Ma petite chérie : What are you looking at? A Small Telepresence System to Support Remote Collaborative Work for Intimate 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 present a telepresence system with a reduced scale face-shaped display for supporting intimate telecommunication. In our previous work, we have developed a real-size face shaped display that tracks and reproduces the remote user's head motion and face image. It can convey user's nonverbal information such as facial expression and gaze awareness. In this paper, 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, and also provide more intimate impression. 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. 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.

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

General Terms
Human Factors

Keywords
Telepresence, CSCW, eye gaze, wearable

1. INTRODUCTION
Several telepresence systems have been proposed and developed to facilitate communication. TV conferencing systems and video chat systems are widely used; however, these systems are often installed at a dedicated location, and thus, they cannot be used in transit. Another drawback is that when a person's face is displayed on a normal 2D display, mutual gaze-awareness is not transmitted correctly. This is known as the "Mona Lisa" effect, if a remote participant looks at the camera, all the local participants feel that the remote participant is looking at them[9, 10].

To overcome these problems, telepresense surrogate systems have been studied and developed. The Talking Head system is a classic example. It has a face-molded screen, and images of the remote participant's face are projected on it[3]. A study has been conducted to compare eye gaze when CG images are projected on screen of different shapes[7]. We have also developed a LiveMask system that tracks head motion and automatically extracts facial images for accurate projection[12].

While these systems are stationary, a series of telepresense systems is making them autonomous and movable. For example, Anybot QB is a movable TV conference robot that can be manipulated by remote participants[1].

We also envisage such surrogate systems becoming smaller and wearable so that people can virtually "bring" their remote participants in a real world environment.

For instance, assume a situation in which a husband is asked by his wife to go to the store to purchase something. As the husband rarely goes there, he spends a long time to find out where the requested goods are displayed, and may be confused by the wide variety of options. At such a time, he may call his wife and ask her where he should go and what he should buy. Without sharing his context, it is difficult for him to explain his situation and the variety of features of products briefly to his wife. However, if his wife can share his point of view, she will be able to quickly understand his situation, give him instructions as needed, and communicate interactively. Thus, a wearable surrogate robot has a capability that enables communication with close partners at one's side, and experience intimate communication which feels like whispering close to the ear.

Figure 1: Ma petite chérie

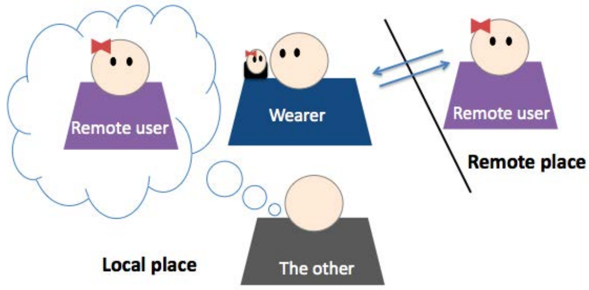

Figure 2: Communication among the remote speaker, wearer and the other

Tsumaki et al. developed a shoulder-mounted robot[15]. It has a display and its shape mimics the human form. Remote participants can represent their intention by controlling the motion of the robot. However, it is bulky owing to its complicated mechanics. Hence, it is not ergonomic. The implementation of bodily gestures increases the complexity and dimensions of the system. Hence, we need to achieve a trade-off between complexity and utility.

In addition, robots with mechanical expressions have some drawbacks. They are not suitable for intimate communication because their expressions are unnatural and different from those of humans. From an outsider's perspective, it is impossible to know who operates the system. The wearable system can be used for communication with not only the wearer and the remote user, but also among the wearer, the remote user, and others in daily situations. (e.g., Shopping, the wearer: husband, the remote user: wife, the third user: the employee whom the wife asks her husband to get help from) In communication between the wearer and the remote user, they can authenticate each other and converse freely. However, suppose the conversation involves the wearer, the remote user and a third person, as in Figure 2. The third user cannot determine the remote user's appearance by seeing the robot, and the mechanical expressions do not convey the remote user's natural expressions.

In this paper, we propose a telepresence system with a scale-down face-shaped screen to convey the remote speaker's presence. This system tracks the remote user's head movement and extracts the facial image, which is projected in real time on a small face-shaped screen that is molded on the basis of the 3D-shape data scanned from the user. As with our previous real-size face display system, we believe this system can enable natural conversation between local and remote participants, and support nonverbal communication such as correct gaze awareness.

2. RELATED WORK

The systems of CSCW (computer-supported collaborative work) are basically used for performing advanced functions such as when an employer receives a report from experts who work in remote locations. WACL displays a remote user's highlights using a laser pointer, which is mounted with a camera and servo motors. The system enables remote veterans to give instructions[14]. Few such systems exist in daily life. Mann identified the same problem and proposed a telepointer[11] with a camera and laser pointer. A telepointer is a system worn around the neck and used for ordinary situations.

In order to support collaborative work, portable robots have also been developed. BlogRobot searches for blogs that aim to provide information linked with the real world to visitors. The system presents suitable information by referring to information in the real world using eye gaze and pointing [13]. Telecommunicator is a miniature robot with a movable head, camera and simple arm. It is teleoperated by a remote user[15]. However, the mechanical expressions of these robots are not human-like, and they do not convey the human natural expressions when using them as their surrogate. Mebot is a semi-autonomous robotic avatar that is designed to convey non-verbal channels[4]. However, the system might encounter the Mona Lisa effect because It use a flat display to show remote user's face.

In humanoid studies, there are studies to make facial expressions using a human-like screen simply not taking approaches of robots or androids[8]. These systems are slightly smaller than a real face or the same size. However, it should be miniaturized when using the system as a portable and wearable device.

However, in order to communicate, it is not clear how such a reduced-size face screen might change the quality of nonverbal communication. It is said that nonverbal information has an important role in communication[5]. In particular, eye gazes can facilitate the flow of conversation[10], and the ability to maintain appropriate levels of intimacy[6]. Hence, we conduct an experiment that eye gaze can be conveyed even if the size is miniaturized.

3. *MA PETITE CHÉRIE*

We summarize the features of this system:

- It is designed to enable people achieve intimacy with others, be accompanied by a buddy, talk with and just regard him/her at any time.
- The telephone or video chat is considered bland. Robots are impersonal. Android robots are difficult to miniaturize. People can easily carry out and feel the presence of a remote user by looking at a small system.
- It is a small device; hence, people use it as a wearable or portable device. Moreover it is capable of putting on the desk without taking up excess the space

System overview: The system consists of a mini face-shaped screen, micro projector, web camera, speaker, and camera platform. Previously, we developed a real-size face-shaped screen that tracks and reproduces the remote user's head motion and facial image. The novel small system has a screen scaled down to 1/14(Figure 4). The face-shaped screen is made from the mold of an actual human face. It is possible to make this type of screen using a 3D scanner and 3D printer. The face-shaped screen moves along two degree-of-freedom (DOF) by reflecting the user's head gestures. The head motion is operated by the 3DOF head tracking data, which is acquired from faceAPI[2].

There are some merits to reducing the size of the face-shaped screen. First, it enables us to miniaturize the size of system without using a fish-eye lens or a first-surface mirror because of a short projected distance. Second, we make use of the projector's brightness enough as required. We adopted the 1/14 scale because it exploits these merits maximally. The size is sufficiently small for mounting it on one's shoulder.

Software overview: 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.

4. EXPERIMENT: PERCEPTION OF EYE GAZE

Figure 3: Comparison of the size: On the left is 1/14 scale face-shaped screen. On the right is the real size face-shaped screen.

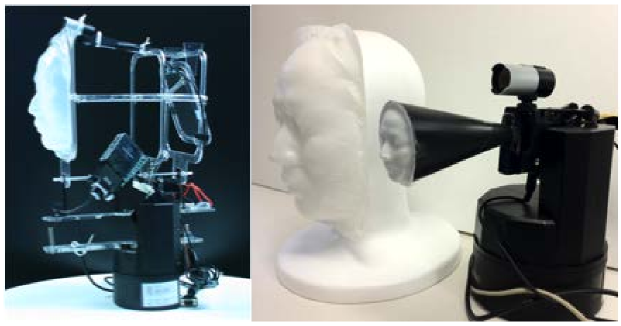

Figure 4: On the left: Telepresence system with a real-size face shaped screen. On the right: Comparison with a real-size face shaped screen and down-sized the face-shaped screen at 1/14.

When people look at the system on their shoulder, they will see it from a diagonal viewpoint. In our previous research, 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(Figure 7). Then, we found that the face-shaped screen improved this problem.

In this experiment, we examine the system can convey the direction of eye gaze even if the face-shaped screen is reduced at 1/14. In our previous research, we tested that under study setting that the participant face on the system. In this experiment, suppose the scene the participants wear

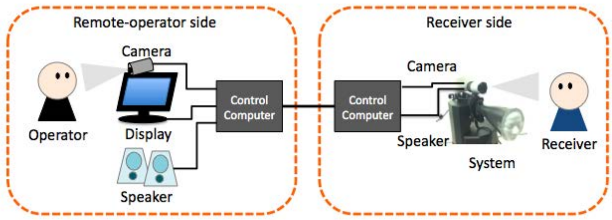

Figure 5: System overview

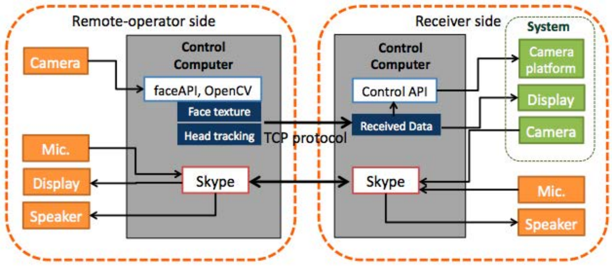

Figure 6: Software overview: The head motion and the face images are extracted by faceAPI and process these data to send to receiver side.

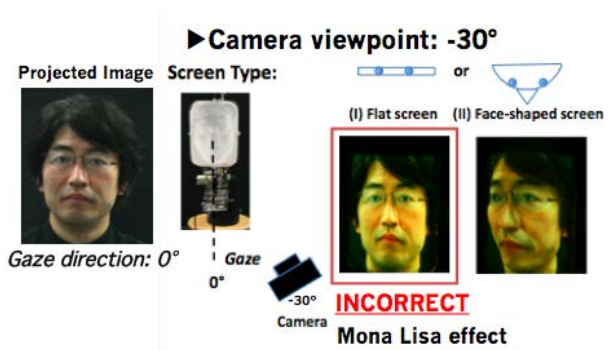

Figure 7: Mona Lisa effect on a flat screen: When the image(gaze direction $0°$) is projected on a flat screen and we see it at -$30°$ viewpoint, we make an eye contact with the subject despite the subject should look at $0°$ angle. In this case, Mona Lisa effect occurs. While, we don't make an eye contact with the image projected on a face-shaped screen.

the system around their shoulders and read the eye gaze by looking into face-shaped screen.

4.1 Purpose

The objective is to recognize which direction the subject looks in by reading his eye gaze using a small face-shaped screen from the wearer's viewpoint.

4.2 Study Setting

Photos are prepared in advance(Figure 8). 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(Figure 9). The number of participants is 5.

4.3 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.

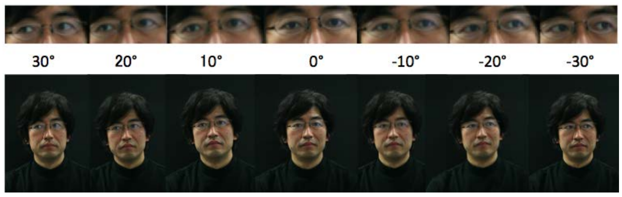

Figure 8: Experiment : Facial position is fixed and eye gaze changed only by using his eyeballs. From left, the direction of gaze is changed $30°$,$20°$... in increments of $10°$

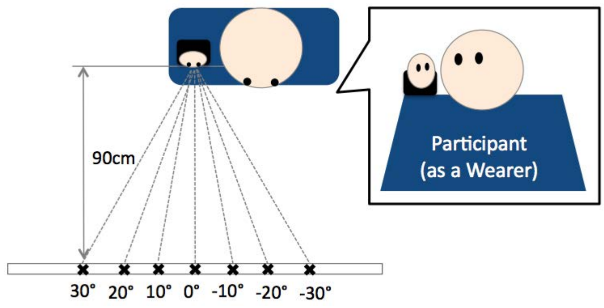

Figure 9: Study Setting: The system on the shoulder displays the Figure 8 images randomly. The participants answer one point which the subject looks at among 7 points.

4.4 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 the Figure 10, which is a record of one participant. 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

From the experimental 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 watching at their shoulder. Some participants were surprised that they felt like making eye contact with this system. Thus, if

Table 1: The result

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

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

Position	1	2	3	4	5	6	7
Answer 1set	△2	△3	△4	④	△6	△7	⑦
Answer 2set	①	②	③	④	△6	△7	⑦
Answer 3set	①	△3	③	④	⑤	△7	△6

Figure 10: One of participant's record: Circle means the answer is correct, and the triangle means the answer is adjacent. In this case, acurracy is 52.3%, and the approximate acurracy is 100%.

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. The tiny system can also be used on a desktop conveying eye gazes.

6. CONCLUSIONS

We proposed a telepresence system with a reduced scale face-shaped display for supporting intimate telecommunication. In this paper, we examined the value and effect of scale reduction of such face-shaped displays. We expected small-size face displays to retain the benefits of real-size talking-head type telecommunication systems, and also provided a more intimate impression. We conducted an experiment using a 1/14 scaled face display, and found critical nonverbal information, such as gaze direction, was still correctly transmitted even when the face size was reduced.

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.

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.

7. REFERENCES

[1] AnyBots. https://www.anybots.com.

[2] Seeing Machine faceAPI. http://www.seeingmachines.com/product/faceapi/.

[3] Talking Head Projection. http://www.naimark.net/projects/head.html.

[4] S. O. Adalgeirsson and C. Breazeal. Mebot: a robotic platform for socially embodied presence. In *Proceedings of the 5th ACM/IEEE international conference on Human-robot interaction*, HRI '10, pages 15–22, New York, NY, USA, 2010. ACM.

[5] M. M. Aguinis, Herman; Simonsen and C. A. Pierce. Effects of nonverbal behavior on perceptions of power bases. In *Journal of Social Psychology 138, no.4 (August)*, pages 455–469, 1998.

[6] M. Argyle, L. Lefebvre, and M. Cook. The meaning of five patterns of gaze. *European Journal of Social Psychology*, 4(2):125–136, 1974.

[7] F. Delaunay, J. De Greeff, and T. Belpaeme. A study of a retro-projected robotic face and its effectiveness for gaze reading by humans. *2010 5th ACMIEEE International Conference on HumanRobot Interaction HRI*, pages 39–44, 2010.

[8] F. Delaunay, J. de Greeff, and T. Belpaeme. Lighthead robotic face. In *Proceedings of the 6th international conference on Human-robot interaction*, HRI '11, pages 101–102, New York, NY, USA, 2011. ACM.

[9] S. A. M. Jens Edlund and B. Jonas. The mona lisa gaze effect as an objective metric for perceived cospatiality. In *Proceedings of the International Conference on Intelligent Virtual Agents.*, IVA'11, pages 439–440, 2011.

[10] A. Kendon. Some functions of gaze-direction in social interaction. *Acta Psychologica*, 26(1):22–63, 1967.

[11] S. Mann. Telepointer: Hands-free completely self contained wearable visual augmented reality without headwear and without any infrastructural reliance. In *Fourth International Symposium on Wearable Computers*, 2000.

[12] K. Misawa, Y. Ishiguro, and J. Rekimoto. Livemask: A telepresence surrogate system with a face-shaped screen for supporting nonverbal communication. In *Proceedings of Interaction.*, 2012.

[13] T. Osumi, K. Fujimoto, Y. Kuwayama, and M. Noda. Blogrobot : Mobile terminal for blog browse. *Progress in Robotics*, pages 96–101, 2009.

[14] N. Sakata, T. Kurata, T. Kato, M. Kourogi, and H. Kuzuoka. Wacl: Supporting telecommunications using wearable active camera with laser pointer. In *ISWC 2003*, pages 53–56, 2003.

[15] Y. Tsumaki, Y. Fujita, A. Kasai, C. Sato, D. Nenchev, and M. Uchiyama. Telecommunicator: a novel robot system for human communications. In *Robot and Human Interactive Communication, 2002. Proceedings. 11th IEEE International Workshop on*, pages 35 – 40, 2002.