

Attention and Social Cognition in Virtual Reality: The effect of engagement mode and character eye-gaze

Brendan ROONEY^a, Katalin BALINT^b, Thomas D. PARSONS^{c*}, Colin BURKE^a, Tess O'LEARY^a, Sharon Chi Tak LEE^a, and Caroline MANTEI^a

^a*University College Dublin, Ireland*

^b*University of Augsburg, Germany*

^c*Computational Neuropsychology and Simulation (CNS) Lab, Department of Psychology, University of North Texas, Denton, Texas USA*

Abstract. Technical developments in virtual humans are manifest in modern character design. Specifically, eye gaze offers a significant aspect of such design. There is need to consider the contribution of participant control of engagement. In the current study, we manipulated participants' engagement with an interactive virtual reality narrative called Coffee without Words. Participants sat over coffee opposite a character in a virtual café, where they waited for their bus to be repaired. We manipulated character eye-contact with the participant. For half the participants in each condition, the character made no eye-contact for the duration of the story. For the other half, the character responded to participant eye-gaze by making and holding eye contact in return. To explore how participant engagement interacted with this manipulation, half the participants in each condition were instructed to appraise their experience as an artefact (i.e., drawing attention to technical features), while the other half were introduced to the fictional character, the narrative, and the setting as though they were real. This study allowed us to explore the contributions of character features (interactivity through eye-gaze) and cognition (attention/engagement) to the participants' perception of realism, feelings of presence, time duration, and the extent to which they engaged with the character and represented their mental states (Theory of Mind). Importantly it does so using a highly controlled yet ecologically valid virtual experience.

Keywords. Virtual Reality; Social Cognition; Virtual Human; Attention

1. Introduction

The current study seeks to examine the contribution of perceptual and conceptual cues on feelings of presence using a virtual environment. The role of perceptual cues will be explored through the manipulation of avatar eye gaze to either give no eye contact or natural semi-regular eye contact to the viewer. Using eye gaze to examine the contribution of perceptual cues on presence is ideal because the avatar's eye gaze behavior can be standardized, meaning that differences between participants can be

* Corresponding Author, thomas.parsons@unt.edu.

attributed entirely to these manipulations. The conceptual influence on presence will be investigated through the provision of instructions that, prior to viewing the virtual reality video, make salient the narrative based aspects or those that encourage attention to the design of the video as an artefact. With the common finding that a user's mental state affects the virtual reality experience [1, 2], the provision of narrative information to the participants should allow them to engage further with the virtual environment. Examining perceptual and conceptual variables together will expand the understanding of how these variables work independently and in combination to affect feelings of presence in the virtual environment. This effect will be further explored by considering the participants perception of time as an index of their feelings of presence[3, 4]. Identifying the conditions under which virtual reality users experience a sense of presence can improve practical applications while yielding valuable insights into human cognition.

Hypothesis one predicts that self-report duration and passage of time judgments will be significantly related to presence as measured by the ITC-SOPI [5]. The second hypothesis predicts that a significant difference will be observed between the four conditions, as created by the use of two conceptual and two perceptual cues, in levels of experienced presence. More specifically, those participants in the third condition who received eye contact as well as the narrative cue will experience the highest degree of self-reported presence. Furthermore, a secondary aim of this research is to utilize participant's experience of time perception

to assess the effectiveness of the ITC-SOPI as a measure of presence. Self-reported time perception is based on the same sense of conscious awareness that can affect levels of presence [6], and exploring time perception and self-reported presence together can contribute to establishing a more reliable measure of presence in virtual environments.

2. Methods

This study used an independent group, experimental design that consists of two independent variables. The first independent variable was eye gaze behavior, which had two levels (eye contact or no eye contact) and the second independent variable was the conceptual cue, which also had two levels (narrative-based or artefact based). Participants were randomly assigned to four conditions (narrative-based conceptual cue with avatar eye contact, narrative-based conceptual cue with no avatar eye contact, artefact-based conceptual cue with avatar eye contact, and artefact-based conceptual cue with no avatar eye contact). Self-report measures of time perception and presence served as the dependent variables. Time perception included duration judgments, accuracy scores and passage of time judgments. Presence was measured using the Spatial Presence, Engagement, Ecological Validity and Negative Effects subscale of the ITC-SOPI [5].

2.1 Participants

The University's Institutional Review Board approved the study. Twenty participants, from a university campus (6 males, 14 females, age range 18-30 years, mean age = 22 years, SD = 8.4 years) volunteered for this study. Participants were randomly allocated to one of the four conditions while maintaining a gender balance.

2.2 Measures and Procedure

The sample video was run on an ASUS computer using the Oculus program and the Oculus Rift headset and the sound was controlled on a One Concept NHT 1051 speaker system. Participants viewed the test video through the Oculus Rift headset, which was presented in a pre-existing virtual environment through the program “Coffee Without Words”.



Figure 1. Screenshot of female avatar during “Coffee Without Words” video in the eye contact condition.



Figure 2. Screenshot of male avatar during “Coffee Without Words” video in the eye contact condition

The video was programmed to differ depending on the condition that participants were assigned to. Participants in conditions 1 and 2 received no eye contact from the avatar,

and the video was programmed so that the avatar “should ignore player” during the entire duration. Participants in conditions 3 and 4 received contact from the avatar, which consisted of full gazes that lasted between 7 and 13 seconds and short gazes that lasted between 1 and 1.5 seconds. The neutral period in between gazes was programmed to be between 10 and 20 seconds. Participants were gender matched to the avatar they viewed in the video, which can be seen in Figures 1 and 2. During the video the user was seated at a table in a coffee shop across from the avatar, no other action took place.

Self-reported time perception: Self-reported perception of time was measured using two retrospective questions immediately after participants watched the virtual reality test video, “Coffee Without Words”: 1) “Estimate how long in minutes and seconds that the video lasted for” was a retrospective duration judgment; and 2) “At what rate did time seem to pass in the virtual environment compared to normal?” was a retrospective passage of time judgment that required participants to indicate, on a 5-point scale (1-5), their experience of time passage during the video.

Presence: The participant’s feelings of presence in the virtual environment were measured using an adapted version of The ITC-Sense of Presence Inventory (ITC-SOPI)[5].

Validity of conceptual cues: To determine whether the conceptual cues, either narrative-based or artefact-based, were having the desired effect on the user’s experience of the mediated environment, participants were asked if during the virtual session they were “mostly thinking about the fictional story of this café scenario” or were “mostly thinking about how this café and the character were designed”. Participants indicated their opinion of both of these statements on a seven-point Likert scale (1-7).

Controlling familiarity: The current study controlled for the potentially confounding familiarity of the avatars used in the “Coffee Without Words” video by asking participants “did the character in the virtual environment look like anyone you know” and if “Yes” was selected they were asked to “please tell us who the character looked like”. This allowed the researcher to consider the effect of the familiarity of the character when evaluating participant’s responses. Four participants indicated that the character in the virtual environment reminded them of someone they knew (3 in the eye contact condition and 1 in the no-eye-contact condition).

2.3 Results

Failing to support hypothesis one, no significant correlation was observed between any time perception and presence variables. Thus, the perceived experience of time for participants during the virtual reality video was not significantly related to their experience of presence within the mediated environment.

No significant group differences were observed in the extent to which participants reported attending to the video as a narrative, $t(43) = -.661, p > .05$, or an artefact, $t(43) = -.216, p > .05$, indicating that the instruction manipulation was unsuccessful. For this reason, the groups based on conceptual cue were collapsed and group differences in perceptual cue (eye-contact) were explored.

No significant difference was observed between the eye-contact conditions in terms of their scores on the Spatial Presence subscale, $F(1,49) = .12, p > .05$. This means that differences in scores on the Spatial Presence subscale did not arise from avatar eye gaze behavior. Scores on the Engagement subscale were also not found to differ significantly due to perceptual cues, $F(1,47) = 3.49, p > .05$.

Results revealed a significant group effect of perceptual cue (avatar eye-contact) on participants' perceived ecological validity of the video, $F(1,49) = 5.54, p < .05$, that is, the degree to which the experienced the video as perceptually natural or representative of the real world. Results revealed that participants in the condition with no eye-contact reported that the video was higher in terms of ecological validity ($M = 4.58$, on a 7-point scale) compared to their eye-contact counterparts ($M = 3.88$ on a 7-point scale). Thus participants in the condition with no eye contact rated the video as more natural. Those participants also reported significantly higher levels of negative effects (e.g. headache, dizziness) from the virtual reality ($M = 2.64$ on a 7-point scale), when compared to those in the eye-contact condition ($M = 2.01$), $F(1,49) = 8.23, p < .01$. Although it is notable that these scores are quite low across the conditions.

3. Discussion

The contribution of perceptual and conceptual cues on feelings of presence within a virtual environment was examined using measures of time perception and self-reported feelings of presence. Results revealed no significant relationship between any time perception and presence variables. Thus, the perceived experience of time for participants during the virtual reality video was not significantly related to their experience of presence within the mediated environment. Also, a negative correlation was observed between passage of time and duration and accuracy scores, suggesting that duration and passage of time judgments reflect different aspects of the experience of time.

Perceptual cues were manipulated by programming gender-matched avatars in the virtual reality video, "Coffee Without Words", to give either semi-regular eye contact or no contact at all to the user. Conceptual cues were manipulated by providing either narrative based information through a relevant background story, or artefact based information relating to the design of the virtual environment, to participants prior to viewing the video. Conceptual cue manipulation was deemed unsuccessful by findings that participants in each group did not differ in terms of how much they attended to narrative or artefact aspects of the video.

Following this, conceptual conditions were collapsed and analyses focused on comparisons between participants who differed in terms of the perceptual cue. In other words, we compared groups of participants who experienced no eye-contact from the avatar with those who experienced natural semi-regular eye contact. Here we observed no effect of perceptual cues on reported spatial presence or engagement. It is possible that the conceptual and perceptual manipulations were not strong enough to shift participants' experience of an already highly immersive virtual experience.

However, participants who experience no eye-contact perceived the video to be more natural than those who experienced eye-contact. This finding is not in line with our predictions based on previous work [7, 8, 9]. Cautiously we suggest that participants may

have rated the lack of eye-contact as more natural in a situation where they sat opposite a stranger they did not know. In this situation perhaps participants would not expect eye contact and so reported the lack of eye-contact as more natural.

In summary, this study observed no effect of manipulating the conceptual context of a virtual reality experience on participants' time perception nor on their self-reported feelings of presence (spatial presence, engagement or ecological validity). Further we report here that manipulating perceptual cues in the form of avatar eye-contact effects participant's perceived realism of the virtual experience, but not their spatial presence nor their engagement. While eye-contact in this case reduced perceived realism, regardless of the direction of the observed effects, our findings demonstrate that perceptual cues are more powerful than conceptual cues in shaping participants' experience of perceived realism in virtual experiences.

References

- [1] M. Avram, D. Opris, and D. David, The Effects of Expectancies on Presence in Virtual Environments: A Brief Research Report. *Transylvanian Journal of Psychology*, **15** (2014) 21–31.
- [2] D. Nunez, and E. Blake, Conceptual Priming as a Determinant of Presence in Virtual Environments, *ACM Afrigraph*. **1** (2003) 101–108.
- [3] T. Sanders, and P. Cairns, Time perception, immersion and music in videogames, *BCS '10 Proceedings BCS HCI* (2010), 160–167.
- [4] J.H. Wearden, and H. Lejeune, Scalar properties in human timing: conformity and violations, *Q. J. Exp. Psychol.* **61** (2008), 569–87.
- [5] J. Lessiter, J. Freeman, E. Keogh, and J. Davidoff, A Cross-Media Presence Questionnaire: The ITC-Sense of Presence Inventory, *Presence*. **10** (2001) 282–297.
- [6] D. Zakay, and R.A. Block, Prospective and retrospective duration judgments: an executive-control perspective, *Acta Neurobiologiae Experimentalis*. **64** (2004) 319–328.
- [7] N. Yee, J.N. Bailenson, M. Urbanek, F. Chang, and D. Merget, The unbearable likeness of being digital: The persistence of nonverbal social norms in online virtual environments, *CyberPsychology & Behavior*. **10** (2007) 115–121.
- [8] J.N. Bailenson, A.C. Beall, J. Loomis, J. Blascovich, and M. Turk, Transformed Social Interaction, Augmented Gaze, and Social Influence in Immersive Virtual Environments, *Human Communication Research*. **31** (2005) 511–537.
- [9] S.Y. Oh, J. Bailenson, N. Krämer, B. Li, M. Cuddon, N. Goharpey, J. Maurer, S. Rosini, A. Tsekouras, and D. Grace, Let the Avatar Brighten Your Smile: Effects of Enhancing Facial Expressions in Virtual Environments, *PLoS One*. **11** (2016) 1–18.