

## Virtual encounters. Creating social presence in net-based collaborations

Sabine Rüggenberg, Gary Bente, Nicole C. Krämer

University of Cologne

{sabine.rueggenberg@uni-koeln.de, bente@uni-koeln.de, nicole.kraemer@uni-koeln.de }

### Abstract

*The paper discusses the experience of "social presence" as a relevant effect dimension of avatar-mediated net-communication. Special attention is paid to measurement issues combining subjective verbal reports of social presence with objective behavioral data relating nonverbal activity, visual attention and person perception. Data will be reported from a study comparing the effects of different real time communication modes (text, audio, video and avatar) in a shared collaborative workspace. Results point to a significant difference between text and all other communication modes, indicating that audio, video and avatar systems worked similarly well in creating an experience of social presence. Analyses of behavioral data yield similar levels of visual attention for both video and avatar conferencing modes, which is however decreasing over time. The data raises critical questions about the added value of avatar systems and the specific requirements those systems have to meet to prove superior to mere audio/video transmissions.*

### 1. Introduction

Technologies for computer mediated communication (CMC) are advancing rapidly, overcoming early restrictions of text-based communication via the internet. Shared workspaces and collaborative virtual environments (CVEs) allow for real time information interchange and the synchronization of distributed working efforts over distance. Especially in the field of net-based collaboration and cooperative learning these developments were mainly driven by one goal: To improve work effectiveness by creating shared knowledge and coordinating problem solving activities. The potential limitations of this primarily task-oriented perspective on mediated collaborations have only recently been noticed. In this line Redfern and Naughton [1] state: "By focusing solely on work effectiveness, we risk missing out on social richness – this has indeed been a problem with technologies such as video conferencing, which typically provide spaces for interaction but not social places as meaningful platforms for communication" (p. 207).

Aiming at a more personalized and emotional communication via the net Redfern and Naughton [1] (p. 206) refer to empirical evidence that the use of avatars can play an important role: „CVEs can provide richness of expression and personality, as well as "identity

persistence" via appropriately detailed and customizable avatars. By fostering users' interest in one another's characters we will support the development of sociability and community". Avatars allow to overcome constraints of mere text-based or audio communication by including nonverbal communication channels. In contrast to video conferencing systems avatar platforms provide additional communication bandwidth without losing specific degrees of freedom which we much appreciate in CMC, i.e. avatars can convey nonverbal cues without necessarily disclosing the person's identity or triggering prejudices based on physical appearance (e.g., gender, culture, age, attractiveness). At the same time avatars as embodied representations allow people to allocate themselves in a shared virtual space [2] and simultaneously handle the shared virtual objects and thus are expected to create an experience of co-presence [3].

Based on such observations the concept of "social presence" has emerged as a central variable in evaluating possible socio-emotional effects of virtual encounters. Biocca, Harms and Burgoon [4] comment: „The assessment of satisfaction with entertainment systems and with its productive performance in teleconferencing and CVEs is based largely on the quality of the social presence they afford“. Social presence is here broadly defined as a "sense of being together", based on „mediated representations of humans via text, images, video, 3D avatars and in artificial representations of humanoid or animal-like intelligence including virtual humans, agents, computers, and robots" (p. 3). Based on previous research we conceptualize social presence as a basic experience of spatial co-location, emotional closeness and social relatedness [5, 6] on which more specific interpersonal effects which are considered relevant for successful net-based collaboration, such as "interpersonal trust" [7] can built up.

The dimensional structure of the complex psychological variable as well as its relation to the other concepts, and its impact on net-based collaborations however has not been explored in detail so far. Also, we lack systematic data on the influence of avatar-mediated nonverbal signals on social presence as compared to other means of computer-based real-time communication. The current study aims to provide empirical clarification regarding the dimensionality of social presence, its interrelation to other relevant social psychological variables in CMC, and the influence of different communication modalities (text, audio, video and avatar conferencing) on social presence and collaborative behavior. In contrast to the existing

approaches our study used behavioral process measures (nonverbal activity and visual attention) to complement subjective verbal assessments of social presence.

**2. Method**

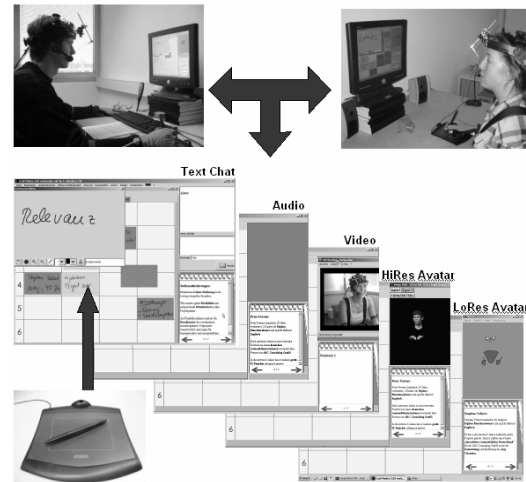
142 participants (68w/ 74m) collaborated in a management assessment centre task, making a decision with respect to selecting the right applicant for a predefined job. The participants interacted in same-sex dyads using a low immersive open desktop communication system consisting of the shared workspace “Cool Modes” [8] and a real time communication window. “Cool Modes” interactions were performed by means of a graphic tablet allowing to make notes, place, edit or remove shared information needed to solve the collaborative task. Cyber Gloves, Polhemus motion trackers and a high resolution eye-tracking system were used to collect nonverbal data: (head and upper body movement, hand and finger movements eye movements and gaze direction). Nonverbal data was stored for all subjects for later analysis. In the avatar conditions the nonverbal data was used to animate the virtual representatives of the communication partners in the communication window in real time. Two types of avatars were available: a cartoon-like low fidelity avatar (LFA), which was rendered in our own rendering software and an anthropomorphic high fidelity avatar (HFA), which was rendered by a commercial 3D animation tool (Kaydara Filmbox©). The participants were randomly assigned to one of five possible communication settings: (1) text only, (2) audio only, (3) audio + video, (4) audio + low fidelity avatar (LFA), and (5) audio + high fidelity avatar (HFA). In the audio mode the communication window was empty, in the text mode it served as a chat window. In the video and in both avatar modes it was used to display the nonverbal behaviour of the vis-à-vis. Figure 1 shows the experimental setting and screen shots of the different interface modalities.

Social presence was measured by means of 58 five-point Likert scale items based on the dimensions introduced by Biocca et al. [4], Kumar and Benbasat [9], and Nowak [10] and Tu [11]. Further a 20 item questionnaire was created to measure interpersonal trust as a two dimensional construct as described by Kanawattanachi and Yoo [7] and Nowak [10]. At last a set of 25 bipolar items was used for the measuring of mutual person perception and a set of 21 items (5 point Likert scale) for the aspect of perceived communication effectiveness. All items sets have already been evaluated and proved their internal consistencies during previous mediated communication studies [5, 12]

**3. Results**

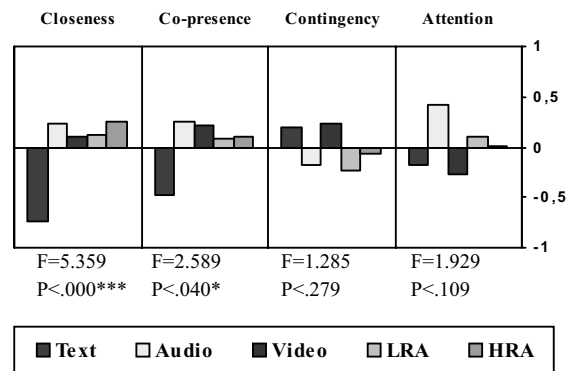
**3.1. Results of the principal component analysis and internal consistency tests of the questionnaires**

Principal component analysis (Varimax rotation) and internal consistency tests were conducted across the four item sets. For the aspect of social presence the analysis yielded a four factor solution explaining 52.14% of the total variance (*closeness, co-presence, contingency, attention*). As expected for interpersonal trust a two factor solution was found, explaining 50,43% of the variance (*cognition based trust, affect based trust*).



**Figure 1 Technical setup and experimental conditions for the media comparison study.**

Three components resulted from principal component analysis for the person perception items, explaining 46,77% of the total variance (*immediacy, assertiveness, competence*). At least the principal component analysis of the perceived communication effectiveness items resulted in a four component solution explaining 55.9% of the variance (*satisfaction, clarity, impression management, relevance*). The Cronbach’s alpha values for all resulting scales were good to excellent. Table 1 shows the factors and the consistency measures.



**Figure 2 Technical setup and experimental conditions for the media comparison study**

**Table 1: ANOVA comparison of five communication modalities (mean factor score)**

	Text	Audio	Video	LRA	HRA	F	P	(Scheffé Test)	Cronbach's Alpha
<b>Social Presence</b>									
<i>Closeness</i>	-.730	.233	.103	.126	.248	5.359	.000	1>2 (p = .015) 1>3 (p = .027) 1>4 (p = .030) 1>5 (p = .003)	.907
<i>Co-presence</i>	-.481	.259	.217	.087	.098	2.589	.040		.804
<i>Contingency</i>	.200	.180	.232	-.240	-.066	1.285	.279		.734
<i>Attention</i>	-.169	.426	-.277	.102	.018	1.929	.109		.653
<b>Interpersonal trust</b>									
<i>Cognition based trust</i>	-.207	.119	.049	.028	.031	.398	.809		.878
<i>Affect based trust</i>	-.626	.408	.018	.202	.088	4.331	.003	1>2 (p = .009) 1>4 (p = .043)	.831
<b>Person perception</b>									
<i>Immediacy</i>	-.830	.433	.048	.126	.248	7.511	.000	1>2 (p = .000) 1>3 (p = .015) 1>4 (p = .010) 1>5 (p = .001)	.904
<i>Assertiveness</i>	-.223	-.003	-.007	.311	-.033	.922	.453		.822
<i>Competence</i>	.014	.339	.087	-.170	-.025	.835	.505		.664
<b>Perceived communication effectiveness</b>									
<i>Satisfaction</i>	-.928	.031	.219	.267	.330	9.818	.000	1>2 (p = .009) 1>3 (p = .000) 1>4 (p = .000) 1>5 (p = .000)	.875
<i>Clarity</i>	-.150	.434	-.205	.076	-.033	1.585	.182		.814
<i>Impression Management</i>	.053	.126	-.095	.062	-.083	.257	.905		.814
<i>Relevance</i>	-.062	-.529	.145	.130	.156	2.120	.082		.659

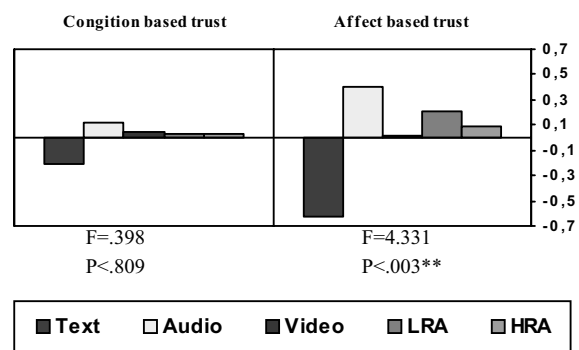
**3.2 Media differences in social presence, interpersonal trust, person perception, and perceived communication effectiveness**

ANOVAs with ex-post Scheffé tests were conducted for all factors to determine the differential influence of the five communication modalities on the social presence, interpersonal trust, person perception, and perceived communication effectiveness aspects.

For the social presence scales significant differences were only found for the factors “closeness” and “co-presence” (see figure 2). With respect to “closeness” the text condition proved to be significantly different from all the other modalities, indicating that the provision of an analogous real-time channel alone – be it audio, video or avatars - was sufficient to increase the experience of emotional closeness, immediacy, and mutual understanding.

The significant difference in the “co-presence” factor revealed by the ANOVA was not reflected in the pairwise post-hoc comparisons. However, the text

condition also proved to be the one that scored lowest on co-presence (see table 1).



**Figure 3 Media differences in the trust factors**

For the aspect “cognition based trust” no significant results could be found. For “affect based trust” the results attained significance, indicating that - in contrast to all other modes - the text did not produce positive

levels of trust. Only the difference between text chat and both audio and high resolution avatar reached significance (see figure 3 and table 1).

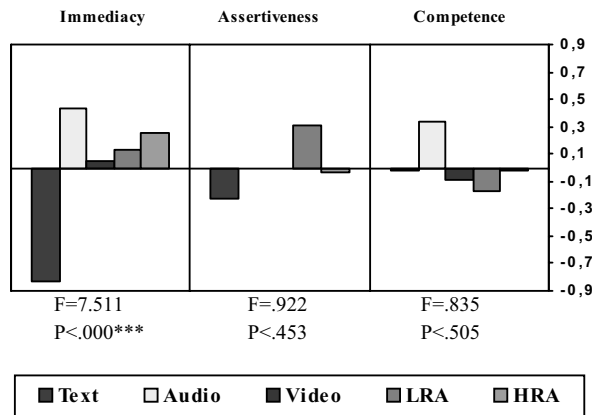


Figure 4 Media differences in person perception

The different communication groups varied significantly with regard to the aspect of “immediacy/social evaluation”. However, direct comparisons show that this is only due to the text condition, as this is the only condition that differs from all other settings (see figure 4 and table 1).

The different communication groups varied only with regard to the aspect of “perceived interaction effectiveness”. Direct comparisons show that again only text was significantly different from the other four communication technology settings (see figure 5 and table 1).

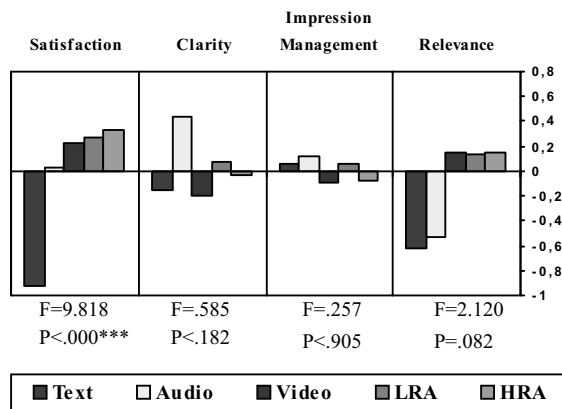


Figure 5 Media differences in the communication effectiveness factors

### 3.3 Media differences in nonverbal activity and visual attention

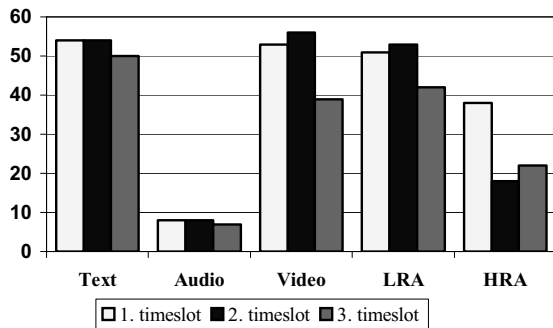
As behavioural indicators of social presence we analysed the parameters head orientation (exposing the own face), gaze direction (being attentive to the vis-à-vis’s appearance and nonverbal activity) and nonverbal effort (defined as overall movement complexity). It was assumed that the relevance of nonverbal information as

provided by the audio-video conditions (video, avatar) would lead to higher levels of exposure (head orientation towards the communication window, higher levels of visual attention to the communication window (directed gaze) and higher levels of nonverbal activity (movement complexity, time spent moving)

To separate the communicative use of the visual channels from orientation responses towards moving stimuli and curiosity effects (innovation effects) comparisons of the media were done for three consecutive time segments at the beginning of the interactions. A decline in visual attention and nonverbal activities could thus indicate a loss of interest over time which would mean that the nonverbal signals were only relevant for first impression checks, or due to orientation responses, which habituated over time or just to the newness of the medium. As illustrated by figure 8 the number of directed gazes towards the communication window was highest in the text mode (reading activity) and lowest for the audio mode (window had no display function). This data was in line with our expectations. For video as well as for the LFA condition the communication window reached nearly the same level of visual attention as in the text mode. HFAs however attracted only about 50% of the attention of the other AV modes. ANOVA result for the whole observation time was highly significant (F=11.317, p<.000). Post-hoc tests showed that the audio mode was significantly different from chat as well as from video and LFA, which reached the highest levels of visual attention during the first 6 minutes of interaction. There was however a significant drop of visual attention over time, indicating a loss of interest after the first inspection. Visual attention dropped by 10% in the video and low resolution avatar condition in the third 3-minute sequence. The decrease in visual attention in the three audio-visual modalities indicates that nonverbal behaviour was not primarily used for interaction fine-tuning, but for first impression formation. In the HFA condition the drop of visual attention was stronger (20%) and already occurred in the second sequence. It has to be mentioned that the average time the gaze was directed towards the communication windows in the first three-minute sequence was highest for the HFA, indicating an increased level of curiosity and a more persistent visual inspection in the beginning. After this, however, the total number of gazes as well as average duration of directed gaze dropped significantly: The data suggests that there might have been higher expectations with respect to the social information provided here than could be met by the avatar. One possible reason for this could be that the lack of facial expressions could be attributed to the technology in case of the low resolution avatar but not in case of the HFA.

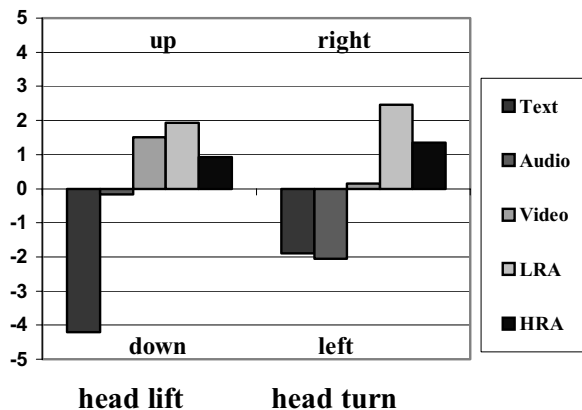
The statistical analysis of head orientation indicated that the heads of the interlocutors were more upright (F = 13.697, p <.000) and right oriented (F = 6.475, p <.000) - i.e. in direction of the communication window: upper right corner - in the audio-visual conditions as compared to the text or audio condition. Figure 7 shows the deviations from the mean position for all conditions,

indicating that the face was more exposed to the partner in the audio-visual conditions as compared to text or audio.



**Figure 6: Number of gazes towards the communication window in three subsequent 3-minute timeslot**

Significant differences were also found in the degree of nonverbal activity as reflected in the parameter movement complexity ( $F=18.397, p < .000$ ). Post-hoc comparisons revealed significant differences between text and all other modes ( $p < .000$ ). Typing activity in the text mode determined the higher levels of activity here. A significant difference was also found for the comparison between audio and video. Although not reaching conventional levels of statistical significance there was a clear tendency for all avatar modes to induce more nonverbal activity than the audio mode. As this behavior was persistent over time, figure 8 only shows the average values for the whole observation time (9 minutes).

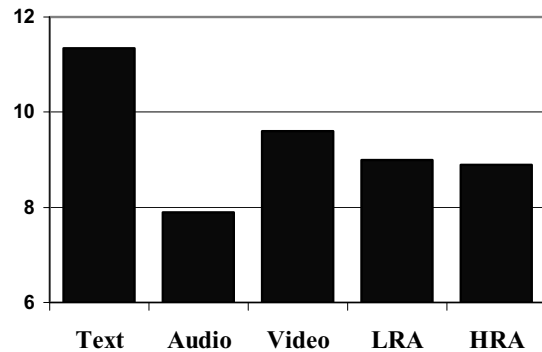


**Figure 7 Sagittal and rotational head positions as deviations from the mean (right side indicates the direction of the communication window)**

**Discussion**

The results of our study indicate that socio-emotional effects of the provided communication modalities are clearly reflected in the subjective verbal reports of the participants. General satisfaction with

interaction outcome, the feeling of being co-present and the experience of emotional closeness as a relevant dimension of social presence as well as the affective component of interpersonal trust and the evaluative component of mutual person perception (liking) all seem to benefit from the provision of real-time audio or audio-visual channels for communication. However, post-hoc tests showed that significant ANOVA results were mainly due to the differences between the text mode and all the other modes, i.e. audio, video and avatar platforms did equally well in producing these desirable interpersonal results. This is notable, given that in earlier studies the most significant differences were found when comparing all types of computer-mediated communication to f2f interactions [13]. Now within CMC we find a clear distinction between text and all other modes [14]. Also, the behavioural data point to similar patterns of nonverbal communication and visual attention in the video and the avatar conditions revealing a certain loss of interest over time. This data challenges the common assumption that social presence and the related concepts are to be conceptualized as a continuum on which media characteristics and psychological effects of media use are co-aligned and on which social presence can be quantitatively tuned.



**Figure 8 Nonverbal activity (movement complexity = number of movement dimensions involved in action) in five media conditions**

Alternatively, qualitatively distinct processes could be posited which refer to different levels of social verification [15]. The first level would be the mere construction a social situation by entering any kind of interpersonal communication or by being addressed by another social entity. Second level verification would refer to the distinction between face to face or mediated communication. This distinction is most relevant with respect to any kind of mutual bodily impact (touch, interpersonal distance). In f2f situations actions and reactions cannot be temporarily or spatially buffered or decoupled. Responses have to take into account social adequacy and possible consequences. Third level verification is concerned with distinctions within mediated encounters. The special effects of text might be due to the fact that written speech is discrete (non-continuous) and is based on an arbitrary code. There is a

natural turn taking structure induced by the syntax of the language. This is different for all the other modes where always at least one continuous and analogous channel is open simultaneously for the interlocutors, i.e. for audio, video and avatars. Similarities between these modes again could be due to the salience and predominant role of speech – as for example posited by Nass and Gong [16]. But while Nass and Gong [16] see speech generally and against an evolutionary background as the most decisive component, we hypothesize that under specific conditions - e.g. in case relational or socio-emotional information cues becomes more important – nonverbal aspects will gain additional salience and other than within the current study visual aspects of behavior will not suffer from a loss of interest over time. To test this hypothesis, subsequent experiments will imply systematic variation of the importance of socio-emotional aspects of interaction and mutual person perception.

All these modalities are still experienced as mediated, i.e. not constituting the experience of co-presence in the sense of expecting immediate bodily consequences. However, while audio and video are limited, avatar platforms offer new possibilities to overcome many of these restrictions. In the words of Foster and Meech [17] the challenge could be defined as to overcome the experience of “here and there” and to create a new shared experience of “elsewhere” (p. 212). Virtual worlds and avatars could thus be seen more as a means to contextualize social interaction and to foster the salience of nonverbal information, rather than just to provide high fidelity transmission channels for visual cues. They are in this sense not just virtual equivalents of a video conferencing system but a possibility for active filtering and contingency management systems. While current psychological research focuses mainly on the measurement of social presence, future research will have to address psychological as well as neurobiological knowledge concerning the specific demand characteristics of highly immersive virtual social realities as derived from human social information.

## References

- [1] S. Redfern, N. Naughton. Collaborative virtual environments to support communication and community in internet-based distance education. *Journal of Information Technology Education*, 1(3), 201-211. 2002.
- [2] J. Steuer. Defining virtual reality: Dimensions determining telepresence. *Journal of Communication*, 42 (4), 73-93. 1992.
- [3] E. Prasolova-Forland. Supporting awareness in education: overview and mechanism. In *Proceedings of the International Conference on Engineering Education*. August 18–21, Manchester, U.K. 2002.
- [4] F. Biocca, C. Harms, J. Burgoon. Criteria and scope conditions for a theory and measure of social presence. *PRESENCE 2001, 4<sup>th</sup> Annual International Workshop May 21-23*, Philadelphia, PA. 2001. URL: <http://nimbus.ocis.temple.edu/~mlombard/P2001/papers.html>
- [5] G. Bente, S. Rüggenberg, N. C. Krämer. Social presence and interpersonal trust in avatar-based, collaborative net-communications. *7th Annual International Workshop Presence 2004*. Valencia: Editorial Universidad Politécnic de Valencia. 2004.
- [6] F. Biocca, J. Burgoon, C. Harms, M. Stoner. Criteria and scope conditions for a theory and measure of social presence. *E. Lansing, MI: Media Interface and Network Design (M.I.N.D.) Lab*. 2001. URL: <http://nimbus.ocis.temple.edu/~mlombard/P2001/Biocca1.pdf>
- [7] P. Kanawattanachai & Y. Yoo. The dynamic nature of trust in virtual teams. In *Journal of Strategic Information Systems*, 11(3), 187-213. 2002.
- [8] Bollen, H.U. Hoppe, M. Milrad & N. Pinkwart. Collaborative Modelling in Group Learning Environments. In P. I. Davidsen, E. Mollona, V. G. Diker, R. S. Langer & J. I. Rowe (ed.), *Proceedings of the XX. International Conference of the System Dynamics Society*, Palermo (Italy). 2002.
- [9] N. Kumar, I. Benbasat. Para-social presence: a re-conceptualization of „social presence“ to capture the relationship between a web site and her visitors. In *Proceedings of the 35th Annual Hawaii International Conference on System Sciences (HICSS-35.02)*, IEEE Computer Society. URL: <http://csdl.computer.org/comp/proceedings/hicss/2002/1435/01/14350013b.pdf>. 2002.
- [10] K. Nowak. Conceptualizing and measuring the sense of access to another mind. *Paper presented at the Presence 2001 Conference, Philadelphia, PA*. 2001. URL: <http://astro.temple.edu/~lombard/P2001/Nowak1.pdf>.
- [11] C.-H. Tu. The impacts of text-based CMC on online social presence. In *The Journal of Interactive Online Learning*, 1(2), 1-24. 2002.
- [12] A. Petersen, *Interpersonale Kommunikation im Medienvergleich*. Münster: Waxmann. 2001.
- [13] A. Petersen, G. Bente, N.C. Krämer. Virtuelle Stellvertreter: Analyse avatar-vermittelter Kommunikationsprozesse. In G. Bente, N.C. Krämer, A. Petersen (eds.), *Virtuelle Realitäten*, 227-253. Göttingen: Hogrefe. 2002.
- [14] S.G. Straus, J.E. McGrath. Does the medium matter? The interaction of task type and technology on group performance and member reactions. In *Journal of Applied Psychology* 79, 1, 87-97. 1994.
- [15] J. Blascovich, J. Loomis, A. Beall, K.R. Swinth, C. L. Hoyt, J. N. Bailenson. Immersive virtual environment technology as a methodological tool for social psychology. In *Psychological Inquiry*, 13, 146-149. 2001.
- [16] N. Nass, L. Gong. (2000). Speech interfaces from an evolutionary perspective. *Communications of the ACM*, 43 (9), 36-43. 2000.
- [17] D. Foster, J. Meech. The Social implications of virtual reality. In K. Carr (ed.) *Simulated and Virtual Realities: Elements of Perception*. Taylor and Francis. USA. 1