Embodied Realistic Avatar System with Body Motions and Facial Expressions for Communication in Virtual Reality Applications

Sahar Aseeri* Sebastian Marin† Richard N. Landers† Victoria Interrante* Evan S. Rosenberg*

*Computer Science and Engineering †Industrial/Organizational Psychology
University of Minnesota

Figure 1: An embodied virtual avatar with realistically rendered user movements and facial expressions for VR real-time communications.

Abstract

Providing effective communication in virtual reality (VR) applications requires body language and facial expressions to transmit users' thoughts and behavior. Embodied VR is an effective way to realistically render users' movements onto a virtual avatar in a virtual environment in real time. There are many VR applications that utilize this technique; however, most of them have limitations in photorealistic realism, body motion, and facial expressions. In this work, we introduce a novel VR communication system that mimics users' movements, facial expressions, and speech in order to render these capture data into different types of avatar representations in real-time. This system could be beneficial for any type of VR application that requires avatar-based communication in a 3D immersive virtual environment.

Index Terms: H.5.1 [Information Interfaces and Presentation]: Multimedia Information Systems—Artificial, augmented, and virtual realities; I.3.7 [Computer Graphics]: Three-Dimensional Graphics and Realism—Virtual reality

1 Introduction

Virtual reality (VR) technology has resulted in techniques with great potential to support communication in recent years by enabling groups to communicate in a 3D immersive virtual environment (IVE). Creating effective communication in a 3D IVE that simulates real communication remains one of the most difficult problems in VR applications. The specific difficulties include capturing human behavioral interaction with movement in terms of facial expressions, body gestures, and eye gaze in real time.

*e-mail: {aseer002, marin343, rlanders, interran, suma}@umn.edu

There are many VR applications that support a low-fidelity embodied avatar for interaction and communication. However, these applications have some limitations in communication with the avatar due to a lack of facial expression or body movement. This problem could affect the immersion and sense of presence in a 3D IVE. Smith and Neff (2018) discussed the fact that having a low-fidelity embodied avatar in VR for interaction resulted in a significant social presence with communication behavior and conversation compared to not having an embodied avatar. They excluded a discussion of facial expressions due to the limitations of the technology, but they mentioned that adding facial expressions into a realistic embodied avatar could provide a much greater sense of the presence of other people; this, in turn, could deliver more effective communication in VR [3].

Therefore, in this work, we propose a system that supports embodied virtual avatars in VR by capturing different types of data in real time (see Figure 1). These data are a combination of a user's body movement, facial expressions, and streaming audio that are applied onto a 3D virtual avatar to support communication in VR in real time. We have used HTC Vive trackers along with an inverse kinematics (IK) solver for body movement, an iPhone X to capture face data, and a microphone for audio input in real time.

The system would be beneficial for creating a 3D interactive environment for different types of VR applications that are based on communication with a virtual avatar. For example, one application could be job interview training with a virtual interviewer in real time. Another application could be reducing implicit biases of protected groups through various interventions (e.g., stereotype replacement, counter-stereotypic imaging, or partnership building) in an IVE. This could be done by providing an opportunity for individuals to experience communication with an embodied virtual avatar in real time with various kinds of appearances and different types of scenarios to enhance user performance or change user perspectives.

978-1-7281-6532-5/20/$31.00 ©2020 IEEE
DOI 10.1109/VRW50115.2020.00141

2 Related Work

There are many relevant studies that have been conducted to improve communication with a virtual avatar in a virtual environment (VE). Lopez et al. (2019) [1] described a system for full-body embodied avatar movements. They explored the implicit gender bias and embodied virtual avatar by using full-body visuomotor synchrony. They used nine HTC Vive trackers and a head-mounted display (HMD) with IK to animate the full-body virtual avatar in real time. Some of the system's limitations are missing facial features and finger movement. Our system includes features for facial expressions, and we intend to add finger motion in real time for future work.

Roth et al. (2017) implemented another system called SIAM-C, which is an immersive avatar-mediated platform for communication in VR [2]. The virtual avatar imitates a real human in real time by tracking body motions and facial expressions, as well as transmitting audio. They installed two OptiTrack systems in two rooms for body tracking and RGB-D sensors for facial expressions. Also, they used a microphone for real-time audio input. The avatar is projected on a big screen in front of the users in the two rooms for communication. Our system is quite similar to this work in terms of tracking body motions and facial expressions, as well as transmitting audio. However, our system is more immersive, with a higher level of realism in the virtual environment.

3 System Setup

This system requires a two-user setup; one user animates the virtual avatar body and facial expressions with audio in a 3D virtual environment, and the other user views the animated virtual avatar through an HMD to communicate with it in real time. The system allows users to communicate with each other in the same place or in different places.

The system involves Reallusion Character Creator 3, which is a software tool that provides numerous features to construct the desired 3D realistic avatar by choosing the type of hair, face, and/or cloth, the color of the skin, and the body shape. We later imported the avatar into iClone 7, which is real-time 3D animation software used with the Motion Live plug-in on the server (the desktop) and the LIVE FACE application by the client (the iPhone X device) for face animation. Additionally, the Unreal Engine (UE4.22) was used to render the live motions of the avatar in a 3D immersive environment. This can be done by using the Unreal-iClone Live Link plug-in to stream the 3D character motions from iClone 7 to the Unreal Engine in real time.

Three features are needed to animate the avatar in a 3D virtual environment: body animation, facial expressions, and audio. In terms of animating the virtual character body, we used the inverse kinematics (IK) solver within the unreal engine to control the 3D character in the virtual environment. The user is able to animate all or part of the 3D character body using three to five HTC Vive trackers. Because our VR interview application involves seated use, we have applied three Vive trackers that are tracked by the lighthouse system. The transformation matrices for each tracker are applied to the head effector, left-hand effector, and right-hand effector per frame. The IK algorithm receives all these effector matrices as input to calculate the other transformation matrices for other character body joints. This technique helps the user control the pose of the avatar in real time.

For virtual character facial expressions, the server (the desktop) and the client (the iPhone X) should be connected to the same Wi-Fi network, which may involve connecting the LIVE FACE application to the Motion Live plug-in in iClone by inserting the client's IP address on the server side. This enables streaming face data with the capture of facial expressions by the iPhone X depth camera, which are then applied onto the virtual character in iClone. Later, we transferred the virtual character with its real-time 3D face animation to the Unreal Engine by using the Live Link plug-in to view the character in a 3D immersive environment using the HMD. In order to reduce latency for the facial expression data, all components are connected using a local-area network provided by a Netgear Nighthawk R6700 gaming router with NAT filtering disabled.

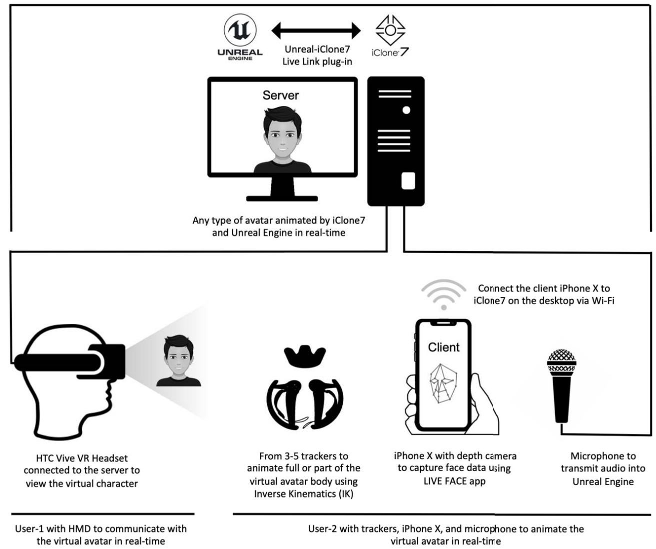

Figure 2: Overview of the prototype avatar-based communication system supporting body movements, facial expressions, and speech.

Audio is also supported through this system by linking the microphone directly to the server to receive audio input data in the Unreal Engine while the avatar is speaking in real time. Figure 2 shows the whole system setup, which could be used for any VR application that needs virtual avatar communication.

4 Limitations and Future Work

As previously mentioned, this system would be useful for creating a 3D interactive environment for different types of VR applications that are based on communication with a virtual avatar. The system could solve the problem of rendering the nature of human motions and apply communication prompts such as verbal and nonverbal cues to avatars to create a real-time communication environment that could increase users' sense of communicating with others in a virtual world.

This system is currently limited to VR scenarios requiring one user to be embodied with a virtual avatar and the other user viewing and communicating with the virtual avatar in the virtual environment. For future work, this system could be improved by allowing both users to be embodied with virtual avatars and both communicating with each other with full-body motions and facial expressions in real time. It may also be possible to use more photo-realistic avatars to increase users' presence in the virtual world.

References

[1] Sarah Lopez, Yi Yang, Kevin Beltran, Soo Jung Kim, Jennifer Cruz Hernandez, Chelsy Simran, Bingkun Yang, and Beste F Yuksel. Investigating implicit gender bias and embodiment of white males in virtual reality with full body visuomotor synchrony. In *ACM Conference on Human Factors in Computing Systems*, page 557. ACM, 2019.

[2] Daniel Roth, Kristoffer Waldow, Marc Erich Latoschik, Arnulph Fuhrmann, and Gary Bente. Socially immersive avatar-based communication. In *IEEE Virtual Reality*, pages 259–260. IEEE, 2017.

[3] Harrison Jesse Smith and Michael Neff. Communication behavior in embodied virtual reality. In *ACM Conference on Human Factors in Computing Systems*, page 289. ACM, 2018.