

Recommended Paper

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

KANA MISAWA^{1,†1,a)} YOSHIO ISHIGURO^{1,b)} JUN REKIMOTO^{1,c)}

Received: November 1, 2012, Accepted: January 11, 2013

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. We expect this system can accurately convey the user's nonverbal information in remote communication. In particular, it can transmit the user's gaze direction in the 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 communication, we conducted three experiments. As the results of these evaluations, we found that the recognizable angles of the face-shaped screen were bigger, and the recognition of the head directions was better than those of the flat 2D screen. More importantly, we also found the face-shaped screen could accurately convey the gaze directions and it solves the Mona Lisa effect problem even when screen size is reduced.

Keywords: Telepresence, nonverbal information, eye gaze

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 systems [2], [3], 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 [16]. The third, avatar operating systems [1], [15], [19]. The remote speaker teleoperates 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 (**Fig. 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 (**Fig. 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,



Fig. 1 Communication applied on LiveMask.

and in particular, the participant's gaze direction, which is not possible using a 2D screen.

In addition, we examine the value and effect of scale reduction of such face-shaped displays. We expect small size face displays retain the benefit of real-size talking-head type telecommunication systems. It is easier to transport or put on a desk, and it can be worn on the shoulder of the local participants so that people bring it like a small buddy [8]. However, it is not clear how such reduced-size face screen might change the quality of nonverbal communication. We thus conducted an experiment using a 1/14 scale face display, and found critical nonverbal information, such as gaze-direction, is still correctly transmitted even when face size is reduced.

¹ The University of Tokyo, Meguro, Tokyo 153-8505, Japan

^{†1} Presently with Dentsu Inc.

^{a)} kana.misawa@gmail.com

^{b)} ishiy@acm.org

^{c)} rekimoto@acm.org

The content of this paper was reported at Interaction 2012 in March 2012 and the paper was recommended to submit for Journal of Information Processing society.

2. Related Work

When studying telepresence, embodying the remote speaker who doesn't exist at same place is an important issue. In recent years, researchers have strived to convey not only realistic sensation and tele-existence, but also psychological elements such like affinity, confidence, and engagement. We focused on non-verbal information, which is important as communication, and developed a "surrogate" set in remote place, which can convey head gestures and eye gaze.

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.

Kidd and Breazeal presented that a robot is perceived more engaging and credible than a computer-graphic agent [14]. An android is a robot designed to look and act like a human and Geminoid [20] is a tele-operated android for tele-communication. 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 [17]. 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 [13], controlling the other person's reaction, and maintaining a level of intimacy [9]. Most current telepresence systems use a flat screen, which causes viewers to feel eye contact at any angle, what is called the Mona Lisa effect [12]. There are some studies on the difference of eye gaze by forms of projected substance [10], [11]. 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

3.1 System Overview

The system is composed of remote-operator side tele-operating LiveMask and receiver side where LiveMask is placed (Fig. 3). Voice communication is supported by using Skype. Operator looks at the remote place from usb camera view mounted on LiveMask (Fig. 4). In future work, usb camera should be mounted corresponding to eye gaze of LiveMask. Also, the size of person showed on monitor is to be adjusted.

3.2 Face-shaped Screen

The face-shaped screen, which is an essential part of this system, 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

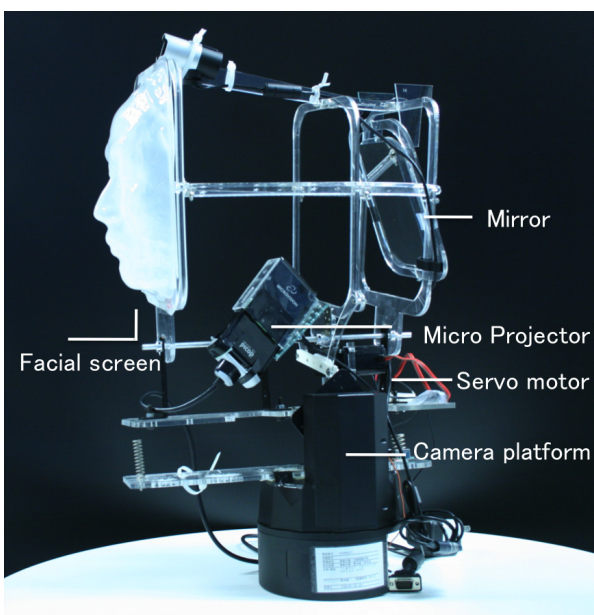


Fig. 2 LiveMask - To make the system small, we examined the angle of field for the projector, and reflect the image using the first-surface mirror. The black object is a camera platform.

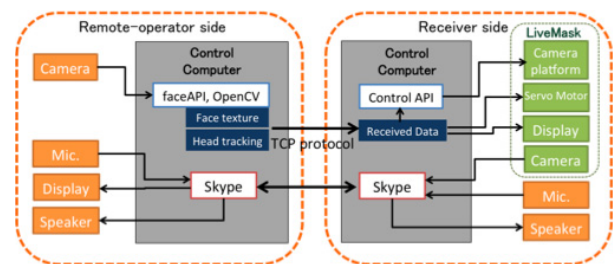


Fig. 3 System overview: The head motion and the face images are extracted by faceAPI and process these data to send to receiver side.

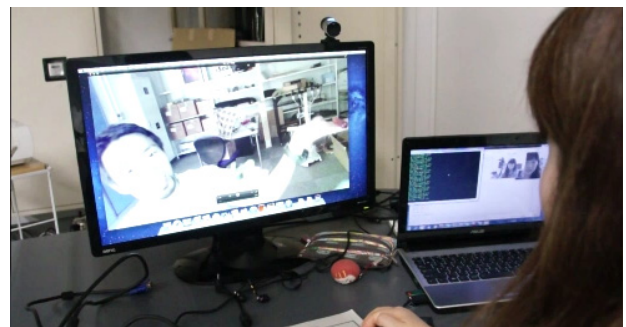


Fig. 4 Remote operator looks at the local user from camera view.



Fig. 5 3DOF - Pan-shaking the head from side to side, Tilt-nodding up and down, Roll-inclining the head from side to side.



Fig. 6 Image Processing - Screenshot of facial recognition. In the image on the left, the state of head inclined has been adjusted, resulting in the right texture image. The image is inverted for projection.

discomforting unless we used a real face-shaped screen to represent the person. At the beginning, we tried to create a moulage from a female facial mold, which required great care. Later, we established another means to make the face-shaped screen, which is the method we follow currently. 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.

3.3 Head Motion Mechanism

This system has a mechanism that permits 3DOF, panning, tilting, and an inclination of the head (**Fig. 5**). We controlled the system by tracking head motion through the use of the FaceAPI library [5], which is a product developed by a company called Seeing Machine. Pan-tilt head motion is operated by using camera platform, and that roll head motion is operated with servo motor. The axis of rotation of human head motion is located under the ears. While, the axis of rotation of LiveMask is around neck, so that pan motion rotates widely. This is because projector is mounted inside the system to shorten the distance of projection and to make structure of reflecting. According to the center of gravity of entire system, current system doesn't correspond to real human movement.

3.4 Extraction of the Face Image

The texture image is created by FaceAPI when performing facial recognition is projected on the system. From participant's face image, head position and angle are estimated, and frontal face image is restored (**Fig. 6**). We make use of this texture and manipulate the image to adjust for errors and deviations due to projection.

3.5 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.

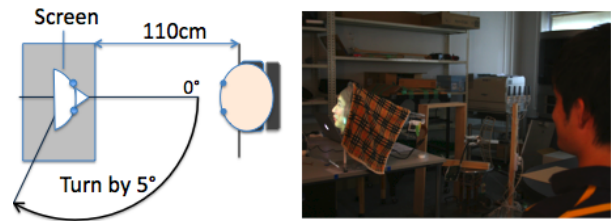


Fig. 7 Experiment 1 setup - On the left: the perspective image. On the right: participants look the screen from a distance and answer the questions.

4. Experiment

We conducted the experiments to investigate how effective the face-shaped screen and head motion of LiveMask contribute communication. The first one is examining the blind angle of a face-shaped screen and a flat screen, and evaluating the ease of distinguishing facial expressions. We expect these 2 kinds of screen have a difference as reading expressions. The second experiment is evaluating how the face directions of the remote user can be correctly transmitted. This experiment is for studying head motion. The third experiment is evaluating how the gaze direction can be correctly transmitted. By using real human face-shaped screen, we tested eye gaze can be transmitted.

4.1 Preparatory Experiment: Blind Angle

In a real meeting, people are seated around a table, so they don't always sit in front of the telepresence system. Imagine a conference situation in which the number of telepresence displays on the table is the same as the number of participants joining from a distance. If screens of system are flat, it's difficult for a person sitting next to the system to read expressions because he/she has to view the remote speaker's face at that angle. On the other hand, LiveMask can be more easily seen because its profile has an irregular surface. If a screen has a three dimensional surface, expressions can be read not only from an anterior view but also from the side and above, just like a real human face. We conducted an experiment to investigate the blind angle, which is the angle at which people stop recognizing the projected face on the screen. There were 8 participants in the experiment.

4.1.1 Purpose

The goal is to examine the blind angle of each of the two types of screen. We define the blind angle that people can't recognize the identity of the person projected on the screen. The criteria of judgement is whether participant can recognize the person projected on screen.

4.1.2 Study Setting

The LiveMask is fixed on the desk, and the seat of participants is in a fixed position relative to the LiveMask (**Fig. 7**). The flat screen for the experiment has been cut to adjust the size of the display. Three people's images are prepared for projection. The angle the system sets to the participant is defined 0° and the experiments start from this angle.

4.1.3 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.

4.1.4 Result

Flat screen: At 90°, the flat screen is invisible to all participants. However, in an inclination range of 5° to 85°, all participants could identify the remote subject (Fig. 8).

Face-shaped screen: All participants could identify the subject at an angle of 90° because of the uneven contours of the 3D facial mold. Next, the operator inclined the face-shaped screen to the opposite side. Participants could accurately identify the subject from approximately 105° to 110°.

4.2 Experiment 1: Facial Expressions

Next, we studied the ease and degree of discrimination between different facial expressions when using face-shaped screens as opposed to flat screens.

4.2.1 Purpose

The goal of this experiment is to study the difference in the degree of recognition as well as the ease of discrimination between facial expressions while using a face-shaped screen as opposed to a flat screen.

4.2.2 Study Setting

We prepared 20 images (Fig. 9), each of which represents one person reacting to a stimulus, with five different kinds of expression. These facial expressions denote joy, sadness, anger, sur-

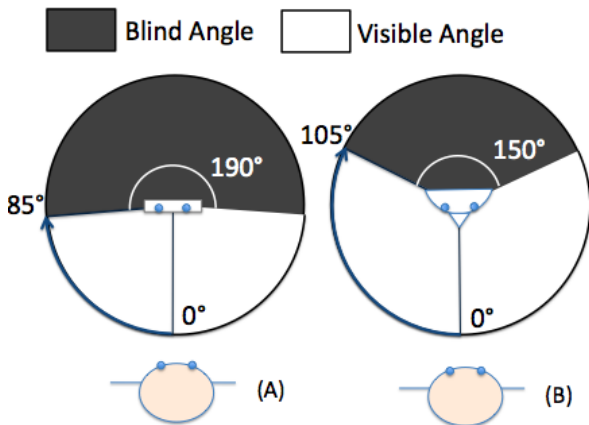


Fig. 8 Blind angle - (A) is a flat screen. More than 85° is a blind angle. (B) is a face-shaped screen. It has a 3D surface, making the visible angle wider. More than 105° is a blind angle. We examined the angle indicated by arrows. The result is assumed to be symmetric, then inverted.



Fig. 9 Five kinds of emotion used for experiment - Each type is prepared by four photos.

prise, 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 (Table 1).

4.2.3 Procedure

1) Before the experiment, we showed each participant five kinds of pictures (joy, sadness, anger, surprise, and fear), to establish a common baseline between all participants, as to which kinds of facial expressions were being shown.

2) From preparatory experiment, the blind angle of the flat screen was obtained as 85°. Therefore, we adopted this viewing angle to establish how often participants identified facial expressions correctly.

3) We then repeated steps 1) to 2) using face-shaped screens and adopting the blind viewing angle for face-shaped screens, which also was obtained in preparatory experiment.

4) After the experiment, we asked participants what their benchmarks are to determine and identify expressions in the study. In addition, we solicited their opinion on which type of screen made it easier to identify expressions.

4.2.4 Result

Flat screens were marked by a greater accuracy in recognition than face-shaped screens, but the difference is minor enough to regard the use of both as yielding the same result. In the case of the face-shaped screens, participants could see the screen (105° to 110°), but it was difficult to recognize the expression (See Table 2).

4.3 Experiment 2: Compare Head Gesture of Livemask to 2D Screen

Furthermore, we also conducted an experiment to recognize the directional instructions indicated by LiveMask implemented head gestures to compare usual 2D display. 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. Where is the remote speaker turning on? We were thus able to conduct an experiment with head gestures, and compared the ease of distinguishing when using 2D display, as opposed to the gestures of LiveMask. We also test with the two kinds of different screen attached on LiveMask.

4.3.1 Purpose

The goal is to compare the head gestures of LiveMask to the images projected on a 2D display. Currently, people use teleconference system on 2D. In this experiment, we compare how this

Table 1 Experiment 1: Condition.

	Shape	Expression	Head gesture
i	Flat screen	5 types	-
ii	Face-shaped screen	5 types	-

Table 2 Experiment 1 Result: The success rate of recognition.

	Average	Std.Dv.
i) Flat screen 85°	71.9%	0.113
i) Face-shaped screen 85°	70.0%	0.963
ii) Face-shaped screen 105°-110°	25.0%	0.134

(i) : Flat screen

(ii) : Face-shaped screen

Table 3 Experiment 2: Condition.

	Shape	Expression	Head gesture
I	2D display	Pointing*	Pointing (on 2D)
II	LiveMask with flat screen	Static	Pointing (on 3D)
III	LiveMask with face-shaped screen	Static	Pointing (on 3D)

Pointing*: User turns on and looks forward to an image. The direction of eye sight is conform with head gesture.

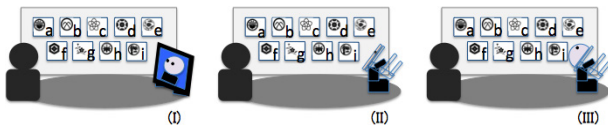


Fig. 10 Experiment 2: Condition - (I) is a image for study setting of 2D display. (II) is a image for study setting of LiveMask system attached a flat screen, and (III) is attached a face-shaped screen.



Fig. 11 2D display condition - Each 9 images is projected on a flat display.

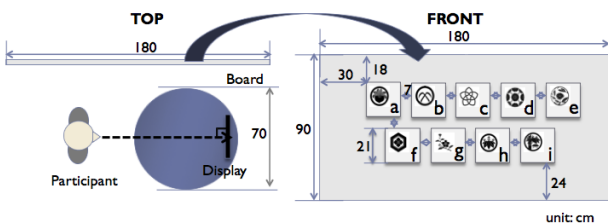


Fig. 12 Experiment 2 Study Setting.

system convey the head gesture of remote speaker to LiveMask system. Furthermore, we investigate the difference of screen shape effects on LiveMask gesture (Table 3).

4.3.2 Study Setting

Nine signs were attached (listed in alphabetical order from *a* to *i*) to the wall. Assuming that the remote speaker is turning to one sign among nine, participants must identify which sign the remote subject is turning to. One way to do this is by studying the two-dimensional images (Fig. 10 (I)). The second way is by checking the direction towards which the LiveMask is pointing. This LiveMask has a flat screen (Fig. 10 (II)). The last is by checking the direction of LiveMask attached on a face-shaped screen (Fig. 10 (III)). These instructions are preconfigured.

2D image (Condition I) : The remote speaker sits on a chair in study setting and looks at the nine signs by using her head. The scenes shown below were pictures taken from the viewpoint of the participants (Fig. 11).

Head gestures (Condition II, III) : 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 on the screen.

4.3.3 Procedure

1) First, start by studying the two-dimensional images. The

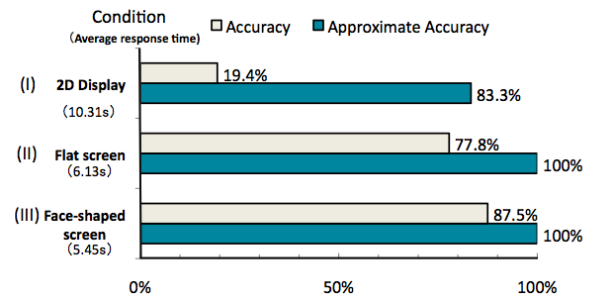


Fig. 13 Experiment 2 Result - It takes about 2 times longer to use 2D display than LiveMask. Approximate accuracy is the rate counted identification of adjacent signs as also correct (e.g., If the right answer was *e*, but *d*, *h* and *i* were also regarded correct answer).

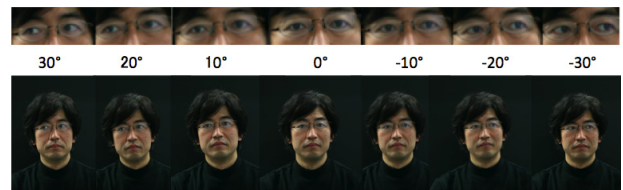


Fig. 14 Experiment 3 Eye Gaze - Facial position is fixed and eye gaze changed only by using his eyeballs. From left, the direction of gaze is changed 30°, 20°... in increments of 10°.

participant sits on left side. While setting up the test, participant is blindfolded (Fig. 12).

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 nine position randomly, it is changed to LiveMask system. Each test conducted again by 1 set (9 times).

4.3.4 Result

When using a LiveMask with a face-shaped screen (Condition III), many participants could complete identification with almost total accuracy, 87.5%. The accuracy of Condition II using flat screen was 77.8%, which was lower than that of Condition III, but higher than that of Condition I. While using the 2D display under Condition I, 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%. The response time to a question was apt to take 2 times longer when using a 2D display (Fig. 13).

4.4 Experiment 3: Perception of eye Gaze

In Experiment 2, we studied the effect of movement of LiveMask. This experiment, we limit condition and only test eye sight.

4.4.1 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.

4.4.2 Study Setting

In advance, prepare the photos (Fig. 14). One person looks from -30° to 30° in increments of 10°, only changing his eye gaze. On the same setting as Experiment 1, these photos are projected onto a flat or a face-shaped screen of LiveMask (Table 4). As Fig. 15, at each position participants look at 7 photos, and perform in total 7 positions (This is 1 set). Each screen is tested by

Table 4 Experiment 3: Condition.

	Shape	Expression	Head gesture
i	Flat screen	Eye sight	-
ii	Face-shaped screen	Eye sight	-

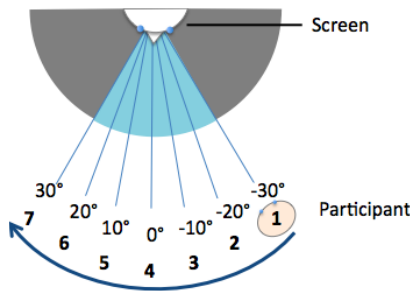


Fig. 15 Experiment 3 Procedure - Participant moves from -30° to 30° by adding 10° and answers whether he/she can make a contact with an image projected on a screen.

Table 5 Experiment 3 Result: The eye sights.

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

Accuracy : percentage of questions answered correctly.

2 sets.

4.4.3 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 it reaches an angle of 30° . Then, repeat steps 1 to 3, using the face-shaped screen instead of the flat screen.

4.4.4 Result

Table 5 shows a summary of the results. When using face-shaped screens the accuracy was 75.9%. On the other hand, when using flat screens accuracy 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° (Fig. 16). 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. Look at Fig. 17, the direction of eye gaze is not different under condition (a): front viewpoint. While, directional eye gazes are different between two screens under condition (b): diagonal viewpoint (Fig. 18). Actually, the person of (1) looks toward front side (0°), but we feel making eye contact with (1). At gaze direction -30° , (4) projected on face-shaped screen seems to make eye contact with us. However we can't make eye contact with (3), which seems to look at the another direction.

4.5 Complementary Experiment: Effect on Perception of Eye Gaze by Downsizing

In our previous experiment, we recognized the Mona Lisa effect occurred when the image subject looking at front side projects, we felt like making an eye contact with the image at any angle, fundamentally we shouldn't have felt. In this experiment,

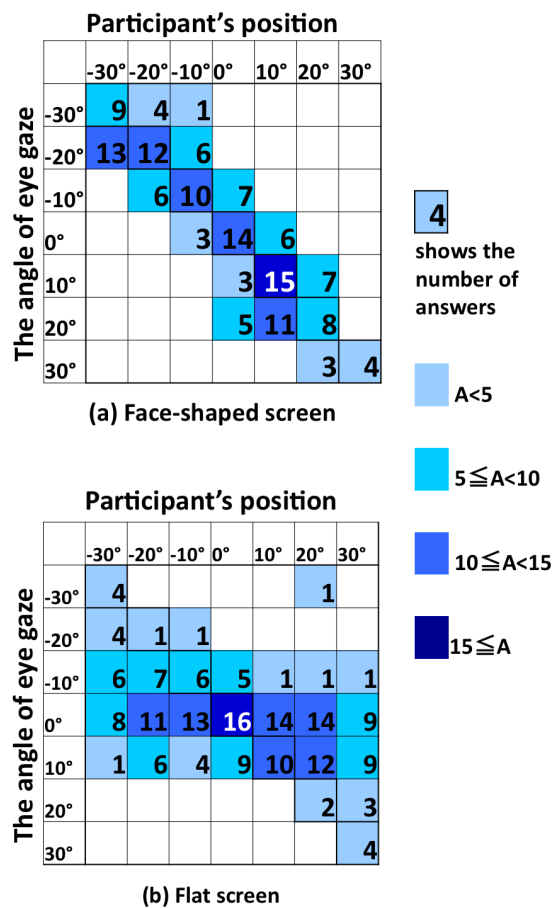


Fig. 16 Experiment 3 - Result of eye gaze : Grid shows the number of answers. The color of grid depends on the number of answers. The answers are distributed diagonally for the face-shaped screen, while the answers for the flat screen are distributed horizontally around 0° .

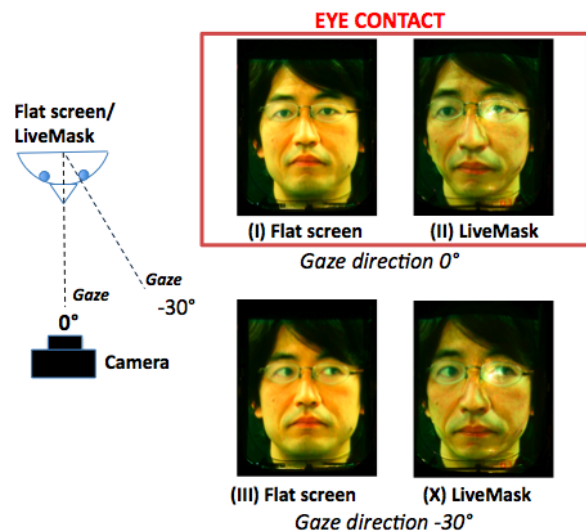


Fig. 17 Condition (a) The camera viewpoint is at 0° - (I) and (II) are projected a same image of the gaze direction 0° . (III) and (X) are also projected a same image of the gaze direction -30° . The direction of eye gaze is the same between flat screens and face-shaped screens. We feel like making eye contact with (I) and (II).

we examine the system can convey the direction of eye gaze even if the face-shaped screen is reduced at 1/14 (Fig. 20). Small system has some merits. It can be placed on the desk without excess space, and that be used as portable and wearable system. In pre-

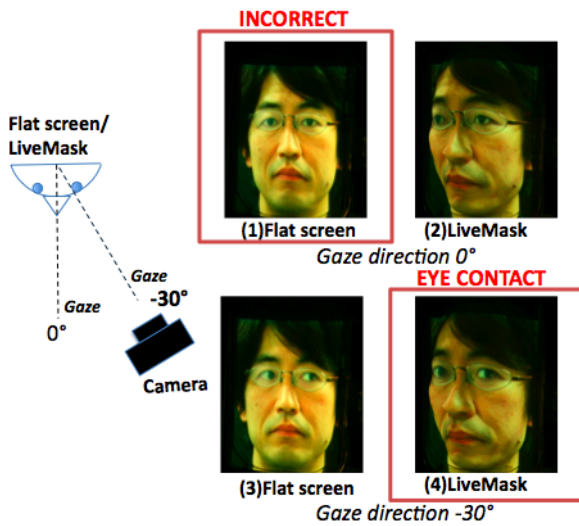


Fig. 18 Condition (b) The viewpoint is at -30° - (1) and (2), (3) and (4) are projected on each image of gaze direction 0° , -30° . The direction of eye gaze is different between flat screens and LiveMask. INCORRECT indicates the remote user's gaze direction is not correctly perceived from the observed point (the Mona Lisa effect). We feel like making eye contact with (1) and (4), but the subject of (1) in fact is 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.



Fig. 19 Comparison of the size: On the left is 1/14 scale face-shaped screen. On the right is the real size face-shaped screen.



Fig. 20 On the right: Comparison with a real-size face shaped screen and downsized the face-shaped screen at 1/14. On the left: Image of mounting around the shoulder.

vious experiment, we tested that under study setting that the participant face on the system by using real size face-shaped screen. In this experiment, suppose the scene the participants wear the system around their shoulders and read the eye gaze by looking into face-shaped screen.

4.5.1 Purpose

The objective is to recognize which direction the subject looks

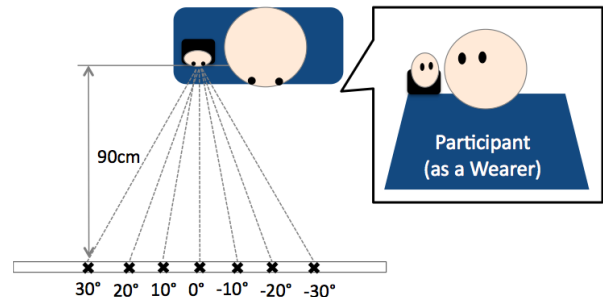


Fig. 21 Study Setting: The system on the shoulder displays the Fig. 14 images randomly. The participants answer one point which the subject looks at among 7 points.

Table 6 The result.

	Average	Std.Dv.
The accuracy	44.5%	0.181
The approximate accuracy	88.6%	0.160

The approximate accuracy : count the identification of adjacent point as also correct.

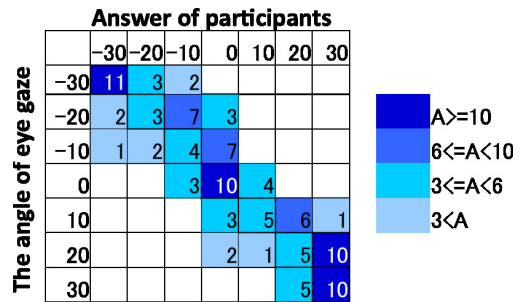


Fig. 22 Distribution of the all answers: The answers are counted in the same way with Fig. 16.

in by reading his eye gaze using a small face-shaped screen from the wearer's viewpoint.

4.5.2 Study Setting

Photos are prepared in advance (Fig. 14). One person looks from -30° to 30° in increments of 10° , but only changes his eye gaze. The subject of the picture is supposed as the remote user. A white board with 7 points is prepared. These 7 points match the direction of the remote user's eye gaze. The participants are supposed to mount the system on the shoulder in the experiment (Fig. 21). The number of participants is 5.

4.5.3 Procedure

- 1) Participants adjust the height so that their shoulder is under the system.
- 2) The operator tells the participant that the person looks in any direction, and let the participants respond if the eye gaze matches. The answer is recorded by the operator.
- 3) The participant views the 7 images randomly, and repeats 3 sets.

4.5.4 Result

44.5% accuracy was achieved. However, if we were to count the identification of adjacent signs as correct, the percentage of approximate accuracy increased to 88.6%. See Fig. 22, which shows distribution of all answers. The answers crowded together diagonally. Thus, we assumed that participants could understand the remote user's gaze direction. We found that the system correctly transmitted even when the face size is reduced.

5. Discussion

According to the **Preparatory experiment**, face projection on a face-shaped screen could be recognized at a wider angle than a flat screen. The face-shaped screen has less of a blind angle, so it can not only be used sitting around a large conference table, but also in a situation where viewers are in random locations in a room. A face-shaped screen can also be useful to effect tele-existence by making use of a wide visible angle. In the result of **Experiment 1**, 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. Two people answered that flat screens are easier to understand and felt that it was difficult to see facial features on a 3D display. They said that features such as the height of nose and curve of cheek made the other side of the face invisible. They tried to read expressions by viewing the full face. The study setting is static in this experiment, but LiveMask is able to turn towards viewers in a real meeting, so these problem will be solved. Participants who saw the LiveMask for the first time said “it has a sense of presence and is more attractive.” The result in **Experiment 2** for judging directional instructions revealed a big difference. Gesture of LiveMask gave easier directions than that of 2D display. About screen type, a face-shaped screen was slightly recognizable than a flat screen. Judging from 2D images, the accuracy is very low. However if one regards images that were adjacent to the correct image to be correct, each participant recorded close to 80% approximate accuracy. There are some reasons for the worse accuracy using a 2D display. From participants’ later 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. If the screen has moving mechanism, the system can give us directional information clearly according to the result of LiveMask. To convey the remote user’s intense, it’s an approach to implement user’s movement. In **Experiment 3**, we tested the differences in the direction of eye gazes when using the two types of screens. In the result, changes in eye gaze were recognizable on the face-shaped screen, but not on the flat screen. It is difficult to recognize small changes in eye gaze and eye contact on the flat screen, so using the face-shaped screen could resolve this problem. When we see the person who looks forward, projected on a flat screen, we failed to make eye contact with him at any angles because of the mona lisa effect. Thus, we misunderstand if the form of screen is flat. On the other hand the effect doesn’t occur on the face-shaped screen, and we established appropriate eye contact with the person at each gaze direction. Most telepresence systems have used a flat screen, but the form should be three dimensional at least part of the eye to convey the remote speaker’s gaze direction (**Fig. 23**).

From the **Complementary experiment** results, even if the face-shaped screen is downsized, the gaze direction can be conveyed. In the experiment, we assumed that the system was mounted on the shoulder, and participants answered while watch-



Fig. 23 Comparison of screens - The same image is projected on 2 kinds of screen. The left is a flat screen, and the right is face-shaped one. The each direction of eye gaze looks different.

ing at their shoulder. Some participants were surprised that they felt like making eye contact with this system. There are previous wearable systems, but they have mechanic expression [18], [21]. If the system with face-shaped screen is downsized, we can take “surrogate” as a wearable system. Then the system conveys the human expression directly. Thus, if it is possible to make eye contacts between the wearer and remote user, the system can be useful for looking at each other to have an intimate conversation with remote user.

6. Conclusion

In this paper, we presented a telepresence system as a surrogate, which has a real face-shaped screen and moves according to the remote participant. We examined the effectiveness of the face-shaped screen and the understandability of the indicated direction with the use of head gestures and eye gaze. The evaluation results indicate that the face-shaped screen has a possibility to use not only for face-to-face communication but also for various interactions because of its wide visible angle. Although there is no measurable difference in the ease of facial expression recognition between flat screen and face-shaped screen, many participants remarked that facial expressions are easy to understand, and the face-shaped screen is more attractive than the flat screen. More importantly, we confirmed that gestures the LiveMask performed was definitely more understandable than images on the 2D display. We could also confirm that a face-shaped screen correctly transmit gaze direction by solving the “Mona-Lisa effect,” a common gaze-recognition problem of face-to-face communication when using a 2D screen.

Moreover, even if a face-shaped screen is downsized at 1/14, the benefits of solving of “Mona-Lisa effect” is still retained. The small telepresence system is easier to bring or put on the desk, and it can be worn on the shoulder of the local participants so that people bring it like a small buddy. We expect that this small system provides a more intimate impression.

In future work, we would like to make it clarify the effect of smallness, for instance, whether a small system provides an impression such as familiarity and an attachment. Studies on how to convey a remote user’s presence are underway.

References

- [1] AnyBots, available from (<https://www.anybots.com>).
- [2] Cisco Telepresence, available from (<http://www.cisco.com/en/US/>)

- products/ps7060/index.html).
- [3] HP Halo, available from (<http://hphalo.org/>).
- [4] Hydra, available from (<http://www.billbuxton.com/hydraNarrative.htm>).
- [5] Seeing Machine faceAPI, available from (<http://www.seeingmachines.com/product/faceapi/>).
- [6] Talking Head Projection, available from (<http://www.naimark.net/projects/head.html>).
- [7] Adalgeirsson, S.O. and Breazeal, C.: Mebot: A robotic platform for socially embodied presence, *Proc. 5th ACM/IEEE International Conference on Human-robot Interaction, HRI '10*, pp.15–22, ACM, New York, NY, USA (2010).
- [8] Kashiwabara, T., Osawa, H., Shinozawa, K. and Imai, M.: TEROOS: A wearable avatar to enhance joint activities, *Proc. 2012 ACM Annual Conference on Human Factors in Computing Systems, CHI '12*, pp.2001–2004, ACM, Austin, Texas, USA (2012).
- [9] Argyle, M., Lefebvre, L. and Cook, M.: The meaning of five patterns of gaze, *European Journal of Social Psychology*, Vol.4, No.2, pp.125–136 (1974).
- [10] Beskow, J. and Moubayed, S.A.: Perception of gaze direction in 2D and 3D facial projections, *Proc. ACM/SSPNET 2nd International Symposium on Facial Analysis and Animation, FAA '10*, pp.24–24, ACM, New York, NY, USA (2010).
- [11] Delaunay, F., Greeff, J.D. and Belpaeme, T.: A study of a retro-projected robotic face and its effectiveness for gaze reading by humans, *2010 5th ACM/IEEE International Conference on Human Robot Interaction HRI*, pp.39–44 (2010).
- [12] Moubayed, S.A., Edlund, J. and Jonas, B.: The mona lisa gaze effect as an objective metric for perceived cospatiality, *Proc. International Conference on Intelligent Virtual Agents, IVA '11*, pp.439–440 (2011).
- [13] Kendon, A.: Some functions of gaze-direction in social interaction, *Acta Psychologica*, Vol.26, No.1, pp.22–63 (1967).
- [14] Kidd, C.D. and Breazeal, C.: Effect of a robot on user perceptions, In *Intelligent Robots and Systems, 2004 (IROS 2004), Proc. 2004 IEEE/RSJ International Conference*, Vol.4, pp.3559–3564 (2004).
- [15] Lee, M.K. and Takayama, L.: “Now, I have a body”: Uses and social norms for mobile remote presence in the workplace, *Proc. 2011 Annual Conference on Human Factors in Computing Systems, CHI '11*, pp.33–42, ACM (2011).
- [16] Nagahara, H., Yagi, Y. and Yachida, M.: Wide field of view head mounted display for tele-presence with an omnidirectional image sensor, *Computer Vision and Pattern Recognition Workshop*, Vol.7, p.86 (2003).
- [17] Nakanishi, H., Kato, K. and Ishiguro, H.: Zoom cameras and movable displays enhance social telepresence, *Proc. 2011 Annual Conference on Human Factors in Computing Systems, CHI '11*, pp.63–72, ACM, New York, NY, USA (2011).
- [18] Osumi, T., Fujimoto, K., Kuwayama, Y. and Noda, M.: Blogrobot: Mobile terminal for blog browse, *Progress in Robotics*, pp.96–101 (2009).
- [19] Paulos, E. and Canny, J.: Social tele-embodiment: Understanding presence, *Auton. Robots*, Vol.11, pp.87–95 (July 2001).
- [20] Sakamoto, D., Kanda, T., Ono, T., Ishiguro, H. and Hagita, N.: Android as a telecommunication medium with a human-like presence, *Proc. ACM/IEEE International Conference on Human-robot Interaction, HRI '07*, pp.193–200, ACM, New York, NY, USA (2007).
- [21] Tsumaki, Y., Fujita, Y., Kasai, A., Sato, C., Nenchev, D.N. and Uchiyama, M.: Telecommunicator: A novel robot system for human communications, *Proc. 11th IEEE International Workshop on Robot and Human Interactive Communication*, pp.35–40 (2002).

Editor’s Recommendation

At Interaction 2012, by the program committee of 87 members, the outstanding 18 papers among 43 submitted papers were adopted as general lecture presentation and 19 papers among 149 submitted papers were selected as finalists for interactive presentation. This paper, chosen from those 37 papers, gained a good assessment as a recommendable paper to journal by the program committee vote; therefore, as a journal editing chairperson, I certainly would like to recommend this paper.

(Interaction 2012 program chairperson, Homei Miyashita)



Kana Misawa received her B.A.Sc. from Tsuda College and M.S. degree in information technology from The University of Tokyo in 2012. She has been engaged in Dentsu inc. since 2012. Her research interest is Augmented Human. She is a member of ACM.



Yoshio Ishiguro received his Associate degree from Toyota National College of Technology and B.E. / M.E. degree, in information technology from Ritsumeikan University and Ph.D. (Information Studies) from The University of Tokyo in 2012. His research interests include mixed reality, augmented reality, and human computer interaction. He is a member of the Virtual Reality Society of Japan and the Association for Computing Machinery. He received the Good Design Award in 2012.



Jun Rekimoto received his B.A.Sc., M.Sc., and Ph.D. in information science from Tokyo Institute of Technology in 1984, 1986, and 1996, respectively. Since 1994 he has worked for Sony Computer Science Laboratories (Sony CSL). In 1999 he formed and directed the Interaction Laboratory within Sony CSL. Since 2007 he has been a professor in the Interfaculty Initiative in Information Studies at The University of Tokyo. Since 2011 he also has been Deputy Director of Sony CSL. He received the Multi-Media Grand Prix Technology Award from the Multi-Media Contents Association Japan in 1998, iF Interaction Design Award in 2000, the Japan Inter-Design Award in 2003, and iF Communication Design Award in 2005. In 2007, he also elected to ACM SIGCHI Academy.