



X . media . press



Christian Lindner (Hrsg.)

X.media.press ist eine praxisorientierte Reihe des Springer-Verlags zur Gestaltung und Produktion von Multimedia-Projekten sowie von Digital- und Printmedien.

Avatare

**Digitale Sprecher
für Business und Marketing**



Springer

Mit Hand und Fuß – Die Bedeutung der nonverbalen Kommunikation für die Emotionalisierung von Dialogführungssystemen

GEORG TROGEMANN

1 Agenten und Avatare

Agenten und Avatare sind viel versprechende und gegenwärtig sehr populäre Konzepte für die Organisation und Realisierung neuartiger Benutzerhilfen und grafischer Bedienoberflächen. Darüber hinaus gelten Agenten als geeignetes theoretisches Fundament, um intelligentes und komplexes Systemverhalten zu modellieren. Festzustellen ist allerdings, dass es derzeit keine allgemein anerkannte Definition des Begriffs „Agent“ gibt. Der sehr lebendige wissenschaftliche Diskurs zu dem Thema hat verschiedene, teilweise unvereinbare Definitionen hervorgebracht. Eine sehr allgemeine Eingrenzung lautet:

„An agent is a computer system that is situated in some environment, and that is capable of autonomous action in this environment in order to meet its design objectives.“²

Unumstritten und allen Definitionsversuchen gemein ist die Bedeutung des Begriffes „Autonomie“. Damit wird das Definitionsproblem allerdings nur verlagert, da dieser Begriff genauso schwierig zu präzisieren ist wie der des Agenten. In einer ersten Näherung können wir Autonomie als die Fähigkeit eines Agenten verstehen, ohne die Intervention eines Menschen oder eines anderen Systems agieren zu können. Entscheidend ist der grundlegende Unterschied zwischen Agenten und dem derzeitig vorherrschenden Software-Konzept des „Werkzeugs“. Während das Werkzeug vom Benutzer geführt wird, handelt der Agent eigenständig, er agiert im Auftrag seines Besitzers.

Im vorliegenden Beitrag soll nicht der Gesichtspunkt der Autonomie im Vordergrund stehen, sondern eine anderes wesentliches Merkmal von Agenten: ihre „sozialen Fähigkeiten“. Agenten sind Software-Systeme, die

² Einen guten Überblick zur Theorie intelligenter Agenten gibt Weiss (2000), aus dessen Buch auch diese Definition entnommen ist.

menschliches Verhalten simulieren. Ihr Verhalten wird durch Begriffe beschrieben, die normalerweise menschlichen Eigenschaften vorbehalten sind, z. B. der Agent „verfolgt“ ein Ziel, er „glaubt“ an etwas, er „verhandelt“, oder er „probiert“ etwas aus. In virtuellen Umgebungen wird es deshalb schwierig, zwischen Menschen und Agenten zu unterscheiden, da beide als so genannte „Avatare“ repräsentiert werden.

Die Bezeichnung Avatar stammt ursprünglich aus dem Hindi und charakterisiert die Inkarnation eines Gottes. Ein Avatar übernimmt im Kontext der neuen Technologien die Aufgabe der Repräsentation eines „inhabitants“. In virtuellen Realitäten steht er für jemand anders, seinen realen Besitzer, auf den er verweist. Diese strenge Definition, dass sich hinter einem Avatar immer eine reale Person verbirgt, wurde allerdings inzwischen im praktischen Gebrauch aufgeweicht. Auch virtuelle Verkäufer und virtuelle Nachrichtensprecher etc., die oft keine Realpersonen repräsentieren, werden als Avatare bezeichnet. Das Spektrum reicht deshalb vom virtuellen Stellvertreter des Benutzers als Protagonist in einem Videospiel, über virtuelle Assistenten, die Hilfestellung bei der Bedienung komplizierter technischer Systeme leisten (z. B. Videorecorderassistent, Fahrzeugassistent), bis hin zum animierten Markenrepräsentanten einer Firma im Internet. Avatare stehen also nicht nur für Personen, sondern auch für Produkte, Ideen, Marken, Firmen, Inhalte und Funktionen. Sie sind Repräsentanten im allgemeinen Sinn.

Zusammenfassend ist ein Avatar charakterisiert durch:

- einen Besitzer,
- einen Körper (z. B. Visualisierung und Sonifikation),
- durch Kommunikationsmöglichkeiten und
- Autonomie (Intelligenz).

Jeder Avatar kann dabei aus mehreren Agenten zusammengesetzt sein. Sofern Avatare reale Personen repräsentieren, sollten sie mit Hilfe ihrer Agenten in der Lage sein, ihren Besitzer – sobald dieser das System betritt – zu identifizieren und zu visualisieren. Um seine Aufgaben zu erfüllen und die im Design des Avatars festgelegten Ziele zu erreichen, muss er in der Lage sein, mit seiner Umgebung (der realen wie der virtuellen) und mit anderen Agenten und Avataren zu kommunizieren. Alternativ zum Begriff Autonomie wird oft Intelligenz als Anforderung an Agenten genannt.³ Ohne Intelli-

³ Die technischen Bedeutungen der Begriffe Autonomie und Intelligenz sind allerdings streng auseinander zu halten. Technische Systeme können autonom sein, ohne intelligent zu sein, und umgekehrt.

genz kann auch der berühmte Heizungsthermostat als Agent gelten und der Drehknopf, der zugleich sichtbare Repräsentation und haptisches Interface ist, als Avatar.

Erst intelligente Agenten und Avatare sind in der Lage, ihre „Umwelt“ wahrzunehmen, und geeignet – d.h. im Sinne ihrer vorgegebenen Ziele auf – Veränderungen zu reagieren (reactivity). Intelligente Agenten und Avatare sind (im Gegensatz zu Heizungsthermostaten) in der Lage zielgerichtetes Verhalten zu zeigen, indem sie selbst die Initiative übernehmen (pro-activeness). Wie bereits oben erwähnt, ist eine weitere entscheidende Forderung, die an intelligente Avatare gestellt wird, „soziale Kompetenz“ (social ability). Unter sozialen Fähigkeiten von Avataren wollen wir mehr verstehen als die übliche technische Kommunikation, d.h. die Fähigkeit, Bit-Ströme auszutauschen und darauf zu reagieren. In der realen Welt müssen wir, um unsere Ziele zu erreichen, kooperieren und verhandeln. Wir können dabei nicht sicher sein, dass unsere Kooperations- und Kommunikationspartner uns verstehen oder gar die gleichen Ziele verfolgen. Es werden – im Gegensatz zur technischen Kommunikation – immer auch Informationen zwischen Sender und Empfänger ausgetauscht, die nicht für beide Seiten verbindliche Bedeutungen haben, sondern von den Kommunikationspartnern unterschiedlich interpretiert werden können. Die Emotionen des Gegenüber, beziehungsweise unsere Einschätzung davon und unsere Reaktion darauf, spielen für das Gelingen der Kommunikation eine entscheidende Rolle. Diese Form der Unzuverlässigkeit der menschlichen Kommunikation ist aber gerade ein zentraler Aspekt der technischen Kommunikation mit autonomen Avataren.

2 Die Macht des Visuellen: Oder warum der User glaubt was er sieht

Das Bild von Norman Rockwell (siehe Abbildung 1) zeigt eindrucksvoll, wie sehr menschliche Kommunikation von Emotionen und Missverständnissen geprägt ist. Die Fragen lauten: Kann es gelingen, die menschliche Kommunikationskompetenz auf Maschinen zu übertragen, um zielgerichtete Arbeitsprozesse zu verbessern? Kann nonverbale Kommunikation funktionalisiert werden, ohne die Fallstricke und Unzulänglichkeiten menschlicher Kommunikation zu implementieren? Werden wir Avatare überhaupt als Verhandlungs- und Geschäftspartner akzeptieren? Können sie so glaubwürdig sein, dass wir ernsthaft mit ihnen verhandeln und Geschäfte mit ihnen abschließen? Psychologische Untersuchungen zeigen zumindest, dass wir auf nonverbale Stimuli (Gestik, Mimik, Haltung, Blick) von virtuellen Darstellern



Abb. 1. Norman Rockwell, Gossip

genauso reagieren wie auf reale. Wird das nonverbale Verhalten zweier Dialogpartner transkribiert und auf virtuelle Figuren übertragen, dann werden Testpersonen die virtuelle Dialogsituation genauso einschätzen wie die Real-situation.

Es gibt auch andere Forschungen die darauf hinweisen, dass wir technischen Systemen sehr bereitwillig menschliche Eigenschaften zuschreiben. Dennett (1987) unterscheidet drei grundsätzliche prädiktionale Einstellungen, die Menschen gegenüber Systemen einnehmen. Der Beobachter eines

Systems wird genau die Einstellung wählen, die für das beobachtete Verhalten die einfachste und zuverlässigste Erklärung liefert.

2.1 Physikalische Einstellung

Gegenüber natürlichen Systemen erweist es sich in der Regel als praktikabel, das Verhalten gemäß physikalischer Eigenschaften und Gesetze vorauszusagen. Die Kollision von Billardkugeln kann zum Beispiel durch die Anwendung von Reflexionsgesetzen am plausibelsten beschrieben werden.

2.2 Design Einstellung

Die Verhaltensweisen von Artefakten, zum Beispiel Fahrzeugen, Küchenmaschinen oder anderen einfachen Werkzeugen, können üblicherweise auf der Basis ihrer Funktionen vorhergesagt werden. Die Frage, wofür bestimmte Teile oder ein System als Ganzes gemacht ist, liefert meist die nötigen Hinweise, um das Verhalten des Artefaktes vorherzusehen

2.3 Intentionale Einstellung

Ab einer bestimmten Komplexität eines Systems greifen die ersten beiden Erklärungsprinzipien nicht mehr. Der Beobachter kann nur noch raten, wie sich das System im nächsten Moment verhalten wird. In diesen Fällen neigt der Mensch dazu, rationales Handeln hinter dem beobachtbaren Verhalten zu vermuten, d. h. dem System eine Intention zu unterstellen.

Im Alltag wechseln wir ganz natürlich und spielerisch zwischen den einzelnen Erklärungsformen. Auch innerhalb eines System werden wir, wenn die Situation es nahe legt, zwischen den Beschreibungsprinzipien wechseln. Wer kennt nicht die Neigung, mit einem Werkzeug zu hadern, als wäre es ein unwilliger Kollege, sobald es sich nicht funktionsgerecht verhält. Wird die erwartete und bisher immer bestätigte Funktion des Artefaktes aus unerwartlichen Gründen verweigert, dann wird dem leblosen Gegenstand schnell böse Absicht unterstellt und damit der Sprung von der Design-Einstellung zur intentionalen Einstellung vollzogen. Wir personifizieren Objekte, d. h. wir statten unbelebte Erscheinungen mit Eigenschaften, Gefühlen, Absichten und Handlungsweisen belebter Wesen aus.

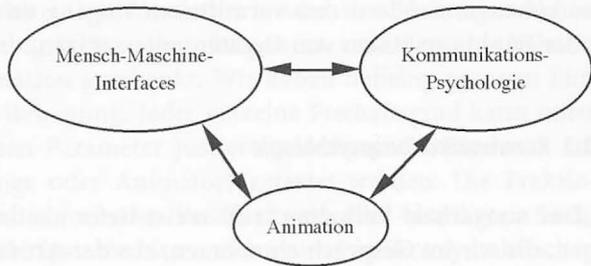
Die intentionale Einstellung gegenüber ein natürliches oder künstliches System hat zwei Quellen, aus denen sie genährt wird, einerseits das nichtkalkulierbare Verhalten des Systems, andererseits die äußere Erscheinung des Systems. Dinge, die wie Lebewesen aussehen, schreiben wir normalerweise solange alle Attribute von Lebewesen zu, solange ihr Verhalten diese Zuschreibung bestätigt. Ohne zusätzliches Wissen ist dieses Vorgehen gewissermaßen die effektivste Arbeitshypothese.⁴ Unsere Neigung, komplexen Systemen, die zudem auch äußerlich noch wie Menschen erscheinen, auch alle anderen menschlichen Attribute zuschreiben, d. h. sie vollständig zu personifizieren, wird im Rahmen der so genannten anthropomorphen Schnittstellen systematisch angewandt und untersucht. Nach dem Vorbild der menschlichen Face-to-Face-Kommunikation – d. h. durch in Verhalten und Aussehen anthropomorphisierte Assistenten – soll die vorhandene Interaktionskompetenz des Anwenders genutzt und die Bedienung von Computern erleichtert und verbessert werden. Der Benutzer soll den Computer als gleichberechtigten Partner erfahren, der Aufgaben entgegennimmt und sie im Sinne des Benutzers erledigt. Der Assistent – selbst eine technische Komponente – wird als aktiver Vermittler zwischen den Menschen und die Funktionen der Technik gestellt. Während die Maschine bedient wird, das Werkzeug geführt wird und das Medium durchlässig ist und wirkt, wird mit dem Avatar kommuniziert. Die Funktion der Maschine ist es zu verarbeiten, die des Werkzeugs zu bearbeiten und die des Mediums zu informieren und zwischen menschlichen Kommunikationspartnern zu vermitteln. Die Funktion des Avatars dagegen ist die aktive Mitgestaltung der Kommunikation. Der Computer ist hier gleichzeitig Sender, Empfänger und Medium.

3 Nonverbale Kommunikation

Die Wirksamkeit von Avataren wird durch ihr Aussehen und ihr nonverbales und kognitives Verhalten bestimmt. Neben der Sprache als Leitmedium bei der Realisierung anthropomorpher Interfaces sind das statische Aussehen und die nonverbale Kommunikation von entscheidender Bedeutung. Gestik, Mimik und Blickverhalten sind wesentliche Komponenten jeder Face-to-

⁴ Letztlich ist dies auch die Kernaussage des Turing-Tests. Wenn eine Testperson bestimmte Verhaltensweisen einer Maschine (z. B. intelligentes Verhalten) nicht mehr von denen eines Menschen unterscheiden kann, muss der Maschine das getestete (bisher menschliche) Attribut zugesprochen werden.

Abb. 2. Bereiche der nonverbalen Kommunikation



Face-Kommunikation. Nonverbale Kommunikation ist in mindestens drei Bereichen von Bedeutung, das zentrale Bindeglied zwischen diesen Bereichen ist der Computer.

Neben den beschriebenen Herausforderungen auf dem Gebiet der Mensch-Maschine-Interfaces gehört die nonverbale Kommunikation zum Gegenstand der Forschungen in der Kommunikationspsychologie. Außerdem spielt sie eine bedeutende Rolle im Bereich der Animation, für die traditionell mimisches und gestisches Verhalten wichtige Ausdrucksmittel sind. Insbesondere die 3D-Computeranimation zeigt natürlich großes Interesse an den Algorithmen der nonverbalen Kommunikation. Die drei Bereiche sind durch den Computer vermittelt und können deshalb nicht vollkommen losgelöst voneinander betrachtet werden (siehe Abbildung 2). Fortschritte in einem der Bereiche nutzen direkt auch den anderen beiden Forschungs- und Anwendungsgebieten. 3D-Animation gehört inzwischen zum Beispiel zum wichtigsten Werkzeug der Kommunikationspsychologie. Umgekehrt kommt psychologisches Wissen über nonverbale Kommunikation anthropomorphen Interfaces und Animationsprojekten zugute.

3.1 Mensch-Maschine Interfaces

Bei der Imitation von Face-to-Face-Kommunikation ist das Leitmedium die Sprache. Die Hinzunahme des nonverbalen Kanals soll vor allem die Zuverlässigkeit und den reibungslosen Dialogverlauf unterstützen. Nonverbales Verhalten kann damit prinzipiell *komplementär*, *supplementär*, *substituiv* oder *disparat* zur Sprache sein. Beim Einsatz nonverbaler Stimuli im Interface-Design steht zunächst der inhaltliche Aspekt der Kommunikation im Vordergrund und nicht der Beziehungsaspekt. Intuitive Kommunikation soll nicht dazu dienen, eine emotionale Beziehung zu virtuellen Assistenten

aufzubauen, sondern den vermittelten Zugang zu Informationen, Diensten oder Funktionalitäten von Geräten unterstützen.

3.2 Kommunikationspsychologie

„Das nonverbale Verhalten trifft meist tiefer als das Wort. Aus den Haltungen, die wir im Gespräch einnehmen, aus der Art und Weise, wie wir uns bewegen, zieht unser Gesprächspartner „unbewusste Schlüsse“. Deren Suggestivkraft ist so zwingend, dass man sich des Eindrucks kaum erwehren weiß. Das nonverbale Verhalten stellt daher das beziehungsstiftende Element in der zwischenmenschlichen Verständigung dar. . . . Um herauszufinden, auf welche spezifischen Elemente des nonverbalen Verhaltens der „unbewusste Schluss“ anspricht, wurde die Methode der Skriptanimation entwickelt“ (Frey 1999, S.93). Die Grundlage der Skriptanimation bildet die Transkription der Körperbewegung in eine dem Computer zugängliche Form.⁵ Das hierbei entstehende Zeitreihenprotokoll erlaubt es, das natürliche menschliche Bewegungsverhalten in seinem ganzen Detailreichtum auf dem Computer mit Hilfe der Methoden der 3D-Animation zu simulieren. Durch gezielte Variation einzelner Bewegungssegmente können dann jene Komponenten ermittelt werden, die für den Eindruck beim Betrachter verantwortlich sind.

3.3 Computeranimation

Zu den Stärken des Computers gehören Repetition und Fraktionierung. Ohne zu ermüden oder ungenau zu werden, ist die digitale Maschine in der Lage, Abläufe beliebig häufig ohne die geringste Abweichung zu wiederholen. Diese Stärke macht sich die Kommunikationspsychologie bei der Untersuchung nonverbalen Verhaltens zu Nutze und die 3D-Animation bei der schrittweisen (zwischen Parameterkorrektur und visueller Kontrolle ständig wechselnden) Erstellung von Animationen. Wozu kein Schauspieler fähig

⁵ Während Frey nach dem von ihm entwickelten *Berner System* notiert, sind auch andere Systeme denkbar, die der automatischen Generierung der Notation durch den Computer besser angepasst sind. Die Beschreibung mit Euler-Winkeln ist die Standardmethode der Computergrafik zur Beschreibung des Bewegungszustandes einer Gliederfigur.

ist, nämlich eine Geste, einen Gesichtsausdruck oder eine Haltung exakt zu wiederholen und in beliebig feinen Abstufungen zu variieren, das bekommen wir bei der 3D-Animation geschenkt. Wir haben beliebig genauen Einfluss auf jedes Detail der Bewegung. Jeder einzelne Freiheitsgrad kann unter Beibehaltung der restlichen Parameter justiert und dann der Eindruck auf den Betrachter (Psychologe oder Animator) getestet werden. Die Fraktionierbarkeit resultiert aus der formalen Beschreibung aller beteiligten Sachverhalte. Komplexe formale Systeme lassen sich immer in ihre syntaktischen Subsysteme zerlegen. Im Extremfall kann das zeitliche Verhalten jedes einzelnen Bits betrachtet werden. Was bei der natürlichen Face-to-Face-Kommunikation untrennbar verbunden ist, kann nun in alle Kanäle einzeln zerlegt und in allen Teilaspekten getrennt untersucht werden. So ist zum Beispiel erstmals die Trennung von Aussehen und Bewegung möglich, oder die exklusive Betrachtung und Analyse von Kopfbewegungen. Die statischen Merkmale einer Person können unabhängig von den dynamischen Merkmalen analysiert und auch synthetisiert werden, sie können aber auch kreuzweise vertauscht werden. Welchen Eindruck hat der Betrachter, wenn die Bewegungen eines Mannes einer weiblichen Figur aufgeprägt werden? Wie wirkt sich die starke Reduktion der statischen Merkmale einer Figur aus, wenn gleichzeitig das nonverbale Verhalten in seiner ganzen Komplexität beibehalten wird?

4 Von der Ausdruckspsychologie zur Eindruckspsychologie und zurück

Psychologische Untersuchungen menschlichen Kommunikationsverhaltens werden von zwei konkurrierenden Paradigmen geleitet: der Ausdruckspsychologie und der Wirkungsforschung (bzw. Eindruckspsychologie). „Speziell die Psychologie, die wie keine andere Disziplin in die Erforschung nonverbaler Verhaltensphänomene investiert hatte, war noch bis weit ins 20. Jahrhundert hinein – nicht anders als der Laie – von der uralten, buchstäblich in die vorgeschichtliche Zeit zurück reichenden Vorstellung beherrscht, die statischen und dynamischen Merkmale des menschlichen Erscheinungsbildes seien sozusagen ein „Spiegel der Seele“, die äußere Bewegung quasi ein Reflex der inneren Bewegung. Wem es gelänge, den Schlüssel zu dieser „Körpersprache“ zu finden, so hatte schon Aristoteles in seinem Werk *Physiognomonika* versichert, für den sei der Mensch ein offenes Buch“ (Frey 1999). Im Mittelalter wurde in Europa die Kunst des Gesichterdeutens von den Astrologen ausgeübt. Die Renaissance brachte einen bedeutenden Erneuerer der

Abb. 3. Beispiel aus: *De humana physiognomia* von Giovanni Battista Della Porta



Physiognomik hervor: Giovanni Battista Della Porta. Er stellte den traditionellen, astrologisch fundierten Ansatz in Frage und stellte die Physiologie und die Beobachtung in den Vordergrund. Außerdem verhalf er einem alten Ansatz zu neuem Leben, der Erlangung von Einsichten in den Charakter durch den Vergleich zwischen Menschen und Tieren (Landau 1995). Sah ein Mensch aus wie eine Ziege (siehe Abbildung 3), wurde er für ziegenähnlich oder dumm gehalten.

In den Jahren 1775–1778 erschien unter dem Titel *Physiognomische Fragmente zur Beförderung der Menschenliebe und Menschenkenntnis* die sehr einflussreiche Anleitung des Schweizer Pastors Johann Caspar Lavater, die einen Höhepunkt des Physiognomisierens in Europa auslöste. Aber auf diesem Höhepunkt meldete sich auch die Gegenbewegung erstmals zu Wort. So kommentierte der bedeutendste Gegner der Physiognomik die Arbeit Lavaters mit den Worten: „Wenn die Physiognomik das wird, was Lavater von ihr erwartet, so wird man die Kinder aufhängen, ehe sie die Taten getan haben, die den Galgen verdienen.“

Ein weiterer bedeutender Widerspruch kam von Helmholtz (1821–1894). Er stellte die damals revolutionäre These auf, dass die sinnstiftende Funktion bei der Deutung nonverbaler Stimuli nicht vom Akteur ausgeht, sondern vom Betrachter. Dennoch erlebte die Ausdruckspsychologie in den zwanziger und dreißiger Jahren unseres Jahrhunderts nochmal eine Hochblüte durch die Arbeiten von Klages, Lersch, Bühler und anderen. Erst nach einer geistigen Umorientierung (Frey 1999) wurde in Jahre 1973 anlässlich der Neuordnung des Studiengangs Diplompsychologie die Ausdruckspsychologie als akademische Disziplin abgeschafft. Die hohen Erwartungen konnten von den empirischen Daten letztlich nie eindeutig bestätigt werden. Erst

die Wirkungsforschung (bzw. Eindruckspsychologie) und der damit verbundene Paradigmenwechsel eröffnete neue Forschungsperspektiven. Endlich wurden die Thesen von Helmholtz und anderen ernst genommen. Gestik, Mimik und Körperhaltung öffnen kein Fenster zur Seele, sie sind Mittel zur Regulierung der zwischenmenschlichen Beziehung. Die Wirkungsforschung betrachtet deshalb nonverbales Verhalten vom Standpunkt der Kommunikation, d. h. die Wirkung nonverbaler Stimuli auf den Betrachter rückt ins Zentrum der Untersuchungen.

Für die Modellierung des nonverbalen Verhaltens von Avataren und virtuellen Assistenten haben die kontroversen Sichtweisen der Ausdruckspsychologie und der Wirkungsforschung beträchtliche Folgen. Ganz generell muss zeitliches Verhalten eines Avatars bei der Realisierung mit der „finite state machine“ Computer auf Zustandsfolgen abgebildet werden. Zu jedem Zeitpunkt befindet sich der Avatar in einem eindeutig definierten Zustand, aus dem zusammen mit der Eingabe (den Werten die an den Sensoren anliegen) der nächste Zustand, der auch die Parameter der Effektoren (Sprache, Gestik, Mimik) beinhaltet, berechnet wird. Die Algorithmen, die für die Berechnung des nächsten Zustandes verantwortlich sind, können dabei beliebig komplex sein und Ziele, Intentionen, Erfahrungen usw. berücksichtigen.

Zunächst ist festzustellen, dass die Modellierung nonverbalen Verhaltens mit Hilfe des Computers die kontroversen Sichtweisen der Ausdruckspsychologie und der Eindruckspsychologie bis zu einem gewissen Grad vereinigt. Es wird der Ausdruckspsychologie Rechnung getragen, da wir innere Zustände benötigen, die das nonverbale Verhalten des Avatars widerspiegeln. Der Eindruck/Ausdruck „Freude“ schlägt sich in bestimmten Parametereinstellungen des 3D-Modells nieder. Die Annotation dieses Parametersatzes als „Freude“ bedeutet ausdruckspsychologisch interpretiert: Die Person (3D-Figur) freut sich. Eindruckspsychologisch interpretiert, bedeutet der gleiche Parametersatz, die Person erweckt beim Betrachter den Eindruck, dass sie sich freut. Das auf dem Monitor zu sehende Bild des Avatars ist in beiden Fällen, also unabhängig von der jeweiligen Interpretation, das Abbild eines inneren Zustands des Programms. Diesen Zustand wollen wir im Weiteren als „*nonverbaler Zustand*“ bezeichnen.

Die Modellierung nonverbalen Verhaltens mit Hilfe des Computers macht aber auch Unterschiede zwischen ausdruckspsychologischer und eindruckspsychologischer Modellierung deutlich. Die Realisierung virtueller Assistenten nach dem Vorbild der Ausdruckspsychologie bedeutet immer die direkte Abbildung des emotionalen Zustands eines Avatars auf das nonverbale Verhalten. Die dynamischen Merkmale des äußeren Erscheinungsbildes sind ge-

mäß der Ausdruckspsychologie ein Spiegel des inneren Zustandes. Grundemotionen wie Überraschung, Traurigkeit, Zorn, Freude, Ärger und Furcht spiegeln sich in inneren Zuständen wider, die quasi „fest verdrahtet“ sind mit dem entsprechenden nonverbalen Verhalten. Der Avatar ist traurig, und diese Traurigkeit zeigt sich auf dem nonverbalen Display.

Die Modellierung nonverbalen Verhaltens gemäß der Eindruckspsychologie ist dagegen komplexer. Das emotionale Verhalten des Avatars, seine Intentionen und Ziele müssen ebenfalls in irgendeiner Weise auf innere Zustände abgebildet werden, die gemäß bestimmter Regeln fortgeschaltet werden. Das nonverbale Display und der nonverbale Zustand sind aber entkoppelt. Der Zustand des Avatars umfasst sehr viel mehr als die für den nonverbalen Display verantwortlichen Parametermuster. Nonverbales Verhalten ist kein Fenster zum inneren Zustand oder gar zum Befinden des Avatars, sondern Mittel zur Regulierung eines Kommunikationsprozesses. Die Modellierung von Face-to-Face-Kommunikation nach dem Vorbild der Eindruckspsychologie bedeutet, es werden die nonverbalen Signale gegeben, die der Kommunikationsabsicht am besten dienen. Der virtuelle Assistent reflektiert sein eigenes nonverbales Verhalten immer im Hinblick auf die Wirkung beim Kommunikationspartner. Der Avatar braucht dazu eine (implizite oder explizite) Repräsentation dessen, was sich im Gegenüber abspielt, um gezielt weitere Handlungen zu planen und auszuführen. Das Kommunikationsproblem wird damit auf eine Metaebene verlagert. Das nonverbale Verhalten kann hier, insbesondere wenn es der momentanen Zielsetzung und dem Kommunikationsprozess dient, „bewusst“ lügen.

Das Spektrum unserer Überlegungen zur Implementierung nonverbalen Verhaltens bewegt sich damit zwischen den Positionen „Emotionen haben“, „Emotionen simulieren“ und „Emotionen beim Anderen erzeugen“. Wir erkennen hier die Grundstruktur der Diskussionen wieder, wie sie im Rahmen der KI schon geführt wurden. Zurecht stellt Kaehr (1999) fest, dass die Unterscheidung von „Haben oder Sein“ keine Seins- sondern eine Reflexionsbestimmung und somit standpunktabhängig ist. Für Implementierungen auf der Basis gegenwärtiger Möglichkeiten sind die praktischen Implikationen der unterschiedlichen psychologischen Forschungsparadigmen aber durchaus wichtig, auch wenn sie sich bei grundlagentheoretischer Betrachtung als unhaltbar erweisen sollten.

Die Verlagerung des Kommunikationsmodells auf eine Metaebene, das reflektierte „so tun als ob“, wird zu leistungsfähigeren virtuellen Assistenten führen als der einfachere ausdruckspsychologische Ansatz. Der Vergleich der hier betrachteten emotiven Modelle mit den kognitiven Modellen der

KI ist aber auch deshalb wenig hilfreich, weil beim nonverbalen Verhalten ein Kommunikationsmodell zugrunde liegt, das immer schon die Reflexion des Gegenübers voraussetzt. Die typische Zielsetzung der KI dagegen ist die intelligente Problemlösung, die auch ohne Reflexion gelingen kann. Wichtig für die Informatik ist in diesem Zusammenhang, dass das Shannonsche Kommunikationsmodell des Dekodierroboters, bei dem der Empfänger eingehende Signale geduldig hinnehmen muss und keine Chance hat, sie anders als gemäß der gemeinsamen Semantik zwischen Sender und Empfänger zu dekodieren, in der Eindruckspsychologie ersetzt wird durch das Modell der inferentiellen Kommunikation. Zentrales Kennzeichen der inferentiellen Kommunikation im Unterschied zur kodierten Kommunikation ist, dass die Beziehung zwischen Zeichen und Bezeichnetem nicht aus einer semantischen Konvention resultiert, sondern das Ergebnis einer pragmatischen Schlussfolgerung des Empfängers ist (siehe z. B. Frey 1999). Es werden immer auch Zeichen zwischen Sender und Empfänger ausgetauscht, die nicht Bestandteil eines verbindlichen Codes sind, sondern für jede Seite etwas anderes bedeuten können.

5 Zur Persönlichkeit virtueller Assistenten

Es ist leicht verständlich, warum Sprachverarbeitung und nonverbale Kommunikation für die Mensch-Maschine-Interaktion von so großer Bedeutung sind. Mehr direkte Kanäle stehen dem Menschen zur unvermittelten Kommunikation mit seiner Umwelt nicht zur Verfügung. Alle anderen Ausdrucksformen werden indirekt, d. h. über das Tun und seine Auswirkungen vermittelt. Die Simulation menschlicher Face-to-Face-Kommunikation wird umso mächtiger, je besser und in sich schlüssiger es gelingt, die Bandbreite und Vielfalt menschlichen Kommunikationsverhaltens im Umgang mit der Maschine auszunutzen. Die große Herausforderung für die Implementierung virtueller Assistenten ist die formale Fassung dessen, was wir Persönlichkeit nennen. Als *Persönlichkeit* verstehen wir hier die Zusammenfassung aller Besonderheiten eines Menschen, seine eigenständigen, von den Rollenmustern, globalen oder lokalen sozialen Benimmregeln und Codes der Gesellschaft unabhängigen Verhaltensmuster. Erst die Persönlichkeit macht uns zu unterscheidbaren Individuen. Jedes Individuum zeigt Wesenszüge, die über lange Zeit aufrechterhalten werden und ihn für andere identifizierbar machen. Die nach außen sichtbaren Zeichen, die Menschen benutzen, um die Persönlichkeit anderer zu lesen, werden von drei Komponenten bestimmt: der ver-

balen Ausdrucksweise, der physischen Erscheinung und dem nonverbalen Verhalten (Cassell 2000). Diese drei Komponenten sind natürlich gleichzeitig Mittel, um soziale Codes zu transportieren. Die Stellung innerhalb der Gesellschaft kann zum Beispiel gut über Accessoires und nichtverbalen Verhalten vermittelt werden. Die formale Behandlung der Triade – verbale, physische und nonverbale Charakteristik – und ihre Implementierung im Computer eröffnet die Möglichkeit die einzelnen Komponenten getrennt zu modellieren und ihren jeweiligen Einfluss zu untersuchen (siehe Figur 4). Sprachverarbeitung, computergrafische Modellierung und die 3D-Computeranimation sind die entsprechenden Forschungsbereiche der Informatik, die für die Untersuchung der technischen Aspekte der Persönlichkeitskomponenten verantwortlich sind.

Im Folgenden wird zwischen Persönlichkeit und Charakter unterschieden. In der hier verstandenen Form bezieht sich Persönlichkeit auf die nach außen sichtbaren Merkmale eines Individuums. Bei den Eigenschaften, die für unsere Handlungsstrategien verantwortlich sind, werden wir von Charakter sprechen. Der Begriff Charakter ist zudem umfassender und unbelasteter als der Begriff Persönlichkeit, den wir im Einklang mit der einschlägigen Literatur (z.B. Cassell) für die drei äußeren Erscheinungsformen – verbaler Stil, statisches Aussehen und nonverbales Verhalten – verwenden wollen. Unter Charakter dagegen verstehen wir den Wert, den wir unseren Entschei-

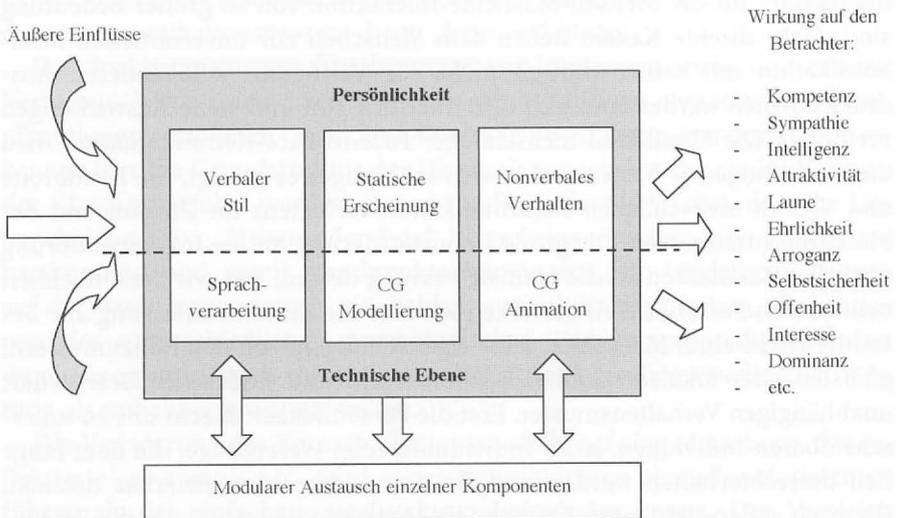


Abb. 4. Komponenten der Persönlichkeit virtueller Assistenten

dungen und Beziehungen zu anderen beimessen. Aus der großen Vielfalt von Handlungsmöglichkeiten zu jedem einzelnen Zeitpunkt wählen wir einige gezielt aus und versuchen sie aufrechtzuerhalten. Diese nachhaltigen Züge werden uns zum Charakter; es sind die Merkmale, die wir an uns selbst schätzen und für die wir die Anerkennung des anderen suchen. Charakter drückt sich durch Verfolgung langfristiger Ziele aus, durch den Verzicht auf schnelle Wechsel zugunsten stetiger Entwicklung. Stabilität erscheint dem Charakter nicht als Lähmung, das Ziel ist wichtiger als das oberflächliche Zeichen des Aufbruchs und der Änderung. Der Charakter bildet den Hintergrund vor dem (1) eine Entscheidung getroffen, (2) Unbekanntes gelernt, (3) Gespeichertes ausgewählt und (4) Neues gedacht und geschaffen wird.

In den unten skizzierten Methoden des Dialogmanagements berühren sich die beiden Bereiche Persönlichkeit und Charakter, die durch Veranlagung und äußere Einflüsse geprägt werden. Hier sind die eigentlichen Herausforderungen der Informatik zu suchen. Wie können virtuelle Assistenten lernen? Wie können langfristige Ziele verfolgt werden und sich dabei langsam verändern? Welche Algorithmen sind in der Lage, ein System in der Interaktion mit dem Benutzer wandlungsfähig zu halten und neue Ziele zu eröffnen? Wie müssen Assistenten aussehen, deren Komplexität nicht schon beim Start determiniert ist, sondern die erst im Zusammenspiel mit dem Benutzer je unterschiedliche Komplexität entwickeln?

5.1 Verbale Ausdrucksweise

Unter verbaler Ausdrucksweise fassen wir die Eigenheiten der Stimme und des inhaltlichen Dialogverhaltens zusammen. Mit den individuellen Unterschieden der menschlichen Stimme musste sich die Spracherkennung seit Beginn ihrer Forschungen auseinander setzen. Bis heute werden die Methoden der sprecherabhängigen, die eine personenbezogene Trainingsphase zur Benutzeradaptation voraussetzen, und die Methoden der sprecherunabhängigen Spracherkennungswerkzeuge unterschieden. Ausgabeseitig verfolgt das Forschungsgebiet der Sprachsynthese das Ziel, geschriebenen Text in individualisierte gesprochene Sprache umzusetzen. Hier bedeutet Individualisierung derzeit in erster Linie, zwischen männlichen und weiblichen, jungen und alten, hohen und tiefen etc. Stimmen zu differenzieren. In den letzten Jahren sind auf diesem Gebiet große Fortschritte zu verzeichnen. Eine sehr viel größere Herausforderung stellt die Dialogmodellierung dar.

Die Ansätze zum Dialogmanagement können grob in *finite-state Methoden* und *selbstorganisierende Methoden* unterschieden werden, wobei die derzeit kommerziell verfügbaren Dialogsysteme auf *finite-state Methoden* beruhen. In endlichen Zustandsmodellen wird die Dialogstruktur in Form von Zustands-Transitions-Netzwerken repräsentiert, wobei den Knoten die Systemfragen entsprechen und die Übergänge den möglichen Pfade durch das Netz, die damit die erlaubten Dialoge definieren. Selbstorganisierende Systeme, zu denen *event-driven methods*, *plan-based management* und *theorem-proving* gehören, zeigen prinzipiell allerdings weitaus größere Flexibilität (Fraser, N. M., Dalsgaard, P. 1996). Intelligentes Verhalten virtueller Assistenten steht andererseits auch im Zentrum der Forschungen zu Multiagenten-Systemen und intelligenten Software-Agenten.

5.2 Physische Erscheinung

Die Modellierung menschlicher Figuren und Gesichter mit Hilfe von 3D-Werkzeugen ist nach wie vor ein schwieriges und zeitraubendes Unterfangen. Der praktische Nutzen und damit die kommerzielle Verwertbarkeit virtueller Assistenten, wie auch die Möglichkeit, systematisch psychologische Versuche durchzuführen, stellt große Anforderungen an den Modellierungsprozess. Wünschenswert wären Werkzeuge, die es erlauben, Figuren auch in hoher Qualität schnell zu generieren und gezielt und effizient Modifikationen an den 3D-Modellen vornehmen zu können. Von besonderem Interesse sind beispielsweise gezielte Einflüsse auf den Abstraktionsgrad, das Alter und das Geschlecht. Durch Variation des Abstraktionsgrades sind einerseits gezielte Experimente möglich, die Aufschluss geben, wie weit das statische Aussehen reduziert werden kann, ohne Verluste im Kommunikationsprozess in Kauf nehmen zu müssen. Andererseits kann z.B. durch Polygonreduktion (bzw. Polygonerhöhung) die Figur an die jeweiligen Leistungsdaten des Computers oder des Netzwerkes angepasst werden. Insbesondere die Frage der Individualisierung virtueller Figuren ist im Zusammenhang mit den üblichen Reduktionen und Abstraktionen der 3D-Modellierung zu sehen. 3D-Modelle von Gesichtern werden üblicherweise aus Gründen der Effizienz achsensymmetrisch angelegt, d.h. die linke Gesichtshälfte wird durch Spiegelung der rechten erzeugt (oder umgekehrt). Die zwei Seiten eines menschlichen Gesichtes sind jedoch asymmetrisch. Auch wenn wir die Unterschiede im Detail nicht benennen können, ist die Differenz der beiden Gesichtshälften für das per-



Abb. 5. Zur Asymmetrie menschlicher Gesichter

sönliche Profil entscheidend. Mit Hilfe einfacher Bildbearbeitungsprogramme lässt sich die Asymmetrie leicht zeigen. Figur 5 zeigt in der Mitte ein Foto des Autors, links ist das Gesicht abgebildet, das aus der rechten Gesichtshälfte durch Spiegelung an der Mittelachse erzeugt wurde und auf der rechten Seite das auf gleiche Weise aus der linken Gesichtshälfte erzeugte.

Einen vielversprechenden Ansatz zur schnellen Generierung und Variation menschlicher Gesichter haben kürzlich Blanz und Vetter (1999) vorgestellt. Auf der Basis von 200 Laserscans der Köpfe junger Erwachsener (100 weibliche und 100 männliche) wird durch Transformation der 3D- und Textur-Daten in eine Vektorraum-Repräsentation ein skalierbares 3D-Modell abgeleitet. Neue Gesichter und sogar Ausdrücke können so effizient durch die Erzeugung von Linearkombinationen der 200 Prototypen gewonnen werden.

5.3 Nonverbales Verhalten

Versuche von Frey (1999) und Bente zeigen eine erstaunliche Dominanz des Bewegungsverhaltens gegenüber den morphologischen Merkmalen. Viele äußerlich einschätzbare Eigenschaften, die auch für virtuelle Assistenten wichtig sind – z.B. Kompetenz, Ehrlichkeit, Mitgefühl, Ruhe und Sympathie –, werden weitgehend über nonverbale Stimuli transportiert. Selbst Eigenschaften wie Attraktivität, die der Laie überwiegend dem statischen Erscheinungsbild zuschreibt, werden offensichtlich zu einem beträchtlichen Teil durch die Bewegungsstimulation determiniert. Erste Untersuchungen

zeigen, dass hierbei der Lateralflexion eine besondere Bedeutung zukommt, die Gründe hierfür sind bisher weitgehend ungeklärt. Diese Ergebnisse machen aber deutlich, dass sich Interface-Agenten als neues Schnittstellenparadigma nur durchsetzen können, wenn wir das nonverbale Verhalten der Figuren überzeugend simulieren können.

Bente und Krämer (1999) benennen drei zentrale Funktionsbereiche des nonverbalen Verhaltens:

- **Inhaltliche Klärung:** Hierzu gehören Zeigegesten, Illustratoren und Embleme.
- **Interaktionssteuerung:** Turntaking (Übergabe von Sprecher und Hörer-Rolle), Backchannel (Bestätigung und nonverbale Kommentierung, z. B. durch Nicken, Kopfschütteln und Handgesten), Interrupts (Blickkontakt herstellen und Stimme anheben).
- **Sozio-emotionale Aspekte:** Aktivierung und Deaktivierung durch Intensivierung bzw. Zurücknahme der Gestik und Lautstärke, Dominanz und Unterordnung etc.

In allen drei Bereichen liegen erste theoretische Ergebnisse vor, andererseits klaffen große Wissenslücken. Unser Arbeitsschwerpunkt wird deshalb in nächster Zeit auf der Programmierung einer Methodenentwicklungs- und Evaluationsplattform für animierte virtuelle Figuren liegen. Die Implementation sollte aufgrund des lückenhaften Wissensbestandes die Vorteile eines theoriegesteuerten Top-down-Ansatzes mit denen eines daten-gesteuerten Bottom-up-Ansatzes verbinden. Langfristiges Ziel ist der Aufbau einer Bewegungsbibliothek für nonverbale Kommunikation, die es gestattet, Dialogsysteme für verschiedene Kontexte effizient generieren zu können.

6 Die Algorithmen der nonverbalen Kommunikation

Beim derzeitigen psychologischen Wissenstand können wir große Bereiche nonverbalen Dialogverhaltens auf absehbare Zeit noch nicht algorithmisch formulieren. Die theoretischen Lücken in den Wissensbeständen sind zu groß, um mit synthetisch generierten Bewegungen überzeugende Kommunikationsabläufe zu simulieren. Bis zur Schließung der Lücken muss deshalb auf Bewegungsprotokolle von natürlichen Bewegungen zurückgegriffen werden. Die präzisen Daten, die optische und magnetische Motion-Capture-Systeme liefern, bieten eine hervorragende Basis um einfach und systematisch

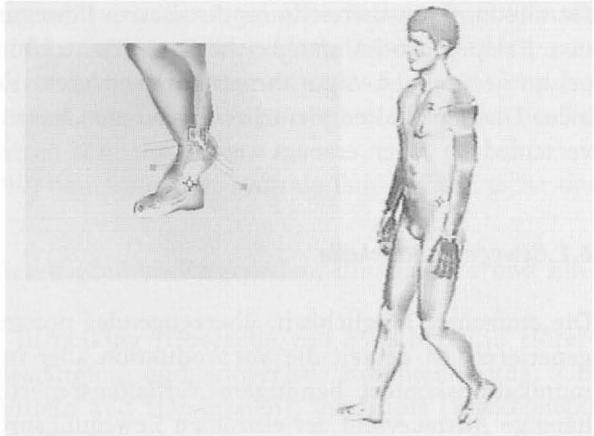
für alle Funktionsbereiche repräsentative Bewegungssequenzen aufzuzeichnen. Entsprechend umfangreiche Datenbanken können so zunächst den Mangel an generativen Algorithmen kompensieren. Kontextabhängiges nonverbales Dialogverhalten virtueller Assistenten kann somit prinzipiell auf drei verschiedene Arten erzeugt werden.

6.1 Bewegungsprotokolle

Die einfachste Möglichkeit, überzeugendes nonverbales Dialogverhalten zu generieren, ist derzeit die Vorproduktion aller in einen bestimmten Kommunikationskontext benötigten Verhaltensrepertoires und die kontextabhängige Ansteuerung der einzelnen Bewegungsprotokolle. Kernstück dieser Methode ist die Entwicklung und Implementierung einer Referenzfigur, auf der ein deskriptives Verhaltensprotokoll als zentrales Format und Bindeglied zwischen Ein- und Ausgabeseite definiert ist. Hier erscheint die Erweiterung des Berner Zeitreihennotationssystems oder die Verwendung eines Standardprotokolls (z.B. Biovision) Erfolg versprechend. Für diese Methode ist es sinnvoll, eine Datenbank von Verhaltensrepertoires anzulegen, da bestimmte Sequenzen – z.B. *turntaking*, *interrupts*, *backchannel*, *emotions* – zur so genannten Metakommunikation gehören, d.h. für jede Face-to-Face Kommunikation verfügbar sein müssen. Da riesige Mengen an Bewegungsdaten zu verwalten sind, muss eine datenbankbasierte Aufzeichnungs- und Abspielmöglichkeit der Verhaltensprotokolle gegeben sein. Es ist darauf zu achten, dass die effiziente Verwaltung der anfallenden Datenmengen und der selektive Zugriff auf einzelne Kanäle (z.B. nur auf Kopfbewegungen) innerhalb der Protokolle gewährleistet ist. Diese Methode ist nur für eingeschränkte Szenarien anwendbar, da alle möglichen nonverbalen Antworten vorproduziert werden müssen. Ein weiteres Problem dieser Methode stellen die Übergänge zwischen den einzelnen Sequenzen dar.

6.2 Bewegungstransformation

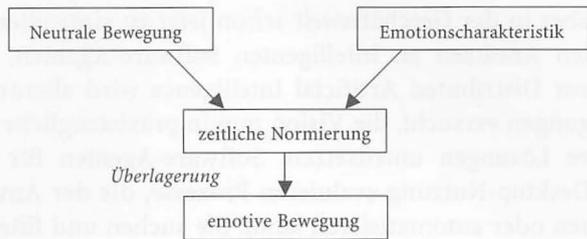
Viele der Schwierigkeiten bei der Simulation menschlicher Bewegungen resultieren aus der großen Anzahl von Freiheitsgraden, die in jedem Zeitschritt kontrolliert werden muss. Eine andere enorme Herausforderung der Animation ist, dass wir alle Experten in Beurteilung menschlicher Bewegun-

Abb. 6. Motion Capture

gen sind.⁶ Auch wenn wir den einzelnen Faktor oft nicht gezielt isolieren können, erkennen wir dennoch sehr zuverlässig, ob etwas „richtig aussieht“ oder nicht. Motion Capturing (siehe Abbildung 6) hat Animationen realisierbar werden lassen, die vorher nicht möglich waren. Diese Methode hat zwar noch immer den Nachteil, dass aufwändiges und teures Equipment erforderlich ist, die entscheidende Limitierung des Verfahrens liegt aber an einer anderen Stelle. Bisher haben wir nur sehr eingeschränkte Möglichkeiten, einmal aufgezeichnete Daten zu editieren und so in verschiedenen Kontexten wieder zu verwenden. Das führt dazu, dass wir den gesamten Aufnahmeprozess wiederholen müssen, auch wenn nur geringfügig andere Bewegungen benötigt werden. Insbesondere Echtzeit-Transformationsmöglichkeiten würden einen wichtigen Fortschritt für die überzeugende Animation virtueller Assistenten bedeuten. Bewegungsbibliotheken für nonverbales Verhalten erlangen erst im Zusammenspiel mit geeigneten Verfahren zur Bewegungs-transformation ihre wirkliche Bedeutung. Hinter dem Ansatz der Bewegungs-transformation verbirgt sich die Hoffnung, dass emotive Bewegungen aus neutralen menschlichen Bewegungen erzeugt werden können (siehe Abbildung 7). Durch die Anwendung von Techniken des *Motion Signal Processing* werden neutrale Bewegungen so transformiert, dass die Grundstruktur

⁶ Die großen Erfolge, die Dinosaurieranimationen in BBC-Fernsehserien oder Hollywoodfilmen feiern, sind nicht zuletzt deshalb möglich, weil niemand weiß, wie sich diese Tiere wirklich bewegt haben. Es gibt niemanden, der hier als Experte zu Rate gezogen werden könnte.

Abb. 7. Überlagerung neutraler Bewegungen mit Emotionen



der ursprünglichen Bewegung erhalten bleibt, die emotionale Signatur sich jedoch gezielt verändert. Die Transformationen sollen hierbei die Differenz zwischen einer neutralen und einer emotiven Bewegung, z. B. einer traurigen oder ärgerlichen Bewegung, erfassen (vgl. Amaya, Bruderlin, Calvert 1996).

6.3 Generative Bewegungs-Algorithmen

Langfristiges Ziel der computerbasierten Animation nonverbalen Verhaltens ist die Entwicklung von Algorithmen zur Generierung menschlicher Bewegungen ohne den Rückgriff auf aufgezeichnete Bewegungsdaten. Im Bereich der motorischen Bewegungen mehrgliedriger Figuren, z. B. zum menschlichen Gangverhalten, liegen bereits umfangreiche Arbeiten vor. Diese Verfahren sind für die Simulation kommunikativer Dialogsituationen nicht geeignet. Im Bereich der Face-to-Face-Kommunikation ist bisher nur der relativ einfache Funktionsbereich der inhaltlichen Klärung, wozu insbesondere Zeigegesten, Illustratoren und Embleme gehören, algorithmisch möglich. Für die Funktionsbereiche Interaktionssteuerung und sozio-emotionale Regulierung liegen bisher keine Implementierungen vor.

7 Ausblick

Die Explosion der Internet-Inhalte und der zunehmende Breitbandzugang zum Netz erfordern neue Ansätze in der Benutzerführung und der Gestaltung des Mensch-Maschine-Interfaces. Die Metapher des intelligenten und autonomen Agenten geht davon aus, dass unsere Bedienprobleme mit dem Computer gelöst sind, sobald der Benutzer sich mit dem Gerät wie mit einem Menschen unterhalten kann. Intelligente Agenten sind heute im Computerbereich zwar immer noch weit von dieser Vision entfernt, kommen

aber in der Geschäftswelt schon jetzt zu sinnvollem Einsatz. Mit den jüngsten Ansätzen zu Intelligenten Software-Agenten, Multiagent Systems und zur Distributed Artificial Intelligence wird allerorten mit großen Anstrengungen versucht, die Vision nun in praxistaugliche und kommerziell lukrative Lösungen umzusetzen. Software-Agenten für Internet-, Intranet- und Desktop-Nutzung evaluieren Prozesse, die der Anwender individuell anpassen oder automatisieren kann. Sie suchen und filtern Informationen, sortieren E-Mails, simulieren menschliche Gespräche und rationalisieren Prozesse für den Anwender. Weitere Aspekte bilden das Lernen und Planen in Multi-Agentensystemen und die Sicherheit und Vertraulichkeit intelligenter Agenten. Ein zentraler Aspekt intelligenter Agenten ist ihre Visualisierung und ihr dynamisches nonverbales Verhalten. Die Leistungsfähigkeit und damit die Akzeptanz virtueller Agenten und Assistenten in der Praxis wird – neben Intelligenz und Autonomie – entscheidend davon abhängen, wie weit es gelingt, überzeugendes nonverbales Verhalten zu implementieren.

Die psychologische Untersuchung nonverbaler Stimuli wird derzeit von zwei konkurrierenden Forschungsparadigmen geleitet, der Ausdruckspsychologie und der Eindruckspsychologie (Wirkungsforschung). Es gilt, die Ausdruckspsychologie, die ihren Ursprung bereits in der *Physiognomonika* des Aristoteles und in den Arbeiten des Schweizer Pastors Johann Caspar Lavater im 18. Jahrhundert haben, durch den reflexiven und inferentiellen Ansatz der Wirkungsforschung abzulösen. Es kann nicht länger darum gehen, den Agenten neben Intelligenz nun auch noch Gefühle zu geben, sondern darum, mit den Gefühlen des Benutzers geeignet umzugehen, d.h. ihn der Zielsetzung entsprechend zu stimulieren. Die Überlegungen zur Implementierung nonverbalen Verhaltens zielen deshalb nicht darauf ab, Emotionen zu implementieren oder zu simulieren, sondern sie beim Benutzer (Betrachter) hervorzurufen und so den virtuellen Assistenten zu unterstützen, seine ihm zgedachten Ziele zu erreichen. Daneben sind die Individualisierung von Information und die Lernfähigkeit der Agenten, d.h. ihre Fähigkeit der Adaption an den jeweiligen Benutzer, große Herausforderungen.