

---

# Transformed Human Presence for Puppetry

**Keisuke Kawahara**

University of Tsukuba  
1-1-1 Tennodai, Tsukuba,  
Ibaraki, Japan  
kawahara@ai.iit.tsukuba.ac.jp

**Mose Sakashita**

*\*Joint first authorship*  
University of Tsukuba  
1-2 Kasuga, Tsukuba, Ibaraki,  
Japan  
mose.sakashita@gmail.com

**Amy Koike**

University of Tsukuba  
1-2 Kasuga, Tsukuba, Ibaraki,  
Japan  
amy23kik@gmail.com

**Ippei Suzuki**

University of Tsukuba  
1-2 Kasuga, Tsukuba, Ibaraki,  
Japan  
1heisuzuki@gmail.com

**Kenta Suzuki**

University of Tsukuba  
1-2 Kasuga, Tsukuba, Ibaraki,  
Japan  
ikuzus.atnek.0626@gmail.com

**Yoichi Ochiai**

University of Tsukuba  
1-2 Kasuga, Tsukuba, Ibaraki,  
Japan  
wizard@slis.tsukuba.ac.jp

**Abstract**

We propose a system for transmitting a human performer's body and facial movements to a puppet with audiovisual feedback to the performer. The system consists of a head-mounted display (HMD) that shows the performer the video recording of the puppet's view, a microphone for voice capture, and photoreflectors for detecting the mouth movements of the human performer. In conventional puppetry, there is also the need for practice in the manipulation of the puppets to achieve good performance. The proposed telepresence system addresses these issues by enabling the human performer to manipulate the puppet through their own body and facial movements. The proposed system is expected to contribute to the development of new applications of puppetry and expand the interactivity of puppetry and the scope of entertainment.

**Author Keywords**

Animatronics; puppet; telepresence; head-mounted display.

**ACM Classification Keywords**

H.5.2 [Information interfaces and presentation (e.g., HCI)]: User Interfaces

---

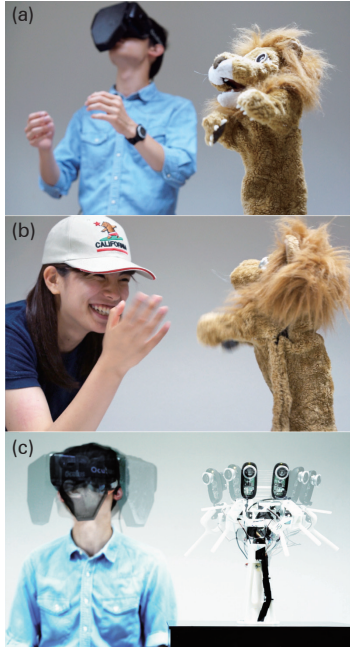
Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s).

Copyright held by the owner/author(s).

ACE2016, November 09-12, 2016, Osaka, Japan

ACM 978-1-4503-4773-0/16/11.

<http://dx.doi.org/10.1145/3001773.3001813>



**Figure 1:** (a) Puppetry using telepresence user interface. (b) Interaction between a puppet and the audience. (c) Synchronization of the performer and puppet through the proposed system.

## Introduction

Puppetry has a long history, being thought to have originated in 3000 BC [1]. A puppet show is a basic form of performance involving the manipulation of inanimate objects with some semblances to humans or animals. One of the simplest methods of puppetry involves the movement of the puppet through the hands or arms of a human puppeteer. The puppet is controlled from above with the rest of the puppeteer's body hidden backstage. This, however, poses a coordination challenge and also requires sufficient training in manipulating the puppet. To address these issues, many user interfaces and novel methods have been developed for computational puppet manipulation. For example, a grove-type user interface [7] and a system for recording the limb-controlled movements of the puppet through a webcam [10] have been presented.

Studies have particularly been conducted on the application of telepresence to puppet manipulation. Telepresence is a technology that enables a user to have a sense of being in a different place. An example of the application of telepresence is Huggable [4], a teddy bear robot designed for social communication applications. These developments used the controlled puppets for communication and education purposes, without particular interest in entertainment. The parameters that are transmitted to the puppets for manipulation are thus often insufficient for the animation of a broad range of expressions and movements.

The concept of human-to-nonhuman telepresence (Figure 2) affords new experiences for performers and observers and has the potential for novel entertainment applications. The telepresence system proposed in this paper transmits the body and facial movements of a performer to a puppet with visual and audio feedback to the performer. In other words, it enables the transmission of a human's presence

to a puppet. The system facilitates easy manipulation of the puppet by the performer, as well as the expression of more lifelike characteristics. This is because the control is accomplished by the performer's own body and facial movements, including those of the mouth. Conventional puppetry has physical limitations, which include the need for the puppeteer to be at the exact location of the puppet. Telepresence eliminates such physical limitations and enables multiple performers at different locations to collaborate in sequential manipulation of the same puppet. It also enables a given performer to seamlessly shift between different puppets. Further, the remote performer is able to interact with the audience as if they were at the location of the puppet. This is facilitated by the feature of real-time visual and audio feedback to the performer. The system enables untrained hands to perform a puppet show, thus eliminating the special practice required by conventional puppetry. The performer's sense of being on location also promises smooth interaction between the puppet and the audience, opening new entertainment frontiers.

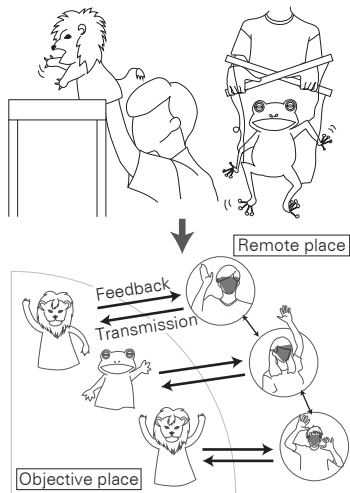
## Related Works

### *Computational puppetry*

Several studies have been conducted on the manipulation of puppets, dolls, and toys. PINOKY [11] is a ring-like device that can be used to animate a toy by movement of its limbs. A grove-type user interface for interactively animating a puppet [7], a system for recording the movements of a puppet by a webcam [10], and a system for animatronics storytelling that enables performers to manipulate puppets by wearing a mask-type device on their faces [8] are other related developments.

### *Facial expression*

In our study, we developed a system for transmitting captured human facial expressions and movements to a real



**Figure 2:** Conventional puppet manipulation (upper), and the concept of the proposed system (lower).

puppet in real time. Hundreds of related virtual reality studies have utilized head-mounted displays (HMDs). However, this creates the difficulty of tracking and capturing the human face because most of the employed facial tracking systems required the captured face to be fully visible. A means of tracking the eye under an HMD was recently developed [5]. The system uses RGB-D cameras to generate an animation of the full face. However, it is heavy and uncomfortable to carry on the user's head because the headset includes a depth sensor in addition to the HMD. Considering that our proposed system also uses an HMD for visual feedback to the user, we employed a mask-type device with a photoreflective sensor module to capture the facial expressions. The mask-type device is lighter than a depth camera, and the employed sensor processing also enables faster calculation compared to the use of image processing to capture facial movements.

### Telepresence

Presentations have been made regarding the application of telepresence to puppet manipulation. An example is a teddy bear robot for social communication applications [4]. The robot is remotely controlled through an input device such as a Wii Remote, with the user receiving audiovisual feedback and being able to communicate with local people at the location of the robot. Another example is Robot-Phone [9], a robotic user interface (RUI) that allows the user to modify the shapes of the components of the connected robot, with the ability to receive feedback on the movements of the components via the Internet and replicate the same in other RobotPhones. PlayPals [2] is yet another wireless system that offers children a means of playfully communicating with other children elsewhere by manipulating a puppet at the remote location. However, the remote puppet can only be used for communication and educational purposes, and not for entertainment. Conversely, the presently

proposed system enables the user to transmit their voice and body, facial, and mouth movements to the remote puppet, with audiovisual feedback, thus enabling application to entertainment.

## Implementation

In this section, we describe the implementation of the proposed system, the configuration of which is shown in Figure 3. The employed mask-type device consists of an HMD [6], a microphone for capturing the performer's voice, and an array of photoreflectors for detecting the mouth movements. A Kinect camera is used to capture the body movements of the performer. The manipulated puppet is equipped with a microphone for capturing ambient sound, two cameras, and seven servomotors that move its arms, mouth, and neck (see Figure 8).

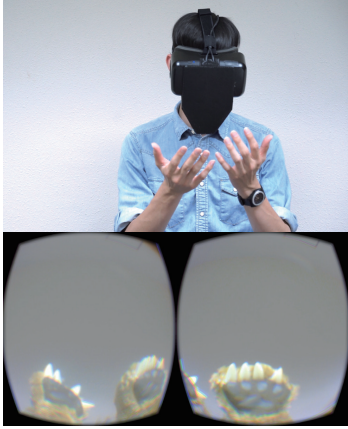
### Head-mounted display and cameras

The eyes of the robot are equipped with two 1.2-million-pixel web cameras (Figure 3 (right)) for streaming live video to the HMD of the performer. This enables the performer to see through the eyes of the puppet and visualize its surroundings (Figures 4 and 5). The distance between the human eyes is approximately 65 mm, and this was applied to the puppet's eyes.

### Facial movements

The robot has a 2-DOF neck that is controlled in accordance with the performer's facial movements. An inertial-measurement unit (IMU) installed in the HMD is used to read the gyro sensor for measuring the Euler angles of the performer's face.

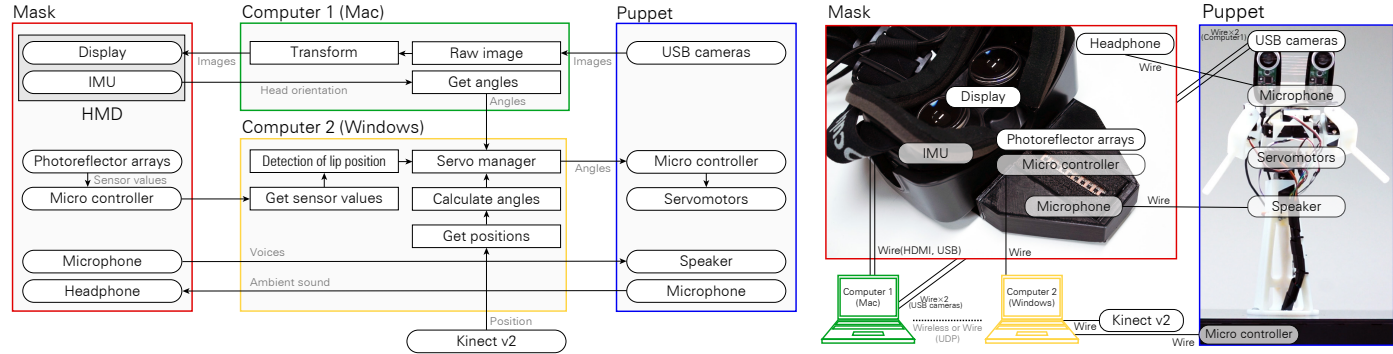
The measured Euler angles are transmitted to the neck servomotors of the robot at 30 fps via a UDP in the software of the employed microcontroller (Arduino Micro).



**Figure 4:** Performer's visualization of the puppet's hands through the HMD.



**Figure 5:** Interaction with the audience (upper), and visual feedback (lower).



**Figure 3:** Configuration of the proposed system.

#### *Detection of mouth state of performer*

We developed an array of photoreflexors for detecting the opened or closed state of the performer's mouth (Figure 6 (e)). 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 photoreflexor sensors (TPR-105; GENIXTEK CORP) with a maximum sensing distance of 20 mm. Figure 6 (d) shows sample results of the photoreflexor sensor measurements for (a) mouth closed, (b) mouth opened, and (c) mouth partly opened. The program for detecting the lower lip position was implemented using a Java processor (Figure 6 (e)). The software could be used to calculate the difference between the measurements of the different sensors, with the maximum measurement appearing to be the valid indication of the actual position of the performer's lower lip.

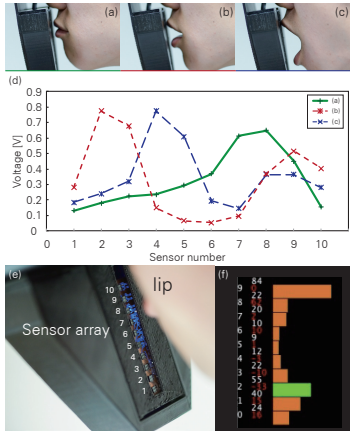
Based on the detected position, the software controlled the mouth servomotor of the puppet. This setup synchronized the mouth movement of the puppet with the speech of

the performer, giving the impression of the puppet actually speaking to the audience. The positions of the sensors on the board can be adjusted to fit different users.

#### *Voice transmission*

**Performer-to-Puppet:** A microphone in the mask is used to capture the performer's speech, which is transmitted and replayed in real-time through a speaker installed in the puppet. An effect is used to change the voice of the performer to that of the puppet character, thus affording a more convincing and attractive show.

**Puppet-to-Performer:** Another microphone installed in the puppet is used to capture ambient sound at the location of the puppet, and this is transmitted back to the performer through earphones. The performer is thus able to monitor acoustic events around the puppet, such as the reaction of the audience and the words of any other puppet character in the show. This enables the performer to interact with the audience through the puppet.



**Figure 6:** (a), (b), and (c) show three states of the performer's mouth, while (d) shows the respective corresponding plots. (e) Lip position detection using the developed sensor array. (f) Software for detecting the performers lips.



**Figure 7:** Mask type device that we developed.

### *Sensing of performer's arms*

As an additional function, a Kinect depth camera is used to capture the arm movements of the performer to manipulate those of the puppet. This is based on a previous work [3]. The present system thus uses a combination of the body and facial movements of the puppeteer to smoothly and synchronously manipulate the puppet. The tilt angles and pan directions are calculated from the positions of the arms captured by the Kinect depth sensor, and transmitted by a UDP at 30 fps to the microcontroller. The software of the microcontroller was developed using SDK version 2.0. A low-pass filter is used to stabilize the low-resolution and noisy input data from the Kinect sensor. The two 2-DOF arms of the puppet are driven by two servomotors each in accordance with the captured motions of the performer's arms.

### *Mask-type device*

As already discussed, the mask-type device is used to display images of the puppet's view to the performer and track the head and mouth movements of the performer. It consists of an HMD and adaptor (Figure 7). Orientations of the performers head can also be acquired by the HMD for further manipulation of the puppet. Because the photoreflexor array captures the mouth movements of the performer, an adaptor was integrated in the headset.

### *Puppet bones*

Within the puppet is a skeletal framework to which the servomotors are attached (Figure 8). The framework was designed such that the servomotors would not interfere with each other. Because the framework can be printed by a 3D printer, the robot can be manufactured at low cost. Considering that the puppet has two 2-DOF arms, a 1-DOF mouth, and a 2-DOF neck, its entire body has seven DOFs.

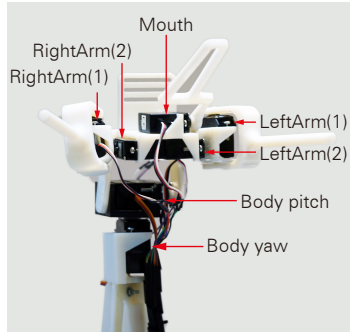
## **Discussion**

Figure 9 illustrates the response of the puppet to the puppeteer's movements as driven by the servomotors. Although the developed telepresence system enables the demonstration of the mouth movements and facial expressions of the puppeteer, there is room for the implementation of additional expressions such as emotion and eye movement to further enhance the performance. This would require the utilization of components such as strain gauges to detect the relevant expressions of the performer under the HMD, as has been done in some previous studies. Further, whereas the developed puppet robot has seven DOFs, this is probably insufficient. Improvements can be achieved by incorporating additional actuators to manipulate the legs and eyes. Additionally, although the weight of the HMD was reduced by the use of photoreflexors, it may still be too heavy for the convenience of some users. Hopefully, lighter HMDs would be developed in the future.

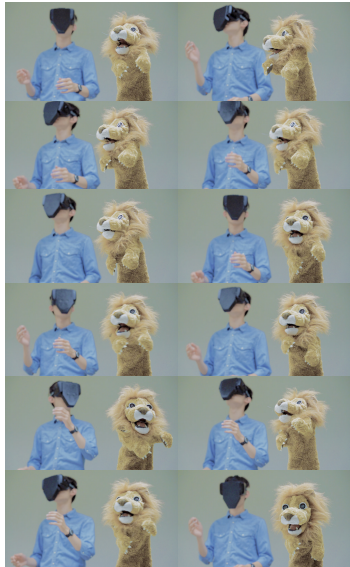
Overall, the proposed system overcomes some of the physical limitations of the conventional puppet manipulation method, enabling the direct transmission of the actual performance of the puppeteer to the puppet. The system has strong potential for application to entertainment. For example, if employed in theme parks, puppets, dolls, and toys can be manipulated by performers at remote locations, eliminating the practice of wearing special character costumes. Multiple performers at different locations, as mentioned earlier, may also manipulate the same puppet in turn, and a given performer can move from one puppet to another. All such shifts can be done almost instantly.

## **Conclusion**

We proposed and described a system for manipulating a puppet using an HMD. The system enables the transmission of the puppeteer's body and facial movements to the



**Figure 8:** Bones and servomotors inside the puppet.



**Figure 9:** Time-series photographs of the puppet animation (intervals of a second).

puppet, with audiovisual feedback to the performer. However, as mentioned in the previous section, the system has some limitations and there is room for improvements through further study to achieve more lifelike animation and explore the potential application to entertainment.

## REFERENCES

1. E. Blumenthal. 2005. *Puppetry and Puppets: An Illustrated World Survey*. Thames & Hudson.
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.
3. 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.
4. 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.
5. 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.
6. Oculus.inc. 2014. Oculus Rift Develop kit 2 (last accessed August 27, 2016). (2014). <https://www.oculus.com/en-us/dk2/>
7. 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.
8. 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.
9. 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.
10. 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.
11. 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.