. Hierfür wurden Probanden gebeten in einem Barkeeper-Szenario dem Roboter einen Gegenstand zu erklären. Es handelte sich um einen multimodalen Dialog, in dem sowohl durch Sprache als auch durch Gestik kommuniziert wurde. Die Probanden wurden in zwei Testgruppen aufgeteilt. Bei der ersten Testgruppe reagierte der Avatar zuerst immer emotional, abhängig vom Dialogerfolg. In einem zweiten Experimentabschnitt blieb der Avatar stets neutral. Die zweite Gruppe führte den Dialog erst mit dem neutralen, dann mit dem emotionalen Avatar durch. Die Emotionen drückten sich durch Mimik und Position des Avatar aus, sowie durch Sprache und Wortwahl. War der Avatar durch Fehler bei der Benutzung verärgert, so wurden in seine Rückfragen Hinweise auf die Bedienung eingebaut. Durch diese Benutzerführung erhoffte man sich ein schnelleres Erlernen des Dialogs und eine daraus folgende Effizienzsteigerung bei der Benutzung.
Die Studie zeigte, dass der Avatar Emotionen gut vermitteln konnte. Die gewünschte

Effizienzsteigerung blieb allerdings aus. Beide Testgruppen schnitten mit dem emotionalen Dialog schlechter ab. Durchschnittlich benötigten sie 9% mehr Dialogturns als mit dem neutralen Dialog. Die emotionale Benutzerführung konnte auch nicht zu mehr Dialogerfolgen führen, als dies bei dem neutralen Dialog der Fall war. Interessanterweise gaben die meisten Probanden an, den emotionalen Dialog besser beherrscht zu haben.

Man kann daraus folgern, dass das entwickelte System funktioniert und Emotionen gut vermittelt werden. Es wurde ebenfalls gezeigt, dass Emotionen durchaus auch Einfluss auf die Benutzung eines Dialogs haben. Durch weitere Verbesserungen sollte eine gezielte emotionale Benutzerführung möglich sein, die das Erlernen eines Dialog vereinfacht.

6.2 Ausblick

Innerhalb dieser Arbeit wurde das bestehende Dialogsystem um die Fähigkeit Emotionen ausdrücken zu können erweitert. Dass diese Emotionen Einfluss auf die Dialogführung haben, wurde durch Experimente bestätigt. Diese Auswirkung entsprach allerdings nicht den Erwartungen. Mögliche Ursachen wurden analysiert und diskutiert.

Die Darstellung der *Affekte* bei Erfolg oder Misserfolg, erwies sich als positiv. Die Probanden erkannten früher, ob sie verstanden wurden oder nicht und stellten sich auf die nächste Frage ein.

Folgende Punkte wurden jedoch an der Konfiguration des emotionalen Dialogs kritisiert und sollten in weiterführenden Untersuchungen beachtet werden. Die Kritikpunkte beziehen sich dabei nicht ausschließlich auf die Emotionalität, sondern gehen auch auf andere Aspekte des Dialogsystems ein.

- Die *synthetische Sprache* wurde von vielen Probanden nicht verstanden. Ein ausgereifteres System wäre für den nächsten Prototypen sehr wichtig. Eine andere Möglichkeit wäre den Dialog in *deutscher Sprache* zu gestalten, um auch die englische Sprachbarriere zu überwinden.

- Die *Wartezeiten* zwischen den Fragen sollten verlängert werden oder zumindest an die Länge der Frage angepasst werden. Lang formulierte Fragen, wie es beim verärgerten Avatar der Fall ist, ließen dem Probanden kaum Zeit für eine Antwort.

- Der Proband sollte eine Möglichkeit des *Neustarts* oder einer *Zurück-Funktion* bekommen. Korrekturen am gewählten Objekt sind sonst nur bis zu einem gewissen Punkt machbar. Ist dieser überschritten, kann bei einem Fehler der Dialog nicht mehr erfolgreich beendet werden.

- Die *Komplexität* stellte sich als zu gering heraus. Die Funktionsweise des Dialogs wurde sehr schnell verstanden. Eine Erweiterung des Szenarios, in dem sich der Proband z.B. dem Roboter vorstellen muss, oder Hinzunahme weiterer Modalitäten würde die Komplexität steigern und somit die Unterschiede beim Lernverhalten weiter hervorheben.

- Die *Verteilung der Objekte* auf dem Tisch erwies sich als ungünstig. Ähnliche Objekte waren zu nah beieinander, so dass die Gestikerkennung Probleme hatte. Durch Hinzunahme von Farben und Objekttypen (z.B. grüne Gabel, weißes Messer) sollte es möglich sein unterschiedliche Schwierigkeitsgrade der Objekte festzulegen. Die Aufgaben des Probanden während den Experimenten könnte somit raffinierter gestaltet werden.

- Die *Position* des Avatar auf dem Bildschirm wurde übertrieben. Ein glücklicher Avatar sollte nicht zu nahe kommen, ein verärgerter sollte sich gar nicht oder zumindest nicht so weit weg bewegen. Diese extreme Darstellung verschreckte eher die Probanden.

Der *negative emotionale Zustand* sollte aber auf jeden Fall beibehalten werden, auch wenn der Ausdruck dieser Emotion nicht zu aggressiv sein darf. Auf diese Weise erweckt der Roboter immer noch den Eindruck eines Butler, beeinflusst aber den Benutzer mit seinen erweiterten Fragen. Die Formulierung sollte aber auf z.B. *„Do you want a cup, plate, or bottle?“* geändert werden. Eventuell lassen sich auch Beispielantworten anhängen oder man kann den Benutzer darauf hinweisen, dass eine Kombination wie „red bottle“ oder „blue plate“ ebenfalls möglich ist und die Bedienung des Dialogs weiter beschleunigt.

Eine größer angelegte Fortsetzung dieser Studie, unter Berücksichtigung der genannten Verbesserungsvorschläge, wird weitere Erkenntnisse über das Lernverhalten eines Benutzers hervorbringen. Diese Erkenntnisse können Aufschluss darüber geben, wie Emotionen eingesetzt werden können, um den Benutzer gezielt durch den Dialog zu führen und dadurch die Effizienz beim Erlernen und bei der Bedienung des Systems zu steigern.

Literatur

[Alan97] Paul Taylor Alan Black (Hrsg.). The Festival Speech Synthesis System: System documentation. Technical report hcrc/tr-83, Human Communciation Research Centre, University of Edinburgh, 1997.

[Alan06] Paul Taylor Alan Black. The Festival Speech Synthesis System. http://www.cstr.ed.ac.uk/projects/festival/, Centre for Speech Technology Research, University of Edinburgh), 2006.

[Bart02] Christoph Bartneck (Hrsg.). Integrating the OCC Model of Emotions in Embodied Characters. Proceedings of the workshop on virtual conversational characters: Applications, methods, and research challenges, melbourne, Department of Industrial Design, Eindhoven University of Eindhoven, 2002.

[Bate94] Joseph Bates (Hrsg.). The Role of Emotion in Believable Agents. Communications of the acm, School of Computer Science, Carnegie Mellon University, 1994.

[Brea00] Cynthia Breazeal (Hrsg.). Sociable Machines: Expressive Social Exchange Between Humans and Robots. Dissertation, Department of Electrical Engineering and Computer Science, MIT, 2000.

[Burk03] Felix Burkhardt (Hrsg.). Emofilt: the Simulation of emotional Speech by Prosody-Transformation. Germany, T-Systems International GmbH, 2003.

[Cole99] Ron Cole (Hrsg.). Tools for research and education in speech science. In proceedings of the international conference of phonetic sciences, Center for Spoken Language Understanding (CSLU), August 1999.

[Dene05] Matthias Denecke (Hrsg.). Rapid Prototyping for Spoken Dialogue Systems. Proceedings of the 19th international conference on computational linguistics, Carnegie Mellon University, 2005.

[DuJo05] Brian Duffy und Gina Joue (Hrsg.). Why Humanoids? 4th chapter conference of applied cybernetics, Affective Social Computing Laboratory, Eorecom, France, 2005.

[Ekma99] Paul Ekman (Hrsg.). Basic Emotions. In t. dalgleish and m. power (eds.). handbook of cognition and emotion. sussex, u.k., University of California, San Francisco, 1999.

[FuCa96] Hiroya Fujisaki und Nick Campbell. *Computing Prosody.* Springer. 1996.

[Gole95] Daniel Goleman. *Emotional Intelligence.* Bantam Books. 1995.

[Hall70] Michael Alexander Kirkwood Halliday. *A Course in Spoken English: Intonation.* Oxford University Press. 1970.

[Hans03] David Hanson (Hrsg.). New robot face smiles and sneers. Newscientist.com news service, The Institute for Interacive Arts and Engineering (The University of Texas at Dallas), Februar 2003.

[Holz05] Hartwig Holzapfel (Hrsg.). Towards Development of Multilingual Spoken Dialogue Systems. Proceedings of the 2nd language and technology conference, Universität Karlsruhe (TH), 2005.

[Hond03] Honda. Technical Information.
http://asimo.honda.com/downloads/pdf/asimo-technical-information.pdf,
American Honda Motor Co.,Inc. and Corporate Affairs & Communications, Januar 2003.

[Izar93] C. E. Izard (Hrsg.). Four systems for emotion activation: cognitive and noncognitive processes. Psychological review, vol 100, no. 1, Department of Psychology, University of Delaware, 1993.

[JaKa03] C. Stephen Jaeger und Ingrid Kasten. *Codierung von Emotionen im Mittelalter.* Walter de Gruyter. 2003.

[jk06] jk. Sony schläfert Aibo ein.
http://www.heise.de/newsticker/meldung/68901,
heise online, Januar 2006.

[JLui06] Jörg Luibl. Heavy Rain, Gefühle & Spielezukunft.
http://www.4players.de/4players.php/dispnews/PlayStation3/Aktuelle_News/53416.html,
4Players online Magazine, Mai 2006.

[JoTh81] Ollie Johnston und Frank Thomas. *The Illusion of Life: Disney Animation.* Abbeville Press, New York. 1981.

[Katz37] David Katz. *Personal Space and its Mastery*, Kapitel 5, S. 94–97. Addison Wesley. 1937.

[Kö05] Martin Kölling. Asimo kann Händchen halten.
http://www.heise.de/newsticker/meldung/67330/,
heise online, Dezember 2005.

[Lab96] TCTS Lab. MBROLA project.
http://tcts.fpms.ac.be/synthesis/mbrola/,
Faculté Polytechnique de Mons (Belgium), 1996.

[LaBC90] Peter J. Lang, M. M. Bradley und B. N. Cuthberg. Emotion, Attention, and the Startle Reflex. *Psychological Review* 97(3), 1990, S. 377–395.

[Lang95] Peter J. Lang. The emotion probe: Studies of motivation and Attention. *American Psychlogist* 50(5), 1995, S. 372–385.

[McTe99] Michael McTear (Hrsg.). Software to Support Research and Development of Spoken Dialogue Systems. Eurospeech, budapest, romania, School of Information and Software Engineering, University of Ulster, September 1999.

[MiKD04] Jan Minar, Dteffen Knoop und Rüdiger Dillmann (Hrsg.). Visualisierung von Zuständen im Serviceroboter Albert. Diplomarbeit, Universität Karlsruhe (TH), 2004.

[MMMI06] Daisuke Matsui, Takashi Minato, Karl F. MacDorman und Hiroshi Ishiguro (Hrsg.). Generating Natural Motion in an Android by Mapping Human Motion. Proc. of ieee/rsj international conference on intelligent robots and systems, Department of Adaptive Machine Systems, Graduate School of Engineering, Osaka University, Intelligent Robotics and Communication Laboratories, Advanced Telecommunications Research Institute International, August 2006.

[MMor70] Masahiro Mori. Bukimi No Tani (The Uncanny Valley). *Energy* 7(4), 1970, S. 33–35.

[Mö01] Bernd Möbius (Hrsg.). German and Multilingual Speech Synthesis. Arbeitspapier, Institut für Maschinelle Sprachverarbeitung (Universität Stuttgart), 2001.

[Mö06] Bernd Möbius. German Festival synthesis system.
http://www.ims.uni-stuttgart.de/phonetik/synthesis/,
Institut für Maschinelle Sprachverarbeitung (IMS Universität Stuttgart), 2006.

[MOer06] Mathias Oertel. Die zehn wichtigsten Erkenntnisse.
http://www.4players.de/4players.php/dispbericht/PC-CDROM/Special/7876/4558/0/E3_2006.html,
4Players online Magazine, Mai 2006.

[Morr03] James H. Morris. Robot Hall of Fame.
http://www.robothalloffame.org/,
powered by Carnegie Mellon University, 2003.

[NaBr05] Clifford Nass und Scott Brave. *Wired for Speech.* The MIT Press. 2005.

[NaSu94] Clifford Nass und S. Shyam Sundar (Hrsg.). Is human-computer interaction social or parasocial? Human communication research, Stanfort SRCT, 1994.

[NiSW03] Kai Nickel, Rainer Stiefelhagen und Alexander Waibel (Hrsg.). Erkennung von Zeigegesten basierend auf 3D-Tracking von Kopf und Händen. Diplomarbeit, Institut für Logik, Komplexität und Deduktionssysteme (ILKD), Universität Karlsruhe, März 2003.

[OrCC88] Andrew Ortony, Gerald L. Clore und Allan Collins. *The Cognitive Structure of Emotions.* Cambridge University Press. 1988.

[PaWa96] Frederick I. Parke und Keith Waters. *Computer Facial Animation.* AK Peters. 1996.

[Pica95] Rosalind W. Picard (Hrsg.). Affective Computing. Technical Report, MIT Media Laboratory Perceptual Computing Section, November 1995.

[Plut80] Robert Plutchik. *Emotion: A Psychoevolutionary Synthesis.* Harpercollins College Div. 1980.

[PoWa98] Thomas Polzin und Alexander Waibel (Hrsg.). Detecting Emotions in Speech. Proceedings of the cmc 1998, School of Computer Science, CMU and Fakultät für Informatik, Universität Karlsruhe, 1998.

[ReNa03] Byron Reeves und Clifford Nass. *The Media Equation: How People Treat Computers, Television, and New Media Like Real People and Places.* Center for the Study of Language and Information. 2003.

[Robo01] SFB 588 Humanoide Roboter. Lernende und kooperierende multimodale Roboter.
http://www.sfb588.uni-karlsruhe.de/,
Deutsche Forschungsgemeinschaft (DFG), Juni 2001.

[Robo06] SONY Dram Robot. QRIOs Technology List.
http://www.sony.net/SonyInfo/QRIO/technology/index_nf.html,
SONY Global, Dezember 2006.

[Sche86] Klaus R. Scherer. Vocal affect expression: a review and a model for future research. *Psychological Bulletin* 99(2), 1986, S. 143–165.

[Schr06] Marc Schröder. OpenMary Text to Speech.
http://mary.dfki.de/,
Deutsches Forschungszentrum für Künstliche Intelligenz GmbH (DFKI), 2006.

[ScTr03] Marc Schröder und Jüurgen Trouvain (Hrsg.). The German Text-to-Speech Synthesis System MARY: A Tool for Research, Development and Teaching. Technical Report, DFKI GmbH, Institute of Phonetics, University of the Saarland, Saarbrücken, Germany, 2003.

[Simm03] Reid Simmons. GRACE: An Autonomous Robot for the AAAI Robot Challenge. *AAAI Magazine* 24(2), 2003, S. 51–72.

[Somm69] Robert Sommer. *Personal Space: Behavioural Basis of Design.* Spectrum Books. 1969.

[Ston99] Patrick Stone (Hrsg.). Revolutionizing Language Instruction in Oral Deaf Education. In proceedings of the international conference of phonetic sciences, Tucker-Maxon Oral School, Portland, Ot, USA, August 1999.

[SvGN04] Wenzel Svojanovksy, Rainer Gruhn und Satoshi Nakamura (Hrsg.). Classification of Nonverbal Utterances In Japanise Spontaneous Speech. Information processing society japan, A. T. R. Institute for Spoken Language Translation, März 2004.

[TAsf01] R. Dillmann Tamin Asfour (Hrsg.). Control of ARMAR for the Realization of Anthropomorphic Motion Patterns. The second ieeeras international conference on humanoid robots, Universität Karlsruhe (TH), Juni 2001.

[Tisc93] Bernd Tischer. *Die vokale Kommunikation von Gefühlen.* Belz Psychologie Verlags Union. 1993.

[Tool06] CSLU Toolkit. Talking Head Baldi.
http://cslu.cse.ogi.edu/toolkit/index.html,
Center for Spoken Language Understanding (CSLU), 2006.

[Tori01] Akira Toriyama. *Manga Zeichenkurs.* Carlsen Comics. 2001.

[TPro06] Hartwig Holzapfel Thomas Prommer (Hrsg.). Rapid Simulation-Driven Reinforcement Lerning of Multimodal Dialog Strategies for Human-Robot Interaction. Technical Report, Universität Karlsruhe (TH) / Carnegie Mellon University, Mai 2006.

[UlMa03] Dieter Ulich und Philipp Mayring. *Psychologie der Emotionen.* Kohlhammer. 2003.

[Waib88] Alexander Waibel. *Prosody and speech recognition.* Morgan Kaufmann Publishers and Pitman. 1988.

[Webe06] Jutta Weber. Der Roboter als Menschenfreund. *c't Magazin für Computer Technik* 26(2), 2006, S. 144–149.

[WhIs05] David Whitehouse und Hiroshi Ishiguro. Japanese develop female android.
http://news.bbc.co.uk/1/hi/sci/tech/4714135.stm,
BBC online, Juli 2005.

[Wiki06a] Wikipedia. Allgemeine Relativitätstheorie.
http://de.wikipedia.org/wiki/Allgemeine_Relativitätstheorie,
Wikipedia, The Free Encyclopedia, 2006.

[Wiki06b] Wikipedia. Uncanny Valley.
http://en.wikipedia.org/wiki/Uncanny_valley,
Wikipedia, The Free Encyclopedia, 2006.

A. Fragebogen

Allgemeiner Fragebogen

Alter: _____

Geschlecht:

- ☐ männlich
- ☐ weiblich

Letzter Bildungsabschluss (Diplom, Vordiplom, Abitur,...):

_____

Berufsbezeichnung:

_____

Fachrichtung (Informatik, Finanzwesen, Sozial, ...):

_____

Englischkenntnisse (mehrfach ankreuzen möglich):

- ☐ Weniger als 7 Jahre Schulausbildung
- ☐ 7 oder mehr Jahre Schulausbildung
- ☐ Weitere Vertiefung in Englisch (LK, Sprachkurse, ...)
- ☐ Muttersprache
- ☐ Mehrwöchiger Aufenthalt im englischsprachigen Ausland
- ☐ Beruflich häufig mit Englisch zu tun

Erfahrungen mit dem System
(1— nie benutzt , 2— geringe praktische Erfahrung, 3— gelegentliche (durchschnittliche) Nutzung, 4— häufige Nutzung, 5—professionelle Nutzung):

Alltäglich Computer (Email, Surfen, Word,...)
1☐ 2☐ 3☐ 4☐ 5☐

Fortgeschrittene Computernutzung (Gerätesteuerung, Netzwerkadministration)
1☐ 2☐ 3☐ 4☐ 5☐

Spracherkennung (Privatcomputer, Kartenreservierung)
1☐ 2☐ 3☐ 4☐ 5☐

Erfahrung mit Robotern (maschinelle Fertigung, ...)
1☐ 2☐ 3☐ 4☐ 5☐

Alternative Eingabemöglichkeiten (Spacemouse, Tablet, Phantom, Datenhandschuh, ...)
1☐ 2☐ 3☐ 4☐ 5☐

Erfahrung mit diesem speziellen System (bei anderen Experimenten mitgewirkt,...)
1☐ 2☐ 3☐ 4☐ 5☐

Abbildung A.1: Fragebogen zum Benutzer

Fragebogen zum emotionalen Dialog

Hier sind jeweils deine persönlichen Eindrücke einzutragen. Bitte überlege bei deinen Antworten nicht lange und antworte spontan.

Wie kamst du insgesamt mit dem System zurecht?
☐ überhaupt nicht ☐ einigermaßen ☐ recht gut ☐ super

Sind dir Veränderungen des Verhaltens während des Dialogs aufgefallen?
☐ ja ☐ nein

Welche Veränderungen waren auffällig?

_____

_____

_____

Welche Emotionen hatte das System?

_____

Wie gut konnten die Emotionen vermittelt werden?
☐ überhaupt nicht ☐ einigermaßen ☐ recht gut ☐ eindeutig

Wirkte das Verhalten dadurch natürlicher?
☐ überhaupt nicht ☐ einigermaßen ☐ sehr natürlich

Wirkte der Roboter sympathischer oder vertrauter?
☐ nein, im Gegenteil ☐ nein ☐ ja, ein wenig ☐ ja, sehr

Haben dir die Emotionen bei der Benutzung geholfen?
☐ nein, im Gegenteil ☐ nein ☐ ja, ein wenig ☐ ja, sehr

Hast du dein Verhalten dem Roboter angepasst?
☐ ja ☐ nein ☐ vielleicht ein bisschen

Bewertung des Systems (warum Emotionen als störend/hilfreich empfunden wurden):

_____

_____

_____

_____

_____

Abbildung A.2: Fragebogen zum emotionalen System

Fragebogen zum neutralen Dialog

Hier sind jeweils deine persönlichen Eindrücke einzutragen. Bitte überlege bei deinen Antworten nicht lange und antworte spontan.

Wie kamst du insgesamt mit dem System zurecht?
☐ überhaupt nicht ☐ einigermaßen ☐ recht gut ☐ super

Wie wirkte der Roboter auf dich?
☐ befremdlich ☐ neutral ☐ sympathisch

Wie empfandest du den animierten Kopf des Systems?
☐ störend ☐ lächerlich ☐ nutzlos ☐ hilfreich

Bewertung des Systems (was fandest du störend/sehr gut, ...):

Abbildung A.3: Fragebogen zum neutralen System

Fragebogen nach beiden Dialogen

Du hast nun beide Systeme benutzt. Hier gilt es nun beide miteinander zu vergleichen und zu bewerten. Auch hier bitte wieder ehrlich und spontan antworten!

Welche Erwartungen hattest du VOR dem Experiment an das System?

Welche Erwartungen konnten erfüllt werden? Welche nicht?

Welches System konnte deine Erwartungen besser erfüllen?
☐ emotionales System ☐ neutrales System ☐ keins von beiden
Warum?

Welches System wirkte auf dich sympathischer?
☐ emotionales System ☐ neutrales System ☐ keins von beiden

Welches System war leichter zu bedienen/erlernen?
☐ emotionales System ☐ neutrales System ☐ keins von beiden

Wie würdest du deine Lernerfolge bzgl. des emotionalen Systems einschätzen?
☐ kein Lernerfolg, Emotionen störten eher
☐ nicht besser durch Emotionen, genau gleich wie neutrales System
☐ vielleicht ein wenig besser
☐ wesentlich besser durch emotionales Feedback

Welches System wird sich eher einmal in den menschlichen Alltag integrieren?
☐ emotionales System ☐ neutrales System ☐ keins von beiden

Persönliche Anmerkungen (Kritik, Verbesserungsvorschläge, ...)

Abbildung A.4: Fragebogen zum Vergleich beider Systeme

B. Tabellarischer Gesamtüberblick über Fragebögen

	absolut	in Prozent alle	absolut	in Prozent Benutzerstudie
Anzahl Teilnehmer	23	-	12	-
männl. Teilnehmer	14	69,6%	6	50,0%
allg. Hochschulreife	23	100,0%	12	100,0%
Studienrichtung				
Informatik	7	30,4%	4	33,3%
Informationswirtschaft	3	13,0%	3	25,0%
Maschinenbau	5	21,7%	1	8,3%
sonst. Ingeneur	5	21,7%	1	8,3%
sonst. nicht Ingeneur	3	13,0%	3	25,0%
Englischkenntnisse				
mind. 7 Jahre Schule	23	100,0%	12	100,0%
erweiterte Englischkenntnisse	12	52,2%	7	58,3%

Tabelle B.1: Infos zu Testpersonen

	absolut	in Prozent alle	absolut	in Prozent Benutzerstudie
Erkannte Emotionsmodalitäten				
Mimik	22	95,7%	12	100,0%
Position	14	60,9%	10	83,3%
Stimmlage	4	17,4%	2	16,7%
Wortwahl	1	4,3%	1	8,3%

Tabelle B.2: Erkannte Modalitäten der Emotionen

	absolut	in Prozent alle	absolut	in Prozent Benutzerstudie
Beherrschung des emotionalen Systems				
überhaupt nicht	0	0,0%	0	0,0%
einigermaßen	11	47,8%	5	41,7%
recht gut	11	47,8%	6	50,0%
super	1	4,3%	1	8,3%
Wie wurden die Emotionen vermittelt?				
überhaupt nicht	1	4,3%	0	0,0%
einigermaßen	5	21,7%	1	8,3%
recht gut	11	47,8%	7	58,3%
eindeutig	6	26,1%	4	33,3%
Wirkte der Roboter natürlicher?				
überhaupt nicht	5	21,7%	1	8,3%
einigermaßen	15	65,2%	10	83,3%
sehr natürlich	3	13,0%	1	8,3%
Wirkte der Roboter sympathischer?				
nein, im Gegenteil	2	8,7%	1	8,3%
nein	6	26,1%	4	33,3%
ja, ein wenig	12	52,2%	6	50,0%
ja, sehr	3	13,0%	1	8,3%
Haben Emotionen bei der Benutzung geholfen?				
nein, im Gegenteil	1	4,3%	0	0,0%
nein	14	60,9%	7	58,3%
ja, ein wenig	5	21,7%	4	33,3%
ja, sehr	3	13,0%	1	8,3%
Wurde eigenes Verhalten an Roboter angepasst?				
ja	9	39,1%	6	50,0%
nein	3	13,0%	1	8,3%
vielleicht, ein bisschen	11	47,8%	5	41,7%

Tabelle B.3: Auswertung des Fragebogens zum emotionalen Dialog

	absolut	in Prozent alle	absolut	in Prozent Benutzerstudie
Beherrschung des neutralen Systems				
überhaupt nicht	2	8,7%	0	0,0%
einigermaßen	10	43,5%	5	41,7%
recht gut	11	47,8%	7	58,3%
super	0	0,0%	0	0,0%
Wie wirkte der Roboter?				
befremdlich	5	21,7%	2	16,7%
neutral	15	65,2%	9	75,0%
sympathisch	3	13,0%	1	8,3%
Wie wird animierter Kopf empfunden?				
störend	1	8,3%	1	8,3%
lächerlich	1	8,3%	1	8,3%
nutzlos	13	56,5%	6	50,0%
hilfreich	8	34,8%	4	33,3%

Tabelle B.4: Auswertung des Fragebogens zum neutralen Dialog

	absolut	in Prozent alle	absolut	in Prozent Benutzerstudie
Welches System konnte die Erwartungen besser erfüllen?				
emotionales System	6	26,1%	2	16,7%
neutrales System	7	30,4%	2	16,7%
keins von beiden	10	43,5%	8	66,7%
Welches System wirkte sympathischer?				
emotionales System	12	52,2%	6	50,0%
neutrales System	6	26,1%	3	25,0%
keins von beiden	5	21,7%	3	25,0%
Welches System war leichter zu bedienen/erlernen?				
emotionales System	7	30,4%	4	33,3%
neutrales System	10	43,5%	5	41,7%
keins von beiden	6	26,1%	3	25,0%
Einschätzung der Lernerfolge bzgl. emotionalen Systems				
schlechter	5	21,7%	2	16,7%
gleich	8	34,8%	4	33,3%
ein wenig besser	6	26,1%	4	33,3%
wesentlich besser	4	17,4%	2	16,7%
Welches System wird sich eher integrieren?				
emotionales System	12	52,2%	6	50,0%
neutrales System	10	43,5%	6	50,0%
keins von beiden	1	4,3%	0	0,0%

Tabelle B.5: Auswertung beider Dialoge

C. Tabellarischer Gesamtüberblick über Dialogauswertung

	Teil 1 (emotional)		Teil 2 (neutral)	
	Turns	Punkte	Turns	Punkte
ID, Geschlecht, Alter, Studienrichtung				
12CM, männl., 24, Technische Redaktion				
Durchlauf 1	10	0	3	4
Durchlauf 2	8	0	10	1
Durchlauf 3	7	3	6	2
Gesamt	25	3	19	7
Score	8,0		24,6	
14CK, weibl., 23, Maschinenbau				
Durchlauf 1	2	5	3	4
Durchlauf 2	9	3	10	1
Durchlauf 3	4	5	6	2
Gesamt	15	13	15	10
Score	57,8		44,4	
16FK, weibl., 22, Informatik				
Durchlauf 1	3	4	8	4
Durchlauf 2	7	1	6	1
Durchlauf 3	8	0	7	2
Gesamt	18	5	21	7
Score	18,5		22,2	
18AP, weibl., 23, Informationswirtschaft				
Durchlauf 1	7	3	5	3
Durchlauf 2	9	1	6	2
Durchlauf 3	8	1	10	1
Gesamt	24	5	21	6
Score	13,9		19,1	
20HS, männl., 26, Informatik				
Durchlauf 1	4	6	8	0
Durchlauf 2	6	2	6	2
Durchlauf 3	8	3	5	4
Gesamt	18	8	19	6
Score	29,6		21.1	
22MV, männl., 26, Informatik				
Durchlauf 1	3	4	2	5
Durchlauf 2	8	3	5	2
Durchlauf 3	6	2	5	2
Gesamt	17	9	16	11
Score	35,4		50,0	

Tabelle C.1: Ergebnisse von Gruppe 1

	Teil 1 (neutral)		Teil 2 (emotional)	
	Turns	Punkte	Turns	Punkte
ID, Geschlecht, Alter, Studienrichtung				
13DC, männl., 24, Informatik				
Durchlauf 1	2	5	8	0
Durchlauf 2	4	3	4	5
Durchlauf 3	8	0	9	1
Gesamt	14	8	21	6
Score	38,1		19,1	
15AU, weibl., 21, Informationswirtschaft				
Durchlauf 1	8	3	4	3
Durchlauf 2	8	1	7	1
Durchlauf 3	11	1	10	1
Gesamt	27	5	21	5
Score	12,4		15,9	
17CH, männl., 22, Informationswirtschaft				
Durchlauf 1	2	5	7	3
Durchlauf 2	7	3	9	0
Durchlauf 3	6	4	7	3
Gesamt	15	12	23	6
Score	53,3		17,4	
19ST, männl., 24, Vertriebsingenieur				
Durchlauf 1	2	5	2	5
Durchlauf 2	7	1	7	2
Durchlauf 3	6	2	8	3
Gesamt	15	8	21	6
Score	35,6		43,1	
21TP, weibl., 32, Kartographie				
Durchlauf 1	6	3	3	4
Durchlauf 2	5	4	6	4
Durchlauf 3	9	1	7	3
Gesamt	20	8	16	11
Score	26,7		45,8	
23JS, weibl., 20, Kartographie				
Durchlauf 1	2	5	4	3
Durchlauf 2	7	3	8	2
Durchlauf 3	9	0	9	2
Gesamt	18	8	21	5
Score	29,6		15,9	

Tabelle C.2: Ergebnisse von Gruppe 1

D. Heavy Rain








Index