The Effects on Presence of Personalized and Generic Avatar Faces

Philipp Ladwig, Christian Geiger

Hochschule Düsseldorf
Münsterstr. 156
40476 Düsseldorf
[philipp.ladwig], [geiger]@hs-duesseldorf.de

Abstract: With today's technology it has become possible to generate and control personalized as well as authentic avatar faces in 3D for social Virtual Reality (VR) applications, as Lombardi et al. [LSSS18] have recently shown. Creating a personalized avatar with facial expressions is expensive in terms of time, computational power and hardware. Against this background, the question arises whether the creation of such a costly avatar with facial expressions is justifiable. A simple, anthropomorphic and generic avatar could be sufficient and probably allows for the same perception of presence. We conducted an initial and brief experiment with 22 participants in two groups and found indications that an arbitrary (in this case a generic) anthropomorphic representation of the dialog partner seems to lead to a lower perception of social presence compared to a personalized representation that resembles the dialog partner. Furthermore, it seems that co-presence is not affected by a personalized avatar. However, further research as well as a more sophisticated experiment design is necessary to finally verify our hypotheses.

Keywords: Social Presence, Co-Presence, Avatar, Facial Expressions, Virtual Reality

1 Introduction

For decades, Virtual Reality research has been exploring remote collaboration, social interaction and the effects on presence. Many different styles of avatars and render fidelity were tested, but creating authentic avatars was always a technical issue and has its limitations.

In 2003, Nowak and Boccia [NB03] were one of the first researchers who investigated the effects of facial expressions of anthropomorphic agents and avatars in VR. Due to the available technology at this time, the faces were far from photo-realistic. More recently, different research groups analyzed social and perceptional effects on realistic self-avatars (e.g. Piryankova et al. [PSR+14], Latoschik et al. [LRG+17], Waltemate et al. [WGR+18]). The scanning systems, used by these research groups, are relative complex and can only be conducted under laboratory conditions and by trained staff. However, none of these systems was used to explore the effect of anthropomorphic personal avatars (those that look like the respective real persons) with facial expressions compared to non-personal avatars that have a generic face with facial expressions during a remote, face-to-face dialog in VR.

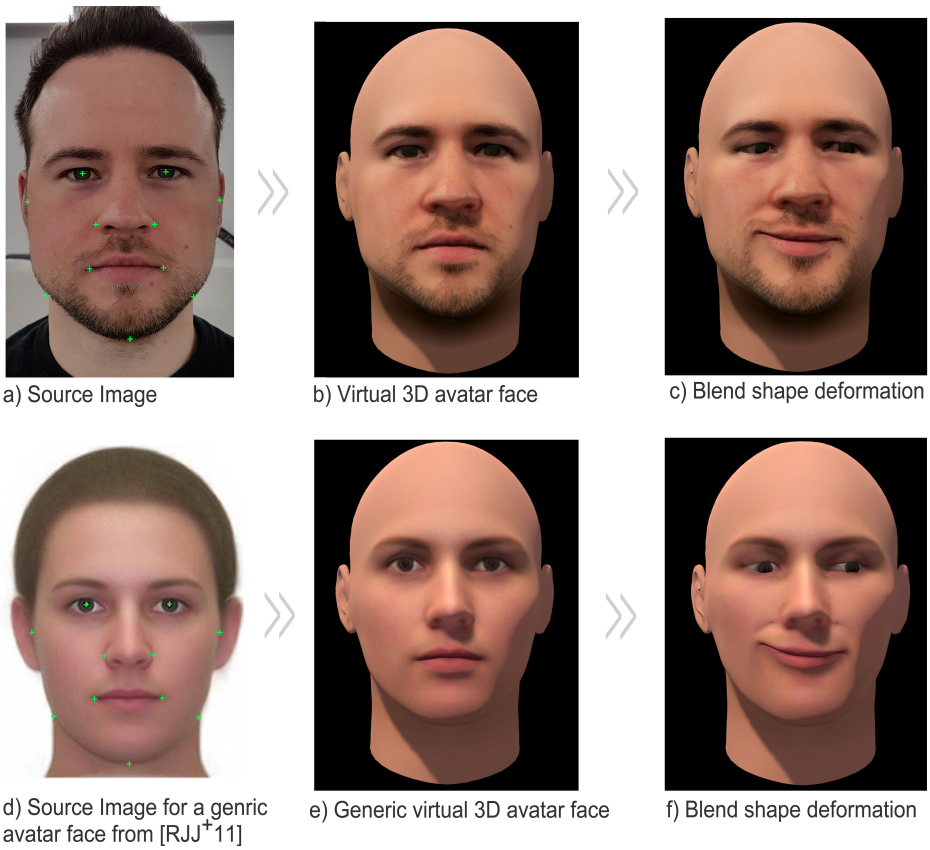

Figure 1: a) Input image for the creation of a personalized avatar. Green crosses are landmarks for the 3D avatar face generation algorithm; **b)** Personalized 3D avatar created from image a); **c)** Personalized avatar deformed by blend shapes; **d)** Generic avatar created from the androgynous norm by Rohdes et al. [RJJ+11]. Generation of avatar head is identical to a) **e)** Generic 3D avatar created from image d); **f)** Generic avatar deformed by identical blend shapes that are shown in c);

1.1 Contribution

The work reported in this paper analyzes the impact on social presence and co-presence induced by the appearance of personalized and non-personalized (in this paper called 'generic') avatar faces with facial expressions. The faces were created by a tool called *FaceGen* [Fac19] by using three photos of the participants face (front, right, left). The generic avatar face is created from an androgynous average image of morphing 48 images (24 female, 24 male) into one. The average image was taken from Rohdes et al. [RJJ+11], as shown in Fig. 1d. A set of 16 different blend shapes is controlled by a facial capture system, which is build into an HMD. The blend shape deformations of each face are identical, as shown in Figure 1c and 1f. The system is capable of real time processing with 90 frames per second.

The topic of this paper is relevant to social interaction and conference systems in Augmented and Virtual Reality. Social presence and co-presence is a key factor in communication and a long term goal for Mixed Reality systems. This work may help understanding the effects during remote meetings with avatars and gives initial indication why the creation of personalized avatars could be beneficial compared to generic ones.

1.2 Hypotheses

To the knowledge of the author, no facial capture system has yet been used with personal avatars in a virtual environment to investigate co-presence and social presence. Therefore, it remains an open question if personalized avatars create more presence than generic ones. The paper investigates the following hypotheses:

H1: A personalized avatar face with facial expressions (based on predefined blend shapes) and driven by a human in real time increases the perception of **co-presence** compared to a generic avatar face with the same predefined facial expressions.

H2: A personalized avatar face with facial expressions (based on predefined blend shapes) and driven by a human in real time increases the perception of **social presence** compared to a generic avatar face with the same predefined facial expressions.

2 Related Research

2.1 Presence

The understanding of presence differs between researchers. Thus, there is a great number of different definitions and meanings [Bio97, BC02, BHB03, NB03, You03, OBW18]. This paper is based on the following definition of Biocca, Youngblut and Oh [BHB03, You03, OBW18]. Co-presence is defined as "... the subjective experience of being together with others in a computer-generated environment, even when participants are physically situated in different sites." [You03, p. 4]. Here, the term "others" does not explicitly mean 'humans', as it is possible to feel co-present with computer-controlled agents or non-living objects.

Social presence excludes agents and objects while it addresses the social interaction with a real human being as well as the "...access to the intelligence, intentions, and sensory impressions of another" [You03]. Biocca stated "social presence occurs when users feel that a form, behavior, or sensory experience indicates the presence of another intelligence" [Bio97].

2.2 Face and Body Scanning

A long line of research has been conducted on the scanning of humans in a wide variety of technical approaches. Even many commercial products for body scanning exists. Simple and inexpensive systems are presented by Nagano et al. [NJJ+17], Straub and Kerlin [SK14], Gesslein et al. [GSG17] and Shapiro et al. [SFW+14]. Casas et al. [CAWF+15] presented a system for the creation of different facial blend shapes. A more sophisticated systems of a body and face scanner were shown and used by Achenbach et al. [AWLB17] and Bogo et al. [BRPMB17]. The latter two systems were also used for social and perceptual studies in Latoschik et al. [LRG+17] and Piryankova et al. [PSR+14]. Some system are even capable of scanning in real time with multiple frames per second, as Orts et al. [ORF+16] have shown. A major difficulty of many system seems to be the detailed capture and animation of the face and the avoidance of the Uncanny Valley effect. This is also a problem of our system.

2.3 Facial Expression Recognition under a Head-Mounted Display

This topic is a relatively new research area. Thies, Zollhöfer and Stamminger [TZS+16] presented a system, called FaceVR, which is capable of reenacting the user's face in a video. For this purpose, the mouth region is tracked by a commodity webcam with a fixed position (not attached to the HMD). Moreover, the eye motion is tracked by a single IR eye-tracking camera attached inside the HMD. A limitation of the system is the recognition of movements in the area around the eye. Thus, eyebrow movements can not be captured.

Li et al. [LTO+15] presented a system using a depth camera attached to the HMD to track the uncovered area around the mouth. Recently, a patent has been granted on this system [Hao17]. Compared to FaceVR [TZS+16] Li et al. do not track the eye movement, but their system is capable of recognizing movements around the eyes by using thin strain sensors in the foam liner of the HMD. Casas et al. [CSF+16] extended FaceVR with the creation and usage of personalized face meshes, textures and blend shapes for facial animation in real time. This system is most closely related to ours.

Lombardi et al. [LSSS18] presented the most advanced system so far. They introduced a deep appearance model for rendering human faces, which employs Generative Adversarial Networks to create personalized avatar faces with personalized facial animations that have photo-realistic look. To achieve this, the system uses cameras attached in- and outside of the HMD to capture the region around the mouth and eyes. Therefore, it is capable of eye and eyebrow tracking and allows for the most expressive avatar faces in this research field.

None of the authors of the aforementioned articles evaluated their system with regard to qualitative or quantitative research of presence.

3 System

The following section describes our work flow and system for creating and animating personalized avatar faces. A video of the system in action can be watched here:
`https://vimeo.com/348100587/96a0571c52`

3.1 Avatar Creation

The creation of personalized avatars is conducted with the software development kit of FaceGen. It requires a minimum of one image of the front of the respective face with a neutral expression. Two further images from the sides of the face improve the visual quality and the personal shape of the resulting avatar head. The process is largely automated, as a script reads the input images and uses them to create an avatar head. The process for the generation of a 3D avatar head requires only the manual tagging of 29 landmarks in the front image and both side images. The landmarks are highlighted as green crosses in Fig. 1a and d. The creation of one personalized avatar face takes less than 90 sec.

FaceGen maps the tagged input images on a generic base head mesh and alters its specific areas according to the images. It uses the idea of statistical shape models (SSM) and

statistical appearance models (SAM). When applied to images of faces, SSMs and SAMs describe the mean shape and mean density distribution of a face within a certain population as well as the main modes of variations of shape and density distribution from their mean values. The availability of this quantitative information regarding the detailed anatomy of faces provides the possibility to deform the base head mesh accordingly to the input image.

Our work flow does not include the creation, capturing or animation of scalp hair or long beard. We only embedded short beard and eyebrows in the facial 2D texture. The reason for this is that, in our experience, our approaches do not conform with the visual quality of face details and lead to less authentic avatars. Furthermore, we do not scan the teeth, because we experienced technical limitations in recreation plausible teeth structure. Therefore, the personalized avatars have generic teeth and are the same for every person.

3.2 Facial Animation

In addition to the mesh and color map, FaceGen exports 112 different blend shapes for animating the face. These blend shapes are generic and not personalized for each person. In our application, that we used for the study, the blend shapes for the eye movements are

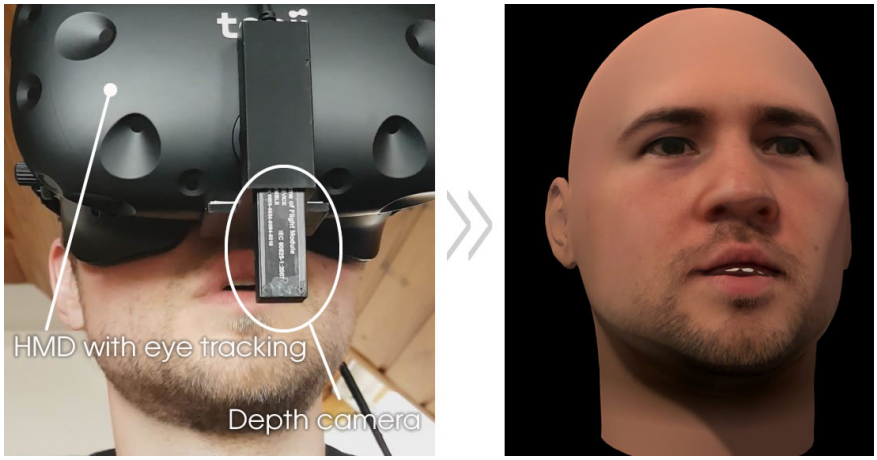

Figure 2: A depth camera is attached to an HMD to track the movements of the lower part of the face. The software by BinaryVR [Bin19] is used to map real facial expressions to virtual ones of an avatar.

driven by a Tobii Pro eye tracker, which is built into an HTC Vive HMD. The eye tracking allows for capturing the gaze direction as well as the blinking for each eye. In addition, the mouth region is captured by a PMD Pico Flexx depth camera attached to the HMD by means of a 3D printed mount (Fig. 2). The software development kit of BinaryVR [Bin19] is used for extracting the position of different parts of the lower face region such as the chin and corners of the mouth. The system is build on Unity version 2018.3.11f1.

Our system focuses on eye and lip tracking. It is not capable of tracking the eyebrows or the movement of the tongue. Furthermore, teeth and tongue are identical for every avatar.

4 Experiment

4.1 Participants

Eleven dyads, in sum 22 persons (two female and 20 male, ages 21–36 years, M = 27.04 years, SD = 4.26), attended in the experiment. All participants were students and members of the local department of computer science and were experienced with VR and AR technology. Both parts of each dyad already knew each other and, therefore, knew the other person's character traits, facial expressions and voice. The average required time per dyad, post questionnaires, unstructured interview and debriefing was 25 minutes. The average time in VR lasted about 10 minutes.

Figure 3: Setup of the study; a) Participant wears a HTC Vive with eye and lip tracking and talks remotely to a person in another room. The test conductor at the table observes the experiment; b) Participant in another room listen during the remote person speaks.

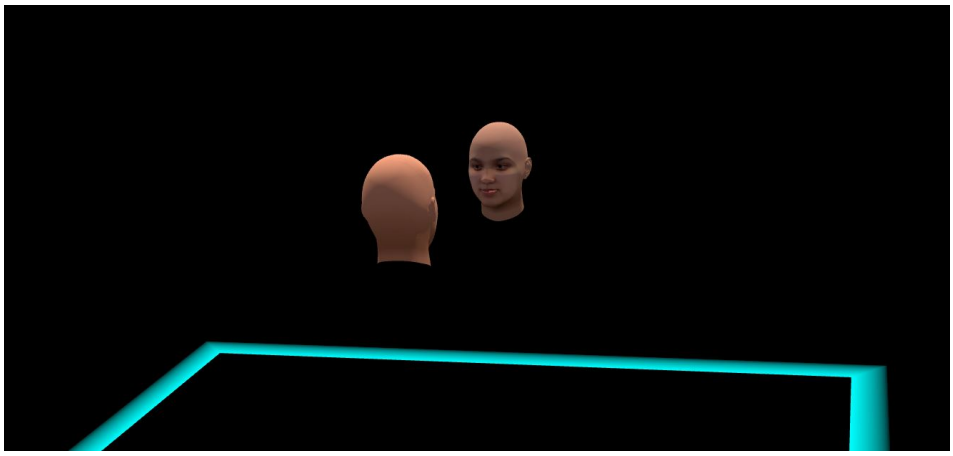

Figure 4: An in-game camera shows the set up within the game engine Unity. No hands, controllers or body is shown.

4.2 Method

The test conductor introduced the participant with a structured procedure to make sure that each participant gets the same information. The test conductor did not inform the participants about the main question of the study. It was only explained that it is a test about the effect of eye and lip tracking in VR. This was made in order to not create a biased understanding of the study. The test was build as a between groups design.

Each dyad completed a consent form and were asked to put on the HMDs and headphones. The participants were in two separated rooms. The setup of the experiment is shown in Figure 3. It was not possible to hear or see the other person physically. The interpupillary distances (IPD) was adjusted; eye and lip tracking were calibrated. For calibration purposes and getting accustomed to the situation a virtual mirror was present in which both participants saw themselves and the avatar face of the remote person with facial animations in real time driven by the eye tracker and depth camera.

The test conductor and the two participants were verbally connected over TeamSpeak [Tea19]. The task was simply to talk about everyday matters such as the last weekend, the next holidays and similar topics. The test itself starts when the test conductor retreated from the Teamspeak chat, retained as listener, and removes the virtual mirror. At this moment, the virtual scene contains only the two avatar faces of the participants, as shown in Figure 4. In many cases the dialog partners immediately starts a conversation. If the dialog partners run out of topics for the conversation, the test conductor introduces new threads by showing a virtual board with questions such as "What are your currently working on?" or ”Tell a joke!”. After an average time span of 10 minutes the test conductor asked to remove the HMDs and presented a post-questionnaire. Questions are summarized in Figure 5 on the next page. The questionnaire contains demographic information and eight five-point Likert scales. The questionnaire is inspired by Nowak and Boccia [NB03] and was also used by Latoschik et al. [LRG+17] for measuring co- and social presence. We selected the relevant components of co-presence and social presence of the Nowak and Boccia's questionnaire for our study.

Ten participants were engaged into 'personalized-face-to-personalized-face' dialogues, another ten participants were engaged into 'generic-face-to-generic-face' dialogues and two participants had a 'generic-face-to-personalized-face' dialog. In sum, we generated eleven personalized 3D avatar heads with FaceGen.

4.3 Results

The test results are visualized in Figure 5 for the conditions 'personalized avatar face' and 'generic avatar face' (bars with yellow line). A Mann-Whitney U test for independent samples was conducted on each question with a significance level of $p = .05$. The test indicates that no statistical difference of the questions regarding to co-presence seems to be present between the two groups. This may be an indication for rejecting H1.

However, the test indicates a significant difference between the test conditions for social

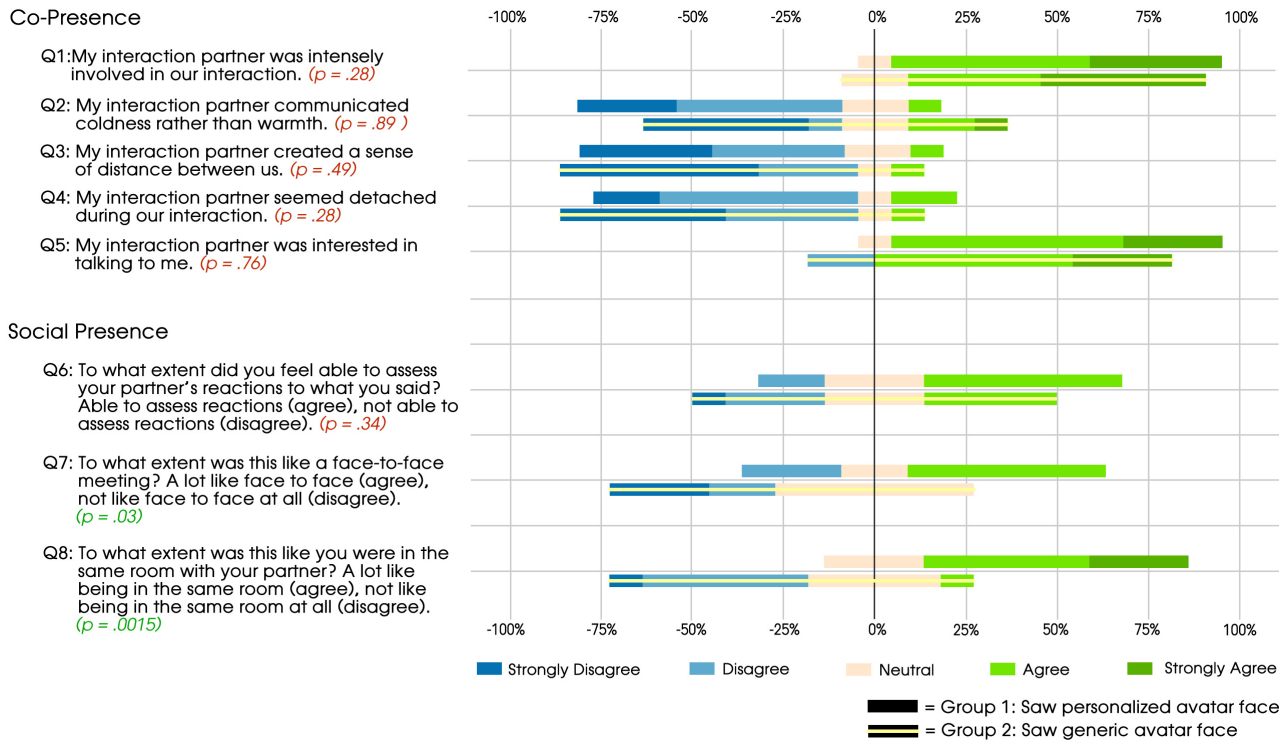

Figure 5: Results of the questionnaires visualized as diverging stacked bar chart. Each bar represents the answers for the Likert scales of eleven participants. The left half of the chart depicts disagreement, the right half shows agreement. 'Neutral' represents the neutral answer of the Likert scale.

presence, as question Q7 and Q8 reveals. The difference for question Q8 is even highly significant with $p < .01$. Two of three questions regarding to social presence shows a significant difference between both groups. This may be an indication to consider H2 is confirmed.

The unstructured interview after each test with both participants of a dyad underpins the confirmation of H2: Four participants of the group, who saw the personalized avatar, praised the quality of the dialog and stated something similar to "The meeting was surprisingly real" while none of the participants of the group, who saw the generic one, highlighted the realism.

5 Discussion and Limitations of the Study

A major limitation of our study is the way of measuring co-presence and social presence. Our questionnaire is inspired from Nowak and Boccia [NB03], but it is not standardized and, therefore, makes the measurement of co- and social presence difficult. While different questionnaire exist for measuring self-presence, it seems it exists no standardized, international and widely accepted questionnaire or procedure to measure co- and especially social presence. Nevertheless, further research should include different questionnaires such as presented by Blascovich et al. [BCBMLNB02], Egerto et al. [EOT64] and Smith and Neff [SN18]. Moreover, Slater claims [Sla04] that questionnaires are only one tool of many. Quantitative

data such as bio signals or measuring subconscious and unintentional behaviour (as it is provoked in the Rubber Hand Illusion experiment [IdKH06]) could reveal further insights and substantiate the findings of the paper.

Although, the use of a not standardized questionnaires may be a limiting factor of the validity of the study, nevertheless, we were surprised by the great difference between answers of the two groups in Q8. Despite of the lack of photo-realistic rendering (missing scalp hair, generic teeth, no tracking of eye brows and tongue as well as floating avatar heads without bodies) the two versions of the avatar faces seems to were subjectively received in very different ways between the groups.

Another limiting factor of the study is the fact that both participants of each trial knew each other. Feng et al. [FLM+14] performed a study with 3D avatars that were scanned and animated by a real person. A set of body gestures were recorded and applied to different human avatars. The study shows that observers found the 3D avatar performance that used the same body gestures as the original human subject were rated as 'more like' the original human subject by groups that knew that subject, than 3D avatars that performed using another human subject's gestures. Facial gestures and expressions were not subject of the study of Feng et al. and is a major difference to our study, however, we believe the reported effect by Feng et al. is also transferable to faces and may have affected the ratings of the questionnaire of our study. There may be a difference between people who know each other and people who do not.

6 Conclusion

We have presented indications that a personalized avatar face with facial expressions, compared to a generic face with the same facial expressions based on blend shapes, may not increases the perception of co-presence while immersed in an virtual environment. However, we found indications that it may increase the perception of social presence.

This paper presented an initial and brief experiment and discussed a possible effect which should be further investigated and proven with a larger amount of participants and by a more sophisticated experiment design.

The findings in this paper gives some first indications to answer the question in the opening paragraph: Is the costly creation of a personalized avatar justifiable? If such an avatar increases the perception of social presence and the mental connection between two persons, than personalized avatars could be vital for vivid and authentic social experiences in Mixed Reality.

6.1 Acknowledgments

This research has been funded by the Federal Ministry of Education and Research (BMBF) of Germany in the framework of Interactive body-near production technology 4.0 (German: Interaktive körpernahe Produktionstechnik 4.0 (iKPT4.0) - project number 13FH022IX6).

References

[AWLB17] Jascha Achenbach, Thomas Waltemate, Marc Erich Latoschik, and Mario Botsch. Fast generation of realistic virtual humans. 2017.

[BC02] Frank Biocca and Harms Chad. Defining and measuring social presence: Contribution to the networked minds theory and measure. *Proceedings of PRESENCE*, 2002.

[BCBMLNB02] Jim Blascovich, Andrew C. Beall, Jack M Loomis, and Jeremy N. Bailenson. Equilibrium theory revisited: Mutual gaze and personal space in virtual environments. *Presence: Teleoperators Virtual Environments*, 2002.

[BHB03] Frank Biocca, Chad Harms, and Judee K. Burgoon. Toward a more robust theory and measure of social presence: Review and suggested criteria. *Presence: Teleoper. Virtual Environ.*, 2003.

[Bin19] BinrayVR - Face tracking for Virtual Reality. `https://binaryvr.com`, 2019. Accessed: 2019-05-24.

[Bio97] Frank Biocca. The cyborg's dilemma: Progressive embodiment in virtual environments. *Journal of Computer-Mediated Communication*, 1997.

[BRPMB17] Federica Bogo, Javier Romero, Gerard Pons-Moll, and Michael Black. Dynamic FAUST: Registering Human Bodies in Motion. 2017.

[CAWF+15] Dan Casas, Oleg Alexander, Andrew W. Feng, Graham Fyffe, Ryosuke Ichikari, Paul Debevec, Ruizhe Wang, Evan Suma, and Ari Shapiro. Blendshapes from commodity RGB-D sensors. 2015.

[CSF+16] Dan Casas, Ari Shapiro, Andrew Feng, Oleg Alexander, Graham Fyffe, Paul Debevec, Ryosuke Ichikari, Hao Li, Kyle Olszewski, and Evan Suma. Rapid Photorealistic Blendshape Modeling from RGB-D Sensors. 2016.

[EOT64] Charles Egerton, George J Suci Osgood, and Percy Tannenbaum. The measurement of meaning. *Proceedings of the Seventh International Conference on Motion in Games*, 1964.

[Fac19] FaceGen - 3D Human Faces. `https://facegen.com/`, 2019. Accessed: 2019-05-24.

[FLM+14] Andrew Feng, Gale Lucas, Stacy Marsella, Evan Suma, Chung-Cheng Chiu, Dan Casas, and Ari Shapiro. Acting the Part: The Role of Gesture on Avatar Identity. *Proceedings of the Seventh International Conference on Motion in Games*, 2014.

[GSG17] Travis Gesslein, Daniel Scherer, and Jens Grubert. BodyDigitizer: An Open Source Photogrammetry-based 3D Body Scanner. 2017.

[Hao17] Kyle Olszewski Hao Li, Joseph J. Lim. High-fidelity facial and speech animation for virtual reality head mounted displays. 2017.

[IdKH06] Wijnand A. IJsselsteijn, Yvonne A. W. de Kort, and Antal Haans. Is This My Hand I See Before Me? The Rubber Hand Illusion in Reality, Virtual Reality, and Mixed Reality. *Presence: Teleoper. Virtual Environ.*, 2006.

[LRG+17] Marc Erich Latoschik, Daniel Roth, Dominik Gall, Jascha Achenbach, Thomas Waltemate, and Mario Botsch. The effect of avatar realism in immersive social virtual realities. 2017.

[LSSS18] Stephen Lombardi, Jason Saragih, Tomas Simon, and Yaser Sheikh. Deep Appearance Models for Face Rendering. 2018.

[LTO+15] Hao Li, Laura Trutoiu, Kyle Olszewski, Lingyu Wei, Tristan Trutna, Pei-Lun Hsieh, Aaron Nicholls, and Chongyang Ma. Facial performance sensing head-mounted display. *ACM Transactions on Graphics*, 2015.

[NB03] Kristine L. Nowak and Frank Biocca. The Effect of the Agency and Anthropomorphism on users' Sense of Telepresence, Copresence, and Social Presence in Virtual Environments. *Presence: Teleoperators and Virtual Environments*, 2003.

[NJJ+17] Koki Nagano, Jaewoo Seo, Jens Fursund, Iman Sadeghi, Carrie Sun, Yen-chun Chen, Liwen Hu, Shunsuke Saito, Lingyu Wei, Jaewoo Seo, Jens Fursund, Iman Sadeghi, Carrie Sun, Yen-Chun Chen, and Hao Li. Avatar Digitization From a Single Image For Real-Time Rendering. *ACM Transactions on Graphics*, 2017.

[OBW18] Catherine S. Oh, Jeremy N. Bailenson, and Gregory F. Welch. A Systematic Review of Social Presence: Definition, Antecedents, and Implications. *Frontiers in Robotics and AI*, 2018.

[ORF+16] Sergio Orts, Christoph Rhemann, Sean Fanello, David Kim, Adarsh Kowdle, Wayne Chang, Yury Degtyarev, Philip L. Davidson, Sameh Khamis, Minsong Dou, Vladimir Tankovich, Charles Loop, Qin Cai, Philip Chou, Sarah Mennicken, Julien Valentin, Pushmeet Kohli, Vivek Pradeep, Shenlong Wang, and Shahram Izadi. Holoportation: Virtual 3d teleportation in real-time. 2016.

[PSR+14] Ivelina Piryankova, Jeanine Stefanucci, Javier Romero, Stephan de la Rosa, Michael Black, and Betty Mohler. Can I Recognize My Body's Weight? The

Influence of Shape and Texture on the Perception of Self. *ACM Transactions on Applied Perception*, 2014.

[RJJ+11] Gillian Rhodes, Emma Jaquet, Linda Jeffery, Emma Evangelista, Jill Keane, and Andrew J. Calder. Sex-specific norms code face identity. *Journal of vision*, 2011.

[SFW+14] Ari Shapiro, Andrew Feng, Ruizhe Wang, Hao Li, Mark Bolas, Gerard Medioni, and Evan Suma. Rapid avatar capture and simulation using commodity depth sensors. *Comput. Animat. Virtual Worlds*, 2014.

[SK14] Jeremy Straub and Scott Kerlin. Development of a Large, Low-Cost, Instant 3D Scanner. *Technologies*, 2014.

[Sla04] Mel Slater. How colorful was your day? Why questionnaires cannot assess presence in virtual environments. *Presence: Teleoperators and Virtual Environments*, 2004.

[SN18] Harrison Jesse Smith and Michael Neff. Communication Behavior in Embodied Virtual Reality. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems - CHI '18*, 2018.

[Tea19] Teamspeak. `https://www.teamspeak.com/`, 2019. Accessed: 2019-05-24.

[TZS+16] Justus Thies, Michael Zollhöfer, Marc Stamminger, Christian Theobalt, and Matthias Nießner. FaceVR: Real-Time Facial Reenactment and Eye Gaze Control in Virtual Reality. 2016.

[WGR+18] Thomas Waltemate, Dominik Gall, Daniel Roth, Mario Botsch, and Marc Erich Latoschik. The impact of avatar personalization and immersion on virtual body ownership, presence, and emotional response. *IEEE Transactions on Visualization and Computer Graphics*, 2018.

[You03] Christine Youngblut. Experience of Presence in Virtual Environments. *Defense Technical Information Center*, 2003.