LiveMask: A Telepresence Surrogate System with a Face-Shaped Screen for Supporting Nonverbal Communication

Kana Misawa
Interfaculty Initiative in Information Studies,
The University of Tokyo
4-6-1 Komaba, Meguro, Tokyo, Japan
kana.misawa@gmail.com

Yoshio Ishiguro
Interfaculty Initiative in Information Studies,
The University of Tokyo
4-6-1 Komaba, Meguro, Tokyo, Japan
ishiy@acm.org

Jun Rekimoto
Interfaculty Initiative in Information Studies,
The University of Tokyo
4-6-1 Komaba, Meguro, Tokyo, Japan
rekimoto@acm.org

ABSTRACT

We propose a telepresence system with a real human face-shaped screen. This system tracks the remote user's face and extracts the head motion and the face image. The face-shaped screen moves along three degree-of-freedom (DOF) by reflecting the user's head gestures. As the face-shaped screen is molded based on the 3D-shape scan data of the user, the projected image is accurate even when it is seen from different angles. We expect this system can accurately convey the user's nonverbal communication, in particular the user's gaze direction in 3D space that is not correctly transmitted by using a 2D screen (which is known as "the Mona Lisa effect"). To evaluate how this system can contribute to the communication, we conducted three experiments. The first one examines the blind angle of a face-shaped screen and a flat screen, and compares the ease with which users can distinguish facial expressions. The second one evaluates how the direction in which the remote user's face points can be correctly transmitted. The third experiment evaluates how the gaze direction can be correctly transmitted. We found that the recognizable angles of the face-shaped screen were larger, and that the recognition of the head directions was better than on a flat 2D screen. More importantly, we found that the face-shaped screen accurately conveyed the gaze direction, resolving the problem of the Mona Lisa effect.

Categories and Subject Descriptors

H.5.3 [**Group and Organization Interfaces**]: Computer-supported cooperative work

General Terms

Human Factors

Permission to make digital or hard copies of all or part 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. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.
AVI'12, May 21–25, 2012, Capri Island, Italy.
Copyright 2012 ACM 978-1-4503-1287-5/12/05... $10.00.

Figure 1: Communication using LiveMask

Keywords

Telepresence, eye gaze, face shape

1. INTRODUCTION

Telepresence is a technology that gives people the feeling of being present and realistic sensory at a remote location. We consider that telepresence systems are of three types in point of focusing on tele-existence or realism: One, the video conferencing systemshphalocisco, which convey these sensory equally. The second is focused on realism. User wears a head mounted display, and acquires a immersion and highly realistic sensory[14]. The third, avatar operating systems[1][13][16]. The remote speaker teleoperate own avatar mounted a projector and a mike to show his presence. Our system is related to the last category.

We consider nonverbal information is very important to convey these sensory to substantialize richer communication. It is said that nonverbal information is an important role in communication. Previous teleconference systems can convey the appearance of user and sound but hardly convey the head motion and gaze direction which needs spatial location awareness. To address these problems, we developed the LiveMask "surrogate" system with a face-shaped screen.(Figure 1) This system tracks the remote participant's face image and head orientation by computer vision, and moves the face screen using the 3 DOF (degree of freedom) mechanism. (Figure 2) Since the face screen is molded based on the 3D data of the participant's real face, the projected image is accurate even when it is seen from the different angles. We expect this system can convey non-verbal information of the participant, and in particular, the participant's gaze direction, which is not possible using a 2D screen.

2. RELATED WORK

Hydra[4] is a table-top telepresence system that facilitates distance communication for 4 people. Each small unit acts as a surrogate for each remote speaker and has a monitor to display their faces. We consider it better to use a face-shaped screen as a surrogate, which can provide a greater sense of fidelity and trust.

In the Talking Heads project, a face image was projected on face-shaped screen[6]. Our study tracks face and head by a camera, so we don't have to wear any devices just as in a real face-to-face situation. And the effectiveness and superiority of 3D form screen is not revealed in comparison to 2D screen.

Geminoid[17] is a tele-operated android for tele-communcation. While it has strong presence because of having a very human-like appearance, which is seen as uncanny. Ishiguro et al. concluded that people felt a stronger presence from a human-like telecommunication medium than from other ones. Android robots have strong existence, but its versatility is low because cost and time of developing each person's robots. The system with a face-shaped screen is making trade-off of existence and versatility.

Moreover, a number of studies have mimicked the remote speaker's gestures, which are controlled remotely by an operator. Nakanishi et al. noted the effectiveness of social telepresence that can approach the viewer[15]. Also Mebot[7] is a telerobot that performs social expressions, making viewers perceive a sense of engagement and familiarity.

It is well known that eye gaze plays a crucial role in dynamics of conversation such as turn-taking, controlling the flow of interaction[12], controling the other person's reaction, and maintaining a level of intimacy[8]. Most current telepresence systems use a flat screen, which causes viewers to feel eye contact at any angle, creating what is called the Mona Lisa effect[11]. There are some studies on the difference of eye gaze by forms of projected substance[9][10]. However, there is a need for more experimental proof concerning the use of human-shaped displays to realize high realistic sensation in telepresence. We couldn't find the previous research on whether the result is the same as CG image projection, so we tested the use of a real human face-shaped screen and human image.

3. *LIVEMASK* : TELEPRESENCE SYSTEM WITH A FACE-SHAPED SCREEN

Face-shaped screen: is made from the mold of an actual human face. We tested our telepresence system with various screen types. Then we found that the resultant projected image was not realistic and was discomforting unless we used a real face-shaped screen to represent the person. We used a 3D scanner to obtain 3D information about the user's face and provided this as input to a 3D printer to make a male mold. Finally, we manufactured a face-shaped screen using thermoplastic sheet.

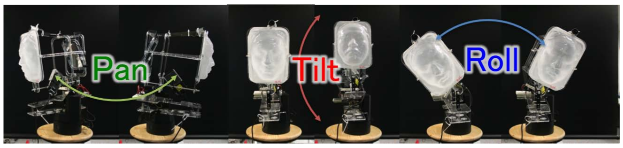

Figure 2: 3DOF: (1) Pan: shaking the head from side-to-side, (2) Tilt: nodding the head up and down, and (3) Roll: inclining the head from side-to-side.

The Driving Mechanism: this system has a mechanism that permits 3DOF, panning, tilting, and roll. We controlled the system by tracking head motion through the use of the FaceAPI library[5].
Face Image: the texture image created by FaceAPI when performing facial recognition is projected on the system. This image is applied to the front face as it warped. We make use of this texture and manipulate the image to adjust for errors and deviations due to projection.
Projector: we use a Pico Projector manufactured by Microvision. This projector is laser-based, so it is suitable for projection on an irregular surface without requiring focus adjustment.

4. EXPERIMENT

Preparatory Experiment: Blind angle

In a real meeting, people seated around a table are not always sitting in front of the telepresence system. Imagine a situation in which the number of telepresence displays on the table is the same as the number of remote participants in the teleconference. In a flat-screen system, it would be difficult for a person sitting beside the screen to read expressions because it is difficult to view the remote speaker's face at that angle. LiveMask, however, can be seen more easily because its features project 3-dimensionally from the screen surface, so that expressions can be read not only from the front, but also from the side and from above, just like a real human face. There were 8 participants in the experiment.
Purpose: The goal is to examine the blind angle of two types of screen. The criterion being evaluated was whether participants could recognize the person projected on the screen.
Study Setting: The LiveMask was set on the desk and the participant's seat was in fixed position 110cm from the LiveMask. The images of three people were prepared for projection. The angle at which the system was set relative to the participant is defined as $0°$ and the experiments began with this angle.
Procedure: 1)Participant sits on the seat placed in front of the LiveMask system and adjusts the height of the seat if needed. 2) A flat screen is fitted onto the front of the LiveMask. 3) The operator projects the images of three people randomly, and participants identify the subjects from different angles. When the participant can identify the projected image with complete accuracy, then that angle of viewing is noted as not being a blind angle. 4) Steps 1) to 3) are repeated with a face-shaped screen fitted onto the front of the LiveMask.
Result: **Flat screen** has wider blind angle $190°$than **face-**

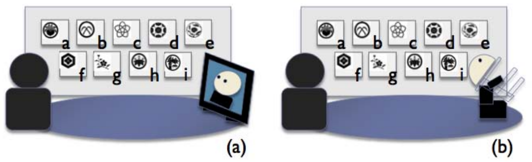

Figure 3: Two conditions of the experiment

shaped screen 150°.

Experiment1: Facial expressions

Purpose: Next, we studied the ease with which different facial expressions could be distinguished using a face-shaped screen versus a flat screen.
Study Setting: We prepared 20 images, each of which represents one person reacting to a stimulus, with five different kinds of expression(joy, sadness, anger, surprise, and fear). Using the same experimental environment as that in Preparatory Experiment, the operator projected those images onto either a flat screen or a face-shaped screen randomly, and participants had to identify the expressions that they saw on the screens.
Procedure is same as preparatory experiment.
Result: Flat screens elicited a greater degree of recognition accuracy than face-shaped screens, but the difference was negligible, such that the screens could be considered the same. (See Table1.)

Experiment2: LiveMask head gesture vs. indication on 2D display

Furthermore, we conducted an experiment to recognize the directional instructions indicated by LiveMask implemented head gestures. In a real meeting, people often discuss while turning in the direction of a speaker and explain by turning towards a whiteboard or wall that includes content.
Purpose: The goal is to compare the ease with which the head gestures of LiveMask could be understood as compared to images projected on a 2D display.
Study Setting: Nine signs were attached to a wall; they were listed in alphabetical order from a) to i). Assuming that the remote speaker is looking at one sign among the nine, participants were asked to identify which sign the remote subject is looking at. One way to do this is by studying the 2D images.(Figure 3(a)) The other way is by checking the direction towards which the LiveMask is pointing. (Figure 3(b)) These instructions are preconfigured.
Head gestures: The system had advance knowledge of the nine different neck positions. In this case, differences in the identification of facial expressions may depend upon the line of sight, so we projected the same facial expressions.
2D image: The remote speaker sits on a chair in study setting and looks at the nine signs by using her head and a direct line of sight. The scenes shown below were pictures taken from the viewpoint of the participants.

Table 1: The success rate of facial recognition

	Average	Std.Dv.
i)Flat screen 85°	71.9 %	0.113
i)Face-shaped screen 85°	70.0 %	0.963

Procedure: 1) While setting up the test, participant is blindfolded. 2)When setting up is complete, the blindfold is removed and the Operator asks participants which signs the subject indicate. The participant answers as soon as possible. The Operator measures the time for each answer. 3) After testing all position randomly, changed conditions. Each condition conducted again by 1 set.
Result: When using a LiveMask, many participants could complete identification with almost total accuracy. While using the 2D images, only 19.4 % accuracy was obtained. However, if we were to count the identification of adjacent signs as also correct, the percentage of approximate accuracy rose to 83.3 %. Approximate accuracy is the rate counted identification of adjacent signs as also correct. (e.g. If the right answer was *e*, but *d*, *h* and *i* were also regarded correct answer.) Further, the response time to questions was about twice as long when a 2D display was used. (Table2)

Experiment3: Perception of eye gaze

Purpose: The goal is to clarify whether there is a difference in direction of eye gaze when using a flat screen and a face-shaped screen.
Study Setting: In advance, prepare the photos. One person looks from -30° to 30° in increments of 10°. On the same setting as Experiment1, these photos are projected onto two kinds of screens. Participants look at 7 photos at each positions from -30° to 30° far from 170cm. The test is conducted 7 photos * 7 position * 2 set.
Procedure: 1) Participants adjust the seat at -30° , matching the height of the flat screen's eyes. 2) The participant views the 7 photos, and answers if the eye gaze matches. After answering, the chair is moved by 10° in clockwise direction. 3) Repeat until 30°, then change the screen.
Result: Table 3 shows a summary of the results. When using face-shaped screens the accuracy was 75.9%. On the other hand, when using flat screens acurracy was lower: 36.6 %. Also for each of the images, the answers are distributed diagonally for the face-shaped screen, while the answers for the flat screen are distributed horizontally around 0°.(See Figure 4) This means that the directions of gaze are conveyed to participants when using a face-shaped screen. For the flat screen, the images 0°(front side) tend to make eye contact at any angle.

5. DISCUSSION

According to the **preparatory experiment**, face projection on a face-shaped screen could be recognized at a wider angle than on a flat screen. The face-shaped screen can be used not only with people sitting around a large conference table, but also with viewers in random locations in a room. In the result of **Experiment1**, there was a slight difference in the ability to read expressions between the 2 screens, but subjective opinions were divided over the ease of reading expression. The result in **Experiment2** revealed a big difference. Judging from 2D images, the accuracy is very low.

Table 2: Experiment 2 Result

Condition(second)	Accuracy	Approximate accuercy
2D display (10.31s)	19.4 %	83.3%
LiveMask (5.45s)	87.5%	100%

Condition(second) : Average response time in each condition

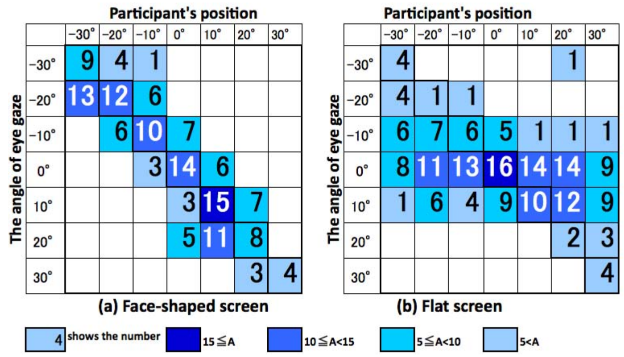

Figure 4: Experiment 3 - Result of eye gaze

Figure 5: Viewpoint is at -30°:(1)(2), (3)(4) are projected on each image of gaze direction 0°, -30°.

However if one regards images that were adjacent to the correct image to be correct, each participant recorded close to 80% approximate accuracy. From participants' comments, "the person's eye gaze was so wide that subject on 2D display appeared to look at 3 or 4 signs instead of looking at just one." 2D images are capable of showing the rough direction, but cannot indicate the precise position and the direction. This is because human eye gazes appear wider when the subject in a 2D image looks at a certain point. In the result of **Experiment3**, the incorrect eye contact occurred when the flat screen was seen from diagonal angle. See Figure 5. By "incorrect", we mean that the remote user ' s gaze direction is not correctly perceived from the point of observation(the Mona Lisa effect). We feel like making eye contact with (1) and (4), but the subject of (1) in fact looking at 0° like (2). If the subject of (3) looks at -30°, we will make eye contact with (3) like (4). However, the subject of (3) seems to look at the different direction. Hence, face-shaped screen resolves the mona lisa effect.

6. CONCLUSION

In this paper, we presented a telepresence system that has a face-shaped screen that serves as a surrogate, which moves in correspondence with a remote participant. We studied the effectiveness of the face-shaped screen and the degree of recognition it elicited for directions indicated by head gestures and eye gaze. The evaluation results indicate that the face-shaped screen may be used not only for face-to-face communication, but also for variety of interactions, due to its wide angle of visibility. Although there was no measurable difference in participants' ease of facial expression recognition on the flat screen versus the face-shaped screen, many remarked that on the face-shaped screen, facial expressions are easy to understand and the screen itself is more attractive. More importantly, we confirmed that gestures made on LiveMask were definitely more understandable than images on the 2D display. We also confirmed that a face-shaped screen correctly transmits gaze direction by resolving the "Mona Lisa effect", a common gaze-recognition problem in face-to-face communication with a 2D screen. In our future work, we would like to clarify the sense of presence and the feasibility of this system in a tele-work situation.

Table 3: The eye sights

	Accuracy	Std.Dv.
Face-shaped screen	75.9 %	0.17
Flat screen	36.6 %	0.30

Accuracy : percentage of questions answered correctly

7. REFERENCES

[1] AnyBots https://www.anybots.com.
[2] Cisco Telepresence. http://www.cisco.com/en/US/products/ps7060/index.html.
[3] HP Halo. http://hphalo.org/.
[4] Hydra. http://www.billbuxton.com/hydraNarrative.htm.
[5] Seeing Machine faceAPI. http://www.seeingmachines.com/product/faceapi/.
[6] Talking Head Projection. http://www.naimark.net/projects/head.html.
[7] S. O. Adalgeirsson and C. Breazeal. Mebot: a robotic platform for socially embodied presence. *Proc. HRI'10*, 15–22, 2010.
[8] M. Argyle, L. Lefebvre, and M. Cook. The meaning of five patterns of gaze. *European Journal of Social Psychology, 4(2)*, 125–136, 1974.
[9] J. Beskow and S. Al Moubayed. Perception of gaze direction in 2d and 3d facial projections. *Proc. FAA'10*, 24, 2010.
[10] F. Delaunay, J. De Greeff, and T. Belpaeme. A study of a retro-projected robotic face and its effectiveness for gaze reading by humans. *Proc. HRI'10*, 39–44, 2010.
[11] S. A. M. Jens Edlund and B. Jonas. The mona lisa gaze effect as an objective metric for perceived cospatiality. *Proc. IVA'11*,439–440,2011.
[12] A. Kendon. Some functions of gaze-direction in social interaction. *Acta Psychologica, 26(1)*,22–63, 1967.
[13] M. K. Lee and L. Takayama. "Now, i have a body": uses and social norms for mobile remote presence in the workplace. *Proc. CHI'11*,33–42,2011.
[14] H. Nagahara, Y. Yagi, and M. Yachida. Wide field of view head mounted display for tele-presence with an omnidirectional image sensor. *CVPR Workshop*, 7:86,2003.
[15] H. Nakanishi, K. Kato, and H. Ishiguro. Zoom cameras and movable displays enhance social telepresence. *Proc. CHI'11*,63–72,2011.
[16] E. Paulos and J. Canny. Social tele-embodiment: Understanding presence. *Auton. Robots*,87–95,2001.
[17] D. Sakamoto, T. Kanda, T. Ono, H. Ishiguro, and N. Hagita. Android as a telecommunication medium with a human-like presence. *Proc. HRI'07*,193–200,2007.