Virtual gestures: Analyzing social presence effects of computer-mediated and computer-generated nonverbal behaviour

Gary Bente & Nicole C. Krämer
Department of Psychology, University of Cologne
Bernhard-Feilchenfeld-Str. 11, D-50969 Köln
bente@uni-koeln.de; nicole.kraemer@uni-koeln.de

Keywords: avatar-based communication, nonverbal communication

Summary

The paper presents recent technologies and basic research paradigms for the investigation of social presence effects of nonverbal behaviour in human-computer-interaction (HCI) and computer-mediated-communication.(CMC). An experimental platform for the coding, editing and experimental computer-animation of human movement behaviour is introduced, which can serve for the experimental variation and offline rendering of nonverbal interactions sequences in fundamental research as well a for the real-time transmission and 3D-animation of nonverbal behaviour in CMC. From both research areas information can be gathered on the implicit grammar and semantics of nonverbal behaviour in human interactions. This basic knowledge is used to implement and evaluate nonverbal encoding and decoding abilities and dialog management systems in so called anthropomorphic interface agents, which can be a first step towards the development of a kind of Artificial Social Intelligence (ASI).

1 Computer-mediated and computer-generated nonverbal behavior

Nonverbal behaviour (NVB), i.e. gestures, facial displays, body postures and movement, plays an important role in face-to-face (ftf) communication (Argyle, 1975; Grammer, Filova & Fieder, 1997; Mehrabian & Ferris, 1967). Along with the development of sophisticated tools for the creation of life-like computer-generated characters the question has been raised repeatedly how computer-mediated communication (CMC) and human-computer-interaction (HCI) can benefit from the introduction of nonverbal signals. Application areas are for example avatar-based conference systems in CMC (Petersen, Bente & Krämer, 2002) and embodied conversational agents in HCI (Cassell et al., 2000; Lester et al., 2000). Although there are common technical problems with respect to realistic 3D-animations, the psychological questions connected to these applications seem to be different. Sundar and Nass (2000) noted: "The basic distinction between CMC and HCI lies in the object of users´ psychosocial attribution: Whereas users respond to other users in CMC, they respond directly to the technology in HCI. The computer is a source in HCI, not just a medium. …HCI is not simply a special case of CMC. The psychology of CMC is quite different - and at least partly independent - from the psychology of HCI (p. 699)". Thus, within the context of HCI we assume nonverbal behaviour to be crucial for creating an illusion of a social interaction, while within CMC the availability of nonverbal channels can be expected to have an influence on the course and the outcome of mediated communication. From a social presence perspective, this distinction points to a conceptual difference between social presence as a qualitative and social presence as a quantitative variable that will be elaborated in our research.. In this line Biocca, Burgoon, Harms, and Stoner (2001) claimed: „The need for a theory of social presence is more pressing as the Internet and virtual environments become increasing social. With time we can observe an increase in social interaction not only among users, but also between users and computer agents. A robust and detailed theory and measure of social presence could contribute to our understanding and explaining social behavior in mediated environments, allow researchers to predict and measure differences among media interfaces, and to guide the design of new social environments and interfaces“ (p. 1).

Although conceptual distinctions are important with respect to the measurement of social presence either in CMC or in HCI, there is also a common final pathway of research concerning the communicational functions of NVB. Four functional levels of NVB have been identified that have to be considered in this context (see Bente, Krämer, Trogemann, Piesk & Fischer, 2001): (1) *Modelling functions* are connected to the fact that humans seem to have clear advantages in performing motor tasks when they can observe somebody else showing the required movements (Bandura, 1977; Rickel & Johnson, 2000), (2) *Discourse functions* are closely related to the verbal production of a communicator and can work as either complements, supplements or substitutes of speech. Pointing gestures, illustrative gestures, beat gestures structuring the speech flow or emblematic gestures belong to this functional category, (3) *Dialogue functions* include turn-taking signals (e.g. eye contact) and back-channel signals (e.g. head nods) and serve to smoothen the flow of interaction when exchanging speaker and listener roles, calling for more explicit information, or redirecting the content line of verbal exchange and (4) *Relational or socio-emotional functions* are the most unexplored in nonverbal communication research. Like in FTF-interaction nonverbal behaviour in animated characters could induce positive feelings in the vis-a-vis, increase motivation and thus facilitate task performance (Lester et al., 2000). This can be true for embodied interface agents as well as for avatar representations of human interlocutors. But also, as in every day life, the behaviour of the virtual partner can evoke negative feelings and hinder task performance.

Although there is a broad consensus about the important communicative functions of NVB there is still little knowledge about its specific 'mechanics': "We respond to gestures with an extreme alertness and, one might almost say, in accordance with an elaborate secret code, that is written nowhere, known to none, and understood by all" (Sapir, 1928). It will be demonstrated that with respect to unravelling the secrets of nonverbal communication CMC and HCI can be more then just challenging application areas. They also provide powerful research tools and suggest paradigmatic research designs. Using for example motion capture devices in an avatar-based CMC-experiments would lead to detailed protocols of nonverbal behaviour. Significant results concerning the influence of specific nonverbal cues or behavioural patterns can then be implemented in embodied interface agents and tested in CMC-applications. The paper will introduce an integrated platform for the analysis of computer-generated and computer-mediated nonverbal behaviour, named Virtual Communication Environment (VCE), and discuss specific research paradigms that are useful for answering basic question on the function of NVB as well as questions concerning the applications in CMC and HCI including social presence effects.

2 The Virtual Communication Environment (VCE)

The VCE is an avatar-based communication platform allowing:

(1) the real time interaction of two interlocutors including nonverbal signals like head movements, body movements, gestures and eye movement,
(2) the experimental variation of the visual appearance of the interlocutors,
(3) the online filtering of behavioural cues,
(4) the recordings of verbal and nonverbal behaviour,
(5) the interactive and/or algorithmic modification of behaviour protocols, and
(6) the offline rendering and display of interactions or selected behavioural batches.

Nonverbal behaviour is detected by means of Cybergloves, Polhemus-trackers and a high resolution eye-tracking system, that we developed for this purpose. Behaviour data is transmitted via Intranet (TCP-IP). Animation is performed by an AVI-CODEC developed for this purpose. This means that the transmitted, as well as the stored data, can be animated with any commercial AVI-player under Windows NT or 2000. The CODEC transforms angular data into 3D-animations. As no pictures but translation and rotation parameters are transmitted 25 times a second there is no speed problem at all.

With respect to the various fundamental as well as applied research questions at stake VCE allows for three distinct approaches. In particular the application of these research designs will depend on the knowledge on nonverbal communication and dialogue management as well as on the availability of technical devices such as systems for the video-based recognition of gestures and facial activities on the input side and for the real time generation and synchronization of verbal and nonverbal behaviour on the output side. Three design types can be distinguished at the moment (see Bente, Krämer, Trogemann, Piesk & Fischer, 2001):

(1) *Third Party Observation:* This is an offline paradigm, in which the human observer is in a passive reception situation. There is no interactivity between user and virtual actors. Interactions between two computer generated characters are presented introducing specific variations in the appearance and/or the nonverbal behaviour of the virtual actors. Observers emotional and cognitive responses to the virtual characters are measured on different levels (psychophysiological responses, gaze direction, ratings, etc.) The design is used to generate basic knowledge on the interpersonal effects of specific nonverbal cues and examine basic influences of technological sophistication (e.g. desktop vs. immersive VR, quality of animation and character design).

(2) *Script Driven Interaction:* The user is polling pre-recorded behaviours from a data base in an interactive problem solving task (e.g. VCR programming). The design allows for semi-interactivity. The 3D-model responds to certain requests of the user (e.g. asking for information on how to start timer-recording). Complete sequences of behaviour then are played from the data base. Systematic variations of static (appearance) and dynamic (nonverbal behaviour) cues can be applied. User requests are send to the system via mouse click or also via natural speech and gestures.

(3) *Contingent Interaction Paradigms:* Three paradigms can be distinguished here, all of them providing full range nonverbal behaviour and real time interactivity:

(a) *Hidden Partner Dialog* (HCI, CMC)*:* The user is interacting with a partner next door, not knowing that this is a real person. The virtual partner is connected to motion capture devices and represented by a real time 3D model to the user. User actions are transmitted via audio/video connection to the expert. The design type guarantees full interactivity and thus allows to explore the effects of anthropomorphic interfaces under optimal conditions. Appearance of the animated figure as well as communication tasks can be varied experimentally.

(b) *Avatar-based Telecommunication* (CMC)*:* This is a classical CMC situation where nonverbal aspects are included in the dialogue. Instead of webcams however, motion capture devices are used to detect the nonverbal behaviour that is time-synchronized with speech and transmitted via network. In contrast to video conferences the partners can stay anonymous and the physical appearance (even gender, age, etc.) can be changed experimentally. Thus the effects of nonverbal behaviour can be isolated from influences of other visual information. Nonverbal behaviour records are stored in data bases and can be implemented in our conversational interface agents.

(c) *Conversational Interface Agents* (HCI): This is the most challenging research paradigm and also a promising future application. Autonomous interface agents will be equipped with basic social intelligence in understanding natural speech and in decoding and encoding nonverbal behaviour. The realization of such a system will depend on the technological advancements in the area of speech and gesture recognition as well as on growing knowledge about the principles of spontaneous production of speech and gestures in face-to-face dialogues. First systems, although restricted in decoding capabilities and communicative repertoire have been introduced (Cassell et al., 2000; http://www.embassi.de).

3 Measuring the effects

Previous research has shown that animated characters on a computer screen can have multiple effects. They can induce anxiety, trigger impression management, or increase the sense of cooperation and trust (see Rickenberg & Reeves, 2000; Sproull et al., 1996). Also, the visual presence of a virtual vis-a-vis is likely to foster social facilitation processes (Zajonc, 1965). To answer questions about the influence of NVB on process and outcome in CMC and HCI in a differentiated way, various measures have to be taken to account for socio-emotional effects as well as for task performance and negative side effects. A focus of our research is on the measurement of social presence (SP) as a key phenomenon. According to Biocca, Harms, and Gregg (2001) three dimensions of SP will be taken into regard: co-presence, psychological involvement, and behavioural engagement. In addition to former approaches, however, we suggest to extend the questionnaire approach to objective behavioural data (gaze direction, facial activity, movement activity, response delays, etc.) and psychophysiological measurements.

4 Research examples

Conducting experimental research within the paradigm of *third party interaction* we were able to show that virtual persons evoke the same person perception processes than video recordings of "real" people. Participants observing the interaction of virtual characters reconstructed from movement protocols of real interactions (see figure 1a) formed the same impression of the interlocutors than those participants watching the original videotape. Additionally, we discovered that quality of movement seems to be more important for person perception than the sophistication of the 3D models - character design had only marginal influence on the judgements (see. Bente, Krämer, Petersen & de Ruiter, 2001).

As Rickenberg and Reeves (2000) noted it is not crucial „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). Against this background we ran a second experiment studying the effects of specific behavioural variations. For this purpose we introduced behavioural changes into real-life movement protocols. We focussed on lateral head position and general level of head movement activity (see figure 1a and 1b, showing the target person with upright head position and head tilted away from the interaction partner). In an experimental setting 160 participants observed and evaluated the computer animated persons. The results indicate that minor cues, such as activity and position of the head have an immense impact on the receivers´ evaluation of the observed character. However, the effect of activity seems to be strongly context-dependent, since increased activity was only evaluated positively when the interaction was not conflict laden (Krämer, 2001).

Figures 1a and 1b: Virtual characters interacting, head position of left interaction partner varied

An experiment using *script driven interaction* was conducted within the EMBASSI project (Multimodal Assistance for Infotainment and Service Infrastructures) funded by the German Ministry of Education and Research (BMBF). 87 participants had to perform various tasks in the programming of a VCR. Instead of a manual, that had to be used by 16 of the subjects, 71 participants were confronted with a virtual assistant on the TV-screen providing instructions for specific tasks (see figure 2). The most important result was that participants were more successful when the instruction was given by the natural speaking anthropomorphic interface agent (see Bente & Krämer, 2001). Ratings and likebility scores for the embodied interface agent also showed a high acceptance.

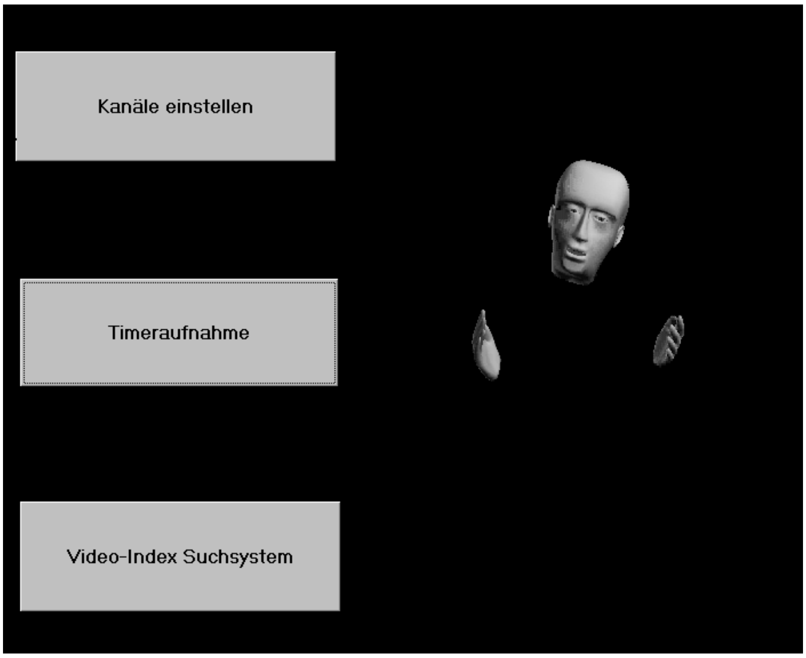

Figure 1: Embodied interface agent Jay

Within the *contingent interaction paraigms* the *hidden partner dialogue* is a very promising paradigm for future research in the realm of human computer interaction. We did a first test using this paradigm in a Wizard of Oz-like way: 90 participants were asked to spend some waiting time with a "virtual entertainment system" and to choose from various entertainment programmes proposed by the system. The interface was either a GUI combined with conventional VCR and TV-devices, a GUI that was projected on the wall with no conventional devices visible, a system producing natural speech, a virtual character that was presented on the screen of conventional devices and finally a virtual character that was projected on the wall without conventional devices being visible. The system and the virtual character were controlled by an experimenter in the next room who could observe the participant. Analyses that focus on the question which kind of representation users prefer are in progress.

Avatar-based communication is supposed to be a promising possibility to investigate social presence effects in CMC. Concerning applied research the use of avatars is thought of as a possibility to include nonverbal channels like gaze, gestures, body movements and facial activities into CMC and thus to render communication process more natural and more satisfying. In contrast to the use of webcams in net-based video-conferences avatar-based systems however bare some features that require special scientific attention with respect to process and outcome of CMC: (1) avatar-systems construct a virtual reality, i.e. an artificial meeting and/or working space that is psychologically remote to both locations of the interlocutors. Due to the common object reference in the virtual space it might be possible to produce higher degrees of social presence (see Biocca et al., 2001) and (2) avatar representations can transmit nonverbal information without disclosing the identity of the communicators or even leaving the choice of the physical appearance to him or her. Both aspects can have specific consequences for the socio-emotional processes involved in CMC, like impression management, mutual person perception, regulation of emotions, social facilitation, etc. In a first experiment we compared the effects of avatar-based interaction (the very reduced avatar displaying only eye movements and hand gestures is presented in figure 3) to face-to-face-interaction, videoconference and audio-conference (for the setting of the avatar-based interaction see figure 4). The dyads interacting in either of these modes did not differ with regard to collaborative joint task performance, content of speech output, person perception and various other dependent variables. The interactions however clearly differed with regard to basic turn taking mechanisms: Here avatar-communication resembled the behaviour during audio-communication, while face-to face- and video-based interaction contrasted with the former two (see Petersen, Bente & Krämer, 2002).

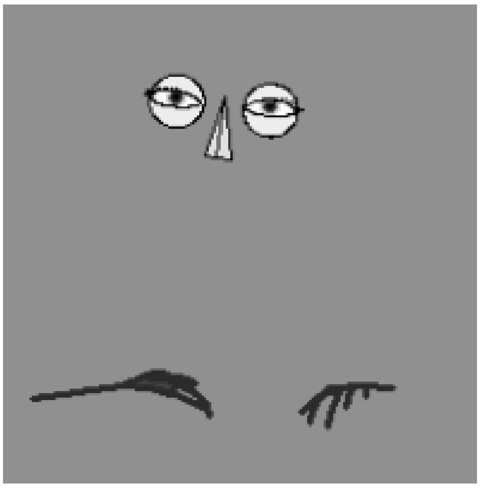

Figure 3: Avatar used in the experimental study

Figure 4: Setting for the interaction in VCE

Conversational interface agents have been successfully developed by the research group of Justine Cassell at the MIT (see Cassell et al., 2000). The German EMBASSI research group developed a home entertainment system recently that allows for the choosing of TV programmes and automated programming of a VCR by natural speech input. First evaluations of the system done in our lab led to promising results that were consistent with the literature (see Sproull et al., 1996; Rickenberg & Reeves, 2000). Although there were no significant differences between the various interfaces with respect to acceptance, we observed significant behavioural effects. So, when confronted with an embodied interface agent (see figure 5) the participants were more likely to use natural speech input instead of the remote control than when using a conventional GUI or a GUI combined with speech output (Krämer & Nitschke, 2002).

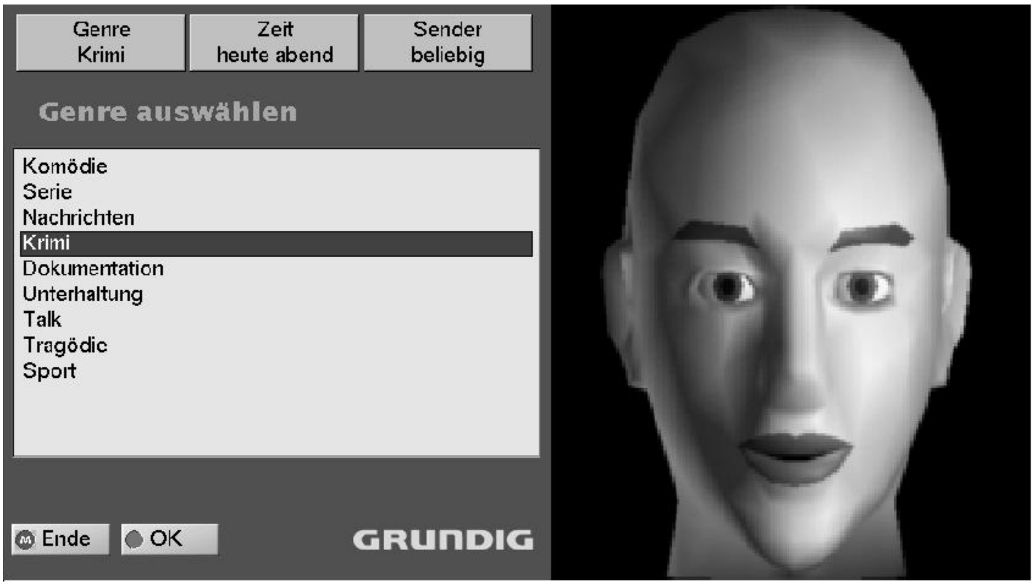

Figure 5: Embodied interface agent (produced by ZGDV, Darmstadt) and GUI (produced by Grundig)

5 References

Argyle, M. (1975). *Bodily communication*. London: Methuen.

Bandura, A. (1977). *Social learning theory*. Englewood Cliffs, NJ: Prentice Hall.

Bente, G. & Krämer, N. C. (2001). Psychologische Aspekte bei der Implementierung und Evaluierung nonverbal agierender Interface-Agenten. In H. Oberquelle, R. Oppermann, J. Krause (Hrsg.), *Mensch und Computer 2001* (S. 275-285). Stuttgart: Teubner.

Bente, G., Krämer, N. C., Petersen, A. & de Ruiter, J. P. (2001). Computer Animated Movement and Person Perception. Methodological Advances in Nonverbal Behavior Research. *Journal of Nonverbal Behavior, 25* (3), 151-166.

Bente, G., Krämer, N. C., Trogemann, G., Piesk, J. & Fischer, O. (2001). Conversing with electronic devices. An integrated approach towards the generation and evaluation of nonverbal behavior in face-to-face like interface agents. In A. Heuer & T. Kirste (Eds.), *Intelligent interactive assistance and mobile multimedia computing. Proceedings of the IMC2000* (pp. 67-76). Rostock: Neuer Hochschulschriftenverlag.

Petersen, A., Bente, G., & Krämer, N. C. (2002). Virtuelle Stellvertreter. In G. Bente, N. C. Krämer & A. Petersen (Eds.), *Virtuelle Realitäten* (pp. 227-253) . Göttingen: Hogrefe.

Biocca, F., Burgoon, J., Harms, C. & Stoner, M. (2001). Criteria and scope conditions for a theory and measure of social presence. Paper presented at *„Presence 2001“, 4th annual international workshop, May 2001, Philadelphia*. Available: http://nimbus.ocis.temple.edu/~mlombard/P2001/Biocca1.pdf [26.3.02].

Biocca, F., Harms, C. & Gregg, J. (2001). The networked minds measure of social presence: Pilot test of the factor structure and concurrent validity. Paper presented at *„Presence 2001“, 4th annual international workshop, May 2001, Philadelphia*. Available: http://nimbus.ocis.temple.edu/~mlombard/P2001/Biocca2.pdf [26.3.02].

Cassell, J., Bickmore, T., Campbell, L., Vilhjálmsson, H. & Yan, H. (2000). Human conversation as a system framework: Designing embodied conversational agents. In J. Cassell, J. Sullivan, S. Prevost & E. Churchill (Eds.), *Embodied conversational agents* (pp. 29-63). Cambridge: MIT Press.

Grammer, K., Filova, V. & Fieder, M. (1997). The communication paradox and a possible solution: Toward a radical empiricism. In A. Schmitt, K. Atzwanger, K. Grammer & K. Schäfer (Eds.), *New aspects of human ethology* (pp. 91-120). New York: Plenum.

Krämer, N.C. (2001). *Bewegende Bewegung. Sozio-emotionale Wirkungen nonverbalen Verhaltens und deren experimentelle Untersuchung mittels Computeranimation.* Lengerich: Pabst.

Krämer, N. C. & Nitschke, J. (2002). Ausgabemodalitäten im Vergleich: Verändern sie das Eingabeverhalten der Benutzer? In R. Marzi, V. Karavezyris, H.-H. Erbe & K.-P. Timpe (Hrsg.), *Bedienen und Verstehen. 4. Berliner Werkstatt Mensch-Maschine-Systeme* (S. 231-248). Düsseldorf: VDI-Verlag.

Lester, J. C., Towns, S. G., Callaway, C. B., Voerman, J. L. & FitzGerald, P. J. (2000). Deictic and emotive communication in animated pedagogical agents. In J. Cassell, J. Sullivan, S. Prevost & E. Churchill (Eds.), *Embodied Conversational agents* (pp. 123-154). Boston: MIT Press.

Mehrabian, A. & Ferris, S. R. (1967). Inference of attitudes from non-verbal communication in two channels. *Journal of Consulting Psychology, 31*, 248-252.

Rickel, J. & Johnson, W. L. (2000). Task oriented collaboration with embodied agents in virtual worlds. In J. Cassell, J. Sullivan, S. Prevost & E. Churchill (Eds.), *Embodied Conversational Agents* (pp.95-122). Cambridge: MIT-Press.

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.

Sapir, E. (1928). The unconscious patterning of behavior in society. In D. G. Mandelbaum (Ed.) (1949), *Selected writings of Edward Sapir* (pp. 544-559). Berkeley: University of California Press.

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.

Sundar, S. S. & Nass, C. (2000). Source orientation in human-computer interaction: programmer, networker, or independent social actor? *Communication Research*, *27* (6), 683-703.

Zajonc, R. B. (1965). Social facilitation. *Science*, *149*, 269-274.