

# Psychologische Aspekte bei der Implementierung und Evaluation nonverbal agierender Interface-Agenten<sup>1</sup>

Gary Bente, Nicole C. Krämer  
Universität zu Köln, Psychologisches Institut

## Zusammenfassung

Der Beitrag behandelt psychologische Aspekte bei der Gestaltung und Evaluation sog. anthropomorpher Schnittstellen. Fokussiert werden dabei Ansätze zur Implementierung nonverbaler Verhaltenskomponenten. Unterschieden werden in diesem Zusammenhang modell-basierte top-down-Ansätze und datenbasierte bottom-up-Ansätze. Vor- und Nachteile der beiden Zugangsweisen werden diskutiert und Integrationsmöglichkeiten im Rahmen einer eigenen Entwicklungs- und Evaluationsplattform für anthropomorphe Interface-Agenten vorgestellt. Erste Forschungsergebnisse aus dem Einsatz der Methodenplattform zur Evaluation eines virtuellen Assistenten in der Unterhaltungselektronik werden referiert und im Hinblick auf weitere Forschungsperspektiven bewertet.

## 1 Intuitive Schnittstellen: Face-to-Face-Kommunikation als neue MCI-Metapher

Aus der wachsenden Komplexität moderner Informations- und Kommunikationstechnologien ergeben sich zunehmend höhere Anforderungen an die Einfachheit und Benutzerfreundlichkeit der verfügbaren Mensch-Technik-Schnittstellen. Vor diesem Hintergrund wird auch vermehrt die Einsetzbarkeit von sog. anthropomorphen Interfaces (embodied interface agents, animated user interface agents, face-to-face like interfaces) diskutiert, die ein- und ausgabeseitig der Humankommunikation nachempfunden sind. Durch den Einsatz natürlicher Kommunikationskanäle, wie Sprache, Gestik, Mimik, etc. sollen solcherart gestaltete Schnittstellen einen intuitiven - d.h. auf allgemein vorhandene Kommunikationsvoraussetzungen rekurrierenden - Zugang selbst zu komplexesten Systemen ermöglichen. Unter Bezug auf die strukturellen Ähnlichkeiten mit der menschlichen face-to-face-Kommunikation wird dieser Interaktionsform generell auch eine leichtere Handhabung sowie größere Effizienz und Akzeptanz zugesprochen (vgl. Brennan, 1990; Laurel, 1990; Bolt, 1987; Thórisson, 1996). So formulieren etwa Takeuchi und Naito (1995, S. 454) in diesem Sinne ihre optimistischen Erwartungen an die neuen Technologien: „*We surmise that once people are accustomed to synthesized faces, performance becomes more efficient, and a long partnership further improves performance. Human-like characterization is one good form of autonomous agents, because people are accustomed to interact with other humans*“.

Bei der Entwicklung derartiger Interfaces sind sowohl auf der Eingabeseite (Spracherkennung, Motion-Capture, Face-tracking, Eye-tracking, etc.) als auch auf der Ausgabeseite (Dialogmodellierung, Echtzeitausgabe realistischer verbaler und nonverbaler Verhaltensmuster) sicherlich noch eine Reihe technischer Probleme zu lösen, in deren Rahmen psychologische Überlegungen zur 'Usability' und kognitiven Ergonomie eine Rolle spielen (vgl. Oviatt, 1999). Ganz besonders mit Blick auf die Ausgabeseite anthropomorpher Interfaces, also auf die Generierung

---

1 Die dargestellten Entwicklungsarbeiten wurden im Rahmen des vom BMB+F geförderten Leitprojektes EMBASSI (Elektronische Multimodale Bedien- und Serviceassistenz; BMB+F Förderkennzeichen 01 IL 904 L) in Kooperation mit der Kunsthochschule für Medien, Köln, durchgeführt.

und Evaluation künstlichen menschlichen Kommunikationsverhaltens werden jedoch in besonderer Weise auch Konzepte und Befunde aus der Sozial- und Kommunikationspsychologie salient. Die Mensch-Computer-Interaktion gewinnt hier eine völlig neue - eine (para-)soziale - Dimension. Konsequenterweise fordern Parise, Kiesler, Sproull und Waters (1999) auch eine Erweiterung der psychologischen Forschungsperspektive: „*As computer interfaces can display more life-like qualities such as speech output and personable characters or agents, it becomes important to understand and assess user's interaction behavior within a social interaction framework rather than only a narrower machine interaction one*“ (S. 123).

Neben der Möglichkeit zur Sprachausgabe eröffnet nun aber gerade die visuelle Präsenz des virtuellen Gesprächspartners eine Reihe zusätzlicher, sog. nonverbaler Kommunikationskanäle, also Körperbewegung, Mimik, Gestik, Körperhaltung, Blickverhalten, etc. deren subtile aber auch nachhaltige Wirkungen aus der Face-to-Face-Kommunikationsforschung bereits bekannt sind. Im Zusammenhang mit der Mensch-Computer-Interaktion (MCI) wird diese Kanalerweiterung im Sinne der Multimodalität per se oft als ein Vorteil gesehen (vgl. etwa Thórisson, 1996), wobei spezifische Problemstellungen, die sich aus der Nutzung nonverbaler Kommunikationsmodalitäten ergeben können, übersehen werden. Wir werden in diesem Beitrag zunächst Konzepte und Befunde zur Wirkung anthropomorpher Interface-Agenten skizzieren und vor diesem Hintergrund zentrale Probleme der Verhaltensmodellierung im Bereich der nonverbalen Kommunikation aufzeigen. Dabei sollen offene Fragen der Grundlagenforschung herausgearbeitet und eine experimentelle Plattform zur Implementation und Evaluation nonverbaler Verhaltenskomponenten vorgestellt werden.

## 2 Nutzen und Wirkung anthropomorpher Interface-Agenten

Die Frage, inwieweit die Bereitstellung anthropomorpher Interfaces überhaupt zu spezifischen, und gegebenenfalls auch zu erwünschten emotionalen, kognitiven und/oder verhaltensmäßigen Reaktion beim menschlichen Nutzer führen, ist bislang empirisch noch wenig erforscht. Auch die theoretische Diskussion ist noch kaum elaboriert zu nennen. Vereinzelt Befunde allerdings scheinen die Fähigkeit der anthropomorphen Assistenten zur Auslösung spezifischer sozio-emotionaler Phänomene zu belegen: So lässt sich etwa die aus der Humankommunikation bekannte, starke Wirkung menschlicher Gesichter hinsichtlich der Attribution von Emotionen auch im Kontext von MCI nachweisen. Takeuchi und Naito (1995) konnten zeigen, dass ein virtuelles Kartenspiel als unterhaltsamer bewertet wird, wenn der Gegner durch ein animiertes Gesicht dargestellt wird (vgl. auch Koda & Maes, 1996). Dass dabei vor allem die bloße Erscheinung des menschlichen Gesichts eine Rolle zumindest im Sinne der Unterhaltung spielt, belegen Untersuchungen von Lester et al. (1997) sowie van Mulken et al. (1998). Lester et al. (1997) konnten hierbei auch zeigen, dass die Effekte der bewegten Gesichter kaum von der Qualität der Animation abhängen. Van Mulken et al. (1998) fanden, dass die Steigerung des Unterhaltungseffekts durch eine anthropomorphe Figur jedoch auch stark kontextabhängig ist, d.h. wenn das Gesicht dem Interface eine besondere Stimulusqualität hinzufügt. So konnten die Vorteile eines virtuellen Gegenübers gegenüber einem Pfeil-Pointer nicht beobachtet werden kann, wenn im Rahmen der Interaktion mit dem System ohnehin die Betrachtung menschlicher Gesichter erforderlich war. Während diese Befunde auf die sozio-emotionale Bedeutung virtueller (Inter-)Faces verweisen, deuten andere Ergebnisse auch auf direkte Verhaltensfolgen für den Benutzer. Rickenberg und Reeves (2000) etwa untersuchen unter dem Stichwort „social facilitation“ (psychologische Prozesse, die im Falle der Anwesenheit anderer Menschen dazu führen, dass unter bestimmten Umständen bessere, je nach Situation aber auch schlechtere Ergebnisse erreicht werden), ob sich durch die Gegenwart eines animierten Charakters Auswirkungen auf das Leistungsverhalten nachweisen lassen. Die Autoren fanden, dass Probanden, die von einem animierten Charakter beobachtet werden, größere Angst erleben und mehr Fehler machen. Die Au-

toren führen dies auf eine durch die soziale Gegenwart bedingte höhere Erregung zurück, aus der sich dann eine Akzentuierung der vorhandenen Handlungs- und/oder Gefühlstendenzen ergeben soll. Auch Sproull, Subramani, Kiesler, Walker und Waters (1996) berichten, dass Personen, die mit einem anthropomorphen Interface agieren, sich selbst als weniger entspannt und selbstsicher beschreiben als beim Umgang mit einem rein textbasierten System. Darüber hinaus konnten die Autoren jedoch auch eine verstärkte Tendenz zur positiven Selbstdarstellung beobachten. Dieses auch als „Impression Management“ oder Selbstpräsentation bezeichnete Phänomen (vgl. Schlenker, 1980; Goffman, 1959), tritt vorrangig in Gegenwart anderer Personen auf und führt häufig zu konformistischem oder sozial erwünschtem Verhalten (Leary, 1995). Ziel ist der möglichst ‘gute Eindruck’, der beim Gegenüber hinterlassen werden und wiederum positive Behandlung nach sich ziehen soll. Sproull et al. (1996) weisen in diesem Zusammenhang auch auf möglicherweise kontrovers zu diskutierende Implikationen dieses Phänomens hin: *„Many people want computers to be responsive to people. But do we also want people to be responsive to computers?“* (S. 119). Tatsächlich finden sich auch weiterreichende Wirkungen sozialer HCI-Schnittstellen, deren Implikationen auch auf gesellschaftlicher Ebene zu diskutieren sind. So fanden Milewski und Lewis (1997) auch Effekte hinsichtlich des Vertrauens der Nutzer in das System bzw. auch in der Delegationsbereitschaft: Eine Delegation an das System scheint leichter zu fallen, wenn ein menschliches Äußeres vorhanden ist. Dies ist konsistent mit den Ergebnissen einiger Akzeptanz- und Bewertungsuntersuchungen, die belegen, dass Systemen mit anthropomorphen Interfaces auch eine besondere Glaubwürdigkeit (trustworthiness) zugeschrieben wird (Rickenberg & Reeves, 2000).

Es wird kaum zu bezweifeln sein, dass der menschliche Nutzer in Interaktion mit einem virtuellen Gegenüber sich durchaus bewusst ist, dass die abgebildete Figur nicht einem realen Menschen gleichkommt und er auch nicht von dieser beobachtet wird. Dennoch sprechen die verschiedenen Untersuchungsergebnisse insofern für das Erleben einer (para-)sozialen Präsenz, als dass Verhaltenstendenzen beobachtet werden können, die denen in realen sozialen Situationen vergleichbar sind. Dennoch bedürfen die Befunde weiterer Überprüfung, da verschiedene Störvariablen und alternative Erklärungsmodelle nicht berücksichtigt wurden: So mochten Sproull et al. (1996) im Nachhinein nicht ausschließen, dass die Personen durch die Abbildung des menschlichen Gesichts lediglich anzunehmen verleitet waren, dass die individuellen Angaben später an eine reale Person weitergeleitet werden. Bezüglich aller Untersuchungen muss ferner in Betracht gezogen werden, dass berichtete Aktivierungserhöhungen auch bereits durch die Bewegtheit des Reizes und nicht unbedingt durch dessen soziale Dimension zustande gekommen sein könnten. Ferner kann erst durch weitere Untersuchungen und besonders durch Untersuchungsreihen im Sinne von wiederholten Testungen oder Längsschnittuntersuchungen eruiert werden, ob es sich bei den gezeigten Befunden tatsächlich um anhaltende Phänomene handelt oder ob die Probanden nach längerer Zeit das Gefühl verlieren, beobachtet zu werden resp. überhaupt mit einem sozialen Wesen zu interagieren.

### 3 Funktionsebenen nonverbalen Verhaltens

Oggleich sich bei der visuellen Gestaltung virtueller Akteure natürlich auch Fragen des Erscheinungsbildes sowie des Darstellungsrealismus ergeben (vgl. Ball et al., 1997; Cassell et al., 1999; Parke, 1991), werden wir uns im folgenden auf die dynamischen Aspekte, d.h. auf das konkrete, visuell wahrnehmbare Verhalten von virtuellen Akteure und dessen mögliche Bedeutung für den Interaktionsverlauf zwischen Mensch und Computer konzentrieren. Tatsächlich scheint aus unserer Sicht die Frage nach dem ‘Wie’ des Verhaltens der prinzipiellen Frage nach dem ‘Ob’ des Einsatzes von anthropomorphen Interfaces vorgeordnet. Auch Rickenberg und Reeves (2000) stellen fest, dass es nicht ausreicht *„... to focus on whether or not an animated character is present. Rather the ultimate evaluation is similar to those for real people - it depends on what the*

*character does, what it says, and how it presents itself (p. 55)*“ und fordern: „... *that decisions concerning the use of animated characters should address the details of execution and social presentation*“. Gerade die Ausgestaltung kommunikativer Verhaltensdetails sowie die Analyse möglicher Wirkungen auf den Betrachter kann nun kaum der Intuition von Animatoren und Designern überlassen werden. Vielmehr ist hier ein systematischer Beitrag der Psychologie gefordert, der sich nicht in der nachträglichen Evaluation alternativer Implementierungen erschöpft, sondern vielmehr frühzeitig im Sinne der Realisationsforschung in die konkrete Technikentwicklung eingreift und Basiswissen für deren Gestaltung bereitstellt. Dabei ist jedoch kritisch zu prüfen, inwieweit diese Wissensvoraussetzung hinreichend für umfassende Implementationen sind bzw. für welche Teilbereiche sie genügen, und welche offenen Forschungsfragen gegebenenfalls noch beantwortet werden müssen.

Tatsächlich ist der Erkenntnisstand innerhalb der Psychologie bisher nur begrenzt geeignet, um umfassende regelbasierte Verhaltensgeneratoren bereitzustellen. Viele der vorhandenen Befunde sind von einem hohen Allgemeingrad oder beruhen auf Aggregatdaten, die zwar statistische Zusammenhänge zwischen der Häufigkeit nonverbaler Phänomene und spezifischen interpersonellen Wirkungen aufzeigen, die jedoch keine Aussagen über die konkrete Einbettung in den Interaktionsverlauf, über Kontextabhängigkeiten, Timing und Dosierung zulassen und entsprechend für die computer-basierte Produktion spontanen Interaktionsverhaltens wenig hilfreich sind. Allerdings lassen sich vor diesem Hintergrund einige Klassifikationsgesichtspunkte herausarbeiten, die eine differenziertere Problemsicht sowie realistische Planungsziele ermöglichen könnten. Prinzipiell lassen sich im Hinblick auf potentielle Wirkungen visuell wahrnehmbaren Verhaltens in der Realinteraktion vier bedeutsame Funktionsaspekte unterscheiden:

1. *Modellfunktionen (modelling functions)*: Hierunter sind instrumentell-kommunikative Operationen zu fassen, d.h. motorische Aktivitäten (im Sinne von Lokomotionen oder manipulativen Eingriffen in der realen bzw. virtuellen Objektwelt) die für den Betrachter Vorbildfunktion haben können (Zusammenbau eines elektronischen Gerätes).
2. *Diskursfunktionen (discourse functions)*: Hier steht die nonverbale Aktivität in engem Zusammenhang mit der Sprachproduktion und übernimmt substituierende, komplementäre oder supplementäre Funktionen (etwa illustrative und emblematische Gesten wie Zeigegesten oder 'beat-gestures', d.h. Kopf- und Blickbewegungen, die zur zeitlichen Strukturierung und Akzentuierung der verbalen Produktion genutzt werden können).
3. *Dialogfunktionen (dialogue functions)*: Hier dient das nonverbale Verhalten der Ablaufsteuerung der Kommunikation im Dienste eines möglichst reibungslosen Sprecher-Hörer-Wechsels und eines wenig überlappenden und unterbrechungsfreien Interaktionsverlaufs (turn-taking Signale wie Blickkontakt bzw. -abwendung, back-channel Signale wie Kopfnicken, Kopfschütteln, Augenbrauen hochziehen, etc.)
4. *Sozio-emotionale Funktionen (relational functions)*: Hierunter sind spezifische Wirkungen des nonverbalen Verhaltens im Bereich der interpersonellen Beziehung zu verstehen. Das nonverbale Verhalten hat einen nachhaltigen Einfluss auf die interpersonelle Eindrucksbildung (etwa im Sinne der Zuschreibung von Attraktivität, Sympathie, Glaubwürdigkeit, etc.) und spielt eine bedeutsame Rolle bei der Vermittlung von Intimität, der Ausübung von Macht und sozialen Kontrolle und der Auslösung von Emotionen.

Die hier vorgeschlagene Klassifikation umfasst in gewissem Sinne auch eine hierarchische Gliederung, wobei von Punkt 1 zu Punkt 4 das explizite Regelwissen ab- sowie die Ambiguität zunimmt. Während der Ablauf instrumenteller motorischer Operationen prototypisch zumeist durch die Aufgabe bzw. das zu manipulierende Objekt determiniert ist und interindividuelle Variationen vernachlässigbar sind, so klären diese Variationen unter Punkt 4 im Sinne des persönlichen Verhaltensstils und der emotionalen Anmutungsqualität einen Großteil der Varianz in der Personwahrnehmung und interpersonellen Eindrucksbildung auf. Im Bereich der Diskursfunktionen, insbesondere bei den Zeigegesten, sind ebenfalls person-externe Situations- und Aufga-

benmerkmale ablaufbestimmend (wo ist das referenzierte Objekt), obwohl bereits größere Freiheitsgrade in der Ausgestaltung der Geste vorhanden sind (Zeigen mit der Hand, mit dem Finger, mit einer Kopfbewegung, etc.). Auch für sprachbegleitende Illustratoren sind die Möglichkeiten durch Objektreferenzen innerhalb einer - wenn auch nicht explizit bestimmten - Grenze eingeschränkt. So lässt sich die Darstellung einer Wendeltreppe eben kaum durch eine lineare Handbewegung von rechts nach links verbildlichen. Personenunabhängige Verhaltensregeln lassen sich auch noch für Punkt 3 explizieren. Diese wurden vor allem auf empirischem Wege gewonnen, und lassen durchaus Implementierungen zu, so z.B. in bezug auf das Blickverhalten beim Sprecher-Hörerwechsel. Übergänge zu Punkt 4 finden sich hier jedoch bereits im Bereich des Back-Channel-Verhaltens. Die Häufigkeit bestätigenden Kopfnickens etwa kann von Person zu Person stark variieren und hat nicht nur Einfluss auf die Sprachproduktion des Gegenübers, sondern natürlich auch auf dessen interpersonelle Eindrucksbildung, die Gefühle des Sprechers und möglicherweise auch die Zu- bzw. Abneigung, die dieser zum Hörer entwickelt. Angesichts der Tatsache, dass die Modellierung instrumenteller Bewegungen kaum der psychologischen Unterstützung bedarf und das Wissen um die sozio-emotionale Komponente noch wenig entwickelt ist, nimmt es nicht Wunder, dass bisherige Implementation vor allem auf Punkt 2 und 3 Bezug genommen haben.

#### **4 Regelbasierte Implementationen sprachbegleitender Gestik und Mimik**

Regelbasierte Implementationen von nonverbalem Verhalten wurden in den letzten Jahren vor allem von der Arbeitsgruppe um Justine Cassell am MediaLab des MIT vorgelegt (Cassell et al., 1994). Sie modellieren ausschnitthaft die Beziehung von sprachlichem Inhalt, Intonation und illustrierenden Gesten. Auf der Grundlage elaborierter Modelle zum Zusammenhang von Gestik und Sprache (Bolinger, 1983; McNeill, 1992) formulieren die Autoren Regelwerke, die die autonome Steuerung einer virtuellen Figur in begrenztem Umfang ermöglichen. Einen vergleichbaren Ansatz verfolgt auch Thórisson (1996). Er umfasst basale turn-taking-Mechanismen (Duncan, 1972) sowie Zeigegesten. Diese werden im Rahmen einer Interaktions-Plattform implementiert, die auch eingabeseitig die Erfassung von Gesten und Blickrichtung (vgl. Bers, 1996) erlaubt. Ausgabeseitig simuliert das System Augenbewegungen, basale Mimik, Kopf- und einfache Handbewegungen. In der Interaktion kann sich der Benutzer durch den humanoiden Agenten „Gandalf“ über das Planetensystem informieren lassen. Neuere Entwicklungen der MIT-Arbeitsgruppe (Cassell et al., 1999) mündeten in eine computergenerierte Immobilienberaterin („Rea“, Real estate agent), die den Benutzer mittels Kameras (videobasiertes Tracking von Kopf und Händen) und Spracheingabe ‘verstehen’ kann und selbst über verschiedene Kommunikationskanäle, wie Sprache, Intonation, Mimik und Gestik verfügt. Doch selbst bei diesem umfassenderen Ansatz, der sein Hauptaugenmerk auf die Interaktionsstruktur im Sinne von Turn-Taking-Prozessen legt, bleiben zahlreiche kommunikative Aspekte unberücksichtigt. Die Modellierung von sozio-emotionalen Signale etwa beschränkt sich auf das Lächeln. Nagao und Takeuchi (1994) leisten bei ihrem Versuch, ein anthropomorphes Interface zu entwickeln, ebenfalls nur die Implementierung von Teilaspekten, in diesem Fall der sprachbegleitenden Mimik, die nach den Befunden von Chovil (1991) modelliert wurde.

Der unbestreitbare Vorteil einer regelbasierten Implementierung, die wir als „Top-Down-Ansatz“ bezeichnen wollen, besteht ganz ohne Zweifel darin, dass Forschungslücken sowie unklar formulierte Modelle durch die konkrete Simulation entlarvt werden. Ein wesentlicher Nachteil der Top-Down-Ansätze besteht nun aber darin, dass im Bereich des nonverbalen Verhaltens nur eine geringe Anzahl an Befunden und Modellen vorliegt, die sich in Algorithmen zur Verhaltensgenerierung umsetzen lassen. Dies liegt zum Teil an der Tatsache, dass die Mehrzahl der empirischen Befunde auf Aggregatdaten beruhen, d.h. dass zwar quantitative Zusammenhänge zwi-

schen bestimmten nonverbalen Verhaltensweisen und sozio-emotionalen Effekten zu finden sind, dass aber keine Prozessinformationen, d.h. Daten zur Kontextabhängigkeit und interaktiven Kontingenz der jeweiligen Verhaltensmuster vorliegen. Diese Beschränkung führt bei Top-Down-Ansätzen zwangsläufig dazu, dass diese nur einen sehr begrenzten Verhaltensbereich abdecken können, was im Effekt zu relativ verarmten grafischen Simulationen führen muss. Nur verhältnismäßig eng mit der Sprache verbundene Teilfunktionen (Zeigegesten, Illustratoren, Turn-Taking) wurden bislang berücksichtigt. Sozio-emotionale Aspekte, die einen bedeutenden Teil des nonverbalen Wirkungsspektrums ausmachen, wurden innerhalb von Top-Down-Ansätzen weitestgehend vernachlässigt. Der Versuch diese in die Entwicklung anthropomorpher Interface-Agenten einzubeziehen, wirft den Forscher allerdings auf das Problem der defizitären Befundlage zurück.

## 5 Das Forschungsprogramm: Entwicklungsplattform für anthropomorphe Interface-Agenten

Angesichts der für eine theoriegeleitete Implementierung von anthropomorpher Interface-Agenten kaum verwertbaren kommunikationspsychologischen Befundlage schlagen wir insbesondere mit Blick auf die sozio-emotionalen Funktionen nonverbalen Verhaltens einen datengestützten „Bottom-Up-Ansatz“ vor. Dabei werden auf der Grundlage detaillierter Verhaltensprotokolle natürlicher Interaktionen, die auf dem Wege spezifischer deskriptiver Kodierverfahren (Berner System zur Zeitreihennotation des Bewegungsverhalten, Facial Action Coding System) und/oder mittels automatischer Bewegungsdetektion (Motion Capture, Mimik-Tracking, Eye Tracking, Data-Gloves) gewonnen werden (vgl. Bente, 1989; Bente, Feist & Elder, 1996; Bente, Krämer, Petersen & Buschmann, 1999) 3D-(Re-)Animationen menschlichen Interaktionsverhaltens erstellt und als annotierte ‘Verhaltenskonserven’ in Datenbanken abgelegt. Wir konnten zeigen, dass auf der Grundlage solcher Datensätze - selbst wenn sie per Video-Transkription gewonnen wurden -, äußerst realistische Animationen möglich sind und dass die sozio-emotionalen Wirkungen derartiger 3D-Animationen jenen vergleichbar sind, die von den Original-Videoaufzeichnungen ausgehen (Bente, Petersen, Buschmann & Krämer, 1999). Der Vorteil dieser Vorgehensweise liegt nun einerseits darin, dass alle Verhaltenskanäle aufgrund der Ausgangsdaten vollständig modelliert werden können, andererseits können - mit Hilfe spezifischer Editoren - systematische experimentelle Variationen an den Ausgangsdatenprotokollen vorgenommen werden. Die auf diesem Wege erstellten - in bezug auf einzelne Verhaltensaspekte divergierende Sequenzen - können nun beurteilt werden, so dass semantische Zuschreibungen auf empirischem Wege gewonnen und entsprechende Annotationen vorgenommen werden können. Es lassen sich auf diesem Wege nicht nur Verhaltensindizierungen für die spätere Dialogmodellierung vornehmen, sondern es können auch systematische Wirkungszusammenhänge aufgedeckt werden, z.B. hinsichtlich des Einsatzes von Blickkontakt, Lächeln, Kopfwendung u.a.m. Diese Erkenntnisse können dann wiederum im Rahmen von Top-Down-Ansätzen zur regelbasierten Verhaltensgenerierung benutzt werden. Eine entsprechende experimentelle Entwicklungsplattform wurde von den Autoren im Rahmen des vom BMBF geförderten Leitprojektes EMBASSI (Elektronische Multimodale Bedien- und Serviceassistenten) in Kooperation mit dem „Laboratory for Mixed Realities“ der Kunsthochschule für Medien, Köln, bereitgestellt. Kernstück der Plattform sind die erwähnten Verhaltenseditoren, die als intuitive Schnittstelle zwischen Experimentator und den Bewegungsdaten ausgelegt sind. Diese Verhaltenseditoren verfügen über multiple Konverterprogramme, die eine Übersetzung in verschiedenste Datenformate zulassen sowie über ein echtzeitfähiges 3D-Render-Modul, in dem Veränderungen der Datenbasis unmittelbar als Animation angezeigt werden können.

Die im Rahmen von Bottom-Up-Modellierung erstellten Animationen gewähren allerdings nur eine eingeschränkte Interaktivität und können somit nur einen Zwischenschritt zur Gestal-

tung anthropomorpher Interfaces darstellen. Auf diesem Wege sieht unsere Entwicklungs- und Evaluationsplattform nun verschiedene experimentelle Testanordnungen vor, die jeweils unterschiedliche Ansprüche an die Interaktivität des Systems stellen:

1. *Passive Observation Paradigm*: In diesem Fall befindet sich der evaluierende Beobachter in einer passiven Beobachtungssituation, bei der er eine computer-animierte Interaktion beobachtet und im Anschluss die gesehenen Personen bewertet. Dieses Paradigma ist besonders geeignet, um unter Einsatz von systematischen Variationen bestimmter Aspekte die spezifische Wirkung eben dieser Verhaltensweisen zu ergründen.
2. *Polled Response Paradigm*: Der Benutzer wird durch einen virtuellen Assistenten auf Anfrage hin informiert und beurteilt die Interaktion im Anschluss. Neben der Möglichkeit zur Gewinnung von Erkenntnissen zur Wirkungsweise spezifischer nonverbaler Signale durch systematische Variation, können in diesem Rahmen erste kontrollierte Studien zur Akzeptanz von anthropomorphen Schnittstellen durchgeführt werden. In einem fortgeschritteneren Stadium werden erste interaktive Elemente ergänzt (z.B. Beachtung der Blickrichtung des Akteurs, Berücksichtigung von Turn-Taking-Signalen).
3. *Hidden Expert Paradigm*: Volle Interaktivität ist derzeit technisch noch nicht zu realisieren. Um Chancen und Risiken des angestrebten Zukunftsszenarios abschätzen zu können, wird Interaktivität simuliert. Das nonverbale Verhalten eines realen Helfers, der den Nutzer über Video beobachtet, wird per motion capture erfasst und unmittelbar auf den virtuellen Assistenten übertragen. Hierbei stehen neben Fragen der Akzeptanz und Effektivität bei der Bearbeitung verschiedenster Aufgaben auch motivations- und sozialpsychologische Aspekte im Vordergrund.
4. *Embodied Interface Paradigm*: Die in den Phasen 1 bis 3 gewonnenen Erkenntnisse zur Funktionsweise nonverbalen Interaktionsverhaltens sollen hier im Rahmen von regelbasierten Implementierungen umgesetzt werden. Umfassende Verhaltensbibliotheken sollen mit Algorithmen zur kontextabhängigen Modifikation bzw. Generierung von Verhaltensprotokollen gekoppelt werden. Hierzu sind Dialogmanagementsysteme aufzubauen und Kontextmodelle bereitzustellen. Wesentliche Komponenten eines solchen Systems werden von den Projektpartnern in oben genanntem Leitprojekt EMBASSI entwickelt.

Um weiteren Einflußgrößen im Zusammenhang mit der Gestaltung anthropomorpher Interfaces Rechnung tragen zu können, stellt die von uns entwickelte Plattform auch Möglichkeiten zur Variation des Erscheinungsbildes der Interface-Agenten bereit, die im Rahmen der experimentellen Untersuchungen als unabhängige Variablen eingesetzt werden. Wesentliche experimentelle Variationen betreffen weiterhin die Aufgabentypen (diese reichen von einfachen Bedienungsaufgaben von elektronischen Geräten bis zu komplexen Problemlösungs- und Entscheidungsaufgaben). Auch unterschiedliche Nutzergruppen (im Hinblick auf Geschlecht, Alter, Expertise, Persönlichkeitsvariablen) können - im Sinne von potentiell moderierenden Variablen - so weit als möglich Berücksichtigung finden. Auf seiten der abhängigen Variablen werden neben objektiven aufgabenorientierten Effektivitäts- und Effizienzkriterien und dem Nutzerverhalten (Aufmerksamkeit, nonverbale Reaktionen) insbesondere auch Akzeptanz sowie sozio-emotionale Bewertungen erfasst.

Erste experimentelle Untersuchungen basieren auf dem Polled Response Paradigm. Konkretes Ziel ist hierbei die Bestimmung der Akzeptanz von virtuellen Helfern im Rahmen einer Videorekorderprogrammierung. Insbesondere wird analysiert, inwieweit diese eine effektive und effiziente Hilfe im Vergleich zu herkömmlichen Bedienungsanleitungen darstellen. Zusätzlich werden dynamische Aspekte des nonverbalen Verhaltens des virtuellen Agenten (z.B. die Wirkung von unterschiedlichen Intensitäten des Bewegungsverhaltens und verschiedener Kopfhaltungen) variiert und deren Wirkung auf die Zuschreibung von spezifischen Personmerkmalen, wie sympathisch, freundlich, kompetent etc. untersucht. Abbildung 1 zeigt die Benutzeroberfläche unseres Programmes zur Benutzerführung bei der Videorekorder-Programmierung mit einem virtu-

ellen Assistenten. Die Interaktionsmöglichkeiten sind beschränkt. Der Nutzer kann lediglich durch Betätigung verschiedener Buttons verschiedene Informationsbatches abrufen oder den Helfer dazu veranlassen, seine Ausführungen zu wiederholen.

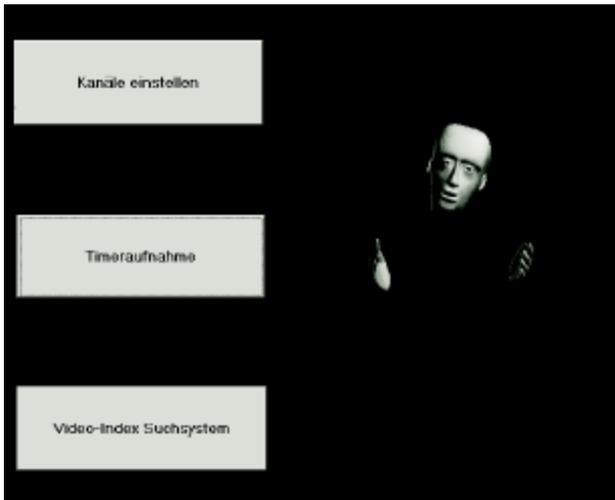


Abb.1: Virtuelle Assistenz bei der Videorekorderprogrammierung

Die Ergebnisse unserer Pilotstudie können hier nicht en Detail wiedergegeben werden. Von besonderer Bedeutung für die aktuelle Diskussion erscheinen uns jedoch neben den inferenzstatistischen Befunden zur differentiellen Wirkung der experimentellen Verhaltensvariationen vor allem auch Beobachtungen bei der Versuchsdurchführung. So zeigen erste Erfahrungen, dass es sich bei der Aufgabe, einen Videorekorder zu programmieren, um eine durchaus anspruchsvolle und für einige Versuchspersonen belastende Anforderung handelt, die mit Hilfe einer anthropomorphen Figur aber in der überwiegenden Zahl der Fälle - auch aus Sicht der Experimentalteilnehmer - befriedigend gelöst werden konnte. Es wird allerdings deutlich, dass erwartungsgemäß eine höhere Interaktivität (variableres Aufsuchen spezifischer Informationen, aktives Eingreifen des Helfers bei falschem Vorgehen im Sinne eines intelligenten Systems) als angenehmer bzw. wünschenswert empfunden wird. Die insgesamt positive Bewertung eines virtuellen Helfers oder Gesprächspartners könnte somit folglich durch größere Interaktivität weiter verbessert werden. Bezüglich der Variation der dynamischen Komponenten lässt sich bereits für die subtile Veränderung der Bewegungsaktivität eine Beurteilungsdifferenz feststellen: Wird z. B. die Kopfbewegung leicht intensiviert, wird die virtuelle Person als femininer ( $T = -2,58$ ,  $df = 48,91$ ,  $p = ,013$ ), weniger zurückhaltend ( $T = 2,35$ ,  $df = 56$ ,  $p = ,022$ ), freundlicher ( $T = 1,733$ ,  $df = 58$ ,  $p = ,088$ ), aufmerksamer ( $T = 1,77$ ,  $df = 58$ ,  $p = ,082$ ) und zugewandter ( $T = -1,82$ ,  $df = 58$ ,  $p = ,073$ ) erlebt. Die Urteilsdifferenzen für beide experimentelle Bedingungen sind in Abbildung 2 dargestellt. Diese ersten Analysen deuten darauf hin, dass bei der Implementation nonverbaler Verhaltensausgaben auf subtilste Variationen geachtet werden muss, wenn vermieden werden soll, dass potentielle Effizienzsteigerungen in der Systembedienung nicht durch massive Akzeptanzprobleme neutralisiert werden.

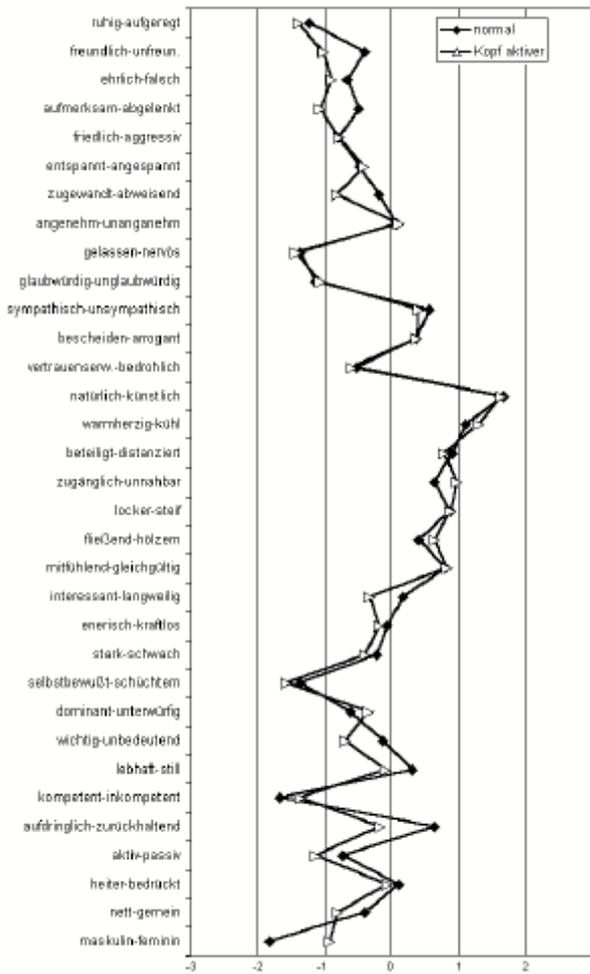


Abb. 2: Urteilsprofile (Mittelwerte) der beiden experimentellen Bedingungen

## 6 Fazit: Aufgabenstellungen für die Psychologie

Bislang wurden viele der Fragen, die mit der Einführung der sozialen Dimension im Rahmen der Mensch-Computer-Interaktion verbunden sind, noch kaum systematisch untersucht (vgl. Dehn und van Mulken, 2000). Auch die Evaluation von anthropomorphen Schnittstellen im Sinne der Überprüfung von Akzeptanz, Effizienz/Effektivität und Verhalten der Benutzer im Umgang mit der neuen Technologie weist noch zahlreiche Lücken auf (vgl. Bente & Krämer, 2000). So vermisst man bislang etwa die kontrollierte Berücksichtigung potentiell Einfluss nehmender Faktoren wie des Aufgabentyps oder eine differenzierende Berücksichtigung unterschiedlicher Nutzergruppen. Doch nicht nur im Rahmen der Evaluation werden psychologisches Wissen sowie entsprechende Untersuchungsmethoden benötigt und bereits eingefordert, auch im Bereich der

Realisation stellen sich zahlreiche Probleme etwa bezüglich der Modellierung der im face-to-face-Kontakt salient werdenden nonverbalen Kommunikation, die einen Beitrag von Psychologie und Kommunikationsforschung zur Interface-Entwicklung angezeigt erscheinen lassen. Es lassen sich somit drei miteinander verschränkte Problembereiche aufzeigen, bei deren Bearbeitung die Psychologie zentral beitragen kann:

1. *Grundlagenforschung*: Theorieentwicklung und experimentelle Modellüberprüfung zur Wirkung nonverbaler Kommunikation; Transfer auf die Mensch-Computer-Interaktion
2. *Evaluationsforschung*: Überprüfung konkreter Implementierungen im Hinblick auf Akzeptanz, Effektivität/Effizienz und Auswirkungen auf den Benutzer
3. *Realisationsforschung*: Implementierung und Validierung von Verhaltensdatenbanken und Algorithmen zur Verhaltensgenerierung und Dialogsteuerung in konkreten Anwendungsszenarien.

## 7 Literatur

- Ball, G., Ling, D., Kurlander, D., Miller, D. Pugh, D., Skelly, T., Stankosky, A., Thiel, D., Van Dantzych, M. & Wax, T. (1997). Lifelike computer characters: the persona project at Microsoft Research. In J.M. Bradshaw (Ed.), *Software agents*. Cambridge, MA: MIT Press.
- Bente, G. (1989). Facilities for the graphical computer simulation of head and body movements. *Behavior Research Methods, Instruments, & Computers*, 21 (4), 455-462.
- Bente, G., Feist, A. & Elder, S. (1996). Person perception effects of computer-simulated male and female head movement. *Journal of Nonverbal Behavior*, 20 (4), 213-228.
- Bente, G. & Krämer, N. C. (2000). Virtuelle Gesprächspartner: Psychologische Beiträge zur Entwicklung und Evaluation anthropomorpher Schnittstellen. In K. P. Gärtner (Hrsg.), *Multimodale Interaktion im Bereich der Prozessführung*. 42. Fachausschusssitzung Anthropotechnik, DGLR-Bericht 2000-02 (S. 29-50). Bonn: Deutsche Gesellschaft für Luft- und Raumfahrt.
- Bente, G., Petersen, A., Krämer, N.C. & Buschmann, J.-U. (1999). Virtuelle Realität im Forschungseinsatz. Ein Wirkungsvergleich videovermittelter und computersimulierter nonverbaler Kommunikation. *Medienpsychologie: Zeitschrift für Individual- und Massenkommunikation*, 2, 95-120.
- Bers, J. (1996). A body model server for human motion capture and representation. *Presence: Teleoperators and virtual environments*, 5 (3), 381-392.
- Bolinger, D. (1983). Intonation and gesture. *American Speech*, 58 (2), 156-174.
- Bolt (1987). The integrated multi-modal interface. *The Transactions of the Institute of Electronic Information and Communication Engineers, (Japan)*, November, 2017-2025.
- Brennan, S. (1990). Conversation as direct manipulation: An Iconoclastic View. In B. Laurel (Ed.), *The Art of Human-Computer Interface Design*. (pp. 393-404). Reading: Addison-Wesley.
- Cassell, J. Bickmore, T., Billinghurst, M., Campbell, L., Chang, K, Vilhjálmsón, H., Yan, H. (1999). Embodiment in conversational interfaces: Rea. *CHI '99 Conference Proceedings* (pp.520-527). Association for Computing Machinery.
- Cassell, J., Steedman, M., Badler, N. Pelachaud, C., Stone, M. Douville, B., Prevost, S. & Achorn, B. (1994). Modeling the interaction between speech and gesture. In A. Ram & K. Eiselt (eds.), *Proceedings on the sixteenth annual conference of the cognitive science* (pp. 153-158). LEA. Verfügbar unter: <http://gn.www.media.mit.edu/groups/gn/publications.html>.
- Chovil, N. (1991). Discourse oriented facial displays in conversation. *Research on Language and Social Interaction*, 25, 163-194.
- Dehn, D. M. & van Mulken, S. (2000). The impact of animated interface agents: a review of empirical research. *International Journal of Human-Computer Studies*, 52, 1-22.
- Duncan, S., Jr. (1972). Some signals and rules for taking speaking turns in conversations. *Journal of Personality and Social Psychology*, 23 (2), 283-292.
- Goffman, E. (1959). *The presentation of self in everyday life*. New York: Doubleday & Company.
- Koda, T. & Maes, P. (1996). Agents with faces: the effect of personification. *Proceedings of the 5<sup>th</sup> IEEE International Workshop on Robot and Human Communication (RO-MAN'96)*, 189-194.
- Laurel, B. (1990). Interface agents: Metaphors with character. In B. Laurel (Ed.), *The Art of Human-Computer Interface Design*, (pp. 355-365). Reading: Addison-Wesley.

- Leary, M. R. (1995). *Self-presentation: Impression management and interpersonal behavior*. Madison: Brown & Benchmark.
- Lester, J. C., Converse, S. A., Kahler, S.E., Barlow, S. T., Stone, B. A. & Bhogal, R. S. (1997). The persona effect: affective impact of animated pedagogical agents. In S. Pemberton (Ed.), *Human Factors in Computing Systems: CHI '97 Conference Proceedings* (pp. 359-366). New York: ACM Press.
- McNeill, D. (1992). *Hand and mind: what gestures reveal about thought*. Chicago: University of Chicago.
- Milewski, A. E. & Lewis, S. H. (1997). Delegating to software agents. *International Journal of Human Computer Studies*, 46 (4), 485-500.
- Nagao, K. & Takeuchi, A. (1994). *Social Interaction: Multimodal Conversation with Social Agents*. 12th National Conference on Artificial Intelligence (AAAI), 22-28.
- Oviatt, S. (1999). Ten myths of multimodal interaction. *Communications of the ACM*, 42 (11), 74-81.
- Parise, S., Kiesler, S., Sproull, L. & Waters, K. (1999). Cooperating with life-like interface agents. *Computers in Human ehavior*, 15, 123-142.
- Parke, F. I. (1991). Techniques of facial animation. In N. Magnenat-Thalmann & D. Thalmann (Eds.), *New trends in animation and visualization* (pp. 229-241). Chichster: John Wiley & Sons.
- Rickenberg, R. & Reeves, B. (2000). The effects of animated characters on anxiety, task performance, and evaluations of user interfaces. *Letters of CHI 2000*, April 2000, 49-56.
- Schlenker; B. R. (1980) *Impression management*. Monterey: Brooks-Cole.
- Sproull, L., Subramani, M., Kiesler, S., Walker, J.H. & Waters, K. (1996). When the interface is a face. *Human Computer Interaction*, 11 (2), 97-124.
- Takeuchi, A. & Naito, T. (1995). Situated facial displays: towards social interaction. In I. Katz, R. Mack, L. Marks, M.B. Rosson & J. Nielsen (Eds.), *Human factors in computing Systems: CHI '95 Conference Proceedings*, pp. 450-455. New York: ACM Press.
- Thórisson, K. R. (1996). *Communicative humanoids. A computational model of psychosocial dialogue skills*. PHD-Thesis, MIT.
- van Mulken, S., André, E. & Müller, J. (1998). The persona effect: how substantial is it? In H. Johnson, L. Nigay & C. Roast (Eds.), *People and Computers XIII: Proceedings of HCI '98*, pp. 53-66. Berlin: Springer.

## Adressen der Autoren

Prof. Dr. Gary Bente / Nicole C. Krämer  
Universität zu Köln  
Psychologisches Institut  
Bernhard-Feilchenfeld-Str. 11  
50969 Köln  
bente@uni-koeln.de  
nicole.kraemer@uni-koeln.de