

Methodological Considerations for Long-Term Experience with Robots and Agents

Tina Ganster, Sabrina C. Eimler, Astrid M. von der Pütten, Laura Hoffmann, Nicole C. Krämer

University of Duisburg-Essen
Forsthausweg 2
47048 Duisburg
Germany

{ tina.ganster@stud., sabrina.eimler@, astrid.von-der-puetten@, nicole.kraemer@, laura.hoffmann@ }uni-due.de

Abstract

This paper concentrates on the integration of methodological approaches, addressing the complexity and potential problems of research on long-term relationships with robots and agents. By presenting different categories of methods and outlining their potential and their benefits with regard to long-term human-agent interactions, solutions are suggested.

1. Introduction

Artificial entities have always been a popular part of science fiction stories. An often recurring aspect of these stories is that those characters are interactive entities that answer to their human conversational partners in an almost natural way and as a consequence, people get attached to them (e.g. Number Five or R2D2 from Star Wars). At the present time, we are still far away from this fictional vision but research already contributes to a similar development by promoting the creation and exploration of interactive agents and robots present in our lives. This includes continuous evaluation and improvement of their interaction with human beings. To achieve this, besides evaluating the agents' performance in terms of usability [Ruttkey et al., 2002; Ruttkey and Pelachaud, 2004], a special focus has to be laid on users' personal perception of a system. Aspects like the systems' helpfulness, believability or the users' satisfaction with and trust in the agent as well as the perceived engagement are important prerequisites for improving system designs over time. In the early days of agent research, researchers concentrated on short-term human-agent interaction. With steady progress in research, the focus more and more shifts towards the development of systems that are not merely temporarily applied within a laboratory environment but rather enable (and encourage) long-term interactions within a user's natural environment. Research on long-term relationships regularly turns out to be

costly and time-consuming since extensive preparation, repeated use of expensive measurement techniques have to be considered. Another crucial point also is the involvement of people who participate in an experiment for a longer time and thus have to be paid and taken care of (because of possible moral and ethical concerns). Since research in this area is only just emerging a lack of expertise cannot be denied. Often, methods and instruments that have proven to be helpful in short-term studies are transferred without knowing whether these are applicable and adequate for long-term interactions.

Against this background we intend to draw attention to several methodological aspects and discuss whether these may be applicable in long-term interactions. With the objective of improving data quality, reducing unnecessary costs and efforts, and increasing comparability of results throughout this research area, we suggest several aspects in need of improvement to motivate and facilitate research particularly on long-term interactions and with regard to an agent's social behavior.

The first aspect we consider to be important is the rather inconsiderate usage of objective and subjective measurement methods. We provide a list of widely established instruments and their respective benefits to the investigation of long-term relationships with agents. The second aspect particularly deals with the need to arrive at and to make use of more standardized measurement instruments. Within the field of objective measurement methods, we point out the possibility to use video and audio analysis to determine relevant changes in the course of the temporal development in the user-agents relationship.

2. Objective vs. Subjective Measurement Methods

According to Krämer [2002], relevant aspects of human-agent interaction can be categorized by distinguishing objectively measurable and subjectively measurable

ones. The former comprise for instance social effects and the user's experience during the interaction (often assessed by means of self-report scales) while aspects like user behavior and efficiency can be measured objectively (through observation).

Each way of measuring covers different phenomena in human-agent interaction but because of their specific strengths and weaknesses, they can be effectively combined to make results more conclusive. With regard to long-term interactions, particularly the usage of subjective measurement methods has to be reconsidered and adapted to serve this special purpose. In the following we will review subjective and objective measurements with regard to their respective benefits and shortcomings in long-term experiments.

Instruments of subjective measurement are commonly used in psychological research. A major part of it can be summarized as surveys [Fife-Schaw, 2000] that are usually filled in by participants at one or several points in time during or at the end of the study. With respect to human-agent interaction it can be stated that surveys are not conducted during the actual interaction but usually afterwards. Few measurement scales are particularly (and widely) used for the evaluation of human-agent interaction as for example the Agents Persona Instrument (API) by Baylor and Ryu [2003] and the Attitude Towards Agents Scale (ATAS) [Van Eck and Adcock, 2003]. With regard to the acceptance of agents, several approaches using survey are conceivable, including a general evaluation (using questionnaires with open- or closed-ended questions), evaluation of appearance [e.g. card sort assignments; e.g. Cowell and Stanney, 2003], perceived efficiency [e.g. Krämer and Nitschke, 2002], believability and trust in a system [e.g. Rickenberg and Reeves, 2000; Sproull et al., 1996; Krämer et al., 2005] as well as the personal feelings associated with the interaction [e.g. Krämer and Nitschke, 2002]. The measurement of social effects may comprise evaluations of socio-emotional aspects and person perception [e.g. Krämer, 2001] as well as para-social interaction [cf. Hartmann et al., 2001].

Surveys used in psychological research can be classified by their respective mode of data collection [Biemer and Lyberg, 2003]. Each mode may offer benefits as well as weaknesses, depending on the kind of data being gathered. Face-to-face interviews may offer good data quality because they provide high flexibility. But interviewer effects and social desirability have to be taken into account [Biemer and Lyberg, 2003]. Hoffmann et al. [2009] for instance showed that interviewer effects also apply to an agent whose performance was evaluated better when participants were confronted with the agent in contrast to filling out a paper-pencil-questionnaire. This also shows that the tendency for interviewer effects and socially desirable answers can be decreased through electronic or paper questionnaires.

Additionally, with respect to long-term interactions, it has to be pointed out that repeated surveys are much cheaper and easier to integrate when not performed face-to-face. In line with this, keeping diaries may as well turn

out to be particularly helpful in assessing interaction placed in the user's natural environment and at the same time provide qualitatively valuable data.

In sum, subjective measurement methods provide the opportunity to capture personal and subjective aspects of human-agent interaction. This kind of evaluation is essential for the aim of improving an agent's social behavior and thus its acceptance and likeability. As illustrated, subjective methods are likely to be of low complexity and more economic in contrast to objective methods (as will be shown later). But possible deficits have to be taken into account, for example interviewer effects, response rates or the fact that a survey normally is taken after the actual interaction (so the response is delayed) and only captures a snapshot in time. The latter aspect particularly represents a problem in long-term interactions since evaluation and memory performance are found to be distorted by primacy and recency effects when referring to a longer period of time. Another potential problem is referred to as reconstructive memory, meaning that the memories about an event are distorted due to incompatible experiences after the event [Atkinson & Shiffrin, 1968]. In the case of long-term interactions this implies that the emotions at the end of an experiment may dominate and overwrite the memory about emotions occurring at the beginning.

Objectively measureable effects of human-agent interaction relate to performance and user behavior from which an implicit measure of acceptance and likeability can be derived. A behavior that may directly refer to acceptance is for example the decision to use an agent for specific tasks or not [Krämer, 2002]. This aspect is particularly relevant in long-term studies where the participant is free to decide. Activation or arousal can also be regarded as indicators for the dimension of acceptance. According to Krämer [2002], the efficiency of a system can be measured through learning effects, task accomplishments or the change of performance through social inhibition [cf. Rickenberg and Reeves, 2000]. Another dimension of objectively measureable aspects covers all kinds of user behavior, such as the use of natural language [cf. Krämer and Nitschke, 2002], impression-management behavior [measurable via social desirable behavior in questionnaire items; e. g. Sproull et al., 1996] - , reciprocal helping [behavior in social dilemmas; cf. Nass and Moon, 2000], attention, physiological arousal, or body movement.

A widely used instrument for objective measurements in research on human-computer and human-agent interaction is *eye tracking*. This term refers to a number of different techniques to conduct research on eye or head movement, respectively, that vary with regard to precision and intrusiveness [Duchowski, 2003]. Since intrusiveness may cause discomfort to the wearer and thus lead to distorted results, modern eye tracking technology tends to become more unintrusive, leaving out head fixation to allow a more natural user behavior. With the help of eye tracking, researchers can determine users' (focus of) attention (e.g. by observing gaze direction) and aspects perceived most interesting or cognitively challeng-

ing (e.g. by measuring duration of eye-gaze behavior). With regard to long-term interaction, it can be pointed out that unintrusive methods of eye tracking may be especially relevant to research but costs of excessive or even permanent usage have to be taken into account. Regular short-term use within a long-term experiment may prove as feasible.

As a second method, *psychophysiology* shall be outlined here. The term refers to a number of instruments measuring physiological arousal [Graham & Schandry, 2009] that are also commonly used for research on human-agent interaction. Among the techniques that are used most often today is Electrodermal Response (EDR), which cannot identify the quality of an emotion but is nevertheless highly sensitive. Furthermore, Electrocardiography (ECG) as well as Electroencephalography (EEG) are non-invasive methods, working with skin electrodes attached to the human body. Since experiments on long-term interactions tend to take place in the user's natural environment, only mobile instruments (e.g. ECG) are efficient in this respect.

One of the most complex and technically advanced instruments for objective measurements is the *Functional Magnetic Resonance Imaging (fMRI)*, a rather recently developed method of neuroimaging [Ogawa et al., 1990] that is able to visualize the level of activity of different parts of the human brain. fMRI-based experiments normally consist of a sequence of different single scans. Although the method is non-invasive, the technique is highly sensitive to movement so the respective body part has to be fixated, extremely decreasing the naturalness of the interaction situation. For this reason, fMRI is probably not functionally adequate for research on long-term interactions.

Another series of methods classified as objective is the *analysis of audio and video material* recorded during the interaction [e.g. Gratch et al, 2006, 2007; Kang et al., 2008a, 2008b]. Audio material can be analyzed to identify certain characteristics of the subject's use of language. Parameters for language analysis may for example be the number or lengths of utterances, hesitations, pauses or the number of incomplete words or conversational fillers [e.g. Gratch et al., 2006]. To increase reliability, the quantitative analysis of natural language can and should be combined with qualitative analysis, such as analysis on Self-Disclosure [Moon, 2000; von der Pütten et al., submitted] and Conversation Analysis [Ten Have, 2007].

Video analysis can be classified as a less intrusive and more accurate variation of direct observation since the subject is not observed during conversation, but subsequently. This also enables the observer to repeatedly look into the material thereby increasing reliability of the assessment. In video analysis, particularly nonverbal behavior is of interest. Since nonverbal communication is a complex phenomenon, one has to carefully select appropriate methods capable of capturing all relevant aspects [Krämer, 2008a]. On the one hand, the users' general behavior can be observed with respect to their way of interacting with an agent which can for instance be sup-

ported by systems like the Bernese System for Time Series Notation [Frey et al., 1980]. On the other hand, the focus can be laid on facial expressions. One of the oldest methods for the measurement and description of facial activity, the Facial Action Coding System (FACS) proposed by Ekman and Friesen [1978] is still used today. However, in the majority of studies, these dynamic aspects of long-term human-agent interaction are not taken into account, (see below).

All of these aspects are objectively observable, but immediate insights into the subjective user experience are not possible. Yet, certain insights can be gained from these observations, for example that a user would only show social behavior (e.g. impression management) to something he/she perceives as a social being (for an overview about social effects of artificial entities see [Krämer, 2008b]).

A large number of objective measurement instruments were presented, including the (potential) shortcomings and strengths for their application in long-term experiments. In contrast to subjective measurements, they allow researchers to capture user behavior and derive emotional reactions along a temporal dimension which represents an important aspect of long-term interactions. When considering a methodological approach, the different objective methods have to be weighted with regard to the respective studies' objective. Some instruments relating to physiological arousal do not allow the participant to move freely (e.g. EDA). This may result in discomfort for the participant who may not be able to act naturally on the one hand; on the other hand, this may interfere with other objectives, e.g. the observation of nonverbal behavior. In general, the complexity of most objective methods represents an obstacle in long-term interactions with respect to natural environments. Here, video and audio analyses seem to be quite unproblematic and unintrusive instruments. With regard to the effects that are observable via objective measurements, it can be stated that they can only indirectly capture the quality of internal emotional states. The level of arousal does not always give an indication of the quality of the corresponding emotion. Similar problems occur within natural language and video analysis. There exist certain methods to identify and quantify the observed behavior but the tendency for interpretation cannot entirely be ruled out.

A rather advantageous aspect of objective research methods that has up to now not captured many researchers' attention is the possibility to make use of the subject's output data (e.g. eye gaze behavior or heart rate response). Since objective measures are taken during interaction, they can be used as inputs in the further course of the interaction. This can be applied to various aspects of human-agent interaction: In human-human communication turn-taking is an important mechanism in the course of which a number of behavioral indicators for holding, taking and giving the turn can be observed. As eye contact is crucial in this regard, eye tracking data might be used as input information for the agent, either to know when he is to speak or in order to signal the user

that he has the floor. This has for example been implemented in the Real Estate Agent that uses eye contact to realize turn-taking [Bickmore, 2003; Cassell et al., 2002] as well as Jonsdottir and Thórisson's [2009] system that analyses pauses in speech flow to regulate turn-taking. With respect to showing interest in a user, eye-gaze behavior can be used to establish mutual eye contact between the agent and the user [Bee et al., 2009] on the one hand. On the other hand, the agent might make use of this information in order to change its behavior and make the user feel more comfortable.

In this respect, objective methods have benefits as well as disadvantages compared to subjective ones. Shortcomings and benefits of each may be outweighed by applying an appropriate mixture of both kinds to cover more relevant aspects of human-agent interaction. A common example would be using psycho physiological instruments to determine the level of arousal while capturing the quality of the corresponding emotion via questionnaire.

3. Using Standardized Instruments

Benefits resulting from the usage of applicable and appropriate methods are dependent on their validity. Since the area of human-agent interaction is a rather new topic in psychological research, only few standardized subjective instruments exist. Some of them were already mentioned at the beginning, such as the Agents Persona Instrument (API) by Baylor and Ryu [2003] and the Attitude Towards Agents Scale (ATAS) [Van Eck and Adcock, 2003; Beun et al., 2003]. Bartneck et al. [2009] stress the need to arrive at more standardized measurement instruments for the categories they have identified as key concepts in human-robot interaction, such as likeability, perceived intelligence, and perceived safety. They propose a set of semantic differential questionnaires, the *Godspeed questionnaires*, covering these categories.

The use of psychological measurement instruments was already suggested, since in human-agent interaction, person perception may play an important role. When it comes to measuring this perception, one can observe a tendency to use ad-hoc questionnaires including single items representing complex dimensions of person perception. Such ad-hoc methods do not only increase difficulties in comparing and contrasting results from different studies, one also has to question the validity of such a questionnaire in contrast to an established and widely used psychological scale. Scales that may be useful in human-agent interaction are for example the Positive and Negative Affect Scale (PANAS) [Watson et al., 1988], which is often used when emotional experiences are evaluated, as well as the Self Assessment Mannequin Scale (SAM) [Lang et al., 1993]. There also exist transferrable questionnaires on person perception [Hurwitz et al., 1975] and the concept of presence (e.g. the Temple Pres-

ence Inventory by Lombard [2005]). Although some standards with respect to objective measurements have been listed, it has to be stated that there is a larger amount of standardized subjective measurements compared to objective ones. Creating objective measurement standards may turn out to be difficult since various different systems can be used to record data. These systems determine which aspects can be analyzed and whether standards can be applied.

4. Process Data

Looking at human-agent relations, we always find the interaction to be a process of a certain duration including a certain number of incidents and actions. As already pointed out, subjective measurements only allow the caption of the experience afterwards and at a specific point in time (except for the diary method). Changes that occurred during the respective period of time are not taken into account as was outlined in section 2 already.

Objective measures are more appropriate when relevant changes should be determined. For instance, Wada and Shibata [2006] repeatedly collected urine samples in the context of their long-term study with seal robots used for therapy in a care house. When presenting different methods of objective measuring, it became clear that a large part of these instruments allow the effects to be measured along a timeline, e.g. psychophysiological measures, and to relate possible changes to the single components of the interaction. The same opportunity is also offered by analyzing natural language and video material. But up to now, this has scarcely been made use of.

Despite researchers stressing the importance of methods enabling a dynamic, procedural mapping of nonverbal behavior [Cappella & Palmer, 1990; Monge & Kalman, 1996], most studies conducted rely on "distributional" instead of "temporal" data [Cappella & Palmer, 1990]. One of the few measuring systems for human movements that incorporate temporal aspects is the Bernese System for Time Series Notation [Frey et al., 1983; for an overview see Donaghy, 1989]. Within the research area of human-agent interaction, the opportunity to make use of "temporal" data with regard to nonverbal and verbal behavior is still widely missing.

5. Conclusion

With respect to long-term human-agent interaction, we stressed the need for the use of applicable research methods capable of handling the complexity of long-term interactions. We suggested possibilities to overcome shortcomings of previous research by presenting different categories of research methods and outlining their benefits with regard to human-agent relationship research.

When it comes to deciding whether to use objective or subjective measurement instruments, one has to consider not only the objective of the study but also the various benefits and disadvantages of methods presented. With regard to objective measurement methods it was shown that many of them still hold hidden potentials advantages to make evaluation more complex and reliable. We suggested that it would be advantageous to combine both kinds of methods to increase data quality and reliability.

The disadvantages of non-standardized, ad-hoc questionnaires were pointed out resulting in the suggestion to make use of already validated psychological scales and those standards existent among objective methods.

However, joint efforts have still to be undertaken to improve all kinds of methods. Advancements in the research area will depend on researchers' willingness to agree on (standardized) measures and to employ process measures even if this means additional effort.

References

- [Atkinson and Shiffrin, 1968] R. C. Atkinson and R. M. Shiffrin. Human Memory: A Proposed System and Its Control Processes. In K.W. Spence & J.T. Spence (Eds.). *The Psychology of Learning and Motivation*. Vol 2. Acad. Press, New York, 1968.
- [Bartneck et al., 2009] C. Bartneck, E. Croft, D. Kulic, and S. Zoghbi. Measurement instruments for the anthropomorphism, animacy, likeability, perceived intelligence, and perceived safety of robots. *International Journal of Social Robotics*, 1(1), pp. 71-81, 2009.
- [Baylor and Ryu, 2003] A. L. Baylor and J. Ryu. *The API (Agent Persona Instrument) for Assessing Pedagogical Agent Persona*. Honolulu, Hawaii: ED-Media, 2003.
- [Bickmore, 2003] T. W. Bickmore. *Relational Agents: Effecting Change through Human-Computer Relationships*. PhD, MIT, 2003.
- [Biemer and Lyberg, 2003] P. Biemer and L. Lyberg. *Introduction to Survey Quality*. John Wiley & Sons, New Jersey, 2003.
- [Bee et al., 2009] N. Bee, E. André and S. Tober. Breaking the Ice in Human-Agent Communication: Eye-Gaze Based Initiation of Contact with an Embodied Conversational Agent Export. *9th International Conference on Intelligent Virtual Agents (IVA 2009)*, pp. 229-242. Springer, Berlin, 2003.
- [Cappella and Palmer, 1990] J. N. Cappella and M. T. Palmer. The structure and organization of verbal and non-verbal behavior: Data for models of production. In H. Giles & W. P. Robinson (Eds.), *Handbook of language and social psychology*, pp. 141-161. John Wiley and sons, Chichester, 1990.
- [Cassell et al., 2002] J. Cassell, T. Stocky, T. Bickmore, Y. Gao, Y. Nakano, K. Ryokai, D. Tversky, C. Vaucelle and H. Vilhjálmsson. MACK: Media lab Autonomous Conversational Kiosk. In *Proceedings of Imagina 2002*. February 12-15 2002, Monte Carlo.
- [Cowell and Stanney, 2003] A. J. Cowell, and K. M. Stanney. Embodiment and interaction guidelines for designing credible, trustworthy embodied conversational agents. In T. Rist, R. Aylett, D. Ballin, and J. Rickel (Eds.), *Proceedings of the Fourth International Working Conference on Intelligent Virtual Agents (IVA2003)*, pp. 301-309. Springer Verlag, Berlin, 2003.
- [Donaghy, 1989] W. C. Donaghy. Nonverbal communication measurement. In P. Emmert and L. Barker (Eds.), *Measurement of communication behaviour*, pp. 296-332. White Plains, NY: Longman, 1989.
- [Duchowski, 2003] A. T. Duchowski. *Eye Tracking Methodology: Theory and Practice*. Springer, London, 2003.
- [Ekman and Friesen, 1978] P. Ekman and W. Friesen. *Facial action coding system: A technique for the measurement of facial movement*. Consulting Psychologists Press, Palo Alto, California, 1978.
- [Fife-Shaw, 2000] C. Fife-Schaw. Surveys and Sampling Issues. In G. M. Breakwell, S. Hammond, & C. Fife-Schaw (Eds.), *Research methods in psychology*, pp. 88-104. Sage Publications, London, 2000.
- [Frey et al., 1983] S. Frey, H.-P. Hirsbrunner, A. Florin, W. Daw, and R. Crawford. A unified approach to the investigation of nonverbal and verbal behavior in communication research. In W. Doise and S. Moscovici (Eds.), *Current issues in European Social Psychology*, pp. 143-199. Cambridge University Press, Cambridge, 1983.
- [Frey et al., 1981] S. Frey, H.-P. Hirsbrunner, J. Pool and W. Daw. Das Berner System zur Untersuchung nonverbaler Interaktion: I. Die Erhebung des Rohdatenprotokolls. In P. Winkler (Ed.). *Methoden der Analyse von Face-to-Face-Situationen*, pp. 203-236. Metzler, Stuttgart, 1981.
- [Graham and Schandry, 2009] K. Graham and R. Schandry. *Lehrbuch Psychophysiologie: Körperliche Indikatoren psychischen Geschehens*. Beltz Psychologie Verlags Union, Weinheim, 2009.
- [Gratch et al., 2006] J. Gratch, A. Okhmatovskaia, F. Lamothe, S. Marsella, M. Morales and R. J. van der Werf. *Virtual Rapport*. Marina del Rey, CA, 2006.
- [Gratch et al., 2007] J. Gratch, N. Wang, J. Gerten, E. Fast and R. Duffy. Creating rapport with virtual agents. *7th International Conference on Intelligent Virtual Agents*. Paris, France, 2007.
- [Hartmann et al., 2001] T. Hartman, C. Klimmt and P. Vorderer. Avatare: Parasoziale Beziehungen zu virtuellen Akteuren. *Medien- und Kommunikationwissenschaft*, 49, pp. 350-268, 2001.
- [Hoffmann et al., 2009] L. Hoffmann, N. C. Krämer, A. Lam-Chi and S. Kopp. Media Equation Revisited: Do Users Show Polite Reactions towards an Embodied Agent? In Z. Ruttkay (Eds.), *IVA 2009*. LNAI 5773, pp. 159-165. Springer-Verlag, Berlin, 2009.
- [Hurwitz, 1975] D. Hurwitz, N. Wiggins, and L. Jones. A semantic differential for facial attribution: The face

- differential. *Bulletin of the Psychonomic Society* (6), pp. 370-372, 1975.
- [Jonsdottir and Thórisson, 2009] G. R. Jonsdottir and K. R. Thórisson. Teaching Computers to Conduct Spoken Interviews: Breaking the Realtime Barrier with Learning. In Z. Ruttkay (Ed.), *IVA 2009*. LNAI 5773, pp. 446-459. Springer-Verlag, Berlin, 2009.
- [Kang et al, 2008a] S.-H. Kang, J. Gratch, N., Wang and J. H. Watts. Does Contingency of Agents' Nonverbal Feedback Affect Users' Social Anxiety? *Proceedings of the 7th international joint conference on Autonomous agents and multiagent systems*, pp. 120-127. Estoril, Portugal: International Foundation for Autonomous Agents and Multiagent Systems, 2008.
- [Kang et al, 2008b] S.-H. Kang, J. Gratch, N. Wang, and J. H. Watts. Agreeable people like agreeable virtual humans. In H. Prendinger, J. Lester, and M. Ishizuk (Eds.), *IVA 2008*. LNAI 5208, pp. 253-261. Tokyo, Japan: Springer Verlag, Berlin, 2008.
- [Krämer, 2001] N. C. Krämer. *Bewegende Bewegung. Sozio- emotionale Wirkungen nonverbalen Verhaltens und deren experimentelle Untersuchung mittels Computeranimation*. Pabst, Lengerich, 2001.
- [Krämer, 2002] N. C. Krämer. Können virtuelle Helfer uns wirklich helfen? – Methoden zur Evaluation von anthropomorphen Mensch-Maschine-Schnittstellen. *Zeitschrift für Medienpsychologie*, 14 (1), pp. 34-37, 2002.
- [Krämer, 2008a] N. C. Krämer. Nonverbal Communication. In J. Blascovich and C. Hartel (Eds.), *Human behavior in military contexts*, pp. 150-188. The National Academies Press, Washington, 2008.
- [Krämer, 2008b] N. C. Krämer. *Soziale Wirkungen virtueller Helfer. Gestaltung und Evaluation von Mensch- Computer-Interaktionen*. Kohlhammer, Stuttgart, 2008.
- [Krämer et al., 2005] N. C. Krämer, I. Iurgel and G. Bente. Emotion und motivation in embodied conversational agents. In L. Canamero (Ed.), *Proceedings of the Symposium "Agents that Want and Like" (AISB 2005)*, pp. 55-61. SSAISB, Hatfield, 2005.
- [Krämer and Nitschke, 2002] N. C. Krämer and J. Nitschke. Ausgabemodalitäten im Vergleich: Verändern sie das Eingabeverhalten der Benutzer? In R. Marzi, V. Karavezyris, H.-H. Erbe and K.-P. Timpe (Eds.), *Bedienen und Verstehen. 4. Berliner Werkstatt Mensch-Maschine-Systeme*, pp. 231-248. VDI-Verlag, Düsseldorf, 2002.
- [Lang et al., 1993] P. J. Lang, M. Greenwald, M. M. Bradley and A. O. Hamm. Looking at pictures: Evaluative, facial, visceral, and behavioral responses. *Psychophysiology*, 30, pp. 261-273, 1993.
- [Lombard, 2005] M. Lombard, M. Temple Presence Inventory (TPI) questionnaire items and factor structures. Retrieved November 26, 2009 from http://astro.temple.edu/~lombard/research/P2scales_11-04.doc
- [Monge and Kalman, 1996] P. R. Monge and M. E. Kalman. Sequentiality, simultaneity and synchronicity in human communication. In J. H. Watt and C. A. VanLear (Eds.), *Dynamic patterns in communication processes*, pp. 71-92. Sage Publications, Thousand Oaks, 1996.
- [Moon, 2000] Y. Moon. Intimate Exchanges: Using Computers to Elicit Self-disclosure from Consumers. *Journal of Consumer Research*, 26 (4), pp. 324-340, 2000.
- [Nass and Moon, 2000] C. Nass and Y. Moon. Machines and mindlessness: Social responses to computers. *Journal of Social Issues*, 56 (1), pp. 81-103, 2000.
- [Ogawa et al., 1990] S. Ogawa, T. M. Lee, A. R. Kay, and D. W. Tank. Brain magnetic resonance imaging with contrast dependent on blood oxygenation. In *Proceedings of the National Academy of Sciences of the United States of America*, 87 (24), pp. 9868-9872, 1990.
- [Rickenberg and Reeves, 2000] R. Rickenberg and B. Reeves. The effects of animated characters on anxiety, task performance, and evaluations of user interfaces. *Letters of CHI 2000*, pp. 49-56, April 2000.
- [Ruttkay and Pelachaud, 2004] Z. Ruttkay and C. Pelachaud (Eds.). *From Brows to Trust: Evaluating Embodied Conversational Agents* (Human Computer Interaction Series, Vol. 7). Kluwer, New York, 2004.
- [Ruttkay et al., 2002] Z. Ruttkay, C. Dormann and H. Noot. Evaluating ECAs – What and how? *Proceedings of the AAMAS02 Workshop on 'Embodied conversational agents - let's specify and evaluate them!'*, Bologna, Italy, July 2002.
- [Sproull et al., 1996] L. Sproull, M. Subramani, S. Kiesler, J. H. Walker and K. Waters. When the interface is a face. *Human Computer Interaction*, 11 (2), pp. 97-124, 1996.
- [Ten Have, 2007] P. Ten Have. *Doing Conversation Analysis* (2nd ed.). London: Sage, 2007.
- [Van Eck and Adcock, 2003] R. Van Eck and A. Adcock. Reliability and factor structure of the Attitude Towards Agent Scale (ATAS). *Paper presented at the AERA*, Chicago, IL, 2003.
- [Von der Pütten et al., submitted] A. M. Von der Pütten, N. C. Krämer, J. Gratch and S. H. Kang. "It doesn't matter what you are!" Explaining social effects of agents and avatars. *Manuscript submitted for publication*.
- [Wada and Shibata, 2006] K. Wada, T. Shibata. Robot Therapy in a Care House - Its Sociopsychological and Physiological Effects on the Residents. *Proceedings of the 2006 IEEE International Conference on Robotics and Automation*, 2006.
- [Watson et al., 1998] D. Watson, L. A. Clark and A. Tellegen. Development and validation of brief measures of positive and negative affect: The PANAS scales. *Journal of Personality and Social Psychology*, 54, 1063-1070, 1998.