Simulation der Emotionsdynamik eines künstlichen humanoiden Agenten

Diplomarbeit an der Technischen Fakultät
der Universität Bielefeld

Juni 2003

vorgelegt von
Christian Becker

Betreuer:
Prof. Dr. Ipke Wachsmuth
Dr. Stefan Kopp
Universität Bielefeld
Technische Fakultät
AG Wissensbasierte Systeme
Universitätsstr. 25
33615 Bielefeld

Christian Becker, Ravensberger Str. 78, 33824 Werther,
30. Juni 2003

Danksagung

An dieser Stelle möchte ich allen danken, die mir während des letzten Jahres mit Rat und Tat beiseite standen. Ich danke meinen Betreuern Ipke Wachsmuth und Stefan Kopp für die Zeit, die sie sich genommen haben, sich meine Ideen anzuhören und sie kritisch zu hinterfragen sowie durch eigene Gedanken zu ergänzen. Bei den Mitarbeitern der AG-WBS bedanke ich mich für die konstruktive Kritik, die mir immer wieder wichtige Denkanstöße gab. Einen besonderen Dank möchte ich meiner Mutter aussprechen, die mit großem Engagement und hoher Sorgfalt meine Diplomarbeit Korrektur gelesen hat. Nicht weniger sei auch Stefan Kopp nochmals für die fachlichen Ratschläge gedankt, mit denen er mir bei der Entstehung dieser 105 Seiten half. Ganz besonders herzlich danke ich meiner Freundin Sonja Fuchß für ihre liebevolle Unterstützung nicht nur während des vergangenen Jahres. Meinem Sohn Jonas möchte ich herzlich dafür danken, dass er mich mit seinem fröhlichen Lachen immer wieder aufgemuntert hat und mir gezeigt hat, wie schön Emotionen sein können.

Inhaltsverzeichnis

1 Einleitung und Motivation....7
2 Emotionstheorien in der Psychologie....11
2.1 Eine Arbeitsdefinition des Begriffes „Emotion“....11
2.1.1 Erster Teil der Arbeitsdefinition – die Liste der Beispiele und das Maß der Ähnlichkeit von Emotionen....12
2.1.2 Zweiter Teil der Arbeitsdefinition – Vier Merkmale von Emotionen....19
2.1.3 Vorläufiges Fazit....26
2.2 Die Theorie von Ortony, Clore und Collins (OCC)....26
2.2.1 Die drei Hauptgruppen der Emotionen....28
2.2.2 Das Gesamtmodell....30
2.2.3 Kritik an den Einschätzungstheorien der Emotionen....33
2.3 Zusammenfassung....35
3 Stand der Forschung bei künstlichen humanoiden Agenten....37
3.1 Systeme, die auf der OCC-Theorie aufbauen....37
3.1.1 Der „Affective Reasoner“ von Clark Elliot....38
3.1.2 „Em“ von W. Scott Neal Reilly....40
3.2 Überblick über weitere Systeme....47
3.3 Zusammenfassung....50
4 Konzeption des Emotionssystems....53
4.1 Eingliederung in die Arbeiten der AG....53
4.1.1 Das visuelle Äußere....53
4.1.2 Die Sprachsynthese....54
4.1.3 Die Gestik....55
4.1.4 Die Sensorik....56
4.1.5 Die kognitiv motivierte Gesamtarchitektur....57
4.2 Anforderungen an das Emotionssystem....58
4.3 Aufbau des Emotionssystems....63
4.3.1 Die Dynamiksimulation....65
4.3.2 Die Kategorisierung....71

4.4 Die Einbindung in die Gesamtarchitektur....75
4.4.1 Die Eingabedaten des Emotionssystems....76
4.4.2 Die Ausgabedaten des Emotionssystems....78
4.4.3 Die modulierenden Auswirkungen des Emotionssystems....78
4.5 Zusammenfassung....79
5 Implementation und erste Demonstration....81
5.1 Entwicklungsumgebung....81
5.2 Umsetzung....82
5.2.1 Die Klassenstruktur und die Emotionskategorien....82
5.2.2 Der Dynamik-Raum....84
5.2.3 Der PAD-Raum....86
5.2.4 Kommunikationsstrukturen....87
5.3 Demonstration....88
5.4 Zusammenfassung....92
6 Zusammenfassung und Ausblick....93
6.1 Zusammenfassung....93
6.2 Ausblick....95
Literaturverzeichnis....99
Anhang....103

1 Einleitung und Motivation

Die vorliegende Arbeit hat die Simulation und Expression der Emotionsdynamik eines virtuellen, situierten und menschenähnlichen Kommunikators namens MAX als Ziel. „Menschenähnlich“ bezieht sich dabei auf drei Aspekte:

1. Seine äußere Erscheinung ist menschenähnlich in dem Sinne, dass er in der speziellen Installation der „AG Wissenbasierte Systeme / Künstliche Intelligenz“ in der Größe und der Form eines realen Menschen dreidimensional projiziert wird. Das trägt bereits zur Illusion eines gleichwertigen Gesprächspartners bei.
2. Er besitzt eine menschenähnlich-reichhaltige Fülle von Modalitäten, derer er sich zur Kommunikation bedienen kann. Er kann sich sprachlich, gestisch und mimisch verständigen und außerdem diese Modalitäten auch in sinnvoller Kombination verwenden. Gerade diese Fähigkeit der multimodalen Kommunikation ist eines der Forschungsziele in der Arbeitsgruppe und das „M“ des Namens „MAX“ steht schließlich für „multimodal“.
3. Er besitzt höhere, kognitiv-motivierte Kontrollstrukturen, welche sich vorhandener Wissensbasen bedienen, um Handlungen zu planen, Verhaltensweisen in der Diskurs-Situation zu steuern und allgemein aus aktuellem Wissen über die Situation Schlüsse zu ziehen. Ergänzt werden diese (kognitiven) Kontrollstrukturen durch die Möglichkeit, reaktive Verhaltensweisen zu erzeugen und aktiv die Umgebung mittels virtuellen Sensoren wahrnehmen zu können.

Durch diese drei Eigenschaften des Systems bekommt der menschliche Benutzer bereits das „Gefühl“, einem einigermaßen ebenbürtigen und vertrauenswürdigen Gegenüber zu begegnen, mit dem er auf natürliche Weise sprach-gestisch kommunizieren kann.

Jedoch gibt es immer wieder unvermeidliche Situationen, in denen sich MAX doch anders „verhält“, als vom Benutzer erhofft bzw. erwartet. Aus unserer alltäglichen Erfahrung –

insbesondere beim Umgang mit Computersystemen – wissen wir, dass eine unerfüllte Erwartungshaltung sehr schnell zu emotionalen Reaktionen führen kann. Das ist insbesondere bei wiederholt auftretenden Missverständnissen zu beobachten und führt in Extremfällen zu Gewalthandlungen psychischer und auch physischer Art. Ausrufe wie „Du blöder Computer!“ sind dabei nicht nur auf die allein schon ärgerlich empfundenen Verständigungsschwierigkeiten bei der Benutzung des Werkzeugs Computer bezogen. Im besonderen Maße verärgert sind die meisten Benutzer wohl über die stoische Ruhe, mit der das System sein Unverständnis zum Ausdruck bringt.

Aber nicht nur aus Sicht der Benutzerfreundlichkeit erscheint die Integration eines für Emotionen sensiblen Subsystems in eine kognitive Gesamtarchitektur sinnvoll. Innerhalb dieser Architektur können simulierte Emotionen dazu dienen, den Raum der Handlungsmöglichkeiten einzuschränken und dem System zu einem konsistenteren Langzeitverhalten zu verhelfen.

Für unseren situierten, künstlichen und virtuellen Kommunikator MAX ergeben sich also zwei viel versprechende Erweiterungsmöglichkeiten in Bezug auf noch natürlichere Verständigungs- und Kooperationsmöglichkeiten:

1. Emotionserkennung: Es sollte MAX möglich sein, den emotionalen Zustand seines (menschlichen) Gegenübers wahrzunehmen. Dieser Teil wird gerade in einer anderen Diplomarbeit auf der Basis von Videobilddaten angegangen und würde MAX eine adäquatere Reaktion auf emotional gefärbte Verhaltensweisen des Kooperationspartners ermöglichen.
2. Simulation einer internen Emotionsdynamik und deren Einfluss auf die Psyche und Physis von MAX: Das ist das Thema dieser Arbeit. Diese Simulation soll es MAX ermöglichen, die Reize aus der Umwelt in Bezug auf den (subjektiven) Emotionsgehalt auszuwerten und zur direkten sowie indirekten Verhaltensänderung zu nutzen bzw. der vorhandenen kognitiven Architektur zur Verfügung zu stellen.

Es soll an dieser Stelle nicht unerwähnt bleiben, dass man in der Literatur immer wieder auf die Meinung trifft, es sei für ein mechanisches System grundsätzlich unmöglich, jemals Emotionen zu besitzen.

Zum Beispiel meint Penrose ([PEN91], S. 436 f):

> „Ist es denn nicht 'offensichtlich', dass bloßes Berechnen weder Lust noch Schmerz hervorrufen kann; dass es weder Poesie wahrzunehmen mag noch die Schönheit eines Abendhimmels oder den Zauber von Klängen; dass es nicht hoffen, lieben oder verzweifeln kann; ...“.

Die Zukunft wird zeigen, ob nicht vielleicht doch das Gegenteil „offensichtlich“ wird; ich halte es für möglich und erstrebenswert.

2 Emotionstheorien in der Psychologie

2.1 Eine Arbeitsdefinition des Begriffes „Emotion“

Mit der Frage „Was ist eine Emotion?“ hat sich die Psychologie seit jeher beschäftigt, da sie den menschlichen Geist, die Seele und das Bewußtsein – also das Meta-physische des menschlichen Wesens – als ureigensten Forschungsgegenstand hat. Erwartet man eine exakte Definition des Begriffes „Emotion“, so führen Meyer, Schützwohl und Reisenzein [MRS01] dazu aus, dass eben „diese exakte Definition von Emotionen [...] nicht die Voraussetzung, sondern vielmehr ein Resultat der wissenschaftlichen Analyse“ (S. 23) sei.
Dazu führen sie weiter aus:

> „Und selbst wenn eine gut begründete Emotionsdefinition dieser Art einmal vorhanden ist, hat sie nicht den Status einer Bedeutungsfestlegung der Begriffs „Emotion“, sondern einer empirischen *Hypothese* über die Natur von Emotionen, die deshalb im Prinzip aufgrund nachfolgender Untersuchungen auch wieder revidiert werden kann.“ [MRS01]

Es bietet sich also nicht nur aus Sicht des Ingenieurs eine pragmatische Arbeitsdefinition von Emotionen an, wie sie von Meyer, Schützwohl und Reisenzein [MRS01] im Folgenden eingeführt wird:

Kasten 1
Eine Arbeitsdefinition von Emotionen.

1. Emotionen sind zeitlich datierte, konkrete Vorkommnisse von zum Beispiel Freude, Traurigkeit, Ärger, Angst, Eifersucht, Stolz, Überraschung, Mitleid, Scham, Schuld, Neid, Enttäuschung, Erleichterung sowie weiterer Arten von psychischen Zuständen, die den genannten genügend ähnlich sind.
2. Diese Phänomene haben folgende Merkmale gemeinsam:
   a. Sie sind aktuelle psychische Zustände von Personen.
   b. Sie haben eine bestimmte Qualität, Intensität und Dauer.
   c. Sie sind in der Regel objektgerichtet.
   d. Personen, die sich in einem dieser Zustände befinden, haben normalerweise ein charakteristisches Erleben (Erlebnisaspekt von Emotionen), und häufig treten auch bestimmte physiologische Veränderungen (physiologischer Aspekt von Emotionen) und Verhaltensweisen (Verhaltensaspekt von Emotionen) auf.

Diese Arbeitsdefinition ist absichtlich recht allgemein gehalten, damit sie „für Emotionsforscherinnen und Emotionsforscher mit unterschiedlichen theoretischen Überzeugungen akzeptabel und deshalb möglichst unkontrovers ist.“ ([MRS01], S. 23)

2.1.1 Erster Teil der Arbeitsdefinition – die Liste der Beispiele und das Maß der Ähnlichkeit von Emotionen

Der erste Teil der Definition ist auf die Problematik zurückzuführen, dass bis heute kein Konsens gefunden wurde über die Anzahl der klar voneinander zu unterscheidenden emotionalen Kategorien. Es wurden Untersuchungen mit dem Ziel durchgeführt, herauszufinden, welche Arten von Zuständen *im Alltag* als Emotionen betrachtet werden [SHA87], [GOO93], und man fand, dass im Alltag mindestens 50 bis 100 Arten von Emotionen unterschieden werden, von denen sich einige recht ähnlich zu sein scheinen.

Ekman [EKM99] geht davon aus, dass es möglich und sinnvoll sei, eine gewisse Anzahl von Basis-Emotionen (*Basic Emotions*) voneinander und von anderen affektiven Phänomenen (*affective phenomena*), wie z.B. Stimmungen, zu unterscheiden. Dabei ist besonders zu erwähnen, dass jede dieser *Basic Emotions* keinesfalls als einzelner affektiver Zustand zu verstehen ist, sondern vielmehr eine Familie von ähnlichen Zuständen bezeichnet. Das Maß der Ähnlichkeit wird hierbei durch folgende Charakteristika gemessen:

1. Distinctive universal signals
2. Distinctive physiology
3. Automatic appraisal, tuned to:
4. Distinctive universals in antecedent events
5. Distinctive appearance developmentally
6. Presence in other primates
7. Quick onset
8. Brief duration
9. Unbidden occurrence
10. Distinctive thoughts, memories images
11. Distinctive subjective experience

Tabelle 1: Kriterien zur Unterscheidung der Basisemotionen voneinander und von anderen affektiven Phänomenen [EKM99], Kapitel 3, S. 9.

Was Meyer et al. [MRS01] als „genügend ähnlich“ beschreiben, versucht Ekman [EKM99] mit diesen Kriterien genauer einzugrenzen. Im Allgemeinen geht Ekman im 3. Kapitel noch von 15 voneinander unterscheidbaren Basis-Emotionen aus. In Kapitel 16 in Bezug auf die visuell am Gesichtsausdruck unterscheidbaren Emotionen reduziert er die Anzahl der Basis-Emotionen auf sechs, die durch spezifische Situationen ausgelöst würden und durch einen ebenso spezifischen Gesichtsausdruck identifizierbar seien:

- happiness (Fröhlichkeit)
- anger (Wut/Ärger)
- disgust (Ekel)

- sadness (Traurigkeit)
- fear (Furcht/Angst)
- surprise (Überraschung)

Wundt [WUN22] vertritt hingegen die Auffassung, dass „nichts unrichtiger“ sei, als die „Gefühlswelt“ als „Summe im Wesentlichen unverändert bleibender Elementargefühle“ aufzufassen. Er stellt die These auf, dass „der qualitative Gefühlsreichtum“ aus der gegenseitigen Beeinflussung sowohl gleichzeitiger wie auch aufeinanderfolgender Gefühle resultiere und deshalb „unerschöpflich“ sei ([WUN22], S. 243). Dieser Ansatz zur Repräsentation von Emotionen (respektive Gefühlen, da Wundt diesen Begriff als Synonym verwendet) geht von einer gänzlich anderen psychologischen Methodik aus: dem im 19. Jahrhundert gängigen Verfahren der Introspektion. Da es sich hierbei um das rein subjektive Verfahren der Selbstwahrnehmung handelt, wird der essentielle Unterschied zu den Ekmanschen Basisemotionen leicht erklärbar. Bei Ekman steht der intersubjektive Verhaltensaspekt im Vordergrund, d.h. die Beobachtung äußerlich wahrnehmbarer Veränderungen, während Wundt über die Methode der Introspektion seinen Schwerpunkt auf den subjektiven Erlebnisaspekt der Emotionen legt. Wundt führt das „Totalgefühl“ ein, das die Resultierende mehrerer, auch gegensätzlicher Gefühle bezeichnet und welches als ein „einheitlicher“ Zustand aufzufassen ist, der „wiederum den Charakter eines Gefühls von bestimmter Qualität und Stärke besitzt“ ([WUN22], S. 239, 241). Oft betont Wundt, dass zwar zwischen beliebigen Einzelgefühlen und auch beliebigen Totalgefühlen eine „Verwandtschafts- und Gegensatzbeziehung“ existiere, da sie jeweils eine „eigentümliche Qualität“ gemeinsam hätten, doch seien „sie niemals [...] in diese [Einzelgefühle] zerlegbar“ ([WUN22], S. 243). Er gelangt schließlich zu einer Theorie der „drei Hauptgegensätze der Gefühlsqualtität, die wir als *Lust* und *Unlust*, *Erregung* und *Beruhigung*, *Spannung* und *Lösung* unterscheiden wollen“ (S. 244). Jede dieser drei bipolaren Achsen gibt dabei eine „Gefühlsrichtung“ für eine „qualitative Gefühlsfärbung“ an, wobei er „Lust und Unlust“ als Gegensätze bezeichnet, die „wir im engeren Sinne Qualitätsrichtungen nennen können, während Erregung und Beruhigung die Bedeutung von Intensitätsrichtungen besitzen“ (S. 244). Obwohl Spannung und Lösung

diesem Zuordnungsschema folgend als Zeitrichtungen benannt werden, stellt Wundt ausdrücklich klar:

> „An sich selbst betrachtet sind Erregung und Beruhigung, Spannung und Lösung gerade so gut Richtungen von qualitativ spezifischer Art wie Lust und Unlust, und **in jeder dieser Richtungen sind** wieder die mannigfachsten **Unterschiede der Intensität und der zeitlichen Dauer möglich**“ ([WUN22], S. 244).

Diese Zeitrichtung ist also auf die Zeitkomponente der die Emotion auslösenden Ereignisse (Wundt fasst diese Ereignisse als „das Substrat der Gefühle bildenden Empfindungen und Vorstellungen“) bezogen. Spannung ist gegeben, wenn ein Ereignis erwartet und durch das dabei entstehende Gefühl evtl. auch erhofft wird, Lösung dementsprechend, wenn eine Erwartung in Erfüllung gegangen ist.

Wundt definiert den Nullpunkt seines Systems als „allgemeine Indifferenzlage des Gefühls“ und führt dann weiter aus:

> „Indem **ein einzelner Punkt** [im durch die drei Achsen aufgespannten Raum] nur **ein momentanes Gefühl** bezeichnet, wird aber **irgendein konkretes Geschehen** immer in einem **bestimmten, stetig zusammenhängenden Gefühlsverlauf** bestehen und im allgemeinen durch **eine Kurve** dargestellt werden können, **die für jeden Augenblick die Gefühlslage angibt**“ ([WUN22], S. 245).

Für Wundt gehört die in sich geschlossene Kurve, die sowohl im Nullpunkt beginnt wie endet und einen „vollständigen, in sich zurückkehrenden Gefühlsvorgang“ darstellt, zu dem häufigsten Typ der Verlaufsformen. Als Illustration präsentiert er folgende Zeichnung:

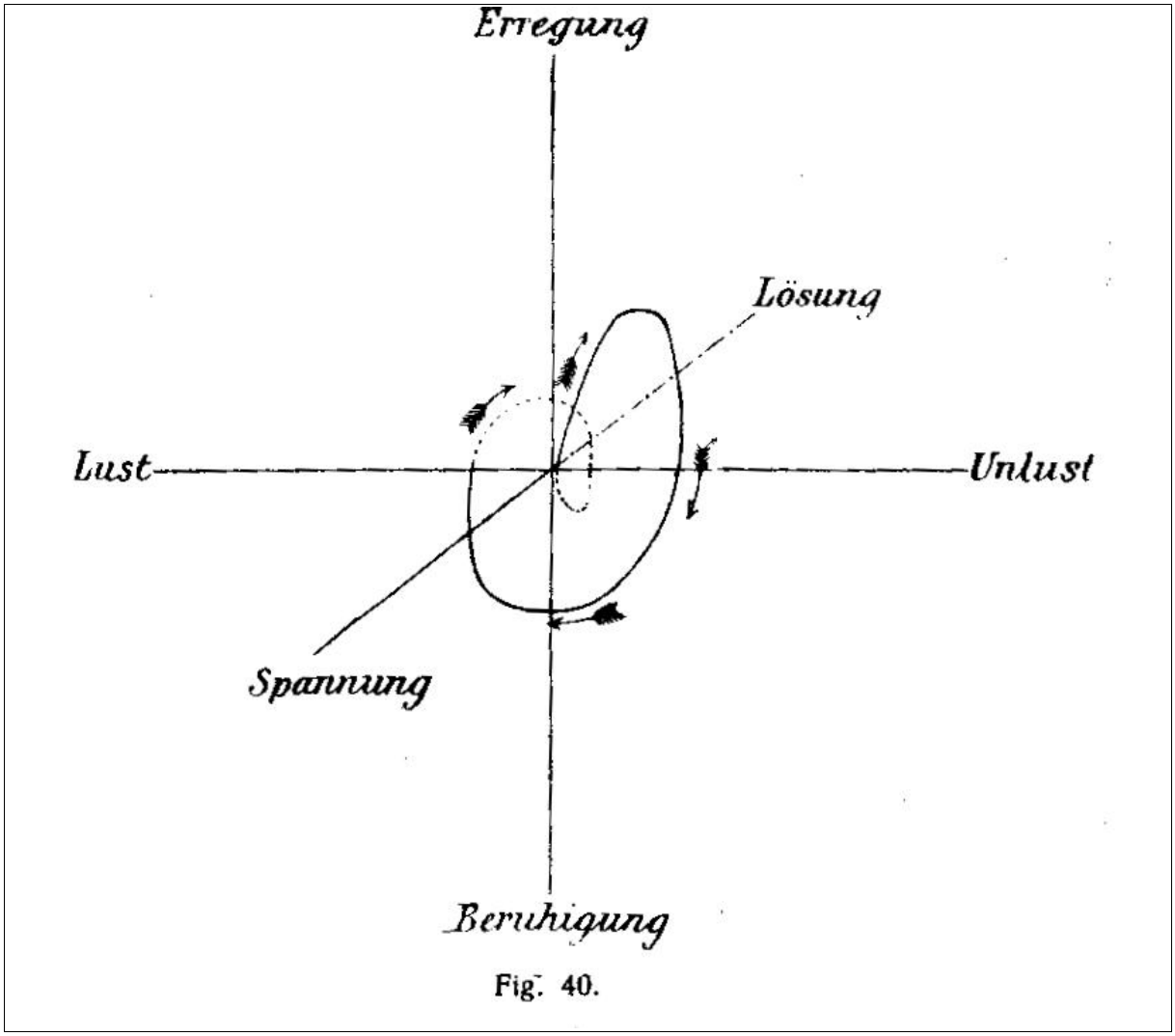


Abbildung 1:
Die drei Hauptrichtungspaare zur Beschreibung von Gefühlsverläufen, [WUN22], S. 246.

Auch möglich sind jedoch „Gefühlsverläufe“, die einen anderen, noch nicht abgeschlossenen Vorgang fortführen.

Mehrabian [MEH95b] legt ein Jahrhundert später ein ganz ähnliches dreidimensionales System zur Kategorisierung von Emotionen vor. Seine drei Achsen nennt er:

- Pleasure und Displeasure (analog zu Wundts Lust und Unlust, kurz P)
- Arousal und Nonarousal (analog zu Wundts Erregung und Beruhigung, kurz A)
- Dominance und Submissiveness (nur annähernd ähnlich zu Wundts Spannung und Lösung, kurz D)

Mehrabian fundiert seine Theorie mit Studien von Osgood, Suci und Tannenbaum, die 1957 bereits nach *basic dimensions of meaning* ([MEH95b], S. 342, Basisdimensionen der Bedeutung) forschten und dabei eine Theorie des semantischen Differenzials (*semantic differential*) aufstellten, die aus den drei Faktoren *Evaluation*, *Potency* und *Activity* bestand. Die Theorie des semantischen Differenzials wurde 1973 bereits im Rahmen einer Untersuchung von Bush dazu verwendet, Versuchspersonen 264 emotionale Begriffe beurteilen zu lassen. Es stellt sich aber die Frage, warum gerade drei Dimensionen am geeignetsten sind, um Emotionen gut kategorisieren zu können. Mit dem Verfahren der Hauptkomponentenanalyse, die Mehrabian auf eine 414 × 34 Felder große Datenmatrix (414 Antwortbögen und 34 diskrete PAD-items (16 Pleasure-Displeasure, 9 Arousal-Nonarousal und 9 Dominance-Submissiveness items)) anwendete, fand er beim vierten Eigenwert einen „clear elbow“ (die ersten fünf Eigenwerte: 10.99, 5.11, 3.74, 1.44 und 1.20; [MEH95b], S. 354). Die ersten drei Eigenwerte decken bereits 58% der gesamten Varianz ab und unter Berufung auf weitere statistische Merkmale (Korrelation u. a. siehe [MEH95b]) sieht Mehrabian damit seine Drei-Faktoren-Lösung als begründet an.

Bei der historischen Einordnung seiner Theorie der *PAD Emotion Scales* bezieht sich Mehrabian nicht auf Wundts Vorarbeiten, obwohl eine grundsätzliche Ähnlichkeit auf der Hand liegt. Die kleinen Unterschiede sind jedoch für diese Arbeit entscheidend.

Erstens benutzt Wundt [WUN22] sein dreidimensionales System zur Beschreibung von Gefühlsverläufen. Auch wenn er an einer Stelle davon spricht, dass „ein einzelner Punkt nur ein momentanes Gefühl bezeichnet“ ([WUN22], S. 245), so hat das Wort „nur“ an dieser Stelle eine entscheidende Bedeutung. Mehrabian hingegen entwickelte seine Theorie auf der Basis der Einordnung emotionaler Begriffe. Er gibt seinem Vorgehen entsprechend folgende Liste von beispielhaften emotionalen Begriffen an, die anhand ihrer Pleasure-, Arousal- und Dominance Werte gruppiert werden:

Pleased, Aroused, Dominant: Admired, bold, carefree, excited, might, triumphant
Pleased, Aroused, Submissive: Amazed, fascinated, grateful, impressed, loved, respectful
Pleased, Unaroused, Dominant: At ease, comfortable, relaxed, satisfied, secure, unperturbed
Pleased, Unaroused, Submissive: Consoled, docile, protected, reverent, sleepy, tranquilized
Displeased, Aroused, Dominant: Angry, catty, defiant, hostile, insolent, nasty
Displeased, Aroused, Submissive: Aghast, bewildered, distressed, in pain, insecure, upset
Displeased, Unaroused, Dominant: Disdainful, indifferent, selfish-uninterested, uncaring, unconcerned
Displeased, Unaroused, Submissive: Bored, despairing, fatigued, lonely, sad, subdued

Tabelle 2: Sample Emotion Terms Grouped According to Pleasure, Arousal, and Dominance Values ([MEH95b], Appendix A).

Auffällig an dieser Aufstellung ist das Fehlen der vier Ekmanschen Basisemotionen happiness, disgust, fear und surprise. Lediglich die beiden negativen Emotionen angry (Displeased, Aroused, Dominant) und sad (Displeased, Unaroused, Submissive) sind vertreten und unterscheiden sich durch die Wertigkeit zweier Achsen voneinander. Außerdem ist es fraglich, ob alle hier aufgeführten Begriffe den in der Arbeitsdefinition genannten „genügend ähnlich“ sind. Eine Abgrenzung „emotionaler“ Begriffe, die für die Emotionspsychologie in Betracht kommen, von Begriffen, die sich auf andere psychische Phänomene beziehen, bleibt Mehrabian schuldig. Er stellt mit seiner Theorie ein Ähnlichkeitsmaß für die diffuse Menge beliebig vieler im Alltag verwendeter „emotionaler“ Begriffe auf.

Zweitens unterscheidet sich Mehrabians Theorie in der Wahl der dritten Achse von Wundts Achse der Spannung und Lösung. Dominanz (zu engl. Dominance) und Unterwürfigkeit (zu engl. Submissiveness) werden von Mehrabian [MEH95b] verstanden als Gefühl der Kontrolle und des Einflusses über andere Menschen oder über Situationen im Fall der Dominanz sowie

im Fall der Unterwürfigkeit als das Gefühl, von den äußeren Umständen kontrolliert und beeinflusst zu werden. Wundt beschreibt seine dritte Komponente der Spannung und Lösung als zeitlichen Aspekt der die Emotion auslösenden Ereignisse. Bei Mehrabian handelt es sich also um eine Achse, die einen Aspekt des Gefühls selbst klassifiziert, wohingegen bei Wundt durch die dritte Achse ein Aspekt des äußeren Reizes klassifiziert wird.

2.1.2 Zweiter Teil der Arbeitsdefinition – Vier Merkmale von Emotionen

1. Merkmal: Emotionen sind aktuelle psychische Zustände von Personen.

Meyer et al. [MRS01] führen dieses Merkmal von Emotionen ein, um:

- lediglich den Menschen als Träger einer Emotion zu betrachten, also ausdrücklich keine Tiere mit in die Analyse einzubeziehen und
- die Emotionen in Form von „aktuellen emotionalen Episoden“ zu verstehen und sie damit „von *emotionalen Dispositionen* abzugrenzen“ (S. 26).

Emotionale Dispositionen sind zum Beispiel Liebe oder generelle Reizbarkeit. Mehrabian entwickelte seine Pleasure-Arousal-Dominance-Theorie ursprünglich zur Kategorisierung von *emotional traits* (emotionale Persönlichkeitseigenschaften, vgl. [MEH95b], S. 355, auch [MRS01], S. 29), die er als zeitlich stabilere bzw. unveränderliche Wesenszüge eines Menschen versteht. Mit der Charaktereigenschaft der generellen Reizbarkeit (*trait arousability*, [MEH95a]) hat sich Mehrabian ebenfalls intensiv experimentell beschäftigt. Vielleicht finden sich deshalb Begriffe wie *loved* (geliebt) und *defiant* (trotzig) in Tabelle 2, die anhand des hier beschriebenen ersten Merkmals nicht als Emotionen, sondern als emotionale Dispositionen gelten. Es wird erneut deutlich, dass bis heute kein Konsens bezüglich des Betrachtungsgegenstandes der Emotionspsychologie gefunden wurde, weshalb eine klare Differenzierung zwischen Emotionen und emotionalen Dispositionen unmöglich ist. Es muss im Einzelfall entschieden werden, ob ein Begriff, der für ein psychisches Phänomen verwendet wird, unter die „globalen und gewöhnlich auch stabilen emotionalen Dispositionen“ [MRS01] fällt oder gerade keinen Teil der Persönlichkeitseigenschaften darstellt, sondern als emotionaler Begriff verstanden werden kann.

2. Merkmal: Emotionen haben eine bestimmte Qualität, Intensität und Dauer.

Diese drei Eigenschaften von Emotionen werden „sowohl im Alltag als auch in der wissenschaftlichen Psychologie benutzt, um konkrete Emotionsvorkommnisse zu gruppieren“ ([MRS01], S. 29). Mit Begriffen wie „Ärger“, „Angst“ und „Freude“ wird sich eigentlich nur auf die *Qualität* einer Emotion bezogen, auch wenn sie im üblichen Sprachgebrauch bereits als Begriffe für Instanzen der Klasse der Emotionen verwendet werden.

Im Hinblick auf den Begriff der Emotionsdynamik unterscheidet Kruse [KRU85] Emotionen und Stimmungen voneinander. Auch Stimmungen haben eine Qualität, Intensität und Dauer, doch lassen sie sich „als langfristig wirkende emotionale 'Hintergrundsignale' im Bewußtsein beschreiben, die *unspezifisch* sind, d.h. ihren Bezug zu einem festen Objekt verloren haben“ ([KRU85], S. 36/37). Nach Kruse [KRU85] könne sich zwar fast jede Emotion zu einer Stimmung entwickeln, doch genüge es aus klinischer Sicht, die Variation der *Grundstimmung* auf einer negativ-positiv-Achse zu betrachten. Die Beinflussung der Emotionen durch die gleichzeitig vorherrschende Grundstimmung fasst Kruse so zusammen:

> „Eine negative Grundstimmung tendiert dazu, positive Emotionen abzuschwächen oder zu unterdrücken. In einer tiefen, depressiven Verstimmung ist ein Klient kaum in der Lage, Freude, Liebe oder Selbstvertrauen zu empfinden. Umgekehrt immunisiert eine Hochstimmung gegen alle negativen Affekte [hier als Synonym für Emotionen verwendet]. Menschen sind damit kaum in der Lage, Angst oder eine Kränkung zu empfinden. Sie sind manchmal praktisch unfrustrierbar.“

Stimmungen beeinflussen also die Intensität der Emotionen und unterscheiden sich maßgeblich durch ihre längere Dauer von ihnen. Es sei allerdings, so Kruse [KRU85], unklar, wodurch Stimmungen hervorgerufen und aufrecht erhalten würden, obwohl der Effekt des „Aufschaukelns von Emotionen“ ([KRU85], S. 39) zu beobachten sei. Durch die Analyse von Krisensituationen beim Menschen, die zu Selbstmordhandlungen führen, fände man oft eine gegenseitige Verstärkung negativer Emotionen, aber auch positive Emotionen könnten in einer Stimmung der „Verliebtheit“ gipfeln.

3. Merkmal: Emotionen sind in der Regel objektgerichtet.

Dieses Merkmal betrifft die Annahme, dass eine aktuell empfundene Emotion oft durch rationales Schlussfolgern begründbar erscheint, auch wenn diese subjektive Begründung falsch sein kann und die Objekte, auf die sich die Begründung stützt, nicht unbedingt real existieren müssen. Die Ereignisse können auch in der Zukunft liegen, was Meyer et al. ([MRS01], S. 30) auf folgende Weise illustrieren:

> „Zum Beispiel kann eine am Bahnsteig wartende Person sich darüber ärgern, dass (wie sie meint) ihr Zug Verspätung hat, während in Wirklichkeit nur die Uhr falsch geht.“

Emotionstheorien können die Emotionen entweder im Subjekt (als dessen Reaktion) oder im Objekt (als dessen *Valenz* oder *Aufforderungscharakter*) lokalisieren ([KRU85], S. 40). Genau in diesem Punkt unterscheiden sich (bei aller sonstigen Ähnlichkeit) die Emotionstheorien von Wundt [WUN22] und Mehrabian [MEH95b]. Wie in Kapitel 2.1.1 bereits ausführlich besprochen, liegt der Unterschied in der Bedeutung der dritten Achse der beiden dreidimensionalen Systeme. Bei Wundt ist die Dimension der „Spannung und Lösung“ objektbezogen, da hiermit die zeitliche Eigenschaft des Ereignisses klassifiziert wird, während die Dimension der „Dominanz und Unterwürfigkeit“ bei Mehrabian eher subjektbezogen verstanden wird. Die eigene emotionale Bewertung, die als Eigenschaft dem Objekt zugeschrieben wird, bezeichnet Kruse [KRU85] als „projektive Bewertung“ und diese Projektion wird durch Begriffe wie „eklig“ erleichtert, die eigentlich eine Eigenschaftszuschreibung an das Objekt darstellen. Somit können Emotionen handlungswirksam werden, ohne bewusst zu werden, indem über das Wahrnehmungssystem die Emotionen auf Objekte projiziert und nicht als eigene Emotionen erkannt werden.

4. Merkmal: Erlebnisaspekt, physiologischer Aspekt und Verhaltensaspekt von Emotionen

Der ***Erlebnisaspekt*** wird in der Literatur häufig als das „Gefühl“ bezeichnet und zum Beispiel von Dörner [DÖR99] vom Begriff „Emotion“ abgegrenzt, indem er feststellt, dass sein rechnergestütztes Simulationssystem Ψ eben nur Emotionen simuliere und *keine Gefühle*. Beim Gefühl handelt sich um die „subjekive Komponente“ von Emotionen und „damit wird

zum Ausdruck gebracht, dass es sich 'in einer ganz bestimmten Weise anfühlt', Emotionen zu haben [...] – eine Weise, die sich von dem Erleben anderer bewusster Zustände [...] auf charakteristische Weise zu unterscheiden scheint“ ([MRS01], S.33).

Im Rahmen der technischen Realisierung eines Emotionssystems muss die Frage geklärt werden, inwieweit dieses „Erleben“ von Emotionen für ein Computersystem möglich ist bzw. ob es für eine kognitiv motivierte Gesamtarchitektur wirklich nötig ist. Manche Psychologen vertreten die Auffassung, „das Spezifische am Erleben von Gefühlen bestehe in der Empfindung der körperlichen (insbesondere physiologischen) Veränderungen, die bei Emotionen auftreten. Wieder andere sehen das Spezifische des emotionalen Erlebens im Gewahrsein von drängenden Handlungsimpulsen, zum Beispiel dem Drang anzugreifen bei Wut oder zu fliehen bei Furcht“ ([MRS01], S.33). Auch bei Wundt ([WUN22], S. 244 ff.) wird der Erlebnisaspekt von Emotionen einer von drei orthogonal zueinander stehenden, unabhängigen Achsen zugeordnet – der Achse der „Lust – Unlust“. Um den Erlebnisaspekt zu erklären, beziehen sich Psychologen also häufig auf einen der anderen beiden Aspekte des vierten Merkmals, weshalb diese – zumindest unter dem Gesichtspunkt einer technischen Implementierung – den Erlebnisaspekt subsumieren können. Dann wäre „das Gefühl“ eine implizite Eigenschaft eines technischen Systems, das den physiologischen wie auch den Verhaltensaspekt von Emotionen berücksichtigt und es wäre eigentlich funktional überflüssig (vgl. [DÖR99]).

Der ***physiologische Aspekt*** von Emotionen wird in [MRS01] breiter gefasst, indem nicht nur die *peripher-physiologischen* Veränderungen bzw. Reaktionen (wie z.B. Erröten, Veränderungen der Atmung, Schwitzen der Hände, Veränderungen in der Stimmqualität), die durch das autonome Nervensystem (ANS) gesteuert werden, sondern auch die Vorgänge im zentralen Nervensystem (Gehirn) gemeint sind. Ekman [EKM84] stellt fest, dass eine Messung der Hauttemperatur und der Herzrate bei willentlich herbeigeführten Gesichtsmuskelaktivitäten (näheres zur Gesichtsmimik siehe *Ausdrucksaspekt* weiter unten) bereits ausreichten, um nicht nur zwischen positiven und negativen Emotionen, sondern auch zwischen *Anger* und *Fear / Sad* nach folgendem Entscheidungsbaum unterscheiden zu können:

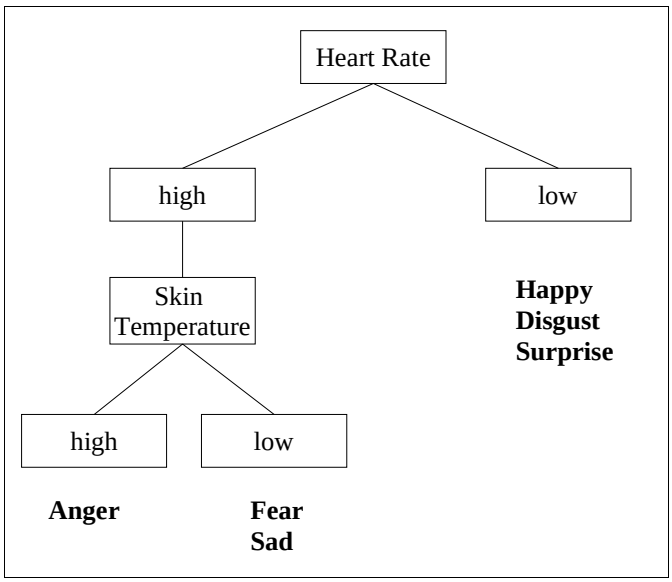


Abbildung 2: „Decision tree for discriminating emotions in Direct Facial Action Task“ (nach [EKM84]).

Neuere Untersuchungen ergaben allerdings, dass auch im Fall von Ekel (*Disgust*) die Hauttemperatur sinkt.

Der ***Verhaltensaspekt*** von Emotionen umfasst nach [MRS01] den expressiven oder Ausdrucksaspekt und den instrumentellen oder Handlungsaspekt.

> „Der *expressive* oder *Ausdrucksaspekt* von Emotionen umfasst verschiedene Arten von meist unwillkürlichem Ausdrucksverhalten: den Gesichtsausdruck (mimischer Ausdruck), Gestik und Körperhaltung (z. B. zusammengesunkene Haltung bei Traurigkeit), die Körperorientierung (z. B. plötzliches Zusammenzucken) sowie bestimmte Merkmale der Sprechstimme“ ([MRS01], S. 35).

Gerade der mimische Ausdruck findet in der gegenwärtigen Emotionspsychologie besondere Beachtung und geht auf Untersuchungen von Darwin [DAR72] zurück. Ekman und Friesen gehen bei ihren Arbeiten von Universalien des menschlichen Emotionsausdrucks aus (vgl. [EKM99]; [EKF71]), die erblich bedingt und somit kulturübergreifend anzutreffen seien. Es wird davon ausgegangen, dass bei allen Menschen die gleiche Verschaltung von Gefühl und Mimik vorhanden ist und aufgrund empirischer Belege für diese Theorie entwickelten Ekman und Friesen schließlich ein deskriptives System zur Gesichtsaktionskodierung FACS

(*Facial Action Coding System*, [EKF78]), welches heute in der Erkennung und Klassifikation von emotionalen Gesichtsausdrücken immer noch Verwendung findet. Es arbeitet dabei auf der Basis der Erkennung von Gesichtsmuskel- bzw. Gesichtsmuskelgruppen-Kontraktionen und lieferte weitere Befunde, die die Autoren als zusätzlichen Nachweis für einen direkten Zusammenhang zwischen Mimik und Emotionen anführen (vgl. [EFA80]).

Vereinzelt konnten auch kulturelle Unterschiede bezüglich Mimik nachgewiesen werden, weshalb die Annahmen um so genannte *display rules* erweitert wurden. Demnach sei es einem Menschen zwar nicht möglich, bei induzierten Emotionen die Aktivierung der efferenten Bahnen zu den korrespondierenden Gesichtsmuskeln zu verhindern, jedoch könne willentlich beeinflusst werden, ob es zu einer Gesichtsbewegung kommt ([EKM84]). Gerade in Bezug auf das Lächeln könnten diese *display rules* immer dann Anwendung finden, wenn ein „aufgesetztes / vorgetäuschtes Lächeln“ zu Tage tritt. Im Kontrast zum echten, so genannten *Duchenne-Smile* (benannt nach dem Neurologen Duchenne de Boulogne, der 1862 die anatomisch-psychologischen Grundlagen der Ausdruckserscheinungen untersuchte), konnte Ekman [EKF82] bei Untersuchungen auf Basis des FACS Unterscheidungsmerkmale feststellen und damit seine These der Differenzierbarkeit von willkürlichen und unwillkürlichen Gesichtsausdrücken weiter festigen. Bei einem vorgetäuschten Lächeln läge nach Ekman [EKF82] ein kürzerer *onset* und ein längerer Apex (also eine längere Verweildauer in extremer Position) vor, sowie eine Asymmetrie beider Gesichtshälften und keine Aktivität des Muskels orbicularis oculi, wodurch die sonst vorhandenen Augenfältchen hier fehlten.

Zum Ausdrucksaspekt gehören ebenso bestimmte Merkmale der Sprechstimme und John Stallo [STA00] führt folgende Zusammenfassung der Effekte von vier Ekmanschen Basisemotionen auf die menschliche Stimme an:

	Anger	***Happiness***	***Sadness***	***Fear***
Speech rate	Faster	Slightly faster	Slightly slower	Much faster
Pitch average	Very much higher	Much higher	Slightly lower	Very much higher
Pitch range	Much wider	Much wider	Slighly narrower	Much wider
Intensity	Higher	Higher	Lower	Higher
Pitch changes	Abrupt, downward, directed contours	Smooth, upward inflections	Downward inflections	Downward terminal inflections
Voice quality	Breathy, chesty tone	Breathy, blaring	Resonant	Irregular voicing
Articulation	Clipped	Slightly slurred	Slurred	Precise

Tabelle 3: Emotionseffekte auf die Stimme (nach [STA00], S. 3).

Die Sprechgeschwindigkeit variiert zum Beispiel unter dem Einfluss von Emotionen in der Weise, dass Traurigkeit sich in einer etwas langsameren Sprechweise niederschlägt, wohingegen Wut zu einer schnelleren Sprechweise führt. Diese Merkmale der Sprechstimme lassen sich auch mit Mehrabians PAD-Modell in Einklang bringen, bei dem *angry* mit großer Erregtheit einhergeht und *sad* als emotionaler Zustand kategorisiert wird, welcher mit großer innerlicher Ruhe zusammenhängt.

Der *instrumentelle* oder *Handlungsaspekt* von Emotionen umfasst nach [MRS01] „instrumentelle (zielgerichtete) Handlungen, zum Beispiel Angriffsverhalten bei Wut oder Fluchtverhalten bei Furcht“. Im Hinblick auf das Anwendungsszenario des in dieser Arbeit entwickelten Emotionssystems wäre sowohl Flucht- als auch Angriffsverhalten zwar nicht sinnvoll – das Emotionssystem soll eine Komponente eines virtuellen Konstruktionsgehilfen sein –, doch für eine möglichst natürlich wirkende Kommunikationssituation wird jede intentionale Handlung des Agenten MAX durch seinen aktuellen emotionalen Zustand „gefärbt“ sein.

2.1.3 Vorläufiges Fazit

Bisher stand die Frage im Vordergrund, was eine Emotion überhaupt ist und es wurde verdeutlicht, dass diese Frage selbst einen bislang ungeklärten Forschungsgegenstand in der Psychologie darstellt. Die Theorien von Wundt [WUN22], Mehrabian [MEH95b] und Ekman [EKM99] wurden in groben Zügen in eine Arbeitsdefinition von Meyer et al. eingeflochten, wodurch der Schwerpunkt auf die Repräsentationsmöglichkeiten von Emotionen und deren Abgrenzung zu anderen psychischen Phänomenen gelegt wurde.

Wenig Beachtung fand die Frage, wodurch Emotionen ausgelöst werden. Es wurde bisher lediglich festgestellt, dass Emotionen „in der Regel objektgerichtet“ sind (siehe Kasten 1, Punkt 2c der Arbeitsdefinition). Obwohl der Fokus dieser Arbeit auf der Emotions**dynamik** liegt, muss auch auf diese zweite wichtige Frage eingegangen werden, weshalb sie im folgenden Kapitel behandelt wird.

2.2 Die Theorie von Ortony, Clore und Collins (OCC)

Die Psychologen Ortony, Clore und Collins [OCC88] verfolgen zwei grundlegende Ziele mit ihrer Arbeit über „Die kognitive Struktur der Emotionen“. Erstens versuchen sie, „wenigstens den Anschein von Ordnung in ein sehr konfuses und verwirrendes Forschungsfeld zu bringen“. Zweitens – und das ist für die vorliegende Arbeit wichtiger – wollen sie die Grundlage für ein Emotionsmodell liefern, welches sich rechnergestützt simulieren läßt. Sie konzentrieren sich dabei auf die Frage, welche Rolle die Kognition bei der Entstehung von Emotionen spielt. Sie erwähnen zwar, das noch andere wichtige Aspekte der Emotionen wie die physiologische und expressive Komponente und der Verhaltensaspekt existieren, stellen aber im gleichen Satz klar, dass sie diesen Aspekten nicht viel Beachtung schenken. Eine ihrer Grundannahmen bezüglich der Natur von Emotionen ist gleichzeitig so reizvoll für Informatiker wie umstritten unter Emotionspsychologen ([OCC88], S. 1):

„Taking the perspective of empirical psychology and cognitive science, we start with

the assumption that emotions arise as a result of the way in which the situations that initiate them are construed by the experiencer.“

Demnach sind für Ortony et al. für die Entstehung von Emotionen drei Dinge notwendig:

1. Situationen, die die Emotionen auslösen.
2. Personen, die die Emotionen empfinden.
3. Die Einschätzung der Situation durch die Person.

Der Vorgang der Einschätzung einer Situation ist dabei nur auf kognitivem Weg denkbar, weshalb dieser Ansatz es ausschließt, dass auf nicht-kognitivem Weg Emotionen entstehen. Da in dieser Theorie der Aspekt der Einschätzung (engl. *appraisal*) von Situationen im Vordergrund steht, fällt sie in die Klasse der *appraisal theories of emotions* (der Einschätzungstheorien der Emotionen), die aufgrund ihrer internen Strukturierbarkeit in Regeln und Schlussfolgerungen bei KI-Wissenschaftlern naturgemäß viel Anklang findet. Im Gegensatz zu anderen Vertretern dieser Klasse (vgl. Scherer [SCH84], Frijda [FRI86]) liegt bei der OCC-Theorie aber ein besonderer Schwerpunkt auf der Implementierbarkeit. Deshalb besitzt sie als einzige nach [MRS03] die folgenden drei positiven Eigenschaften:

- Sie ist umfassend, d.h. sie gilt als Erklärungsmodell für **alle** Emotionen und nicht nur für einen ausgewählten Teil.
- Sie ist detailliert.
- Sie ist systematisch aufgebaut.

Ortony et al. definieren Emotionen folgerichtig als *valenced reactions* (velanzierte Reaktionen), die sich auf Ereignisse, Handlungen oder Objekte beziehen.

2.2.1 Die drei Hauptgruppen der Emotionen

Ortony et al. unterscheiden drei Hauptgruppen der Emotionen anhand des Gegenstandes, auf den sie sich beziehen:

1. ereignisfundierte Emotionen
2. handlungsfundierte Emotionen
3. objektfundierte Emotionen

Tabelle 4: Die drei Hauptgruppen der Emotionen (nach [OCC88]).

Unterschieden werden diese Gruppen durch **zwei Merkmale** voneinander:

1.) **Unterschiedliche Arten von Gegenständen**, auf die sie (bzw. die ihnen zu Grunde liegenden Kognitionen) sich beziehen.

2.) **Unterschiedliche Formen der Bewertung**.

Bei einem Ereignis spielt die *Erwünschtheit* auf einer Skala zwischen erwünscht und unerwünscht eine Rolle, eine Handlung wird hinsichtlich ihrer *Lobwürdigkeit* zwischen lobenswert und tadelnswert eingestuft und bei einem Objekt wird dessen Attraktivität als anziehend oder abstoßend beurteilt.

Um die Eigenschaften dieser Hauptgruppen genauer zu beschreiben, wird an dieser Stelle erneut [MRS03] zitiert, indem die dort gegebenen Zusammenfassungen präsentiert werden:

Tabelle 5:

Die ereignisfundierten Emotionen (nach Ortony et al., 1988)

Allgemeine Beschreibung: Gefühle der Zufriedenheit über ein erwünschtes Ereignis und der Unzufriedenheit über ein unerwünschtes Ereignis

Gegenstand: Ereignis = Sachverhalt ohne verantwortlichen Urheber

Bewertungskriterium: ein persönlicher, d.h. nicht-normativer Wunsch

Resultierende Bewertung: Erwünschtheit (erwünscht-unerwünscht)

Untergruppen:

- Wohlergehensemotionen: Freude, Leid
- Empathieemotionen: Mitfreude, Mitleid, Schadenfreude, Neid/Missgunst
- Erwartungsfundierte Emotionen:
  - Ungewissheitsemotionen: Hoffnung, Furcht
  - Emotionen der Erwartungsbestätigung (Befriedigung, bestätigte Furcht) und der Erwartungsentkräftung (Enttäuschung, Erleichterung)

Die **ereignisfundierten Emotionen** bilden die größte Gruppe der Emotionen, da 12 der insgesamt 22 Emotionskategorien auf sie entfallen. Die zweitgrößte Gruppe mit 4 Emotionskategorien bilden die **handlungsfundierten Emotionen**, die in [MRS03] auf Seite 138 wie folgt zusammengefasst werden:

Tabelle 6:

Die handlungsfundierten Emotionen (nach Ortony et al., 1988)

Allgemeine Beschreibung: Gefühle der Billigung einer lobenswerten Handlung und der Missbilligung einer tadelnswerten Handlung

Gegenstand: Handlung = das Herbeiführen oder Verhindern eines Ereignisses durch einen verantwortlichen Urheber

Bewertungskriterium: eine internalisierte soziale oder moralische Norm

Resultierende Bewertung: Lobwürdigkeit (lobenswert-tadelnswert)

Untergruppen:

- Emotionen des Selbstlobs (z.B. Stolz) und des Selbstvorwurfs (z.B. Schuld)
- Emotionen des Lobs (z.B. Bewunderung) und des Vorwurfs (z.B. Empörung)

Mit lediglich 2 Emotionskategorien sind die **objektfundierten Emotionen** die kleinste Gruppe. Sie werden in [MRS03] zusammengefasst durch:

Tabelle 7:
Die objektfundierten Emotionen (nach Ortony et al., 1988)

Allgemeine Beschreibung: Gefühle des Mögens eines anziehenden Objekts und des Nichtmögens eines abstoßenden Objekts

Gegenstand: Objekt = ein Einzelding wie eine Person, ein Tier oder ein nichtbelebtes Ding

Bewertungskriterium: Einstellungen (Meinungen über die positiven und negativen Eigenschaften des Objekts)

Resultierende Bewertung: Attraktivität (anziehend-abstoßend)

Untergruppen:

- positive objektfundierte Emotionen (z.B. Zuneigung, Liebe)
- negative objektfundierte Emotionen (z.B. Abneigung, Ekel)

Als Spezialfall mit vier Emotionen sind die **Verbundemotionen** zu nennen, die in der OCC-Theorie eine Sonderstellung einnehmen. Die Verbundemotionen können immer dann entstehen, wenn Wohlergehensemotionen und handlungsfundierte Emotionen zusammen auftreten. Ausgehend von der Vermutung, dass es einem Menschen möglich ist, sich sowohl auf die Handlung, als auch auf das daraus resultierende Ereignis zu konzentrieren, bilden sich dann Verbundemotionen wie z.B. Selbstzufriedenheit, Reue, Dankbarkeit oder Ärger.

2.2.2 Das Gesamtmodell

Der genaue kognitive Vorgang wird von Ortony et al. [OCC88] durch Abbildung 3 illustriert, die von Meyer et al. [MRS03] ins Deutsche übersetzt wurde. Für die vorliegende Arbeit ist sie mit arabischen Zahlen, hebräischen Buchstaben, griechischen Buchstaben und römischen Zahlen indiziert worden, um sie besser in Form eines Entscheidungsbaumes interpretieren zu können.

Davon ausgehend, das eine Emotion vorliegt, müssen folgende Fragen beantwortet werden, um diese in eine der 22 Emotionsklassen klassifizieren zu können (siehe Seite 32):

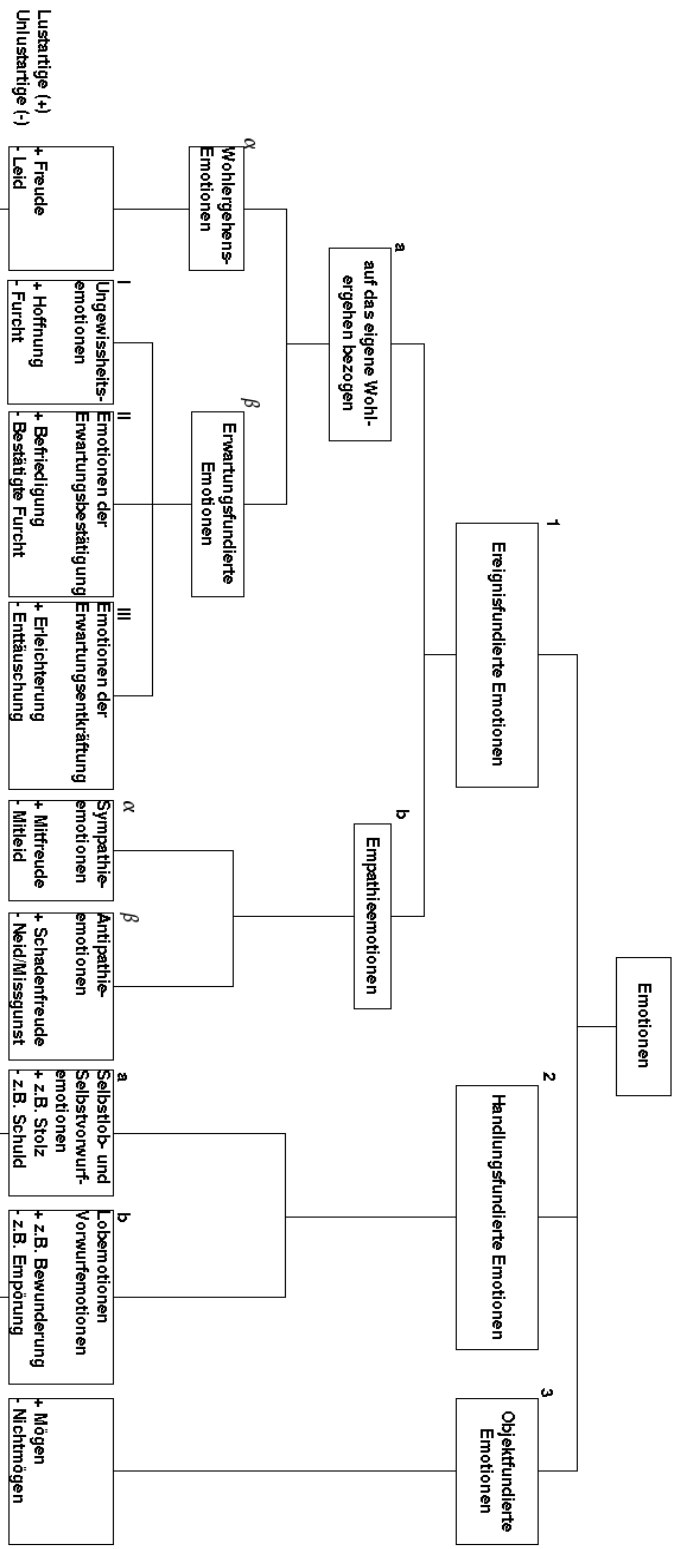


Abbildung 3:
Die Klassifikation der Emotionen nach ihren kognitiven Grundlagen (nach Ortony et al. [OCC88])

Tabelle 8: Entscheidungsfragen zur OCC-Theorie.
Was ist die Ursache bzw. die Grundlage der Emotion? 1. Ereignis, 2. Handlung oder 3. Objekt
1) (Ereignis) Wen betrifft die Emotion? a. Mich selbst oder b. eine andere Person
1a) (Ereignis, Mich selbst) Hatte ich irgendeine Erwartungshaltung gegenüber diesem Ereignis? α. keine bzw. eine bewusst unterdrückte oder β. Erwartungshaltung vorhanden
1aα) (Ereignis, Mich selbst, keine Erwartungshaltung) ⇒ Wohlergehensemotionen: Freude (+) oder Leid (-)
1aβ) (Ereignis, Mich selbst, Erwartungshaltung) Wie sicher war ich mir und ist das Ereignis schließlich entsprechend meiner Erwartungen eingetreten? I. unsicher, II. Erwartung bestätigt oder III. Erwartung verletzt
1aβI) (Ereignis, Mich selbst, Erwartungshaltung, unsicher) ⇒ Ungewissheitsemotionen: Hoffnung (+) oder Furcht (-)
1aβII) (Ereignis, Mich selbst, Erwartungshaltung, bestätigt) ⇒ Emotionen der Erwartungsbestätigung: Befriedigung (+) oder bestätigte Furcht (-)
1aβIII) (Ereignis, Mich selbst, Erwartungshaltung, verletzt) ⇒ Emotionen der Erwartungsentkräftung: Erleichterung (+) oder Enttäuschung (-)
1b) (Ereignis, andere Person) Welche Einstellung habe ich dieser Person gegenüber? α. (eher) Zuneigung oder β. (eher) Abneigung
1aα) (Ereignis, andere Person, Zuneigung) ⇒ Sympathieemotionen: Mitfreude (+) oder Mitleid (-)
1aβ) (Ereignis, andere Person, Abneigung) ⇒ Antipathieemotionen: Schadenfreude (+) oder Neid/Missgunst (-)
2) (Handlung) Wen betrifft die Emotion? a. Mich selbst oder b. eine andere Person
2a) (Handlung, Mich selbst) ⇒ Selbstlob- und Selbstvorwurfemotionen: z.B. Stolz (+) oder Schuld (-)
2a) (Handlung, andere Person) ⇒ Lob- und Vorwurfemotionen: Bewunderung (+) oder Empörung (-)
3) (Objekt) ⇒ Mögen (+) oder Nichtmögen (-)

Ortony et al. [OCC88] weisen ausdrücklich darauf hin, dass es sich bei diesen 22 Emotionskategorien lediglich um sprachliche Repräsentanten für Gruppen ähnlicher Emotionen handelt. Sie verdeutlichen dies anhand des Begriffs *fear* (Angst/Furcht), den sie als gute Bezeichnung einer Emotionskategorie herausstellen, da im Englischen das Wort *terrified* (zu Deutsch etwa: „zu Tode erschrocken") nach Meinung von Ortony et al. (S. 8) lediglich eine sehr starke Form von *fear* bezeichnet und deshalb nur einen qualitativen Unterschied ausmacht. Zudem läge ihr Ziel nicht darin, emotionale Begriffe zu definieren, sondern so sprach-neutral wie möglich die Charakteristika derjenigen Emotionen zu spezifizieren, die voneinander unterscheidbarer (*distinct*) seien.

Des Weiteren stellen Ortony et al. [OCC88] ein Modell zur Intensitätsberechnung auf, welches auf globalen Variablen (wie z.B. *sense of reality* (interessanter Weise ins Deutsche am ehesten als „Realitäts*gefühl*" zu übersetzen) oder *unexpectedness* („Plötzlichkeit")) und – bezogen auf die einzelnen Emotionskategorien – lokalen Variablen (wie z.B. für die erwartungsfundierten Emotionen *likelihood* („Wahrscheinlichkeit")) basiert. Diese Variablen werden zusätzlich noch gewichtet, bevor sie miteinander verrechnet werden, um z.B. zwischen *hope* (hoffen) und *hopeful* (hoffnungsvoll) unterscheiden zu können (vgl. [REI96], S. 83). Sie machen aber keine Angaben über die Dauer respektive den zeitlichen Verlauf der erzeugten Emotionen, was im Hinblick auf den Schwerpunkt der vorliegenden Arbeit sehr bedauerlich ist.

2.2.3 Kritik an den Einschätzungstheorien der Emotionen

Als theoretische Argumente für die Einschätzungstheorien werden die hohe Erklärungskraft und die Übereinstimmung mit zentralen Annahmen der Alltagspsychologie von Meyer et al. [MRS03] angeführt. Zudem haben empirische Untersuchungen die Theorie im Großen und Ganzen bestätigt, wobei darauf hingewiesen wird, dass bei den meisten Untersuchungen allerdings keine „realen Emotionen" ([MRS03], S. 153) im Labor erzeugt wurden. Die Versuchspersonen wurden vielmehr gebeten, sich an eine emotionale Situationen zu erinnern oder sich eine vorgegebene emotionale Situation vorzustellen und diese dann auf postulierten

Einschätzungsdimensionen zu beurteilen. Die Ergebnisse dieser Studien lassen sich kurz wie folgt zusammenfassen:

1. Es konnte bestätigt werden, dass unterschiedliche Gefühle mit unterschiedlichen Mustern von Einschätzungen auf einer begrenzten Anzahl von Dimensionen einhergehen, was eine Grundannahme der Einschätzungstheorien darstellt.
2. Die Ergebnisse der einzelnen Untersuchungen waren sich insofern sehr ähnlich, als dass sich die Einschätzungsmuster für die verschiedenen Emotionen stark ähnelten und die meisten den theoretischen Vorhersagen entsprachen.
3. Auch interkulturell konnten diese Übereinstimmungen zumindest für einen Teil der Emotionen nachgewiesen werden.
4. Der Grad der Korrelation zwischen Einschätzungsmustern und Emotionen entsprach zwar nicht den Vorhersagen durch die Einschätzungstheorien, bei denen jeder Emotion ein anderes Einschätzungsmuster zugrunde liegt, doch Meyer et al. [MRS03] führen dies auf methodische Probleme zurück. Durch Beseitigung dieser methodischen Mängel erhöhe sich der Grad der Korrelation hingegen deutlich.

Trotz dieser nicht zu leugnenden Stärken wurden die Einschätzungstheorien zum Teil sogar von ihren Vertretern selbst (z.B. [FRI93a]) zu Recht kritisiert. Neben einer Kritik an den Untersuchungsmethoden, deren Darstellung den Rahmen dieser Arbeit sprengen würde (siehe bei Interesse [MRS03], S. 158f.), wird immer wieder angeführt, dass Emotionen auch auf nicht-kognitivem Weg entstehen können. Die kognitiven Emotionstheorien – für die die OCC-Theorie nur ein Beispiel ist – enthalten aber alle das Grundpostulat, dass Emotionen kognitive Einschätzungen voraussetzen und lassen deshalb keinen Platz für eine nicht-kognitive Emotionsentstehung. Ein Beispiel für so eine nicht-kognitive emotionale Anregung ist jedoch das Behaglichkeitsgefühl, das durch die Wärme eines Ofens, vor dem man gemütlich sitzt, direkt erzeugt wird ([MRS03], S. 160).

Außerdem findet der Aspekt des zeitlichen Verlaufs von Emotionen in der OCC-Theorie keinerlei Berücksichtigung, obwohl dieser Aspekt nach der anfangs eingeführten Arbeitsdefinition als ein (für die vorliegende Arbeit sogar „der wesentliche“) Bestandteil von Emotionen angesehen wird (vgl. Kasten 1, Punkt 2b).

2.3 Zusammenfassung

In diesem Kapitel wurde der emotionspsychologische Hintergrund in groben Zügen dargestellt.

Zunächst wurden entlang einer möglichst unkontrovers gehaltenen Arbeitsdefinition ([MRS01]) die Theorien von Paul Ekman ([EKM99], [EKM84]) zum Ausdrucksaspekt der Emotionen, Wilhelm Wundt ([WUN22]) zur Dynamik von Emotionen und Albert Mehrabian ([MEH96], [MEH95b], [MEH95a]) zur Klassifikation emotionaler (Alltags-)Begriffe in Teilen vorgestellt und zueinander in Beziehung gesetzt. Ergänzt wurden die Ausführungen durch theoretische Überlegungen zur Abgrenzung der Emotionen von Stimmungen und Persönlichkeitseigenschaften.

Anschließend wurde als ein Beispiel für die Klasse der kognitiven Emotionstheorien die Theorie von Ortony, Clore und Collins ([OCC88]) vorgestellt und einzelne Kritikpunkte wurden nachvollzogen.

Folgende Punkte müssen festgehalten werden:

1. Es existiert bis heute keine allgemein gültige und exakte Definition des Forschungsgegenstands „Emotion“. Anhand der in der Arbeitsdefinition aufgeführten vier Merkmale von Emotionen muss von Fall zu Fall entschieden werden, ob ein psychisches Phänomen zu der Klasse der Emotionen gehört.
2. Der Aspekt der Dauer einer Emotion im Allgemeinen genauso wie ihr zeitlicher Verlauf im Speziellen findet in neueren Emotionstheorien wenig Beachtung. Dies kann zumindest teilweise mit der Entwicklung hin zu kognitiven Emotionstheorien in den letzten 20 Jahren begründet werden, die diesen Aspekt weitgehend unbeachtet lassen.

3. Für die Kategorisierung von Emotionen existieren in der Psychologie mehrere umstrittene Ansätze, die von zum Teil sehr unterschiedlichen Grundannahmen ausgehend verschiedene Ziele verfolgen. Es scheint daher gerechtfertigt, die Stärken einzelner Ansätze zu kombinieren, um eigene Ideen zu verwirklichen. Auf welche Weise andere Wissenschaftler dies bereits getan haben, wird im folgenden Kapitel umrissen.

3 Stand der Forschung bei künstlichen humanoiden Agenten

Nachdem im vorangegangenen Kapitel der emotionspsychologische Hintergrund dargestellt wurde, folgt nun eine Beschreibung derjenigen Software-Systeme, die auf der Basis einer kognitiv-adäquaten Modellierung die Erschaffung eines künstlichen humanoiden Agenten zum Ziel haben. Aufgrund des Themas dieser Arbeit wird dabei der Schwerpunkt auf diejenigen Module der Architekturen gelegt, die sich im weitesten Sinne mit Emotionen befassen.

3.1 Systeme, die auf der OCC-Theorie aufbauen

Es existieren mehrere Systeme, die auf der OCC-Theorie aufbauen (siehe Kapitel 2.2). Da diese Theorie auf der Einschätzung von Ereignissen, Handlungen und Objekten beruht, müssen die Systeme breit angelegt sein. Erst wenn Wissensbasen existieren, die Möglichkeit zur Schlussfolgerung gegeben ist, über Sensoren die Umwelt wahrgenommen werden kann und das System im Allgemeinen die Fähigkeit besitzt, Dinge bezogen auf eigene Wünsche, Normen und Einstellungen zu bewerten, kann eine kognitive Emotionstheorie implementiert werden. Daher wundert es nicht, dass in den Anfängen der computergestützten Emotionssimulation der *broad-but-shallow*-Ansatz (breit-aber-flach) gewählt wurde, bei dem zwar viele Komponenten vorhanden sind, diese aber im einzelnen nicht sehr differenziert ausgearbeitet werden.

Zwei Systeme dieses Typs werden im Folgenden vorgestellt, da sie besonders umfassend angelegt worden sind.

3.1.1 Der „Affective Reasoner“ von Clark Elliot

Clark Elliot [ELL92] entwickelte eine *emotion reasoning platform* (Emotions-Schlussfolgerungs-Plattform) auf der Basis der OCC-Theorie, mit der er die Möglichkeiten ergründen will, Emotionen im Computer zu repräsentieren. Innerhalb dieses Projektes wurde der *Affective Reasoner* als *emotion engine* erarbeitet. Es handelt sich dabei um eine Architektur zum abstrakten, domänenunabhängigen Schlussfolgern über Emotionen. Die 22 Emotionskategorien der OCC-Theorie wurden beim Affective Reasoner durch die Emotionskategorien *love* und *hate* auf 24 erweitert und in ein deskriptives Prozessmodell der Emotionen eingefügt, das auf sogenannten *emotion eliciting conditions* (emotionsauslösende Bedingungen) beruht. Durch den *Affective Reasoner* bekommen die Agenten eine „rudimentäre Persönlichkeit“, die aus einer „interpretativen Persönlichkeitskomponente“ – welche für ein gewisses Maß an Individualität bei der Interpretation von Situationen sorgt – und einer „manifestativen Persönlichkeitskomponente“ – verantwortlich für eine persönliche Note bei der Art, wie der Agent seine Emotionen zeigt – besteht. In Anlehnung an die OCC-Theorie ist dabei die „interpretative Persönlichkeitskomponente“ durch eine hierarchische Datenbank von *goals, standards and preferences* (kurz GSP; bei OCC: Wünsche, Normen und Einstellungen, siehe Kapitel 2.2.1) und die „manifestative Persönlichkeitskomponente“ durch eine *action*-Datenbank realisiert.

Der Informationsfluss startet bei einem Ereignis in einer simulierten Welt, die von mehreren Agenten bevölkert wird, durchläuft eine Phase der Emotions- und Aktionsgenerierung und endet schließlich bei der simulierten Wahrnehmung der Aktionen des Agenten durch die anderen Agenten. Ein Agent interpretiert die anfängliche Situation im Hinblick auf eine der *eliciting condition classes* (kurz ECC), die den emotionalen Hauptgruppen in der OCC-Theorie entsprechen: Zielrelevante Ereignisse; Handlungen eines anderen Agenten; attraktive oder unattraktive Objekte. Durch jede erfolgreiche Interpretation entsteht ein neun-Tupel (ECC, siehe oben), das die für den weiteren Verarbeitungsprozess wichtigen Charakteristika der interpretierten Situation enthält und im nächsten Schritt auf eine der 24

Emotionskategorien abgebildet wird. Staller und Petta [STP98] weisen an dieser Stelle auf das Problem hin, dass diese Abbildung einige ad-hoc Lösungen benötige, da sie nicht immer eindeutig sei. Andernfalls käme es vor, dass mehrere widersprüchliche Emotionen gleichzeitig erzeugt würden.

Die instanziierten Emotionen werden ihrerseits auf Aktionen aus der *action-manifestation*-Datenbank abgebildet. Für jede der 24 Emotionskategorien enthält diese Datenbank 17 *action-response*-Kategorien – wie z.B. *somatic responses* (körperliche Reaktionen) oder *behavioral responses* (reaktive Verhaltensweisen) – für die jeweils mehrere (z.B. drei, nach der Intensität der ihr zu Grunde liegenden Emotion unterschiedene) explizite Reaktionen für eine konkrete Anwendung angegeben werden müssen. Damit enthält diese Datenbank mehr als 1000 individuelle Aktionen für jede Anwendungsdomäne. Zudem müssen *conflict sets* (Konfliktmengen) definiert werden, um die gleichzeitige Instanziierung zueinander inkompatibler Aktionen explizit zu vermeiden. Den Agenten ist es mit Hilfe des Konzepts der *social emotions*, die auf einer *Concerns-of-Other*-Datenbank beruht, auch möglich, Empathieemotionen zu repräsentieren.

Folgendes Beispiel ([ELL92], zitiert nach [RUE98]) illustriert die Fähigkeiten des *Affective Reasoners* deutlich:

> „'Toms Auto sprang nicht an, und Tom verpasste deshalb eine Verabredung. Er beschimpfte sein Auto. Harry beobachtete diesen Vorfall.'
>
> Ein klassisches AI-System würde aus dieser Geschichte Schlußfolgerungen folgender Art ziehen: Tom sollte sein Auto reparieren lassen. Harry hat erfahren, daß Toms Auto defekt ist. Tom konnte ohne sein Auto nicht pünktlich zu seiner Verabredung kommen. Harry schlägt Tom vor, in Zukunft früher zu seinen Verabredungen aufzubrechen.
>
> Der Affective Reasoner dagegen würde zu völlig anderen Schlußfolgerungen kommen: Tom macht sein Auto für seine verpaßte Verabredung verantwortlich. Tom ist ärgerlich. Harry kann nicht verstehen, warum Tom auf sein Auto ärgerlich ist, da man ein Auto doch für nichts verantwortlich machen kann. Harry rät Tom, sich wieder zu beruhigen. Harry hat Mitleid mit seinem Freund Tom, weil dieser so erregt ist.“

Staller und Petta [STP98] kritisieren den *Affective Reasoner* in folgenden Punkten:

- Abstrakte, domänenunabhängige Architekturen wie der *Affective Reasoner* bekommen schwerwiegende Probleme, sobald sie in interaktiven virtuellen Szenarien eingesetzt werden. Elliot [ELL94] selbst weist darauf hin, wie wichtig vielfältige Kommunikationskanäle für die Funktion des Gesamtsystems in der Mensch-Maschine-Kommunikation sind.
- Die flache Architektur des *Affective Reasoners* zusammen mit der Repräsentation einer endlichen Anzahl diskreter Emotionen demonstrieren die Probleme der Brüchigkeit und Folgerichtigkeit, die bereits von Expertensystemen bekannt sind.
- Durch die Vergegenständlichung der Emotionen als identifizierbare Systemkomponenten, durch die der Informationsfluss hindurchgeführt wird, entsteht das Problem der anschließenden Informationsverarbeitung von diesen Einheiten weg, hin zum restlichen System, das zu ad-hoc Lösungen zweifelhafter Validität führt.
- Die vollständige Diskretisierung der Emotionen macht ein glattes, fein-schrittiges Systemverhalten unmöglich.

3.1.2 „Em“ von W. Scott Neal Reilly

Reilly [REI96] möchte mit seiner Arbeit dem Künstler ein Werkzeug an die Hand geben, mit dem dieser *believable and social agents* (glaubwürdige und soziale Agenten) erschaffen kann, wie sie zum Beispiel für überzeugende interaktive Computerspiele benötigt werden. Zwar habe die KI bereits gute Werkzeuge entwickelt, mit denen interaktive Agenten erschaffen werden könnten, doch habe sie bisher den Aspekt der Emotionalität dabei unberücksichtigt gelassen.

Auch sein Emotionsmodul *Em* ist Teil eines größeren Ganzen: dem *Oz*-Projekt. Ziel des Projektes ist es, eine Architektur zur Generierung von *interactive dramas* [MAT97] zu entwickeln. Dabei liegt der Schwerpunkt auf der Integration von „Persönlichkeit“ und „Emotionen“, die ein Charakter innerhalb einer interaktiven Geschichte besitzen muss, um glaubwürdig zu sein. Reilly [REI96] weist darauf hin, dass bei der Entwicklung des *Oz*-Sys-

tems aus der Perspektive des Künstlers begonnen wurde, um diesem dann geeignete Werkzeuge aus dem Forschungsbereich der KI bereitzustellen. Um die grundsätzliche Natur des Begriffes *believability* (Glaubwürdigkeit) aus Sicht der Kunst zu beschreiben und damit die manchmal ungewöhnlich anmutenden Teile des *Em*-Moduls nachvollziehbarer zu machen, weist Reilly auf folgende Punkte hin:

- **Glaubwürdige Agenten müssen nicht unbedingt intelligent sein.**
  Als Beispiel für einen glaubwürdigen und trotzdem unintelligenten Charakter nennt Reilly „Forrest Gump“.
- **Glaubwürdige Agenten müssen nicht realistisch sein.**
  Auch ein animierter Stuhl aus Disneys „Die Schöne und das Biest“ wirkt glaubwürdig.
- **Glaubwürdige Agenten müssen starke Persönlichkeiten haben.**
  Diese Persönlichkeit sollte den Agenten in jeder Hinsicht beeinflussen:
  - Wie der Agent geht.
  - Wie er denkt.
  - Wie er spricht.
  - Wie er seine Emotionen zeigt.
  - Auch Eigentümlichkeiten eines Agenten sind erwünscht.

Tabelle 9: Eigenschaften glaubwürdiger Agenten (nach [REI96], S. 10f.).

Mit diesen Anforderungen an „glaubwürdige Charaktere“ begründet Reilly den *broad-but-shallow*-Ansatz, der ebenfalls (vgl. Abschnitt 3.1.1) für diese Architektur gewählt wurde.

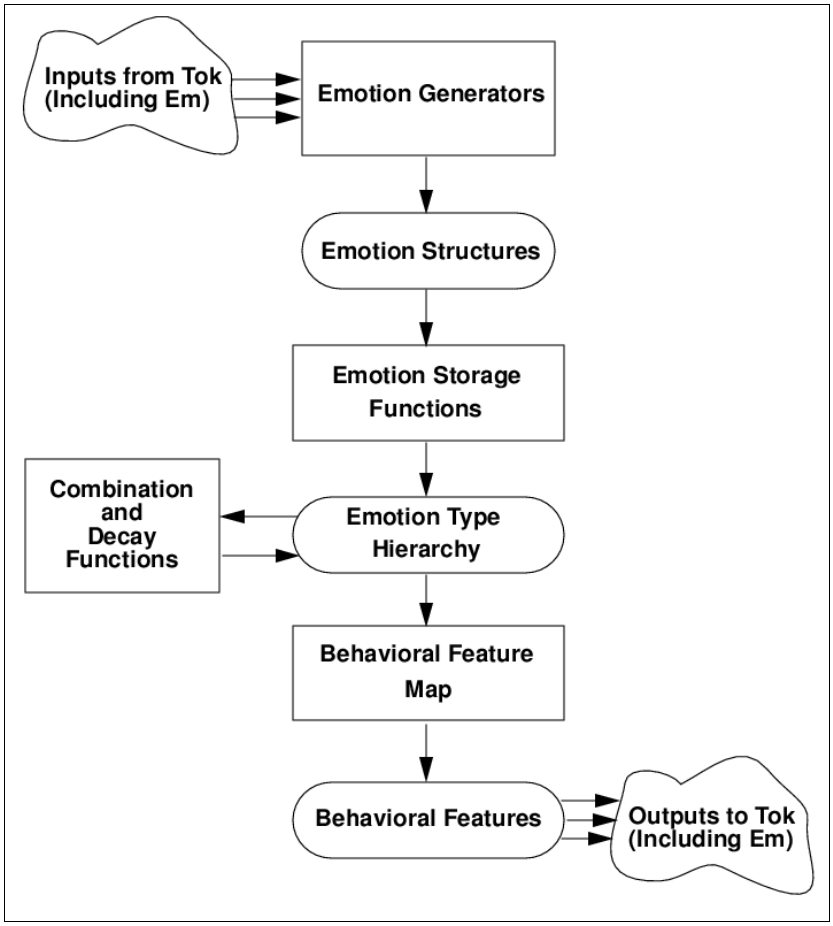


Abbildung 4: Die Em Architektur (nach [REI96], S. 33).

Die ***Em Emotion Architecture*** besteht aus 4 Teilen:

1. *Emotion Generators*: Hier werden *Emotion Structures* regelgeleitet aus den Eingangsdaten abgeleitet, die vom Tok-Agentensystem kommen. Zum Beispiel: Wenn ein Ziel nicht erreicht wird und das Ziel Wichtigkeit X hat, dann generiere eine Emotion Structure vom Typ „Leid“ mit Intensität X.
2. *Emotion Storage Functions*: Diese legen die *Emotion Structures* in einer *Emotion Type Hierarchy* ab. Dabei werden *Emotion Structures* mit allgemeineren Effekten höher einsortiert, als solche mit spezifischeren Effekten. Zum Beispiel könnte „Leid“ in der Hierarchie weiter oben in einer allgemeineren Form mit den Effekten „weinen“ oder „finster blicken“ repräsentiert sein und weiter unten in der Form des „Heimwehs“ mit spezifischeren Effekten wie „Gedanken an zu Hause“, wobei die übergeordneten Effekte geerbt

würden.

3. *Emotion Decay Functions*: Über diese Funktionen kann der Künstler das Abklingen der Emotionen festlegen. Er kann sie explizit für jede *Emotion Structure* (z.B. Wut durch eine Beleidigung vergeht langsamer als eine andere *Emotion Structure*), auf der Ebene der *Emotion Types* (jede Art von Wut vergeht langsam) oder global angeben.
4. *Behavioral Feature Map*: Diese bildet *Emotion Structures* auf *Behavioral Features* ab, wobei nur letztere Einfluss auf das Verhalten des Tok-Agenten nehmen können. Die Abbildung wird dabei als „arbitrary“ (frei wählbar) bezeichnet.

Die in Em repräsentierten Emotionstypen und die Grundlagen für ihre Entstehung werden in ([REI96], S. 58f.) wie folgt zusammengefasst:

Emotion Type	***Cause in Default Em System***
Distress	Goal fails or becomes more likely to fail and it is important to the agent that the goal not fail.
Joy	Goal succeeds or becomes more likely to succeed and it is important to the agent that the goal succeed.
Fear	Agent believes a goal is likely to fail and it is important to the agent that the goal not fail.
Hope	Agent believes a goal is likely to succeed and it is important to the agent that the goal succeed.
Satisfaction	A goal succeeds that the agent hoped would succeed.
Fears-Confirmed	A goal failed that the agent feared would fail.
Disappointment	A goal failed that the agent hoped would succeed.
Relief	A goal succeeds that the agent feared would fail.
Happy-For	A liked other agent is happy.
Pity	A liked other agent is sad.
Gloating	A disliked other agent is sad.
Resentment	A disliked other agent is happy.
Like	Agent is near or thinking about a liked object or agent.
Dislike	Agent is near or thinking about a disliked object or agent.

Emotion Type	***Cause in Default Em System***
Other attitude-based emotions	Agent is near or thinking about an object or agent that the agent has an attitude towards (e.g., awe).
Pride	Agent performs an action that meets a standard of behavior.
Shame	Agent performs an action that breaks a standard of behavior.
Admiration	Another agent performs an action that meets a standard of behavior.
Reproach	Another agent performs an action that breaks a standard of behavior.
Anger	Another agent is responsible for a goal failing or becoming more likely to fail and it is important that the goal not fail.
Remorse	An agent is responsible for one of his own goals failing or becoming more likely to fail and it is important to the agent that the goal not fail.
Gratitude	Another agent is responsible for a goal succeeding or becoming more likely to succeed and it is important that the goal succeed.
Gratification	An agent is responsible for one of his own goals succeeding or becoming more likely to succeed and it is important to the agent that the goal succeed.
Frustration	A plan or behavior of the agent fails.
Startle	A loud noise is heard.

Tabelle 10: Emotionstypen und ihre Entstehung in Em (nach [REI96], S. 58f.).

Reilly weist ausdrücklich darauf hin, dass diese Emotionstypen nicht den Anspruch erheben, psychologisch korrekt zu sein, sondern lediglich dazu dienen, glaubwürdige Charaktere zu erschaffen.

Zur Ermittlung der **generellen Stimmungslage** des Agenten summiert Em zuerst die Intensitäten der positiven (I_p) und negativen (I_n) Emotionen auf [RUE98]:

$$I_p = \log_2\left(\sum_e (2^{I_e})\right), e \in \{positive\,emotions\}$$
$$I_n = \log_2\left(\sum_e (2^{I_e})\right), e \in \{negative\,emotions\}$$

Dann wird nach folgendem Schema die Stimmungslage bestimmt:

IF $I_p > I_n$ THEN set good-mood = I_p AND set bad-mood = 0
ELSE set good-mood = 0 AND set bad-mood = $-I_n$

Bei der Konzeption seiner *Em Emotion Architectur* orientiert sich Reilly zwar an der OCC-Theorie, doch vereinfacht er sie an zwei Stellen auf folgende Weise:

1. In die Berechnung der Intensität einer Emotion fließen nach der OCC-Theorie 4 globale Variablen und 12 lokale Variablen mit ein. Elliot und Siegle [ELS93] erweitern diese sogar auf 24 und geben an, dass es sich immer noch um eine unvollständige Liste handele.

   Reilly verwendet lediglich eine Untermenge dieser Intensitätsvariablen und bezeichnet diese als ausreichend, um glaubwürdige Charaktere zu erschaffen und als einfacher zu verwenden. Außerdem ist es in *Em* möglich, weitere *Emotion Generators* nachträglich zu definieren, um das System einem anderen Szenario anzupassen.
2. Der Entstehungsprozess in *Em* unterscheidet sich ebenfalls von dem strikt kognitiven Entstehungsprozess in der OCC-Theorie. Durch den *broad-but-shallow*-Ansatz ergibt sich die Möglichkeit, einige der umständlichen kognitiven Prozesse der OCC-Theorie zu vereinfachen und allgemeingültiger zu implementieren, da Informationen über die Motivation und die Perzeption des Agenten in den Entstehungsprozess mit einfließen. Um Mitleid zu erzeugen, reicht es in *Em* aus, dass Agent A Agent B mag und dass Agent A denkt, Agent B sei traurig. Dabei kann Agent A allein schon deshalb der Meinung sein, Agent B sei traurig, weil er *sieht*, dass Agent B weint. In der OCC-Theorie müsste dazu Agent A erst zu dem Schluss gelangen, dass ein Ereignis in der Welt von Agent B als unangenehm empfunden wird.

Reilly äußert sich auch zu Problemen bezüglich der *Decay*-Funktionen, die für die vorliegende Arbeit besonders wichtig sind. Da für ihn jede Emotion eine eigene „Zerfallsrate“ und eine eigene „Zerfallsfunktion“ besitzen kann, stehen ihm viele Möglichkeiten offen. Seine Standard-*Decay*-Funktion ist eine lineare Abnahme der Intensität um einen Punkt pro Zeitschritt. Bei „Erschrecken“ ist eine exponentielle Abnahme vorgesehen und „Hoffnung“

und „Furcht“ halten so lange an, wie der Grund der Emotion in der Nähe ist, um danach mit der Rate 1/Zeitschritt abzunehmen.

Im Fall von „Furcht“ wäre es auch möglich, die Standard-*Decay*-Funktion anzuwenden und den *Emotion Generator* erkennen zu lassen, dass die aktuelle Intensität zu gering ist für die gegebene Situation, um dann den Wert erneut zu setzen. Allerdings, so bemerkt Reilly kritisch, führe dies in machen Fällen zu einem sonderbaren Verhalten. Er illustriert dies mit folgendem Beispiel ([REI96], S. 79):

> „Zum Beispiel hat ein Gangster ein „schieße-auf-Officer“-Verhalten, das bei einer Furchtintensität von 7 aktiviert und bei einer Intensität von unter 7 wieder deaktiviert wird. Das Verhalten beginnt mit dem Zielen auf den Officer und endet mit dem Schuss. Wenn die Furchtintensität mit 7 beginnt, so beginnt auch der Gangster mit dem Zielen. Wenn die Intensität fällt, stoppt auch das Verhalten. Wenn die Intensität erneut steigt, wird das Verhalten (wahrscheinlich) wieder gestartet. Dieser Neustart mag oder mag nicht mit einem weichen Übergang beginnen, doch wenn dem nicht so ist, so wird er erneut zielen, anstatt gleich zu schießen. Sorgt man dafür, dass die Intensität auf 7 stabil gehalten wird, entsteht das Problem nicht.“

Im Bewertungsteil seiner Arbeit berichtet Reilly von Versuchspersonen, die sein System in Bezug auf die *Decay*-Funktionen in zwei Punkten kritisieren [REI96], S. 133ff.:

1. *Goal-specific decay*: Zwei Versuchspersonen wollten die Möglichkeit haben, die decay-Funktionen in Abhängigkeit von den Zielen festzulegen, die der Agent gerade verfolgt. Wird der Agent beleidigt, so sollte der resultierende Ärger schneller vergehen, als wenn der Grund für den Ärger darin liegt, angeschossen worden zu sein.
2. *Context-Dependent Decay Functions*: Nachdem der Gangster angeschossen wurde, sollte er nicht nur besonders „ärgerlich“ sein, sondern die Intensität der anderen Emotionen sollte dann schnell abnehmen.

Die Signifikanz der Ergebnisse schwankt allerdings deutlich, so dass Reilly einräumt, *Em* sei lediglich „moderately successful“ ([REI96], S. 129).

3.2 Überblick über weitere Systeme

Es existieren eine Vielzahl weiterer Systeme, von denen einige in folgender Aufstellung zusammengefasst werden:

- TABASCO, [STP98]:

  Die Abkürzung TABASCO steht für: „Eine anwendungsorientierte, bewertungsbasierte Architektur für einen situierten Wahrnehmungsapparat“. Den in Abschnitt 3.1.1 bereits zitierten Kritikpunkten am *Affective Reasoner* wollen Staller und Petta mit dieser *three-level*-Architektur begegnen und bauen dabei auf Theorien von Leventhal [LEV84] auf (siehe Abbildung 5). Diese Architektur soll bei situierten Software-Agenten, wie z.B. MAX, Anwendung finden.

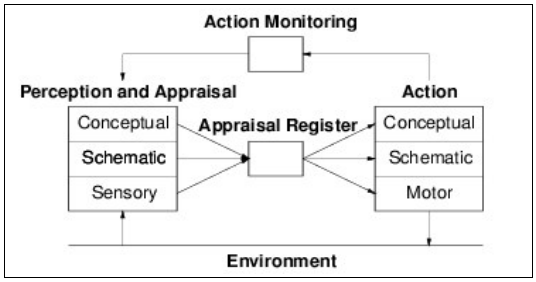


Abbildung 5: Die TABASCO Architektur (nach [STP98], S. 9).

- (Virtual) Yuppy, [VEL98]:

  Velásquez beschäftigt sich mit *Emotion-based Decision Making* und baut dabei auf Arbeiten von Damasio [DAM94] auf, der aus neuropsychologischen Befunden ableitet, dass die Entscheidungsfindung beim Menschen auch durch emotionale Prozesse beeinflusst wird. Velásquez Software-Architektur selbst macht Anleihen bei Minskys *Society of Mind* Modell [MIN86], indem dort den sechs *Basic Emotions* (*Anger*, Fear, *Distress/Sadness*, *Enjoyment/Happiness*, *Disgust* und *Surprise*, vgl. auch Kapitel 2.1.1) einzelne Agenten zugeordnet werden, die sich gegenseitig verstärken und abschwächen können (vgl. Abbildung 6). Einsatz findet diese Architektur sowohl in virtuellen Simulationen als auch in Robotern.

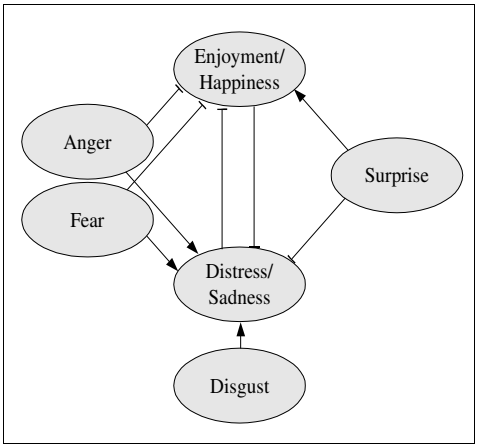


Abbildung 6: Das Agentennetz (nach [LES02]).

- „Ψ“ (Psi), [DÖR99]:

  Am Institut für theoretische Psychologie der Universität Bamberg wird Ψ entwickelt, um zu erforschen, ob ein einfaches Grundelement der Informationsverarbeitung, das (theoretische) Neuron, ausreicht, um komplexes und unter anderem auch emotionales Verhalten zu erzeugen. Dabei werden „Emotionen“ nicht explizit repräsentiert, sondern lediglich Mechanismen vorgegeben, die das beobachtbare Verhalten in einer Weise modulieren (vgl. Abbildung 7), dass man dem System „von außen betrachtet“ Emotionen zuschreiben kann.

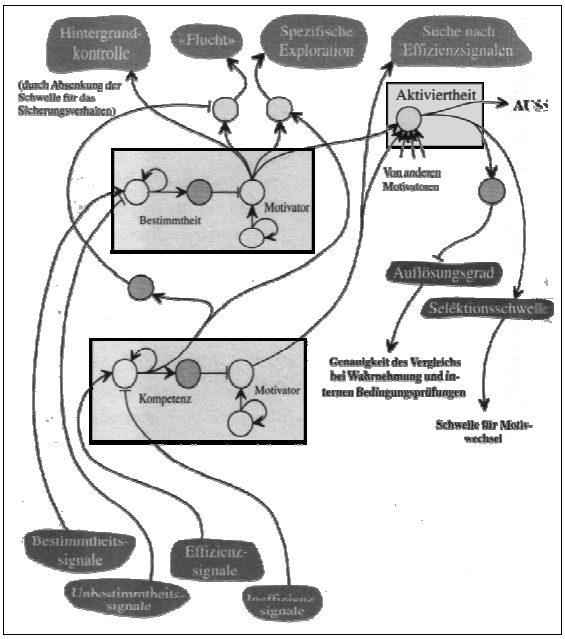


Abbildung 7: Die Einstellung der Modulatoren (nach [DÖR99], S. 538).

- EMOTE, [BAD00]:

  Mit der *Expressive MOTion Engine* werden Bewegungsdaten, die zuvor mittels des *motion-capture*-Verfahrens aufgenommen wurden, mit Hilfe der *Laban Movement Analysis* (LMA) analysiert, um sie dann qualitativ zu verändern. Nach der Analyse liegen u.a. die Parameter *Effort* (Aufwand) und *Shape* (Form) vor, die vor einer Resynthese der Bewegung variiert werden. Schließlich können die Daten durch einen virtuellen 3D-Charakter visualisiert werden.

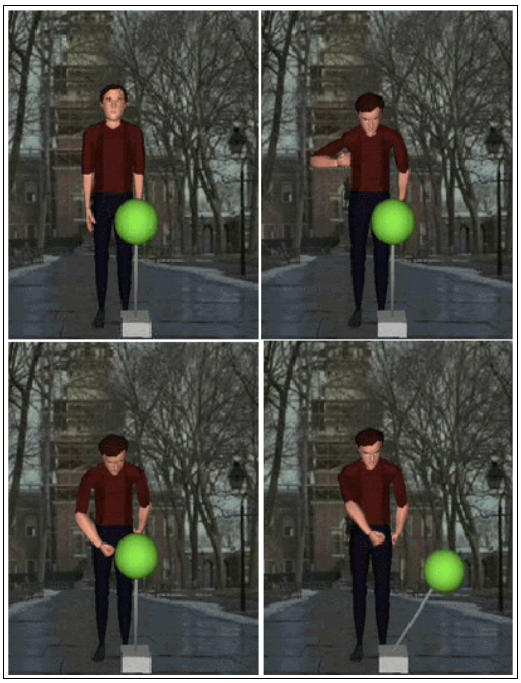


Abbildung 8: "EMOTE alterations of hitting a balloon." (aus [ALL02], S. 4).

- Kismet, [BRE98]:

  Im Rahmen des *Sociable Machines Project* am MIT wird Kismet als *Sociable Human Robot* entwickelt. Er kann seine Umwelt visuell und auditiv wahrnehmen und soll auf natürliche und intuitive Weise sozialen Kontakt aufnehmen und aufrechterhalten können. Dazu besitzt er ein emotionales System, das auf Arbeiten von Itzard und Ekman (vgl. Kapitel 2.1.1, S. 13) aufbaut. Da es den Entwicklern vorrangig um die Simulation kleinkindhafter Emotionen geht, halten sie es für ausreichend, die Ekmanschen Basisemotionen um drei Emotionskategorien (*accepting, stern* und *tired*) zu erweitern. Diese werden durch so genannte *Drives* angeregt, durch die ein nicht näher beschriebener zeitlicher Verlauf der Emotionen in einem dreidimensionalen Raum entsteht (siehe Abbildung 9).

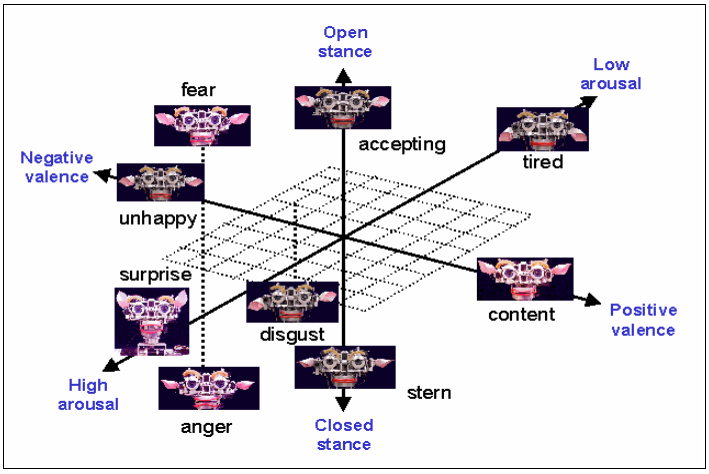


Abbildung 9: "Visual Expressions" (nach [KIZ03])

Prototype	***Function of the Associated Behavior***	***Emotion Associated***	***Activation Conditions for Kismet***
Incorporation	Accept environmental stimulus	acceptance, calm	Acceptance of a desired stimulus
Rejection	Get rid of something harmful already accepted	disgust	Attend to a salient but *undesired* stimulus
Protection	Avoid being destroyed	fear, distress	Appearance of a threatening, overwhelming stimulus
Deprivation	React against important loss	sorrow	Loss of a desired stimulus
Orientation	React to a new or strange object	interest	Appearance of new or *salient* stimulus
Exploration	Explore environment	boredom	Need of a desired yet absent stimulus
Reward	Reinforce beneficial behavior	joy	Success in achieving goal of active behavior
Destruction	Remove barrier to achieve some need	anger, frustration	Delay, difficulty in achieving goal of active behavior
Alert	Stratle Response	surprise	Sudden, unexpected stimulus

Tabelle 11: Kismets emotive Reaktionen und ihre Entstehungsbedingungen (nach [KIZ03]).

3.3 Zusammenfassung

In diesem Kapitel wurden zwei Implementierungen der OCC-Theorie detailliert vorgestellt. Dabei wurden folgende Punkte deutlich:

- Beide Autoren haben die Theorie nicht in ihrer ganzen Komplexität implementiert. Dies ist zum Teil nachvollziehbar begründet worden und doch lässt es vermuten, dass die OCC-Theorie einige konzeptionelle Probleme besitzt, die einer umfassenden Implementation im Wege stehen. Zudem wurde in Kapitel 2 bereits kritisiert, dass die OCC-Theorie eventuell nicht alle Arten der Emotionsentstehung beschreiben kann.
- Große Probleme bereitete beiden Autoren die Tatsache, dass mehrere widersprüchliche Emotionen zur selben Zeit auftreten können. Da diese sich in beiden Systemen weder bei ihrer Entstehung noch in Bezug auf ihren zeitlichen Verlauf gegenseitig beeinflussen, werden am Ende ad-hoc Lösungen verwendet, um ein konsistentes und nachvollziehbares Verhalten der Agenten zu erreichen.
- Nur Reilly beschäftigt sich mit dem zeitlichen Verlauf der erzeugten Emotionen und stößt dabei auf zwei konzeptionelle Probleme. Erstens scheint die Zeitlichkeit einer Emotion unter anderem von dem Ziel abzuhängen, das der Agent verfolgt und das der Grund für die Emotion ist. Zweitens scheinen sich die Emotionen auch in ihrem zeitlichen Verlauf gegenseitig zu beeinflussen.
- Alles in allem wird durch beide Systeme die Erklärungskraft und Mächtigkeit der OCC-Theorie deutlich.

Im Anschluss daran wurden fünf weitere Systeme kurz beschrieben, die sich mit unterschiedlichen Aspekten von Emotionen beschäftigen und ihrer jeweiligen Problemstellung mit unterschiedlichen Konzepten begegnen. Die Architekturprinzipien reichen dabei von einem Drei-Schichten-Modell über ein Agentennetzwerk bis hin zu einem neuronalen Netz, und so unterschiedlich wie diese Ansätze sind auch die theoretischen Hintergründe der modellierten, psychischen Phänomene. Die einen wählen den Weg der OCC-Theorie und versuchen die konzeptionellen Schwächen anderer Systeme mit ihrem Ansatz zu vermeiden. Ein anderer

startet bei den Ekmanschen Basisemotionen und repräsentiert diese in Form von einzelnen Agenten, die sich über ein Agentennetzwerk gegenseitig beeinflussen können. Wieder andere gehen von einem einzigen Grundelement der Informationsverarbeitung aus und vermeiden es, in welcher Form auch immer, Emotionen explizit zu repräsentieren, sondern postulieren, dass der Beobachter dem System anhand seines modulierten Verhaltens Emotionen zuschreiben wird.[1]

Die letzten beiden Systeme sind für diese Arbeit von besonderem Interesse, da sie den Ausdrucksaspekt der Emotionen auf technisch naheliegende Weise behandeln. Bei EMOTE werden Bewegungen des gesamten Körpers durch die Parameter *Effort* und *Shape*, die unter anderem zum Ausdruck von Emotionen verwendet werden können, qualitativ beeinflusst, was im Hinblick auf die Arbeiten zur Synthese und Erkennung von Gestik innerhalb der AG von besonderem Interesse ist. Kismet ist in Form eines Roboterkopfes realisiert und seine Emotionen treten nicht nur in Form seiner Mimik, sondern auch durch Hin- und Abwendung des gesamten Kopfes in Erscheinung. Der Schwerpunkt der Arbeiten bei Kismet liegt dabei auf der sozialen Interaktion mit einem Gegenüber, dabei wird sich auf kleinkindhafte Emotionen beschränkt. Leider konnte nichts über die Dynamik seiner Emotionen in Erfahrung gebracht werden, außer der Tatsache, dass eine solche vorhanden ist.

Abschließend bleibt zu sagen, dass auch in der anwendungsorientierten Emotionsforschung eine Vielzahl unterschiedlicher Ansätze existiert, die nur schwer miteinander vergleichbar sind. Drei Gruppen lassen sich trotzdem unterscheiden:

1. Die kognitiv motivierten Ansätze: Sie wählen meistens die „broad-but-shallow“ Architektur, um genügend Informationskanäle für logisches Schlussfolgern zur Verfügung zu haben. Leider haftet ihnen zu Recht der Vorwurf der „Überkognitivierung“ der Emotionen an.
2. Die ganzheitlichen, nicht-kategorialen Ansätze: Bei ihnen wird davon ausgegangen, dass es nicht nötig sei, Emotionen explizit zu repräsentieren. Vielmehr verhielte sich ein System mit genügend hohem Komplexitätsgrad automatisch „emotional“.

1 Hier muss in Bezug auf Ψ kritisch angemerkt werden, dass für die Gesichtsmimik von Ψ dann doch auf diskrete (Ekmansche) Basisemotionen abgebildet wird.

3. Die ausdrucksbasierten Ansätze: Ihnen kommt es weniger auf kognitive Plausibilität, als vielmehr auf ein überzeugendes, nach außen hin sichtbares Verhalten des Gesamtsystems an. Sie bauen meistens auf den Ekmanschen Basisemotionen auf, da diese besonders geeignet zur Ansteuerung der Gesichtsmimik eines Agenten sind.

Da der in der AG bereits entwickelte virtuelle Agent MAX als Grundlage der vorliegenden Arbeit für eine Referenzinstallation dient und dieser sich mimisch, gestisch und sprachlich äußern sowie in unmittelbaren Kontakt mit einem menschlichen Gegenüber treten kann, liegt eine konzeptionelle Nähe zum dritten Ansatz auf der Hand. Trotzdem wird im nun folgenden Teil der Arbeit versucht, auch den ersten Ansatz mit in die Überlegungen einzubeziehen, da das Emotionssystem Teil einer kognitiv motivierten Gesamtarchitektur ist. Aus diesem Grund scheidet der nicht-kategoriale Ansatz prinzipiell aus, da es Komponenten gibt, die mit Hilfe konkreter symbolischer Repräsentationen von Emotionskategorien logische Schlüsse ziehen können, die schließlich zu intentionalen Handlungen des Agenten führen. Zudem ist ein nicht-kategorialer Ansatz konzeptionell zu weit von dem Ziel der Arbeit entfernt, einen Schwerpunkt auf den Ausdrucksaspekt der Emotionen im zeitlichen Verlauf zu legen.

4 Konzeption des Emotionssystems

4.1 Eingliederung in die Arbeiten der AG

Zur Konzeption der „Simulation der Emotionsdynamik eines künstlichen humanoiden Agenten“ ist es als erstes notwendig, die Rahmenbedingungen zusammenzufassen. Wie bereits in der Einleitung angedeutet, handelt es sich bei MAX [KOP02] um einen multimodalen Assemblierungsexperten, der einem menschlichen Gegenüber bei Konstruktionsaufgaben in einer virtuellen Umgebung Hilfestellung leistet. In diesem Kapitel werden die im Hinblick auf die vorliegende Arbeit relevanten Komponenten von MAX kurz beschrieben, da sie in nicht unerheblichem Maße die Anforderungen für zu entwerfende Emotionssystem prägen.

4.1.1 Das visuelle Äußere

Im Rahmen des Projektseminars „Computeranimation virtueller Charaktere“ wurden die Grundlagen der computergraphischen Animation einer menschenähnlichen virtuellen Figur erarbeitet und fanden bei der Erschaffung von MAX exemplarisch Anwendung. Die Modellierung seines Äußeren sowie die Erarbeitung und Implementierung eines Ansatzes zur Animation der Lippenbewegung synchron zu synthetisch generierter Sprache waren unter anderem die Aufgaben der Teilnehmer. Zur Ansteuerung der Gesichtsmimik wurden auf Arbeiten von Ekman und Friesen [EKF78a] und Fleming und Dobbs [FLD99] aufbauend 21 virtuelle Gesichtsmuskeln definiert, deren Auslenkung das Drahtgittermodell des Gesichtes deformiert. Aus computergrafischen Gründen wurden dabei auch „Muskeln“ definiert, die in einem menschlichen Gesicht nicht vorhanden sind, und auf eine Vielzahl von Muskeln wurde verzichtet, da sie nicht darstellbar sind. So wird das Prinzip der teilweisen, gewichteten De-

formation auch für die Bewegung des gesamten Unterkiefers und der Augenlider ausgenutzt, wodurch MAX auch „kauen“ und „blinzeln“ kann.

Abbildung 10:
Das äußere Erscheinungsbild von MAX

Mit Hilfe dieser 21 Muskeln wurden 8 Gesichtsposturen definiert, die MAX den Ausdruck von Emotionen im Gesicht ermöglichen. Dabei entsprechen 5 Posturen den Ekmanschen Basisemotionen „Freude“, „Wut/Ärger“, „Ekel“, „Traurigkeit“ und „Überraschung“ (siehe Abbildung 11) und 2 Posturen entsprechen den psychischen Zuständen „gelangweilt“ und „konzentriert“. Eine Postur ergibt sich aus technischen Gründen als „neutral“ und bezeichnet die unveränderte Geometrie des Gesichtes, wie es ursprünglich modelliert wurde.

Abbildung 11: Emotionale Gesichtsausdrücke von MAX

So genannte *secondary actions* sorgen für eine gewisse Belebtheit von MAX, indem sie ihn in leicht unregelmäßigen Abständen atmen und blinzeln lassen. Sowohl die zeitlichen Abstände zwischen diesen Animationen als auch die Dauer jeder einzelnen Animation sind dabei parametrisierbar.

4.1.2 Die Sprachsynthese

Ein System zur Sprachsynthese [STÖ02] befähigt MAX, zur Laufzeit geplante Äußerungen

akustisch-sprachlich wiederzugeben. Dabei wurde darauf geachtet, dass diese synthetische Sprache qualitativ durch Parameter verändert werden kann. Im Einzelnen sind dies folgende Parameter der Sprechstimme:

- Grundfrequenz
- Sprechrate
- Grad der Abweichung von der Grundfrequenz (Modulation)

Kasten 2:
Parameter zur globalen, qualitativen Beeinflussung der Sprechstimme

Außerdem kann MAX einzelne Satzteile bzw. Wörter betonen und zeitlich synchron dazu Gesten ausführen. Die verbalen Äußerungen können mit den oben beschriebenen emotionalen Gesichtsausdrücken kombiniert werden.

4.1.3 Die Gestik

Bei der Entwicklung des „Artikulierten Kommunikators“ MAX wurde von Anfang an ein Augenmerk auf eine möglichst natürlich wirkende und auf abstraktem Niveau spezifizierbare Gestik gelegt, deren Timing- und Formparameter flexibel genug sind, damit die durch sie erzeugte Gestik in eine dynamische Umwelt integrierbar ist [KOP02]. Auf abstraktester Ebene können Gesten aus einem Gestenlexikon durch dynamisch aus dem Weltwissen extrahierte Parameter, wie z.B. die Lage eines virtuellen Objektes, angereichert und sprachsynchron ausgeführt werden. Die Gesten des Gestenlexikons selbst werden dabei mit Hilfe einer XML-konformen Sprache namens MURML [KKW02] beschrieben, die auch für die Spezifikation des sprachlichen Anteils einer Äußerung (angereichert durch eventuelle Synchronisationsinformationen für die Gestik) und der emotionalen Gesichtsmimik verwendet werden kann.

Bei der Synthese der Gesten wird ein System verwendet, das einzelne Teile der Gesten zur Laufzeit plant, die sich gegenseitig aktivieren. Dadurch ist dieses System flexibel genug, um Rücksicht auf die gegenseitige Beeinflussung aufeinanderfolgender Gesten zu nehmen und der Interaktivität des Einsatzszenarios gerecht zu werden.

4.1.4 Die Sensorik

MAX kann seine Umwelt auditiv, taktil und visuell wahrnehmen. Über den auditiven Kanal kann MAX natürlich-sprachlichen Äußerungen folgen, die von einem menschlichen Kommunikationspartner stammen. Seine virtuelle Sensorik ermöglicht es ihm, sowohl Kollisionen mit anderen virtuellen Objekten als auch mit den Händen oder dem Kopf seines menschlichen Gegenübers, der dafür mit spezieller *Tracking*-Hardware versehen wird, zu detektieren und deren Position und Entfernung zu ermitteln.

Unter Verwendung verschiedener Wissensbasen hat er Zugriff auf zusätzliche Informationen über die ihn umgebenden Objekte und kann mit Hilfe einer kognitiv motivierten Gesamtarchitektur planvoll und der Situation angemessen handeln.

4.1.5 Die kognitiv motivierte Gesamtarchitektur

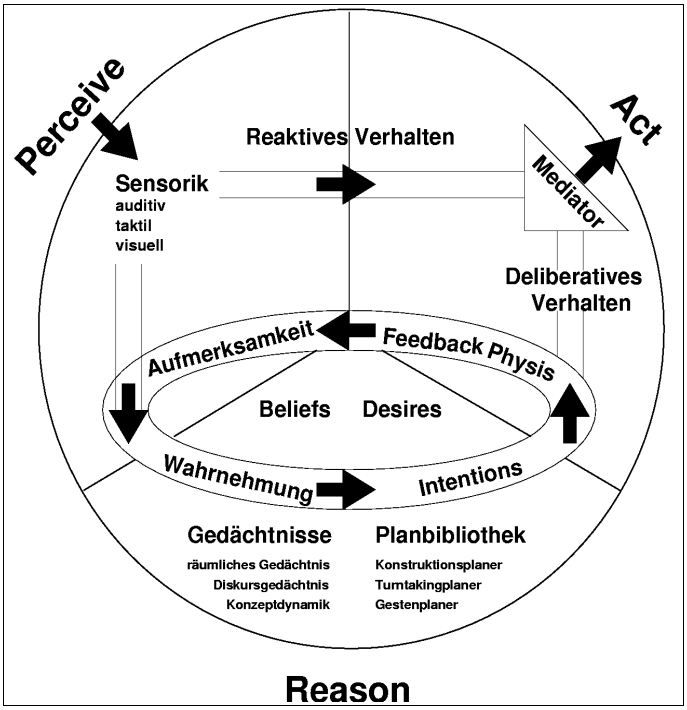


Abbildung 12:
Der Aufbau der Gesamtarchitektur von MAX (aus [LEW03])

Um die technischen Möglichkeiten sinnvoll für die zum Ziel gesetzte Mensch-Maschine-Kommunikation ausnutzen zu können und dabei gleichzeitig zu verhindern, dass konkurrierende Verhaltensweisen zu einem unnatürlichen Verhalten von MAX führen, wurde eine kognitiv motivierte, hybride Gesamtarchitektur entworfen ([LEW03], siehe Abbildung 12). Diese besteht aus einem reaktiven und einem deliberativen Teil sowie einem Mediator, der zwischen konkurrierenden Verhaltensweisen schlichtet.

Das reaktive Verhalten

Reaktive Verhaltensweisen erfüllen zwei konzeptionell unterscheidbare Funktionen:

1. MAX kann datengetrieben reagieren, falls die eintreffenden Sensordaten ein schnelles Handeln erzwingen. Diese Verhaltensweisen haben eine hohe Priorität, um eine unmittelbare Systemreaktion wie ein Ausweichverhalten bei einer drohenden Kollision zu

ermöglichen.

2. Durch die Implementierung der *Secondary Behaviours* werden Teile der Funktionen eines vegetativen Nervensystems simuliert. So „atmet“ und „blinzelt“ MAX, um den Grad der Glaubwürdigkeit zu steigern; jedoch haben diese Verhaltensweisen eine niedrige Priorität, um deliberativ geplante Verhaltensweisen nicht zu überstimmen.

Das deliberative Verhalten

Deliberative Verhaltensweisen ermöglichen MAX ein zielorientiertes, planvolles Handeln. Dabei wird ein BDI-Kernel[2] mit dynamischen, eigenständig operierenden Planern eingesetzt, der durch *Beliefs* Annahmen über den Zustand der Welt repräsentiert, über *Desires* Wünsche über den zu erreichenden Weltzustand enthält und dessen Planungsprozesse Pläne instantiieren, von denen schließlich einer als Intention wirksam wird. In diese Planungsprozesse werden Vorbedingungen, Constraints und Effekte sowie spezialisierte Gedächtnisse mit einbezogen, wobei die durch sie instantiierten Pläne hierarchisch aufgebaut sind.

Die Aufmerksamkeitssteuerung sowie Aspekte des Turntaking und der Feedback-Signale werden vom deliberativen und reaktiven Teil der Gesamtarchitektur gemeinsam übernommen, wodurch MAX zu einer situierten Kommunikation befähigt wird.

4.2 Anforderungen an das Emotionssystem

Die Anforderungen an dieses Emotionssystem ergeben sich aus den theoretischen Aspekten, die in Kapitel 2 behandelt wurden, genauso wie aus den praktischen Aspekten, die im vorangegangenen Kapitel anhand der Rahmenbedingungen thematisiert wurden. Trotz dieser Rahmenbedingungen sollte die Allgemeingültigkeit des Ansatzes nicht zu sehr eingeschränkt werden.

Die zu konzipierende Emotionsdynamik soll Bestandteil eines künstlichen humanoiden Agenten sein und das Adjektiv „künstlich“ deutet an, dass das Subjekt selbst, welches in den

2 *(B)elief-(D)esire-(I)ntention*-Kernel, für eine ausführliche Darstellung siehe [LES02].

Besitz dieser Emotionsdynamik kommen soll, auch nur in Form einer Simulation existiert. Allerdings soll dieses Subjekt **menschenähnlich** (humanoid) wirken und fähig sein, **im Auftrag einer anderen Person** zu **handeln** (Agent). Dieser Agent muss also zwangsläufig glaub- und vertrauenswürdig genug erscheinen, damit das menschliche Gegenüber ihn trotz seines Wissen um seine Künstlichkeit akzeptiert.

Reilly [REI96] weist darauf hin, dass glaubwürdige Agenten, die sich bei ihm im Rahmen von *interactive dramas* bewegen, nicht unbedingt intelligent sein und auch kein realistisches Äußeres besitzen müssen. Um Glaubwürdigkeit zu erreichen, sei vielmehr eine „starke Persöhnlichkeit“ zwingend notwendig (vgl. Tabelle 9 auf Seite 41). Dabei solle sich diese Persönlichkeit sowohl direkt als auch indirekt auf das Verhalten des Agenten auswirken; direkt durch Modulation der Gangart und der Sprechweise des Agenten und indirekt durch die Art, „wie er seine Emotionen zeigt“ und „wie er denkt“. Wie schnell eine emotionale Erregung abklingt, hängt unter anderem von der Persönlichkeit eines Menschen ab. Das hier zu konzipierende Emotionssystem sollte in der Weise parametrisierbar sein, dass der Teil der Persönlichkeit, welcher mit dem individuell unterschiedlichen zeitlichen Verlauf der Emotionen zu tun hat, repräsentiert wird.

Bei der Berechnung der **generellen Stimmungslage** beachtet Reilly nicht die konzeptionellen Unterschiede zwischen Emotion und Stimmung, wie sie in Abschnitt 2.1.2 (S. 20) herausgearbeitet wurden. Eine einfache Summation der Wertigkeiten aller zu einem Zeitpunkt vorhandenen Emotionen (vgl. S. 44f.) wird der psychologischen Theorie in drei Punkten nicht gerecht:

1. Die Art der Emotionen, die ein Mensch zu einem gewissen Zeitpunkt empfinden kann, hängt unter anderem von seiner Stimmungslage zu diesem Zeitpunkt ab.
2. Ebenso führen kurz aufeinanderfolgende Emotionen ähnlicher Valenz (d.h. z.B. mehrere positive Emotionen) zu einer Stimmung ähnlicher Valenz (z.B. positiv). Dies wurde in Kapitel 2 auf Seite 20 mit dem „Aufschaukeln der Emotionen“ beschrieben.
3. Eine Stimmung ist ein zeitlich stabilerer Zustand als die meisten Emotionen. Aber auch zwischen den einzelnen Emotionskategorien existieren Unterschiede in ihrer Dauer und

schließlich wird davon ausgegangen, dass sich fast jede Emotion zu einer Stimmung entwickeln kann.

Sicherlich trägt Reillys Vorgehensweise dem zweiten und dritten Kritikpunkt in gewisser Weise Rechnung, doch fehlt in seinem Modell die Auswirkung der Stimmungslage auf die Emotionen. Wenn sich jemand in einer positiven Stimmung befindet, dann ist es nicht so einfach, ihn „aus der Ruhe zu bringen“, also zum Beispiel wütend zu machen. Analog dazu wird sich jemand, der „schlecht drauf ist“ (negative Stimmung), nur schwer erfreuen lassen (vgl. Zitat auf S. 20). Für die Konzeption einer Emotionsdynamik, bei der ein **Schwerpunkt auf den zeitlichen Verlauf der repräsentierten Emotionen** gelegt wird und die außerdem die **gegenseitige Beeinflussung der Emotionen** beachten will, erscheint es deshalb sinnvoll, Stimmungen mit in die Überlegungen einzubeziehen. Dadurch wird es möglich, die oben genannten zeitlichen Aspekte sinnvoll zu integrieren.

Der Roboter Kismet besitzt eine große konzeptionelle Nähe zu dem für diese Arbeit vorgesehenen Zielszenario eines künstlichen humanioden Agenten, der mit seiner Umwelt in direkten sozialen Kontakt treten kann. Da ihm lediglich seine Gesichtsmimik als Ausdrucksmittel seiner Emotionen zur Verfügung steht, wird sich bei ihm auf die Simulation derjenigen Emotionskategorien beschränkt, die sich im Gesicht widerspiegeln lassen. Das vorgestellte dreidimensionale Emotionssystem (vgl. S. 49, Abbildung 9), in dem die neun Emotionskategorien verortet werden, besitzt eine bemerkenswerte Ähnlichkeit zu der in Kapitel 2 ab Seite 16 dargestellten Emotionstheorie von Mehrabian. Der künstliche humanoide Agent MAX, wie er in Abschnitt 4.1 beschrieben wurde, muss ebenfalls für den Dialog mit dem Menschen die Fähigkeit besitzen, soziale Kontakte aufzubauen und zu erhalten. Auch MAX besitzt die Möglichkeit, auf nonverbale Weise zu kommunizieren, mit Hilfe seiner **Gestik**, seiner **Körperhaltung** und seiner **Gesichtsmimik**. Darüber hinaus kann über **sichtbare Anteile eines simulierten vegetativen Nervensystems** (Atmung, Lidschlag u.ä.) schon vor der kognitiven Verarbeitung durch das BDI-Modul eine Emotion Wirkung zeigen.

Es muss also eine Möglichkeit gefunden werden, die Simulation einer Emotionsdynamik,

welche auf der Unterscheidung und gegenseitigen Beeinflussung von Emotionen und Stimmungslage beruht, mit dem dreidimensionalen System zur Kategorisierung von emotionalen Begriffen, wie Wundt und Mehrabian es in der psychologischen Emotionstheorie einführen und wie es für Kismet bereits in ähnlicher Weise eingesetzt wird, zu verbinden. Dadurch würde der Emotionsdynamik des Emotionssystems ein „Aufschaukeln der Emotionen“ implementiert werden können, durch das ein sprunghaftes Wechseln der emotionalen Qualität vermeidbar wäre und somit MAX über ein konsistenteres emotionales Langzeitverhalten verfügen würde. Eine anschließende Abbildung auf theoretisch fundierte Emotionskategorien kann nur in einem zweiten dreidimensionalen Raum vorgenommen werden, da dabei die Stimmungslage und Emotion der Emotionsdynamik in geeigneter Weise auf solche Parameter abgebildet werden müssen, die sich in den psychologischen Theorien als hinreichend für die Beschreibung beliebiger Emoionskategorien erwiesen haben.

Folglich besteht das hier zu entwickelnde System aus **zwei Komponenten**:

1. Eine Komponente ist für die **Simulation der Dynamik** verantwortlich. In diese müssen die Überlegungen zum zeitlichen Verlauf der Emotionen und zum Zusammenhang zwischen Stimmungslage und Emotionen einfließen.
2. Die **Abbildung** der emotionalen Valenzinformation und der Stimmungslage aus der ersten Komponente **auf konkrete emotionale Begriffe** sollte eine zweite Komponente liefern. Dabei scheint es sinnvoll, sich auf die Repräsentation der in der Gesichtsmimik unterscheidbaren Ekmanschen Basisemotionen (siehe Abschnitt 2.1.1) – evtl. erweitert durch die Kategorien „Langeweile“ und „Ausgeglichenheit“, die aus historischen Gründen bereits vorhanden sind – zu beschränken.

Mit dieser Zweiteilung wird eine Trennung der Dynamik von der Kategorisierung ermöglicht, wodurch einer der beiden Teile zu einem späteren Zeitpunkt leichter durch ein anderes Konzept, das zum Beispiel eine andere Theorie als Grundlage hat, ersetzt werden kann. Das Gesamtsystem sollte **möglichst offen realisiert** sein, damit die Möglichkeit für spätere Ergänzungen vorhanden ist und mit den einzelnen Parametern experimentiert werden kann. Alle wichtigen **Parameter** der Simulation sollten **intuitiv zur Laufzeit veränderbar**

und die **Eigendynamik** des Systems sollte **leicht beobachtbar** sein, was nur durch eine adäquate Visualisierung der Daten erreicht werden kann.

Die Simulation der Emotionsdynamik ist wie beschrieben nur ein Teil einer kognitiv motivierten Gesamtarchitektur, die aus mehreren Komponenten besteht und sich noch in der Entwicklung befindet. So ist es nicht auszuschließen, dass später hinzukommende Komponenten in der Lage sein müssen, direkt Einfluss auf das Emotionssystem nehmen zu können. Außerdem sollte es allen Komponenten unabhängig voneinander möglich sein, Informationen über die emotionale Lage des Agenten direkt zu erlangen. Deshalb sollte dieses Emotionssystem die Form eines eigenständig ablaufenden Software-Agenten haben, wie er durch die Software-Infrastruktur der Arbeitsgruppe als *WBS-Agent* realisierbar ist. Hierbei muss betont werden, dass es sich bei dieser Entscheidung um die Auswahl eines softwaretechnischen Design-Prinzips handelt und nicht um eine inhaltlich konzeptuelle Entscheidung, wie sie im Fall des Agentennetzwerks von Velásquez (vgl. Abschnitt 3.2, S. 47) getroffen wurde. Die WBS-Agenten, wie sie in der Arbeitsgruppe Wissensbasierte Systeme entwickelt wurden, sind als eigenständige Prozesse realisiert, die auf dem TCP/IP-Protokoll aufbauend KQML-konforme Nachrichten (und damit Daten) untereinander austauschen können. Diese Agenten können dynamisch zu einem Agentennetz verschaltet werden, womit auch ein „An- und Abschalten der Emotionen“ zur Laufzeit möglich wird.

4.3 Aufbau des Emotionssystems

Die vorangegangenen Überlegungen zum hier zu entwickelnden Emotionssystem lassen sich wie in Abbildung 13 darstellen:

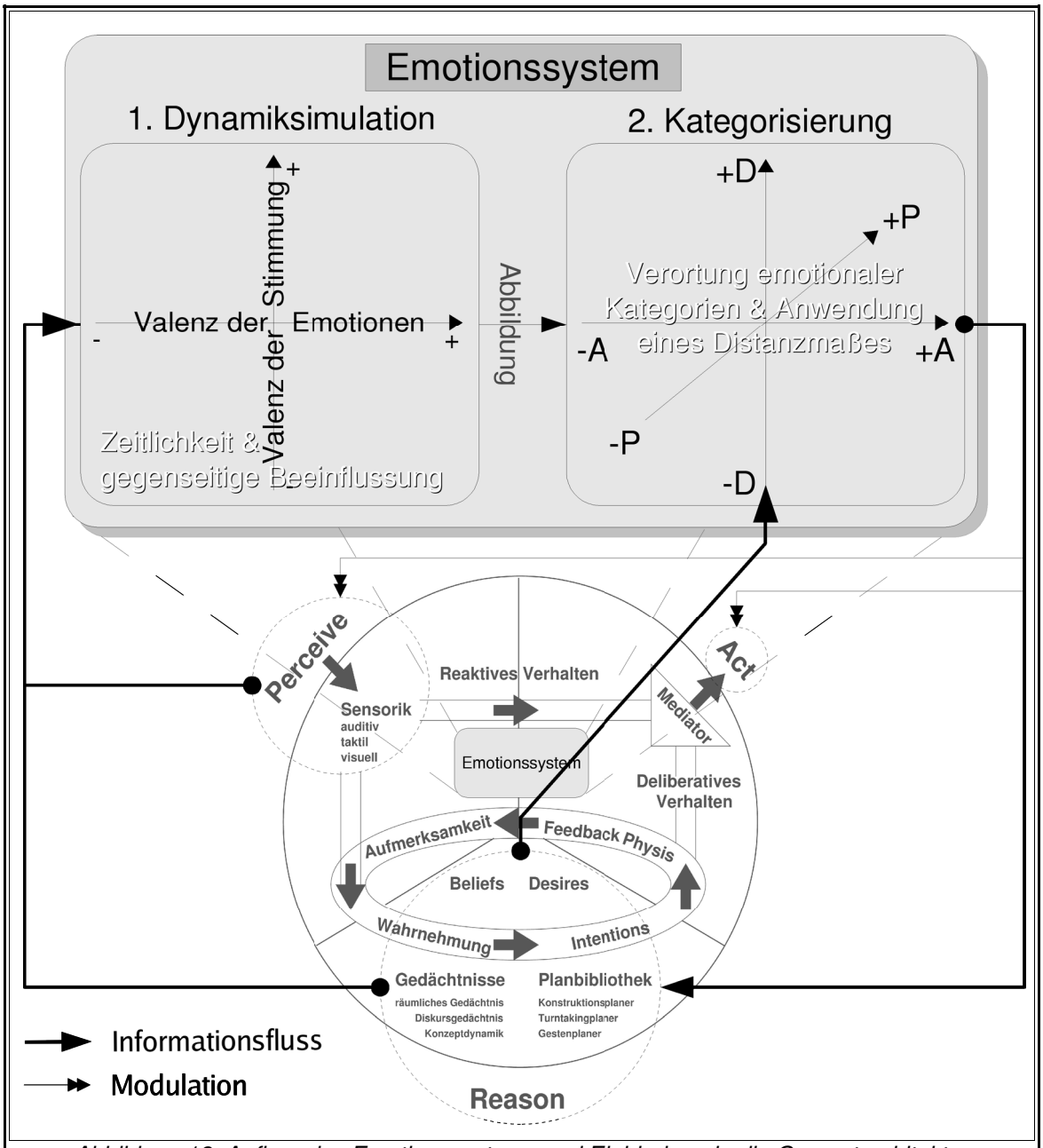


Abbildung 13: Aufbau des Emotionssystems und Einbindung in die Gesamtarchitektur

Die Gesamtarchitektur findet sich in Abbildung 13 in Form des dreigeteilten Kreises wieder. Da das Emotionssystem einen Teil der Gesamtarchitektur ausmacht und mit den anderen Komponenten hochgradig vernetzt ist, befindet es sich zentral innerhalb des Kreises. Um seinen inneren Aufbau verdeutlichen und den Informationsfluss leichter wiedergeben zu können, wurde es nach oben hin vergrößert, was durch die gestrichelten Linien angedeutet wird.

4.3.1 Die Dynamiksimulation

Die erste Komponente des Emotionssystems ist für den zeitlichen Verlauf und die gegenseitige Beeinflussung von Emotionen und Stimmungen verantwortlich. In diese fließen Valenzinformationen sowohl der Emotionen als auch der Stimmung mit ein. Durch die orthogonale Anordnung der beiden Achsen „Valenz der Emotionen“ und „Valenz der Stimmung“ wird auf den unterschiedlichen zeitlichen Verlauf dieser beiden psychischen Phänomene Rücksicht genommen. Die beiden Achsen sind prinzipiell unabhängig voneinander, so dass das Abklingen der emotionalen Valenz unabhängig von der Stimmungslage und ihrer eventuellen Rückkehr zur Ausgeglichenheit implementiert werden kann.

Ein **Aufschaukeln der Emotionen**, wie es auf Seite 20 theoretisch begründet wird, kann nun realisiert werden, indem die „Valenz der Emotionen“ zugleich als Gradient interpretiert wird und dieser Gradient zu einem Anstieg bzw. Abfall der Stimmung führt. In Abbildung 14 wird der Ansatz schematisch dargestellt. Die Länge der gestrichelten Pfeile deutet den Betrag der Gradienten an und die Richtung, in die die Pfeile zeigen, geben das Vorzeichen der Gradienten wieder. Der an einer Stelle vorhandene Gradient ist also proportional zu dem Wert der Abweichung (also zu der Valenz) auf der X-Achse.

Diese Abhängigkeit lässt sich formal beschreiben durch:

$$\dot{y} = m \cdot x$$

Formel 1: "Aufschaukeln der Emotionen".

wobei der Proportionalitätsfaktor m als ein Persönlichkeitsmerkmal angesehen werden kann. „Launische“ Agenten würden durch ein größeres m modelliert, als zum Beispiel Agenten, die „durch nichts aus der Ruhe zu bringen“ sind.

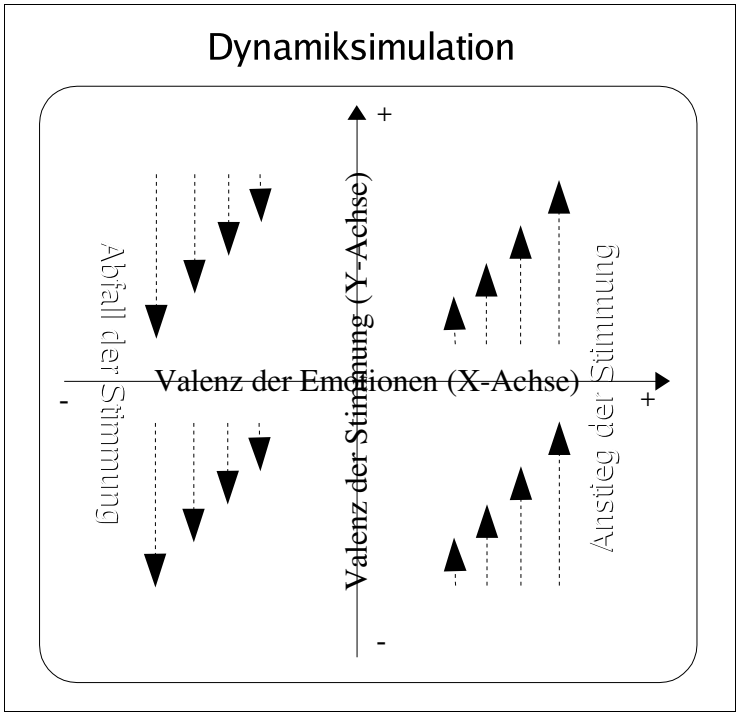


Abbildung 14: Das "Aufschaukeln der Emotionen".

Als nächstes stellt sich **die Frage des zeitlichen Verlaufs der Emotionen bzw. der Stimmung**. In Anlehnung an die Arbeiten von Reilly (vgl. Abschitt 3.1.2, S. 40ff.) wird auch in der vorliegenden Arbeit davon ausgegangen, dass die Intensität der beiden Valenzen mit der Zeit kontinuierlich abnimmt. Da dieses Prinzip sowohl in Bezug auf die Valenz der Emotionen als auch in Bezug auf die Valenz der Stimmung Anwendung findet, ergibt sich im Zusammenspiel mit den oben geschilderten Gradienten ein „Gefühlsverlauf“, der eine gewisse Ähnlichkeit zu den in der Wundtschen Theorie postulierten Gefühlsverläufen aufweist (vgl. Abbildung 1, S. 16). Zwar muss dazu erst noch in den der Wundtschen Theorie ähnlichen dreidimensionalen Raum zur Kategorisierung abgebildet werden, jedoch wird diese Abbildung zu jedem Zeitpunkt unter Erhaltung der Stetigkeit durchgeführt (Details dazu siehe Abschnitt 4.3.2). Zur exakten Spezifikation des zeitlichen Verlaufs schlägt Reilly [REI96] drei Funktionen vor:

1. Lineare Abnahme der Intensität (f(t), „one per tick“)
2. Exponentielle Abnahme der Intensität (g(t), ständige Halbierung der Intensität)
3. Reziproke Abnahme der Intensität (h(t), 1/(Anzahl Zeitschritte seit Maximalintensität))

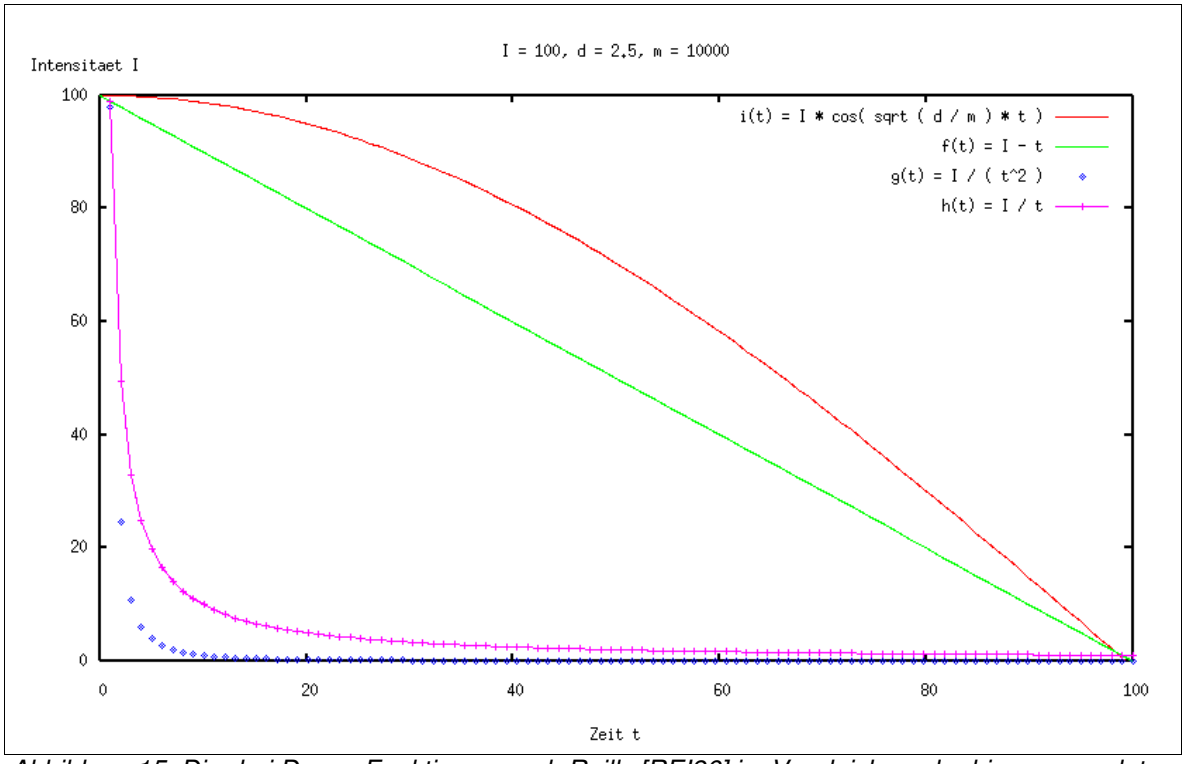


Abbildung 15: Die drei Decay-Funktionen nach Reilly [REI96] im Vergleich zu der hier verwendeten Decay-Funktion i(t)

Die Funktionen f(t), g(t) und h(t) entsprechen den von Reilly beschriebenen Decay-Funktionen und erscheinen alle unvernünftig:

- Bei der Funktion f(t) ist die Zeit, die benötigt wird, um den Intensitätswert zu halbieren (die „Halbwertszeit“), linear abhängig von der Anfangsintensität der Emotion. Außerdem befindet sich der Agent gleich lange in jeder Abstufung der emotionalen Intensität.
- Die Funktionen g(t) und h(t) ähneln sich stark, wobei g(t) die extremere Form darstellt. Die Halbwertszeiten dieser Funktionen sind konstant und wären nach Erweiterungen der Formeln auch parametrisierbar. Warum das Maximum der Intensität so schnell verlassen werden sollte, wie es diese Funktionen vorgeben, wird allerdings von Reilly nicht begründet. Um den Effekt des Aufschaukelns der Emotionen zu erleichtern, ist ein

umgekehrtes Geschwindigkeitsprofil besser geeignet.

Mit der **Funktion i(t)** werden diese Probleme vermieden. Erstens besitzt auch sie eine konstante Halbwertszeit, die über den Parameter d festzulegen ist und zweitens nimmt der Grad der Abnahme (also die „Geschwindigkeit“) mit der Zeit langsam zu, anstatt schnell zu beginnen und dann rapide abzunehmen. Dadurch liegt die Intensität der Emotion eine längere Zeitspanne über der Hälfte der Anfangsintensität als darunter.

$$i(t) = I \cdot \cos\left(\sqrt{\frac{d}{m}} \cdot t\right)$$

Formel 2: Die Decay-Funktion i(t)

Der Verlauf der Funktion i(t) entspricht dabei exakt der **Bewegung einer Spiralfeder**, die zu Anfang auf den Intensitätswert ausgelenkt wird und auf die dann eine Rückstellkraft einwirkt, die abhängig von der Federkonstante d und der Auslenkung x ist.

$$F(t) = -d \cdot x(t-1)$$
$$a(t) = \frac{F(t)}{m}$$
$$v(t) = a(t) \cdot \Delta t + v(t-1)$$
$$s(t) = v(t) \cdot \Delta t + s(t-1)$$
$$\text{mit } t \equiv \text{Zeit}, m \equiv \text{Masse}, d \equiv \text{Federkonstante}$$

Formel 3: Formeln zur Simulation einer Spiralfeder

Würde die in Formel 3 aufgeführte rekursive Berechnung ohne Randbedingungen iterativ für jeden Zeitschritt durchgeführt, so erhielte man eine gleichmäßige ungedämpfte Schwingung. Geht man davon aus, dass die simulierte Spiralfeder im Nullpunkt des Koordinatensystems verankert ist, so entstünde nach einer Auslenkung der Spiralfeder ein gleichmäßiges Pendeln um den Nullpunkt, was in keiner Weise dem zeitlichen Verlauf einer Emotion oder einer Stimmung gerecht würde. Deshalb wird die Simulation der beiden Spiralfedern im Nullpunkt abgebrochen und die physikalischen Simulationsparameter Geschwindigkeit v(t) und Beschleunigung a(t) auf Null zurückgesetzt. Vereinfachend wird davon ausgegangen, dass die Spiralfedern selbst keine Masse besitzen und sie unabhängig voneinander Einfluss auf den simulierten Massepunkt nehmen. Lediglich durch die Wahl der

Federkonstanten d_x und d_y können ihnen unterschiedliche physikalische Eigenschaften und damit unterschiedliche Halbwertszeiten verliehen werden.

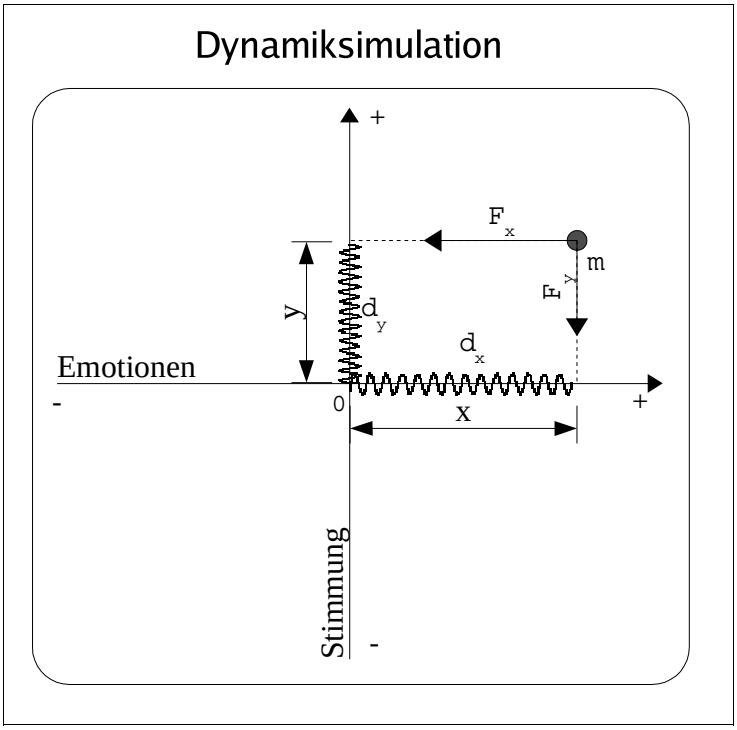


Abbildung 16:
Schematische Darstellung der Spiralfeder-Simulation

Um „**Langeweile**“ repräsentieren zu können, muss eine **dritte orthogonale Achse** eingeführt werden. Das psychische Phänomen „Langeweile“ wird beim Menschen dadurch hervorgerufen, dass er über einen längeren Zeitraum hinweg keine neuen oder für ihn nicht relevante Informationen wahrnimmt. Die Tatsache, dass **das Fehlen von Stimuli** zu „Langeweile“ führt, macht Zeitinformationen und eine interne Dynamik für die Modellierung dieses Phänomens unumgänglich. In der Dynamik-Komponente wird der Grad der Langeweile als eigenständiger Simulationsanteil mit unabhägiger Dynamik vorgesehen. Unter der Annahme, dass einem Menschen nur bei emotionaler Ausgeglichenheit und neutraler Stimmungslage langweilig werden kann, wird eine **ϵ-Umgebung um den Nullpunkt** des bisherigen Systems eingeführt (siehe Abbildung 17). Ist also der Zustand des Agenten als annähernd neutral zu einem Zeitpunkt zu bezeichnen, so nimmt der Grad der Langeweile stetig linear zu. Um dies zu visualisieren, wird eine Z-Achse eingeführt, die (analog zu einem rechtshändigen Koordinatensystem) aus der X-Y-Ebene herauszeigt. Da es für die Abbildung in den Raum der Kategorisierung sinnvoll ist (wie später erläutert werden wird), wird der **Grad der Langeweile mit negativem Vorzeichen** notiert. Dadurch sinkt der Referenzpunkt im Dynamik-Raum bei zunehmender Langeweile „nach unten“ unter die X-Y-Ebene. Da dieses Absinken auf lineare Weise erfolgt, ist ein Parameter für die (konstante) Geschwindigkeit, mit

der dem Agenten langweilig wird, ausreichend. Auch dieser Parameter kann als **Persönlichkeitsmerkmal** angesehen werden. Sollte sich der Referenzpunkt außerhalb der durch ϵ_x und ϵ_y beschriebenen ϵ-Umgebung befinden, so wird der Grad der Langeweile **automatisch auf Null zurückgesetzt**, so dass dem Agenten per Definition bei größerer emotionaler Erregung oder nicht annähernd neutraler Stimmungslage nicht langweilig sein kann.

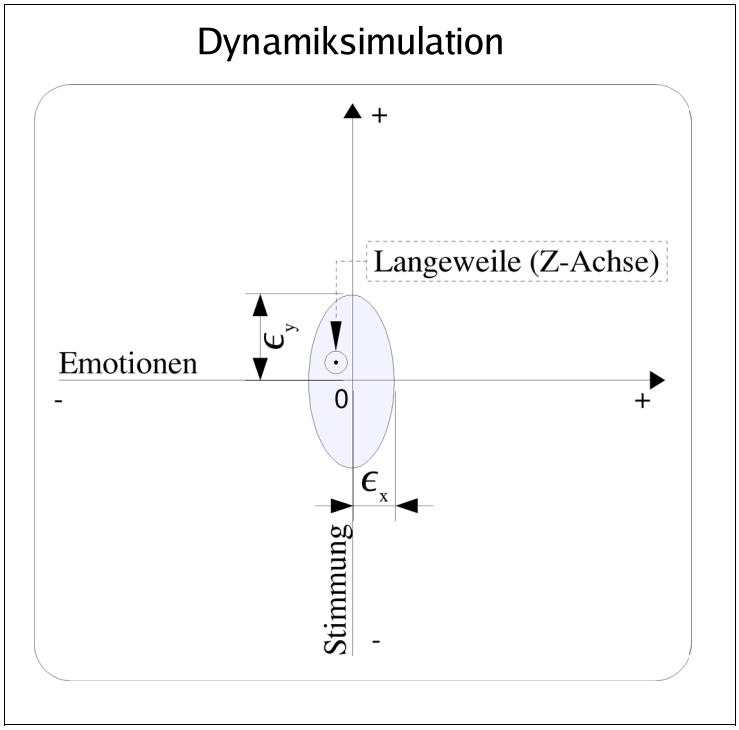


Abbildung 17:
Schematische Darstellung der ϵ-Umgebung

Die Dynamiksimulation liefert zu jedem Zeitpunkt t folgendes Tripel:

$$D(t)=(x_t, y_t, z_t), \text{ mit } x_t=[-1,1], y_t=[-1,1], z_t=[-1,0]$$

Formel 4: Das Ausgabetripel der Dynamiksimulation

Der Wert x_t bezeichnet dabei die emotionale Valenz zum Zeitpunkt t und wird innerhalb des geschlossenen Intervalls [-1, 1] liegen. Folglich bezeichnet y_t den Wert der zum Zeitpunkt t herrschenden Stimmung und dieser liegt ebenfalls innerhalb des geschlossenen Intervalls [-1, 1]. Für z_t, das den Wert der „Langeweile“ bezeichnet, gilt als Einschränkung des Wertebereichs hingegen das geschlossene Intervall [-1, 0], da die „Langeweile“ einpolig repräsentiert wird.

4.3.2 Die Kategorisierung

Die **Abbildung $K(x_t, y_t, z_t, t)$** führt vom Dynamik-Raum in den für die Kategorisierung verwendeten PAD-Raum. Die Bedeutungen der drei Achsen des PAD-Raumes sind (vgl. Abschnitt 2.1.1, S. 12ff):

- (P)leasure ↔ Displeasure (zu dtsch. Lust (Freude) ↔ Unlust (Verdruss)):
  Der Informationsgehalt dieser Achse betrifft hauptsächlich die **Valenz der Emotionen**.
- (A)rousal ↔ Nonarousal (zu dtsch. Erregung ↔ Beruhigung):
  Auf dieser orthogonalen und unabhängigen Achse wird das **Maß der körperlichen und mentalen Erregung** notiert. Es wird vermutet, dass Emotionen tendenziell durch einen höheren Grad der Erregung gekennzeichnet sind als Stimmungen. Bei absoluter Beruhigung liegt Schläfrigkeit/Schlaf vor.
- (D)ominance ↔ Submissiveness (zu dtsch. Dominanz ↔ Unterwürfigkeit (Ergebenheit)):
  Diese dritte orthogonale und unabhängige Achse ist ein Maß dafür, wie sehr der Agent meint, **die Situation zu dominieren/kontrollieren**. Durch sie wird es möglich, zwischen den Emotionskategorien „Wut“ und „Angst“ sowie „Unzufriedenheit“ und „Traurigkeit“ zu unterscheiden. Wut und Unzufriedenheit werden im Zustand der (scheinbaren) Dominanz empfunden, wohingegen Angst und Traurigkeit im Falle von (scheinbarer) Unterwürfigkeit empfunden werden. Die Pleasure- und Arousal-Werte sind bei diesen Paaren gleich. Beispielsweise könnten die Emotionskategorien auf folgende Weise repräsentiert werden:
  - Wut ⇐ (-P, +A, **+D**) ↔ Angst ⇐ (-P, +A, **-D**)
  - Unzufriedenheit ⇐ (-P, 0, **+D**) ↔ Traurigkeit ⇐ (-P, 0, **-D**)

Unter Beachtung dieser Annahmen erscheint folgende Abbildung vom Dynamik-Raum in den PAD-Raum sinnvoll:

$$K(x_t, y_t, z_t, t) = (p(x_t, y_t), a(x_t, z_t), d(t)),$$
$$\text{mit } p(x_t, y_t) = \frac{1}{2} \cdot (x_t + y_t) \text{ und } a(x_t, z_t) = |x_t| + z_t$$

Formel 5: Die Funktion $K(x_t, y_t, z_t, t)$

Die Funktion $K(x_t, y_t, z_t, t)$ liefert ein Tripel, das aus den Funktionen $p(x_t, y_t)$ für die Berechnung des Pleasure-Wertes, $a(x_t, z_t)$ für den Arousal-Wert und $d(t)$ für den Dominance-Wert besteht.

- Die **Funktion $p(x_t, y_t)$** verwendet zur Berechnung des Pleasure-Wertes die **normierte Summe der Valenzen von Emotion und Stimmung** der Dynamiksimulation. Die Normierung mit dem Faktor 0.5 wird notwendig, damit der Wertebereich innerhalb des geschlossenen Intervalls [-1, 1] liegt. Die „maximale Lust“ läge also bei maximaler positiver Stimmung und einer maximal-positiven Emotion vor und „„maximale Unlust“ im gegenteiligen Extrem von Stimmung und Emotion.
- In die **Funktion $a(x_t, z_t)$** fließen der **Betrag der emotionalen Valenz x_t** und der **Grad der Langeweile z_t** mit ein. Jede Abweichung von neutraler emotionaler Valenz führt somit zu einem Anstieg der Erregung. Im Gegensatz dazu hat eine eventuelle negative oder positive Stimmungslage keinen Einfluss auf das Arousal. Im Zusammenhang mit der Simulation der Langeweile verhindert eine nicht neutrale Stimmung jedoch, dass der Erregungswert des Agenten negativ wird, da in diesem Fall der Grad der Langeweile z_t nicht (in negativer Richtung) zunehmen kann. Anders ausgedrückt kann das Arousal des Agenten nur im Zustand relativer Ausgeglichenheit (in Bezug auf Emotion und Stimmung) einen negativen Wert annehmen. Hierbei steht die Vermutung im Hintergrund, dass das Fehlen von neuen bzw. für den Agenten relevanten Stimuli nicht nur zu „Langeweile“ sondern auch zu „Schläfrigkeit“ führt.
- Durch die **Funktion d(t)** wird der Dominance-Wert zu jedem Zeitpunkt bestimmt und wie anhand des Arguments der Funktion ersichtlich, hat die Dynamiksimulation auf diesen Wert keinen Einfluss. Welche Faktoren diesen Wert bestimmen könnten, wird im nächsten

Kapitel im Zusammenhang mit der Darstellung des Informationsflusses näher erläutert.

Die Abbildung $K(x_t, y_t, z_t, t)$ ist **weder injektiv noch surjektiv**. Injektivität würde voraussetzen, dass jedem Punkt des PAD-Raumes *höchstens* ein Punkt im Dynamik-Raum entspricht. Leider können sich aber die Variablen x_t und y_t in der Funktion $p(x_t, y_t)$ beliebig oft gegeneinander aufheben, weshalb entlang beliebiger Geraden mit der Steigung -1 im Dynamik-Raum auf den gleichen Punkt im PAD-Raum abgebildet wird. Außerdem führt die Betragsfunktion in $a(x_t, z_t)$ zu einer Verletzung der Voraussetzung für Injektivität. Surjektivität würde bedeuten, dass einem Punkt des PAD-Raumes *mindestens* ein Punkt des Dynamik-Raumes zugeordnet würde. Das PAD-Tripel (p=1, a=0, d) ist aber zum Beispiel nicht in den Dynamik-Raum abzubilden, wie leicht nachvollziehbar ist.

Diese Eigenschaften der Abbildung lassen es nicht zu, eine allgemein gültige **inverse Abbildung K^{-1}** anzugeben. Dies ist für die Funktion des Emotionssystems so lange unproblematisch, wie nicht die Notwendigkeit besteht, aus dem PAD-Tripel einer im PAD-Raum repräsentierten Emotionskategorie einen Referenzpunkt im Dynamik-Raum bestimmen zu wollen. Aus konzeptioneller Sicht sind die Eigenschaften der Abbildung K jedoch kritisch zu bewerten. Wenn der Wertebereich des PAD-Raumes in allen drei Achsen im Sinne der psychologischen Emotionstheorien voll ausgeschöpft würde, dann würde die fehlende Surjektivität der Abbildung K ernst zu nehmende Konsequenzen für die Allgemeingültigkeit dieses Emotionssystems haben. Aufgrund der in der psychologischen Literatur vorhandenen Beispiele für empirisch ermittelte PAD-Tripel liegt allerdings die Vermutung nahe, dass der Wertebereich nicht in derart kritischem Maße ausgeschöpft wird. Da es außerdem im Rahmen einer wirkungsvollen Simulation weniger auf psychologische Adäquatheit ankommt, kann in dem hier konzipierten System die räumliche Lage der Emotionskategorien im PAD-Raum entsprechend den äußeren Anforderungen gewählt werden. Somit wird vermieden, dass Emotionskategorien vom Dynamik-Raum aus nicht erreicht werden können und es kann in Richtung einer visuell nachvollziehbaren und sinnvoll erscheinenden Emotionsdynamik gearbeitet werden.

Zusätzlich zu der Abbildung K ist es notwendig, **emotionale Kategorien im PAD-Raum**

„zu verankern“. Dazu wird für jede Emotionskategorie (bzw. eine sie repräsentierende Emotion) ein **Referenzpunkt** angegeben, um den herum die Intensität der Emotion maximal ist. Die Distanz, die für eine maximale Intensität der Emotion noch toleriert wird, wird durch den **Sättigungsschwellwert** Δ angegeben. Durch den **Aktivierungsschwellwert** Φ wird die maximale Entfernung zu einem Referenzpunkt bestimmt, bei deren Unterschreitung die Emotion mit geringer Intensität aktiviert wird. Zwischen den Schwellwerten Δ und Φ ist ein **stetiger, linearer Übergang** der Intensität vom Minimum bei Φ zum Maximum bei Δ denkbar. Dadurch würde der durch die Dynamik generierte zeitliche Verlauf der Emotionen nicht nur über die Abfolge verschiedener Emotionen zum Ausdruck kommen, sondern sich auch in dem zeitlichen Verlauf der Intensität jeder einzelnen Emotion widerspiegeln (siehe Abbildung 18).

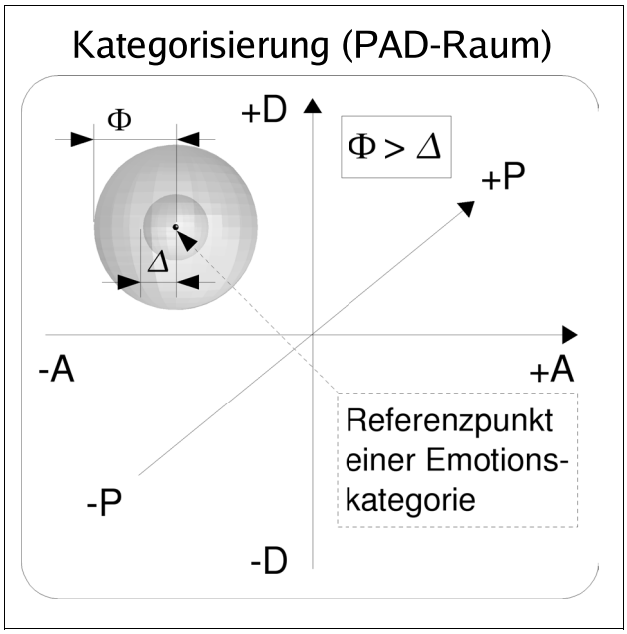


Abbildung 18: Schematische Darstellung des Aktivierungs- und Sättigungsschwellwertes

Da es sich bei Δ und Φ um **globale Schwellwerte** handelt, die für jede Emotionskategorie gültig sind, ist es prinzipiell möglich, dass mehrere Emotionkategorien gleichzeitig eine von Null verschiedene Intensität besitzen. Auch wenn man für jede Emotionskategorie eigene Schwellwerte vorsehen würde, könnte dieses Problem auftreten. Dadurch würde nur die Anzahl der „Stellschrauben“ erhöht werden, weshalb hier lediglich zwei globale Schwellwerte vorgesehen sind. Im Fall der **gleichzeitigen Aktivierung mehrerer Emotionskategorien** wird die **„Winner-takes-all“-Strategie** vorgeschlagen. Die Emotionskategorie mit dem größten Intensitätswert „gewinnt“ in dem Sinne, dass sie alle anderen Emotionskategorien unterdrückt und ihre Intensität als einzige Information nach außen sichtbar ist.

Es kann Bereiche im PAD-Raum geben, in denen **keine Emotionskategorie** aktiviert wird.

Denkbar ist dies zum Beispiel im Fall positiver Stimmung zusammen mit negativer Emotion im Dynamik-Raum. Diese Konstellation würde zu dem PAD-Tripel

$$K_{PAD}(x_t=-1,y_t=1,z_t=0,t) = \langle p(-1,1),a(-1,0),d(t)\rangle = \langle 0,1,d(t)\rangle$$

führen. Dies entspräche neutraler Lust bei maximaler Erregung und ist ebenso schwer zu interpretieren wie die statische Information einer negativen Emotion im Zustand positiver Stimmung. Eine derartige Konstellation ist vielmehr mit dem psychischen Phänomen der (kurzzeitigen) Verwirrung eines Menschen zu erklären, dem im Zustand „guter Laune“ etwas Negatives widerfährt. Bezogen auf das Emotionssystem wird deshalb angenommen, dass sich der Agent im **Zustand der Verwirrung** befindet, wenn keine Emotionskategorie aktiv ist.

4.4 Die Einbindung in die Gesamtarchitektur

Durch die Pfeile in Abbildung 13 wird die Vernetzung des Emotionssystems mit den anderen Komponenten der Gesamtarchitektur verdeutlicht. Dabei wird zwischen Informationsfluss und Modulation unterschieden. Der Informationsfluss ist in Form eines Kreislaufs realisiert, wobei jedoch die Art der Information, die dem Emotionssystem als Eingabe dient, sich von der Art der Information unterscheidet, die das Emotionssystem der Gesamtarchitektur als Ausgabe zur Verfügung stellt.

1.) Eingabedaten: In das Emotionsystem hinein fließen hauptsächlich Valenzinformationen. Zusätzlich können deliberative Komponenten direkt Einfluss auf die Dominance-Achse nehmen, was durch den schräg nach oben verlaufenden Pfeil angedeutet wird.

2.) Ausgabedaten: Das Emotionssystem stellt den restlichen Komponenten der Gesamtarchitektur folgende Informationen über die aktuelle emotionale Lage des Agenten zur Verfügung:

- Die Valenz der Stimmung und den Grad der Langeweile aus dem Dynamik-Raum.
- Das PAD-Tripel aus dem PAD-Raum.
- Falls vorhanden, die Intensität der aktiven Emotionskategorie aus dem PAD-Raum, sonst „Verwirrung“.

Kasten 3:
Struktur der Ein- und Ausgabedaten des Emotionssystems

4.4.1 Die Eingabedaten des Emotionssystems

Die Valenzinformationen, die als Eingangsdaten für die Dynamiksimulation des Emotionssystems entscheidend sind, können auf unterschiedliche Weise gewonnen werden:

- Die Sensoren des Agenten könnten hart-verdrahtet auf Annäherung eines fremden Objektes mit negativer Valenz reagieren. Sanfte Berührungen hingegen könnten zur Generierung positiver Valenzinformationen führen und ein lautes Geräusch könnte negativen Einfluss auf das Emotionssystem haben. So können auf reaktive Weise Emotionen angeregt werden.
- Die Präsentation eines erwünschten Objektes durch das Gegenüber könnte durch deliberative Prozesse zu der Ausschüttung von positiver Valenz führen. Verhielte sich das Gegenüber jedoch nicht kooperativ, so könnte dies negative Auswirkungen auf das Emotionssystem zur Folge haben. Auch selbstbezügliche Gründe für Emotionen könnten modelliert werden, indem z.B. das Erreichen eines (Teil-)Zieles zur Generierung positiver Valenz führen würde.

Von außen eintreffende Valenzinformationen werden zu der aktuellen in der Dynamik-

simulation vorhandenen „Valenz der Emotionen“ hinzu addiert. Sollte dadurch der Wertebereich verlassen werden, so wird der Wert auf das Maximum resp. das Minimum des Wertebereichs gerundet. Da das Emotionssystem als eigenständiger Prozess realisiert ist, führt eine interne Taktung zu ständigen Veränderungen der Werte, wie durch die oben beschriebene interne Dynamik vorgegeben. Dabei ergeben sich zwei positive Eigenschaften:

1.) Es wird vermieden, dass Max allzu sprunghaft auf die einströmenden Valenzinformationen reagiert, da diese nur die Valenz der Emotionen beeinflussen können. Durch die interne Verknüpfung der Emotionen mit der Stimmung kann letztere nur indirekt modifiziert werden und wirkt dabei wie ein emotionales Gedächtnis.

2.) Kurz nacheinander eintreffende, gegensätzliche Valenzinformationen heben sich gegenseitig auf. Würde zum Beispiel eine positive emotionale Valenz mit der Intensität +0.5 im Emotionssystem eintreffen, so würde die „Valenz der Emotionen“ unverzüglich um 0.5 ansteigen. Sollte kurz danach, also noch bevor die simulierte Dynamik den Wert wesentlich verändert hat, dem Emotionssystem eine negative emotionale Valenz von -0.5 mitgeteilt werden, so würde die „Valenz der Emotionen“ lediglich wieder auf den Anfangswert zurückgesetzt werden. In anderen Systemen würde die Valenz sprunghaft das Vorzeichen wechseln, da ihr Wert überschrieben würde.

Durch die im vorigen Kapitel beschriebene Abbildung $K(x_t, y_t, z_t, t)$ bleibt die Dominance-Achse des PAD-Raumes unbeeinflusst. Deshalb müssen die anderen Komponenten der Gesamtarchitektur die Möglichkeit haben, den Dominance-Wert direkt zu verändern. Zum Beispiel könnte die Information darüber, wer in der Gesprächssituation den Turn besitzt (also gerade die Rolle des Sprechers einnimmt), verwendet werden, um den Dominance-Wert entsprechend zu setzen. Sollte MAX den Turn besitzen, so ist er dominant, andernfalls unterwürfig. Andere Heuristiken wären ebenfalls denkbar und dieser Punkt wird später erneut aufgegriffen, wenn es um die Vereinbarkeit dieses Emotionssystems mit einer kognitiven, einschätzungsbasierten Emotionskomponente gehen wird.

4.4.2 Die Ausgabedaten des Emotionssystems

Der Zustand des Emotionssystems kann auf verschiedene Weise beschrieben werden. Die abstrakteste Information wird über die Intensität der Emotionskategorie bzw. den Zustand der „Verwirrung“, falls keine Emotionskategorie aktiv ist, nach außen sichtbar. Für manche Komponenten der Gesamtarchitektur könnten allerdings auch die PAD-Werte selbst von Nutzen sein, die zu einem Zeitpunkt vorliegen. Wieder andere Komponenten könnten lediglich an der aktuellen Stimmungslage des Agenten oder dem Grad seiner Langeweile interessiert sein. Da noch nicht abzusehen ist, auf welche Informationsform eventuell verzichtet werden kann, werden den anderen Komponenten vorerst alle Informationsformen zur Verfügung gestellt.

Die BDI-Komponente könnte die bereitgestellten Informationen über die Emotionen des Agenten als Beliefs repräsentieren und in die Vorbedingungen ihrer Pläne integrieren. Damit wären die intentionalen Handlungen des Agenten „emotional gefärbt“ und dem Agenten wäre sein emotionaler Zustand bewusst. Zusätzlich könnte auf reaktivem Weg das Emotionssystem unbewusst die Verhaltensweisen und die Wahrnehmung direkt modulieren, wie im nächsten Abschnitt dargestellt.

4.4.3 Die modulierenden Auswirkungen des Emotionssystems

Im Gegensatz zum Informationsfluss werden bei der Modulation lediglich außerhalb des Emotionssystems vorhandene Datenstrukturen und Simulationsparameter modifiziert. Es werden hierbei also keine Informationen in dem Sinn übermittelt, dass sie vom Empfänger gespeichert würden, sondern die hier übermittelten Informationen dienen lediglich der Modulation vorhandener Kontrollmechanismen. Es geht hier also nicht um die Frage, was MAX tut, sondern wie er etwas tut. Dass MAX blinzelt und atmet, wird nicht vom Emotionssystem angeregt, sondern die Art, wie er dies tut, wird durch seine Emotionen moduliert. Ebenso wird nicht der Wahrnehmungsprozess selbst durch das Emotionssystem gesteuert, sondern Parameter der Wahrnehmung, wie die Genauigkeit der Sensorauswertung, werden durch die

Emotionen moduliert. Sicherlich wird dies technisch so realisiert, dass vom Emotionssystem die gleichen Informationen ausgesendet werden, wie im Fall des Informationsflusses, doch die empfangende Komponente wird diese Informationen auf konzeptionell andere – nämlich modulierende – Weise verwenden. Erste Versuche zur Modulation der reaktiven Verhaltensweisen werden in Abschnitt 5.3 beschrieben und die dafür vorgenommenen Zuordnungen der Emotionskategorien zu den Modulationseffekten in Tabelle 12 auf Seite 88 zusammengefasst.

4.5 Zusammenfassung

Das hier vorgestellte Emotionssystem zur Simulation der Emotionsdynamik eines künstlichen humanoiden Agenten wurde unter Berücksichtigung der durch MAX vorgegebenen Rahmenbedingungen konzipiert. Dennoch ist es abstrakt genug, um als allgemeiner Ansatz zur Simulation einer Emotionsdynamik auch anderer künstlicher humanoider Agenten genutzt werden zu können. Im Einzelnen:

1.) Das System ist in sich abgeschlossen.

Es besitzt eine Eingabeschnittstelle, die auf sehr intuitive Weise durch andere Komponenten des Agenten genutzt werden kann. Da nur emotionale Valenzinformationen – also Intensitätswerte der Emotionen innerhalb des geschlossenen Intervalls [-1, 1] – als Eingabedaten benötigt werden, ist es auch den reaktiven Komponenten, die keinen Zugriff auf Reasoning-Mechanismen haben, prinzipiell möglich, die Emotionen des Agenten zu beeinflussen. Die Ausgabeschnittstelle hingegen stellt Informationen über den emotionalen Zustand des Agenten auf unterschiedlichem Abstraktionsniveau bereit, wodurch sowohl reaktive als auch deliberative Komponenten adäquat bedient werden.

2.) Das System ist technisch unabhängig von anderen Komponenten.

Es benötigt zwar Eingabedaten wie unter Punkt eins erwähnt, doch da es technisch als eigenständiger Prozess konzipiert wurde, besitzt es seine eigene interne Taktung und kann sogar – entsprechende Kommunikationsschnittstellen vorausgesetzt – zur Laufzeit zu einem Agenten hinzugeschaltet werden. Sollten die Sensoren und die Reasoning-

Komponenten ausfallen, so würde dem Agenten einfach langweilig werden, da das Emotionssystem keine Eingabedaten mehr erhielte.

3.) Das System ist (sogar zur Laufzeit) in allen Details konfigurierbar.

Alle entscheidenden Parameter sowohl der Dynamiksimulation als auch der Kategorisierung sind zur Laufzeit veränderbar und abzuspeichern. Dadurch lässt es sich den Anforderungen des jeweiligen Agenten, für den es eingesetzt wird, flexibel anpassen.

5 Implementation und erste Demonstration

Nachdem der Aufbau des Emotionssystems im vorangegangenen Kapitel konzipiert wurde, wird in diesem Kapitel eine beispielhafte Implementation des Emotionssystems als Komponente der kognitiv motivierten Gesamtarchitektur des in der AG Wissenbasierte Systeme entwickelten virtuellen humanoiden Agenten MAX (vgl. Kapitel 4.1) vorgestellt. Dabei wird zum einen die technische Umsetzung des Software-Agenten beschrieben, durch den das Emotionssystem realisiert wird, und zum anderen eine Reihe denkbarer Kommunikationsstrukturen innerhalb der Gesamtarchitektur vorgeschlagen, von denen einige für die erste Demonstration implementiert wurden.

5.1 Entwicklungsumgebung

Da das Emotionssystem als WBS-Agent innerhalb des WBS-Agentensystems realisiert wurde und die Basisklassen des verwendeten Agenten in C++ geschrieben wurden, war C++ die Grundlage der Implementierung. Um dem Anspruch der detaillierten Visualisierung sowohl der veränderlichen Parameter als auch der Eigendynamik des Systems gerecht zu werden, wird Qt 3 [QT03] in Kombination mit OpenGL [OGL03] verwendet. Mit Hilfe des *Qt Developers*, einem Werkzeug zur Gestaltung auf Qt basierender Benutzeroberflächen, wird die grafische Benutzerschnittstelle (grafical user interface, GUI) des Agenten entworfen, die als zentrales Element zur Visualisierung der dynamischen Daten ein *OpenGLWidget* enthält, in dem über OpenGL-Kommandos eine dreidimensionale Darstellung der beiden Komponenten des Emotionssystems erreicht wird. Die vorhandenen Kommunikationsstrukturen des WBS-Agentensystems werden ausgenutzt, um die Ein- und Ausgabeschnittstellen auf flexible Weise zu realisieren.

5.2 Umsetzung

In den folgenden Abschnitten werden zunächst die Klassenstruktur und die implementierten Emotionskategorien beschrieben. Danach wird anhand von Bildschirmfotos der grafischen Benutzeroberfläche zuerst auf die Umsetzung des Dynamik-Raumes und anschließend auf die Umsetzung des PAD-Raumes eingegangen.

5.2.1 Die Klassenstruktur und die Emotionskategorien

Der *AC_Qt3EmotionsAgent* wird von einem *ActiveQt3Agent* abgeleitet und enthält als private Komponente einen Pointer auf ein Objekt des Typs *AC_Qt3EmoGui*. Dadurch wird die Agentenkommunikation von der grafischen Benutzerschnittstelle und der Simulation getrennt. Die Berechnungen zur Simulation der Dynamik werden durch einen *QTimer* alle 20 Millisekunden innerhalb von *AC_Qt3EmoGui* angestoßen. Die physikalischen Daten zur iterativen Berechnung der Bewegung im Dynamik-Raum werden in der Klasse *AC_EmotionsContainer* gekapselt, von der eine Instanz vorhanden ist, auf die über einen Pointer von der Klasse *AC_Qt3EmoGui* aus zugegriffen werden kann. Diese Klasse besitzt ebenfalls einen Pointer auf ein Objekt des Typs *AC_Qt3EmotionsAgentOpenGLGui*, welches von der Klasse *QGLWidget* abgeleitet ist und die dreidimensionale Visualisierung übernimmt. Von der abstrakten Basisklasse *AC_EmotionsConverter* wird die konkrete Klasse *AC_EmotionsConverterPAD* abgeleitet, in der eine Funktion zur Konvertierung der Daten des Dynamik-Raumes in das PAD-Tripel des PAD-Raumes, wie in Formel 5 auf Seite 71 konzipiert, implementiert wird. Diese Klasse übernimmt auch die Verortung von 10 Emotionskategorien, von denen 3 doppelt vorhanden sind. Tabelle 12 listet die Emotionskategorien auf, die in einer zentralen Konfigurationsdatei („init.emo_pad“) festlegbar sind:

Emotionskategorie	***1. PAD-Tripel***	***(2. PAD-Tripel)***
ausgeglichen	(0, 0, 0)	
aufmerksam	(0, 0, 1)	(0, 0, -1)
gelangweilt	(0,-.85, 1)	
deprimiert	(0,-.85, -1)	
überrascht	(.80, .80, 1)	(.80, .80, -1)
erfreut	(.50, 0, 1)	(.50, 0, -1)
wütend	(-.80, .80, 1)	
ängstlich	(-.80, .80, -1)	
missgelaunt	(-.50, 0, 1)	
traurig	(-.50, 0, -1)	

Tabelle 12: Die 10 Emotionskategorien des PAD-Raumes

Sollte es nach empirischen Untersuchungen sinnvoll erscheinen, auch bezüglich der Emotionskategorien „erfreut“, „überrascht“ und „ausgeglichen“ eine Unterscheidung hinsichtlich der Dominanz vorzunehmen, so würde es ausreichen, die Initialisierungsdatei entsprechend abzuändern. Ebenso existiert eine Initialisierungsdatei „init.emo_dyn“, die zur Initialisierung der Parameter des Dynamik-Raumes bei Programmstart geladen wird. Die Simulationsparameter können jederzeit in Textdateien gesichert werden, wodurch es möglich wird, mit verschiedenen „Persönlichkeitseigenschaften“ des Agenten, wie sie sich in den dynamischen Parametern der Simulation widerspiegeln, zu experimentieren.

5.2.2 Der Dynamik-Raum

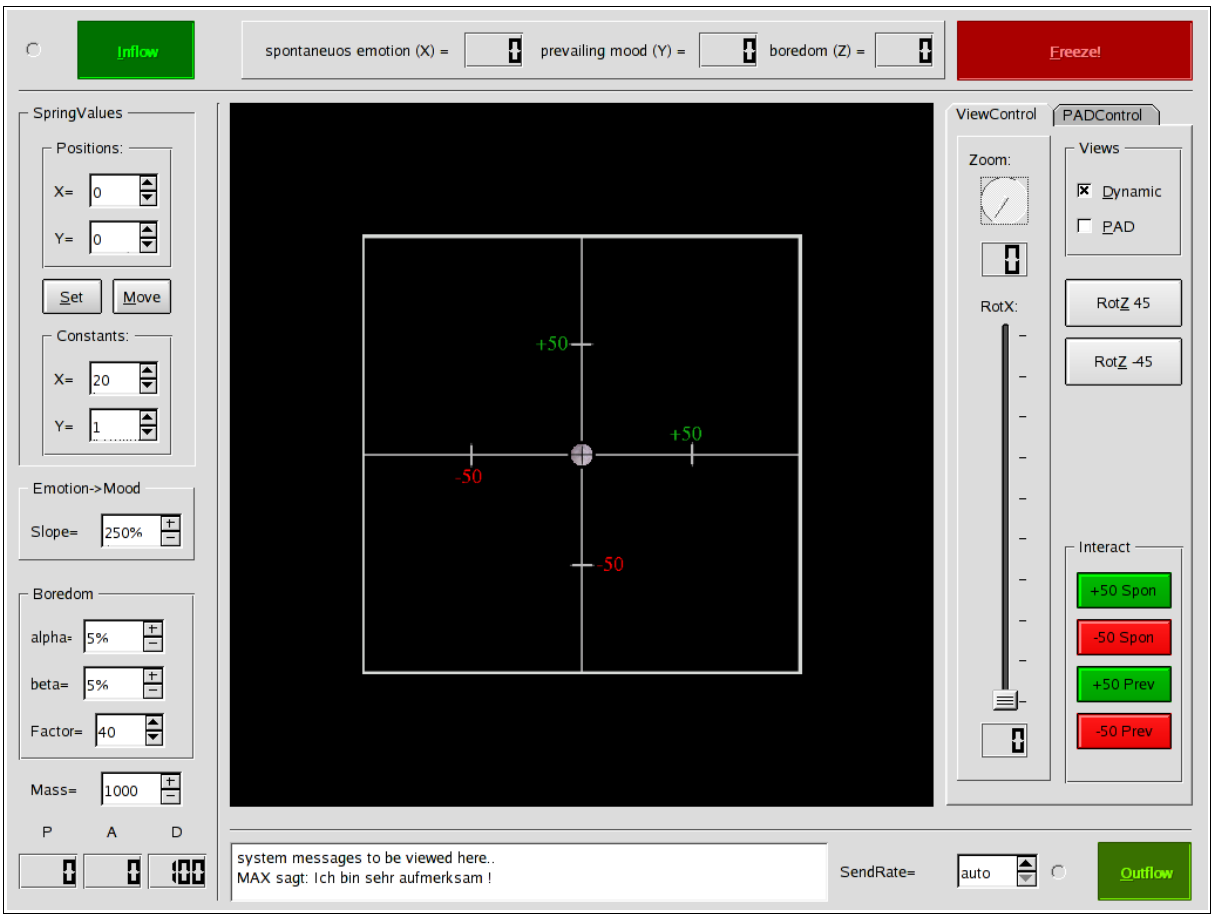


Abbildung 19: Die Darstellung des Dynamik-Raumes.

Abbildung 19 zeigt die grafische Benutzerschnittstelle (GUI) des Agenten direkt nach Programmstart. Im Zentrum der GUI befindet sich die dreidimensionale Darstellung des Dynamik-Raumes, dessen Ansicht über die rechts daneben vorhandenen Kontrollelemente verändert werden kann. Oberhalb dieser dreidimensionalen Darstellung sind die aktuellen Werte des Dynamik-Raumes ablesbar und es kann der Eingabestrom durch Deaktivierung des Knopfes *Inflow* unterbunden werden sowie die Simulation über den Knopf *Freeze!* eingefroren bzw. angehalten werden.

Der linke Bereich der GUI teilt sich in fünf Gruppen auf:

1.) *SpringValues*: Im oberen Teil dieser Gruppe können zu Testzwecken die Positionen des Referenzpunktes entweder absolut (über *Set*) oder relativ (über *Move*) gesetzt werden. Im unteren Teil können die physikalischen Konstanten d_x (X) und d_y (Y) der beiden simulierten Spiralfedern eingestellt werden (vgl. Formel 3 und Abbildung 16, S. 68f.). Dabei wird auf die Angabe von Einheiten verzichtet, da es lediglich auf den relativen Unterschied beider Konstanten ankommt.

2.) *Emotion->Mood*: Hier ist die Steigung m (engl. *Slope*) der Formel 1 (S. 66) einstellbar.

3.) *Boredom*: In dieser Gruppe sind die Parameter der „Langeweile“ veränderbar. Durch *alpha* (ϵ_x) und *beta* (ϵ_y) wird die Region um den Nullpunkt des Dynamik-Raumes angegeben (vgl. Abbildung 16, S. 70) und mit *Factor* die Geschwindigkeit, mit der dem Agenten langweilig wird.

4.) *Mass*: Mit dieser *QSpinBox* kann die Masse m des Referenzpunktes im Dynamik-Raum (die weiße Kugel im Zentrum von Abbildung 19) angegeben werden (vgl. Formel 3 und Abbildung 16, S. 68f.).

5.) *P A D*: Hier werden die aktuellen Werte des (P)leasure-(A)rousal-(D)ominance-Raumes angezeigt.

Am unteren Rand der GUI erscheinen *system messages* in Form von „MAX sagt: ...“-Meldungen, die eine linguistische Umschreibung des emotionalen Zustandes von MAX darstellen. Dadurch ist zusammen mit den *Interact*-Knöpfen, die aus Bequemlichkeit implementiert wurden, eine erste Kontrolle der Funktionsweise des Emotionssystems auch ohne andere Komponenten der Gesamtarchitektur möglich. In der mit „sendrate“ beschrifteten *QSpinBox* lässt sich die Anzahl der verschickten Nachrichten pro Sekunde einstellen; ist dieser Wert auf *auto* gesetzt, so wird nur bei Schwellwertüberschreitung (vgl. Abbildung 18, S. 73) eine Nachricht verschickt. Durch Deaktivierung des Knopfes *Outflow* kann das Verschicken von Nachrichten unterbunden werden.

5.2.3 Der PAD-Raum

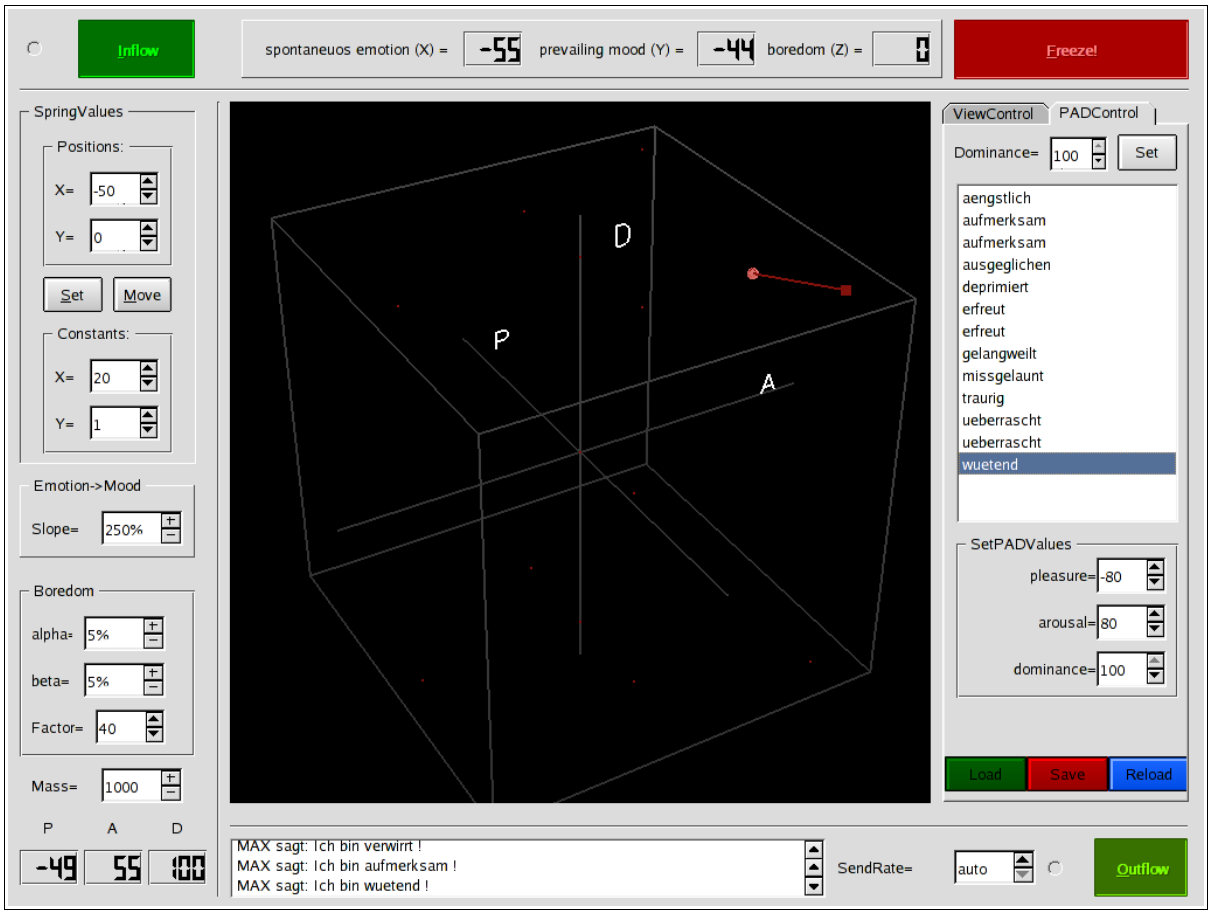


Abbildung 20: Die Darstellung des PAD-Raumes.

Abbildung 20 zeigt die dreidimensionale Darstellung des PAD-Raumes und auf der rechten Seite die zweite Registerkarte *PADControl*, über dessen Elemente man die in Tabelle 12 aufgelisteten Emotionskategorien anwählen und bezüglich ihrer PAD-Werte verändern kann. Am unteren Rand der Registerkarte befinden sich drei Knöpfe zum Laden, Speichern und erneuten Laden der in Abschnitt 5.2.1 auf Seite 82 erwähnten Konfigurationsdateien. Am oberen Rand der Registerkarte befindet sich eine mit *Dominance* beschriftete *QSpinBox*, deren Wert über den *Set*-Knopf dem System mitgeteilt werden kann. Eine mit dem Referenzpunkt des PAD-Raumes (rote Kugel) verbundene Spiralfeder wird dann in dem zu erreichenden Dominanzwert verankert und ihre Rückstellkraft wirkt so lange auf den Referenzpunkt ein, bis der

gewünschte Dominanzwert vom System angenommen wurde. Dadurch werden sprunghafte Veränderungen vermieden und die Stetigkeit bleibt auch in dieser unabhängigen Dimension erhalten.

Die Entfernung zu einer evtl. aktivierten Emotionskategorie wird in der Darstellung des PAD-Raumes durch die in Abbildung 20 zu erkennende rote Verbindungslinie visualisiert. Zusammen mit der Hervorhebung der auf der rechten Seite angewählten Emotionskategorie in der Darstellung des PAD-Raumes wird somit eine visuelle Kontrolle der Abfolge unterschiedlicher Emotionskategorien erleichtert.

5.2.4 Kommunikationsstrukturen

Als **Eingabeschnittstelle** des Agenten wird die *MaxEmotionMessage* verwendet, in der über einen Aufzählungtypen symbolisch angegeben werden kann, welche Aktion mit den übermittelten Daten vom Sender beabsichtigt ist. Bisher vorgesehen wurden:

- *Undefined*, falls vergessen wurde, den Wert zu setzen.
- *Setspontanous*, um den Wert der emotionalen Valenz beeinflussen zu können. Die „Emotionsachse“ wurde in der Implementierung zur besseren Unterscheidung der Zeitlichkeit mit „spontaner Emotion“ (engl. *spontaneous emotion*) bezeichnet.
- *Setprevailing*, um Einfluss auf die Valenz der Stimmung nehmen zu können. Auch hier wurde der Begriff „Grundstimmung“ (engl. *prevailing mood*) im Kontrast zu „spontaner Emotion“ (s.o.) gewählt. Die Valenz von außen direkt verändern zu können, ist für spätere Ergänzungen vorgesehen (siehe Diskussion und Ausblick).
- *Setdominance*, um die Dominanz verändern zu können.
- *Setdynamic*, um von außen die Simulationsparameter der Emotionsdynamik setzen zu können.
- *Setpad*, um eine neue Menge von Emotionskategorien angeben zu können.
- *Getdata*, um dem Agenten mitzuteilen, dass er dem Sender die Simulationsdaten (in Form einer MaxEmotionMessage) mitteilen soll.
- *Reset*, um den Agenten anzuweisen, erneut den Startzustand einzunehmen.

Für die **Ausgabeschnittstelle** wird eine *MaxEmotionMessage* mit der Aktion *Undefined* verschickt, durch die die aktuellen Daten des Emotionssystems den restlichen Komponenten des Gesamtsystems bekannt gegeben werden. Diese Nachricht wird über den Verteiler *MaxEmoData* verschickt, auf dem sich alle anderen Komponenten eintragen können, wenn sie an Informationen über den emotionalen Zustand des Agenten interessiert sind.

5.3 Demonstration

Für eine erste Demonstration des Emotionssystems werden die reaktiven Komponenten von MAX durch Funktionen ergänzt, die es dem Emotionssystem ermöglichen, ausgewählte visuelle Aspekte der durch Emotionen verursachten physiologischen Veränderungen zu simulieren. Es werden die in Abschnitt 4.1.1 beschriebenen acht emotionalen Gesichtsposturen indirekt durch das Emotionssystem angesteuert, indem die bei der reaktiven Komponente eintreffenden *MaxEmotionMessages* hinsichtlich der Emotionskategorie ausgewertet und auf die Gesichtsposturen abgebildet werden. Die in Kasten 2 (S. 55) angegebenen globalen Parameter der Sprechstimme werden in Anlehnung an die in Tabelle 3 (S. 25) vorgestellten Emotionseffekte auf die Stimme moduliert (siehe Tabelle 13).

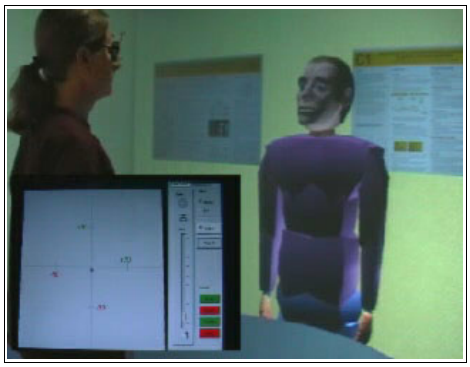
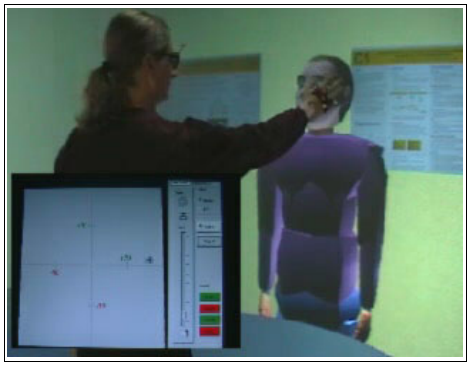
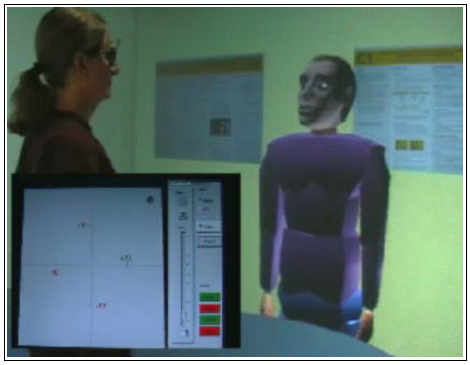
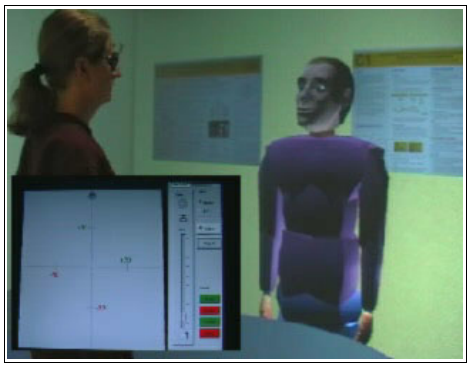
Emotionskategorie	***Atmung***	***Blinzeln***	***Gesichtsröte***	***Gesichtspostur***	***Frequ., Rate, Mod.***
ausgeglichen	+/-	+/-	0	neutral	(-), (+/-), (+/-)
aufmerksam	+/-	+/-	0	konzentriert	(-), (+/-), (+/-)
gelangweilt	--	--	0	gelangweilt	(--), (--), (--)
deprimiert	--	--	0	traurig	(-), (--), (--)
überrascht	+	+	0	überrascht	(-), (+), (+/-)
erfreut	-	-	+ (Wangen)	freundlich	(-), (-), (+)
wütend	++	++	+ (Gesicht)	verärgert	(-), (++), (+/-)
ängstlich	++	++	+ (Gesicht)	verärgert	(-), (++), (+/-)
missgelaunt	-	-	+ (Gesicht)	traurig	(-), (-), (--)
traurig	--	--	0	traurig	(-), (--), (--)

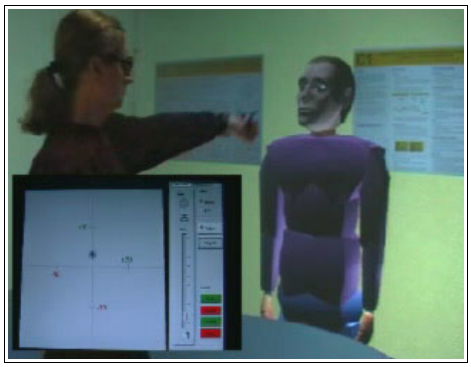
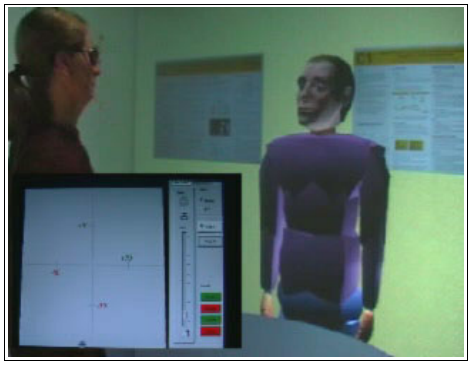
Tabelle 13: Die Emotionskategorien und ihr Einfluss auf MAX

Atmungs- und Blinzelfrequenz sowie Sprechrate werden auf qualitativ gleiche Weise durch die Emotionskategorien moduliert und bei der Gesichtsröte wird zwischen einem Einfärben der Wangen bei der positiven Emotion „erfreut“ und dem Einfärben des gesamten Gesichtes bei den drei negativen Emotionen „wütend“, „ängstlich“ und „missgelaunt“ unterschieden. Die von Max eingenommene Gesichtspostur alleine ließe zum Beispiel keine Unterscheidung zwischen „traurig“ und „missgelaunt“ zu, doch durch Hinzunahme der drei physiologischen Parameter wird eine Unterscheidung möglich.

Die folgenden Bildschirmfotos wurden einem Video entnommen, das während einer Demonstration (am 25.02.2003) in der CAVE-ähnlichen Installation der AG WBS aufgenommen wurde. Wenn der Benutzer seine rechte Hand in die Nähe des Kopfes von MAX hält, so wird dies von MAX als „Streicheln“ interpretiert und er beginnt, *MaxEmotionMessages* mit der *action* s*etspontaneous +10* an das Emotionsmodul zu senden. Wird die Hand wieder weggenommen, so hört auch die Ausschüttung positiver Valenz durch MAX auf. Den gegenteiligen Effekt hat die linke Hand des Benutzers, wenn sie nahe an den Kopf von MAX geführt wird. Dies wird als „Schlagen“ von MAX interpretiert und führt, so lange die Hand in der Nähe des Kopfes bleibt, zur wiederholten Ausschüttung negativer (emotionaler) Valenz (*setspontaneous -10*). MAX bleibt bei dieser ersten Demonstration immer dominant (*Dominance* = 1), da erstens in diesem Szenario nicht genug Informationen für die Bestimmung der Dominanz vorhanden sind und zweitens diese erste Demonstration vorrangig als Test der Emotionsdynamik gedacht war, die sich durch die Berechnungen innerhalb des Dynamik-Raumes und die anschließende Abbildung in den PAD-Raum ergibt.

Die Bildschirmfotos wurden anhand markanter Merkmale ausgewählt, die jeweils nebenstehend erläutert werden. Das Video, aus dem die Bildschirmfotos stammen, befindet sich auf der beigefügten CD (im mpg1-Format).

	Während der gesamten Demonstration verfolgt MAX den Benutzer, indem er ihm in die Augen schaut. Hier ist MAX aufmerksam, da nicht genug Zeit verstrichen ist, damit ihm hätte langweilig werden können. Unten links ist die grafische Benutzerschnittstelle des Emotionssystems eingeblendet.
	Durch das „Streicheln“ des Benutzers wird positive emotionale Valenz ausgeschüttet, die innerhalb des Dynamik-Raumes zu einem Ausschlag nach rechts führt.
	Durch die positive emotionale Valenz steigt die Stimmung von MAX stark an und die Kombination aus positiver Stimmung und positiver emotionaler Valenz führt zu „Überraschung“, die sich im Gesichtsausdruck von MAX widerspiegelt.
	Wird MAX nicht weiter emotional angeregt, so erreicht er aufgrund der unterschiedlichen Federkonstanten (die Dynamikeinstellungen entsprechen Abbildung 19) seine emotionale Ausgeglichenheit schneller als seine neutrale Stimmungslage. Er ist „erfreut“.

	Nachdem MAX in eine beinahe neutrale Stimmungslage zurückgekehrt ist, beginnt der Benutzer, ihn zu schlagen.
	Das „Schlagen“ führt zur Ausschüttung von negativer emotionaler Valenz und durch diese ist die Stimmung von MAX ins Negative gesunken. Die Emotionskategorie „wütend“ wird aktiv, wodurch sich das Gesicht von MAX rötlich färbt und er den entsprechenden Gesichtsausdruck zeigt.
	Nachdem die emotionale Valenz beinahe wieder die Neutralität erreicht hat, bedingt seine extrem negative Stimmungslage, das MAX „missgelaunt“ ist, was sich an seinem Gesicht ablesen lässt.
	Später wird MAX zuerst positiv angeregt, jedoch kurz darauf „geschlagen“. Dadurch wird im Dynamik-Raum ein Punkt im oberen linken Quadranten erreicht, der bei positiver Stimmung eine negative emotionale Valenz repräsentiert. Wie in Abschnitt 4.3 auf Seite 74 beschrieben, führt diese Konfiguration bei MAX zum Zustand der „Verwirrung“.

5.4 Zusammenfassung

Bei der Implementation des Emotionssystems in Form eines WBS-Agenten wurden folgende Punkte berücksichtigt:

- Die Anforderungen an die grafische Benutzerschnittstelle wurden durch die Kombination zweidimensionaler Elemente zur manuellen Kontrolle aller Simulationsparameter auf der Basis von Qt3 und einer zentralen Komponente zur dreidimensionalen Visualisierung der Eigendynamik des Emotionssystem mittels OpenGL und objektorientierter Programmierung in C++ erfüllt.
- Die Eigendynamik des Emotionssystems wurde erfolgreich umgesetzt und zeichnet sich durch ein stabiles Laufzeitverhalten aus.
- Die Kommunikationsstrukturen wurden entsprechend der Konzeption in Form einer eigenen Nachrichtenklasse mit entsprechenden Datenstrukturen angelegt und werden den anderen Komponenten der Gesamtarchitektur über einen Verteiler bereitgestellt.
- In einer ersten Demonstration konnten allgemein überzeugende Resultate erzielt werden. Die Anregung des Emotionssystems wurde dabei durch die Simulation taktilen Feedbacks realisiert und die Effekte des Emotionssystems spiegelten sich in der Gesichtsmimik und der Modulation der Atem- und Blinzelfrequenz sowie der Gesichtsrötung wider.

Trotz dieser ersten überzeugenden Demonstration müssen in Zukunft empirische Untersuchungen zur Evaluierung des Emotionssystems durchgeführt werden.

6 Zusammenfassung und Ausblick

6.1 Zusammenfassung

Um eine „Simulation der Emotionsdynamik eines künstlichen humanoiden Agenten“ zu verwirklichen, wurde als erstes der emotionspsychologische Hintergrund in groben Zügen dargestellt. Dabei wurden hauptsächlich diejenigen Theorien betrachtet, die entweder den Emotionsverlauf – also die Emotionsdynamik – oder den Ausdrucksaspekt der Emotionen schwerpunktmäßig behandelten. Um den Begriff „Emotion“ so weit wie möglich von anderen psychischen Phänomenen abzugrenzen, wurde eine Arbeitsdefinition übernommen, die möglichst unkontrovers gehalten ist. Durch sie wird deutlich, dass im Einzelfall anhand von vier Merkmalen entschieden werden muss, ob ein psychischer Zustand zu der Klasse der Emotionen gehört oder nicht. Zusammen mit den empirischen Untersuchungen über den Effekt von Emotionen auf die Gesichtsmimik, den Arbeiten zur Kategorisierung von emotionalen Begriffen in einem dreidimensionalen Raum und der konzeptionellen Unterscheidung von Stimmungen und Emotionen führte dies zu der Überlegung, dass es zur Simulation einer Emotionsdynamik sinnvoll ist,

1.) Stimmungen in das System zu integrieren. Damit kann der Effekt des „Aufschaukelns der Emotionen“ simuliert und ausgenutzt werden, um die zeitliche Abfolge der Emotionen psychologisch adäquat zu modellieren.
2.) Einen dreidimensionalen Raum zur Kategorisierung zu verwenden. Für diesen Ansatz sprechen umfangreiche empirische Untersuchungen und es ist prinzipiell möglich, beliebige Emotionskategorien in diesem Raum zu verorten.

Um den Entstehungsprozess von Emotionen zu behandeln, wurde eine kognitive Emotionstheorie vorgestellt. Die Erklärungskraft dieser Klasse von Theorien ist groß, jedoch wurde

festgestellt, dass auch der nicht-kognitive Entstehungsweg für Emotionen existiert, der durch diese Theorien nur mit Mühe abgedeckt werden kann.

Im nächsten Kapitel wurden zuerst zwei Implementationen und die dabei auftretenden Probleme der kognitiven Emotionstheorien beschrieben. Es wurde insbesondere herausgestellt, wie schwierig der zeitliche Verlauf der Emotionen in diese Theorien zu integrieren ist, weshalb sie für die vorliegende Arbeit ungeeignet erschienen. Anschließend wurde ein Überblick über fünf weitere Systeme gegeben, die sich in der einen oder anderen Weise mit der Simulation von Emotionen beschäftigen, und es wurden drei Gruppen von anwendungsorientierten Ansätzen in der Emotionsforschung aufgestellt, die sich anhand ihrer zentralen Vorgehensweise und ihrer Zielsetzungen unterscheiden lassen.

Die Konzeption des Emotionssystems wurde im darauf folgenden Kapitel vorgenommen. Dazu wurden zunächst die vorhandenen Arbeiten der AG vorgestellt, in die das Emotionssystem anschließend konzeptionell eingefügt wurde. Dabei wurden Erkenntnisse aus dem Theorieteil und den zuvor dargestellten Systemen ausgenutzt und eigene Ideen zur Modellierung des zeitlichen Verlaufs mit in die Konzeption einbezogen. Zentral war dabei die Idee, die physikalische Simulation zweier Spiralfedern, die Einfluss auf den als Massepunkt aufgefassten Referenzpunkt des Dynamik-Raumes nehmen, zur Generierung des voneinander unabhängigen zeitlichen Verlaufs der Emotion und der Stimmung zu verwenden. Dadurch wurde der ebenfalls simulierte Effekt des „Aufschaukelns der Emotionen“ begünstigt und eine intuitiv nachvollziehbare Emotionsdynamik erreicht. Die Rahmenbedingungen legten es nahe, auch den Aspekt der „Langeweile“ in die Emotionsdynamik zu integrieren. Um den Systemzustand auf psychologisch fundierte Emotionskategorien abbilden zu können, wurde erörtert, wie die Daten anschließend in einen, auf den Theorien von Wundt und Mehrabian basierenden, dreidimensionalen Raum zur Kategorisierung transformiert werden. Schließlich wurde die Struktur der Ein- und Ausgabedaten des Emotionssystems angegeben und der Informationsfluss sowie die modulierenden Effekte des Emotionssystems erklärt.

Kapitel 5 behandelte die Implementation der theoretischen Überlegungen der vorangegangenen Kapitel, wobei ein Schwerpunkt auf die Erklärung der grafischen

Benutzerschnittstelle gelegt wurde, da eine der Anforderungen an das Emotionssystem in einer detaillierten Visualisierung der veränderlichen Simulationsparameter und der Dynamik des Systems bestand. Am Ende des Kapitels wurde eine erste Demonstration beschrieben, in der reaktive Verhaltensweisen und die Gesichtsmimik des Agenten MAX durch die Emotionen moduliert wurden. Zur Generierung von Eingabedaten für das Emotionssystem wurde die Sensorik von MAX erweitert, so dass er körperliche Nähe als positiv bzw. negativ bewertete.

Damit wird das vorgestellte Emotionssystem sowohl den theoretischen Anforderungen gerecht, die sich aus dem psychologischen Hintergrund ergeben, als auch den praktischen Anforderungen, die aus der Integration in ein echtzeitfähiges Gesamtsystem zur Mensch-Maschine-Kommunikation resultieren. Durch die erste Komponente des Emotionssystems, in dem zwischen Emotion und Stimmung unterschieden wird, erhält der Agent bezüglich seiner Emotionen ein konsistentes Langzeitverhalten. Die zweite Komponente ermöglicht die Abbildung auf Emotionskategorien, wie sie unter anderem für die Ansteuerung der Gesichtsmimik psychologisch fundiert sind und wozu sie auch in anderen Systemen bereits verwendet werden. Die Implementierung des Emotionssystems in Form eines nebenläufigen Prozesses (als Software-Agent) mit eigener Taktung, die intuitive Konfigurierbarkeit zur Laufzeit und die gut strukturierten Ein- und Ausgabeschnittstellen ermöglichen außerdem eine schnelle Anpassung an andere Agentenachitekturen, weshalb dieses Emotionssystem im Ergebnis als allgemeingültige Simulation der Emotionsdynamik eines künstlichen humanoiden Agenten gelten kann.

6.2 Ausblick

Das vorgestellte Emotionssystem zur Simulation der Emotionsdynamik eines künstlichen humanoiden Agenten legt seinen Schwerpunkt auf den Teilaspekt der Dynamik der Emotionen. Für die Ansteuerung des Emotionssystems werden lediglich emotionale Valenzen benötigt, deren Entstehungsgrund für das System irrelevant ist. Leider geht damit der

Objektbezug verloren und die sinnvoll im PAD-Raum zu repräsentierenden Emotionskategorien sind auf solche beschränkt, deren Entstehungsprozess weniger von Bedeutung ist. Mit den kognitiven Emotionstheorien können hingegen Emotionskategorien wie „Neid“ und „Scham“ sehr gut erklärt werden, da diese Theorien mit Zugriff auf Wissensbasen den Entstehungsprozess jeder einzelnen Emotionskategorie nachvollziehen. Leider wird dabei die gegenseitige Beeinflussung der Emotionen in ihrem Entstehungsprozess und ihrem zeitlichen Verlauf kaum berücksichtigt, da dies konzeptionell schwierig in die kognitiven Emotionstheorien zu integrieren ist. Doch detaillierte, kognitiv motivierte Modelle zur Berechnung der Intensitäten einzelner Emotionskategorien sind genauso vorhanden wie Informationen über die Qualität (das „Vorzeichen“) jeder einzelnen Emotionskategorie. Deshalb liegt der Gedanke nahe, das in dieser Arbeit vorgestellte Emotionssystem mit einem weiteren, eventuell eigenständigen und nebenläufigen Emotionssystem zu kombinieren, welches auf einer kognitiven Emotionstheorie (wie z.B. der OCC-Theorie) beruht. Dabei könnten beide Systeme voneinander profitieren. Die Valenzinformationen des kognitiven Emotionssystems könnten als Eingabedaten für das hier konzipierte Emotionsdynamiksystem dienen und die Informationen über den zeitlichen Verlauf der Emotionen, die durch die simulierte Emotionsdynamik automatisch generiert werden, könnten ausgenutzt werden, um die Intensität und Dauer der kognitiven Emotionskategorie zu bestimmen.

Zum Beispiel könnte die auf kognitive Weise entstandene Emotionskategorie „Scham“ zum Eintreffen negativer emotionaler Valenz im Emotionsdynamiksystem führen. Wenn die Stimmungslage des Agenten bei Eintreffen dieser Valenzinformation positiv ist, so würde die Stimmung zwar negativ beeinflusst, doch dem kognitiven Emotionssystem würde zurückgemeldet, dass (aufgrund der positiven Stimmungslage) dem Agenten die Scham nicht bewusst wird. Ist die Stimmungslage beim Eintreffen der Valenzinformation hingegen neutral oder negativ, so würde der Agent bewusst Scham empfinden können. Der zeitliche Intensitätsverlauf (und damit gleichzeitig die Dauer) der Scham könnte ebenfalls durch das Emotionsdynamiksystem bestimmt werden. Da die eintreffende negative Valenzinformation zu einem negativen Ausschlagen der emotionalen Valenz führen würde und diese ihrerseits einen

negativen Effekt auf die Stimmung des Agenten hätte, könnte zum Beispiel zu jedem Zeitpunkt die Differenz der Stimmung S_e beim Eintreffen der emotionalen Valenzinformation und der aktuellen Stimmung S_{e+t} so lange als Maß der Intensität gelten, bis diese wieder Null erreicht hat. In Abbildung 21 wird die Bestimmung der Intensität einer kognitiven Emotionskategorie mit Hilfe der Dynamiksimulation im Emotionsdynamiksystem schematisch verdeutlicht.

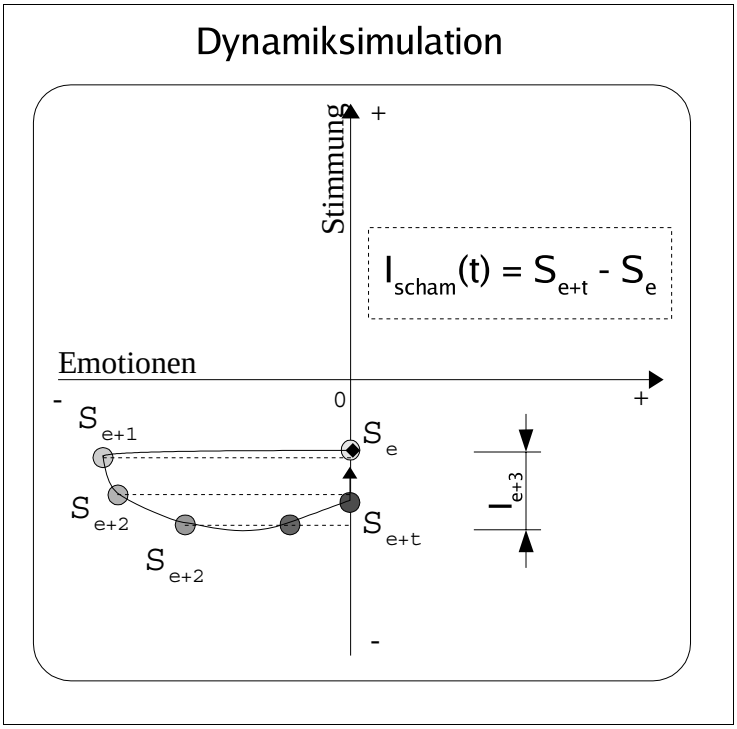


Abbildung 21:
Schematische Darstellung der Bestimmung des Intensitätsverlaufs einer kognitiven Emotionskategorie wie z.B. „Scham“ mit Hilfe der Emotionsdynamik.

Analog könnten die anderen 21 kognitiven Emotionskategorien in das System eingebunden werden, wodurch sich die konzeptionellen Stärken beider Systeme sehr gut ergänzen würden.

Eine zweite Erweiterungsmöglichkeit des vorliegenden Emotionssystems betrifft die Valenz der Stimmung. Sie ist im Emotionsdynamiksystem bisher nur indirekt über den Effekt des „Aufschaukelns der Emotionen“ von außen zu beeinflussen. Mit einer Methode zur direkteren Beeinflussung der Stimmungslage des Agenten wurde bereits experimentiert. Dazu wurde die Verankerung der simulierten Spiralfeder, die für die Rückstellkraft entlang der Stimmungs-

achse (vgl. Abbildung 21) verantwortlich ist, zeitweise in den von außen modifizierten Stimmungswert verschoben. Dadurch wird vermieden, dass die Stimmung ebenso sprunghaften Veränderungen unterliegt, wie dies bei Eintreffen von Valenzinformationen für die emotionale Valenz der Fall ist. Nach Erreichen des gewünschten Stimmungswertes wurde die physikalischen Simulationparameter Geschwindigkeit und Beschleunigung auf Null zurückgesetzt und die Verankerung der Stimmungsachse wieder in den Koordinatenursprung verschoben. Es wurde leider deutlich, dass es ungünstige Fälle gibt, in denen dieses Vorgehen zu unerwünschten Ergebnissen bezüglich des Verlaufs der Stimmung führt, doch prinzipiell sollte das Emotionssystem auch in dieser Hinsicht erweiterbar sein.

Bevor das Emotionssystem allerdings durch die beschriebenen Erweiterungen ergänzt wird, sollten empirische Untersuchungen zur Evaluation durchgeführt werden, um die überzeugenden Ergebnisse der ersten Demonstration statistisch zu untermauern.

Literaturverzeichnis

[ALL02] Allbeck, J. & Badler, N., *Toward Representing Agent Behaviors Modified by Personality and Emotion*, 2002

[BRE98] Breazeal, C., *Infant-like Social Interactions between a Robot and a Human Caretaker*, 1998

[DAM94] Damasio, A., *Descartes' Error: Emotion, Reason, and the Human Brain*, New York, 1994

[DAR72] Darwin, C., *The expression of the emotions in man and animals*, London, 1972

[DÖR99] Dörner, Dietrich, *Bauplan für eine Seele*, 1999

[EFA80] Ekman, P., Friesen, W. V. & Ancoli, S., *Facial signs of emotional experience.*, in Journal of Personality and Social Psychology, 39, S. 1125-1134, 1980

[EKF71] Ekman, P. & Friesen, W. V., *Constants across cultures in the face and emotion.*, Journal of Personality and Social Psychology, 29, S. 124-129, 1971

[EKF78] Ekman, P. & Friesen, W. V., *The facial action coding system*, Palo Alto, CA, 1978

[EKF78a] Ekman, P. & Friesen, W.V., *Facial action coding system: A technique for the measurement of facial movement*, Palo Alto, Calif., 1978

[EKF82] Ekman, P & Friesen, W. V., *Felt, false, and miserable smiles.*, in Journal of Nonverbal Behavior, 6, S. 238-252, 1982

[EKM84] Ekman, P, *Expression and the nature of emotion*, in Approaches to emotion, K. R. Scherer & P. Ekman, editors, S. 319-323, Hillsdale, 1984

[EKM99] Ekman, P., *Handbook of Cognition and Emotion*, University of California, San Francisco, CA, USA, New York, 1999

[ELL92] Elliot, C., *The Affective Reasoner: A Process Model of Emotions in a Multi-agent System*, Northwestern University, 1992

[ELL94] Elliot, C., *Research problems in the use of a shallow AI model of personality and emotion*, 1994

[ELS93] Elliot, C. & Siegle, G., *Variables Influencing the Intensity of Simulated Affective States*, Menlo Park, CA, 1993

[FLD99] Fleming, B. & Dobbs, D., *Animating facial features and expressions*, 1999

[FRI86] Frijda, N.H., *The emotions*, Cambridge University Press, 1986

[FRI93a] Frijda, N.H., *The place of appraisal in emotion.*, in Journal: Cognition and Emotion, 7, S. 357-387, 1993

[GOO93] Van Goozen, S. & Fridja, N. H., *Emotion words used in six European countries*, European Journal of Social Psychology, 23, S. 89-95, 1993

[KIZ03] Breazeal, C. et al., *Kismet Hompage*, http://www.ai.mit.edu/projects/humanoid-robotics-group/kismet/kismet.html, 2003

[KKW02] Kranstedt, A., Kopp, S. & Wachsmuth, I., *MURML: A Multimodal Utterance Representation Markup Language for Conversational Agents*, Universität Bielelfed, 2002

[KOP02] Kopp, S., *Synthesis and Coordination of Speech and Gesture for Virtual Multimodal Agents,* Infix DISKI-265, Amsterdam: Akademische Verlagssgesellschaft Aka GmbH, Universität Bielefeld, 2002

[KRU85] Kruse, Otto, *Emotionsdynamik und Psychotherapie*, 1985

[LES02] Leßmann, N., *Eine kognitiv-motivierte Architektur für einen situierten Agenten*, Universität Bielefeld, 2002

[LEV84] Leventhal, H., *A perceptual-motor theory of emotion*, in Journal: Advances in Experimental Social Psychology, 17, S. 117-182, 1984

[LEW03] Leßmann, N. & Wachsmuth, I., *A Cognitively Motivated Architecture for an Anthropomorphic Artificial Communicator*, in Proceedings of the Fifth International Conference on Cognitive Modeling, 2003

[MAT97] Mateas, M., *An Oz-Centric Review of Interactive Drama and Believable Agents*, 1997

[MEH95a] Mehrabian, A., *Theory and Evidence Bearing on a Scale of Trait Arousability*, in Current Psychology, 1, S. 3-28, 1995

[MEH95b] Mehrabian, A., *Framework for a Comprehensive Description and Measurement of Emotional States*, in Genetic, Social, and General Psychoilogy Monographs, 121, S. 339-361, 1995

[MEH96] Mehrabian, A., *Pleasure-Arousal-Dominance: A general framework for describing and measuring Individual Differences in Temperament*, in Current Psychology, 4, S. 261-292, NJ, 1996

[MIN86] Minsky, M., *The Society of Mind*, New York, 1986

[MRS01] Meyer, W.-U., Reisenzein, R. & Schützwohl, A., *Einführung in die Emotionspsychologie, Band I: Die Emotionstheorien von Watson, James und Schachter*, Bern, 2001

[MRS03] Reisenzein, R., Meyer, W.-U. & Schützwohl, A., *Einführung in die Emotionspsychoilogie, Band III: Evolutionspsychologische Emotionstheorien*, Bern, 2003

[OCC88] Ortony, A., Clore, G. L. & Clore, A., *The cognitive structure of emotions*, Cambridge, 1988

[OGL03] *OpenGL Organisation Homepage*, http://www.opengl.org, 2003

[PEN91] Penrose, R., *Computerdenken : Des Kaisers neue Kleider oder die Debatte um künstliche Intelligenz, Bewußtsein und die Gesetze der Physik*, 1991

[QT03] *Qt: A multiplatform, C++ application development framework*, http://www.trolltech.com/, 2003

[REI96] Reilly, W.S.N., *Believable Social and Emotional Agents*, Carnegie Mellon University, Pittsburgh, 1996

[RUE98] Rübenstrunk, G., *Emotionale Computer: Computermodelle von Emotionen und ihre Bedeutung für die emotionspsychologische Forschung*, http://www.ruebenstrunk.de/emocompstart.htm, 1998

[SCH84] Scherer, K.R., *On the Nature and Function of Emotion: A Component Process Approach*, in Approaches to emotion, K. R. Scherer & P. Ekman, editors, S. 293-317, Hillsdale, 1984

[SHA87] Shaver, P., Schwartz, J., Kirson, D. & O´Connor, C., *Emotion knowledge: Further exploration of a prototype approach.*, Journal of Personality and Social Psychology, 52, S. 1061-1086, 1987

[STA00] Stallo, John, *Simulating Emotional Speech for a Talking Head*, Chapter 3, Curtin University of Technology, Western Australia, 2000

[STÖ02] Stößel, D., *Ein System zur Sprachsynthese für multimodale virtuelle Agenten*, Universität Bielefeld, 2002

[STP98] Staller, A. & Petta, P., *Towards a Tractable Appraisal-Based Architecture for Situated Cognizers*, Vienna, 1998

[VEL98] Velásquez, J.D., *Modeling Emotion-Based Decision Making*, Cambridge, Massachusetts, 1998

[WUN22] Wundt, Wilhelm, *Vorlesungen über die Menschen und Tierseele*, Siebente und Achte Auflage 1922, 1863/1922

Anhang

Versicherung

Versicherung gemäß Paragraph 19, Absatz 9 der Diplomprüfungsordnung für den Studiengang Naturwissenschaftliche Informatik an der Technischen Fakultät der Universität Bielefeld vom 10. Januar 1996.

Hiermit versichere ich, dass ich die vorliegende Diplomarbeit selbstständig erarbeitet und keine anderen als die angegebenen Quellen und Hilfsmittel benutzt sowie Zitate kenntlich gemacht habe.

Werther, 30. Juni 2003

Christian Becker