Effects of Avatar and Background Representation Forms to Co-Presence in Mixed Reality (MR) Tele-conference Systems

Dongsik Jo*
ETRI and Korea University

Ki-Hong Kim
ETRI

Gerard Jounghyun Kim†
Korea University

Figure 1: *Comparing co-presence in tele-conferencing with a virtual human: in a virtual environment (left) and, in an augmented environment (right).*

Abstract

With continued technology innovation, the traditional 2D video based tele-conferencing and collaborative systems are evolving into ones that are immersive, 3D, more interactive and even augmented in real environment. One important quality of a tele-collaboration system is the sense of co-presence as felt by the participating users. Virtual and augmented reality based implementations and media presentation will have different ramification toward the sense of co-presence and effectiveness of the communication. In this paper, we propose to carry out a preliminary study comparing various mixed reality based 3D collaborative media in two dimensions: (1) the form of the background (real vs. virtual) and (2) the form of user (photo-realistically reconstructed vs. pre-built 3D avatar). The traditional video based system is also compared as a reference for which both the background and user is represented in real but in 2D flat screen. We present the experimental design and report the results.

Keywords: tele-conference, mixed reality, augmented reality, virtual reality, collaborative system, co-presence, virtual human

Concepts: • **Computing methodologies ~ Graphics systems and interfaces;** *Mixed/augmented reality;*

*e-mail:dongsik@etri.re.kr
† e-mail:gjkim@korea.ac.kr (Correspondence)

VR Meets PR 2016, December 05-08 2016, Macao
ACM 978-1-4503-4548-4/16/12.
http://dx.doi.org/10.1145/2992138.2992146

1 Introduction

Tele-conference systems (mostly in the form of 2D networked live/streamed videos) are widely used for communication among two or more participants [Firestone et al. 2007]. While it is a huge improvement over the non-visual voice based communication, video-based tele-conference systems still fall short of realizing "tangible" multi-party meetings offering only flat 2D upper body imageries, a fixed viewpoint, inconsistent gaze direction, and the restricted movement in the meeting room [Jo et al. 2014].

On the other hand, virtual reality (VR) based tele-conference, in which remote participants seem "teleported" as avatars, can bring about a higher sense of co-presence and share virtual space ("being there"), and thereby a more realistic communication experience [Gross et al. 2003; Beck el al. 2013]. Augmented reality (AR) based method adds on to such an improvement by sharing the "real" local user space, supporting the sense of other participants of "being here" [Pejsa et al. 2016; Maimone et al. 2013; Ranieri et al. 2016]. Such an AR based method was found to have positive effects on the power of persuasion, which is one measure for evaluating tele-conference systems [Jo et al. 2015]. Despite these expected merits, the communication effectiveness will depend on and vary according to the fidelities of the represented user and background (shared space) [Garau et al. 2003].

In this paper, we propose to carry out a preliminary study comparing VR and AR based 3D collaborative media in two dimensions of component representation fidelity: (1) the form of the background (real vs. virtual) and (2) the form of user (reconstructed vs. pre-built 3D avatar). By such a comparison, the study really compares various forms of mixed reality in terms of its functional and affective effects to the task of collaborative tele-conferencing. The traditional video based system is also compared as reference for which both the background and user is represented in real but in 2D flat screen. Then, we present the experimental design and report the partial results.

2 Related Works

To put our work into a proper context, we review mainly three areas of related literatures: (1) previous implementations of VR/AR/MR based tele-conference, social or collaborative systems, (2) approaches to virtual character creation and representation and varying effects toward social interaction, and (3) methods and results of evaluating co-presence in tele-conference situations.

VR/AR/MR Tele-conference Networked multi-user VR and AR technologies are increasingly being applied to the tele-conference systems that can situate the participants in remote shared spaces, virtual or real [Jo et al. 2014]. Just to cite few notables, Beck et al. presented a remote meeting system with the two communicating parties visualized in life-sizes in a virtual environment and investigated various interaction scenarios [Beck et al. 2013]. Many different form of VR/AR displays have been tested as for their effectiveness, such as the kinetic multiple mirror displays controlled by the remote user's head motions [Otsuka 2016], a life-size transparent 3D display [Ranieri et al. 2016], an optical see-through head-worn display [Maimone et al. 2013], a video see-through head-mounted display [Jo et al. 2015], digital projectors to offer the image (e.g. the captured person in the remote) [Pejsa et al. 2016], and an auto-stereoscopic 3D display [Jones et al. 2009]. While most of these implementations have looked into the basic system development and how particular component technologies (e.g. display type/size, immersive, projective) might affect co-presence and communication efficiency, our paper will focus on the qualitative form of the user and scene (background) representation, which roughly will also categorize the genre of the given medium (VR, AR, augmented VR (AVR), effectively real, see Section 3).

Virtual human for social interaction Virtual humans have been used as the teleported avatars in tele-conference and collaborative systems [Maimone and Fuchs 2011]. Raij et al. has found that users did not find interacting with the virtual humans to be much different from interacting with real humans at least with respect to information transfer [Raij et al. 2007]. Researchers have been striving to also equalize the sense of co-presence and the aspect of affect (aside from just exchanging information effectively) e.g. by adding physicality such as props [Chuah et al. 2012] and self-body representation [Steed et al. 2016] and using minimally realistic reconstructed avatars both visually [Robb et al. 2015] and behaviorally [Guadagno et al. 2007]. Shapiro et al. developed a rapid virtual human capture system that mixed the reconstructed 3D model mesh with texture blending using a depth camera [Shapiro et al. 2014]. Feng et al. introduced an auto-rigging method for adapting avatar motion to body scanned virtual characters [Feng et al. 2015]. In particular, one very important behavior in multi-user communication is that of the mutual gaze (or head direction) and body gestures [Robb et al. 2016]. Hart and Proctor investigated the features of the conversational virtual human and its effects [Hart and Proctor 2016]. In our experiment, the avatars (photo-realistically reconstructed or avatar) both employ a basic gaze behavior.

Co-presence evaluation for tele-conference The most prevalent method in assessing presence or co-presence has been the use of surveys and questionnaires [Bailenson et al. 2003; Slater 1999; Witmer and Singer 1998]. Physiological signals (e.g. heart rate, skin conductance) have been suggested [Meehan et al. 2003] by only for singular presence. Presence (and co-presence) studies are concerned about various elements that might singularly or collectively promote (or demote) the sense of presence. Many such presence elements have been identified [Cho et al. 2003] and in particular, Garau at al. looked at the affective effects of virtual human appearance and media type (e.g. 2D or 3D environments) [Garau et al. 2003]. As co-presence and even communicative performance are believed to be connected to the user emotion, such qualities are increasingly being assessed under various conditions, e.g. by the form of the avatar (virtual human or unaugmented one) [Robb et al. 2013], by the avatar visualization type (realistic, cartoon like, and sketch like) [Volante et al. 2016], the level of behavior fidelity [Kim and Welch 2015], and the display type (stereoscopy and display size) [Camporesi and Kallmann 2016].

However, in all of these previous studies, no comprehensive work has been done in assessing the potential difference between VR and AR multi-user communication system as related to co-presence in terms of the combination of avatar and background representation forms.

3 Experimental Design

Table 1 illustrates the main factors (and levels) in the experimental designs for assessing the level of co-presence in the MR based tele-conference. Each factor (especially factors 1 and 2) is mixed and matched making each treatment represent a particular form of mixed reality (e.g. AR1, AR2, VR, and AVR). All the treatments use an MR based 3D tele-conference system that can be configured in terms of different forms of full sized avatars and scene backdrop (see Figure 2). For all the four cases, a head mounted display (HMD) is used and the video see-through (VST HMD) for AR1 and AR2.

Table 1: *The main factors (backdrop type and avatar type) for the experimental design assessing co-presence in various MR based tele-conference environment.*

Backdrop Type / Avatar Type	Real Environment	Virtual Environment
2D video (Reference)	Conventional Video based (Monitor)	-
Pre-built 3D Model	AR 1 (VST HMD)	VR (HMD)
Near-real Reconstructed 3D Model	AR 2 (VST HMD)	AVR (HMD)

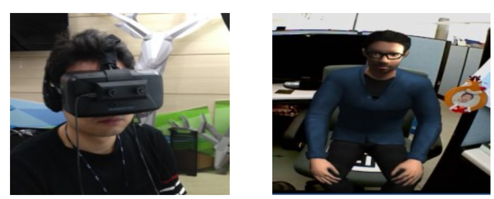

Figure 2: *The experimental MR based 3D teleconference system implemented using an HMD (attached with a dual camera) and video see-through (for AR1 and AR2) to accommodate various types of properly scaled life-sized avatars and backdrops.*

We regard the "near-real reconstructed 3D model" in AR2 to represent a pseudo-real character as compared to the "pre-built 3D model." The difference is the latter has no visual resemblance (except few basic features like skin color, wearing of a glass, approximate size) to the actual tele-conference counterpart participant, whereas the former is a 3D model of the same reconstructed as realistically as possible. The current experimental platform (e.g. avatar and background) is implemented using

Unity3D and runs on a 64-bit MS Window 8. The "near-real" avatar (used for AR2 and AVR condition) was reconstructed using the full-body scanning technology (http://www.3dplus.cn/) in the offline process (see Figure 3).

4 Expected Results and Discussion

We put forth two hypotheses that (1) AR1 and AR2 with real background (vs. VR and AVR) will exhibit a higher sense of co-presence and it will in turn contribute to an enhanced information transfer and understandability, and (2) between AR1 and AR2 or between VR and AVR, no significant differences in co-presence or task performance will be found. The former is expected as such as the communication occurs in the very same physical space and as for the latter, as already demonstrated in several related research [Ranieri el al. 2016] (that is "being here" is stronger than "being there"), with minimum facial and bodily features, the behavioral cues, such as gaze and gestures, play a more significant role in sensing co-presence and communicating effectively than the visual details.

Figure 3: *The actual user (left) represented in two forms, simple pre-built 3D avatar (center) with just the basic corresponding features and near-realistic reconstructed one (right).*

To assess co-presence and measure how much the teleported participant are felt to be "here" in the same location (as one indicator of MR tele-conference effectiveness), we conducted a simple informal test by showing imageries from the four different MR based tele-conference visualizations (see Figure 4), and asked of the expected level of co-presence (answered in 7 point Likert scale). Five subjects with the mean age of 38.2 years participated in the experiment. Results concurred with our first hypotheses in that significant differences in score were found among AR1/AR2 (average of 6.2), VR/AVR (4.4), and 2D (3.6). The second hypothesis was also supported with not much score difference between VR (4.3) and AVR (4.5) and between AR1 (6.1) and AR2 (6.3). Our plan is to conduct a formal controlled experiment as described in Table 2 and report the final results. Because realistic reconstruction of full sized human user - for AR2 - is costly and inconvenient, our expectation is that AR1 would be both the most effective (that showed the similar score with AR2) and economic form for future tele-conference or tele-collaboration systems.

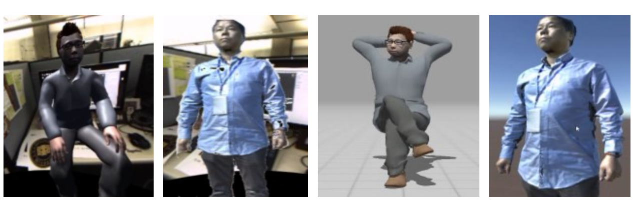

Figure 4: *Four sample imageries from the four different MR based tele-conference system, from the left, AR1, AR2, VR and AVR.*

Table 2 shows the possible experimental tasks, and both subjective and quantitative methods for measuring the level of co-presence. The experimental task, aside from just experiencing the tele-conference and interacting with a counterpart participant (from which the sense of co-presence could be induced and measured at the basic level), was designed to appraise the degree of information convey. Simple scripted (and not completely open-ended and not known to the subject) inquisitive conversations in a job interview, self-introduction, and role-based negotiation setting are exchanged. Later the subject would be questioned about the content of the conversation and one's understanding evaluated.

Table 2: *Possible experimental tasks and dependent variables (responses to the subjective presence survey and physiological signal differences).*

Experimental tasks	Job interviews	Negotiation (Role-playing)	Self-disclosure
	Passive discussion	Active discussion	
Measuring methods	Questionnaire (Co-presence, and Emotion response)	Physiological tests (Skin conductance, heart rate, skin temperature, EMG, and/or EEG)	

5 Conclusion

In this paper, we proposed to carry out a preliminary study of the effectiveness of mixed reality based 3D collaborative systems in two dimensions: (1) the form of the background (real vs. virtual) and (2) the form of user (photo-realistically reconstructed vs. pre-built 3D avatar). We also hope to demonstrate advantages of 3D and mixed reality over the conventional 2D video based tele-conference. We hypothesize that using real local background scene and behaviorally faithful avatars will be the most effective way as tentatively confirmed by the informal pre-survey. Therefore, in the future, the formal full blown experiment will be carried out.

Acknowledgements

This work was supported by the ICT R&D program of MSIP/IITP (I5501-16-1016, Instant 3D based Join & Joy content technology).

References

BAILENSON, J. N., BLASCOVICH, J., BEALL, A. C., AND LOOMIS, J. M. 2003. Interpersonal distance in immersive virtual environments. *Personality and Social Psychology Bulletin 29, 7,* 819-833.

BECK, S., KUNERT, A., KULIK, A., AND FROEHLICH, B. 2013. Immersive group-to-group telepresence. *IEEE Transactions on Visualization and Computer Graphics 19, 4,* 616-625.

CAMPORESI, C., AND KALLMANN, M. 2016. The Effects of Avatars, Stereo Vision and Display Size on Reaching and Motion Reproduction. *IEEE transactions on visualization and computer graphics, 22, 5,* 1592-1604.

CHO, D., PARK, J., KIM, G. J., HONG, S., HAN, S., AND LEE, S. 2003. The dichotomy of presence elements: The where and what. *In Proceedings of IEEE Virtual Reality*, 273-274.

CHUAH, J. H., ROBB, A., WHITE, C., WENDLING, A., LAMPOTANG, S., KOPPER, R., AND LOK, B. 2012. Increasing agent physicality to raise social presence and elicit realistic behavior. In *IEEE Virtual Reality Workshops (VRW),* 19-22.

FIRESTONE, S., RAMALINGAM, T., AND FRY, S. 2007. Voice and video conference fundamentals. Cisco Press.

FENG, A., CASAS, D., AND SHAPIRO, A. 2015. Avatar reshaping and automatic rigging using a deformable model. In *Proceedings of the 8th ACM SIGGRAPH Conference on Motion in Games,* 57-64.

GARAU, M., SLATER, M., VINAYAGAMOORTHY, V., BROGNI, A., STEED, A., AND SASSE, M. A. 2003. The impact of avatar realism and eye gaze control on perceived quality of communication in a shared immersive virtual environment. In *Proceedings of the SIGCHI conference on Human factors in computing systems,* 529-536.

GROSS, M., WÜRMLIN, S., NAEF, M., LAMBORAY, E., SPAGNO, C., KUNZ, A., AND STREHLKE, K. 2003. Blue-c: a spatially immersive display and 3D video portal for telepresence. *ACM Transactions on Graphics (TOG) 22, 3,* 819-827.

GUADAGNO, R. E., BLASCOVICH, J., BAILENSON, J. N., AND MCCALL, C. 2007. Virtual humans and persuasion: The effects of agency and behavioral realism. *Media Psychology 10, 1,*1-22.

HART, J. L., AND PROCTOR, M. D. 2016. Framework and Assessment of Conversational Virtual Humans as Role-players in Simulated Social Encounters with People. *Journal of the International Association of Advanced Technology and Science*.

JO, D., KIM, K. H., AND KIM, G. J. 2014. Avatar motion adaptation for AR based 3D tele-conference. In *Collaborative Virtual Environments (IEEE VR 3DCVE),* 1-4.

JO, D., KIM, K. H., AND KIM, G. J. 2015. SpaceTime: adaptive control of the teleported avatar for improved AR tele-conference experience. *Computer Animation and Virtual Worlds (CAVW) 26, 3-4,* 259-269.

JONES, A., LANG, M., FYFFE, G., YU, X., BUSCH, J., MCDOWALL, I. AND DEBEVEC, P. 2009. Achieving eye contact in a one-to-many 3D video teleconferencing system. *ACM Transactions on Graphics (TOG) 28, 3,* Article No. 64.

KIM, K., AND WELCH, G. 2015. Maintaining and Enhancing Human-Surrogate Presence in Augmented Reality. In *IEEE International Symposium on Mixed and Augmented Reality Workshops (ISMARW),* 15-19.

MAIMONE, A., AND FUCHS, H. 2011. A first look at a telepresence system with room-sized real-time 3d capture and life-sized tracked display wall. In *Proceedings of ICAT*, 4-9.

MAIMONE, A., YANG, X., DIERK, N., STATE, A., DOU, M., AND FUCHS, H. 2013. General-purpose telepresence with head-worn optical see-through displays and projector-based lighting. *In IEEE Virtual Reality (VR),* 23-26.

MEEHAN, M., RAZZAQUE, S., WHITTON, M. C., AND BROOKS, F. P. 2003. Effect of latency on presence in stressful virtual environments. In *Proceedings IEEE virtual reality,* 141-148.

OTSUKA, K. 2016. MMSpace: Kinetically-augmented telepresence for small group-to-group conversations. *In IEEE Virtual Reality (VR),* 19-28.

PEJSA, T., KANTOR, J., BENKO, H., OFEK, E., AND WILSON, A. 2016. Room2Room: Enabling Life-Size Telepresence in a Projected Augmented Reality Environment. In *Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing.* 1716-1725.

RAIJ, A. B., JOHNSEN, K., DICKERSON, R. F., LOK, B. C., COHEN, M. S., DUERSON, M., AND LIND, D. S. 2007. Comparing interpersonal interactions with a virtual human to those with a real human. *IEEE transactions on visualization and computer graphics (TVCG) 13, 3,* 443-457.

RANIERI, N., BAZIN, J. C., MARTIN, T., LAFFONT, P. Y., POPA, T., AND GROSS, M. 2016. An Immersive Bidirectional System for Life-size 3D Communication. In *Proceedings of the 29th International Conference on Computer Animation and Social Agents (CASA),* 89-96.

ROBB, A., KOPPER, R., AMBANI, R., QAYYUM, F., LIND, D., SU, L. M., AND LOK, B. 2013. Leveraging virtual humans to effectively prepare learners for stressful interpersonal experiences. *IEEE transactions on visualization and computer graphics (TVCG) 19, 4,* 662-670.

ROBB, A., CORDAR, A., LAMPOTANG, S., WHITE, C., WENDLING, A., AND LOK, B. 2015. Teaming Up With Virtual Humans: How Other People Change Our Perceptions of and Behavior with Virtual Teammates. *IEEE transactions on visualization and computer graphics (TVCG) 21, 4,* 511-519.

ROBB, A., KLEINSMITH, A., CORDAR, A., WHITE, C., LAMPOTANG, S., WENDLING, A., AND LOK, B. 2016. Do Variations in Agency Indirectly Affect Behavior with Others? An Analysis of Gaze Behavior. *IEEE transactions on visualization and computer graphics (TVCG) 22, 4,* 1336-1345.

SHAPIRO, A., FENG, A., WANG, R., LI, H., BOLAS, M., MEDIONI, G., AND SUMA, E. 2014. Rapid avatar capture and simulation using commodity depth sensors. *Computer Animation and Virtual Worlds (CAVW) 25, 3-4,* 201-211.

SLATER, M. 1999. Measuring presence: A response to the Witmer and Singer presence questionnaire. *Presence: Teleoperators and Virtual Environments 8, 5,* 560-565.

STEED, A., PAN, Y., ZISCH, F., AND STEPTOE, W. 2016. The Impact of a Self-Avatar on Cognitive Load in Immersive Virtual Reality. *In IEEE Virtual Reality (VR)*.

VOLANTE, M., BABU, S. V., CHATURVEDI, H., NEWSOME, N., EBRAHIMI, E., ROY, T., AND FASOLINO, T. 2016. Effects of Virtual Human Appearance Fidelity on Emotion Contagion in Affective Inter-Personal Simulations. *IEEE transactions on visualization and computer graphics 22, 4,* 1326-1335.

WITMER, B. G., AND SINGER, M. J. 1998. Measuring presence in virtual environments: A presence questionnaire. *Presence: Teleoperators and virtual environments 7, 3,* 225-240.