Virtual Reality (2020) 24:53–73
https://doi.org/10.1007/s10055-019-00382-0

ORIGINAL ARTICLE

Cinemacraft: exploring fidelity cues in collaborative virtual world interactions

Siddharth Narayanan[1] · **Nicholas Polys**[1] · **Ivica Ico Bukvic**[1]

Received: 30 July 2018 / Accepted: 3 April 2019 / Published online: 16 April 2019

Abstract
The research presented in this paper explores the contribution of avatar fidelity to social interaction in virtual environments and how sensory fusion can improve these interactions. Specifically, we vary levels of interaction fidelity to investigate how responsiveness and behavioural realism affect people's experience of interacting with virtual humans. This is accomplished through the creation of Cinemacraft, a technology-mediated immersive platform for collaborative human–computer interaction. Cinemacraft leverages a voxel game engine similar to Minecraft to facilitate collaborative interaction in a virtual 3D world and incorporates sensory fusion to improve the fidelity of real-time collaboration. The primary hypothesis of the study is that embodied interactions result in a higher degree of presence, and that sensory fusion can improve the quality of presence and co-presence. We tested our hypothesis through a user-study of 24 participants. Based on suggestions from existing literature, we sidestep the uncanny valley effect through the use of low fidelity avatars (a la Minecraft) and identify cues that impact users ratings of presence, co-presence and successful collaboration. The findings and ensuing data in this research can be applied to produce a more compelling platform for live collaborative interactions, performances, and empathetic storytelling. This research contributes to the field of immersive, collaborative interaction by making transparent the platform, methodology, instruments and code accessible for team members with less technological expertise, as well as developers aspiring to use interactive 3D media to promote further experimentation and conceptual discussions.

Keywords Information systems · Software engineering · Virtual worlds software · Computing methodologies · Motion capture

1 Introduction

Recent revolution in the area of off-the-shelf immersive technologies and phenomenology has changed the way users interact with games, media, and the arts. Creative projects have actively adopted the Microsoft Kinect (Microsoft 2017a), along with an array of affordable alternative all-in-one consumer-level motion-capture devices to explore novel interactions and perceptions. Experiments have integrated human interaction into performance, either as stylized body movements (Ratcliffe 2014) or through the use of virtual interfaces (Ahmaniemi 2010). Video games have also served as a rich foundation for artistic expression using immersive devices through machinima (Kastelein 2013) and digital puppetry (Polyak 2012). These new modes of interaction are also a significant boon to the level of immersion in video game technology.

Studies have shown that an embodied interaction is correlated to players' increased engagement (Bianchi-Berthouze 2013) and a stronger affective experience (Bianchi-Berthouze et al. 2007). Most modern video games however, continue to be played using low fidelity interaction devices—interactions between video game characters and the user are often accomplished through keyboard presses or joystick manipulation for various body motions, ranging from simple actions such as walking or jumping to more elaborate tasks like opening doors or pulling levers.

Such an approach can often result in non-natural and potentially limiting interactions. Such interactions also lead

✉ Siddharth Narayanan
nsiddh3@vt.edu

Nicholas Polys
npolys@vt.edu

Ivica Ico Bukvic
ico@vt.edu

[1] Virginia Polytechnic Institute and State University, Blacksburg, USA

to profoundly different bodily experiences for the user and may detract from the sense of immersion (Parker 2008). Modern day game avatars are also loaded with body, posture, and animation details in an unending quest for realism and compelling human representations that are often stylized and limited due to the aforesaid limited forms of interaction. These challenges lead to the uncanny valley problem that has been explored through the human likeness of digitally created faces (Lay et al. 2016; Seymour et al. 2017; Kätsyri and de Gelder 2018), and the effects of varying degrees of realism in visualizing players' hands (Argelaguet and Hoyet 2016). More recent studies such as Makled and Abdelrahman (2018) also show the influence of body animation (excluding head animation) on viewers perception and realism of the computer generated human.

Some titles have taken an alternative approach towards sidestepping the uncanny valley by utilizing simplified cartoon-like graphics. For instance, in Minecraft, a sandbox-like gaming environment that has attained an unprecedented level of popularity, the simple avatar movements and highly stylized cartoon-like and 8 bit-like graphics (Garrelts 2014) make it "work" because of a strong mix of the game's aesthetic sensibility, open-ended design, mechanics, development history, and the creative activities of its players (Duncan 2011).

A growing number of research projects explore Minecraft as a platform for enhanced immersion (Choney 2016; Huh et al. 2018) along with full body and facial immersion (Viniconis 2011). However, with the exception of HTC Vive's VR implementation of Minecraft (Vivecraft 2016) they have remained by and large confined to complex setups that are difficult to reproduce. In addition, such projects often require motion-capture data to be aligned with other information, resulting in a complicated endeavour when utilizing a combination of devices (Carey and Ulas 2016). This limitation impedes the level of immersion the platform is capable of delivering.

Immersion has also been shown to significantly depend on aspects of virtual presence like emotion (Lombard and Ditton 1997; Mousas et al. 2018). Such emotional immersion (Thon 2008), along with storytelling (Shin 2018) promotes the building of a bond between the user and in-game avatars in the virtual world. This concept is leveraged by multi-sensory immersive virtual environments, which have been shown to be capable of inducing emotional responses (Bailey et al. 2012).

1.1 Motivation

The relevance of (various forms of) presence for immersive technologies and virtual worlds in particular has been repeatedly emphasized (Nah et al. 2011; Schultze and Orlikowski 2010; Animesh and Pinsonneault 2011). However, the question remains—what creates and contributes to the sense of presence. Recent work such as Seymour et al. (2018) also points to the need for a better understanding of how user response to avatars with different degrees of realism, or which factors might contribute positively to the creation of natural and believable interactions for realistic visual presence.

Additionally, notable conclusions on the impact of individual performance on presence drawn by various studies differ. Some results suggest that a higher level of presence is always preferable to a lower level when intending to increase the performance of individuals, for instance during memorization tasks (Hirose et al. 2009; Ragan et al. 2010).

In contrast, some suggest that there is no evidence for a causality between presence and in-world task performance (Sacau and Laarni 2008; Schultze and Orlikowski 2010). Moreover, when participants of virtual worlds are not on their own but rather interact with (and experience) others in the virtual space, the feeling of being there is supplemented by the feeling of being with others (Schultze 2011; Schroeder 2012).

A major challenge for avatar interaction is finding an acceptable balance between complexity and control. As for instance, attempting to exactly control the position of objects in a virtual world and aiming at generating natural-looking movements at the same time are conflicting objectives (Sims 1994). Additionally, latency effects may cause serious consistency problems (Bainbridge 2007; Fritsch et al. 2005). In this context, research on the appearance of and reactions to avatars or virtually embodied agents can draw on experiences similar to the Uncanny Valley (Mori et al. 2012). Additionally, while some work such as Narang et al. (2018) has focused on programmed avatar-agents for more realistic interactions to enhance user co-presence, this is yet to fully explored in user-controlled embodied interactions. Related studies have also examined whether the uncanniness of an animated character, that is, the extent of how awkward a nearly human looking character is being perceived by others due to slight derivations from true human behaviour may depend on which emotion is being communicated by that character (Tinwell et al. 2011).

It appears that presence and co-presence with respect to the avatar's behavioural fidelity in immersive virtual environments need to be further explored. Additionally, the challenges of avatar design within social virtual environments along with the balance between avatar complexity and behavioural control need to be further investigated. Moreover, the potential for an immersive environment to serve as a compelling platform for expression, empathy, and storytelling with an emphasis on ease of use and accessibility remains largely underutilized, especially in virtual collaborative environments. This is in good part due to the aforesaid uncanny valley challenge, as well as due to commonly

complex, site-specific, and/or cost-prohibitive nature of the current solutions. Inspired by the successes of projects like Minecraft, we see this as an opportunity to introduce an accessible and affordable alternative that seeks to sidestep uncanny valley by employing cartoon-like environment and increasing the avatar's emotional and expressive depth by increasing the interaction fidelity. The role of interaction in uncanny valley has also been explored in studies such as Seymour et al. 2017), who propose that the simple relationship between affinity and realism in the Uncanny Valley theory needs to be rethought to account for complexity of a situation where interactivity is introduced. Recent work (Roth et al. 2016) suggests that despite demands for the improvement of avatar appearance, realism and sensory modalities are still limited in current immersive systems, as user's facial expression and eye gaze are typically not faithfully replicated.

1.2 Research problem

One of the major drawbacks of collaborative virtual environments (CVEs) is the relative paucity of avatar expressiveness compared with live human faces on video. Avatars in graphical chat platforms vary widely in appearance and can exhibit lively behaviours, however they have been critiqued for serving merely as placeholders and failing to contribute meaningfully to the conversation. The avatars used in collaborative laboratory-based studies are typically visually simplistic and have limited behavioural capabilities, such as the movement of a single arm for object manipulation. A significant challenge in developing CVEs as a communications medium is the development of expressive avatars capable of contributing to the interaction. Although CVEs can offer the benefits of spatial interaction and immersive experience, they remain low fidelity compared with video-mediated communication (VMC); where VMC portrays objects and events from the real world, CVEs portray an artificial environment populated with artificial representations of people.

In increasing avatar fidelity there are technical challenges as well as theoretical goals to consider. These affect both the avatar's static appearance (visual fidelity) and dynamic animation (behavioural fidelity). This research focuses on the latter, the behavioural fidelity of the avatar, by changing the interaction modes while keeping the avatar's visual fidelity to be constant. We study the contribution of the avatar's behavioural fidelity to presence and co-presence in the collaborative virtual environment. We also contribute to previous work focusing on users' immersive tendencies and how they relate to the level of presence and co-presence they experience. The research questions are addressed through a combination of post-test questionnaires and analysis of user-study data. The items for each questionnaire are detailed in "Appendix 1".

1.3 Presence and co-presence

One of the distinct features of virtual worlds is embodiment, which allows a user to "be" in the virtual world through avatars. This experience resembles a phenomenon rooted in media research which is referred to, among others, as presence. Presence has also been circumscribed as the extent to which users physically attribute themselves to a virtual world by means of their avatar as their mental representation (Nash et al. 2000). Through re-embodiment and avatar identification, users not only experience their avatar as an extension "of an actual human mind translated into a virtual body," but also receive the actual feedback of "a human mind seeing oneself as a body present in a virtual world" (Bray and Konsynski 2007).

Social presence is defined as the degree to which a user perceives other people to be physically present when interacting with them (Carlson and Davis 1998). Generally, social presence theory assumes that the more social cues a medium conveys, the more it will be perceived as warm, personal and sociable (Yoo and Alavi 2001). A study in the context of virtual worlds found that social presence was the only social outcome which had a significant impact on users' intention to use this technology (Mäntymäki and Riemer 2011). Self-presence relates to avatar-identification aspects, thus a state where "the virtual self is experienced as if it were the actual self" (Park et al. 2010).

Co-presence lies at the intersection between tele and social presence and refers to a sense of collocation or "the sense of being in a shared virtual setting with remote others" (Schultze and Orlikowski 2010). In the context of virtual worlds, Saunders et al. (2011) have focused on two interpretations of presence, namely in the form of social richness (based on social presence theory) and in the form of immersion. As indicated above, social richness appraises the perception of media according to a medium's ability to establish a personal connection through the amount of human warmth, intimacy, and sociability transmitted (Sia et al. 2002; Zhu et al. 2010). Immersion qualifies the extent of perceptual and psychological immersion of a person into a virtual environment, thus "the extent to which the person seems to be immersed or engaged in the virtual world" (Saunders et al. 2011). Using "bodily practices such as sitting, gesturing, smiling, and dressing", virtual worlds are considered "potentially more immersive than other media" (Schultze and Orlikowski 2010).

1.4 Research questions

Our research comprises user-based experiments addressing two nested questions:

Question 1: What is the relationship between the avatar's behavioural fidelity and presence?

This question addresses the assumption made by numerous researchers, that behavioural fidelity should be prioritized over visual fidelity in the development of expressive avatars. The avatar's functionality is modified through different interaction modes for each interaction exercise varying from standard keyboard + mouse interaction and audio chat and then moving on to incorporating upper body and full-body real-time motion capture to study whether improvements in behavioural fidelity benefit the constant low fidelity avatars regardless of their appearance.

Question 2: Does Sensory Fusion increase presence and co-presence?

This question is addressed by adding improved mouth detection through an audio input fusion layer. When measuring these improvements, we focused on components of avatar behaviour that contribute to co-presence to study people's sense of being with others in a shared VE. The question also explores whether there is indeed a correlation between presence and co-presence. Previous studies on a smaller groups (Slater and Sadagic 2000; Schroeder and Steed 2001) found a positive relation between presence and co-presence however, others have proposed an alternative trade-off where users can experience high presence or high co-presence, but not both (Spante and Heldal 2003).

2 Cinemacraft

Cinemacraft is a novel technology-mediated immersive machinima platform for collaborative performance and musical human–computer interaction. It innovates on a custom, reverse-engineered version of Minetest (2016), an open-source version of the ubiquitous Minecraft, to offer a collection of live theatrical and cinematic production tools, and leverages the Microsoft Kinect for Windows v2 (Kinect 2017) for embodied interaction, including posture, arm movement, facial expressions, and through the sensory fusion lip syncing based on captured voice input. It is designed as an out-of-box turnkey solution that side-steps the uncanny valley by utilizing a cartoon-like appearance for its gaming environment, for simple yet compelling storytelling along with multiple live camera views, scene changes, subtitles, lip sync, production-centric stage cues, and virtual audience. The platform is aimed at extending the frontiers of collaborative content creation as well as broadening audience impact to enhance creativity and emotional experiences. Recent research (Anderson et al. 2017) also supports avatar actions that map to established social norms in the physical world for more efficient communication and avatar movement was effective at communicating nonverbal information, even when done so unintentionally. This is especially relevant to our study when trying to capture a user's natural movements and voice cues to improve avatar behaviour.

2.1 The OPERAcraft lineage

The OPERAcraft platform (Bukvic et al. 2014) along with Mirrorcraft (Barnes et al. 2016), a precursor to Cinemacraft, were envisioned as environments to aid creativity and thinking skills and better self-expression, with particular focus on the K-12 education opportunities. OPERAcraft was built as an arts+technology education platform where students could write a story and libretto, build a virtual set, costumes or virtual character skins, and ultimately control the characters within the virtual setting in a live performance accompanied by live singers and musicians. Many of these affordances are inherent to the Minecraft platform–users can easily sculpt the landscape, interact with it, and change their own appearance. Others were added as part of the reverse engineering effort, resulting in a mode that is deeply integrated into Minecraft's core. These include character lip syncing based on the singer's input processed through the Pd-L2Ork (Bukvic 2012) and forwarded to a FUDI-based parser via a UDP socket embedded inside reverse-engineered version of Minecraft, audience subtitles and stage cues only visible to the actors, ability to change between discrete arm positions and interpolate between them to provide rudimentary body language, and near-instantaneous scene changes through coordinated character teleportation and scene cross-fades. In an ongoing pursuit of building a compelling real-time machinima production platform, the second generation of OPERAcraft introduced in the fall 2015 as part of the second opera production offers additional affordances, including multiple camera views and cameras that are only visible to the actors, invisible bystanders, as well as stability improvements and optimization that allowed the mode to scale beyond the original limit of five actors.

2.2 New platform

Cinemacraft builds on the live performance capture aspects of the existing system through the integration of the Microsoft Kinect for Windows v2 device to provide a more immersive and expressive embodied experience in a collaborative virtual world including both kinematic data and facial expressions. It in many ways supplants previously keyboard-controlled arm expressions and instead provides full-body immersion to the extent allowed by the simple skeletal structure of the Minecraft avatars that lack hands, elbows, and

knees, and further enhances expressivity by tracking facial expressions. Based on user tests and audience feedback, the avatar remains compelling despite the minimal character design due to the reported feeling of sentience offered by the user's real-life body motion and facial expressions. As a result, both user avatars can show a dynamic range of emotional reactions and responses. In cinematic terms, the avatar no longer appears to be merely acting. Rather, it is the actor who is responding to their projection in and the situational awareness of the virtual environment. Spontaneous reactions like squinting against a sudden bright light help to humanize characters and make them more compelling than current game characters that seem shallow and with whom we have a hard time forming compelling, coherent relationships (Buecheler 2010).

Sensory fusion and multi-sensory stimulation has been shown to help in illusorily experiencing a virtual body and virtual body parts (Lecuyer 2017). Its use within the context of embodied interaction can be particularly useful in situations where one sensory input does not provide adequate resolution and could benefit from secondary inputs that improve its accuracy and fidelity. In Cinemacraft, such a situation can be observed when trying to monitor mouth movement. Kinect for Windows v2, particularly when a user is located farther away from the sensor, so as to allow for full-body capture, is not capable of providing accurate capture of fast and varied shapes generated by the mouth. In this case, sensory fusion through the use of voice analysis can be an appropriate way of complimenting visual capture. Similarly, in situations where a user has their mouth opened and are making no sound, visual capture needs to take precedence as the only reliable source of information. While, the previous Cinemacraft version implements OPERAcraft-based avatar lip syncing based on the simple transient and melisma detection that is cross-pollinated with the Kinect for Windows v2-based capture with the two swapping precedence based on the context, the new version supplants this with the audio sensory fusion layer.

2.3 Modes of interaction

Synchronized movements between the user and their avatar have been shown to have a positive effect on both the users cognitive ability and feeling of agency over the virtual avatar (Kokkinara and Slater 2015). Additionally, the ownership of another person's body, or the "embodiment illusion", can be induced via multi-sensory correlation (Maister and Slater 2015). However, it's important to investigate such anatomical control systems in more depth, particularly the potential link between motion-capture functionalities and embodiment, in this case, in first person. Studies have found that participants' upper body movement being mirrored alone was a strong tool to provoke the illusion of both agency and body ownership towards the virtual body even without full-body tracking (Collingwoode-Williams and Gillies 2017).

Cinemacraft offers different modes of embodied interaction captured by Kinect for Windows v2, namely regular, mirrored, and upper torso only. For instance, if the user prefers to both control the avatar and act out the gestures and facial expressions they can gesticulate to the other players to complement their speech and chat messages and thereby increase the effectiveness of the conversations, while still being able to navigate the expansive landscape outside the range afforded by the area monitored by Kinect for Windows v2 using more conventional controls (e.g. keyboard). While some studies such as Makled and Abdelrahman (2018) point to the importance of body animations over head motion, another similar study on avatar appearance and mapping user movements to the avatar movements (Heidicker et al. 2017) found that while motion-controlled avatars with full representation of the avatar body lead to an increased sense of presence, avatars which have only head and hands visible produced an increased feeling of co-presence, i.e. we do not need a complete avatar body in social VR. This would be interesting to explore as part of interaction modes in our study. The same can be also applied in the gaming scenarios, as well as hybrid situations where a separate user controls the avatar while an operatic singer, for instance, provides only upper body language. Similarly, the mirroring mode has been added to explore illusory experience interactions with the avatar, most notably through the Mirrorworlds research project focusing on the study of integration of physical and virtual mirrored presence (Polys et al. 2015). These modes provide an opportunity to draw parity between the different approaches to machinima and also open new exciting possibilities for sensory fusion, with the introduction of HMDs (head-mounted displays) along gesture-based and haptic controllers. Cinemacraft also inherits a battery of OPERAcraft's cinematic tools, empowering user to explore the methods of live machinima production, including live theatrical play, as well as produced cinema. The virtual audience feature, that enables audience members to freely roam the scene, or the ensuing world in which the storytelling takes place, offers new research opportunities in the study of perception of storytelling, drama, and empathy as a function of vantage point.

3 Implementation

The avatar and scene rendering are performed through the Minetest client on two networked high-end graphics PCs. At each location, a display peripheral can be inserted in the set to project the screen for a larger field of view and immersive experience. Each set is also equipped with Microsoft Kinect cameras and microphones to capture the actor. Cinemacraft

captures the user's motions and facial expressions in real time within Minecraft using the Kinect for Windows v2 sensor and a custom C# WPF (Microsoft 2017b) client that leverages Kinect for Windows v2 API. The C# WPF Kinect application and Cinemacraft Minetest client are packaged as a drop-in mode and an independent executable file, respectively. A major consideration for our setup was accessibility. For this reason, although earlier prototypes relied on two first-generation Kinects, due to their observed inability to simultaneously do body and face tracking, the current implementation relies on just one second generation Kinect for Windows v2 device responsible both for kinematic and facial tracking. The actor is tracked using the custom WPF application for simultaneous body and face motion tracking system calibrated to transmit skeleton information to each computer. The microphone data are captured through a regular audio chat application while the sensory fusion audio data are captured through a custom PurrData patch which sends UDP packets to the WPF application.

In order for the Kinect for Windows v2 client to interface with the Minetest mod, core Minetest was reverse engineered and retrofitted with a FUDI-compliant protocol (FUDI 2017) capable of parsing remote messages. Its simpler version is already found in OPERAcraft and previous versions of Cinemacraft where it was used to coordinate various aspects of the production, including switching camera angles, lip syncing as detected by the singers' microphones, subtitles, and stage cues. As a result these could be handled remotely through multiple distributed Pd-L2Ork clients. The ensuing UDP based protocol can be seen as a simplified counterpart to the Open Sound Control (OSC) (Wright et al. 2003). All communication is relayed through UDP packets between the microphone and WPF application and Minetest clients. On the client's monitor, users are able to see the reactions of the other participant as well as their own avatars in the virtual world during the interaction, allowing them to monitor how their actions affect both the physical and the virtual worlds. Because the performance is driven by real-time motion data, this information need to be transmitted via network. The virtual interaction must be synchronized in real-time on each user's monitor as well as on the server screen. The virtual world in this case is a selection of neutral Minetest game maps that both participants can choose from. A media layer where the interaction between audience through the server and clients can also be integrated and manipulated at will. Cinemacraft, handles positions through real-time processing by effectively updating the avatar's motion in game. This allows support for multiple clients that communicate with other users along with out-of-box multiplayer support with chat and other core functionality.

3.1 Sensory fusion

The emphasis on ease of use and reliance on a single Kinect for Windows v2 device requires our implementation to essentially stretch the limits of the current Kinect for Windows v2 API. While the Kinect for Windows v2 does provide improved resolution over the first generation, it is still best suited for face tracking in close proximity which limits its ability to track body. In turn, our implementation offers accurate simultaneous full body and facial tracking. Further still, we have identified problems with Kinect's machine learned library of postures and facial expressions that have resulted in a prevalent number of false positives pertaining to eye winks, eyebrow movement, and eyeglass detection.

We have envisioned a platform with parallel pipelines of Audio Inputs, Kinect API and Computer Vision optimization and learning for improving facial Expressions, with all three working together to further refine the platform's capabilities through sensory fusion. To address Kinect for Windows v2's limited ability to detect mouth shapes, we merged the depth camera data with the captured audio input. Here, the sensory fusion allowed us to use voice detection to combine the user's audio with the facial tracking data and thereby improve detection of minor gestures and facial expressions which may not be otherwise captured due the technical limitations of the two distinct approaches to monitoring user's input. For instance, doing so enabled us to animate mouth motion through captured audio that exceeds the resolution of 30 frames per second, as well as audio-centric outliers, such as the cartoon-like quivering of lips in a sung operatic melisma. The ensuing implementation utilizes a simple logic by which the two sensory inputs are given precedence (Fig. 1).

The sensory fusion layer uses switching and heuristics to allow the audio input to take precedence over incoming mouth data in the event of incoming audio data. The

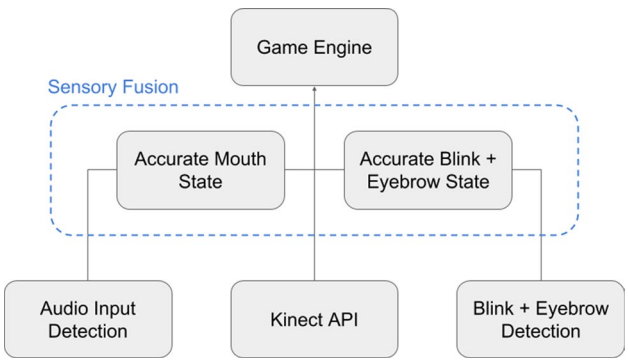

Fig. 1 Sensory fusion design

additional audio input allowed the purrdata patch to capture the user input sounds through a microphone and translate them into numeric values that were sent to the WPF application. Additionally, the specific sounds and pronunciations can be mapped to unique corresponding facial expressions to further enhance the realism of the avatar detection. Therefore when a threshold audio detection is crossed, a certain numeric value is generated through a purrdata patch for that loudness and enunciation. This numeric value is then sent to the WPF application which then maps the appropriate eyebrow state and mouth state using a face matrix and send the packet to the game. The data is then parsed to check whether the fusion layer must be activated and with which state. Further, there are also a fail-safe default values in cases a new unknown value for the face matrix is generated that is not found in the game textures. Additions to the existing Kinect WPF C# application were made to relay the incoming audio data. This was implemented through a new socket and thread to read and parse the data into a compound packet for the Minetest client. The new version of the application also supports much larger range of facial expressions which are mapped to corresponding textures in that change according to the user's facial expression in the real world.

3.2 Future enhancements

The sensory fusion layer can also be leveraged to address existing challenges with Kinect API. We have identified problems with Kinect for Windows v2's machine learned library of postures and facial expressions that have resulted in a prevalent number of false positives pertaining to eye winks, eyebrow movement, and eyeglass detection. While we had explored further enhancing face detection with infrared video feed inherent to Kinect for Windows v2, the low reflectivity of eye pupils when captured by the Kinect for Windows v2 camera from a distance makes the task extremely difficult. A separate module can be added to run a low-latency computer vision algorithm on the facial capture output of the Kinect for Windows v2 to improve the tracking of the eye and eyebrow states. Preliminary testing with a zoom-in capture of the user's eyes to track reflectivity using infrared has shown some promise. However, the version of application used for the experiments described in this paper only used audio input detection as part of the sensory fusion layer.

The focus of our study is the modular methodology of the sensory fusion layer to sidestep limitations in the existing Kinect API for improved accuracy in the final application rather than to direct improvements to Kinect API.

4 Experiment

All experiments had a common theme, namely the visual impact of avatar fidelity on the interaction. However, different aspects of behaviour fidelity as well as different responses were explored in each experiment through increasing the level of interaction fidelity. The general expectation was that the greater the level of interaction fidelity, the more the virtual humans would be seen to contribute to the experience and the more they would elicit social and co-presence responses from participants. However, one challenge in this area of research is that, just as there exist many questions about the impact of virtual humans, so are there open questions about what constitutes a social and co-presence response. The first step in designing the experiments was therefore to define the specific research question in terms of the response variables of interest.

The hypothesis to test here is first, that avatars with higher interaction fidelity will enhance the sense of presence and co-presence in a CVE. Secondly, sensory fusion for more accurate facial expressions would yield the highest presence and co-presence scores. This is done by exploring the extent to which embodiment through body synchronization and mouth synchronization (lip sync) using the sensory fusion layer influence presence and co-presence. Our hypothesis was that there would be a stronger effect on presence and co-presence when both lip and body motions are synchronized due to more access to control over the body. The expectation was that the mouth detection with sensory fusion interaction exercise would lead to an improvement in perceived communication quality regular mouth detection, based on the logic that its mouth movements were related to an aspect of the conversation taking place. In order to test the above hypotheses, the following response variables were constructed from n questionnaire items, each on a 1 to 7 Likert (Albaum 1997) scale with the score adjusted for analysis so that the higher score represented a higher response.

- Presence score, P: This variable is measured by making use of Slater's presence questionnaire (Slater 1999). It measures the degree of personal presence experienced by the participant.
- Co-presence score, CO-P: This variable measures the co-presence experienced by the user. The Co-presence score is further divided into contributing components adapted from the questionnaire.
- The immersive tendencies score, IT: This variable is measured using Witmer and Singer's immersive tendencies questionnaire (Witmer and Singer 1998). It meas-

ures the tendencies of individuals to become involved and immersed in the experience.

The scores for the response variables are measured using four interaction exercises are listed below in increasing order of their expected contribution to presence and co-presence:

- Interaction: Keyboard, Mouse + Inter-user communication: Audio chat
- Interaction: Kinect for face and upper torso, Keyboard + Inter-user communication: Audio chat
- Interaction: Full face and body motion + Inter-user communication: Audio chat
- Interaction: Full face and body motion with Sensory Fusion + Inter-user communication: Audio chat

Previous work has found that the IT predicts, within a given virtual environment, the level of presence felt by participants (as measured by their presence questionnaire). We need to check whether there indeed is a positive correlation with the immersive tendencies score using our presence and co-presence questionnaires. It is also important to see if there is a correlation between the P score and the CO-P score since previous research has indicated a positive correlation between personal presence and co-presence.

4.1 Design

This experiment investigated avatar behavioural fidelity along the interaction dimension and used a within-group experimental design. The experiment required pairs of participants who did not know each other prior to the experiment. An effort was made to remedy this by randomly allocating participants to each condition using a counterbalanced latin squares methodology to remove any input and ordering biases in the data collection based on their assumptions about what the experiment is about (demand characteristics). 12 pairs of participants were assigned to one of four conditions representing different methods of mediated communication. The conversations took place within the same building over a network link separated by a physical barrier. As mentioned the previous sections, a deliberate choice was made not to make use of the 3D potential of the avatar and retain the inherent low fidelity presence of the avatar and use affective animations. The players could either choose to manipulate the avatar in first person or third person. While some previous work has pointed to an increased sense of presence in first person point of view due to a greater sense of ownership (Slater et al. 2010; Denisova and Cairns 2015), research has also shown that third-person perspective generally helps users to better evaluate distances and anticipate the trajectory of mobile objects. Thus, for this study we chose to give users the option of choosing between both perspectives depending on their preference. Literature suggests that conceptualizing users as social actors puts researchers in a better position to ask with whom an actor is interacting, about what issues, under what conditions, for what ends, with what resources, etc. This approach particularly provides opportunities for advancing our understanding of virtual worlds' communication effects. Thus, role-playing various social interaction between the pairs of participants, co-located in the virtual world but separated in the physical world was chosen to be the best interaction exercise for the experiment (Fig. 2).

4.2 Avatars

Participants in each pair were represented by a visually similar avatar as differences in facial geometry and texture mapping could potentially impact on the visual effect of the animations. The only significant change was that a female avatar was used for female participants, and a male avatar for male participants. The participants could either choose to see their own avatar in third person or only choose to see the other person' avatar on the screen using the first person view. Each avatar was independently controlled for each user. The avatars are capable of a selection of behaviours such as smiling, frowning, looking sad, shrugging, pointing, waving, jumping etc (Figs. 3, 4).

4.3 Apparatus

The experiment space consisted of two conjoined rooms separated by a physical barrier. Each room contained a projector, PC, Kinect for Windows v2, microphone and peripherals for the user and the participants completed questionnaires following each interaction exercise in the same space. The rooms were equipped with identical equipment as described below. The rooms were purposefully bare in order to avoid providing visual distractions during the conversation. The two rooms in which participants were present were audio channel link through the microphone and a visual link through the Minetest game. The Kinect is placed at a sufficient distance from the participant to ensure that it can capture the entire user skeleton moving in the physical space while also being able to discern the user's facial expressions in sufficient detail. Participants sat 4 metres from a projector so that as the interaction exercise list changes with different input modes, they would be able to get up and move within the space without much trouble.

4.4 Interaction exercises

Since the two participants were expected to speak for several minutes and did not know each other prior to the experiment, it was necessary to give them a topic of conversation.

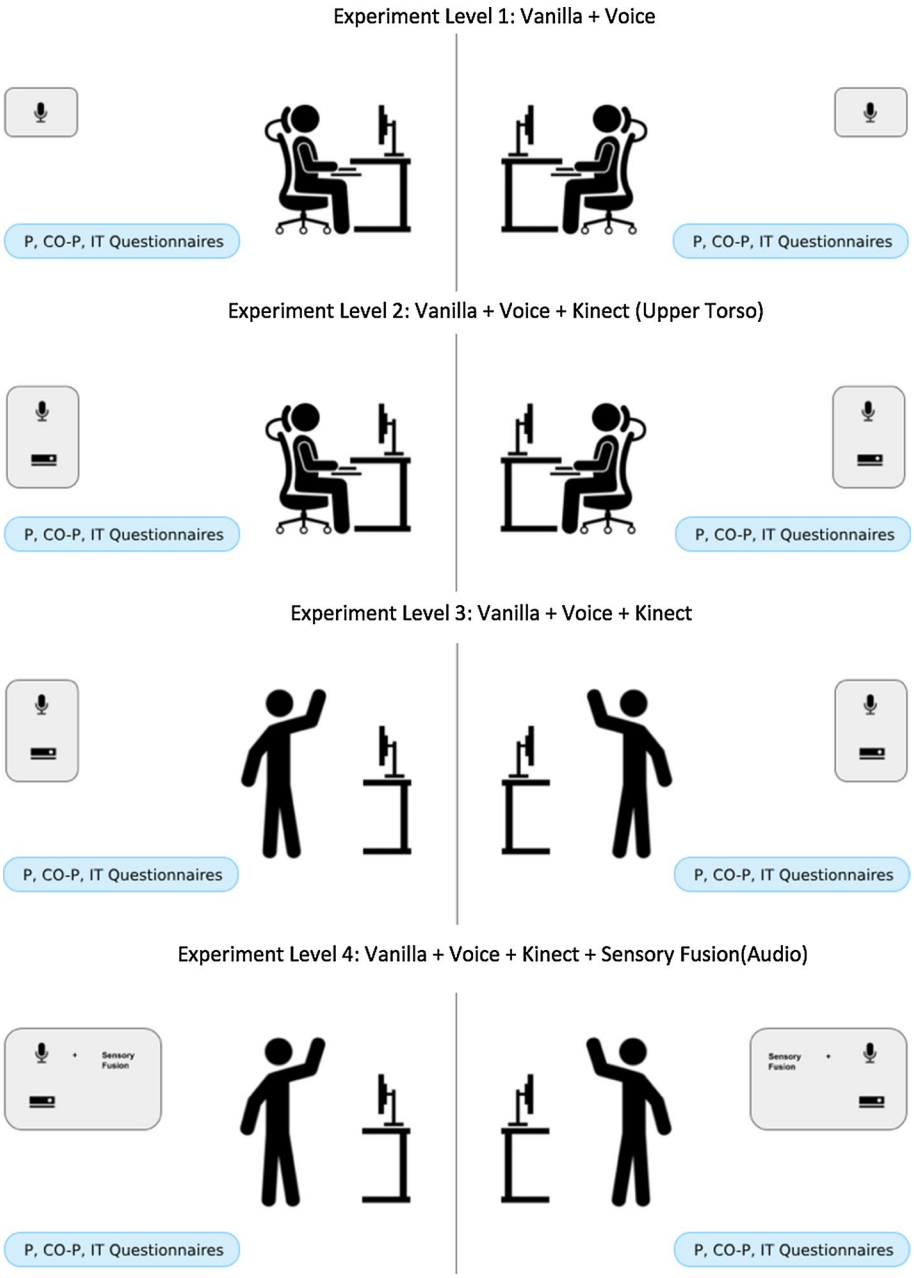

Fig. 2 The 4 conditions (interaction exercises) of the experiment with different modes of interaction. The vanilla version of Cinemacraft runs an unmodified version of the Minetest game engine. Users were provided Presence (P), Co-Presence (CO-P) and Immersive Tendencies (IT) questionnaires after each interaction exercise

The first two sessions were conducted using a simple and contemporary script that the users had to read out to each other, inspired by speech impediment treatment narratives. A notable deficiency that became apparent was that while the scripts seemed interesting by themselves, the conversations between the avatars seemed uninteresting since participants often remained stationary to converse and used minimal head and body motions. Thus the full range of 3D avatar facial expressions and gestures remained unused even at higher levels of embodiment and interaction fidelity. This led to the adoption of a second script designed as a guessing game where each participant had unknown object placed behind them that was only visible to the other participant. This was done in order to elicit stronger gestures, movements and audio input (for the sensory fusion layer) to generate more expressive avatars. While this led to a significant improvement in avatar facial expressions, the players still spent a sizable portion of the experiment standing still and the full potential of the full-body motion capture remained underutilized (Fig. 5).

Finally, a set of common and most recognizable body expressions was compiled in the form of a game where each participant must enact the designated body expression from a sheet, for the other person to guess within a stipulated time limit. A full list of these body expressions is provided in "Appendix 2". Users were given identical interaction

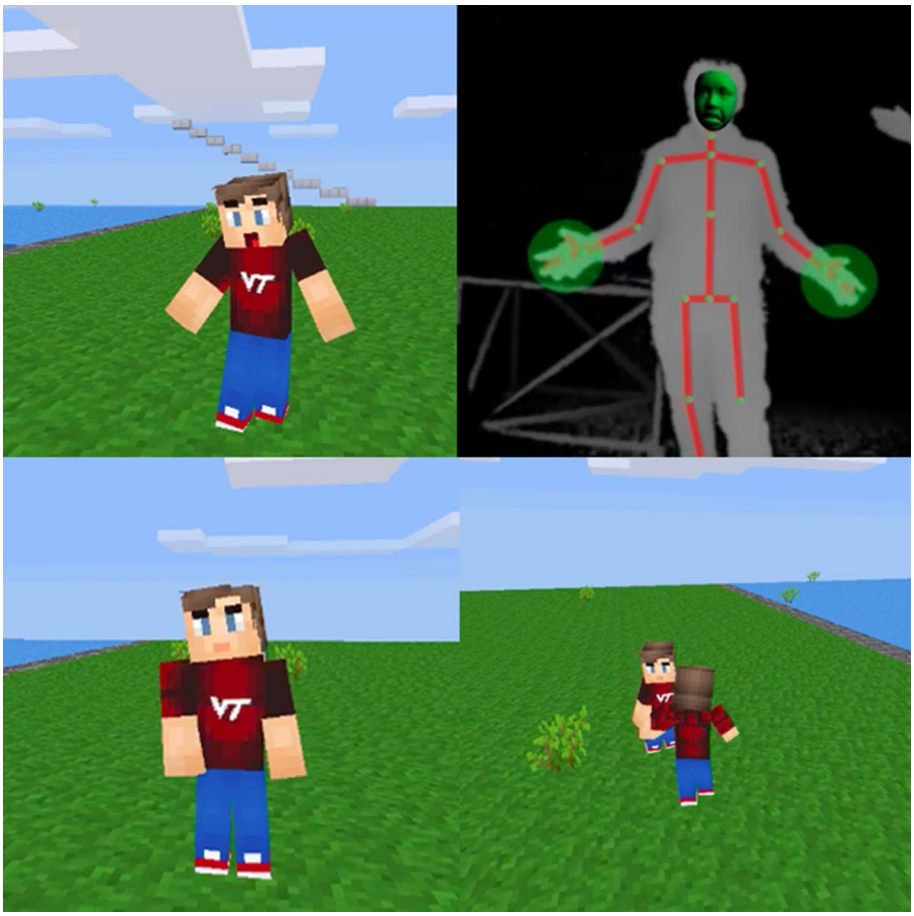

Fig. 3 Top left: participant 1—Avatar talking in sensory fusion mode; Top Right: Motion and Audio capture; Bottom Left: Participant 2—Avatar interacting in virtual world; Bottom Right: Both participants can view the scene in third person

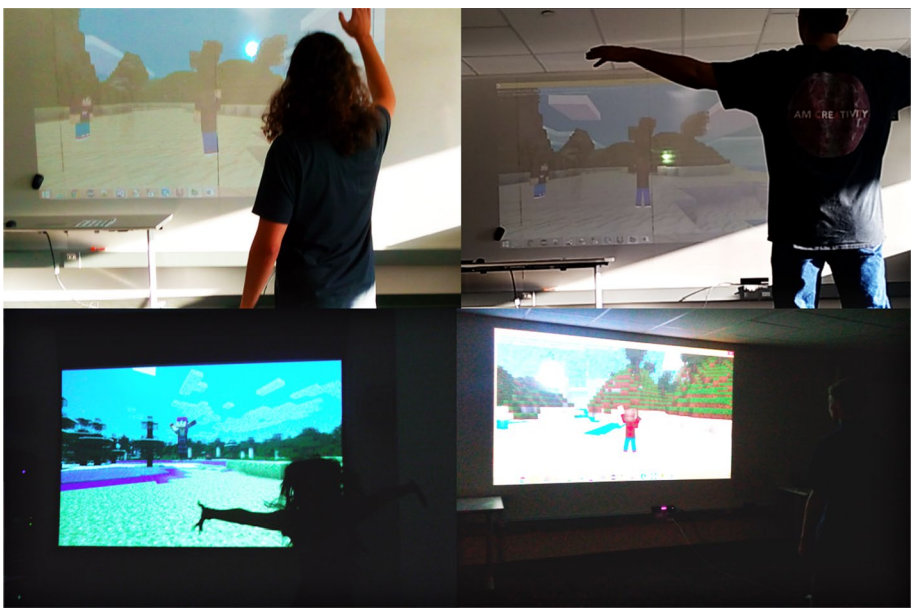

Fig. 4 Pairs of participants communicating with each other using embodied interaction and audio

exercise sheets for each experimental interaction exercise and were expected to enact out the body expressions without stating or explicitly alluding to the caption on the list. The goal of the game was to guess as many body expressions successfully between them within a stipulated amount of time, similar to a game of charade where users enact body expressions with the help of voice cues and phrases that do not directly include the displayed caption for the body

Fig. 5 Sample body expressions as part of the experimental interaction exercise list. A full list of these body expressions is provided in "Appendix 2"

expression. The set of selected body expressions accounted for the fidelity and range of available body expressions for a specific interaction mode. For instance, body expressions with lower body movements were excluded from the vanilla and upper torso interaction modes. The vanilla version also offers limited avatar interaction using the traditional mouse, keyboard setup and programmed key-frame avatar animations so participants relied more on voice cues. The participants were given 1 interaction exercise sheet each with a list of these body expressions for each experimental interaction exercise involving a specific interaction mode. The expectation was that participants' interaction exercise performance, i.e. the number of body expressions successfully guessed and enacted from their designated lists, would increase with higher interaction fidelity. The audio sensory fusion layer was expected to give the best results, i.e. the users would be able to guess the most number of body expressions successfully with synchronized mouth and body motions.

4.5 Procedure

Upon arrival, participants were greeted in a reception area by two experimenters (the author and a colleague). One experimenter was assigned to 'mind' each participant for the duration of the session. Participants were explained the experimental procedures and given the interaction exercise sheets. Participants were informed that all data would be confidential and would only be used for the purpose of data analysis. They were also instructed that they were free to withdraw from the experiment at any time and without giving a reason for withdrawing. Each participant was asked to sit down and the chair height was adjusted so that their face and shoulders were clearly visible on Kinect camera. All applications and the audio channel were pre-configured and running prior to participants' arrival. Participants were then given a few minutes to prepare for their interaction exercises. This included greeting each other and initiating a brief conversation through the audio channel. Once they felt ready to proceed, they were reminded of the amount of time they would have to perform their experimental interaction exercise, and that at the end of the interaction exercise the experimenter would return to guide them through the next stage. During the interaction exercise, the experimenters quietly observed participants. In the interests of a standardized procedure, participants were stopped at the end of the assigned time period regardless of whether the interaction exercise had been completed. After completing each interaction exercise, participants filled out questionnaires about their experience.

5 Results

The mean and standard deviations of the counts of responses across the n questions in each condition are presented as descriptive statistics for the questionnaire data in Table 1.

Our findings are presented in the form of response scores of each questionnaire variable for each user for each interaction mode, along with the cumulative mean and deviation for all users within a specific interaction mode. An alternate approach could be to observe the mean and deviation for each user for each questionnaire response variable across all interaction modes. However, we wanted our results to capture the differences in interaction modes and improvements in higher interaction fidelity.

The questionnaire responses for presence (Fig. 6) display a progressive increase with increasing interaction fidelity. This observation is in line with our hypothesis for our first research question, that increasing interaction fidelity through embodied interactions results in a higher degree of presence.

The x-axis represents the data point for each user which have been connected (instead of discrete data-points) to show the overall trend in data for that interaction mode.

Similarly, mean scores also show a progressive increase (Fig. 7); the mean for only Keyboard + Mouse was 72.5 with a standard deviation of 8.69. The mean for Kinect for face and upper torso + Keyboard was 92.54 with a standard deviation of 11.80. For Full face and body motion, the mean was 160.41 with a deviation of 7.56, and finally for Full face

Table 1 Mean ± standard deviations of count response variables

	Modality: keyboard, mouse ($N = 24$)	Modality: kinect for face and upper torso, keyboard ($N = 24$)	Modality: full face and body motion ($N = 24$)	Modality: full face and body motion with Sensory Fusion ($N = 24$)
Co-presence ($n = 28$)	50 ± 5.08	67.33 ± 4.13	123.41 ± 3.95	164.5 ± 4.96
Empathy ($n = 5$)	11.08 ± 0.58	11.416 ± 1.63	18.16 ± 1.00	25.20 ± 0.58
Mutual awareness ($n = 6$)	13.66 ± 2.31	16.12 ± 0.53	30.45 ± 1.10	36.16 ± 1.88
Attentional allocation ($n = 3$)	7.04 ± 0.69	12.29 ± 1.19	16 ± 0.29	18.04 ± 0.46
Presence ($n = 34$)	72.5 ± 8.69	92.54 ± 11.80	160.41 ± 7.56	191.45 ± 9.97
Immersive tendencies ($n = 14$)	70.16 ± 9.34			

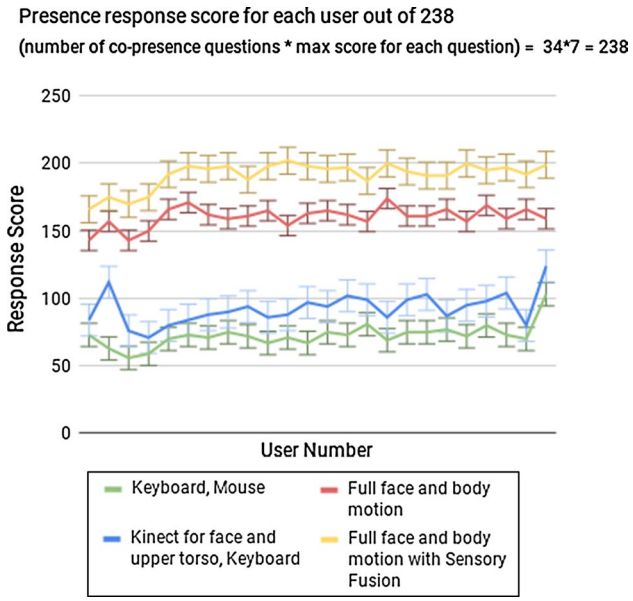

Fig. 6 Response scores for Presence Questionnaire for each user. *X*-Axis represents the data point for each user

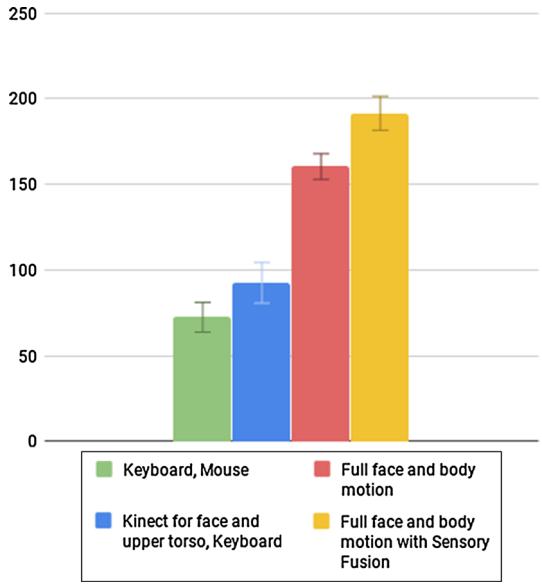

Fig. 7 Mean and deviation of cumulative questionnaire response scores for presence

and body motion with Sensory Fusion, the mean was 191.45 with deviation of 9.97.

The response scores for co-presence questionnaire items for each user also displayed an increasing trend with higher interaction fidelity modes in the response scores (Fig. 8). The mean for cumulative questionnaire response scores for co-presence (Fig. 9) for only Keyboard + Mouse was 50 with a standard deviation of 5.08. The mean for Kinect for face and upper torso + Keyboard was 67.33 with a standard deviation of 4.13. For Full face and body motion, the mean was 123.41 with a deviation of 3.95, and finally for Full face and body motion with Sensory Fusion, the mean was 164.5 with deviation of 4.96.

The co-presence scores are further divided into contributing factors—Mutual Awareness, Attentional Allocation and Empathy, which are also scored separately to reveal their individual trends with changing interaction fidelity. The Response scores for Mutual Awareness, Attentional Allocation and Empathy questionnaire items for each user are shown in Figs. 10, 11 and 12, respectively. The mean for cumulative questionnaire response scores for Mutual Awareness (Fig. 13) for only Keyboard + Mouse was 13.79 with a standard deviation of 2.32. The mean for Kinect for face and upper torso + Keyboard was 16.16 with a standard deviation of 0.81. For Full face and body motion, the mean was 30.66 with a deviation of 1.23, and finally for Full face and body motion with Sensory Fusion, the mean was 36.25 with deviation of 1.93. The positive trend in presence scores and contributing factors of co-presence with increasing interaction fidelity helps corroborate our hypothesis for the second research question by showing that sensory fusion helps to increase presence and co-presence.

The mean for cumulative questionnaire response scores for Attention Allocation (Fig. 14) for only Keyboard + Mouse was 7.04 with a standard deviation of 0.95. The mean

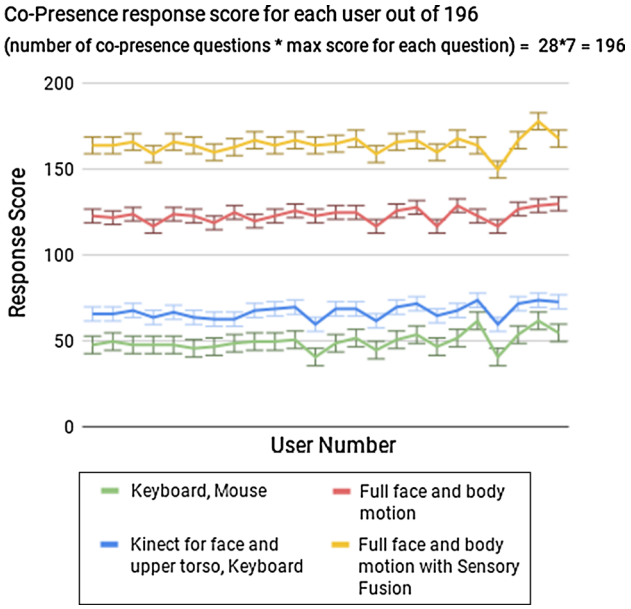

Fig. 8 Response scores for Co-Presence questionnaire items for each user. *X*-Axis represents the data point for each user

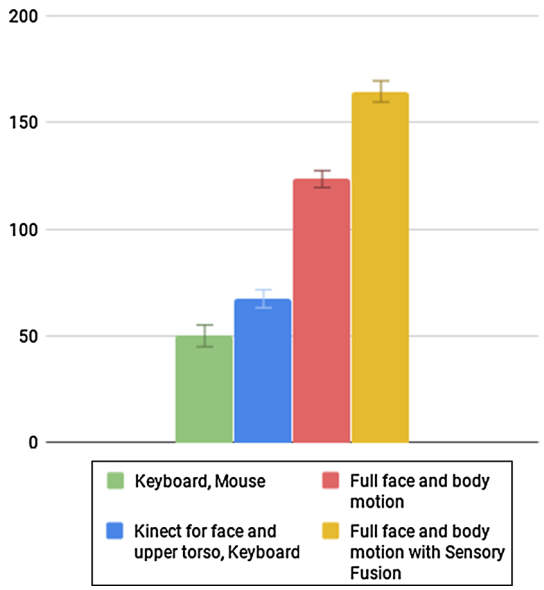

Fig. 9 Mean and deviation of cumulative questionnaire response scores for Co-Presence

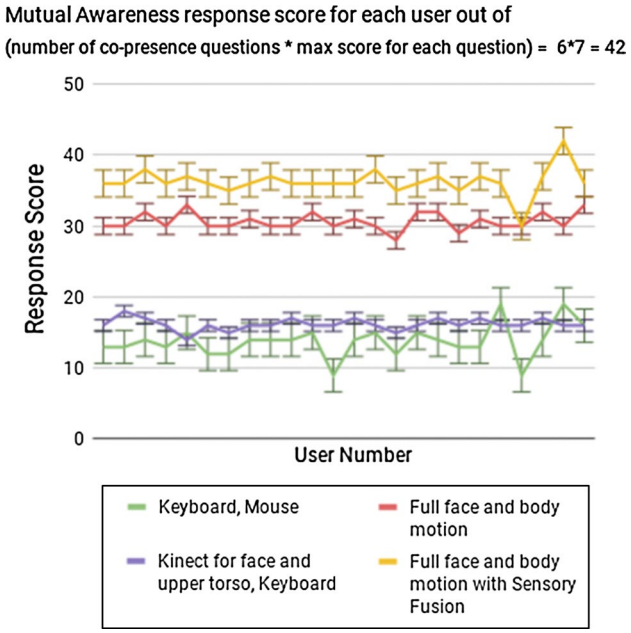

Fig. 10 Response scores for Mutual Awareness questionnaire items for each user. *X*-Axis represents the data point for each user

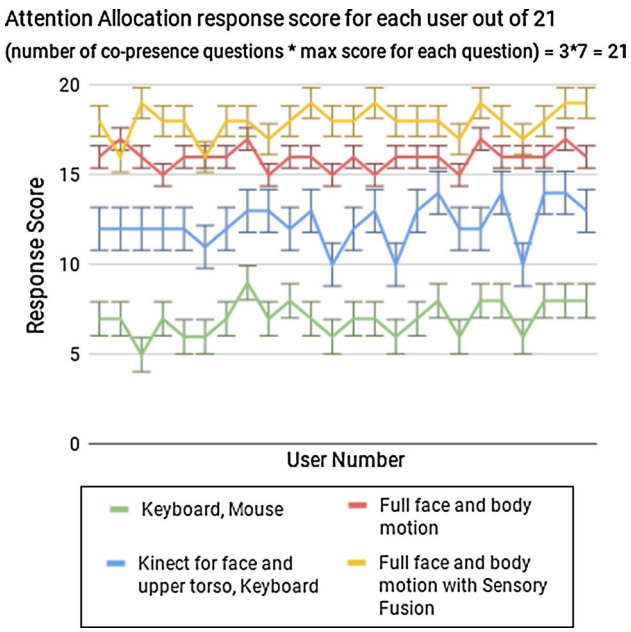

Fig. 11 Response scores for Attention Allocation questionnaire items for each user. *X*-Axis represents the data point for each user

for Kinect for face and upper torso + Keyboard was 12.29 with a standard deviation of 1.19. For Full face and body motion, the mean was 15.95 with a deviation of 0.62, and finally for Full face and body motion with Sensory Fusion, the mean was 17.95 with deviation of 0.85.

The mean for cumulative questionnaire response scores for Empathy (Fig. 15) for only Keyboard + Mouse was 11.29 with a standard deviation of 0.85. The mean for Kinect for face and upper torso + Keyboard was 11.5 with a standard deviation of 1.66. For Full face and body motion, the mean was 17.79 with a deviation of 1.64, and finally for Full face and body motion with Sensory Fusion, the mean was 24.66 with deviation of 1.57.

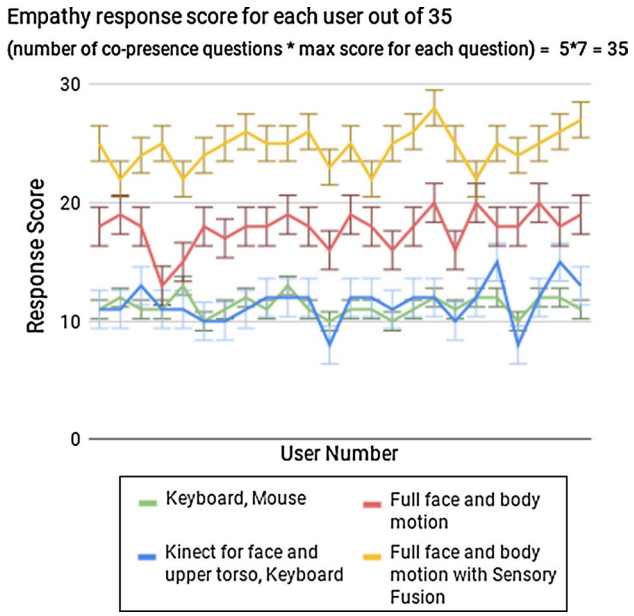

Fig. 12 Response scores for Empathy questionnaire items for each user. *X*-Axis represents the data point for each user

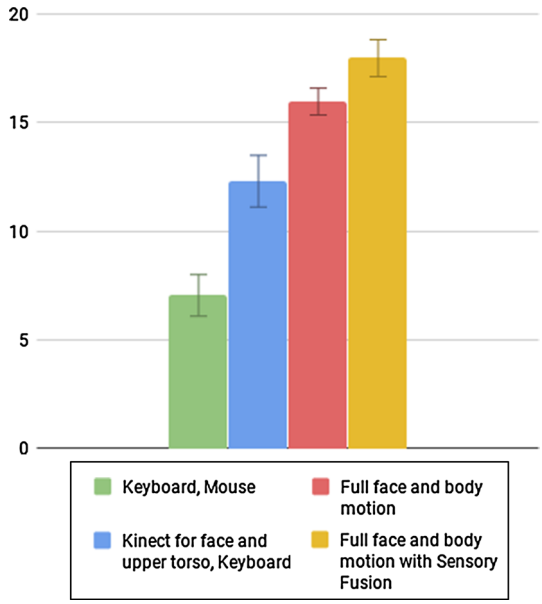

Fig. 14 Mean and deviation of cumulative questionnaire response scores for attention allocation

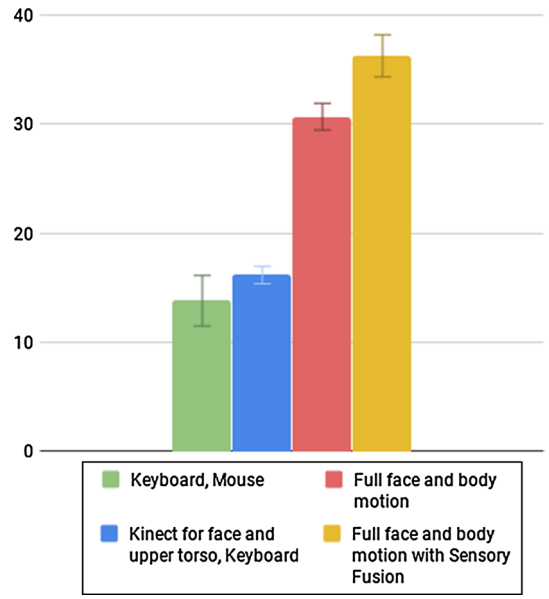

Fig. 13 Mean and deviation of cumulative questionnaire response scores for mutual awareness

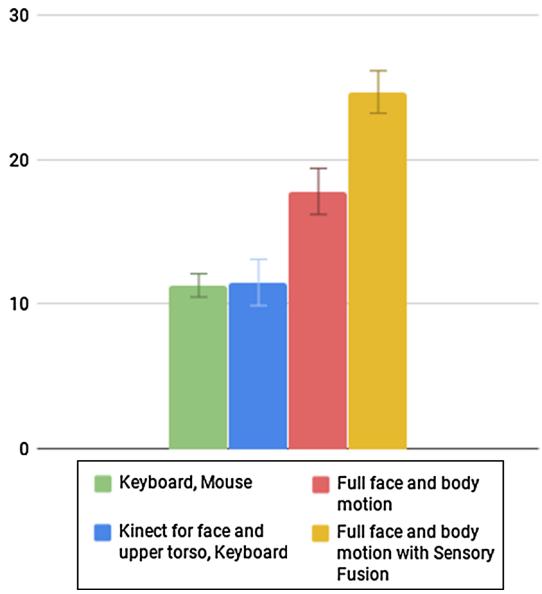

Fig. 15 Mean and deviation of cumulative questionnaire response scores for empathy

5.1 Analysis

For each interaction mode, we measured the presence score (P), the co-presence score (CO-P) , and the immersive tendencies score (IT). We performed a one-way analysis of variance (ANOVA) on interaction mode and each response variable score.

5.1.1 Presence

We compared the difference in the P (Presence) scores between the interaction modes (Table 2) and we found that there was a significant difference at the 0.001 alpha level. The rows indicate user observations and the columns correspond to the independent variable conditions, i.e. rows

Table 2 ANOVA test for presence scores

Source of variation	SS	*df*	MS	*F*	*P* value	*F* crit
ANOVA for P scores						
Rows	4650.95833	23	202.21558	3.5787691	2.18E−05	1.686897
Columns	225,823.208	3	75,274.4028	1332.1907	5.60E−61	2.73749231
Error	3898.79167	69	56.5042271			
Total	234,372.958	95				

Table 3 ANOVA test for co-presence scores

Source of variation	SS	*df*	MS	*F*	*P* value	*F* crit
ANOVA for CO-P						
Rows	1531.625	23	66.592391	12.042918	4.20E−16	1.686897
Columns	198,451.46	3	66,150.486	11963.002	1.23E−93	2.73749231
Error	381.54167	69	5.5295894			
Total	200,364.63	95				

Table 4 ANOVA test for avatar empathy scores

Source of variation	SS	*df*	MS	*F*	*P* value	*F* crit
ANOVA for empathy scores						
Rows	49.65625	23	2.15896739	2.9037564	3.41E−04	1.686897
Columns	3210.94792	3	1070.31597	1439.5478	4.04E−62	2.73749231
Error	51.3020833	69	0.74350845			
Total	3311.90625	95				

Table 5 ANOVA test for mutual awareness scores

Source of variation	SS	*df*	MS	*F*	*P* value	*F* crit
ANOVA for mutual awareness scores						
Rows	111.458333	23	4.8601449	2.6165634	1.12E−03	1.686897
Columns	8603.70833	3	2867.90278	1548.4992	3.39E−63	2.73749231
Error	127.791667	69	1.85205314			
Total	8842.95833	95				

refer to sources of variation within a group and columns correspond to between groups. Thus if *F*-value for columns (between group) is greater than the *F*-critical value for the alpha level selected (0.001), we have evidence to reject the null hypothesis. Likewise, if the *P* value is less than the alpha level selected, we reject the Null Hypothesis. The effect on presence was measured to be ($F1, 24 = 1332$, $p < .001$) which was greater than the *F*-critical value, supporting the rejection of the null hypothesis.

5.1.2 Co-presence

A statistically significant difference was also observed in the CO-P scores (Table 3) across the interaction modes with ($F1, 24 = 11963$, $p < .001$).

As previously stated, the co-presence was divided into Empathy, Mutual Awareness and Attentional Allocation to better relate the specific improvements in co-presence with respect to each interaction mode. A statistically significant difference at the 0.001 alpha level was observed for avatar empathy (Table 4) with ($F1, 24 = 1439$, $p < .001$) between groups or each interaction mode.

Mutual awareness scores were observed to be significant (Table 5) with ($F1, 24 = 1548$, $p < .001$) between groups.

The difference in attentional allocation scores (Table 6) for the players was also a statistically significant at ($F1, 24 = 1645$, $p < .001$). As mentioned before, since *F*-value for columns (between group) is greater than the *F*-critical value and *P* value is less than the selected alpha level (0.001), we have evidence to reject the null hypothesis.

Table 6 ANOVA test for attention allocation scores

Source of variation	SS	*df*	MS	*F*	*P* value	*F* crit
ANOVA for attention allocation scores						
Rows	27.40625	23	1.19157609	3.5033289	2.95E−05	1.68689696
Columns	1678.78125	3	559.59375	1645.2503	4.31E−64	2.73749231
Error	23.46875	69	0.34012681			
Total	1729.65625	95				

Table 7 Correlation matrix for CO-P and P scores with respect to IT scores

	IT	Presence	Co-presence
Keyboard, mouse			
ImmTend	1		
Presence	0.4168244	1	
Co-presence	0.0897477	0.12492112	1
Kinect for face and upper torso, keyboard			
ImmTend	1		
Presence	0.1560843	1	
Co-presence	− 0.195071	0.1315403	1
Full face and body motion			
ImmTend	1		
Presence	0.2955511	1	
Co-presence	0.0009021	0.01407258	1
Full face and body motion with sensory fusion			
ImmTend	1		
Presence	0.2412165	1	
Co-presence	0.0431559	0.1014519	1

Thus, we can conclude that there is a statistically significant difference in the attentional allocation scores between the interaction modes, which shows an increasing trend with interaction fidelity, as observed in Figs. 11 and 14.

5.1.3 Immersive tendencies

A correlation analysis was performed on the P, CO-P, and IT variables, and no significant relationships between were observed between them (Table 7). At a significance level of 0.001, with $n = 24$, we observed a Pearson's r correlation coefficient of 0.41 for P while for CO-P, it was found to be negligible at 0.08 for the Keyboard and mouse mode. Similarly, a Pearson's r value of 0.15 for P and negligible for CO-P was observed for upper torso embodiment + keyboard control. Pearson's r of 0.29 (P) and 0 (CO-P) for full face and body motion and finally, 0.24 (P) and 0.04 (CO-P) for full face and body motion with sensory fusion were observed, respectively. All values point to almost no or very weak-correlation between the immersive tendencies score and presence and co-presence scores. The corresponding scores for all modes are presented in the table below.

5.2 Discussion

The results show that there was a significant difference in the co-presence scores and presence scores with increasing interaction fidelity, i.e. interaction modes with the Kinect for Windows v2 using full-body immersion for embodied interactions and additional sensory fusion audio input yielded the highest scores, which was picked up by the co-presence and presence questionnaires. This supports our hypothesis that increasing the avatar's functionality through a higher interaction fidelity results in increasing presence. This may be explained by the fact that since the high-collaboration interaction exercise was more challenging, it required the participants to be more involved in the experience and hence enhanced the sense of personal presence. This also supports previous work suggesting behavioural fidelity should be prioritized over visual fidelity in the development of expressive avatars. Our study also shows that improvements in behavioural fidelity benefit the constant low fidelity avatars regardless of their appearance. The Co-Presence and Presence scores were also observed to be the highest in the interaction exercises with sensory fusion, which support the second hypothesis (Table 7).

6 Conclusion

We have designed a new platform for immersive performance-centric interaction inspired by the success of Minecraft and builds on its approach to sidestep the uncanny valley effect based on suggestions from existing literature. Our results so far are promising and we were able to create high level of immersion by combining cartoon-like low fidelity visuals and multiple interaction techniques into a single system. We have also extended sophisticated technology like immersive VR and gesture tracking to easy markerless motion capture for performers to control their avatar

with relative ease and accuracy without extended training sessions.

7 Recognition and broader impact

Cinemacraft was showcased as part of two high profile exhibitions with two additional opportunities pending. The team has used such opportunities to iteratively improve upon and refine the design, as informed by the outcomes demonstrations and real-world user feedback. In particular, the prototype was showcased at Virginia Tech's official exhibit at South by Southwest 2016 (Tech 2016), and as part of ICAT day showcase at the Moss Arts Center in Virginia Tech (Institute for Creativity and Technology 2016). More recently, Cinemacraft has also been chosen to be displayed at the Science Museum of Southwest Virginia (Tech 2017). As a result of the strong response to and interest in the tool, it has also been selected to be integrated in the Virginia Tech Visitor Center. Both exhibits are scheduled to open in the winter of 2017.

Cinemacraft also offers opportunities to extend virtual presence and consequently outreach by allowing audiences to engage with the production directly in game. The team envisions the ensuing implementation being appropriate in a broad range of live and post-production scenarios, beyond its original intent, from machinima movie-making to theatre. Studies have shown that learning movie and theatre production skills help to instill the sense of ownership, confidence and self-belief in students (Swainston et al. 2015). Minecraft has already been effectively used as an education tool through the successful MinecraftEdu (Microsoft 2016). Our shift towards the Minetest platform and an increased reliance on sensory fusion has a potential to improve such an educational experience as a live learning tool.

Appendix 1 : Questionnaires

All 24 participants recorded their responses to each experimental interaction exercise using a presence questionnaire based on Slater's (Slater 1999) presence questionnaire, a co-presence questionnaire based on the Networked Minds (Biocca and Harms 2001) and Nowak's (Nowak and Biocca 2003) co-presence questionnaires and finally, the immersive tendencies questionnaire (Witmer and Singer 1998).

1.1 Presence questionnaire

Please rate your sense of being in the virtual environment, on a scale of 1 to 7, where 7 represents your normal experience of being in a place. How much were you able to control events? How responsive was the environment to actions that you initiated (or performed)? How natural did your interactions with the environment seem? How completely were all of your senses engaged? How much did the visual aspects of the environment involve you? How much did the auditory aspects of the environment involve you? How natural was the mechanism which controlled movement through the environment? How aware were you of events occurring in the real world around you? How aware were you of your display and control devices? How compelling was your sense of objects moving through space? How inconsistent or disconnected was the information coming from your various senses? How much did your experiences in the virtual environment seem consistent with your real-world experiences ? Were you able to anticipate what would happen next in response to the actions that you performed ? How completely were you able to actively survey or search the environment using vision? How well could you identify sounds? How well could you localize sounds? How compelling was your sense of moving around inside the virtual environment? How closely were you able to examine objects? How well could you examine objects from multiple viewpoints? How well could you move or manipulate objects in the virtual environment? To what degree did you feel confused or disoriented at the beginning of breaks or at the end of the experimental session? How involved were you in the virtual environment experience? How distracting was the control mechanism? How much delay did you experience between your actions and expected outcomes? How quickly did you adjust to the virtual environment experience? How proficient in moving and interacting with the virtual environment did you feel at the end of the experience? How much did the visual display quality interfere or distract you from performing assigned interaction exercises or required activities? How much did the control devices interfere with the performance of assigned interaction exercises or with other activities? How well could you concentrate on the assigned interaction exercises or required activities rather than on the mechanisms used to perform those interaction exercises or activities? Did you learn new techniques that enabled you to improve your performance? Were you involved in the

experimental interaction exercise to the extent that you lost track of time? To what extent were there times during the experience when the virtual environment was the reality for you? When you think back to the experience, do you think of the virtual environment more as images that you saw or more as somewhere that you visited?

1.2 Co-presence questionnaire

I often felt as if I was all alone. I think the other individual often felt alone. I hardly noticed another individual. The other individual didn't notice me in the room. I was often aware of others in the environment. Others were often aware of me in the room. I think the other individual often felt alone. I often felt as if I was all alone. I sometimes pretended to pay attention to the other individual. The other individual paid close attention to me I paid close attention to the other individual. My partner was easily distracted when other things were going on around us. I was easily distracted when other things were going on around me When I was happy, the other was happy. When the other was happy, I was happy. My interaction partner seemed to find our interaction stimulating. My interaction partner communicated coldness rather than warmth. My interaction partner seemed detached during our interaction. My interaction partner was unwilling to share personal information with me. My interaction partner created a sense of closeness between us. My interaction partner was interested in talking to me. I wanted to maintain a sense of distance between us. I was interested in talking to my interaction partner I perceive that I am in the presence of another person in the room with me. I feel that the person is watching me and is aware of my presence. The thought that the person is not a real person crossed my mind often. The person appears to be sentient (conscious and alive) to me. I perceive the person as being only a computerized image, not as a real person.

1.3 Immersive tendencies questionnaire

Do you easily become deeply involved in movies or tv dramas? Do you ever become so involved in a television program or book that people have problems getting your attention? How mentally alert do you feel at the present time? Do you ever become so involved in a movie that you are not aware of things happening around you? How frequently do you find yourself closely identifying with the characters in a story line? Do you ever become so involved in a video game that it is as if you are inside the game rather than moving a joystick and watching the screen? How physically fit do you feel today? How good are you at blocking out external distractions when you are involved in something? When watching sports, do you ever become so involved in the game that you react as if you were one of the players? Do you ever become so involved in a daydream that you are not aware of things happening around you? Do you ever have dreams that are so real that you feel disoriented when you awake? When playing sports, do you become so involved in the game that you lose track of time? How well do you concentrate on enjoyable activities? How often do you play arcade or video games? (OFTEN should be taken to mean every day or every two days, on average.)

Appendix 2: List of body expressions

Participants were given the interaction exercise sheets for each experimental interaction exercise and were expected to enact out and guess the body expressions within a fixed amount of time for each experimental interaction exercise, without stating or explicitly alluding to the caption on the list. The expectation was that the number of body expressions successfully guessed and enacted from their designated lists, would increase in interaction exercises with higher interaction fidelity.


Balanced
Beneudent
Brilliant
Bright
Brave
Brainy
Blessed
Big-Hearted
Bubbly
Blissful
Boastful
Liar
Self Satisfaction
Exaggerate
Shy
Arrogant
Kind

References

Ahmaniemi T (2010) Gesture controlled virtual instrument with dynamic vibrotactile feedback. In: NIME, pp 485–488

Albaum G (1997) The likert scale revisited. Mark Res Soc J 39(2):1–21

Anderson A, Dossick CS, Iorio J, Taylor JE (2017) The impact of avatars, social norms and copresence on the collaboration effectiveness of aec virtual teams. J Inf Technol Constr (ITcon) 22(15):287–304

Animesh SB, Pinsonneault OH (2011) An odyssey into virtual worlds: exploring the impacts of technological and spatial environments on intention to purchase virtual products. Mis Q 35(3):789–810

Argelaguet F, Hoyet (2016) The role of interaction in virtual embodiment: effects of the virtual hand representation. In: Virtual reality (VR), 2016 IEEE. IEEE, pp 3–10

Bailey J, Bailenson JN, Won AS, Flora J, Armel KC (2012) Presence and memory: immersive virtual reality effects on cued recall. In: Proceedings of the international society for presence research annual conference, Oct, Citeseer, pp 24–26

Bainbridge WS (2007) The scientific research potential of virtual worlds. Science 317(5837):472–476

Barnes B, Elsi G, Kiseleva M (2016) Cinemacraft: virtual minecraft presence using operacraft. Inst Creat Arts Technol (ICAT), pp 11–32

Bianchi-Berthouze N (2013) Understanding the role of body movement in player engagement. Hum Comput Interact 28(1):40–75

Bianchi-Berthouze N, Kim WW, Patel D (2007) Does body movement engage you more in digital game play? and why? In: International conference on affective computing and intelligent interaction. Springer, pp 102–113

Biocca F, Harms (2001) The networked minds measure of social presence: pilot test of the factor structure and concurrent validity. In: 4th annual international workshop on presence, Philadelphia, pp 1–9

Bray DA, Konsynski BR (2007) Virtual worlds: multi-disciplinary research opportunities. SIGMIS Database 38(4):17–25. https://doi.org/10.1145/1314234.1314239

Buecheler C (2010) Character: the next great gaming frontier? http://www.chicagotribune.com/sns-gamereview-feature-characters-story.html. Accessed 22 July 2018

Bukvic I (2012) A behind-the-scenes peek at world's first linux-based laptop orchestra–the design of l2ork infrastructure and lessons learned. In: Linux audio conference, Stanford, California, pp 55–60

Bukvic II, Cahoon C, Wyatt A (2014) Operacraft: blurring the lines between real and virtual. In: ICMC, pp 6–7

Carey B, Ulas B (2016) Vr'space opera': mimetic spectralism in an immersive starlight audification system. arXiv preprint arXiv:161103081, pp 4–5

Carlson PJ, Davis GB (1998) An investigation of media selection among directors and managers: from“self” to“other” orientation. MIS Q 22(3):335–362. https://doi.org/10.2307/249669

Choney S (2016) Microsoft stores offering free minecraft vr demos on oculus rift. https://blogs.microsoft.com/firehose/2016/09/29/microsoft-stores-offering-free-minecraft-vr, http://dl.acm.org/citation.cfm?id=2910632. Accessed 4 Oct 2018

Collingwoode-Williams T, Gillies M (2017) The effect of lip and arm synchronization on embodiment: a pilot study. In: Virtual reality (VR), 2017 IEEE. IEEE, pp 253–254

Denisova A, Cairns P (2015) First person vs. third person perspective in digital games: do player preferences affect immersion? In: Proceedings of the 33rd annual ACM conference on human factors in computing systems. ACM, pp 145–148

Duncan SC (2011) Minecraft, beyond construction and survival. Well Played J Video Games Value Mean 1(1):1–22

Fritsch T, Ritter H, Schiller J (2005) The effect of latency and network limitations on mmorpgs: a field study of everquest2. In: Proceedings of 4th ACM SIGCOMM workshop on Network and system support for games. ACM, pp 1–9

FUDI (2017) Fudi. https://en.wikipedia.org/wiki/FUDI. Accessed 1 June 2018

Garrelts N (2014) Understanding Minecraft: essays on play, community and possibilities. McFarland, Jefferson

Heidicker P, Langbehn E, Steinicke F (2017) Influence of avatar appearance on presence in social vr. In: IEEE symposium on 3D user interfaces (3DUI), 2017. IEEE, pp 233–234

Hirose M, Schmalstieg D, Wingrave CA, Nishimura K (2009) Higher levels of immersion improve procedure memorization performance. In: Proceedings of the 15th joint virtual reality eurographics conference on virtual environments, pp 121–128

Huh Y, Duarte GT, El Zarki M (2018) Minebike: Exergaming with minecraft. In: 2018 IEEE 20th International conference on e-health networking, applications and services (Healthcom). IEEE, pp 1–6

Institute for Creativity A, Technology (2016) Icat day 2016. https://www.icat.vt.edu/content/icat-day-2016

Kastelein R (2013) The rise of machinima, the artform. http://insights.wired.com/profiles/blogs/the-rise-of-machinima-the-artform. Accessed 12 May 2018

Kätsyri J, de Gelder B (2018) Uncanny slope instead of an uncanny valley: testing the uncanny valley hypothesis in painted, computer-rendered, and human faces, pp 4–9

Kinect M (2017) Kinect kinect. http://www.xbox.com/en-US/xbox-one/accessories/kinect. Accessed 2 Jan 2019

Kokkinara E, Slater M (2015) The effects of visuomotor calibration to the perceived space and body, through embodiment in immersive virtual reality. ACM Trans Appl Percept (TAP) 13(1):3

Lay S, Brace N, Pike G, Pollick F (2016) Circling around the uncanny valley: design principles for research into the relation between human likeness and eeriness. i-Perception 7(6):2–6. https://doi.org/10.1177/2041669516681309

Lecuyer A (2017) Playing with senses in vr: alternate perceptions combining vision and touch. IEEE Comput Gr Appl 37(1):20–26. https://doi.org/10.1109/MCG.2017.14

Lombard M, Ditton T (1997) At the heart of it all: the concept of presence. J Comput Mediat Commun 3(2):0–0

Maister L, Slater M (2015) Changing bodies changes minds: owning another body affects social cognition. Trends Cogn Sci 19(1):6–12

Makled E, Abdelrahman (2018) I like to move it: investigating the effect of head and body movement of avatars in vr on user's perception. In: Extended abstracts of the 2018 CHI conference on human factors in computing systems, ACM, New York, CHI EA '18, pp LBW099:1–LBW099:6. https://doi.org/10.1145/3170427.3188573

Mäntymäki M, Riemer K (2011) How social are social virtual worlds? an investigation of hedonic, utilitarian, social and normative usage drivers. In: PACIS, p 126

Microsoft (2016) Impact minecraft education edition is making in classrooms. https://education.minecraft.net/. Accessed 2 August 2018

Microsoft (2017a) Kinect 360. https://support.xbox.com/en-US/xbox-on-windows/accessories/kinect-for-windows-info. Accessed 3 July 2018

Microsoft (2017b) Windows presentation foundation (wpf) is a next-generation presentation system for building windows client applications. https://msdn.microsoft.com/en-us/library/aa970268(v=vs.100).aspx. Accessed 3 July 2018

Minetest (2016) Meet minetest. http://www.minetest.net/. Accessed 6 August

Mori M, MacDorman KF, Kageki N (2012) The uncanny valley [from the field]. IEEE Robot Autom Mag 19(2):98–100
Mousas C, Anastasiou D, Spantidi O (2018) The effects of appearance and motion of virtual characters on emotional reactivity. Comput Hum Behav 86:99–108
Nah FFH, Eschenbrenner B, DeWester D (2011) Enhancing brand equity through flow and telepresence: a comparison of 2d and 3d virtual worlds. MIs Q 35(3):731–747
Narang S, Best A, Manocha D (2018) Simulating movement interactions between avatars & agents in virtual worlds using human motion constraints. In: 2018 IEEE conference on virtual reality and 3D user interfaces (VR). IEEE, pp 9–16
Nash EB, Edwards GW, Thompson JA, Barfield W (2000) A review of presence and performance in virtual environments. Int J Hum Comput Interact 12(1):1–41
Nowak KL, Biocca F (2003) The effect of the agency and anthropomorphism on users' sense of telepresence, copresence, and social presence in virtual environments. Presence Teleoper Virtual Environ 12(5):481–494
Park N, Lee KM, Jin SAA, Kang S (2010) Effects of pre-game stories on feelings of presence and evaluation of computer games. Int J Hum Comput Stud 68(11):822–833
Parker JR (2008) Buttons, simplicity, and natural interfaces. Loading 2(2)
Polyak E (2012) Virtual impersonation using interactive glove puppets. In: SIGGRAPH Asia 2012 posters, ACM, New York, SA '12, pp 31:1–31:1. https://doi.org/10.1145/2407156.2407191
Polys NF, Knapp B, Bukvic I (2015) Fusality: an open framework for cross-platform mirror world installations. In: Proceedings of the 20th international conference on 3D web technology. ACM, pp 171–179
Ragan ED, Sowndararajan A, Kopper R, Bowman DA (2010) The effects of higher levels of immersion on procedure memorization performance and implications for educational virtual environments. Presence Teleoper Virtual Environ 19(6):527–543
Ratcliffe J (2014) Hand motion-controlled audio mixing interface. Proc New Interfaces Musical Expr (NIME) 2014:136–139
Roth D, Lugrin JL, Galakhov D, Hofmann (2016) Avatar realism and social interaction quality in virtual reality. In: Virtual reality (VR), 2016 IEEE. IEEE, pp 277–278
Sacau A, Laarni J (2008) Influence of individual factors on presence. Comput Hum Behav 24(5):2255–2273
Saunders C, Rutkowski AF, van Genuchten M, Vogel D, Orrego JM (2011) Virtual space and place: theory and test. MIS Q 35(4):1079–1098
Schroeder R (2012) The social life of avatars: presence and interaction in shared virtual environments. Springer, Berlin
Schroeder R, Steed (2001) Collaborating in networked immersive spaces: as good as being there together? Comput Gr 25(5):781–788
Schultze U (2011) The avatar as sociomaterial entanglement: a performative perspective on identity, agency and world-making in virtual worlds. In: Proceedings of the international conference on information systems, ICIS 2011, Shanghai, China
Schultze, Orlikowski (2010) Virtual worlds: a performative perspective on globally distributed, immersive work. Inf Syst Res 21(4):810–821
Seymour M, Riemer K, Kay J (2017) Interactive realistic digital avatars–revisiting the uncanny valley. In: Hawaii international conference on system sciences, HICSS-50, Honolulu
Seymour M, Riemer K, Kay J (2018) Actors, avatars and agents: potentials and implications of natural face technology for the creation of realistic visual presence. J Assoc Inf Syst 19(10):953–981
Shin D (2018) Empathy and embodied experience in virtual environment: to what extent can virtual reality stimulate empathy and embodied experience? Comput Hum Behav 78:64–73
Sia CL, Tan BC, Wei KK (2002) Group polarization and computer-mediated communication: effects of communication cues, social presence, and anonymity. Inf Syst Res 13(1):70–90
Sims K (1994) Evolving virtual creatures. In: Proceedings of the 21st annual conference on computer graphics and interactive techniques. ACM, pp 15–22
Slater M (1999) Measuring presence: a response to the witmer and singer presence questionnaire. Presence Teleoper Virtual Environ 8(5):560–565
Slater M, Sadagic (2000) Small-group behavior in a virtual and real environment: a comparative study. Presence Teleoper Virtual Environ 9(1):37–51
Slater M, Spanlang B, Sanchez-Vives MV, Blanke O (2010) First person experience of body transfer in virtual reality. PloS one 5(5):e10564
Spante M, Heldal (2003) Is there a tradeoff between presence and copresence. In: Proceedings of presence 2003: 6th international workshop on presence
Swainston A, Jeanneret N, et al (2015) Wot opera: a joyful, creative and immersive experience. In: Music: educating for life. ASME XXth national conference proceedings, australian society for music education, p 99
Tech V (2016) South by southwest 2016. http://www.vt.edu/sxsw.html. Accessed 7 May 2018
Tech V (2017) Science museum of western virginia. http://www.icat.vt.edu/smwv. Accessed 7 May 2018
Thon J-N (2008) Immersion revisited: on the value of a contested concept. In: Leino O, Wirman H, Fernandez A (eds) Extending experiences: structure, analysis and design of computer game player experience. Lapland University Press, Lapland, pp 29–43
Tinwell A, Grimshaw M, Nabi DA, Williams A (2011) Facial expression of emotion and perception of the uncanny valley in virtual characters. Comput Hum Behav 27(2):741–749
Viniconis N (2011) Minecraft + kinect : building worlds! http://www.orderofevents.com/MineCraft/KinectInfo.htm. Accessed 14 Oct 2018
Vivecraft (2016) Vivecraft. http://www.vivecraft.org/. Accessed 24 Oct 2018
Witmer BG, Singer MJ (1998) Measuring presence in virtual environments: a presence questionnaire. Presence Teleoper Virtual Environ 7(3):225–240
Wright M, Freed A, Momeni A (2003) Opensound control: state of the art 2003. In: Proceedings of the 2003 conference on new interfaces for musical expression. National University of Singapore, pp 153–160
Yoo Y, Alavi M (2001) Media and group cohesion: relative influences on social presence, task participation, and group consensus. MIS Q 25(3):371–390
Zhu L, Benbasat I, Jiang Z (2010) Let's shop online together: an empirical investigation of collaborative online shopping support. Inf Syst Res 21(4):872–891