



**Emotionale Aspekte der Mensch-Roboter-Interaktion und ihre
Realisierung in verhaltensbasierten Systemen**

Natalia Esau, Universität Paderborn

C-LAB Report

Vol. 9 (2010) No. 02

Cooperative Computing & Communication Laboratory

ISSN 1619-7879

C-LAB ist eine Kooperation
der Universität Paderborn und der Siemens AG
www.c-lab.de
info@c-lab.de

C-LAB Report

**Herausgegeben von
Published by**

**Dr. Wolfgang Kern, Siemens AG
Prof. Dr. Franz-Josef Rammig, Universität Paderborn**

Das C-LAB - Cooperative Computing & Communication Laboratory - leistet Forschungs- und Entwicklungsarbeiten und gewährleistet deren Transfer an den Markt. Es wurde 1985 von den Partnern Nixdorf Computer AG (nun Siemens AG) und der Universität Paderborn im Einvernehmen mit dem Land Nordrhein-Westfalen gegründet.

Die Vision, die dem C-LAB zugrunde liegt, geht davon aus, dass die gewaltigen Herausforderungen beim Übergang in die kommende Informationsgesellschaft nur durch globale Kooperation und in tiefer Verzahnung von Theorie und Praxis gelöst werden können. Im C-LAB arbeiten deshalb Mitarbeiter von Hochschule und Industrie unter einem Dach in einer gemeinsamen Organisation an gemeinsamen Projekten mit internationalen Partnern eng zusammen.

C-LAB - the Cooperative Computing & Cooperation Laboratory - works in the area of research and development and safeguards its transfer into the market. It was founded in 1985 by Nixdorf Computer AG (now Siemens AG) and the University of Paderborn under the auspices of the State of North-Rhine Westphalia.

C-LAB's vision is based on the fundamental premise that the gargantuan challenges thrown up by the transition to a future information society can only be met through global cooperation and deep interworking of theory and practice. This is why, under one roof, staff from the university and from industry cooperate closely on joint projects within a common research and development organization together with international partners. In doing so, C-LAB concentrates on those innovative subject areas in which cooperation is expected to bear particular fruit for the partners and their general well-being.

ISSN 1619-7879

C-LAB
Fürstenallee 11
33102 Paderborn
fon: +49 5251 60 60 60
fax: +49 5251 60 60 66
email: info@c-lab.de
Internet: www.c-lab.de

© Siemens AG und Universität Paderborn 2010

Alle Rechte sind vorbehalten.

Insbesondere ist die Übernahme in maschinenlesbare Form sowie das Speichern in Informationssystemen, auch auszugsweise, nur mit schriftlicher Genehmigung der Siemens AG und der Universität Paderborn gestattet.

All rights reserved.

In particular, the content of this document or extracts thereof are only permitted to be transferred into machine-readable form and stored in information systems when written consent has been obtained from Siemens AG and the University of Paderborn

Emotionale Aspekte der Mensch-Roboter-Interaktion und ihre Realisierung in verhaltensbasierten Systemen

Natalia Esau
C-LAB Report
Cooperative Computing & Communication Laboratory

1 Einleitung

Im Bereich der sozialen Robotik werden unterschiedliche Fragestellungen zur Mensch-Roboter-Interaktion mit dem Ziel behandelt, den Umgang mit Robotern für den Menschen zu erleichtern. Im Rahmen dieser Arbeit werden emotionale Aspekte der Interaktion zwischen Menschen und Robotern betrachtet, die zur Erweiterung der Kommunikationsfähigkeiten eines Roboters beitragen und einen Roboter zu einem emotional-kompetenten Kommunikationspartner machen.

Ziel dieser Arbeit ist die Entwicklung eines ganzheitlichen emotionalen Robotersystems, das über unterschiedliche emotionale Fähigkeiten bei der Kommunikation mit einem menschlichen Gegenüber verfügt und diese in einer realen Situation adäquat umsetzt. Diese Fähigkeiten werden durch die vier Komponenten Emotionsmodellierung, Emotionserkennung, Emotionsbewertung und Emotionsdarstellung charakterisiert. Aufgrund dieser Komponenten lässt sich die Roboteremotionalität in einem Robotersystem realisieren. Solche Robotersysteme können in einer Vielzahl von Lebenssituationen die Rolle von menschlichen Partnern einnehmen und als benutzerfreundliche Maschinen dem Menschen jederzeit zur Verfügung stehen.

Das in dieser Arbeit entwickelte Emotionsmodell, das Adaptive Fuzzy Emotionsmodell, stellt den emotionalen Zustand des Roboters dar und bildet den konzeptuellen Rahmen, auf dem die weiteren drei Komponenten der Erkennung, Bewertung und Darstellung von Emotionen basieren. Zu den wichtigen Anforderungen dabei gehört die Übereinstimmung mit psychologischen Emotionsmodellen, die sowohl eine sinnvolle Grundlage zur Entwicklung eines informationstechnischen Emotionsmodells als auch eine gute Plattform zur Realisierung der Roboteremotionalität bieten.

Basierend auf diesem informationstechnischen Emotionsmodell wird eine weitere Komponente der Roboteremotionalität, die Emotionserkennung, entwickelt und realisiert. Da die meisten Robotersysteme über unterschiedliche Sensoren wie optische und akustische Sensoren verfügen, werden zwei Systeme VISBER (VISION Based Emotion Recognition) und PROSBER (PROSody Based Emotion Recognition) zur visuellen und sprachbasierten Emotionserkennung entworfen. Die beiden Systeme, die auf Fuzzy Logik basieren, arbeiten in mehreren Schritten von der Merkmalsextraktion bis zur Emotionsklassifizierung und erfüllen die Anforderung der Realzeitfähigkeit.

Zum Verständnis und zur Verarbeitung sowohl „eigener“ Emotionen eines Roboters als auch der Emotionen seines menschlichen Kommunikationspartners wird die Komponente zur Emotionsbewertung entwickelt. Hierfür wird der zeitliche Verlauf von Emotionen und Bedürfnissen in einem technischen System betrachtet, die den inneren Zustand des Roboters darstellen. Der innere Roboterzustand wird von unterschiedlichen äußeren und inneren Reizen beeinflusst, die entweder aus der Umgebung wahrgenommen werden oder in dem System selbst entstehen. Deshalb

wird eine kognitive Bewertung der äußeren und inneren Reize erarbeitet, mit Hilfe derer die Anregungsfunktionen zur dynamischen Entwicklung der Emotionen und Bedürfnisse in einem Robotersystem definiert werden.

In Kombination mit den äußeren Reizen ist der innere Roboterzustand für die Auswahl adäquater Verhaltensweisen verantwortlich, die zur emotionalen Reaktion des Roboters führt und die vierte Komponente der Roboteremotionalität, die Emotionsdarstellung, realisiert.

Die vier emotionalen Aspekte eines Robotersystems werden anhand des Roboterkopfes MEXI (Machine with Emotionally eXtended Intelligence) umgesetzt, der auf Basis einer verhaltensbasierten Architektur realisiert ist. Durch Erweiterung der vorhandenen Architektur um die emotionalen Komponenten verfügt MEXI über die Fähigkeiten zur Emotionsmodellierung, -erkennung, -bewertung und -darstellung. Ausgestattet mit diesen Fähigkeiten sowie mit unterschiedlichen Sensoren und Aktoren stellt MEXI ein ganzheitliches emotionales Robotersystem dar und kann als emotional-kompetenter Roboter im Bereich der Mensch-Roboter-Interaktion eingesetzt werden.

2 Überlegungen zu der emotionalen Kompetenz in der Mensch-Roboter-Interaktion

Die Bezeichnung *emotionale Intelligenz* prägten die Psychologen Salovey und Mayer im Jahre 1990. Sie beschreiben emotionale Intelligenz als Form der sozialen Intelligenz, die die Fähigkeit mit einbezieht, eigene Gefühle und Emotionen und Gefühle und Emotionen anderer zu überwachen, diese zu unterscheiden und zu verwenden, um das eigene Denken und Handeln zu leiten [SM90]. Im Jahre 1995 veröffentlichte der Psychologe Daniel Goleman den populären Bestseller *Emotional Intelligence* [Gol96], in dem er anhand zahlreicher Beispiele die Bedeutung der emotionalen und sozialen Faktoren in der menschlichen Kommunikation zeigt. Für Robotikanwendungen ist der Begriff der emotionalen Kompetenz besonders interessant, der in der Psychologie als Synonym für emotionale Intelligenz verwendet wird. Im Bereich der Sozialpsychologie, Berufspädagogik und Personalberatung wird er aber häufig als Fähigkeit bezeichnet, die eigenen und fremde Gefühle im konkreten Kontext richtig zu bewerten und so Konflikte und Stress zu vermeiden.

Zusammengefasst beinhaltet emotionale Kompetenz die folgenden vier Faktoren:

- die Fähigkeit zur Wahrnehmung von Emotionen,
- die Fähigkeit zum Verständnis eigener und fremder Emotionen,
- die Fähigkeit zur Regulation eigener und fremder Emotionen und
- die Fähigkeit zu einem der Situation angemessenen Ausdruck von Emotionen.

Der erste Faktor bedeutet die Fähigkeit, Emotionen einer fremden Person anhand ihrer Mimik, Gestik und Sprache wahrzunehmen. Der zweite Faktor umfasst die Fähigkeit, Veränderungen der fremder Emotionen zu bewerten. Der dritte Faktor beschreibt die Fähigkeit, den eigenen emotionalen Zustand entsprechend dieser Bewertung anzupassen bzw. zu regulieren. Der vierte Faktor umfasst die Emotionsdarstellung durch adäquate Gesichtsausdrücke, Körperhaltung und Stimme [MSC04]. Hierdurch kann implizit auch eine Regulation fremder Emotionen erfolgen.

Das Konzept der emotionalen Kompetenz beschreibt alle Schritte der emotionalen Kommunikation zwischen Menschen, von der Emotionserkennung bis zur Emotionsdarstellung. Wird das Konzept der emotionalen Kompetenz auf ein Robotersystem übertragen, das über die vier Fähigkeiten bei der Kommunikation mit einem menschlichen Gegenüber verfügt, kann dieses System als emotional-kompetent bezeichnet werden.

Da die Roboteremotionalität im Unterschied zur menschlichen Emotionalität nicht an sich besteht (ein Roboter hat kein subjektives Empfinden und zeigt keine physiologischen Reaktionen), sondern nur durch die menschliche Interpretation des Roboterhaltens während einer Interaktion mit seinem menschlichen Gegenüber, bietet sich die Definition der emotionalen Kompetenz als Grundlage für eine Realisierung der Roboteremotionalität an [EKK08a].

Betrachtet man als mögliche Interaktionen eines Roboters zunächst nur die Möglichkeiten zur Emotionsdarstellung durch Mimik, Gestik etc., so lässt sich basierend auf dieser Grundlage die Roboteremotionalität durch die folgenden Komponenten in einem Rückkopplungskreis realisieren:

- die Emotionserkennung realisiert die Fähigkeit zur Wahrnehmung von Emotionen,
- die Emotionsbewertung ist für das Verständnis eigener und fremder Emotionen verantwortlich,
- die Emotionsdarstellung durch Mimik, Sprache oder Gestik erlaubt einen Ausdruck eigener Emotionen und damit implizit auch die Beeinflussung/Regulation fremder Emotionen,
- das Emotionsmodell erfüllt mehrere Aufgaben. Es stellt den emotionalen Zustand des Roboters dar und „ersetzt“ damit das subjektive und physiologische Befinden des Menschen. Neben dieser statischen Sicht beinhaltet es auch dynamische Aspekte zur Regulation eigener „Emotionen“. Desweiteren bildet es den konzeptuellen Rahmen, auf dem die obigen drei Komponenten Emotionserkennung, -bewertung und -darstellung arbeiten.

Abbildung 1 stellt das Zusammenwirken dieser vier Komponenten in einem Rückkopplungskreis dar.

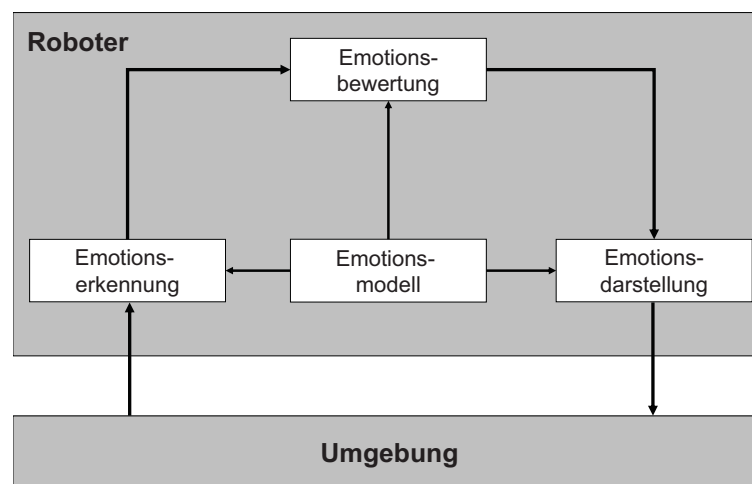


Abbildung 1: Zusammenwirkung der Grundkomponenten der Roboteremotionalität

In einem Robotersystem müssen zuerst von außen einwirkende Stimuli durch unterschiedliche Sensoren wahrgenommen werden. Da für ein emotional-kompetentes Robotersystem insbesondere die Erkennung des emotionalen Zustandes seines menschlichen Gegenübers wichtig ist, wird die Umwelt mit einer Person assoziiert, die ihre Emotionen mittels Gesichtsausdrücke, Sprachäußerungen oder Gesten repräsentieren kann. Danach werden die wahrgenommenen Stimuli situationsabhängig bewertet. Dabei wird beachtet, was die Wahrnehmung für die aktuellen Bedürfnisse und Emotionen des Robotersystems bedeutet. Die Resultate der Stimuli-Bewertung werden den emotionalen Zustand des Roboters beeinflussen. Dabei muss der Roboter in der Lage sein, seine Emotionen situationsabhängig anzupassen bzw. zu regulieren. Letztendlich wird das Verhalten des Systems bestimmt, das durch den inneren Emotionszustand unter Berücksichtigung der Sichtbarkeitsinformationen gesteuert werden. So werden adäquate Verhaltensweisen generiert wie beispielsweise emotionale Gesichtsausdrücke oder emotionale Sprachäußerungen. Die Umwelt wirkt somit auf das System ein, und das System versucht seinerseits über die Aktorik die Umwelt zu beeinflussen. Dementsprechend entsteht ein Rückkopplungskreis, den das System mit der Umwelt bzw. einem menschlichen Kommunikationspartner bildet.

Die Betrachtung der Emotion eines Roboters als Ausdruck dynamischer Prozesse, an denen mehrere Subsysteme beteiligt sind, führt zu einem neuen Konzept einer emotional-kompetenten Mensch-Maschine-Schnittstelle, die durch ihre emotionale Kompetenz Konflikte oder stressige Situationen zwischen Mensch und Maschine zu vermeiden ermöglicht und eine vielschichtige natürliche Mensch-Roboter-Interaktion unterstützt.

3 Der Roboterkopf MEXI

Ziel dieser Arbeit ist die Entwicklung eines ganzheitlichen emotionalen Systems, das sowohl zur Erkennung, Bewertung, Darstellung als auch zur Modellierung von Emotionen im Bereich der Mensch-Roboter-Interaktion eingesetzt werden kann. Als Grundlage für die Realisierung eines solchen Systems wurde der Roboterkopf MEXI (Machine with Emotionally eXtended Intelligence) verwendet, der als Forschungsergebnis einer Gruppe von Informatikern unter der Leitung von Dr. Bernd Kleinjohann im C-LAB entstand (siehe Abbildung 2) [Kle01]. MEXI war bereits in der Lage, künstliche Emotionen durch Gesichtsausdrücke, spezifische Ohrenbewegungen und emotional-betonte Aussprache aufgrund von einer verhaltensbasierten Architektur zu simulieren sowie eigene Bedürfnisse proaktiv zu verfolgen. Zur Umsetzung eines ganzheitlichen emotionalen Systems fehlten jedoch einige Aspekte wie Emotionserkennung, -modellierung und -bewertung sowie die Realisierung der adäquaten Reaktionen auf bestimmte Reize, die MEXI aus seiner Umgebung erhält. Um diese Herausforderungen zu lösen, sollte die vorhandene verhaltensbasierte Architektur um die fehlenden emotionalen Komponenten erweitert werden. Die erweiterte Architektur wird emotionsgesteuerte verhaltensbasierte Architektur genannt.

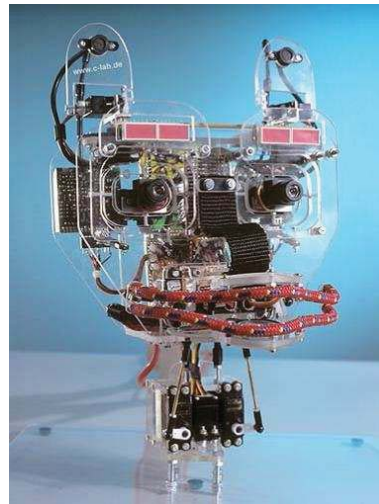


Abbildung 2: MEXI - Machine with Emotionally eXtended Intelligence

4 Emotionsgesteuerte verhaltensbasierte Architektur

Die Software Architektur des Roboterkopfs MEXI basiert auf Nilsson's Triple Tower Architektur [Nil98] und besteht aus drei Hauptkomponenten: Wahrnehmung, Handlungssteuerung und Aktorik (siehe Abbildung 3).

Der Roboterkopf nimmt seine Umgebung mittels eingebauter Sensoren wahr. Visuelle Informationen über die Welt bekommt MEXI über seine Kamera-Augen. Die sprachliche Information kann per Mikrofon oder Tastatur wahrgenommen werden. Für die Verarbeitung der über diese Sensoren aufgenommenen Daten ist das Wahrnehmungsmodul zuständig.

Das Wahrnehmungsmodul besteht aus drei Komponenten: Bildvorverarbeitung, Sprachvorverarbeitung und Emotionserkennung. Die Bildvorverarbeitungskomponente umfasst Bearbeitungsprozeduren zur Detektion menschlicher Gesichter bzw. der Spielobjekte aus den Eingangsbildern. Die Resultate der Bildvorverarbeitung werden entweder zu VISBER (VISION Based Emotion Recognition) oder direkt zum Verhaltenssystem weiterleitet. Die Komponente VISBER ist für die Emotionserkennung mittels Gesichtsanalyse aus Bildsequenzen zuständig.

Die Aufgabe der *Sprachvorverarbeitung* besteht darin, die sprachlichen Eingaben zu verarbeiten. Als sprachliche Eingaben werden einerseits über die Tastatur eingegebenen Sätze verstanden, andererseits die natürliche Sprache,

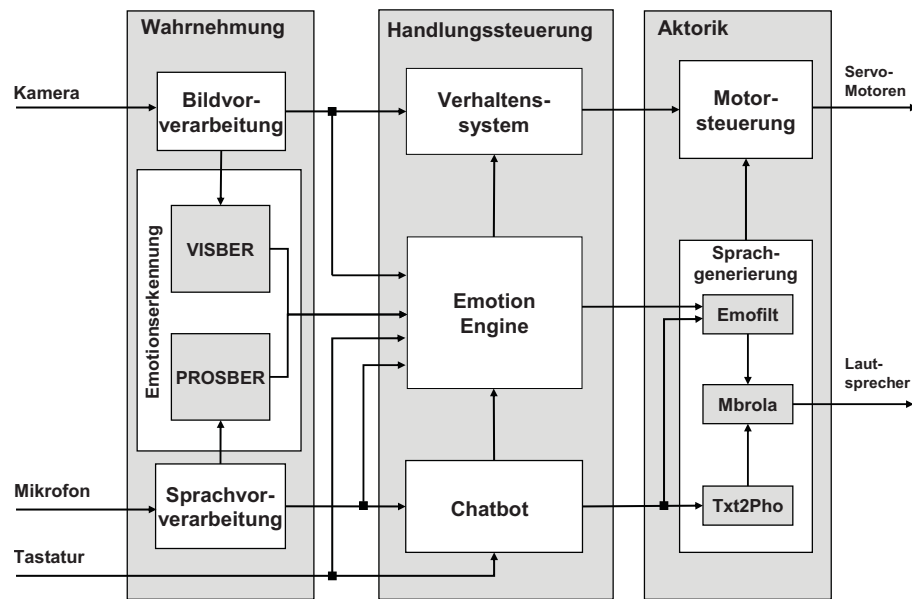


Abbildung 3: Überblick über die Architektur des Roboterkopfs MEXI

die mittels des in das System integrierten kommerziellen Tools *ViaVoice* erkannt werden kann. Die Sprachdaten, die über Tastatur aufgenommen werden, werden dem Chatbot in Textform übergeben. Die mittels Mikrofon wahrgenommenen Audiodaten werden von dem Spracherkennung *ViaVoice* verarbeitet und als Text zu einem Chatbot weitergeleitet. Außerdem werden die Audiodaten durch *PROSBER* (PROsody Based Emotion Recognition) analysiert, um aus der Prosodie eines gesprochenen Satzes die Emotion des Sprechers zu erkennen.

Die beiden Systeme *VISBER* und *PROSBER* stellen die Emotionserkennungskomponente der MEXI-Architektur dar. Die Resultate der beiden Systeme können miteinander kombiniert werden und werden als gemeinsames Ergebnis zum *Handlungssteuerungsmodul* weitergeleitet.

Das Handlungssteuerungsmodul erlaubt MEXI auf die audio-visuellen Eingaben mit entsprechenden Kopf-, Ohren- bzw. Augenbewegungen und adäquater Sprachausgaben zu reagieren. MEXIs Verhalten wird aber nicht nur rein reaktiv durch sein Verhaltenssystem bestimmt, wie es bei verhaltensbasierten Architekturen üblich ist. In seinem Handlungssteuerungsmodul besitzt der Roboterkopf zusätzlich die Emotionsbewertungskomponente *Emotion Engine*, die das Verhaltenssystem durch den internen Roboterzustand steuern kann. Im Rahmen der *Emotion Engine* wird der interne Zustand von MEXI mittels seiner Emotionen und Bedürfnisse modelliert, die sich in Abhängigkeit von den wahrgenommenen Reizen ändern.

Das Verhalten des Roboterkopfs wird im Rahmen des Verhaltenssystems als Kombination einiger Basisverhaltensweisen realisiert, wie beispielsweise *Look Around*, *Follow Face*, *Smile* oder *Follow Toy*, die als adäquate Reaktion auf die Änderungen des internen Roboterzustands und/oder auf bestimmte Reize (Wahrnehmungen aus seiner Umgebung) ausgelöst werden.

Das *Aktorik-Modul* realisiert einerseits die Steuerung der Servo-Motoren durch die Komponente *Motorsteuerung* und andererseits die Generierung der Sprachausgaben durch die Komponente *Sprachgenerierung*.

5 Adaptives Fuzzy Emotionsmodell

Im Folgenden wird ein Emotionsmodell vorgestellt, das basierend auf Fuzzy Logik definiert ist und sich an unterschiedliche psychologische Emotionsmodelle der diskreten Emotionstheorien adaptieren lässt. Dieses Emotionsmodell wurde als zentrale Grundkomponente der Roboteremotionalität im Rahmen dieser Arbeit entwickelt und

realisiert [EKK05]. Das Ziel dabei war, eine theoretische bzw. mathematische Modellierung der Emotionen zu erstellen, die zur Anwendung für unterschiedliche Aspekte der emotionalen Mensch-Maschine-Interaktion geeignet ist. Das Modell soll die Ideen psychologischer Forscher repräsentieren und sie in seiner Struktur widerspiegeln. Außerdem soll das Emotionsmodell die einzelnen Elemente der emotionalen verhaltensbasierten Architektur zur Mensch-Roboter-Interaktion miteinander verbinden. Dazu gehören vor allem das Wahrnehmungsmodul, das zur Emotionserkennung die Eingangsdaten liefert, und die Emotion Engine, in der der Emotionszustand bestimmt wird und die Konfiguration der Koeffizienten zur nachfolgenden Auswahl von emotionspezifischen Verhaltensweisen stattfindet. Somit bildet das adaptive Fuzzy Emotionsmodell die Grundlage zur Emotionserkennung, zur Emotionsdarstellung sowie zur Bewertung des emotionalen Roboterszustandes und kann als Leitfaden des gesamten Systems angesehen werden.

Als psychologische Grundlage für die Definition des Emotionsmodells wird das Modell des amerikanischen Psychologen Plutchik gewählt [Plu80]. Plutchik definiert acht fundamentale Verhaltensweisen, denen er acht primäre Emotionen bzw. Basisemotionen zuordnet. Alle primären Emotionen charakterisiert Plutchik mit einem zusätzlichen Merkmal, das er als Intensität beschreibt. In seinem Emotionsmodell unterscheidet Plutchik zwischen drei Emotionsintensitäten: geringe, mittlere und starke. Mit abnehmender Intensität unterscheiden sich die Basisemotionen zunehmend weniger voneinander. Alle anderen Emotionen sind Mischungen oder Kombinationen der primären Emotionen, die Plutchik als sekundäre Emotionen bezeichnet. Aus der Mischung zweier oder dreier primärer Emotionen entstehen sogenannte Dyaden oder Triaden. Es gibt eine große Anzahl psychologischer Emotionsmodelle, denen die Idee der Basisemotionen zugrunde liegt. Es besteht eine erhebliche Uneinigkeit über die Anzahl bzw. die Intensität der Basisemotionen zwischen den Emotionstheoretikern. Somit gehört die Heterogenität der Basisemotionen zu den größten Problemen der diskreten Emotionstheorien.

Das im Rahmen dieser Arbeit entwickelte informationstechnische Emotionsmodell erlaubt, all diese Uneinigheiten zu umgehen, da die Anzahl der Basisemotionen bzw. die Abstufung der Intensität bei der auf der Fuzzy Logik basierten Emotionsmodellierung frei wählbar ist. Somit ist das Modell an mehrere psychologische Emotionsmodelle der diskreten Emotionstheorie adaptierbar und wird dementsprechend *Adaptives Fuzzy Emotionsmodell* (AFEM) genannt.

5.1 Definition der Emotionen über Fuzzy-Mengen

Zunächst wird ein beliebiger Emotionszustand als Fuzzy-Menge definiert, die die Intensitäten der differenzierten Basisemotionen beschreibt. Im Rahmen der Fuzzy Logik wird eine Fuzzy-Menge A durch ihre Zugehörigkeitsfunktion μ_A charakterisiert, die jedem Element x in der Grundmenge X eine reelle Zahl im abgeschlossenen Intervall $[0, 1]$ zuordnet. A ist dann vollständig definiert als eine Menge von Paaren:

$$A = \{(x, \mu_A(x)) | x \in X\}.$$

Um eine Emotion als Fuzzy-Menge definieren zu können, wird angenommen, dass die Intensität einer Basisemotion keinen unendlichen Wert annehmen kann und deshalb auf das Intervall $[0, 1]$ normiert werden kann.

Emotion als Fuzzy-Menge. Seien BE eine endliche Menge von n Basisemotionen e_1, e_2, \dots, e_n und $\mu_{FE_j} : BE \rightarrow [0, 1], j = 1, 2, \dots$ eine unendliche Menge von Zugehörigkeitsfunktionen, die jeder Basisemotion e_i ihre Intensität aus dem Intervall $[0, 1]$ zuordnen. Jede Fuzzy-Menge FE_j , definiert als

$$FE_j := \{(e_i, \mu_{FE_j}(e_i)) | e_i \in BE\}, j = 1, 2, \dots, \quad (1)$$

korrespondiert genau zu einer Emotion E_j . Der Wert von $\mu_{FE_j}(e_i)$ ist also eine beliebige reelle Zahl zwischen 0 und 1 und gibt damit für jede Basisemotion $e_i \in BE$ den Grad des Zutreffens bzw. die Intensität der Basisemotion e_i an. Im Folgenden wird die Fuzzy-Menge (1) als Fuzzy-Emotionsmenge bezeichnet.

5.2 Geometrische Darstellung des Adaptiven Fuzzy-Emotionsmodells

Die Idee des Adaptiven Fuzzy-Emotionsmodells basiert auf der geometrischen Darstellung der Fuzzy-Emotionsmenge, die die geometrische Fuzzy-Menge-Notation von B. Kosko [Kos92] zugrunde liegt. Dabei wird jede Fuzzy-Emotionsmenge als Punkt eines n -dimensionalen Hyperwürfels dargestellt, falls die Emotionsmenge $BE = \{e_1, e_2, \dots, e_n\}$ mit n Elementen als Grundmenge genommen wird.

In einer geometrischen Darstellung wird jedem Element e_i eine Dimension zugeordnet. Auf der zugehörigen Achse, die mit e_i bezeichnet wird, wird der Zugehörigkeitsgrad des Elementes e_i zur Menge FE_j angetragen. Im Falle einer scharfen Menge steht für ein vorhandenes Element eine „1“, andernfalls eine „0“.

Die Definition einer Fuzzy-Emotionsmenge (1) basiert auf einer begrenzten Anzahl n von Basisemotionen, die laut psychologischer Emotionsmodelle eine begrenzte Anzahl von Emotionszuständen bestimmen. Die maximale Anzahl aller möglichen Emotionen, die mit n Basisemotionen definierbar sind, ist gleich $2^n - 1$. Das sind entweder die Basisemotionen oder deren Mischungen. Im Rahmen des Fuzzy-Emotionsmodells werden sie durch $2^n - 1$ Eckpunkte des n -dimensionalen Hyperwürfels präsentiert. Der Eckpunkt $(0, 0, \dots, 0)$ stellt dabei einen neutralen Zustand dar. Dementsprechend kann der Hyperwürfel in 2^n Bereiche unterteilt werden, die je eine Emotion bzw. einen neutralen Zustand aus 2^n möglichen emotionalen Zuständen repräsentieren und im Rahmen des Fuzzy-Emotionsmodells die entsprechenden Emotionsklassen bestimmen. Jeder beliebige Punkt, der sich in einem bestimmten Bereich des Hyperwürfels befindet, wird somit der entsprechenden Emotionsklasse zugeordnet und repräsentiert eine bestimmte Emotion bzw. einen neutralen Zustand.

Mit der geometrischen Darstellung des Fuzzy-Emotionsmodells ist sowohl eine anschauliche Repräsentation der Basisemotionen als auch ihrer abgeleiteten Emotionen bzw. Mischungen möglich. Außerdem ist das Fuzzy-Emotionsmodell gleichzeitig ein adaptives Emotionsmodell, das die Emotionsmodelle der diskreten Emotionstheorie unabhängig von der Anzahl und Auswahl der Basisemotionen mathematisch mit Hilfe der Fuzzy Logik beschreibt bzw. sich an ein beliebiges Emotionsmodell anpassen lässt. Im Folgenden wird das AFEM als Grundlage der Emotionsklassifizierung zur Emotionserkennung im Wahrnehmungsmodul sowie der Emotionsdarstellung im Handlungssteuerungsmodul der emotionsgesteuerten verhaltensbasierten Architektur verwendet.

6 Automatische Emotionserkennung in einem Robotersystem

Es gibt verschiedene Möglichkeiten, den emotionalen Zustand eines menschlichen Kommunikationspartners bei der Mensch-Maschine-Interaktion zu ermitteln. Besondere Bedeutung für den Menschen hat die Emotionserkennung aus gesprochener Sprache und anhand der Mimik, wie sie auch in dieser Arbeit verwendet wird. Im Folgenden werden zwei Systeme VISBER (VISION Based Emotion Recognition) und PROSBER (PROSody Based Emotion Recognition) zur automatischen Emotionserkennung eines menschlichen Kommunikationspartners vorgestellt. Bei der Entwicklung der beiden Systeme wurde das Ziel verfolgt, funktionsfähige Prototypen zu erstellen, die Anforderungen der Realzeitfähigkeit erfüllen und in die MEXI-Architektur integrierbar sind. Somit werden die Fähigkeiten des Roboterkopfes um die automatische Erkennung von Emotionen erweitert.

6.1 VISBER - die visuelle Emotionserkennung

Wie kann man den Gesichtsausdruck eines Menschen messen? Zu diesem Zweck wurde das Facial Action Code System (FACS) im Jahre 1978 von den amerikanischen Psychologen Ekman und Friesen entwickelt [EF78]. Die Gesichtsmuskeln bewirken sichtbare Veränderungen der Mimik, beispielsweise das Heben der Augenbrauen. Diese kleinsten Einheiten der Mimik werden Action Units (AUs) genannt. Jede AU kann dabei in drei verschiedenen Intensitätsstufen auftreten, gleichzeitiges Auftreten mehrerer AUs ist gestattet. Eine bestimmte Kombination von einzelnen Aktionseinheiten (AUs) erzeugt einen bestimmten Gesichtsausdruck, der dann auf eine konkrete Emotion hinweist.

Das System VISBER basiert auf der Kodierung von Gesichtsausdrücken durch Aktionseinheiten (AUs) und arbeitet in mehreren Schritten von der Bildaufnahme bis zur Ausgabe der vom System berechneten Emotionswerte [EWKK07b]. Die Emotions-Ausgabewerte werden dabei durch Emotionsklassifizierung bestimmt, die auf Basis des Adaptiven Fuzzy-Emotionsmodells aufgebaut ist.

Nach Gesichtslokalisierung werden zunächst die charakteristischen Gesichtspunkte aus den Videobildern extrahiert, die einige Anforderungen erfüllen müssen. Zum einen müssen sie möglichst robust im Gesicht gefunden werden können, zum anderen müssen sie charakteristische Merkmale darstellen, mit denen die emotionalen Gesichtsausdrücke bestimmt werden können. Auf Basis von charakteristischen Merkmalspunkten werden die Gesichtsmerkmale definiert, die sich bei emotionalen Gesichtsausdrücken sichtbar verändern. In vielen bekannten Ansätzen zur visuellen Emotionserkennung [PP06],[CHFT06],[KNP06],[KB03] werden als Gesichtsmerkmale die Distanzen zwischen den charakteristischen Punkten benutzt, die zuerst normiert und erst dann in die Emotionsklassifizierung einbezogen werden. In VISBER werden jedoch zum ersten Mal Winkel als Gesichtsmerkmale zur Mimikerkennung verwendet, die mit Hilfe von 2D-Positionen der charakteristischen Gesichtspunkte berechnet werden. Im Vergleich zu Distanzen sind Winkel bei der Kameraprojektion invariant und können deshalb ohne Normierung in die Bewertung einbezogen werden.

VISBER berechnet die Veränderungen der Gesichtsmerkmale in jedem einzelnen Bild der aufgenommenen Bildsequenz und wertet sie mittels der Fuzzy-Inferenz auf Basis von Fuzzy-Regeln aus. Die Fuzzy-Regeln wurden dabei aufgrund von psychologischen Arbeiten und selbständiger Analyse der unterschiedlichen Gesichtsausdrücke formuliert und anschließend zur Anpassung der Parameter von Zugehörigkeitsfunktionen automatisch trainiert.

Einige personenabhängigen Mimikerkennungssysteme werden mittels neuronaler Netze trainiert [SBHN04]. Das erfordert vor allem umfangreiche Trainingsdaten, deren Erstellung sich in der Praxis als schwierig erweist. Im Vergleich zu neuronalen Netzen braucht man in VISBER zur Anpassung der Fuzzy-Regeln nur wenige Messungen der charakteristischen Gesichtsmerkmale, die auf Basis einiger Bildaufnahmen gemacht werden.

Der große Vorteil von VISBER gegenüber existierenden visuellen Emotionserkennungssystemen besteht jedoch darin, dass VISBER unter Benutzung des AFEMs in der Lage ist, sowohl die Basisemotionen als auch ihre unterschiedlichen Mischungen zu erkennen [EWKK07a].

VISBER erzielte gute Erkennungsraten der personenabhängigen visuellen Emotionserkennung auf Basis der Fuzzy Logik. Die durchschnittlichen Erkennungsraten sind mit den bereits existierenden Ansätzen zur visuellen Emotionserkennung vergleichbar und liegen bei 60% bei der personenunabhängigen Emotionserkennung und bei 79% für die personenabhängige Erkennung. Neben den guten Erkennungsraten erfüllt VISBER die Anforderungen der Realzeitfähigkeit und kann deshalb für die visuelle Emotionserkennung im MEXI-System eingesetzt werden.

6.2 PROSBER - die prosodiebasierte Emotionserkennung

PROSBER stellt das System zur automatischen Emotionserkennung aus gesprochener Sprache dar. Bei der Entwicklung von PROSBER wurde das Ziel verfolgt, einen funktionsfähigen Prototyp für eine prosodiebasierte Emotionserkennung zu erstellen, der in die MEXI-Architektur integrierbar ist. Somit werden die Fähigkeiten des Roboterkopfes um die Erkennung von Emotionen aus gesprochenen Sprache erweitert.

Zur Erstellung des Systems PROSBER wurden sowohl die Erkenntnisse aus der Psycholinguistik, die die Darstellung von Emotionen in Sprache beschreiben, als auch Ergebnisse aus bereits existierenden Systemen zur Emotionserkennung berücksichtigt. Aus der Analyse dieser beiden Bereiche gingen in erster Linie prosodische Sprachmerkmale hervor, die sich besonders gut eignen, um den Emotionsgehalt eines Sprachsignals zu bestimmen. Prosodie ist die Gesamtheit sprachlicher Eigenschaften, die über den gesprochenen Text hinaus gehen. Zu prosodischen Eigenschaften gehören solche akustischen Merkmale wie Betonung, Sprechrhythmus und Sprachmelodie. Die Prosodie umfasst damit Dauer, Lautheit und Tonhöhenverlauf (Intonation) einer Sprache [Bur01].

Das Verfahren zur prosodiebasierten Emotionsklassifizierung basiert auf Fuzzy Logik und ist ähnlich aufgebaut wie die Fuzzy-Klassifizierung zur Mimikerkennung in VISBER. Im Unterschied zu VISBER ist PROSBER das erste System, das Fuzzy Logik zur prosodiebasierten Emotionsklassifizierung einsetzt. Während bei VISBER sechs

Winkel als Merkmale zur visuellen Klassifizierung ausreichen, ist die Anzahl zu berechnender Sprachmerkmale bei PROSBER wesentlich größer¹. Dadurch wurde eine manuelle Erstellung der Fuzzy Regelsysteme unmöglich. Zudem wäre es viel zu umfangreich, wenn alle 20 Merkmale zur Klassifizierung verwendet würden. Aus diesem Grund wurde ein automatisches Training realisiert, das für jede Emotion aus einer Menge von Trainingsdaten die aussagekräftigsten Merkmale (in der Regel 4-6) bestimmt und eine Regelbasis basierend auf diesen Merkmalen generiert. Die automatische Merkmalsauswahl zur prosodiebasierten Emotionserkennung gab es zum Zeitpunkt der Realisierung von PROSBER noch nicht.

Bei den existierenden Systemen zur Emotionserkennung aus der Sprache handelt es sich hauptsächlich um Verfahren, die auf einer statistisch orientierten Grundidee basieren. Im Unterschied hierzu ist das System PROSBER zur prosodiebasierten Emotionserkennung aus gesprochener Sprache auf der Basis von Fuzzy Logik realisiert, deren klare Vorteile in der Nachvollziehbarkeit und Verallgemeinerbarkeit der Fuzzy-Klassifizierungsverfahren liegen. Während maschinellen Lernverfahren, wie Support Vector Machines, Hidden Markov Modellen oder Neuronalen Netzen ein so genanntes Black-Box-Modell zugrunde liegt, ermöglicht das Fuzzy-Regelsystem Rückschlüsse darüber zu machen, aufgrund welcher Eigenschaften Daten verschiedenen Emotionsklassen zugeordnet werden. Da der fuzzy-basierte Regelsatz nicht mit festen Schwellwerten arbeitet, um bestimmte Merkmalsausprägungen eines emotionalen Sprachausdrucks einer bestimmten Emotion zuzuordnen, hat er den Vorteil, die interpersonale Variabilität des Emotionsausdrucks bei verschiedenen Sprechern insbesondere für sprecherunabhängige Systeme besser modellieren zu können. Der geringere Bedarf an Trainingsdaten für das Berechnen von Fuzzy-Regelsysteme ist ein weiteres Argument für die Entscheidung für Fuzzy Logik gegenüber den anderen maschinellen Lernverfahren.

Die bereits existierenden Emotionserkennungsverfahren aus der Sprache sind nur in der Lage, eine kleine Anzahl der ausgewählten Basisemotionen zu modellieren. Das war ein weiterer wichtiger Grund für die Entscheidung, in PROSBER Fuzzy-Logik einzusetzen. Die Benutzung des AFEMs ermöglicht PROSBER die Klassifizierung nicht nur über einige Basisemotionen, sondern auch über deren Kombinationen, was das System flexibler gegenüber den anderen existierenden Ansätzen macht.

Insgesamt weist PROSBER, das seine Emotionsklassifizierung auf Basis der Fuzzy-Logik aufbaut, gute Erkennungsergebnisse auf. Die durchschnittlichen Erkennungsraten sind mit den existierenden Systemen zur sprachbasierten Emotionserkennung vergleichbar und liegen zwischen 76% und 79% für die sprecherabhängige Erkennung und zwischen 54% und 60% für die sprecherunabhängige Erkennung. Die beste durchschnittliche Erkennungsrate von 79% erreicht das System bei Verwendung der sprecherabhängigen Emotionsklassifizierung mit sechs Merkmalen.

7 Emotionen und Bedürfnisse eines Roboters

Emotionen und Bedürfnisse gehören zu den grundlegenden Vorgängen des menschlichen Handelns, die in der Psychologie als aktivierende Komponenten bezeichnet werden. Aus der Interaktion zwischen den aktivierenden emotionalen und bedürfnisspezifischen Komponenten sowie verschiedenen kognitiven Prozessen, die zur Bestimmung von Zielen und Handlungsprogrammen führen, erwächst die Motivation des Handelns. Im Rahmen der Emotion Engine, die in der Architektur des Roboterkopfs MEXI eine zentrale Rolle spielt, werden psychologische Prozesse wie Emotionen, Bedürfnisse und Auswahl von Verhaltensweisen modelliert. Zeitlich aufeinander folgende Zustände werden dynamisch abgebildet und im zeitlichen Verlauf quantitativ dargestellt.

MEXI hat kein internes Umgebungsmodell, um sein Verhalten zu planen bzw. zu steuern. Für diesen Zweck benutzt MEXI vielmehr seine Emotionen und Bedürfnisse. Im Prinzip hat MEXIs System zwei wichtige Zielsetzungen, um seine Verhaltensweisen zu steuern. Zum einen versucht MEXI positive Emotionen zu erreichen und negative Emotionen zu vermeiden. Zum anderen werden seine Bedürfnisse in einem komfortablen (homöostatischen)² Bereich gehalten, der sich durch gewisse Schwellwerte definieren lässt.

¹Insgesamt werden in PROSBER 20 Sprachmerkmale zur Emotionserkennung analysiert.

²Als Homöostase wird in der Biologie das fließende Gleichgewicht zwischen einem Organismus und seiner Umwelt bezeichnet, das als Grundvoraussetzung für eine Leistungsanpassung gilt.

7.1 Dynamische Entwicklung der Emotionen und Bedürfnissen

Für MEXI wird eine Menge von Basisemotionen eingeführt, welche bei dem Roboterkopf mit Hilfe von Gesichtsausdrücken und Sprachausgaben leicht repräsentiert werden können. Das sind „Ärger“, „Freude“, „Traurigkeit“ und „Angst“. „Freude“ ist eine positive Emotion, welche MEXI anstrebt. Die anderen negativen Emotionen versucht MEXI zu vermeiden.

Um einen einfachen Steuerungsmechanismus für MEXI zu realisieren, werden hauptsächlich zyklische homöostatische Bedürfnisse betrachtet. Homöostatische Bedürfnisse motivieren das Verhalten, einen bestimmten Gleichgewichtszustand zu erreichen. Beispiele sind Hunger oder Durst beim Menschen. Homöostatische Bedürfnisse von MEXI sind beispielsweise das Kommunikationsbedürfnis oder ein Spiel-Bedürfnis.

In MEXIs Emotion Engine wird jede Emotion durch eine Emotionsintensität zwischen 0 und 1 und jedes Bedürfnis durch eine Bedürfnisstärke von -1 bis 1 repräsentiert. Die grundlegenden Eigenschaften der zeitlichen Entwicklung der Emotionen bzw. Bedürfnisse werden mit Hilfe der grafischen Darstellung in Abbildungen 4 und 5 beschrieben. In Abbildung 4 ist der zeitliche Verlauf einer Emotion e , beispielsweise „Freude“, als durchgezogene Linie zu sehen. Aus der Abbildung 5 lässt sich der zeitliche Verlauf eines Bedürfnisses d , beispielsweise des „Kommunikationsbedürfnisses“, als durchgezogene Linie erkennen. Die punktierte Linie zeigt den Verlauf der Bewertung der Stimuli, die vom System wahrgenommen werden. Falls der aktuell wahrgenommene Stimulus positiv bewertet wird, liegt die punktierte Linie über der Zeitachse. Bei „negativer“ Bewertung fällt diese Linie unter die Zeitachse.

Für jede Emotion ist ein Schwellwert θ_e , $0 < \theta_e < 1$, definiert. Solange der Emotionswert über diesem Schwellwert liegt, wird MEXIs Verhalten so konfiguriert, dass sich diese Emotion in der Roboter-Mimik widerspiegelt. In Abbildung 4 sind solche Situationen als schattierte Flächen dargestellt.

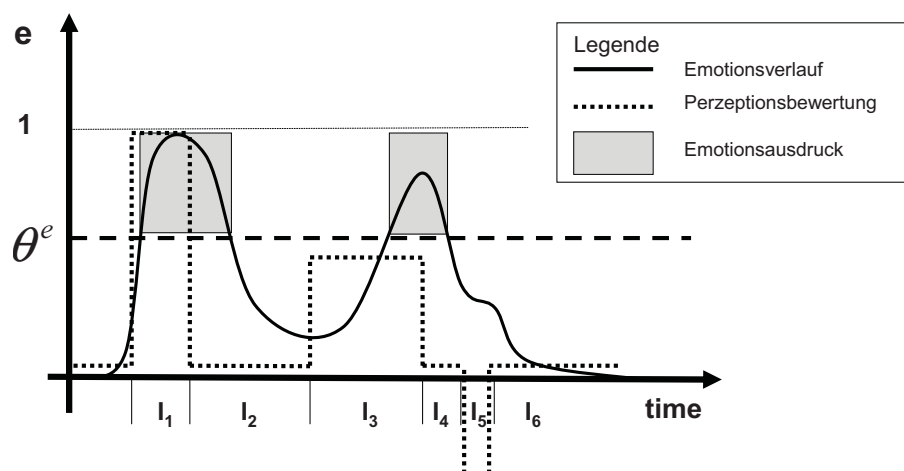


Abbildung 4: Zeitliche Entwicklung der Emotionen

Der Anstieg einer positiven Emotion ist zu beobachten, wenn ein „positiver“ Stimulus erkannt wird (siehe Intervall I_1, I_3). Falls das Emotionssystem im Zusammenhang mit dem Bedürfnissystem betrachtet wird, kommt noch die Bedingung hinzu, dass ein entsprechendes Bedürfnis sich im unbefriedigten Zustand befindet. Ansonsten befindet sich das System in einem neutralen Zustand. Diese Tatsache lässt sich an einem psychologischen Beispiel mit der Emotion „Freude“ und dem Kommunikationsbedürfnis leicht erklären. Ein Mensch freut sich, einen interessanten Kommunikationspartner zu sehen und mit ihm kommunizieren zu können, falls das Bedürfnis nach Kommunikation vorhanden ist. In diesem Fall sind beide Bedingungen vorhanden: Erstens befindet sich das Kommunikationsbedürfnis im unbefriedigten Zustand, und zweitens wird der Stimulus „Kommunikationspartner“ wahrgenommen. In der Annahme, dass nur eine Bedingung erfüllt ist, gibt es zwei unterschiedliche Möglichkeiten. Bei fehlendem Kommunikationsbedürfnis freut man sich kaum, einen möglichen Kommunikationspartner zu treffen. In dieser Situation zeigt man den neutralen Gesichtsausdruck, was in diesem Fall als „keine Interesse an

Kommunikation“ interpretiert werden könnte. Das vorhandene Kommunikationsbedürfnis ohne den gewünschten Kommunikationspartner wird von einem neutralen oder gar einem verärgerten Emotionszustand begleitet.

Wenn kein positiver Stimulus wahrgenommen wird (siehe Intervalle I_2, I_4 und I_6), nimmt eine positive Emotion mit der Zeit automatisch ab. Durch eine negative Perception kann der Abfall des Emotionsverlaufs entweder beschleunigt oder verlangsamt werden (Intervall I_5).

Im Unterschied zu den Emotionen hat ein Bedürfnis d einen unteren Schwellwert θ_d^l und einen oberen Schwellwert³ θ_d^u , die entsprechend aus den Intervallen $[-1, 0[$ und $]0, 1]$ gewählt werden ($-1 \leq \theta_d^l < 0$ und $0 < \theta_d^u \leq 1$). Durch diese beiden Schwellwerte wird der homöostatische Bereich definiert, in dem die Bedürfniszustände ausgeglichen sind. Der Wert 1, der die maximale Bedürfnisstärke darstellt, repräsentiert ein sehr großes Streben nach Bedürfnisbefriedigung, während -1 dem überbefriedigten Bedürfniszustand entspricht.

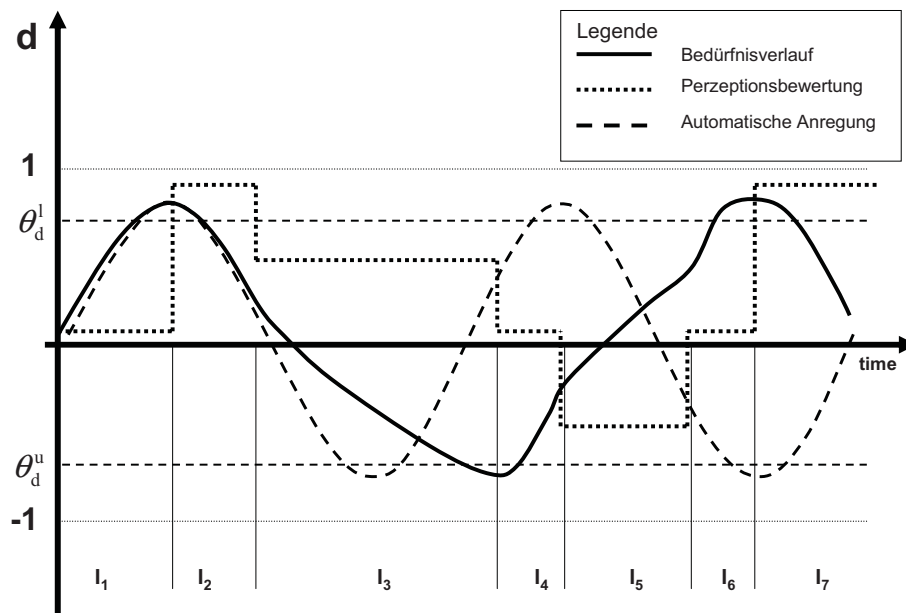


Abbildung 5: Zeitliche Entwicklung der Bedürfnisse

Damit man ein proaktives Verhalten realisieren kann, werden die Bedürfnisse wie eine Schwingung realisiert: Sie wechseln ihren Abfall mit dem Anstieg periodisch mit der Zeit. Dieser Verlauf wird durch eine Funktion verursacht, die Anregungsfunktion genannt wird. Die Anregungsfunktion ist als gestrichelte Linie in Abbildung 5 dargestellt. Die Priorisierung von Bedürfnissen verhindert, dass MEXIs ausgewählte Verhaltensweisen ständig wechseln, um konkurrierende Bedürfnisse zu befriedigen.

Einige Stimuli können den Anstieg der Anregungsfunktion entweder beschleunigen oder verlangsamen (Intervall I_5). Andere Stimuli, die zu Bedürfnisbefriedigung führen, verursachen einen „normalen“ Abfall der Anregungsfunktion (Intervalle I_2 und I_7). Außerdem gibt es auch solche Stimuli, die in der Lage sind, den Abfall entweder zu beschleunigen oder zu verlangsamen (Intervall I_3). In anderen Zeitintervallen wird MEXIs Perception keinen Einfluss auf die Anregungsfunktion haben.

7.2 Emotionsbewertung in der Emotion Engine

Laut der Definition der Roboteremotionalität wird der emotionale Zustand eines Roboters von äußeren und inneren Faktoren beeinflusst. Äußere Faktoren sind Reize, die ihre Quelle außerhalb des Systems haben bzw. aus der äußeren Umgebung wahrgenommen werden. Unter inneren Faktoren sind solche Einwirkungen zu verstehen, die

³Für MEXIs Bedürfnisse gilt: $|\theta_d^l| = |\theta_d^u|$.

im Inneren des Systems entstehen und als innere Systemzustände interpretiert werden können. Psychologische Untersuchungen haben gezeigt, dass zu den wichtigsten inneren Reizen, die den Ablauf von Emotionen beeinflussen, vor allem Bedürfnisse gehören. Die Befriedigung von Bedürfnissen wird beispielsweise als angenehm empfunden bzw. von einer positiven Emotion begleitet. Falls die Bedürfnisse längere Zeit nicht befriedigt werden können, führt dies zur Auslösung von negativen Emotionen. Emotionen stellen also das Ergebnis einer komplexen Informationsverarbeitung dar, die im Vergleich zu Bedürfnissen nicht nur die kognitive Bewertung der äußeren Reize umfasst, sondern auch die Schritte der kognitiven Bewertung des Verlaufs von Bedürfnissen beinhalten, die den emotionalen Prozess als innere Reize beeinflussen.

Die Emotion Engine bearbeitet mehrere Aufgaben von der Perzeptionsbewertung bis zur Konfiguration der Verhaltensweisen. Ihre Hauptaufgabe besteht jedoch darin, den inneren Zustand von MEXI zu verwalten, der durch aktuelle Werte seiner Emotionen und Bedürfnisse repräsentiert wird [EKK08b].

Die Emotion Engine empfängt Daten vom Wahrnehmungsmodul, die zunächst bedürfnis- bzw. emotionsspezifisch bewertet werden. Danach wird MEXIs Bedürfniszustand mit Hilfe der bedürfnisspezifischen Perzeptionsbewertung aktualisiert. Für jedes Bedürfnis wird dabei eine aktuelle Stärke anhand der entsprechenden Anregungsfunktion berechnet, die vom vorherigen Bedürfniswert bzw. seiner Änderung rekursiv beeinflusst wird.

Als nächstes wird MEXIs Emotionszustand mittels der emotionsspezifischen Perzeptionsbewertung und des aktuellen Bedürfniszustandes aktualisiert. Im Unterschied zum Bedürfniszustand, der durch mehrere nebeneinander existierende Einzelbedürfnisse gekennzeichnet ist, wird der Emotionszustand in eine bestimmte Emotionskategorie der Basisemotionen oder ihrer Mischung anhand des AFEMs eingeordnet.

Im letzten Schritt werden die Verhaltensweisen des Roboters konfiguriert. Die Konfiguration erfolgt durch Verstärkungsfaktoren, die für jede Verhaltensweise in Abhängigkeit vom aktuellen Emotions- bzw. Bedürfniszustand des Roboters berechnet werden. Die nachfolgende Auswahl von Verhaltensweisen zur Darstellung adäquater emotionaler Reaktionen findet innerhalb des Verhaltenssystems statt.

8 Wechselwirkungen zwischen Erkennung, Bewertung und Darstellung von Emotionen

Zwischen den Grundkomponenten der Roboteremotionalität Emotoinserkennung, -bewertung und -darstellung bestehen verschiedene Wechselwirkungen. Durch Erkennung von unterschiedlichen Stimuli, die MEXI in seiner Umgebung wahrnimmt, verändern sich seine inneren Emotions- und Bedürfniszustände. Unter Einbeziehung der eingehenden bedürfnisspezifischen Stimulusinformationen werden zunächst die Bedürfnisse aktualisiert. MEXIs aktualisierte Bedürfnisse in Kombination mit emotionsrelevanten Stimulusinformationen wirken auf seine Emotionen und passen sie situationsabhängig an. Angepasste Emotionen und aktualisierte Bedürfnisse beeinflussen die Auswahl der Verhaltensweisen, die zu emotions- und bedürfnisspezifischen Handlungstendenzen und Sprachäußerungen führt.

Falls MEXI beispielsweise Bedarf nach Kommunikation hat und einen potenziellen Kommunikationspartner erkennt, so ist MEXI glücklich. Er reagiert daher auf seinen menschlichen Gegenüber und folgt ihm durch Augen- und Kopfbewegungen. Dauert ihm die Kommunikation mit dem menschlichen Partner zu lange, wird MEXI ärgerlich und ignoriert sofort den Menschen. Das zeigt MEXI durch einen ärgerlichen Gesichtsausdruck, den er mittels entsprechender Mund- und Ohrenbewegungen erzeugt, und durch Drehen seines Kopfes in eine dem Menschen entgegengesetzte Richtung. Die Emotion „Ärger“ nimmt automatisch ab, nachdem ihre Intensität den maximalen Wert erreicht hat.

MEXI ist so realisiert, dass er einen freundlichen Charakter besitzt. So freut sich MEXI, wenn er ein Gesicht mit einem neutralen oder freudigen Gesichtsausdruck erkennt. Ebenso reagiert MEXI mit Freude auf seine Spielzeuge, die er in seinem Blickfeld sieht. MEXIs Emotion „Freude“ steigt in diesen Fällen an, wobei die Intensitäten aller anderen negativen Emotionen schnell abfallen, bis sie auf dem Nullpunkt ankommen und dort verbleiben. Allerdings sinkt MEXIs Emotionswert für „Freude“ sofort ab, falls während der Kommunikation mit einem Menschen Stimuli wie „Trauer“, „Angst“ oder „Ärger“ erkannt werden. Sein eigener Emotionszustand entspricht nach

kurzer Zeit einem neutralen Zustand. In dieser Situation versucht MEXI, seinen Kommunikationspartner durch unterschiedliche Aktivitäten „aufzumuntern“ oder „zu beruhigen“.

9 Einordnung MEXI zu den existierenden Ansätzen

Es sind noch keine Ansätze bekannt, die alle Aspekte der emotionalen Robotik wie Emotionserkennung, Emotionsmodellierung, Emotionsbewertung und Emotionsdarstellung in einem sogenannten ganzheitlichen System zusammenführen. Es gibt allerdings Anwendungen, die die beschriebenen Aspekte zumindest teilweise umsetzen. Solche Anwendungen findet man entweder im Bereich der sozialen Robotik oder im Forschungsgebiet der virtuellen Agenten.

Als weltweit bekannteste Roboter aus dem Bereich der sozialen Robotik können die Hunderoboter AIBO der Firma Sony genannt werden. Die AIBO-Roboter sind in der Lage, mit Hilfe einer Vielzahl der Sensoren und einer Kamera ihre Umwelt akustisch, visuell und haptisch zu erfassen und eine zur wahrgenommenen Umgebung passende Verhaltensweise zu berechnen. Das ermittelte Verhalten wird sehr „hundenähnlich“ anhand der Aktoren nach außen demonstriert. Einige primäre Emotionszustände wie Freude, Trauer oder Ärger können allerdings nur mittels farbiger LEDs ausgedrückt werden. Das humanoide Modell QRIO der Firma Sony gehört zu der Nachfolgegeneration des AIBO. Ausgestattet mit einem menschenähnlichem Körper ist der Roboter in der Lage, Treppen zu steigen oder zu tanzen. Mit seiner akustischen und visuellen Wahrnehmung ist es QRIO möglich, bekannte Personen wieder zu erkennen. Wird eine Person vom Roboter als unbekannt klassifiziert, werden die Daten von dieser Person selbständig in die Datenbank aufgenommen und bei der nächsten Kommunikation verwendet. Allerdings kann QRIO seine Aktionen nur eine sehr kurze Zeit und in begrenztem Ausmaß ausführen. Einen richtigen Gesichtsausdruck sucht man auf seinem Kugelkopf auch vergeblich. Ähnlich wie AIBO kann der Roboter QRIO seine emotionalen Zustände nur mittels farbiger LEDs ausdrücken.

Im Gegensatz zu den Entwicklern der Firma Sony konzentriert sich der Roboterforscher Hiroshi Ishiguro von der Universität Osaka auf die Realisierung menschenähnlicher Gesichtsausdrücke für verschiedene emotionale Zustände. Die Roboter, die von seiner Forschungsgruppe IRL (Intelligent Robotics Laboratory) geschaffen werden, werden Androide genannt. Einer der berühmtesten ist der Android Repliee Q1, der mit einem menschenähnlichen Körper, Händen, Kopf, Silikonhaut und Haaren ausgestattet ist und einer Frau sehr ähnlich sieht. Angetrieben von insgesamt 31 Aktoren kann der Android menschliche Körperbewegungen und Gesichtsmimik täuschend echt nachahmen. Bei den Forschern der Universität Osaka, die sich auf die Darstellung menschenähnlicher Verhaltensweisen spezialisiert haben, kommen zunächst solche Aspekte der emotionalen Mensch-Machine-Interaktion wie Erkennung oder Bewertung fremder Gefühle nicht in Betracht.

Zu den bekanntesten Ansätzen aus der emotionalen Robotik gehört der Roboterkopf Kismet. KISMET ist ein weit entwickeltes System, das sich sowohl mit der Emotionsregulation als auch mit der Ausdrucksfunktion des emotionalen Zustandes beschäftigt. Allerdings bleibt der erste Punkt der Erkennungsfunktion in diesem Projekt offen. Kismet erkennt keine Emotionen im engeren Sinne, sondern extrahiert stattdessen Parameter aus dem Sprachsignal, mit deren Hilfe er Rückschlüsse auf die Intention des Sprechers zieht. So erkennt Kismet beispielsweise, ob sein menschliches Gegenüber ihn lobt oder ein Verbot ausspricht und reagiert darauf entsprechend. Die Forscher stellen auch nicht die Anforderung, dass er durch adäquates Verhalten Konflikte beseitigen oder stressige Situationen umgehen kann.

Im Vergleich zu vielen existierenden Ansätzen setzt der Roboterkopf MEXI alle Aspekte der Roboteremotionalität um und kann sogar als emotional kompetent bezeichnet werden. MEXI ist in der Lage, sowohl die visuelle Emotionserkennung als auch die Emotionserkennung aus der Sprache durchzuführen. Im Gegensatz zu den anderen Ansätzen, die auf die Erkennung einiger ausgewählten Basisemotionen abzielen, ist es im MEXI-System vorgesehen, nicht nur die Basisemotionen sondern auch deren Mischungen zu erkennen. Dafür wird das für MEXI entwickelte adaptive Fuzzy Emotionsmodell (AFEM) benutzt, das es ermöglicht, ausgewählte Basisemotionen und deren Kombinationen innerhalb der Emotionserkennung zu klassifizieren.

Der innere emotionale Roboterzustand, der von der erkannten Emotion beeinflusst wird, kann mittels dieses Emotionsmodells bestimmt werden und anhand seiner Aktoren repräsentiert werden. Im MEXI-System sind außerdem

die Mechanismen eingearbeitet, die es ihm ermöglichen, auf die erkannten Emotionen seines menschlichen Gegenübers mit adäquatem Verhalten zu reagieren. So kann MEXI beispielsweise verärgerte Menschen beruhigen, traurige Menschen ermutigen oder die Freude seines Gegenübers einfach unterstützen.

Zusammenfassend stellt MEXI einen Ansatz vor, der ein neues Klassifikationssystem mit unterschiedlichen Modalitäten emotionsbehafteter Eingaben kombiniert und alle vier Grundkomponenten der Roboteremotionalität wie Emotionserkennung, Emotionsmodellierung sowie Bewertung und äußere Repräsentation des eigenen Emotionszustandes vereinigt (siehe Abbildung 1). Somit trägt der in dieser Arbeit vorgestellte Ansatz zur Entwicklung eines ganzheitlichen emotionalen Systems bei.

10 Zusammenfassung und Ausblick

Im Rahmen dieser Arbeit wurde ein ganzheitliches emotionales System entwickelt, das alle emotionalen Aspekte der Mensch-Roboter-Interaktion, die Emotionsmodellierung, -erkennung, -bewertung und -darstellung, miteinander verbindet und am Beispiel des Roboterkopfs MEXI in einer emotionsgesteuerten verhaltensbasierten Architektur zusammenführt.

Zahlreiche Erfahrungen mit MEXI auf verschiedenen öffentlichen Ausstellungen und Messen⁴ haben gezeigt, dass Menschen den Roboterkopf als einen möglichen Kommunikationspartner und nicht nur als eine Maschine empfinden. Für Besucher der Ausstellungen wirkte MEXI wie ein emotionaler Partner, der in der Lage ist, nicht nur seine eigenen Emotionen zu zeigen, sondern auch die Emotionen des menschlichen Gegenübers durch seine emotional-kompetente Art und Weise, mit Menschen zu kommunizieren, positiv zu beeinflussen. Gleichzeitig brachten die Erfahrungen auch neue Fragestellungen für die weitere Entwicklung solcher emotionsfähiger Maschinen wie MEXI mit.

Eine mögliche Erweiterung der Wahrnehmung von MEXI ist beispielsweise die Kombination der sprachbasierten Emotionserkennung mit der mimikbasierten Emotionserkennung. Informationen aus unterschiedlichen Sensoren sollten zur Erhöhung der Sicherheit der gewonnenen Aussage kombiniert und zusammengeführt werden. Die Verknüpfung der Ergebnisse beider Systeme VISBER und PROSBER ermöglicht dann die Verbesserung der Gesamterkennungsrate. Ein weiterer Ansatz wäre die Realisierung einer Personenerkennung. Hierfür würde MEXI im ersten Schritt der Emotionserkennung die konkrete Person erkennen und erst dann die Emotionserkennung mit entsprechenden personenabhängigen Regelbasen durchführen. Dieser Ansatz wäre bei einem Szenario denkbar, in dem MEXI beispielsweise die Rolle eines Haushaltsroboters mit einer begrenzten Anzahl möglicher Kommunikationspartner übernimmt, mit denen er öfter in Kontakt tritt.

Weiterhin ist es wünschenswert, die Emotion Engine noch zu verfeinern. So ist es sinnvoll, ein Gedächtnis zur Speicherung von Erfahrungen mit bestimmten inneren und äußeren Reizen zu realisieren, die beim Auftreten der gleichen emotionsspezifischen Situation wieder zur Verfügung gestellt werden. Durch Lernverfahren könnte MEXI dann die richtige Reaktion auf Emotionen seines menschlichen Gegenübers erlernen und in entsprechenden Situationen wiederholen.

Eine sinnvolle Ergänzung zu MEXI wäre eine „körperliche“ Konstruktion, so dass der emotionale Roboter sich frei im Raum bewegen könnte und mit einem beliebigen Kommunikationspartner in Kontakt treten könnte. Der Gesamteindruck des Roboters als möglicher emotionaler Partner würde durch diese Option drastisch gesteigert.

Allerdings sollten parallel zu der technischen Entwicklung der emotionalen Robotik auch moralische und ethische Aspekte untersucht werden. Eine enge Zusammenarbeit der Roboterentwickler mit Psychologen und Gesellschaftstheoretikern sollte stattfinden, um Aspekte wie den Aufbau von sozialen Bindungen zum Roboter, psychologische Wirkung emotionaler Systeme auf menschliche Benutzer oder fehlende zwischenmenschliche Kommunikation im Fall einer übermäßigen Roboterverbreitung betrachten zu können.

⁴C-BIT 2003 in Hannover, Sonderausstellung „Computerwelten - Vom Abakus zum Avatar“ 2003 SiemensForum Erlangen, Kölner Abiturientenmesse 2004, Tag der offenen Tür 2005 Universität Paderborn, Hausmesse Siemens Innovationstag 2006 in Hannover, Ausstellung „Paderborn überzeugt“ NRW-Tag 2007 in Paderborn sowie verschiedene Sonderausstellungen im Heinz Nixdorf MuseumsForum in Paderborn.

Literatur

- [Bur01] F. Burkhardt. *Simulation emotionaler Sprechweise mit Sprachsyntheseverfahren, Dissertation an der TR–Berlin*. Shaker Verlag, 2001.
- [CHFT06] Y. Chang, C. Hu, R. Feris, and M. Turk. Manifold based analysis of facial expression. 24(6):605–614, 2006.
- [EF78] P. Ekman and W. V. Friesen. *Facial Action Coding System*. Consulting Psychologists Press Inc., California, 1978.
- [EKK05] N. Esau, L. Kleinjohann, and B. Kleinjohann. An adaptable fuzzy emotion model for emotion recognition. In *Proceedings of the 4th Conference of the European Society for Fuzzy Logic and Technology (EUSFLAT–LFA 2005)*, Barcelona, Spain, 2005.
- [EKK08a] N. Esau, L. Kleinjohann, and B. Kleinjohann. Emotional competence in human–robot communication. In *ASME 2008 International Design Engineering Technical Conferences & Computers and Information in Engineering Conference (IDETC/CIE 2008)*, New York, USA, 2008.
- [EKK08b] N. Esau, L. Kleinjohann, and B. Kleinjohann. Integrating emotional competence into man–machine collaboration. In *IFIP Conference on Biologically Inspired Collaborative Computing (BICC 2008)*, Milan, Italy, 2008.
- [EWKK07a] N. Esau, E. Wetzel, L. Kleinjohann, and B. Kleinjohann. A fuzzy emotion model and its application in facial expression recognition. In *ASME 2007 International Design Engineering Technical Conferences & Computers and Information in Engineering Conference (IDETC/CIE 2007)*, Las Vegas, USA, 2007.
- [EWKK07b] N. Esau, E. Wetzel, L. Kleinjohann, and B. Kleinjohann. Real–time facial expression recognition using a fuzzy emotion model. In *IEEE International Conference on Fuzzy Systems (FUZZ–IEEE 2007)*, London, UK, 2007.
- [Gol96] D. Goleman. *Emotionale Intelligenz*. München, Hanser, 1996.
- [KB03] D. Kim and Z. Bien. Fuzzy neural networks (fnn)–based approach for personalized facial expression recognition novel feature selection method. In *Proceedings of the IEEE International Conference on Fuzzy Systems*, volume 2, pages 908–913, 2003.
- [Kle01] B. Kleinjohann. *MEXI, Roboter mit Gefühl*, pages 33–40. Jahresbericht, C–LAB, Paderborn, 2001.
- [KNP06] I. Kotsia, N. Nikolaidis, and I. Pitas. Fusion of geometrical and texture information for facial expression recognition. In *ICIP*, pages 2649–2652, 2006.
- [Kos92] B. Kosko. *Neural Networks and Fuzzy Systems; A Dynamical Systems Approach to Machine Intelligence*. Prentice Hall, Englewood Cliffs, NJ, 1992.
- [MSC04] J.D. Mayer, P. Salovey, and D.R. Caruso. Emotional intelligence: Theory, findings and implications. *Psychological Inquiry*, (15):197–215, 2004.
- [Nil98] N. J. Nilsson. Artificial intelligence – a new synthesis. In *Proceedings of the International Conference on Spoken Language Processing (ICSLP 96)*. Morgan Kaufmann Publishers, 1998.
- [Plu80] R. Plutchik. *Emotion: A Psychoevolutionary Synthesis*. Harper and Row, Publishers, New York, 1980.
- [PP06] M. Pantic and I. Patras. Dynamics of facial expression: Recognition of facial actions and their temporal segments from face profile image sequences. 36(2):433–449, 2006.
- [SBHN04] R. Schweiger, P. Bayerl, and H. Heiko Neumann. *Neural Architecture for Temporal Emotion Classification*, pages 49–52. Springer Berlin / Heidelberg, Berlin/Heidelberg, 2004.

[SM90] P. Salovey and J. D. Mayer. Emotional intelligence. *Imagination, Cognition and Personality*, (9):185–211, 1990.