Research article

Daniel Roth*, Marc Erich Latoschik, Kai Vogeley, Gary Bente

Hybrid Avatar-Agent Technology – A Conceptual Step Towards Mediated "Social" Virtual Reality and its Respective Challenges

DOI 10.1515/icom-2015-0030

Abstract: Driven by large industry investments, developments of Virtual Reality (VR) technologies including unobtrusive sensors, actuators and novel display devices are rapidly progressing. Realism and interactivity have been postulated as crucial aspects of immersive VR since the naissance of the concept. However, today's VR still falls short from creating real life-like experiences in many regards. This holds particularly true when introducing the "social dimension" into the virtual worlds. Apparently, creating convincing virtual selves and virtual others and conveying meaningful and appropriate social behavior still is an open challenge for future VR. This challenge implies both, technical aspects, such as the real-time capacities of the systems, but also psychological aspects, such as the dynamics of human communication. Our knowledge of VR systems is still fragmented with regard to social cognition, although the social dimension is crucial when aiming at autonomous agents with a certain social background intelligence. It can be questioned though whether a perfect copy of real life interactions is a realistic or even meaningful goal of social VR development at this stage. Taking into consideration the specific strengths and weaknesses of humans and machines, we propose a conceptual turn in social VR which focuses on what we call "hybrid avatar-agent systems". Such systems are required to generate i) avatar mediated interactions between real humans, taking advantage of their social intuitions and flexible communicative skills and ii) an artificial social intelligence (AIS) which monitors, and potentially moderates or transforms ongoing virtual interactions based on social signals, such as performing adaptive manipulations of behavior in intercultural conversations. The current article sketches a respective base architecture and discusses necessary research prospects and challenges as a starting point for future research and development.

Keywords: Computer-Mediated Communication, Human-Computer Interaction, Interpersonal Synchronization, Virtual Reality

***Corresponding author: Daniel Roth,** Media and Communication Psychology, Department of Psychology, University of Cologne, Cologne, Germany; and Human-Computer Interaction, Institute for Computer Science, University of Würzburg, Würzburg, Germany, e-mail: daniel.roth@uni-koeln.de
Marc Erich Latoschik: Human-Computer Interaction, Institute for Computer Science, University of Würzburg, Würzburg
Kai Vogeley: Cognitive Neuroscience, Forschungszentrum Jülich, Jülich, Germany; and Department of Psychiatry, University Hospital of Cologne, Cologne, Germany
Gary Bente: Media and Communication Psychology, Department of Psychology, University of Cologne, Cologne, Germany; and Department of Communication, Michigan State University, East Lansing, USA

1 Introduction

Envisioning the near future, we will soon be able to retrieve online information on real life objects with Augmented Reality (AR), Mixed Reality (MR), or VR devices and permanently connect to friends and family or their virtual representatives. Interaction metaphors will rely on sensors and actuators that are ambient and non-intrusive (Izadi et al. 2011) or wearable (e. g. Google Glass or Microsoft HoloLens). These metaphors will certainly depart from classical and artificial device-based input methods (Bowman et al. 2005) fostering more natural device-less input methods such as speech and gestures (Latoschik 2005, Latoschik and Fischbach 2014) or thoughts (Millán et al. 2010). These technologies have the potential to make the mobile phone disappear, and instead of "removing one away from the world of others" in a detached mode of observation (Katz 2014), they will be able to provide a shared MR or VR environment that is easily and instantaneously accessible for all communication partners. Although we do not know which particular technologies will be accepted by a broad range of users, it seems obvious that we will soon enter a new age of communication in which VR will become a mass consumer product (Slater 2014). This will result in an unpredictable impact that will even supervene and transcend the social transitions triggered by mobile phones or the internet. *But how can interpersonal understanding and virtual interactions benefit in a social sense from these technologies?* In our view, benefits can only be taken from new concepts and interaction metaphors that place a strong focus on the social dimension, *enhancing interpersonal communication with the help of multimodal artificial mediation, analyzing, interpreting and transforming the social dynamics of communication*. For instance, in a conversation between a Japanese and a German this could mean to translate the culture specific greetings like the bow and the hand shake into the other's behavioral space displayed via an avatar / agent, resulting in higher rapport and better understanding. While avatars are defined as digital representatives with human controlled behavior, the behavior of (digital) embodied agents is controlled by the machine and the algorithms (Bailenson and Blascovich 2004). In contrast to current VR systems, which mostly focus on either avatar or agent based approaches, this article presents a new conceptual approach for a social VR which we call "hybrid avatar / agent technology" (HAAT) that takes into account both, human and technological aspects in social interactions to actively mediate virtual communication in the social dimension.

2 Hybrid Systems – The Conceptual Turn

So far, developers have been taking different routes to approach socially enriched VR systems. On the one hand developers of shared virtual environments (SVE), such as Second Life, aim to provide communication platforms, using avatars with high physical and behavioral realism. These concepts allow virtual interactions to take full advantage of verbal and nonverbal channels available in face-to-face encounters. The main focus of these approaches is to provide plausible characters, displaying the full range of nonverbal behavior in great detail to prevent these cues from "being filtered" out (Culnan and Markus 1987) as it is assumed that they have the unique capacity of providing information on who others are, individual emotional expression and individual personalities (Walther et al. 2015). On the other hand developers, focusing on artificial intelligence, aim at the creation of autonomous computer agents, which can act as social partners or conversational interfaces in human-machine interaction (Latoschik 2014b). The main challenge here is to develop an artificial social intelligence (ASI) core which possesses the perceptual, cognitive and behavioral skills typical for humans. Both approaches bare technological and psychological challenges, which might be difficult or even impossible to meet. However, we claim that both approaches miss out on exploiting the real potential of social VR, which we see in the combination of human capacities (avatar component) and skills with those of the machine (agent component), taking into account the specific strengths and weaknesses of both components. We conceptualize such systems as "hybrid avatar-agent technologies" (HAAT). In fact, this idea is not completely novel. Many researchers use so-called Wizard-of-Oz paradigms combining human and artificial intelligence to study the effects of conversational agents (Bradley, Mival and Benyon 2009) or robots (Riek 2012). While those setups use hidden human interactors in human-machine interactions, the HAAT approach inverts this rationale, making use of artificial intelligence to actively mediate interpersonal understanding, fully transmitting and enhancing the human physical and social dimensions to computer-mediated communication (CMC). Figure 1 illustrates the basic technological concept of the mediation process. The ASI controlled sync engine acts as artificial mediator analyzing and interpreting behavioral patterns visible or hidden to the human user. The sync engine then decides whether behavior needs to be mediated in order to increase social understanding. While autonomous agent technologies would require both, a more or less complete "social intelligence" as well as the technical real-time capabilities in sensing, processing and producing social behavior, our concept allows to keep these aspects separate: (i) *capture and synthesis components* that record communicative behavior perceivable by humans, (ii) *a sync engine* that records behavioral, audiovisual, physiological as well as underlying neural signals, identifying and inducing means of manipulation and communicative mediation.

3 Signals and Dimensions of VR Mediation

The HAAT approach draws upon the fact that encounters in SVEs are different from real life interactions in significant ways. They are characterized by unprecedented

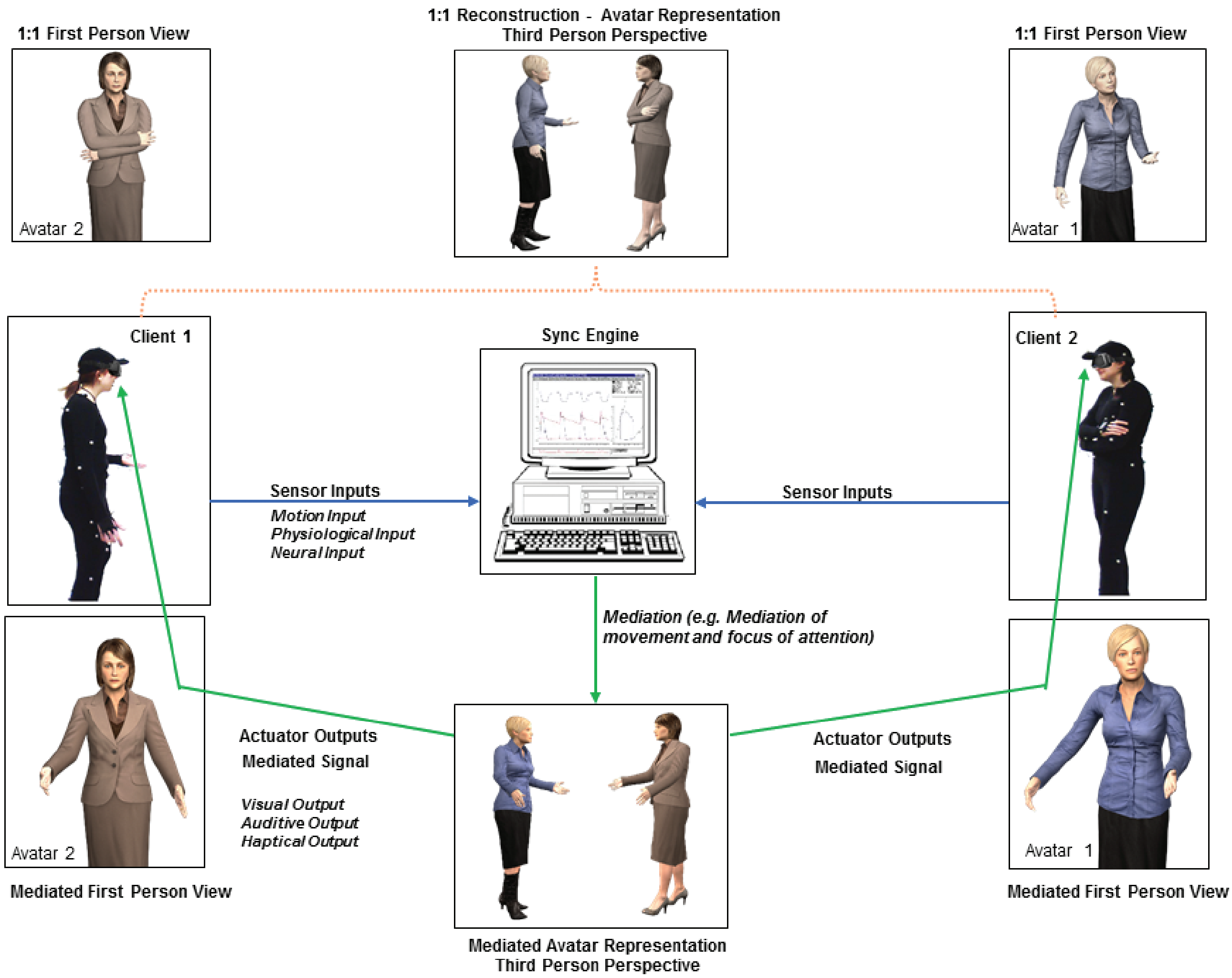

Figure 1: The sync engine describes the intelligent core of the system referring to the results from integrated data analysis, deciding about intervention strategies and targeted communication channels. Interventions could reach from mere push notifications over hints for behavior change to direct transformations.

degrees of freedom in the physical appearance, as well as the dynamic manipulation of their behavior (Bente, Krämer and Eschenburg 2008). Furthermore, sensing and measurement technologies can seamlessly be added to provide access to information potentially relevant to but not observable or not consciously perceived by humans. This enables the ASI component to make use of signals, such as neural activity measured through portable EEG devices (Lievesley, Wozencroft and Ewins 2011), physiological arousal or subtle behavioral dynamics which are not or only unconsciously processable by human interlocutors (Betella et al. 2014, Wagner et al. 2013). These possibilities comprise the enormous potential for VR mediation to enhance or augment interpersonal communication and to provide experiences which are impossible in real-life interactions (Bailenson 2004, Yee and Bailenson 2006). The ASI component thus serves as a supervisor or moderator in human-human interactions taking a third person perspective rather than a first-person view. Real-time technology allows to capture, analyze, transmit and render human behavior, by monitoring crucial instances in subtle verbal and nonverbal variations making them available to the human interactors by explicit feedback or direct modification. In contrast to real-life mediation HAAT is able to actively transform the behavioral input to a mediated output in addition to or instead of just reporting / displaying the original signals. Although computational approaches to interpret social signals exist (e. g. Wagner et al. 2013) there are important factors to be considered on the interpersonal level.

4 Psychological Backbones of VR Mediation

To understand the basic rules of conversational mediation it is important to note that behavioral cues and phenomena in social understanding are time-critical. This does not imply that social understanding processes require simultaneity. It rather presupposes that behavioral contingencies occur in certain time frames and that patterns of social behaviors and understanding are in essence only observable if we look at the dyad. In other words the smallest unit of observation to identify behaviors to manipulate using the ASI is the dyad and not the experience or behavior of one individual. A simple example for a possible manipulation is eye contact. It could be shown that the latencies for reciprocating the direct gaze of a partner are critical for the social meaning of eye contact as are the latencies for interrupting eye contact again (Bente, Donaghy and Suwelack 1998, Bente, Eschenburg and Krämer 2007, Kuzmanovic et al. 2009, Pfeiffer et al. 2011, 2012, 2014). If we reciprocate too late or we look too long this might be taken as disinterest or seen as inappropriate staring. However, eye contact cannot be established without both partners showing partner directed gaze. Effects of facial or motor mimicry are time sensitive as well. For instance, mirroring the body posture of the partner is only perceived as a relevant social signal if there is a certain time delay. Otherwise it is perceived as an arbitrary coincidence (Krämer 2010). Such contingency patterns of social understanding and dyadic rapport can even occur with complete phase shifts. It could be shown that partners in conflict often show overlaps in increased motor activity while good rapport is visible in nonverbal turn taking patterns just as it is the case for verbal utterances in conversations (Frey 1984). In fact there are innumerous examples of socially meaningful patterns only observable on the dyadic level, whose ultimate meaning is coded in temporal contingencies. These aspects of interpersonal synchronization represent a rising field and the biggest challenge for social psychology and social neuroscience (Chanel, Kivikangas and Ravaja 2012, Dumas et al. 2010, Lakens and Stel 2011, Lakin and Chartrand 2003, Sebanz, Bekkering and Knoblich 2006, Spapé et al. 2013).

The scope of possible intervention scenarios for HAAT is broad and amongst others includes the mediation of intercultural encounters in SVEs. Imagine a negotiation between a German and a Japanese business team. There is evidence on differences in the production and perception of nonverbal signals in Western and Eastern cultures (Matsumoto 2006). Endrass, André, Rehm, and Nakano (Endrass et al. 2013) could also show differences in cultural preference of nonverbal behavior (Japanese or German) mapped to virtual characters. Andersen (2015) commented on the implications for cross-cultural communication: “Because we are usually not aware of our own nonverbal behavior, it becomes extremely difficult to identify and master the nonverbal behavior of another culture” (p. 258). HAAT could help to monitor and adjust critical nonverbal cues in virtual encounters. For instance, nonverbal expressiveness (quantity and intensity of bodily and facial expressions; Riggio 2006) or level of immediacy (interpersonal distance and eye contact; Sanders and Wiseman 1990) or dominance gestures (expansivity of gestures, directed gaze, raised head; Bente et al. 2010) can be analyzed identifying possible sources of misunderstanding and conflict and modified to achieve better rapport. Limitations to these manipulations of course imply physical borders and appropriateness as well as the level of intervention with respect to interindividual differences and “accents” in nonverbal signals. Finding these levels of intervention represents a challenging task for interdisciplinary approaches. The transfer of the conceptual idea to the techno-psychological implementation unveils yet unsolved issues.

5 Techno-Psychological Challenges

The implementation of the HAAT interaction metaphor challenges psychologists, engineers and computer scientists. First, the AIS and the sync module have to know about relevant behavioral and neurophysiological signatures and interpersonal phenomena of successful communication within the dyad. This “supervising knowledge” includes access to relevant data from multiple sources, including AI-based algorithms specialized to derive knowledge and insight about the ongoing interaction (Latoschik 2014b, Latoschik and Blach 2008). Second, the platform has to provide real-time capacities to meet the perceptual task-deadlines in human communication. Real-time capacity is crucial because two time critical aspects depend on the real-time responsivity: (i) *the intrapersonal aspect* of action-perception coupling resulting in perceived body ownership and agency experience and (ii) *the interpersonal aspect* of social contingencies and interplay of different signal types as well as close coupling of action and reaction time within temporal synchronization phenomena that induce co-presence, social presence (see Lee 2004) and synchronization phenomena. The interpersonal aspect of social contingencies requires multimodal signal processing to be highly responsive, allowing multiple simultaneous dependent and independent processing loops to enable coherent manipulation (Latoschik and Fischbach 2014). For example, lower level (apparent) motion is perceived rapidly by the human observer (Baker and Braddick 1985, Thornton 1998) and thus could be manipulated during a refresh cycle. However, information carried through higher-level social signals that last longer (e. g. gestures, facial expression) involve social information processing and evoke time delayed receiver responses and adaptions, e. g. hundreds of milliseconds in the case of joint attention (Pfeiffer et al. 2012). Furthermore, longer lasting social experiences or reactions like empathy may even last over days (Vinciarelli, Pantic and Bourlard 2009), and are thus impossible to be interpreted and mediated in a single refresh cycle. As these signaling systems have different frequencies HAAT is required to be an *interactive real-time system, combining multiple modalities with different frequencies that are partly subject to enhanced real-time timing conditions due to the high reactivity requirements* (Latoschik 2014a).

6 The Requirement for Timeliness

For the realization of a multimodal HAAT concept, one of the major factors lacking in current systems is a reliable timeliness. *With timeliness we denote all relevant temporal aspects of the algorithms and processes involved*, e. g., real-time constraints, latency, or jitter. Task-deadlines for interactive systems are ultimately defined by necessary display frequencies (e. g., 60hz or higher for graphics, 1000hz for haptics; Hale and Stanney 2014) which consequently define a complete refresh cycle of a mediating VR system including signal manipulations. In the early days of video games, a field strongly related to VR, AR, and MR, developers were still enabled to fit all of the *functional aspects*, the necessary algorithms and computation, into the video refresh cycles. For one part, the functional aspects were very limited compared to today's computer games or VR systems. For the other part, developers had full control over the underlying hardware and operating systems (if there were any in the first case), and thus had a much lower level approach than current best practices. Today, the wide-spread usage of commodity (PC) hardware in combination with multi-user multi-tasking operating systems with no preemptive capabilities renders control over the timeliness an impossibility for developers. The best one can achieve under these circumstances is monitoring and black-box optimizing timeliness (Rehfeld, Tramberend and Latoschik 2014). In addition, the usage of game engines on top of these systems on an application layer can produce soft real-timeliness at best which is probably why end-to-end latency is rarely reported.

Physiological and psychological properties of the users also determine *certain non-functional requirements, specifically concerning the temporal aspects* of the closely-coupled human-computer-interaction loops typical for VR, AR, and MR. For example, visuo-motor contingency is a crucial *intrapersonal bottom-up factor* in several aspects, ranging from the acceptance of avatar representations (Lugrin, Latt and Latoschik 2015) to the generation of reliable perspectives and view corrections (e. g., head tracking) and the avoidance of simulator sickness. However, today's chipset designs and operating systems mostly gave up the challenge to achieve a real-time operating system for commercial use, although concepts and designs arise from other perspectives, e. g. the mobile platform Android (Yan et al. 2013). While current VR approaches can respectively be considered as *soft real-time* systems (not meeting the task deadline still has some utility for the system) it is essential for a HAAT alike concept to reach a *firm* or even better *hard real-time* state (task results must meet the deadline, see: Buttazzo 2011). Continuous developments thus must achieve real-time conditions for VR that result in refresh rates sufficiently *faster* than the respective sensory system (e. g., the visual system) of the user. Focus on technological developments in progress or used in other industries (e. g. embedded systems in the automation industry or robotics) may help to identify architectural concepts for *functional aspects*. For example, a look at time critical applications such as the Air Bag may help implementing functional aspects that enable HAAT alike approaches, leaving aside soft real-time embodied agent technologies. To assure real-time transmission and computation, the introduction of Quality of Service parameters to future VR systems and programming languages seems necessary in order to evaluate and ensure communication with the environment, providing access to a notion of time, the support of redundancy and diversity as well as the facilitation of executable code with deterministic predictable behavior (see Gumzej and Halang 2010, p. 3). Overall, the conceptual shift of HAAT would also imply a conceptual shift of the AR / VR / MR architectural approach and the respecting layer specifications.

7 Future Perspectives

In the present article we presented a concept for a social virtuality that builds on a hybrid approach for intelligent CMC in SVEs. The biggest psychological challenge for HAAT is the identification of interpersonal phenomena to build a taxonomy for interpretation. Technologically, we are still not in full control over the necessary timeliness for the realization and suggest to rethink methods to implement functional and non-functional requirements. Without doubt, one possible next generation of social communication involves HAAT alike technologies for shared VR, AR and MR environments. Projecting current developments in VR on the Gartner Hype Cycle it seems appropriate that VR is at the "through of disillusionment" (Rivera and van der Meulen 2014) while social VR may just have received the "trigger". Picking up the idea of Buxton's long nose of innovation (Buxton 2008), rethinking and refining approaches and concepts might just be as important as billion dollar investments of stakeholders. This includes critical thoughts on ethical, societal and security manners. HAAT is based on an artificial mediation of social behavior, thus bases on manipulating conversations and interpersonal understanding, and respectively manipulating the individual. Further exploration of the concept must include ethical debates on normativity of "successful" and "adequate" encounters as well as the justification of intervention. The latter being important concerning i) moral justification and ii) security measurements, as online real-time applications imply high security standards. For researchers and technologists the future of social VR is a challenge for human understanding and creativity, research and development that will eventually lead to another reality that is supporting interpersonal adjustments by amplifying, augmenting or enhancing reality rather than just trying to reconstruct it, a reality that is mediating understanding and interpersonal communication rather than just displaying it. It may lead to an *Ultra Reality*.

References

Andersen, P. A. 2015. In different dimensions: nonverbal communication and culture. In: (L. A. Samovar and R. E. Porter, eds) *Intercultural communication: a reader* (Fourteenth Edition and Fortieth Anniversary Edition. Cengage Learning, Boston, MA, pp. 229–241.

Bailenson, J. N., A. C. Beall, J. Loomis, J. Blascovich and M. Turk. 2004. Transformed social interaction: decoupling representation from behavior and form in collaborative virtual environments. *Presence-Teleop. Virt.* 13(4): 428–441.

Bailenson, J. and J. Blascovich. 2004. Avatars. In: *Encyclopedia of Human-Computer Interaction*, Berkshire Publishing Group, 64–68.

Baker, C. L. and O. J. Braddick. 1985. Eccentricity-dependent scaling of the limits for short-range apparent motion perception. *Vis. Res. 25*(6): 803–812.

Bente, G., W. C. Donaghy, and D. Suwelack. 1998. Sex differences in body movement and visual attention: an integrated analysis of movement and gaze in mixed-sex dyads. *J. Nonverbal Behav.* 22(1): 31–58.

Bente, G., F. Eschenburg and N. C. Krämer. 2007. Virtual gaze. A pilot study on the effects of computer simulated gaze in avatar-based conversations. In: (R. Shumaker, ed) *Virtual reality, HCII 2007, LNCS 4563,* Springer, 185–194.

Bente, G., N. C. Krämer and F. Eschenburg. 2008. Is there anybody out there? In: (E. A. Konijn, S. Utz, M. Tanis and S. B. Barnes, eds) *Mediated interpersonal communication*, 131–157.

Bente, G., H. Leuschner, A. A. Issa and J. J. Blascovich. 2010. The others:universals and cultural specificities in the perception of status and dominance from nonverbal behavior. *Conscious. Cogn.* 19(3): 762–777.

Betella, A., D. Pacheco, R. Zucca, X. D. Arsiwalla, P. Omedas, Lanatà, A. et al. 2014. Interpreting psychophysiological states using unobtrusive wearable sensors in virtual reality. *ACHI 2014, The Seventh International Conference on Advances in Computer-Human Interactions,* 331–336.

Bowman, D. A., E. Kruijff, J. Joseph, J. LaViola and I. Poupyrev. 2005. *3D user interfaces – Theory and practice*. Addison-Wesley.

Bradley, J., O. Mival and D. Benyon. 2009. Wizard of Oz experiments for companions. *Proceedings of the 23rd British HCI Group Annual Conference on People and Computers: Celebrating People and Technology*, British Computer Society, 313–317.

Buttazzo, G. C. 2011. *Hard real-time computing systems*. Springer, Boston, MA.

Buxton, B. 2008. The long nose of innovation. *Insight*, *11*: 27.

Chanel, G., Kivikangas, J. M. and N. Ravaja. 2012. Physiological compliance for social gaming analysis: Cooperative versus competitive play. *Interact. Comput.* 24(4): 306–316.

Culnan, M. J. and M. L. Markus. 1987. Information technologies.

Dumas, G., J. Nadel, R. Soussignan, J. Martinerie and L. Garnero. 2010. Inter-brain synchronization during social interaction. *PLoS ONE*, *5*(8): e12166.

Endrass, B., E. André, M. Rehm and Y. Nakano. 2013. Investigating culture-related aspects of behavior for virtual characters. *Auton. Agent Multi Agent Syst.* 27(2): 277–304.

Frey, S. 1984. *Die nonverbale Kommunikation (The nonverbal communication)*. SEL-Foundation for technical and economical communication science in the society of funding for german science.

Gumzej, R. and W. A. Halang. 2010. *Real-time systems' quality of service*. Springer, London.

Hale, K. S. and K. M. Stanney. 2014. *Handbook of virtual environments: Design, implementation, and applications*. CRC Press.

Izadi, S., D. Kim, O. Hilliges, D. Molyneaux, R. Newcombe, P. Kohli, et al. 2011. KinectFusion: real-time 3D reconstruction and interaction using a moving depth camera. In: *Proceedings of the 24th Annual ACM Symposium on User Interface Software and Technology*, ACM, pp. 559–568.

Katz, J. E. 2014. *Living inside mobile social information*. Greyden Press, Dayton, OH.

Krämer, K. 2010. *Real or fake – how does motor mimicry influence the perception of human interactions?* (Diploma thesis). University of Cologne, Cologne.

Kuzmanovic, B., A. L. Georgescu, S. B. Eickhoff, N. J. Shah, G. Bente, G. R. Fink and K. Vogeley. 2009. Duration matters: dissociating neural correlates of detection and evaluation of social gaze. *Neuroimage*, 46(4), 1154–1163.

Lakens, D. and M. Stel. 2011. If they move in sync, they must feel in sync: movement synchrony leads to attributions of rapport and entitativity. *Soc. Cognition*, 29(1): 1–14.

Lakin, J. L. and T. L. Chartrand. 2003. Using nonconscious behavioral mimicry to create affiliation and rapport. *Psychol. Sci.* 14(4): 334–339.

Latoschik, M. E. 2005. A User Interface Framework for Multimodal VR Interactions. In *Proceedings of the IEEE seventh International Conference on Multimodal Interfaces, ICMI 2005*, 76–83. Trento, Italy.

Latoschik, M. E. 2014a. *Lecture notes – real-time interactive systems – Julius Maximilian University Würzburg*. Würzburg.

Latoschik, M. E. 2014b. Smart Graphics / Intelligent Graphics. *Informatik Spektrum*, Vol. 37 (1), pp. 36–41. Springer.

Latoschik, M. E. and Blach, R. 2008. Semantic modelling for virtual worlds – a novel paradigm for realtime interactive systems? In: *Proceedings of the ACM VRST 2008*, 17–20.

Latoschik, M. E. and M. Fischbach. 2014. Engineering variance: software techniques for scalable, customizable, and reusable multimodal processing. M. Kurosu (Hrsg.), *Human Computer Interaction, Part I, HCI 2014* (Vol. 8510, 308–319). Springer International Publishing.

Lee, K. M. 2004. Presence, explicated. *Commun.Theory*. 14(1): 27–50.

Lievesley, R., M. Wozencroft and D. Ewins. 2011. The Emotiv EPOC neuroheadset: an inexpensive method of controlling assistive technologies using facial expressions and thoughts? *J. Assist. Technol.*5(2): 67–82.

Lugrin, J.-L., J. Latt and M. E. Latoschik. 2015. Avatar anthropomorphism and illusion of body ownership in VR. *Proceedings of the IEEE VR 2015*.

Matsumoto, D. 2006. Culture and nonverbal behavior. *Handbook of nonverbal communication*, 219–235.

Millán, J., R. Rupp, G. R. Müller-Putz, R. Murray-Smith, C. Giugliemma, M. Tangermann, C. Vidaurre, F. Cicotti, A. Kübler, R. Leeb, C. Neuper, K. R. Müller and D. Mattia. 2010. Combining brain-computer interfaces and assistive technologies: state-of-the-art and challenges. *Front. Neurosci*. 4, 161.

Pfeiffer, U. J., L. Schilbach, M. Jording, B. Timmermans, G. Bente and K. Vogeley. 2012. Eyes on the mind: investigating the influence of gaze dynamics on the perception of others in real-time social interaction. *Front. Psychol. 3*.

Pfeiffer, U. J., L. Schilbach, B. Timmermans, B. Kuzmanovic, A. L. Georgescu, G. Bente and K. Vogeley. 2014. Why we interact: on the functional role of the striatum in the subjective experience of social interaction. *NeuroImage*, 101, 124–137.

Pfeiffer, U. J., B. Timmermans, G. Bente, K. Vogeley and L. Schilbach. 2011. A non-verbal turing test: differentiating mind from machine in gaze-based social interaction. *PLoS ONE*, 6(11), e27591.

Rehfeld, S., H. Tramberend and M. E. Latoschik. 2014. Profiling and benchmarking event- and message-passing-based asynchronous Realtime Interactive Systems. *ACM VRST 2014 Conference Proceedings*.

Riek, L. D. 2012. Wizard of Oz studies in HRI: a systematic review and new reporting guidelines. *J. Hum.-Rob. Interact.* 1(1).

Riggio, R. E. 2006. Nonverbal skills and abilities. *The Sage handbook of nonverbal communication*, 79–96.

Rivera, J. and R. van der Meulen. 2014. Gartner's 2014 Hype Cycle for emerging technologies maps the journey to digital business. Retrieved from http://www.gartner.com/newsroom/id/2819918 (April 2015).

Sanders, J. A. and R. L. Wiseman. 1990. The effects of verbal and nonverbal teacher immediacy on perceived cognitive, affective, and behavioral learning in the multicultural classroom. *Communic. Educ.*39(4): 341–353.

Sebanz, N., H. Bekkering and Knoblich, G. (2006). Joint action: bodies and minds moving together. *Trends Cogn. Sci.* 10(2): 70–76.

Slater, M. (2014). Grand Challenges in Virtual Environments. *Front. Robot. Artif. Intel*. 1(3).

Spapé, M. M., J. M. Kivikangas, S. Järvelä, I. Kosunen, G. Jacucci and N. Ravaja. 2013. Keep your opponents close: social context affects EEG and fEMG linkage in a turn-based computer game. *PLoS ONE*, *8*(11): e78795.

Thornton, I. M. (1998). The visual perception of human locomotion. *Cogn. Neuropsychol.* 15(6–8): 535–552.

Vinciarelli, A., M. Pantic and H. Bourlard. 2009. Social signal processing: Survey of an emerging domain. *Image. Vis. Comput.* 27(12): 1743–1759.

Wagner, J., F. Lingenfelser, T. Baur, I. Damian, F. Kistler and E. André. 2013. The social signal interpretation (SSI) framework: multimodal signal processing and recognition in real-time. *Proceedings of the 21st ACM international conference on Multimedia*, ACM, 831–834.

Walther, J. B., B. Van Der Heide, A. Ramirez Jr, J. K. Burgoon and J. Peña. 2015. Interpersonal and hyperpersonal dimensions of computer-mediated communication. In: (S. S. Sundar, ed) *Handbook of the psychology of communication technology*. John Wiley & Sons, Inc.

Yan, Y., S. H. Konduri, A. Kulkarni, V. Anand, S. Y. Ko and L. Ziarek. 2013. Rtdroid: a design for real-time android. In: *Proceedings of the 11th International Workshop on Java Technologies for Real-time and Embedded Systems*, ACM, 98–107.

Yee, N. and J. N. Bailenson. 2006. Walk a mile in digital shoes: the impact of embodied perspective-taking on the reduction of negative stereotyping in immersive virtual environments. *Proceedings of PRESENCE*, 24–26.

Bionotes

Daniel Roth
University of Cologne, Germany
University of Würzburg, Germany
daniel.roth@uni-koeln.de

Daniel Roth, M. Eng. received his Master's degree in Media and Imaging Technology from the University of Applied Sciences, Cologne and is currently working as research associate at the Professorship for Media- and Communication Psychology of the University of Cologne. He started his doctoral thesis on interpersonal synchronization in Virtual Realities at the Human-Computer Interaction Group of the University of Würzburg.

Prof. Dr. Marc Erich Latoschik
University of Würzburg, Germany
marc.latoschik@uni-wuerzburg.de

Marc Erich Latoschik holds the chair for Human-Computer Interaction of the University of Würzburg. His work is interdisciplinary oriented with a strong focus on multimodal interaction and intelligent graphics interconnecting real-time 3D simulations, Virtual and Augmented Reality, Artificial Intelligence, and cognitive sciences.

Prof. Dr. Dr. Kai Vogeley
Forschungszentrum Jülich, Germany
University Hospital of Cologne, Germany
kai.vogeley@uk-koeln.de

Kai Vogeley is a Full Professor at the Department of Psychiatry of the University of Cologne and affiliated at the Institute of Neuroscience and Medicine – Cognitive Neuroscience, Research Center Jülich, where he leads the Social Cognition Group. His research focuses on social neuroscience with a special focus on nonverbal communication and interaction. Clinically, his main interest is autism spectrum disorder in adulthood.

Prof. Dr. Gary Bente
University of Cologne, Germany
Michigan State University, USA
bente@uni-koeln.de

Gary Bente is a psychologist and Full Professor at the Department of Psychology of the University of Cologne and head of the Unit for Media and Communication Psychology. He is appointed as a Visiting Professor at the Department of Communication at Michigan State University, USA. His research focuses on the analysis of movement and gaze behavior in interpersonal communication including face-to-face as well as mediated interactions.