You as a Puppet: Evaluation of Telepresence User Interface for Puppetry

Mose Sakashita
University of Tsukuba
mose.sakashita@gmail.com

Tatsuya Minagawa
University of Tsukuba
minatatu0901.happy@gmail.com

Amy Koike
University of Tsukuba
amy23kik@gmail.com

Ippei Suzuki
University of Tsukuba
1heisuzuki@gmail.com

Keisuke Kawahara
University of Tsukuba
kawahara@ai.iit.tsukuba.ac.jp

Yoichi Ochiai
University of Tsukuba
wizard@slis.tsukuba.ac.jp

Figure 1. (a) Synchronization between a performer and a puppet with our system. (b) Users manipulating puppets with our system. (c) Puppet-show using telepresence user interface for a puppet.

ABSTRACT
We propose an immersive telepresence system for puppetry that transmits a human performer's body and facial movements into a puppet with audiovisual feedback to the performer. The cameras carried in place of puppet's eyes stream live video to the HMD worn by the performer, so that performers can see the images from the puppet's eyes with their own eyes and have a visual understanding of the puppet's ambience. In conventional methods to manipulate a puppet (a hand-puppet, a string-puppet, and a rod-puppet), there is a need to practice manipulating puppets, and there is difficulty carrying out interactions with the audience. Moreover, puppeteers must be positioned exactly where the puppet is. The proposed system addresses these issues by enabling a human performer to manipulate the puppet remotely using his or her body and facial movements. We conducted several user studies with both beginners and professional puppeteers. The results show that, unlike the conventional method, the proposed system facilitates the manipulation of puppets especially for beginners. Moreover, this system allows performers to enjoy puppetry and fascinate audiences.

UIST 2017, October 22–25, 2017, Quebec City, QC, Canada.

ACM ISBN 978-1-4503-4981-9/17/10$15.00
https://doi.org/10.1145/3126594.3126608

ACM Classification Keywords
H.5.2 Information interfaces and presentation (e.g.,HCI): User Interfaces

Author Keywords
Animatronics; puppet; telepresence; head-mounted display.

INTRODUCTION
Puppets and characters are used in various contexts, as communication tools among family [28], education tools [8, 23], TV-shows like "The Muppets" produced by Jim Henson [5], attractions or shows in theme parks, and so on. When a puppet expresses thoughts and feelings as a personality, it is easy to recognize it as a 'real' person, i.e., a person with whom we can communicate. The idea that an object can become alive and lifelike has always attracted people and challenged their imagination. In a long history of puppetry, which is thought to have originated in 3000 BC [1], various methods for manipulating puppets have been developed. Some simple ways to perform puppetry include moving puppets by wearing a puppet directly on the hands (glove-puppet) and connecting strings or rods between it and the puppeteer's hands or arms (string-puppet, rod-puppet). In most cases, performers manipulate puppets with their hands above a table while the rest of their body is below it. This poses a challenge in coordinating the actions. Consequently, puppetry requires considerable training to manipulate puppets.

Many user interfaces have been proposed that expand the rage of methods used to manipulate puppets computationally. For example, research into puppet manipulation has recently

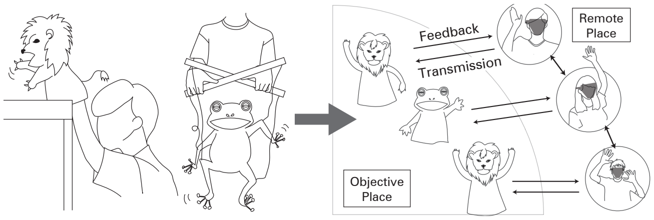

Figure 2. Conventional method to manipulate puppet (left) vs. concept of our system (right).

collaborated with the field of telepresence. The term "telepresence" refers to technologies that enable users to get a sense of being in a different place. These studies pertain to the manipulation of remote puppets for the sake of communication or education, rather than focusing on performances like puppet-shows as entertainment. Therefore, the parameters transmitted to the puppet for its manipulation are insufficient for an abundance of animated expressions and movement. Besides, they utilize 2D-displays such as tablets or smartphones for manipulation rather than using head-mounted displays (HMDs). These systems have an absence of immersion in a puppet.

The concept of human-to-nonhuman immersive telepresence (Figure 2) affords new experiences to performers and observers, with the potential for novel applications. The conventional method for manipulating a puppet has physical limitations: puppeteers need to be positioned and have to perform exactly where the puppet is situated. In theme parks, actors are required to be inside costumes to change expressions or motions. When communicating with children or grandchildren through storytelling using a puppet, their parents and grandparents must be present with them. Combining puppet-shows and telepresence technology can remove the physical limitation, and puppets can be manipulated at a place where the puppeteers want to be.

The telepresence system proposed in this paper transmits the performer's body and facial movements to a puppet and provides visual and audible feedback to the performer (Figure 1). The cameras located in place of the puppet's eyes stream live video to the HMD worn by the performer. Hence, performers are able to see the images from the puppet's eyes for themselves and have a visual understanding of the puppet's environment. This immersive experience enables performers to act like a puppet, since performers can make eye contact with the audience and other puppets. This allows users to enjoy and perform more lifelike characters, as if they themselves have become the puppet. Moreover, beginners who are not trained to manipulate puppets can also perform puppet shows in simple ways by using the proposed system, whereas intricate skill is needed to manipulate puppets conventionally.

We investigated the effectiveness of the proposed telepresence user interface for puppetry in terms of its ability to both manipulate a puppet (performer side) and facilitate a puppet-show (audience side). For performers, it is crucial that the system is useful and enjoyable, and that it is not overly exhausting for the performer. In terms of the audience, it is important that the puppetry facilitated through the telepresence system is fun and lifelike.

Therefore, in this paper, we consider the effectiveness of our system by evaluating the proposal based on familiarity with puppetry. We conducted several user studies with both beginners and professional puppeteers. This enabled evaluation of both the advantages and limitations of a telepresence system for puppetry. The results of our user study show that the proposed system facilitates the manipulation of puppets, especially for beginners. As we demonstrate, the system allows performers to enjoy performing with a puppet. We discuss the limitations of the proposed system, and conclude with possibilities for future work, both in our application and in telepresence technology beyond puppetry.

RELATED WORK

Computational Puppet

Several studies have been conducted on the manipulation of physical puppets, dolls, and toys. PINOKY [33] is a ring-like device that can be used to animate a toy by moving its limbs. Moreover, in the context of costumes in theme parks, a "Tongue" joystick device was proposed for use inside an articulated-head character costume [32], and "Mimicat" enabled a costume's facial expressions to be synchronized with a performer's facial action [30]. Several other related studies have been proposed, including the following: an expressive musical doll capable of conversation [36]; and a system for animatronic storytelling that enables performers to manipulate puppets by wearing a mask-type device on their faces [9, 27].

Input systems for animating a puppet on a display have also been proposed. These include: a hand-manipulated interactive motion control interface that utilizes LeapMotion [24]; a system for recording the movements of puppets using a webcam [31]; a grove-type user interface for interactively animating a puppet [25]; an ergonomic hand-mapping system for digital puppetry [13]; and a controller to create digital puppetry using body motion [12].

Facial Recognition

There have also been studies pertaining to the detection of facial movements and expressions, with some based on depth maps [14] and others based on video facial tracking [15]. A wearable-type approach that uses EMG signals was proposed by Suzuki et al. [6]. A glasses-type device [18] is another wearable method that uses photo-reflective sensors. Accordingly, many works on facial representation, tracking, mapping, and animation have been developed, such as [17]. Moreover, there have been collaborations between puppetry and technology for sensing facial expressions and movements: Voice Puppetry [3] enables the faces in pictures or drawings to move according to the user's voice; and Face/Off [34] enables high-resolution real-time facial-expression tracking that is transferred to another person's face. However, in most studies of facial sensing, facial expressions and animations are generated in 2D or 3D and exclusively displayed on screens. We developed an application that associates the captured facial expressions with a real puppet's face in real-time by moving the puppet robot. In the field of virtual reality, the use of HMDs has been ubiquitous. With such devices, however, there is difficulty tracking and capturing facial expressions. Indeed, the user's face must

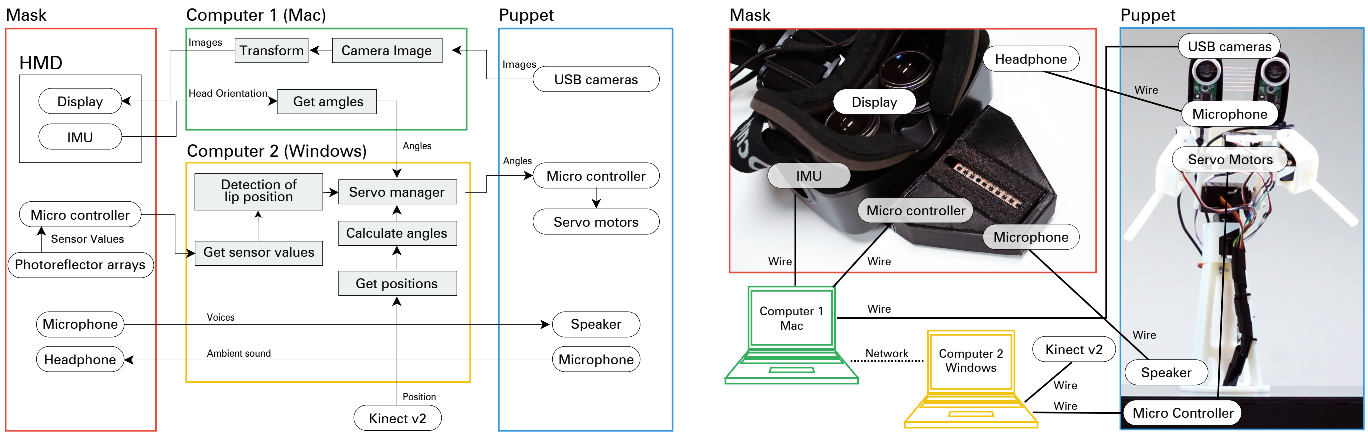

Figure 3. Configuration of the proposed system.

be fully visible in almost all previous facial-tracking systems. Nevertheless, one solution has recently been proposed for eye-tracking under an HMD: [16] uses RGB-D cameras to generate an animation of the entire face. However, the device is heavy and uncomfortable to wear, insofar as it includes both a depth sensor and an HMD. Given that our proposal uses an HMD to provide the user with visual feedback, we considered a mask-type device with a photoreflector sensor module to capture facial expressions. Indeed, our mask device is lighter than a depth camera, and it can process sensor data at higher speeds than image processing.

Telepresence

Telepresence refers to technology that enables users to experience the feeling of being in a different place. A remote user is connected to a removed environment in order to communicate and interact with people in that environment. A number of systems have been proposed to achieve telepresence after Marvin Minsky articulated the theory [19]. For instance, LiveSphere [21] is an attempt at human-to-human telepresence with a wearable omnidirectional camera, and ChameleonMask [20] allows one user to act as a surrogate for another, by displaying the face of a remote user onto the surrogate.

Some teleoperation robots have also been proposed, which allow users to control robots from remote places [4, 35]. Human-like remote-controlled android robots have also been developed [26]. Likewise, research has been conducted into the combination of non-humanlike puppet manipulation and telepresence. In particular, [11] is a robotic teddy bear designed for social communication applications. It can be moved from remote places using an input device such as a Wii remote controller. Users can receive visual and auditory feedback and communicate with local people. The RobotPhone RUI [29] allows the user to modify the shape of connected robotic components. Motion directed to these components can be sent via the Internet to other RobotPhones. PlayPals [2] is a wireless system that provides children with a playful way to communicate between remote locations. The above works aim to transmit particular user information to a puppet. However, in all of these works, the manipulation of a remote puppet is designed for communication or education. They are not focused on entertainment, such as a puppet-show. Moreover, these systems do not provide an immersive experience. Our proposal transmits the performer's voices, along with body, facial, and mouth movements, to a puppet and provides visual and audible feedback to the performer. In effect, the performer's presence is transmitted to a puppet. All parameters are transmitted to the puppet merely by moving the body and face. As such, it is easy to manipulate the puppet and there is no need for intricate skill to do so.

IMPLEMENTATION

In this section, we describe the implementation of the system. The configuration of the system is shown in Figure 3. Our mask-type device consists of an HMD [22], a microphone to capture the performer's voice, and photoreflector arrays to detect movements of the mouth, which are then transmitted to the puppet. We used a Kinect camera to capture the movements of the performer. The puppet robot has a microphone that captures ambient sound, and two cameras and seven servos to move the arms, mouth, and neck.

HMD and Cameras

The robot carries two web cameras with 9.2 million pixels in place of its eyes. These cameras stream live video to the HMD worn by the performer. Therefore, performers are able to see the images from the puppet's eyes with their own eyes in order to have a visual understanding of the puppet's environment (see Figure 4).

Figure 4. Interaction with audience (left), visual feedback (center), and two web cameras on the puppet (right).

Head Rotation

The robot's 2-DOF neck is controlled according to performer's facial movements. We used an inertial measurement unit (IMU) in the HMD to acquire values for the gyro sensor, in order to measure the Euler angles of the performer's face.

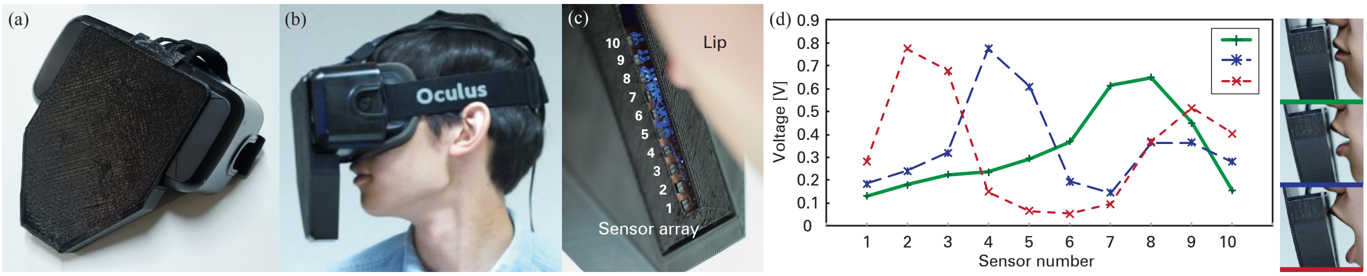

Figure 5. (a)(b) Mask-type headset worn by the performer for manipulation, (c) array of photoreflectors for detecting the mouth state, and (d) measurements of the photoreflector sensor for mouth closed (green), mouth partly open (blue), and mouth open (red).

The Euler angles are sent to the servomotors in the robot's neck using the user datagram protocol (UDP) at 30 fps with software encoded in a microcontroller.

Detection of Mouth State

We developed an array of photoreflectors to detect the open or closed state of the performer's mouth (Figure 5 (c)). The sensors were spaced at 6 mm on a board and connected via a general-purpose input/output (GPIO) to a control photo-emitter and the analog-to-digital converter of the microcontroller. We employed 10 photoreflector sensors (TPR-105; GENIXTEK CORP.) with a maximum sensing distance of 20 mm. Figure 5 (d) shows sample measurements from the photoreflector sensor for mouth closed (green), mouth partly open (blue), and mouth open (red). The program for detecting the lower lip's position was implemented using a Java processor. The software is used to calculate the difference between the measurements of different sensors, with the maximum measurement considered the valid indication of the actual position of the performer's lower lip:

$$\mathrm{Diff}_{value}(i) = value(i+1) - value(i), i = 0 \sim 8, i \in N$$

Based on the detected position, the software controls the mouth servomotor of the puppet. This setup synchronized the mouth movement of the puppet to the speech of the performer, giving the impression that the puppet is actually speaking to the audience. The positions of the sensors on the board can be adjusted to fit different users.

Voice Transmission

Inside the mask, there is a microphone to record the performer's voice. A speaker is installed in the puppet and replays the voice in real-time. The puppet moves its mouth as the performer does, giving the impression that the puppet is speaking to the audience. Furthermore, we can use an effect that changes the voice, depending on the puppet's character. This is designed to make the puppet-show more attractive.

Furthermore, another microphone is employed in the puppet to record the ambient sound from the audience, and outputs this sound to performer's earphones. Thus, the performer is provided with an aural experience of what is happening around the puppet—e.g., the audience's reaction—allowing for interaction with the audience.

Sensing of the Performer's Arms

As an additional function, we used the Kinect depth camera to the capture movements of the user's arms. This was motivated by [7], a proposal for manipulating a puppet with data from the performer's hands captured by Kinect. Our system combines body and facial movements to manipulate puppets and support a smooth synchronization between performers and puppets in puppet-play. The tilt angles and pan directions are calculated from the positions of the arms captured by the sensor. These angles are transferred using UDP at 30 fps to software encoded in a microcontroller. We developed this software using SDK Version 2.0 and Processing 3.0.1. We applied a low-pass filter to stabilize low-resolution and noisy input data from the Kinect sensor. The puppet has two 2-DOF arms, each with two servos. The puppet thus synchronizes the captured motions of the performer's arms.

Mask-Type Device

As shown in Figure 5 (a, b, c), the mask device consists of an HMD and an adaptor. It is used to track the head direction and capture mouth movements of the performer. The video and head orientation are acquired using the HMD. Because the photoreflector arrays capture mouth movements, we designed an adaptor integrated to the headset.

"Bones" of the Puppet

We designed a framework (or "bones") inside the puppet upon which the servos could be attached (Figure 6). The framework was designed such that the servos do not interfere with each other. All components of the puppet's framework were printed with a 3D-printer, meaning that the robot can be designed inexpensively. The puppet robot has two 2-DOF arms, a 1-DOF mouth and a 2-DOF neck. Therefore, the entire body of the puppet has 7-DOFs.

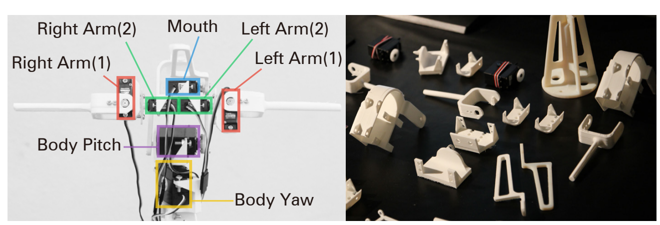

Figure 6. "Bones" and servos in the puppet (left), 3D-printed components (right).

SYNCHRONIZATION AND LATENCY

Our system had a slight time-lag owing to the motors and network connections among software components. To evaluate the accuracy of synchronization between the user and puppet,

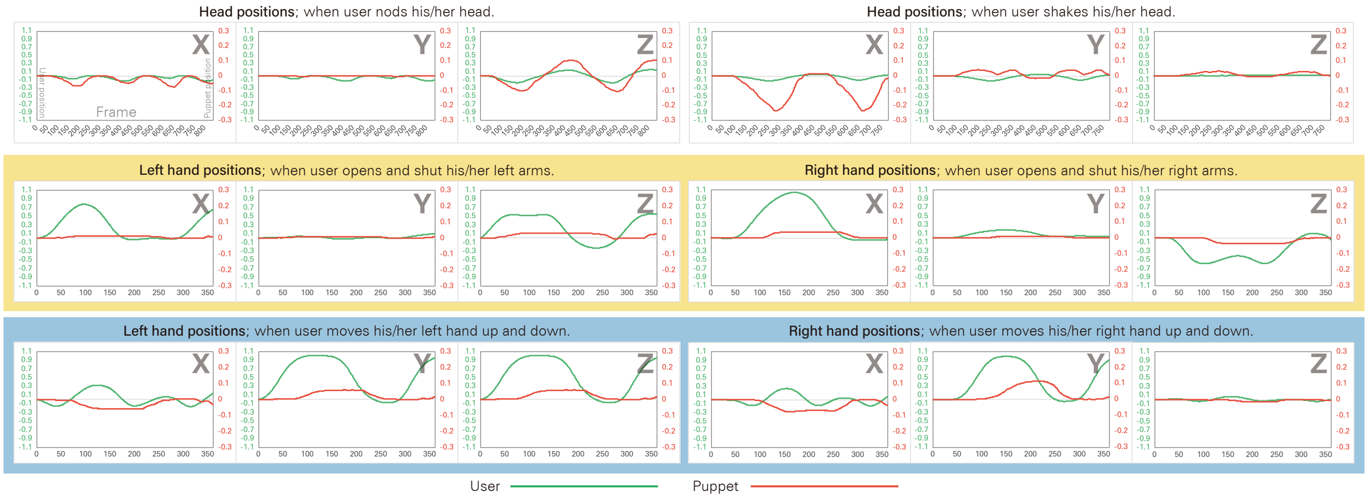

Figure 7. Result from captured 3-axes position of the puppet and user with 4 kinds of movements. Note that the maximum values of the vertical lines differ between the green line (user) and red line (puppet).

we measured the latency of the system with a motion capture system (OptiTrack). We focused on the 2-DOF (yaw and pitch) movement in parts of the body and the 4-DOF (pan and tilt in the left and right) movement in the hands.

Study Process

We set six cameras (four Prime 41, two Prime17W) in a wide room. Markers for tracking the positions were attached to the top of the head and palms of both the puppet and user. We recorded the positional data of the puppet and user at the same time while observing four types of movements (viz., nodding head, shaking head, opening and shutting arms, and moving hands up and down). We calculated the response time of the system by recoding motion data as the user controlled the puppet.

Results

Figure 7 shows the results of the captured positions of the user and puppet with motion capture. Note that the maximum values of the vertical lines differ between the green line (user) and the red line (puppet). Although the distance in the puppet movement is not the same as that of the human—insofar as the scale between the human's body and the puppet's body differs—the results show that the puppet synchronized with the human body's movements. We measured latency in each joint by taking the difference between the user movement's start frame and the movement of the system's frame: the latency of the body was 232.4 ms, the head was 257.3 ms, the right tilt was 639.1 ms, the right pan was 398.4 ms, the left tilt was 664 ms, and the left pan was 589.3 ms. These results show that the system's reaction was within approximately 650 ms.

USER STUDY

The effectiveness of the proposed system was investigated in terms of both manipulating the puppet (performer side), and watching the puppet-show (audience side). In this section, we show the results from two user studies, i.e., the usability of the system, and the comparison of puppetry. For these studies, we used the amplitude of the user's voice to move the puppet's mouth, rather than photoreflectors. We implemented the system in a wired network environment for laptops.

Usability in Manipulating a Puppet

For performers, it is crucial that the system is useful and enjoyable, and that it does not result in fatigue. Thus, we investigated how performers felt when they performed puppetry using our system.

Study Process

We prepared a simple stage with two puppets (a lion and a frog) for puppetry (see Figure 8). Each participant manipulated a puppet freely for approximately five minutes using our system. Ordinary users who were not trained as puppeteers and were not familiar with the experience of manipulating a puppet using a typical method, were allowed a few minutes to manipulate a puppet conventionally, by hand. To understand how useful, enjoyable, or stressful participants felt when they used our system, we asked them to rate each question on a seven-point scale, from very useful/enjoyable/stressful to not at all useful/enjoyable/stressful. In addition, we asked for details with free-description questions. We analyzed the scores of each question on a 7-point Likert scale except for questions about fatigue, which was analyzed using two-tailed Mann–Whitney U tests evaluated at an alpha level of 0.05.

Figure 8. Stage and puppets we prepared for the user study (left), and participants and puppets divided by a wall (right). Note that P and G denote professional and general, respectively.

Participants

Fourteen participants (7 females, 7 males) aged between 18 and 24 years ($M = 20.9$, $SD = 1.46$) participated in this user study. Of the participants, 5 were professional puppeteers, and 9 were general participants who were not trained to manipulate a puppet by hand. We recruited general participants by advertising on a few social network services, and recruited professional participants from a specific professional puppeteer group.

Results

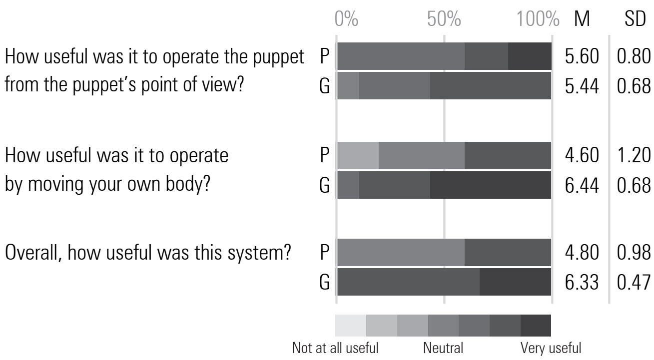

Figure 9. Rating the usefulness of manipulating a puppet using our system. Note that P denotes the professional participants and G denotes the general participants.

To understand the usefulness of our system at manipulating a puppet, we asked participants three questions. Figure 9 shows the results from all three questions.

The first question pertained to operating the puppet from the puppet's point of view. In conventional puppetry, performers can only imagine the puppet's point of view when they manipulate a puppet. Therefore, this question was designed to understand the visual feedback provided by our system.

Professionals gave an average rating of 5.60, and general participants gave an average rating of 5.44. Thus, both participants rated it positively on an average.

We conducted a U test for this question, and no significant difference (U = 22.0, $p > 0.05$) between general (M = 5.44, MD = 6, SD = 0.68) and professional participants (M = 5.60, MD = 5, SD = 0.80) was observed.

According to the participants' free-descriptions, there were positive opinions with reference to benefits of operating the puppet from the puppet's point of view. Specially, opinions were as follows: actors can see exactly what the puppet sees (P4); actors can make it easy to position its line of sight as actors expected (P1, P2, and P4); as performers can see the scene that the puppet sees, they can express feelings (G1), it enables users to behave naturally (G4).

P(professional)1: *"Usually, I manipulate a puppet by imagining the puppet's point of view, but it's difficult to do that, and the puppet's point of view sometimes goes in a different direction, contrary to my intention. So I feel it was useful that I could see things from the puppet's point of view."*

Participants commented on the reasons why it was useful to see from the puppet's point of view. From the above comment by a professional puppeteer, we confirmed that puppeteers usually try to be conscious of the puppet's point of view during performances.

G(general)4: *I felt like I could became a character that was not me. It was similar to the feeling when I wear a costume.*

On the other hand, there were also negative opinions: the viewing angle is narrow (G4 and G5); the video delayed and was blurred (P2, P4, G2, and G7).

G7: *When I moved my body, the video that I saw was shaken a lot and I felt sick.*

The second question pertained to the usefulness of the system in terms of manipulating their bodies. Conventionally, performers use their hands to manipulate a puppet, whereas with our system, users manipulate the puppet by moving their bodies. Unlike with the previous question, most professional participants rated this aspect as neutral.

The U test for this question showed that the difference was significant (U = 5.0, $p \leq 0.05$). The result indicated that the general participants perceived the system to be more useful for manipulating the puppet by moving their bodies (M = 6.44, MD = 7, SD = 0.68) than professional participants did (M = 4.60, MD = 4, SD = 1.20).

P5: *"I could experience it as if I had become the puppet itself. I couldn't understand how much I could move a puppet when I moved my body, but I think it has more potential if the performers practice, since it is easier to use than by hand."*

In puppetry, the appearance of the puppets from the audience's point of view is an important aspect. Our system provides feedback to allow the puppeteer to understand the orientation of the puppet, but it does not provide feedback regarding the movements of the puppet's hands. Therefore, the performers have to move the puppet by inferring its appearance from the point of view of the audience. We believe that the professional puppeteers' rating was neutral because the appearance of the puppet to the audience is important to them, but with our system it must be inferred.

G1: *"When I was asked to manipulate a puppet in a conventional manner, I attempted to move its hands using my thumb and little finger. However, I did not have a technique for moving its hands up and down."*

G8: *"It felt joyful to actually exercise."*

In general, the comments from the general participants offered more subjective opinions compared to those from professionals. The professional participants rated the system more objectively, by considering it from the appearance of the audience. By contrast, the general participants rated it based on their experience of using the system.

The third question we asked participants pertained to the overall usefulness of the system. The U test showed that the difference was significant between general and professional participants (U = 6.0, $p \leq 0.05$), and the result indicated that the general participants perceived the system to be more useful as a rule (M = 6.33, MD = 6, SD = 0.47) than professional participants did (M = 4.80, MD = 4, SD = 0.98). In comparison with the rating of general participants, that of professionals was lower by 1 point. The professional puppeteers pointed out issues with our system, as suggested by the following quotations:

P5: *"This robot cannot move its position, because it doesn't have legs. I will get troubled when I use some props for detailed performance."*

P1: *"Movements seem artificial."*

P2: *"I feel that the mechanical sound that the robot makes when it moves is noisy."*

Overall, the professional puppeteers rated the system neutrally, and we believe that this is because they were good at manipulating a puppet by hand but unfamiliar with our system.

G7: *"With your system, I could express my feelings in the puppet. When I manipulate a puppet directly by hand, it feels like I 'manipulate' a puppet. But with your system, I felt like I became a puppet."*

For general participants, it seemed very natural and useful to manipulate a puppet with our method, and we believe that this is because they were unfamiliar with manipulating puppets conventionally.

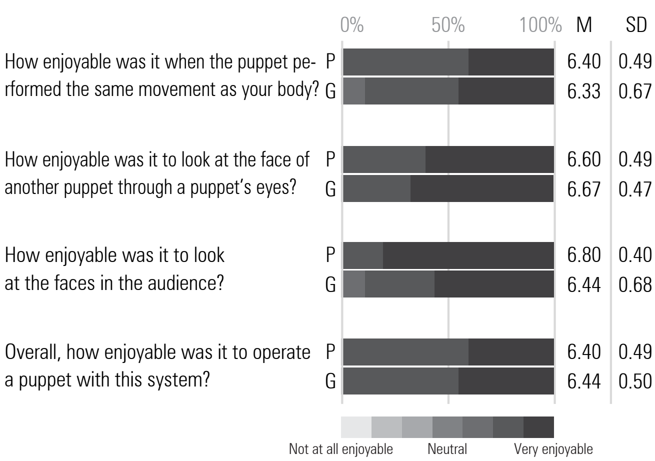

Figure 10. Rating the joy from manipulating a puppet using our system.

As shown in Figure 10, we asked four questions pertaining to the enjoyment felt when participants experienced the proposed system.

First, we asked participants, "How enjoyable was it when the puppet performed the same movement as your body?" The U test for this question showed that the difference was not significant (U = 22.0, $p > 0.05$). The averages of the scores were very high for both general participants (M = 6.33, MD = 6, SD = 0.67) and professionals (M = 6.40, MD = 6, SD = 0.49).

We also asked "How enjoyable was it to look at the faces of another puppet through a puppet's eyes?" For this question also, we found no significant difference between the general and professional participants (U = 21.0, $p > 0.05$). The average scores were very high for both general participants (M = 6.67, MD = 7, SD = 0.47) and professional participants (M = 6.60, MD = 7, SD = 0.49).

Regarding the third question, we found no significant difference (U = 16.5, $p > 0.05$) between general (M = 6.44, MD = 7, SD = 0.68) and professional participants (M = 6.80, MD = 7, SD = 0.40). The U test for the final question showed that the difference was not significant (U = 21.5, $p > 0.05$) between general (M = 6.44, MD = 6, SD = 0.50) and professional participants (M = 6.40, MD = 6, SD = 0.49).

In both professional and general participants, all averages for each item were over 6.0 (for enjoyment). This result shows that all participants felt enjoyment using the system.

P2: *"I enjoyed that the mouth of the puppet moved in accordance with my voice."*

G1: *"It was joyful to see the audience, because people come closer to the robot puppet, and they don't normally do that. This made me laugh when our puppets communicated with each other."*

G7: *"It was good that I could make a voice. Communication between puppets through the system was joyful, and I was able to get along with strangers."*

Participants commented on their enjoyment of the system from different perspectives. One of professional participants enjoyed the experience of having the puppet's mouth synchronized with the user's mouth. In conventional puppetry, it is difficult to move both the mouth and hands at the same time. With the proposed system it was easy to manipulate the puppet's hands and mouth independently. A general participant raised a point about communication as a reason for enjoyment. Even although most participants had not met before, they had conversations in a friendly atmosphere during the user study. Regardless of whether they were professionals, participants seemed to enjoy the user study.

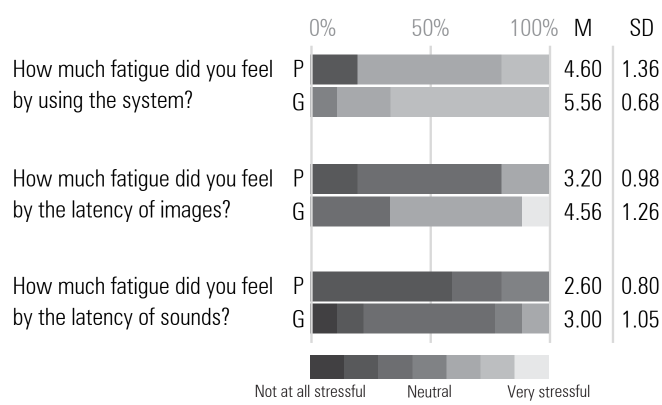

Figure 11. Rating fatigue from manipulating the puppet using our system.

It is well known that exposure to a virtual environment for a long time causes symptoms, called VR sickness, such as motion sickness [10]. This is mainly caused by the discrepancy between visual and bodily sensations. In the proposed system, servo motors move according to values acquired by the IMU, and then images on the web camera are displayed in the HMD.

It takes longer than 200 ms before head rotations are reflected, as we described in the section Synchronization and Latency. The system thus risks inducing VR sickness and physical fatigue.

Therefore, we asked three questions to investigate how much fatigued participants felt using the system. Figure 11 shows our results. None of the participants had used the system previously, so we collected all of the answers together. The overall fatigue from our system was rated at 5.21 on average, but the coefficient of variation was 20.7% and there were individual differences. In the free-description question, several participants noted "lack of experience" (P1, G2, and G8), "weight of HMD" (G2 and P5) and "VR sickness" (P3, P4, G1, G3, G4, and G7) as causing fatigue. In terms of the latency of the image and sound, there were also individual differences. The coefficient of variation in the latency of image was 34.7%, and that of sound was 34.7%, and participants were more positive in rating the latency of the sound than that of the image.

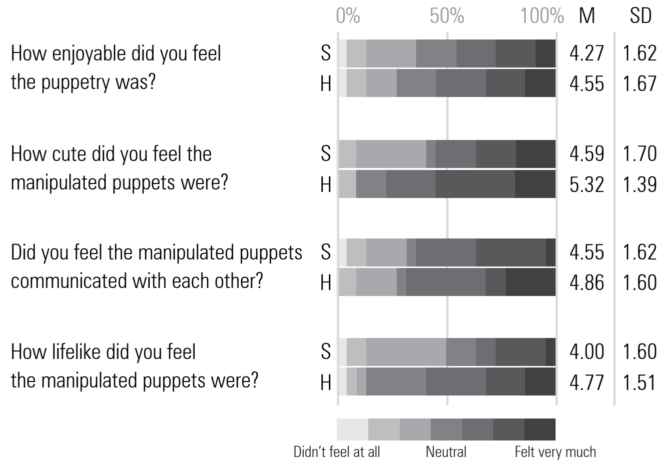

Figure 13. Rating two kinds of puppetry. Note that S denotes the proposed system and H denotes hand manipulation.

Comparison of Telepresence Puppetry and Conventional Puppetry

Because we here propose a new method to perform puppetry, we need to confirm how puppetry acted by our system differs from conventional puppetry. In this user study, we investigated how the audience felt when they watched puppetry acted through different methods.

Figure 12. Puppetry acted with conventional method (left), and puppetry acted using our system (right).

Study Process

We invited professional puppeteers and asked them to perform puppetry using both our method and the conventional one. We recorded video of their puppetry assigning them the same story, and asked participants to rate these videos in order to compare them (Figure 12). We used a web-based question form to ask the participants to rate the puppetry itself and the puppet manipulation, both on a seven-point scale, from "Did not feel at all" to "Felt very much". We analyzed the scores of each question on a 7-point Likert scale using two-tailed Mann–Whitney U tests evaluated at an alpha level of 0.05.

Participants

This study involved 22 participants (11 females, 11 males) aged between 19 and 24 years (M = 21.5, SD = 1.40). We recruited general participants by advertising on a few social network services.

Results

Figure 13 shows the results of the questionnaire with four questions related to the two kinds of puppetry. The results of the first question indicate how enjoyable the acted puppetry was. In both methods, the range extended from 1 ("Did not feel enjoyable") to 7 ("Felt very much"). The U test for this question showed that the difference between puppetry using our system (M = 4.27, MD = 4, SD = 1.62) and the conventional one (M = 4.55, MD = 5, SD = 1.67) was not significant (U = 218.0, $p > 0.05$), and the overall evaluation did not differ significantly with either approach.

The second question asked how "cute" the manipulated puppets were. The U test for this question showed that no significant difference existed (U = 185.5, $p > 0.05$) between puppetry using our system (M = 4.59, MD = 5, SD = 1.70) and the conventional one (M = 5.32, MD = 6, SD = 1.39). The skin of both puppets was the same, but the puppet manipulated with our system included a 7-DOF robot under that skin. Whereas 90.9% of participants rated the puppet that was manipulated conventionally as positive or neutral, only 59.0% of participants rated the puppet controlled with our system as positive or neutral. It is therefore clear that linear movements and the sound of servo motors negatively affected the results.

The third question pertained to the communication between the two puppets. The U test for this question showed that there was no significant difference (U = 222.0, $p > 0.05$) between puppetry using our system (M = 4.55, MD = 5, SD = 1.62) and the conventional one (M = 4.86, MD = 5, SD = 1.60). The results of this question also varied. The coefficient of variation with the conventional method was 33.3%, and 35.0% with the proposed method. Although there was variation, approximately 63% of participants felt that puppets communicated with each other.

The fourth question investigated how lifelike the puppets acted. The U test for this question showed that there was no significant difference (U = 171.0, $p > 0.05$) between puppetry using our system (M = 4.00, MD = 3.5, SD = 1.60) and the conventional one (M = 4.77, MD = 5, SD = 1.51).

In the free-description question, opinions about expressions of puppetry differed among participants.

In terms of the puppet's expressions, there were more negative opinions than positive ones. While a few participants felt that the puppetry controlled using our system seemed novel and realistic (participants 1 and 15), other participants felt that the hand-manipulated puppetry appeared lifelike (P1 and P7), realistic (P16), comical (P15), dynamic (P2), hearty (P21), cute (P20), and fun (P19), and the puppetry controlled using our system appeared mechanical (P6 and P13), and not alive (P19). The opinions of a few participants are provided below.

P(participant)1: *"I felt that the conventional puppets were more lifelike than the system-controlled ones. However, the system-controlled puppets seemed more novel to watch, because I could easily notice the 'hand-manipulation' of the conventional puppets. I suggest that it would be so interesting if the system could manipulate a puppet in a way that manual hand motion could never accomplish."*

P7: *"The hand-manipulated puppets' movements were subtle and seemed more lifelike."*

P15: *"I felt that the movements performed using the system were more realistic; however, the movements performed using the hand were comical and cuter. I think both methods have advantages, i.e., movements are either realistic or comical."*

P19: *"I couldn't feel the life of the puppet controlled by the system."*

A few participants referred to the mechanical sound of motors (P4, P10, P17 and P22).

P10: *"In the puppetry manipulated using the system, the operating sound was annoying."*

The participants' opinions differed as to which method better represented puppet communication.

P22: *"Although smoothness and expression are better with hand-manipulated puppetry, I did not feel that the puppets communicated with each other, because I could see that there were humans manipulating the puppets. On the contrary, the puppetry with the telepresence system made noisy sounds and had awkward movements. Even so, I felt strongly that the puppets communicated."*

P18: *"With hand-manipulated puppetry, the puppets seemed conscious of each other."*

A participant (P1) mentioned that it is interesting that the system manipulates and moves a puppet in a manner in which a hand cannot.

LIMITATIONS

Overall, our findings suggest that the telepresence user interface facilitates the manipulation of a puppet and allows the performers to enjoy it. The participants rated our system highly, whether they were ordinary users or professional puppeteers. At the same time, however, the results suggest several limitations to our system, especially from the perspective of the audience. In this section, we point out limitations and problems with our system, and we suggest further technologies to resolve them.

The Position of a Puppet on Stage Cannot be Changed

In conventional puppetry, a puppet can move around on the table freely. It can hide and later appear, and it can move left and right on stage. However, puppets with our system are fixed on stage, as shown in Figure 1 (c). They cannot move from one end to another. Professional participants mentioned the problem. In future telepresence puppetry, performers will need to move the puppets, in order for telepresence puppetry to be able to tell a complicated story to audiences for the sake of entertainment. To address this, we shall consider a function for detecting the center of gravity (e.g., with a Nintendo Wii Fit) as a solution to move a puppet in a 1-axis direction. Furthermore, other orientations detected by Kinect v2 can be adopted for movement on the table.

Performer Cannot Confirm How a Puppet Moves

Professional puppeteers try to be conscious of the audience. In other words, they really care about the appearance of the puppet that the audience sees. Therefore, it is useful for them to see how a puppet looks using our system. Performers who use the system can only see it from a first-person viewpoint. Thus, they cannot visualize the entire body of the puppet. The current system does not have any function to provide feedback for the state of the puppet's movements, except in terms of its head orientation. It is possible to design software to calculate and infer the state of the puppet's body, however, and to display that state in the HMD. We could also address the problem with images from a fixed-point camera that capture the puppet objectively and display it in the HMD.

Movements Look Somewhat Mechanical or Artificial

As we described in the section of USER STUDY, around half of the participants did not feel that a puppet was lifelike. As a reason, some commented that the puppet looked mechanical or artificial, because of its linear movements and the sound of the motors. The puppet is a 7-DOF robot, and indeed this is not sufficient to express subtle motions made with conventional hand manipulation. The system can address this issue by adding extra servo motors to the robot. By reproducing with servo motors the way actual animals move, the system can manipulate the puppet in a more lifelike manner. Regarding the sound of motors, we believe that playing some background music can mitigate this somewhat.

Stress and Fatigue on a Performer

We determined that the main cause of user fatigue was VR sickness due to the latency of video displayed in the HMD and owing to the low-resolution displays of the HMD. We guess that the video latency with our system is from the approximately 200 ms latency when reflecting the head orientations. In the user study, we asked participants how many hours they thought they could use the system. Indeed, 12 of the 14 participants answered that they could use it for longer than 30 minutes. For short puppet shows, then, we believe that the fatigue from the system can be tolerated by most performers. However, time using the system may last longer than 30 minutes; for example, when the telepresence system is used for communication or attending a conference as an avatar. In our system, we use two-web cameras which stream the video to the display in the HMD; however, as they are attached to the robot bone, they follow the movements of servo motors, and this can induce nausea. Thus, in future applications of telepresence, omnidirectional cameras, such as RICOH THETA, equipped around the head of the robot or puppet can be utilized to display a video remotely because the video can be cropped and displayed to users without being affected by the latency of servo motors. However, immersion can be lost to an extent because the viewpoint becomes slightly different from what a puppet actually sees. How users feel the loss of immersion using omnidirectional cameras can be investigated in future study.

Cost and Space

One of the benefits of typical puppetry is that it can be performed at an extremely low cost and in small spaces. The proposed system requires approximately 1.5 m $\times$ 1.5 m space for the puppeteer, even though the area required for the audience is the same as that in the case of the typical method. This is because the depth camera for capturing motions requires space and the system contains laptops and some other components. The cost of the system is high because of the HMD, laptops, servo motors, etc. For the general public, the HMD can be replaced with a low-cost version such as Google Cardboard.

CONCLUSION AND FUTURE WORK

We implemented a novel system for controlling puppets remotely using several techniques. The system tracked the body, mouth, and head of the performer, while providing visual feedback through cameras and an HMD. This allowed the performer to judge the audience's reaction to the puppet, and facilitated puppet-to-puppet interaction. We measured the latency of our system, and the results showed that our system can control every joint of a robotic puppet within approximately 650 ms. Further, we conducted several user studies to assess the usability and enjoyment of the proposal, and to compare conventional puppetry and telepresence control from the perspective of both the audience and controller. The user study showed that the telepresence system allows performers to enjoy manipulating a puppet, provides and immersive experience, and improves the easiness of manipulating a puppet, especially for beginners. On the other hand, the current system has limitations such as fatigue caused by the HMD, mechanical movements and sound, and lack of servo motors. Our hope is that the proposed method will be useful and assistive in designing new systems that combines puppets and telepresence.

Though we have demonstrated its utility in the puppetry, our telepresence system may be adapted for use in a number of ways.

Communication Tools Using Telepresence Puppets

Communication tools using puppets have been proposed to talk to children. In future work, our immersive telepresence system would be useful and enjoyable to parents who are in a remote place and have trouble talking to a screen, yet want to communicate with their children through a puppet. In such cases, it is critical to investigate how children react to parents acting as puppets, and how parents feel when they communicate as a puppet to their children.

Communication systems using the proposed telepresence method also can be used for attending a conference from a remote place. Attending a conference remotely can cause failure in recreating the atmosphere at the place, conveying opinions, displaying information to the attendees, and communicating their feelings with others. We believe that a telepresence puppet can be developed as an avatar for attending a conference remotely in future work. A physical puppet allows remote attendees to express their thoughts, share specific information with others, and express their opinions.

Puppet Show in Theme parks

In theme parks, there are stage shows, parades, and attractions with puppets and costumed characters. Interfaces for changing the facial expressions of a character in costume have been proposed previously. However, actors must be inside the costume in order to change expressions and motions. Our system can address this physical limitation and allow actors to control a puppet and interact with the audience from a distance. Moreover, interactive attractions with displays are increasingly common, but interactive attractions with robotic puppets are very uncommon. Using our telepresence system in such an attraction could encourage repeat customers by changing the content each time they ride the attraction. In addition, the proposed system utilizes a normal sized puppet and few servos on its face and body. Our system is, however, applicable to both very tiny and very huge scaled puppets, and unique shaped puppets that are usually difficult for actors to manipulate in a conventional method. For example, in theme park's shows or parades, a lifelike tiny character or large sized character that can have interactions with guests will fascinate and attract more guests.

ACKNOWLEDGMENTS

We would like to express our sincere appreciation to NEU Puppetry Troupe, which played puppet-shows for our experiment. I would also like to thank the following members of FUJITSU SOCIAL SCIENCE LABORATORY LIMITED: Tsuyoshi Shimada, Chieko Hiramatsu, Yasuji Katase, Ichizo Yonezawa, Shin Katou, and Kosaku Ogawa, for their unfailing support and assistance.

REFERENCES

1. E. Blumenthal. 2005. *Puppetry and Puppets: An Illustrated World Survey*. Thames & Hudson. https://books.google.co.uk/books?id=VkV-QgAACAAJ

2. Leonardo Bonanni, Cati Vaucelle, Jeff Lieberman, and Orit Zuckerman. 2006. PlayPals: Tangible Interfaces for Remote Communication and Play. In *CHI '06 Extended Abstracts on Human Factors in Computing Systems (CHI EA '06)*. ACM, New York, NY, USA, 574–579. DOI: http://dx.doi.org/10.1145/1125451.1125572

3. Matthew Brand. 1999. Voice Puppetry. In *Proceedings of the 26th Annual Conference on Computer Graphics and Interactive Techniques (SIGGRAPH '99)*. ACM Press/Addison-Wesley Publishing Co., New York, NY,

USA, 21–28. DOI: http://dx.doi.org/10.1145/311535.311537

4. L. Fritsche, F. Unverzagt, J. Peters, and R. Calandra. 2015. First-Person Tele-Operation of a Humanoid Robot. In *15th IEEE-RAS International Conference on Humanoid Robots*. 997–1002. http://www.ausy.tu-darmstadt.de/uploads/Site/EditPublication/Fritsche_Humanoids15.pdf
5. J.C. Garlen and A.M. Graham. 2009. *Kermit Culture: Critical Perspectives on Jim Henson's Muppets*. McFarland, Incorporated, Publishers. https://books.google.co.jp/books?id=Rac9xis0BIgC
6. A. Gruebler and K. Suzuki. 2010. Measurement of distal EMG signals using a wearable device for reading facial expressions. In *2010 Annual International Conference of the IEEE Engineering in Medicine and Biology*. 4594–4597. DOI: http://dx.doi.org/10.1109/IEMBS.2010.5626504
7. Robert Held, Ankit Gupta, Brian Curless, and Maneesh Agrawala. 2012. 3D Puppetry: A Kinect-based Interface for 3D Animation. In *Proceedings of the 25th Annual ACM Symposium on User Interface Software and Technology (UIST '12)*. ACM, New York, NY, USA, 423–434. DOI: http://dx.doi.org/10.1145/2380116.2380170
8. Tamara Hunt and Nancy Renfro. 1982. *Puppetry in early childhood education*. N. Renfro Studios.
9. Keisuke Kawahara, Mose Sakashita, Amy Koike, Ippei Suzuki, Kenta Suzuki, and Yoichi Ochiai. 2016. Transformed Human Presence for Puppetry. In *Proceedings of the 13th International Conference on Advances in Computer Entertainment Technology (ACE2016)*. ACM, New York, NY, USA, Article 38, 6 pages. DOI:http://dx.doi.org/10.1145/3001773.3001813
10. Joseph J. LaViola, Jr. 2000. A Discussion of Cybersickness in Virtual Environments. *SIGCHI Bull.* 32, 1 (Jan. 2000), 47–56. DOI: http://dx.doi.org/10.1145/333329.333344
11. Jun Ki Lee, R. L. Toscano, W. D. Stiehl, and C. Breazeal. 2008. The design of a semi-autonomous robot avatar for family communication and education. In *RO-MAN 2008 - The 17th IEEE International Symposium on Robot and Human Interactive Communication*. 166–173. DOI: http://dx.doi.org/10.1109/ROMAN.2008.4600661
12. Luís Leite and Veronica Orvalho. 2012. Shape Your Body: Control a Virtual Silhouette Using Body Motion. In *CHI '12 Extended Abstracts on Human Factors in Computing Systems (CHI EA '12)*. ACM, New York, NY, USA, 1913–1918. DOI: http://dx.doi.org/10.1145/2212776.2223728
13. Luis LEite and Veronica Orvalho. 2017. Mani-Pull-Action: Hand-based Digital Puppetry. *Proc. ACM Hum.-Comput. Interact.* 1, 1, Article 2 (June 2017), 16 pages. DOI:http://dx.doi.org/10.1145/3095804
14. B. Y. L. Li, A. S. Mian, W. Liu, and A. Krishna. 2013. Using Kinect for face recognition under varying poses, expressions, illumination and disguise. In *Applications of Computer Vision (WACV), 2013 IEEE Workshop on*. 186–192. DOI: http://dx.doi.org/10.1109/WACV.2013.6475017
15. H. Li, P. Roivainen, and R. Forcheimer. 1993. 3DD Motion Estimation in Model-Based Facial Image Coding. *IEEE Trans. Pattern Anal. Mach. Intell.* 15, 6 (June 1993), 545–555. DOI:http://dx.doi.org/10.1109/34.216724
16. Hao Li, Laura Trutoiu, Kyle Olszewski, Lingyu Wei, Tristan Trutna, Pei-Lun Hsieh, Aaron Nicholls, and Chongyang Ma. 2015. Facial Performance Sensing Head-mounted Display. *ACM Trans. Graph.* 34, 4, Article 47 (July 2015), 9 pages. DOI: http://dx.doi.org/10.1145/2766939
17. Peter Lincoln, Greg Welch, Andrew Nashel, Adrian Ilie, Andrei State, and Henry Fuchs. 2009. Animatronic Shader Lamps Avatars. In *Proceedings of the 2009 8th IEEE International Symposium on Mixed and Augmented Reality (ISMAR '09)*. IEEE Computer Society, Washington, DC, USA, 27–33. DOI: http://dx.doi.org/10.1109/ISMAR.2009.5336503
18. Katsutoshi Masai, Yuta Sugiura, Masa Ogata, Katsuhiro Suzuki, Fumihiko Nakamura, Sho Shimamura, Kai Kunze, Masahiko Inami, and Maki Sugimoto. 2015. AffectiveWear: Toward Recognizing Facial Expression. In *ACM SIGGRAPH 2015 Posters (SIGGRAPH '15)*. ACM, New York, NY, USA, Article 16, 1 pages. DOI: http://dx.doi.org/10.1145/2787626.2792632
19. Marvin Minsky. 1980. Telepresence. OMNI magazin, June 1980. (1980).
20. Kana Misawa and Jun Rekimoto. 2015. ChameleonMask: Embodied Physical and Social Telepresence Using Human Surrogates. In *Proceedings of the 33rd Annual ACM Conference Extended Abstracts on Human Factors in Computing Systems (CHI EA '15)*. ACM, New York, NY, USA, 401–411. DOI: http://dx.doi.org/10.1145/2702613.2732506
21. Shohei Nagai, Shunichi Kasahara, and Jun Rekimoto. 2015. LiveSphere: Sharing the Surrounding Visual Environment for Immersive Experience in Remote Collaboration. In *Proceedings of the Ninth International Conference on Tangible, Embedded, and Embodied Interaction (TEI '15)*. ACM, New York, NY, USA, 113–116. DOI: http://dx.doi.org/10.1145/2677199.2680549
22. Oculus.inc. 2014. Oculus Rift Develop kit 2 (last accessed August 27, 2016). (2014). https://www.oculus.com/en-us/dk2/
23. E.M.B.J. O'Hare. 2005. *Puppetry in Education and Therapy: Unlocking Doors to the Mind and Heart*. AuthorHouse. https://books.google.co.jp/books?id=7X0CodZfjrcC

24. Masaki Oshita, Yuta Senju, and Syun Morishige. 2013. Character Motion Control Interface with Hand Manipulation Inspired by Puppet Mechanism. In *Proceedings of the 12th ACM SIGGRAPH International Conference on Virtual-Reality Continuum and Its Applications in Industry (VRCAI '13)*. ACM, New York, NY, USA, 131–138. DOI: http://dx.doi.org/10.1145/2534329.2534360

25. Emil Polyak. 2012. Virtual Impersonation Using Interactive Glove Puppets. In *SIGGRAPH Asia 2012 Posters (SA '12)*. ACM, New York, NY, USA, Article 31, 1 pages. DOI: http://dx.doi.org/10.1145/2407156.2407191

26. Daisuke Sakamoto, Takayuki Kanda, Tetsuo Ono, Hiroshi Ishiguro, and Norihiro Hagita. 2007. Android As a Telecommunication Medium with a Human-like Presence. In *Proceedings of the ACM/IEEE International Conference on Human-robot Interaction (HRI '07)*. ACM, New York, NY, USA, 193–200. DOI: http://dx.doi.org/10.1145/1228716.1228743

27. Mose Sakashita, Keisuke Kawahara, Amy Koike, Kenta Suzuki, Ippei Suzuki, and Yoichi Ochiai. 2016. Yadori: Mask-type User Interface for Manipulation of Puppets. In *ACM SIGGRAPH 2016 Emerging Technologies (SIGGRAPH '16)*. ACM, New York, NY, USA, Article 23, 1 pages. DOI: http://dx.doi.org/10.1145/2929464.2929478

28. C.E. Schaefer and L.J. Carey. 1994. *Family Play Therapy*. J. Aronson. https://books.google.co.jp/books?id=kElcV84iH-4C

29. Dairoku Sekiguchi, Masahiko Inami, and Susumu Tachi. 2001. RobotPHONE: RUI for interpersonal communication. In *CHI'01 Extended Abstracts on Human Factors in Computing Systems*. ACM, 277–278.

30. Rika Shoji, Toshiki Yoshiike, Yuya Kikukawa, Tadahiro Nishikawa, Taigetsu Saori, Suketomo Ayaka, Tetsuaki Baba, and Kumiko Kushiyama. 2012. Mimicat: Face Input Interface Supporting Animatronics Costume Performer's Facial Expression. In *ACM SIGGRAPH 2012 Posters (SIGGRAPH '12)*. ACM, New York, NY, USA, Article 72, 1 pages. DOI: http://dx.doi.org/10.1145/2342896.2342983

31. Ronit Slyper, Guy Hoffman, and Ariel Shamir. 2015. Mirror Puppeteering: Animating Toy Robots in Front of a Webcam. In *Proceedings of the Ninth International Conference on Tangible, Embedded, and Embodied Interaction (TEI '15)*. ACM, New York, NY, USA, 241–248. DOI: http://dx.doi.org/10.1145/2677199.2680548

32. Ronit Slyper, Jill Lehman, Jodi Forlizzi, and Jessica Hodgins. 2011. A Tongue Input Device for Creating Conversations. In *Proceedings of the 24th annual ACM Symposium on User Interface Software and Technology (UIST '11)*. ACM, 117–126.

33. Yuta Sugiura, Calista Lee, Masayasu Ogata, Anusha Withana, Yasutoshi Makino, Daisuke Sakamoto, Masahiko Inami, and Takeo Igarashi. 2012. PINOKY: A Ring That Animates Your Plush Toys. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '12)*. ACM, New York, NY, USA, 725–734. DOI: http://dx.doi.org/10.1145/2207676.2207780

34. Thibaut Weise, Hao Li, Luc Van Gool, and Mark Pauly. 2009. Face/Off: Live Facial Puppetry. In *Proceedings of the 2009 ACM SIGGRAPH/Eurographics Symposium on Computer Animation (SCA '09)*. ACM, New York, NY, USA, 7–16. DOI: http://dx.doi.org/10.1145/1599470.1599472

35. John P Whitney, Tianyao Chen, John Mars, and Jessica K Hodgins. 2016. A hybrid hydrostatic transmission and human-safe haptic telepresence robot. In *Robotics and Automation (ICRA), 2016 IEEE International Conference on*. IEEE, 690–695.

36. Tomoko Yonezawa and Kenji Mase. 2002. Musically Expressive Doll in Face-to-Face Communication. In *Proceedings of the 4th IEEE International Conference on Multimodal Interfaces (ICMI '02)*. IEEE Computer Society, Washington, DC, USA, 417–. DOI: http://dx.doi.org/10.1109/ICMI.2002.1167031