The Belated Guest: Exploring the Design Space for Transforming Asynchronous Social Interactions in Virtual Reality

Portia Wang*
Department of Management Science and Engineering
Stanford University

Mark R. Miller†
Department of Computer Science
Stanford University

Jeremy N. Bailenson‡
Department of Communication
Stanford University

Abstract

Social meetings in Virtual Reality (VR) are fundamentally different from videoconferencing because the VR tracking data can be used to render scenes as if they were in real-time, enabling people to go back in time and experience discussions that they may have not attended. Moreover, the tracking data allows people to visit the meeting from any location in the room, as opposed to the single camera angle used for video conferences. The current paper describes the methodology of transforming tracking data around proxemics and head orientations of recorded avatars to nonverbally assimilate new users into a recorded scene.

Index Terms: Human-centered computing—Human-centered computing (HCI)—Interaction paradigms—Virtual reality;

1 Introduction and Prior Work

The proliferation of social communication platforms highlights the appeal of enabling nonverbal social interactions [6]. Many have underscored the importance of addressing proxemics, people's use of space [7, 8], and eye gaze when constructing synchronous virtual social gatherings [2, 9]. Meanwhile, a typical consumer currently records and reconstructs social events through videos. To that end, researchers have also explored manipulations of avatar appearance of pre-recorded single-person videos [4].

While there have been a handful of research projects which transform gaze and distance between avatars or agents [1, 5], the opportunity to apply such algorithms to recorded VR situations is unique given the rise of social VR, and the ability to record the tracking data from any collaborative discussion. Consequently, there remain untapped opportunities for augmenting and transforming pre-recorded meetings for asynchronous participation.

To address this, we present an exploratory work on the design space of asynchronous social interactions, where a participant joins a pre-recorded group discussion in VR. By systematically analyzing gaze and proxemic patterns of small group discussion recordings, we transform these recordings to use as renderings for individuals to nonverbally "join" in the future. We first used previous theory and findings to determine how one might transform nonverbal behavior, and then used a bottom-up analysis of a large-scale social VR dataset to instantiate those parameters.

2 Recordings of Social VR Meetings

We studied recordings of social VR discussions in the ENGAGE social virtual reality platform, where 137 university students engaged in a medium-group sized discussion with 5 to 8 people while wearing Meta Oculus Quest 2 headsets and holding hand-held controllers [3]. We recorded the positions and orientations of each participant's head and hands. This large-scale dataset enabled us to extrapolate typical gaze and proxemics insights, which allowed us to further design and implement transformations on any recording. Since eye gaze direction is not adjustable in ENGAGE, transformations on eye gaze direction were performed on head orientation (yaw and pitch). Figure 1 shows an example transformation.

*e-mail: portiaw@stanford.edu
†e-mail: mrmillr@stanford.edu
‡e-mail: bailenso@stanford.edu

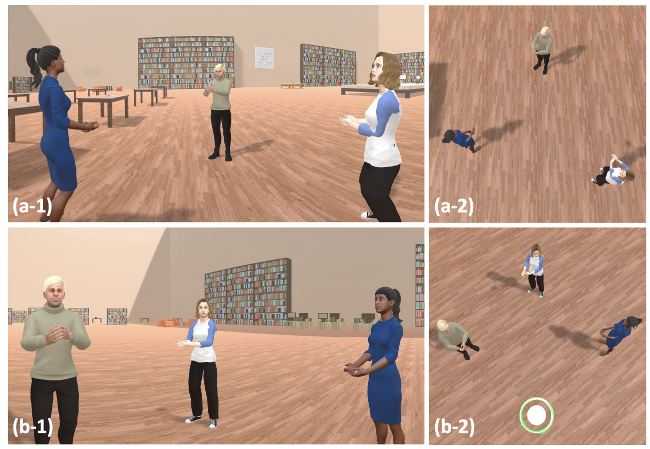

Figure 1: Our transformations of asynchronous VR social interactions for a three-person discussion. (a-1) User view of the original scene. (a-2) Top-down view of the original scene. (b-1) User view of the scene after proxemics and gaze transformations. (b-2) Top-down view of the scene after proxemics and gaze transformations, where the green circle indicates the position of the user.

3 Design Factors

3.1 Proxemics Manipulation

In order to spatially include the *belated guest* while maintaining appropriate social distances between all participants, we manipulated proxemics. Furthermore, seeing that users typically gathered around a circle for each recorded discussion, we implemented a procedure that places the user as part of an enclosed circle. To do this, we begin by picking an arbitrary belated guest position, "the gap", in close proximity to the pre-recorded group discussion. Following this, we perform a gradient-based optimization scheme such that the distance between each pair of users (including the guest) is iterated to approach a parameterized social distance d_{social}. As this procedure transforms user positions into points evenly spaced on a circle with a diameter roughly matching d_{social}, the choice of d_{social} depends on group size, with larger groups using larger d_{social}s as to maintain appropriate interpersonal distances for social interactions.

3.2 Gaze Manipulation

3.2.1 Remapping Gaze after Proxemics Manipulation

Changes in participant positions following proxemics manipulations could distort relative gaze directions. To correct distorted gaze directions, we implemented a gaze remapping procedure. Figure 2 shows a schematic example of this procedure. We first define the physical boundary by connecting lines emitted from each participant's position that is perpendicular to that participant's position and the center of the group (red dotted lines in Figure 2a). We then emit a ray (dotted rays in Figure 2a) from each participant (dots in Figure 2a) in the direction of the gaze and record the position of the collision point between the ray and the boundary, which we refer to as the original gaze targets (xs in Figure 2a). We proceed to calculate the transformed gaze targets (xs in Figure 2c) using a triangulation-based cubic interpolation transformation derived from the mapping between the original positions (small dots in Figure 2b) and the transformed positions (large dots in Figure 2b, dots in Figure 2c) outlined in Section 3.1. Finally, we derive the updated head orientations (yaw) from the new gaze targets.

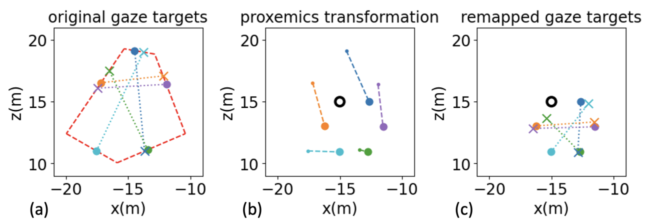

Figure 2: Gaze Remapping of one frame for a five-person discussion. The black circles in (b) and (c) show the position of the new user.

3.2.2 Inserting Gaze towards the Uninvited Guest

We consider two questions for inserting gazes toward the guest: how long should each gaze be, and how fast should the head turn prior to and after each inserted gaze? To answer these questions, we examined the original tracking data. An analysis of gaze instances from unaltered recordings revealed a large variation of gaze durations and a wide range of head rotation speeds for frames prior to and following each gaze. Further, we found that the duration of gazes from one avatar to another varied depending on the distance between the dyads as well as their roles in the conversation (speaker vs. non-speaker). Thus, to insert realistic gazes, we leverage the rich gaze information in each recording by sampling existing gaze instances, and biased sampling towards gazes occurring at similar distances and conversation roles.

More concretely, we start by extracting each gaze instance from a reference recording, noting its gaze duration, the roles of the person looking and the person being looked at (speaker vs. non-speaker), and the distance between the dyad. Then, for a group of size n, we select evenly spaced timestamps for inserting gazes such that a gaze is assigned to each user every n inserts, where the order of users within each block of n inserts is further randomized. For each insert, a reference gaze instance is randomly selected from a set of k samples with the most similar distances with the same pair of conversation roles (non-speaker towards non-speaker, speaker towards non-speaker). The gaze is then added by first orienting the avatar's head towards the gap position for the same duration as the reference instance. We finally interpolate frames prior to and after the added gaze using the same head orientation speed as the reference instance until the transformed orientation converges with the original orientation.

We quantify the amount of added attention for inserting x gazes for every user, $r_{attention}$, as the average ratio of time for added fixed gaze to the total amount of transformable time across all participants.

4 Implementation

We transformed group discussions using the proposed set of techniques by directly altering the recordings of VR meetings described in Section 2. Each participant's avatar position in the horizontal plane is manipulated following the proxemics manipulation detailed in Section 3.1. To ensure that the new user is standing at the desired location, we placed an invisible chair at "the gap" to anchor the user. While the current work does not feature formal user testing, the authors examined about ten of the 137 conversations across various group sizes in the belated guest conditions and iterated on the algorithm design and choice of parameters between the ten sessions.

5 Conclusions and Future Work

VR allows people to attend previous meetings anywhere in the virtual meeting space. For this reason, we see VR as a powerful tool for not only reconstructing social interactions but also providing experiences that are uniquely personalized and augmented from what had actually occurred. In this ongoing project, we reimagine how users can experience social VR meetings they had not attended previously by exploring the design space for transforming proxemics and gaze. While this work scratches the surface of what is possible of asynchronous VR interactions, much research is needed to understand and quantify the potential benefits of such augmentations.

Acknowledgments

We thank Cyan DeVeaux, Eugy Han, and Hanseul Jun for collecting the large-scale dataset of VR recordings used for our analysis.

References

[1] 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(4):511–537, 01 2006. doi: 10.1111/j.1468-2958.2005.tb00881.x

[2] J. N. Bailenson, J. Blascovich, A. C. Beall, and J. M. Loomis. Equilibrium theory revisited: Mutual gaze and personal space in virtual environments. *Presence*, 10(6):583–598, 2001. doi: 10.1162/105474601753272844

[3] E. Han, M. Miller, C. DeVeaux, H. Jun, K. Nowak, J. Hancock, N. Ram, and J. Bailenson. People, Places, and Time: A Large-scale, Longitudinal study of Transformed Avatars and Environmental Context in Group Interaction in the Metaverse. *Journal of Computer-Mediated Communication*, 28(1), 1 2023.

[4] T. T. Nguyen, Q. V. H. Nguyen, D. T. Nguyen, D. T. Nguyen, T. Huynh-The, S. Nahavandi, T. T. Nguyen, Q.-V. Pham, and C. M. Nguyen. Deep learning for deepfakes creation and detection: A survey. *Computer Vision and Image Understanding*, 223:103525, 2022. doi: 10.1016/j.cviu.2022.103525

[5] D. Roth, G. Bente, P. Kullmann, D. Mal, C. F. Purps, K. Vogeley, and M. E. Latoschik. Technologies for social augmentations in user-embodied virtual reality. In *25th ACM Symposium on Virtual Reality Software and Technology*, VRST '19. Association for Computing Machinery, New York, NY, USA, 2019. doi: 10.1145/3359996.3364269

[6] M. Slater, A. Sadagic, M. Usoh, and R. Schroeder. Small-group behavior in a virtual and real environment: A comparative study. *Presence: Teleoper. Virtual Environ.*, 9(1):37–51, feb 2000. doi: 10.1162/105474600566600

[7] K. Takahashi, Y. Inoue, and M. Kitazaki. Interpersonal distance to a speaking avatar: Loudness matters irrespective of contents. In *2022 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)*, pp. 774–775, 2022. doi: 10.1109/VRW55335.2022.00238

[8] J. Williamson, J. Li, V. Vinayagamoorthy, D. A. Shamma, and P. Cesar. Proxemics and social interactions in an instrumented virtual reality workshop. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, CHI '21. Association for Computing Machinery, New York, NY, USA, 2021. doi: 10.1145/3411764.3445729

[9] L. Yoon, D. Yang, J. Kim, C. Chung, and S.-H. Lee. Placement retargeting of virtual avatars to dissimilar indoor environments. *IEEE Transactions on Visualization and Computer Graphics*, PP:1–1, 08 2020. doi: 10.1109/TVCG.2020.3018458