VR-All Presence: A system for real-time asymmetric interactions in VR and physical space enhanced with eye-tracking

Elnaz Tafrihi Bailey *
Berkeley, USA
elnaz_tafrihi@berkeley.edu

Jiexin Lyu *
Berkeley, USA
jessie_lyu@berkeley.edu

Yejun Wu *
Berkeley, USA
wuyejun@berkeley.edu

Keebaik Sim *
Berkeley, USA
keebaik_sim@berkeley.edu

Niloufar Salehi
Berkeley, USA
nsalehi@berkeley.edu

ABSTRACT

Virtual reality (VR) has made communication and collaboration easier by providing an opportunity for us to meet remotely in a fully immersive environment. However, VR technology still has serious limitations regarding connecting people emotionally in the physical and virtual space in collaborative settings. By researching various VR applications that are designed for asymmetric interactions, we discovered that people who are participating through VR often become a passive participant among people who are gathering in the physical space due to two factors: 1) Lack of physical interaction in the space. 2) Inability to provide eye contact between the two user groups. In response to these issues, we designed an interactive device equipped with eye tracking technology and connected to the VR space using Google Firebase to signify the physical presence of the VR participant. We conducted two experiments using this device, word games and a mock up online interview to test the user experience. Due to the COVID-19 challenges, our user study group was limited, and therefore we were not able to test users with different levels of familiarity towards VR which could have helped with more accurate findings. Our user study gave us insight on how to improve the user interface of a smartphone application in the physical space as well as avatar options both in the VR and physical space.

Author Keywords

Virtual Reality; Telecommunication; Video Conference; Eye Gazing; Interpersonal Communication; Remote Collaboration; Live Streams

*All four authors contributed equally to this project.

ACM ISBN 978-1-4503-2138-9.
DOI: 10.1145/1235

PROBLEM

Participating remotely in collaborative events such as conferences or parties has inevitable interaction gaps. Telepresence has always been a popular topic in the Human-Computer Interaction field. Emerging technologies such as virtual reality and augmented reality could open up new possibilities for telepresence in a collaborative event.

Nonetheless, collaborative immersive environments can cause discrepancies in terms of the experience among users in VR and people in the physical space. In our research, we are interested in understanding these discrepancies in asymmetric collaborations, and came up with potential strategies to mitigate that. What possible interactions could we design for people wearing VR goggles when they are participating in collaborative events?

INSIGHT AND THEORETICAL CONTRIBUTION

VR telepresence is a very recent research area in the VR/AR domain. For this project, we are trying to understand what possible interactions could be created between people in VR headsets and people in physical space to reduce the experience discrepancies in a collaborative environment. For this research project, we are trying to make the following unique contributions:

- Contribution 1: research on VR experience for various types of collaborative events, and find interactions that could be enhanced by eye-contacting.
- Contribution 2: creating new collaborative VR experience among users in the physical space vs people in VR (VR telepresence), with focus on eye-gazing interaction.
- Contribution 3: (potential) improving emotional connection and physical interaction of users in VR and physical space through our system.

PREVIOUS WORK

Enhancing users' experiences in asymmetric collaborative environments have been studied from different perspectives in the field of HCI. These collaborations include displaying digital content that can be manipulated simultaneously by all users, or allowing multiple user games among HMD and non HMD users in the same physical space, and combining tangible user interfaces in such collaborations [7]. As an example, ShareVR enables co-located experiences among HMD and non-HMD users to enjoy and interact with the same digital content [4]. The authors ran a user study comparing ShareVR to a baseline condition showing how the interaction using ShareVR led to an increase of enjoyment, presence and social interaction which was an inspiration for our project to create enjoyment in asymmetric experiences. Another example of shared digital content is Black Hat Cooperative [8], which is a two player VR game that users in asymmetric conditions work together, with one player having access to the VR person's point of view. Furthermore, according to Jeronimo Gustavo Grandi et. al's research in "Characterizing Asymmetric Collaborative Interactions in Virtual and Augmented Realities", users in the asymmetric collaborative environments (CVEs) had a better performance compared to users both in AR or VR [3].

DESIGN PROCESS

Problem Finding

In order to understand people's current virtual meeting experience, we conducted a quantitative survey through Google Form. 47 people with previous virtual meeting experience responded to our survey. According to the survey result, we found that Zoom (63.8%), Webex (51.1%) and Google Hangout (23.4%) were the three most popular virtual meeting platforms. Our participants used those platforms mainly for professional meetings (78.7%) and academic meetings (48.9%). On average, over half (53.2%) of the participants spent 2-3 hours on those platforms every day. On average, about half (44.7%) of the virtual meetings were small scale meetings with 1-5 people.

Although participants were generally satisfied (61.7% rated their usual platforms as 4 and 5 out of a score of 5) with their usual virtual meeting platforms, we were able to recognize a few universal key factors that led to people's unsatisfactory virtual meetings. Besides technical and connectivity issues that were out of the scope of our research, lack of social cues and engagement, for example, gestures, facial expressions and emotional reactions, was recognized as the biggest challenge in virtual meetings currently. Similarly, sufficient social interaction and engagement led to some of the most satisfactory virtual meetings. As a result, we narrowed down our research scope and decided to focus on how to effectively integrate and highlight social cues to improve virtual meeting experience.

Solution

Our solution is to create a system that engages participants in physical space and VR together through eye-gazing. We are especially interested in improving the experience of asymmetrical collaborative events when participants are joining either in-person or virtually.

For participants joining in-person in physical space, we are interested in knowing to make them feel more engaged with the users joining virtually with eye-gazing technology. We designed a physical stand that could rotate based on the direction of the VR participant eye movement. At the same time, the stand is equipped with a sound sensor and camera to capture activities in the physical space. The sound sensor will detect the person speaking, and the camera will detect the physical participants' eye movements. These data from physical space will be sent to the VR environment to create simulations of activities in the real world.

In order to render the representation of VR participant in physical space, we design an mobile application which could be mounted on the physical stand top. The application is rendering the VR participant's activities as an avatar.

In VR space, we designed a virtual meeting environment with all participants shown as avatars. The data sent from the physical stand will be read inside the VR application and create activities simulation on those avatars. At the same time, the VR application is sending data of VR participant's activities to the physical stand and the mobile application.

APPLICATION DEVELOPMENT

The Physical Stand

We propose a new product that will help connect users in VR and in physical space through eye-tracking technology and help them engage better. Currently the physical prototype is composed of three components: 1) The bottom piece that includes a Raspberry Pi and sound sensors. 2) The middle piece equipped with a servo motor. 3) The top piece that includes a Raspberry Pi camera module V2 and a smartphone mount. (see Figure 1)

Voice Detection & Eye Tracking Development

For the initial prototype, we used sound sensors to detect voices of users in the physical space, and initialize the eye tracking. For eye tracking in the physical space we used the Raspberry Pi module 2 camera, and OpenCV library to detect faces and eyes in the physical space(see Figure 2). In the next step, we used CLAHE (Contrast Limited Adaptive Histogram Equalization) in OpenCV to increase the contrast of the video, and made the video black and white (see Figure 3). Then by analyzing the data that was extracted from detected eyes in the video, we set rules to identify left eye and right eye based on their corner points location compared to the face coordinates.

In order to eliminate any false positives, we added additional rules in terms of length dependencies between the two eyes, since we had moments where other parts of the face with darker features were identified as eyes, so our method helped us eliminate any wrong eyes detected. In the next step, we used the Threshold binary, inverted in OpenCV to find the placement of the pupil by eliminating any pixels that did not match our defined value [1]. We set up three rules based on placement of the pupil compared to the width of the detected eyes, to find if the user is looking straight, left or right.

User's eye location is then sent from Raspberry Pi to Firebase database, using Firebase library. User's eye movement can then be retrieved by Unity, which updates the animation associated with the eye gazing of avatars in VR in real-time. The Raspberry Pi also retrieves VR user's eye gazing from Firebase database to update the rotation degree of the servo motor to match the head movements of the user in VR, and to

Figure 1. Final prototype of the physical product

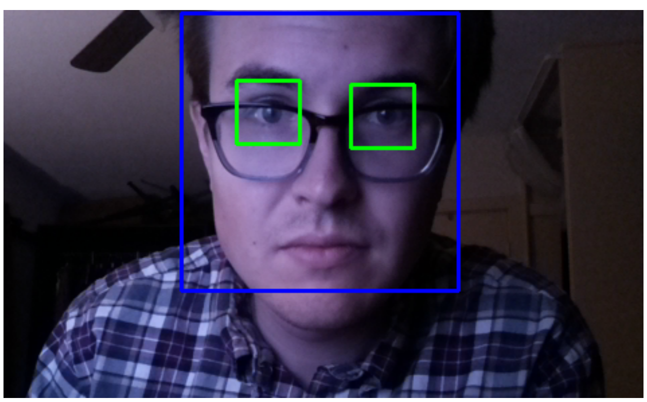

Figure 2. Face and eye tracking using OpenCV

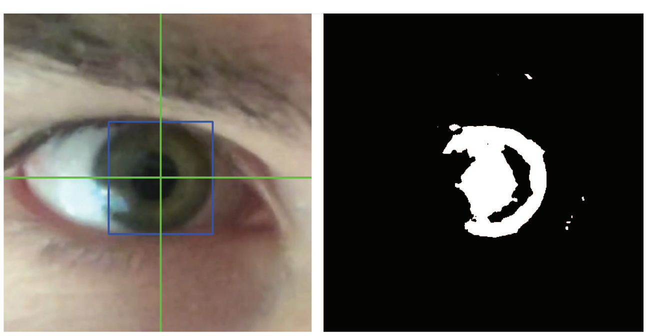

Figure 3. Eye pupil detection using threshold binary inverted method

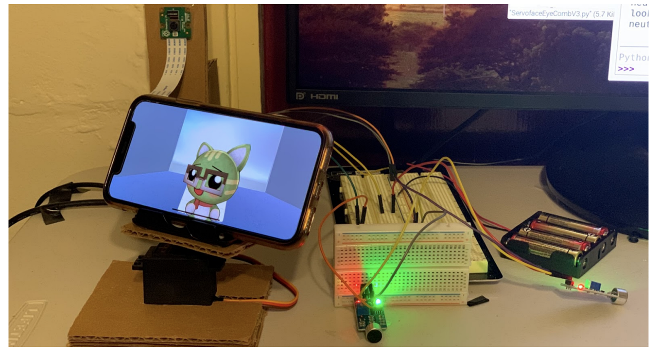

Figure 4. Early stage physical prototype, servo rotation synced with eye movement

create more engagement among asymmetric users. We also developed an application on the smartphone placed on our product to display the avatar of the user in VR, that was built in Unity using iOS settings for an iPhone device (see Figure 4).

VIRTUAL SPACE DEVELOPMENT

VR Eyegazing Implementation

We followed the steps from immersivelimit.com[5] to implement the eye-tracking mechanism in VR. The eye-gazing mechanism is attached to the main camera in Unity. In a VR setting, the main camera represents users' eyes. By incorporating eye gazing tracking logic in the main camera we could detect if a user is looking at an object, which in our case represents users in physical space.

Cartoon Animal Models VS Realistic Human Models

Initially we designed a virtual work environment with realistic human models using Unity game engine to represent the physical space meeting participants.The animated characters used Mixamo animations that are created using motion capture technology. There are several possibilities in extracting facial expressions in real time using tools such as Mixamo Face Plus, that could lead to a more realistic virtual environment [2].

However, we were interested in Jaron Lanier's "uncanny valley[6]" theory, that is introduced in his book called "Down of the New Everything: Encounters with Reality and Virtual Reality." In this book he stated that "if the world you are wondering in is not real, that should be made obvious" [Lanier, 2017]. Lanier specifically mentioned the design of 3D human models in VR and wrote "when an avatar is just slightly off, then the brain panics"[2017]. In our research, we are inspired by Lanier's theory about uncanny valley in VR. After trying different human 3D models, we found it's impossible to make the models "acting" like a real person. And the facial expressions and gestures were not in sync with reality, even creepy. So we decided to go with the other extreme of 3D modeling in VR – if we can't make things real, we can just make things very unreal so people know that they are in a virtual world.

We implemented animal avatars asset called "Yippy Kawaii[9]" from Unity Asset Store since we discovered this asset package could fulfill most of the animations we required for this

Figure 5. Final Scene of VR Meeting Space

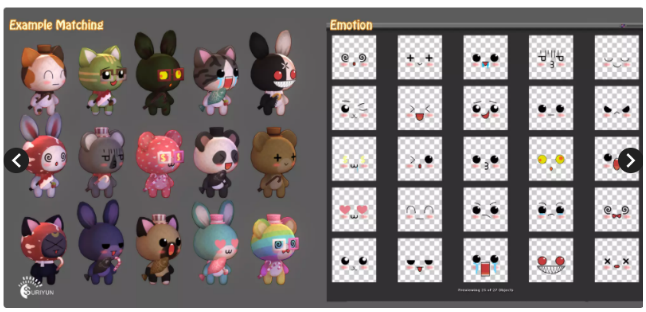

Figure 6. Yippy Kawaii Unity Asset

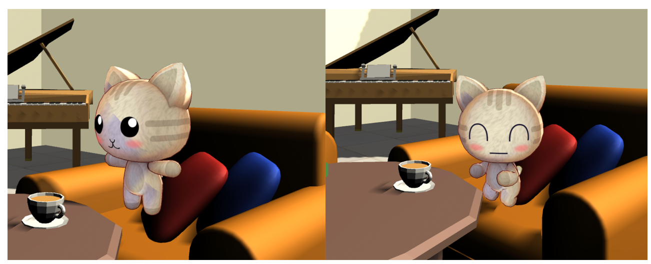

Figure 7. Two states of animations we created for meeting participants. Left: neutral; Right: smile and rotate head towards the VR person

prototype, such as head rotation and facial expression change (see Figure 6).

Animations

For the initial version of this system prototype, we decided to implement only two animations: head rotation and facial expression change (see Figure 7). In typical verbal communication, the most common interactions are facial expression changes (speaking, smiling, etc.), and upper body gestures (head rotation, hand gestures, etc.). Since we downloaded the asset from Unity Web Store, we were not able to customize the animations as detailed as our specific needs. For this version, we decided to change facial expression from neutral to smile, and change head rotation to indicate the avatar is responding to verbal communication.

CONNECTING VR AND PHYSICAL SPACE

In order to connect VR and physical space, we used Google Firebase as the central hub to store all the eye movement data from both VR participants and physical participants (see Figure 8).

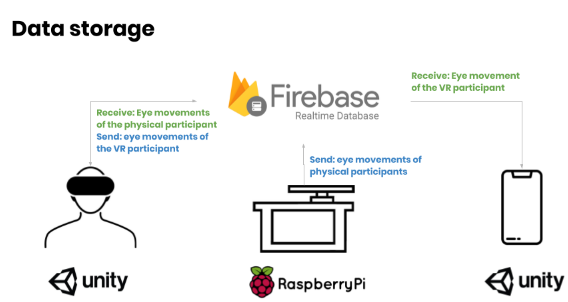

Figure 8. Connecting VR and Physical Space with Firebase

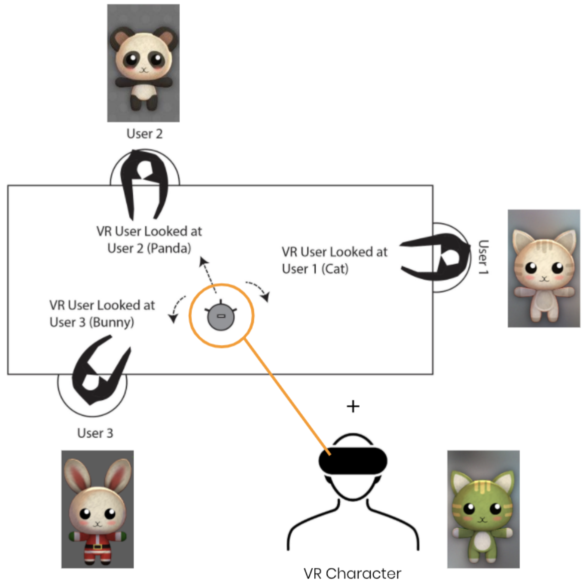

Figure 9. User Testing of physical meeting setting diagram

Figure 10. Top: Participants in the physical meeting space; bottom: participant in the VR meeting space

TESTING

As we managed to connect the eye-gazing part from the physical meeting with that of the VR space through Firebase, we recruited four participants to conduct a user testing. Three people were asked to meet in person, with one person joining their meeting remotely by VR. We set up the physical meeting space similar to the VR scene, with three people sitting around a table and the physical product set in the middle to detect their voices and eye-gazing. The participant joining through VR was represented as an animated avatar in the phone app, which would turn to a particular participant when he was looking at and talking to him (see Figure 9).

We divided the user testing into two different experiments. The first one was a more formal interview-like QA session. As the interviewer, the participant in VR called a participant in each round and asked him a question. The second experiment was a more casual verbal game. Participants were asked to say a food in a particular order, with the first character of the word as the last character of the word of the previous participant. Each experiment went on for around 10 minutes. We recorded the VR scene with our cell phone and the physical meeting with the 360 camera for our research group only. Both groups of participants did not know what each other would see during the meeting (see Figure 10).

After the user testing, we conducted a discussion-based post-test interview with both groups of participants. We were interested in their overall meeting experience and their feedback on our product. Specifically, we asked about the difference of experience between a regular VR conference with our product in terms of engagement, interaction, enjoyment, focus and participation. We were also interested in the difference of experience between using our product in a gaming context versus a more formal interview context. Lastly, we inquired about the difference of experience between meeting with live streaming of participants and cartoon animal avatars to represent the physical participants.

LIMITATION AND FUTURE WORK

After testing VR-All Presence, we gained feedback from the participants who joined on-site and the participants who engaged through VR. Due to the COVID-19 situation, we were unable to conduct more in-person user testing for our project as we initially planned. Although the feedback received from the initial testing might be biased due to the small sample size, we were able to gain some interesting feedback to guide our future work.

First of all, both groups of participants said that using animal characters as avatars were confusing. The participant joining through VR answered that it was difficult for him to identify which animal avatar was speaking and who was actually in the meeting. Also, the participants in the physical space mentioned that the VR participant's animal avatar had not delivered the person's facial expression which made interpersonal communication difficult. For the future testing, on the one hand, we will try another round of testings using animal characters with more elaborate physical cues such as moving mouth, arms, and nodding head. Besides, we will test with real human-like 3D rendered avatars, still images of actual participants, or even live stream video of an event in order to find the ideal tool for evoking asymmetry communication. On the other hand, we also consider conducting tests using other technologies such as Varjo's XR-1 mixed reality device.

Furthermore, our testing hypothesized a setting in which one person in VR engages in a meeting of multiple people. What if more than one person participates remotely? Or, what if more people join the meeting from various places and fewer people are physically present? We need to conduct future usability testing to answer these questions. The answers may lead to redesigning not only the virtual space but also the actual objects in physical space such as a desk which might need to be designed in a specific form to stream multiple participants, a detailed guide about where participants in actual space need to sit or a real time detection of participants locations that will be translated into VR, or where VR-All Presence physical product should be placed.

CONCLUSION

Remote communication using traditional digital technology results in the emotional distance among the participants. In order to reduce the emotional gap and overcome the limitations that exist in current remote communication technology, we designed a system using VR and a device signifying a remote participant among people who are meeting in-person. In order to increase emotional empathy and provide remote participant's presence, we implemented an eye-gazing mechanism both in VR and in physical space. We found that eye-gazing could be a useful add-on to virtual collaboration in asymmetric VR-real space environments. Through our test results, we are offering various possibilities and directions on how scholars and academics in the HCI field can develop enriched real-time asymmetric interactions in VR and physical space in the future. As we become more global and given the situation of COVID-19 that most of us have to work remotely, this field will become a promising research area.

ACKNOWLEDGMENTS

For this project, we want to thank our Professor Niloufar Salehi, who gave us a lot of advice in defining our research scope and provided us with resources for VR research. We also want to thank our families and friends for being our usability testing participants during the unprecedented COVID-19 time.

REFERENCES

[1] Sergio Canu. 2019. Eye Motion Tracking - Opencv with Python. (2019). https://pysource.com/2019/01/04/eye-motion-tracking-opencv-with-python/ Library Catalog: pysource.com.

[2] Emiliano Gambaretto and Charles Piña. 2014. Real-Time Animation of Cartoon Character Faces. In *ACM SIGGRAPH 2014 Computer Animation Festival (SIGGRAPH '14)*. Association for Computing Machinery, New York, NY, USA, 1. DOI: http://dx.doi.org/10.1145/2633956.2658830

[3] Jeronimo Gustavo Grandi, Henrique Galvan Debarba, and Anderson Maciel. 2019. Characterizing Asymmetric Collaborative Interactions in Virtual and Augmented Realities. In *2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*. IEEE, Osaka, Japan, 127–135. DOI: http://dx.doi.org/10.1109/VR.2019.8798080

[4] Jan Gugenheimer, Evgeny Stemasov, Julian Frommel, and Enrico Rukzio. 2017. ShareVR: Enabling Co-Located Experiences for Virtual Reality between HMD and Non-HMD Users. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*. ACM, Denver Colorado USA, 4021–4033. DOI: http://dx.doi.org/10.1145/3025453.3025683

[5] Adam Kelly. 2015. Detect When the Player is Looking at an Object. (2015). https://www.immersivelimit.com/detect-when-looking-at-object Library Catalog: www.immersivelimit.com.

[6] Jaron Lanier. 2017. *Dawn of the New Everything: Encounters with Reality and Virtual Reality*. Henry Holt and Company. Google-Books-ID: 8MCuDgAAQBAJ.

[7] Austin Lee, Hiroshi Chigira, Sheng Kai Tang, Kojo Acquah, and Hiroshi Ishii. AnnoScape: remote collaborative review using live video overlay in shared 3D virtual workspace. In *Proceedings of the 2nd ACM symposium on Spatial user interaction - SUI '14* (2014). ACM Press, 26–29. DOI: http://dx.doi.org/10.1145/2659766.2659776

[8] Mark Sullivan, Nick Garza, and Rick Cody. 2016. Black Hat Cooperative. (2016). http://www.teamfuturegames.com/gameplay.html Library Catalog: www.teamfuturegames.com.

[9] Suriyun. 2018. Yippy Kawaii. (2018). https://assetstore.unity.com/packages/3d/characters/humanoids/yippy-kawaii-88488 Library Catalog: assetstore.unity.com.