Effects of Facial and Voice Similarity on Presence in a Public Speaking Virtual Environment

Laura Aymerich-Franch, Cody Karutz and Jeremy N. Bailenson,

Virtual Human Interaction Lab, Stanford University
{laura.aymerich@gmail.com, ckarutz@stanford.edu, bailenson@stanford.edu}

Abstract

Self-presence examines how much a user extends features of his or her identity into a virtually mediated world while represented by an avatar. Additionally, social and spatial presence measure how a user treats actors and environments in mediated space as if they were real. In this study, we examined the effects of facial and voice similarities of participants on self-reported presence measures in a public speaking task that takes place in an immersive virtual environment. Experimental participants (N = 51) were instructed to give a five-minute speech in front of a virtual audience in a two (voice similarity) by two (face similarity) design. For facial similarity, participants saw their avatar's face reflected back in a virtual mirror with either a similar or dissimilar face. For voice similarity, participants either gave their own speech out loud, or had a previous participant's speech emanate from their avatar's mouth. Results showed participants in the similar voice condition reported significantly higher self-presence and social presence than those in the dissimilar voice condition. Facial similarity did not significantly affect any measures of presence. We discuss implications for the study and design of avatars.

Keywords: **Presence, virtual reality, virtual environments, voice, face, avatar**

1. Introduction

Media is constantly expanding its ability to capture both behaviors and appearances of users. Devices such as Nintendo Wii, Microsoft Kinect, and other accessible mobile technologies are able to capture a tremendous amount of data about a user in a way that was never before possible. As a result, the amount of customization that users can experience in a virtually mediated environment is increasing.

What are the implications of this customization on behaviors? As the time users spend in virtual environments continues to grow, it is important to look at exactly how the concept of the self is extended into those mediated environments. The term encompassing this digital self is often called an *avatar*, and can include variable amounts of photorealism or lifelike behaviors. Now, users can customize an avatar that contains similar hair, eyes, face, or body type. On the opposite end of the spectrum, users can control avatars that look nothing their actual selves. There is a need for research that examines the effects of extending the self into a virtual environment when the avatar is customized to look highly similar or dissimilar to the user.

In spite of that, research needs to pay more attention to how users' relation to the created self influences their connection to the virtual environment and mediates their feeling of presence (Morie, 2008). The concept of presence can be used to examine how much users extend their identity to avatars in mediated environments. Overall presence can be broadly defined as a psychological state in which virtual objects are experienced as actual objects in either sensory or non-sensory ways (Lee, 2004).

Following Lee (2004), we can consider three types of presence: self, social and physical. Self presence is the psychological state in which virtual self is experienced as the actual self. Social presence is the psychological state in which virtual social actors are experienced as actual social actors. Finally, physical or spatial presence is that in which virtual objects are experienced as actual physical objects.

Among the three types, self-presence is the most relevant to the purpose of our study. Self-presence provides a standardized framework and operationalization for describing how users connect to their mediated self-representations (Ratan & Hasler, 2010). This sub-type of presence can help in understanding how users extend features of their identity into a virtually mediated environment. The concept was introduced by Biocca in the late nineties to describe the effect of virtual environment on the perception of one's body, physiological states, emotional states, perceived traits, and

identity (Biocca, 1997). Lee (2004) and Ratan (2010, 2011a; Ratan & Hasler, 2009) have contributed to the operationalization of it. However, the construct lacks a standard definition and it has not received rigorous empirical treatment.

Beyond the theoretical design work, previous experimental work has examined how changes in the user's avatar (the mediated identity) affect presence. These studies fall into one of two categories: behavior or appearance manipulations.

In a mediated virtual environment, the synchronous connection between a user's physical movements and the user's avatar's virtual movements play a critical factor in inducing self-presence. In a study by Yee and Bailenson (2009) users with high similarities between physical and virtual movements showed an increased psychological connection between the user and his or her virtual self. This relationship also extends to hand movements, where high synchronous movements between a user's physical and virtual hands created an increased sense of ownership of the user's virtual body (Slater, Spanlang, Sanchez-Vives & Blanke, 2010). Users also showed increased learning of performance in a Tai Chi task when viewed a 3D point cloud of their physical movements as opposed to conditions in which they did not see their own body movements on an avatar (Bailenson et al., 2008).

When looking at appearance, users with similar or dissimilar avatars also demonstrated changes in associated behaviors while operating those mediated identities. Facial similarity caused participants to change their exercise behaviors in a study where users saw their avatars gain or lose weight in a virtual environment (Fox & Bailenson, 2009). Additionally, when using the Nintendo Wii, customization of a Mii avatar affected levels of self-presence. Users that were able to customize their avatars to make them similar to their physical selves reported higher levels of self-presence than dissimilar conditions (Hoshi & Waterworth, 2009; Ratan, 2011b; Ratan, Santa Cruz & Vorderer, 2007). Also, Eastin (2006) found that gender matching between female users' physical and virtual selves in a video game resulted in higher levels of presence than when gender was mismatched.

Previous research is still lacking in experimental manipulations that examine how avatar similarities of users specifically affect levels of self-presence. In particular, there is very little research on looking at voice as a manipulated variable for affecting self-presence. Nass and colleagues have found a similarity-attraction effect with voice interfaces that share a similar personality or accent with the user (Dahlbäck, Swamy, Nass, Arvidsson & Skågeby, 2001; Dahlbäck, Wang, Nass & Alwin, 2007; Nass & Lee, 2000). But these studies looked at the voice of another social actor, not the voice of the self instantiated in an avatar.

The current study aimed to further establish the relationship between mediated similarities of a user and self-reported levels of presence. More specifically, voice and facial similarities were manipulated in a virtual public speaking task. Since the construct of self-presence is based on the similarities of the user's virtual avatar to the user's physical self, we predicted that increasing the similarity of the user's face and voice would increase self-presence. For social presence, we expected there would be some relationship between similarity of face and voice and the manner in which a user interacts with others, but viewed the manipulation as exploratory. For environmental presence, we did not predict differences by condition as the virtual scene was highly realistic in all conditions.

2. Method

2.1. Sample

Fifty-one students of a medium-sized West Coast university participated in the experiment in exchange of course credit. Seven participants were discarded due to technical failure or motion sickness. The final sample (N=44) consisted of 19 male and 25 female who ranged in age from 19 to 23 (M=20.4, SD=1.08).

2.2. Design

We created a virtual classroom. Participants gave a speech in front of a virtual audience of twelve virtual humans. We combined face and voice variables in a two by two between-subjects design. We assigned participants an avatar with a similar face modeled from a photograph of them or a dissimilar face of a previous participant of the same gender and skin color. Also, participants gave the speech either using their own voice or acted out the gestures along to the playback of a previous participant's speech audio. We placed a mirror on the back of the virtual classroom so participants were able to see their virtual representation at all times during the speech.

Participants were assigned to one of the four conditions combining similar face versus dissimilar face and own voice versus other participant's voice.

For correct avatar representation, we created eight categories considering gender (men/women) and skin color of participants (i.e. white Caucasian, white Asian,

Figure 1. Virtual classroom environment. A participant sees her avatar at the mirror during her performance.

Brown and Black). We assigned participants to one of the four experimental conditions in their category when they arrived, following this order: own face and own voice, own face and other's voice, other's face and own voice, and other's face and other's voice. In all dissimilar conditions, gender and ethnicity matched the participant. In the other's face other's voice condition we assigned the voice from a previous participant and the face from a different previous participant. This way, we ensured that all faces and speeches were used only once as a dissimilar condition. Also, this allowed us to control for public speaking skills and idiosyncratic facial features.

2.3. Apparatus

Participants wore an nVisor SX111 head-mounted display (NVIS, Reston, VA) with a resolution of 2056 x 1024 and a refresh rate of 120 frames per second. An optical tracking system (Worldviz PPTH) along with an orientation sensor (Intersense3 Cube) provided tracking on 6 degrees of freedom (x, y, z position and pitch, yaw, and roll) for the head. In addition, the participants wore trackers on the hands that tracked the x, y, z position of each hand. The virtual environment was generated and programmed using Worldviz's Vizard VR Toolkit.

2.4. Procedure

Participants partook in the experiment individually. They completed a pre-survey when they arrived. Afterwards, we asked them to give a five-minute speech

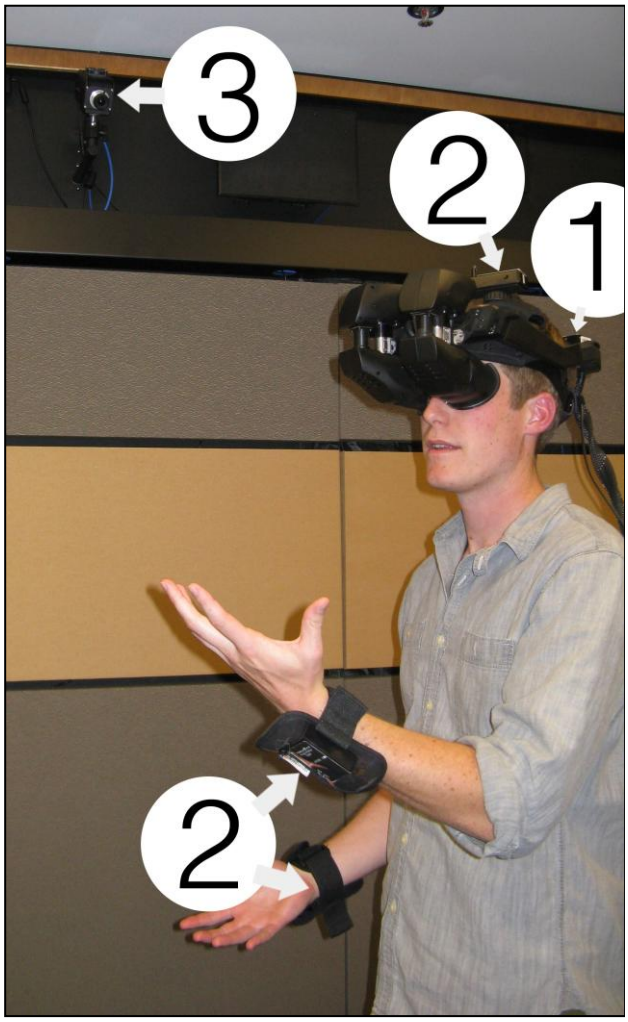

Figure 2. A participant during the experiment wearing the HMD (1), infrared tracking sensors on the head and hands (2), and cameras (3) that detect the position of the trackers.

about Stanford University in front of a virtual audience. Participants in the own voice condition were given five minutes time to prepare their speech.

When they were ready to begin, we accompanied them to the experimental VR room. We equipped them with the Head Mounted Display (HMD) and optical sensors on their wrists to detect hand movement. Participants were told that an avatar would represent them in the virtual environment. In the first stage of the VR task, a curtain opened and an empty classroom appeared. Participants had some time to walk along the stage in the classroom to become comfortable with the space. Afterwards, we asked them to look at the virtual mirror placed on the back of the room. They were able to see

their avatar in the mirror. We told them to lift both arms to make sure they were aware of their avatar self-representation. As soon as they felt ready, the curtain closed. We informed participants that in the next stage, the audience would arrive in the classroom. After a few seconds, the curtain opened again and the virtual audience was sitting watching the participant. In total, the audience was represented by four males and eight females of different races. Once the curtain was fully open, participants were either able to start their speech or the pre-recorded voice started automatically. For own voice conditions, speeches were recorded using an external microphone. After four and a half minutes, the researcher alerted participants that they had 30 seconds remaining. After the speech concluded, the curtain was closed one final time. We helped participants to take off the helmet and the sensors and accompanied them to the adjacent survey room. Finally, they completed a post-survey and were thanked for their participation.

2.5. Measurement

Among other measures, participants completed a 15-item presence scale (adapted from Nowak & Biocca, 2003) consisting in five items for self-presence (α=.850, M=3.56, SD=.786), five items for social presence (α=.894, M=2.89, SD=.852) and five items for spatial presence (α=.914, M=3.06, SD=.917). The exact wording of the questions is depicted in Appendix A. Also, we asked participants how similar was their face to the avatar's face they saw at the mirror, how similar was their body to the avatar's body they saw at the mirror and how much the voice of their avatar sounded like theirs, as manipulation checks. We considered subjects failed manipulation checks whether they answered *extremely* in other's conditions or *not at all* at own's conditions for face or voice questions.

3. Results

Univariate General Linear Model was used in SPSS to analyze the effect of voice and face on each type of presence: self-presence, social presence and spatial presence. Six subjects were excluded because they failed manipulation checks. The final analyses had 38 subjects. Appendix B shows means and SD of each group for self-presence, social presence and spatial presence.

3.1. Self-Presence

In order to analyze self-presence, the self-presence scale was included as a dependent variable with voice and face variables as fixed factors. The effect of voice was significant in determining self-presence (F=4.38, p=.044, partial η^2=.114). Participants that gave the own speech felt greater self presence compared to subjects that pretended to talk. Neither *face* (F=1.77, p=.192, partial η^2=.050) nor the interaction between *face* and *voice* (F=.09, p=.754, partial η^2=.003) were significant.

3.2. Social Presence

We ran an ANOVA similar to the previous analysis but with social presence as the dependent variable. Voice and face variables were included as fixed factors. The effect of voice was significant in determining social-presence (F=4.73, p=.037, partial η^2=.122). Participants that gave their own speech felt greater social presence compared to subjects that pretended to talk. However, neither *face* (F=.003, p=.957, partial η^2=.000) nor the interaction between *face* and *voice* (F=.22, p=.642, partial η^2=.006) were significant.

3.3. Spatial Presence

Spatial presence scale was included as dependent variable with face and voice as fixed factors. No significant effects were found. Neither *voice* (F=.540, p=.468, partial η^2=.016), nor *face* (F=.009, p=.926, partial η^2=.000) or the interaction between *face* and *voice* (F=2.06, p=.160, partial η^2=.057) were significant.

4. Discussion

This study preliminarily reveals that having an avatar with a similar voice can contribute to an increased sense of presence. In our experiment, using one's own voice to give a speech in a virtual environment increased the sense of self-presence and social presence. However, having an avatar with a similar or dissimilar face did not significantly affect self or social presence. In all conditions, voice or face similarities did not affect spatial presence.

One study limitation was that in the virtual mirror, we did not render elbow or leg movements of the user's avatar. Thus, the reflection of avatar in the mirror was limited in range of movements. Although a participant reported in the open-ended response that "the mirror in the back of the room helped me feel present because the

reflection was a face that looked like me and was always in my line of sight", many subjects underlined this constraint by stating, for example, that the "avatar arms didn't bend which was odd" or "the hand movements of my avatar were very stiff". In addition, we gave participants a standard male or female body that possibly differed from their actual body, potentially decreasing identification with the avatar in the similar face conditions. Resolution and realism of the participant's face in the mirror could be improved in a future version of the environment. Some participants commented on this, for example, "the picture of me however made me feel uncomfortable because it was just a 2D straight faced shot that was awkward to look at and made me feel a little foolish/embarrassed as it didn't show any emotion or movement" or "even when I saw myself in the mirror, because the avatar did not match my mouth movements, facial expressions, and hand gestures during my speech, I felt like I couldn't look into the mirror and see myself talking". These limitations in the avatar's appearance could have contributed to attenuate the effect of the variable *face* on the model.

On the other hand, the *voice* manipulation might be qualitatively different from the face manipulation. Participants had or did not have their own voice or their own face. However, there was an extra confound in the case of *own voice* conditions, as in that condition participants also were tasked with preparing and giving the own speech. Also, participants in the *non voice* condition acted out a recording of another participant. This allows for the possibility that other variables besides voice similarity were involved in the process.

Future studies should take into account these limitations. A follow up study could technologically manipulate voice so that participant's voices did not sound like their own to produce the "other voice" condition.

Our work contributes to better understand how avatar similarities of a user affect levels of presence, with particular focus on self-presence. Future follow-up studies include further analyzing the influence of avatar degrees of freedom on self-presence. For example, examining if having elbows, leg or head movements tracked and reflected in the virtual mirror increases self-presence. As virtual reality and digital technologies improve the ability to realistically represent a user, experiments focusing on the effects of similarities between the real and the digital self on presence are critical.

Technology that enables the creation of highly customizable avatars representing their users is rapidly improving. This convergence between the physical and the virtually mediated self has important repercussions on attitudes and behaviors of the user. However, consequences of these transformations have not been thoroughly studied yet since they represent a novelty in the media landscape. Thus, further theoretical and experimental studies of presence are needed to advance the understanding of self-representations in mediated environments.

References

Bailenson, J.N., Patel, K., Nielsen, A., Bajscy, R., Jung, S.-H., & Kurillo, G. (2008). The effect of interactivity on learning physical actions in virtual reality. Media Psychology, 11, 354–376.

Biocca, F. (1997). The cyborg's dilemma: Embodiment in virtual environments, Cognitive Technology. *'Humanizing the Information Age'. Proceedings., Second International Conference on* (pp. 12-26).

Dahlbäck, N., Wang, Q., Nass, C., & Alwin, J. (2007). Similarity is more important than expertise: Accent effects in speech interfaces. *Proc. SIGCHI conference on human factors in computing systems 2007*, 1553–1556.

Dahlbäck, N., Swamy, S., Nass, C., Arvidsson, F., Skågeby, J. (2001). Spoken Interaction with Computers in a Native or Non-Native Language - Same Or Different?*Proceedings INTERACT 2001*.

Eastin, M. S. (2006). Video game violence and the female game player: Self-and opponent gender effects on presence and aggressive thoughts. *Human Communication Research, 32*, 351-372.

Fox, J., & Bailenson, J.N. (2009). Virtual self-modeling: The effects of vicarious reinforcement and identification on exercise behaviors. *Media Psychology, 12*, 1-25.

Hoshi, K., & Waterworth, J. A. (2009). Tangible Presence in Blended Reality Space. *Proceedings of the 12th Annual International Workshop on Presence.* Los Angeles, CA, USA.

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

Morie, J.F. (2008). The Performance of the Self and Its Effect on Presence in Virtual Worlds. *Proceedings of the 11th Annual International Workshop on Presence*, Padova, Italy.

Nass, C., & Lee, K.. (2000). Does Computer-Generated Speech Manifest Personality? An Experimental Test of Similarity-Attraction. *Proceedings of CHI 2000*, The Hague, 329-336.

Nowak, K.L., & Biocca, F. (2003). The effect of agency and anthropomorphism on users' sense of telepresence, copresence, and social presence in virtual environments. *Presence: Teleoperators and Virtual Environments, 12,* 481-494.

Ratan, R. (2010). *Self-Presence, Explicated*. Paper presented at the 60th Annual Conference of the International Communication Association, Singapore.

Ratan, R.A. (2011a). *Self-Presence: Body, Emotion, and Identity Extension into the Virtual Self.* Unpublished doctoral dissertation. University of Southern California.

Ratan, R. (2011b). *Wii Experiment With Self-Presence: When Mii is Me.* Paper presented at the 61st Annual Conference of the International Communication Association, Boston.

Ratan, R.A., & Hasler, B. (2009). Self-Presence Standardized: Introducing the Self-Presence Questionnaire (SPQ). *Proceedings of the 10th Annual International Workshop on Presence.* Los Angeles, USA.

Ratan, R., & Hasler, B.S. (2010). Exploring Self-Presence in Collaborative Virtual Teams. *PsychNology Journal, 8*(1), 11-31.

Ratan, R. A., Santa-Cruz, M., & Vorderer, P. (2007). Multitasking, Presence & SelfPresence on the Wii. *Proceedings of the 10th Annual International Workshop on Presence.*

Slater, M., Spanlang, B., Sanchez-Vives, M. V., & Blanke, O. (2010). First person experience of body transfer in virtual reality. *PLoS ONE, 5*(5).

Yee, N., & Bailenson, J.N. (2009). The difference between being and seeing: The relative contribution of self perception and priming to behavioral changes via digital self-representation. *Media Psychology, 12*(2), 195-209

Appendix A

Self-presence Scale items

To what extent did you feel that… (1= very strongly – 5= not at all)

1. If something happened to the avatar, it was happening to you.
2. The avatar's body was your own body.
3. You were in the avatar's body.
4. The avatar was an extension of you.
5. The avatar was you.

Social presence Scale items

To what extent did you feel that… (1= very strongly – 5= not at all)

1. The audience was present.
2. You were in the same room with the audience.
3. The audience was watching you.
4. The audience was aware of your presence.
5. The audience was real.

Spatial presence Scale items

To what extent did you feel that… (1= very strongly – 5= not at all)

1. You were really inside the virtual classroom.
2. You were surrounded by the virtual classroom.
3. You really visited the virtual classroom.
4. The virtual classroom seemed like the real world.
5. You could reach out and touch the objects in the virtual classroom.

Appendix B

Estimated Marginal Means and Std. Error for Self-presence, Social Presence and Spatial Presence

	E.M. Mean	*Std. Error*
Self-Presence*		
Own face & own voice	3.2	.205
Own face & other's voice	3.65	.213
Other's face & own voice	3.46	.279
Other's face & other's voice	4.07	.301
Social Presence*		
Own face & own voice	2.66	.232
Own face & other's voice	3.15	.241
Other's face & own voice	2.54	.316
Other's face & other's voice	3.3	.341
Spatial Presence*		
Own face & own voice	3.14	.256
Own face & other's voice	2.92	.266
Other's face & own voice	2.71	.348
Other's face & other's voice	3.4	.376

* 5-point Presence scales: 1 = very strong presence – 5 = not at all