ThirdEye: Simple Add-on Display to Represent Remote Participant's Gaze Direction in Video Communication

Mai Otsuki Taiki Kawano Keita Maruyama Hideaki Kuzuoka
University of Tsukuba
1-1-1 Tennodai, Tsukuba, Ibaraki, Japan
{otsuki@emp|kuzuoka@iit}.tsukuba.ac.jp

Yusuke Suzuki
OKI Electric Industry Co., Ltd.
Warabi, Saitama, Japan
suzuki543@oki.com

ABSTRACT

A long-standing challenge in video-mediated communication systems is to represent a remote participant's gaze direction in local environments correctly. To address this issue, we developed ThirdEye, an add-on eye-display for a video communication system. This display is made from an artificial ulexite (TV rock) that is cut into a hemispherical shape, enabling light from the bottom surface to be projected onto the hemisphere surface. By drawing an appropriate ellipse on an LCD and placing ThirdEye over it, this system simulates an eyeball. Our experiment proved that an observer could perceive a remote Looker's gaze direction more precisely when the gaze was presented using ThirdEye compared to the case in which the gaze was presented using the Looker's face on a flat display.

Author Keywords

Telecommunication; gaze awareness;

ACM Classification Keywords

H.4.3 Communications Applications: Computer conferencing, teleconferencing, and videoconferencing.
H.5.3. Information interfaces and presentation (e.g., HCI): Group and Organization Interfaces

INTRODUCTION

Gaze awareness, i.e., "the ability to monitor the direction of a partner's gaze and thus his/her focus of attention" [12], is an important factor in human communication [8, 11, 15]. For example, eye contact plays a critical role during turn taking in a multiparty conversation [22], and eye-gaze toward an object is one of the main cues for achieving joint attention [18]. However, in video communication, a remote participant's gaze direction cannot be represented properly on a display especially when his/her face is viewed from oblique angles; this is common when there are multiple participants in a local environment and some of them have to view the display from oblique angles. A typical phenomenon, known as the Mona Lisa effect [4], is one in which the eyes in a portrait appear to follow the observers as they move. This phenomenon has been noted to undermine turn taking and joint attention significantly [9].

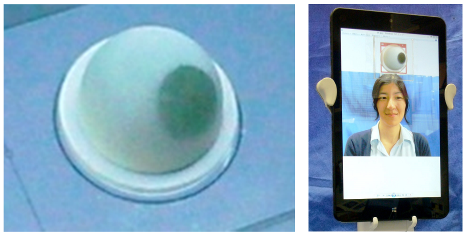


Figure 1. ThirdEye (left) and its use case (right).

To address this problem, various telepresence robots that act as the remote participant's surrogate have been proposed [1, 14]. Typically, these robots have movable flat displays that show the remote participant's face. In spite of the expectation that the orientation of the display will indicate the remote participant's gaze direction, previous studies determined that a rotatable display inevitably results in the Mona Lisa effect and an observer still has difficulty in perceiving the remote participant's gaze direction properly [13, 21].

Delaunay and Misawa proposed systems that use a 3D face-shaped screen with a 3D motion platform [3, 17]. Theoretically, such screens effectively reduce the Mona Lisa effect. However, because each screen needs to match a remote participant's face, general versatility of these systems is limited. A simpler approach is to attach a motor-operated eye [20] to a video conferencing system. However, adding extra actuators makes the system more complex.

Using eye tracking technology, some studies directly overlay gaze representation over a target object; e.g., Higuchi et al. proposed to use a video projector or a head mounted display (HMD) [10]. However, using a projector is not suitable when the target is not placed on a flat projectable surface and a HMD is a burden on a user.

We propose "ThirdEye," which is an eye display that mimics a human eyeball (Figure 1 left) and serves as an add-on to a video communication system (Figure 1 right). We propose to use ThirdEye to represent a remote participant's gaze direction toward a target in a local environment such as for local participants or other physical objects. We chose eyeball representation because we think

CHI 2017, May 06-11, 2017, Denver, CO, USA
© 2017 ACM. ISBN 978-1-4503-4655-9/17/05…$15.00
DOI: http://dx.doi.org/10.1145/3025453.3025681

that it is the most intuitive method that can represent a remote participant's gaze direction. We assume that this method may increase the local participants' perception of gaze direction in video communication. Brockmeyer et al. also created an eye display using 3D printed light pipes and used it as interactive characters' eyes [2]. However, they have not tested its accuracy for representing the character's gaze direction. This paper describes the development of ThirdEye and the evaluation of its accuracy for representing a remote participant's gaze direction.

IMPLEMENTATION

ThirdEye was made by cutting commercially available artificial Ulexite into a hemispherical shape. Ulexite can project an image from its bottom surface to an opposite surface. As a result, the image appears to float on the spherical surface. Consequently, if a moving eye is displayed on a flat LCD and the hemisphere is placed on the LCD, the eyeball appears to rotate and look around the surrounding area.

The average diameters of a human eyeball and iris are 24 and 12 mm, respectively [5]; our hemispherical display and its iris were made to match these sizes. The shape of the iris displayed on the LCD is decided such that it becomes a perfect circle when it appears on the surface of the hemisphere. ThirdEye requires no mechanical movements; it requires only the hemispherical display and software modules. Therefore, the system is quite responsive when presenting a remote participant's saccadic eye movement. Furthermore, it consumes much less electricity compared to a motor-operated eye robot. Because of these features, ThirdEye can be used with a small mobile terminal.

Figure 2 shows a typical system configuration. The in-camera of the mobile terminal at a local site captures an image of the local environment; this image is displayed at the remote site. As a remote participant looks at an object in the image, an eye tracker measures his/her gaze position and sends this data to the local terminal. The terminal then calculates an appropriate gaze direction for ThirdEye and displays the iris at the appropriate position on the mobile terminal's LCD. If the ThirdEye is placed closed to the in-camera, its gaze direction can be calculated by considering the field-of-view of the in-camera.

EXPERIMENT

We conducted an experiment to investigate how a local participant perceives a remote participant's (Looker's) gaze direction in the vertical plane (task 1) and horizontal plane (task 2). In both tasks, an experimenter played the Looker's role.

Experimental design

For both tasks, we tested three types of gaze presentations and two observation directions (front and oblique). The three types of gaze presentations are:

(a) *Face-to-face* (F2F): The actual Looker looked at the designated targets (Figure 3 (a)).

(b) *ThirdEye* (TE): ThirdEye, focusing on the designated targets, was presented to the participant by attaching a hemispherical display to a 17'' flat display (Figure 3 (b)). The Looker's face was not presented on the display.

(c) *Flat display* (FLT): A 17-inch flat display showed prerecorded still images of the Looker focusing on the designated targets in the same environment as that in F2F. The face image had the same size as the actual face (Figure 3 (c)). This condition simulates traditional video-mediated communication systems.

Our hypothesis is that compared to flat display conditions, ThirdEye is effective in alleviating the Mona Lisa effect, i.e., participants estimate the target position more precisely in TE than in FLT condition, especially in oblique conditions.

Experimental setup

As shown in Figure 4, a participant sat facing a Looker (i.e., an actual human, ThirdEye, or a face on the flat display) at a distance of 80 cm. For the front condition, the participant sat straight in front of the Looker. For the oblique condition, the participant sat at an offset of 20° in the clockwise direction from the front condition. The height of the participant's eyes was adjusted to match the height of the Looker's eyes. We asked the participant not to move his/her head too much. However, we did not use a chin-rest

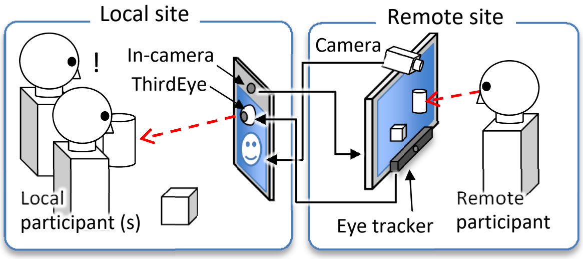
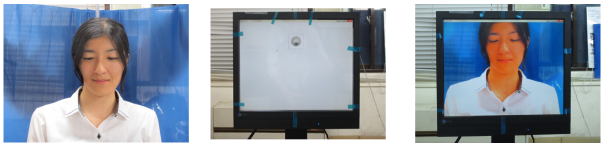


Figure 2. Typical system configuration using ThirdEye



(a) Face-to-face (b) ThirdEye (c) Flat display

Figure 3. Gaze presentation conditions

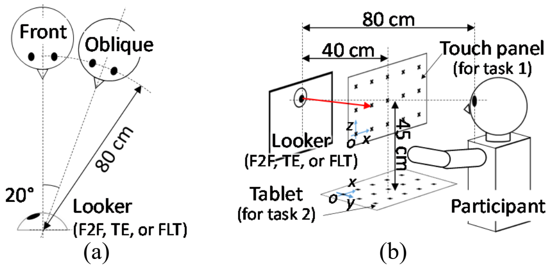


Figure 4. Experimental setup. (a) Observation direction conditions; (b) experimental apparatus.

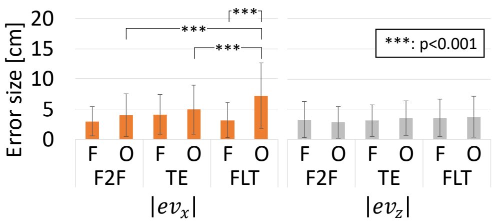


Figure 5. Average of $|ev_x|$ and $|ev_z|$ in task 1.

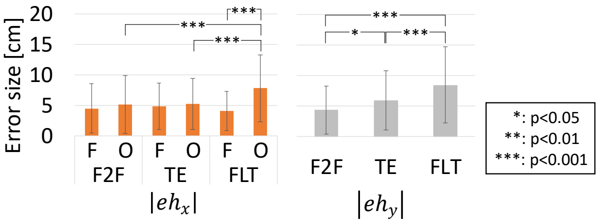


Figure 6.Average $|eh_x|$ and $|eh_y|$ in task 2.

because we thought that it was unnatural to fix the position of the participant's head [11]. In the F2F and FLT conditions also, the Looker did not fix the position of her face but tried not to move her head to a great extent.

Figure 4 (b) shows the apparatus used for the two tasks. In task 1, we vertically placed a transparent touch panel (Awesome Electronic, ATP-2150) in the middle of the display and the participant. The height of the touch panel was configured such that its center matched the height of the Looker's eyes. In task 2, we horizontally placed a graphic tablet (WACOM Intuos 4 Extra large) on the desk (the surface of the tablet was 45 cm below the Looker's eyes). The center of the panel was in the middle of the display and the participant. The two tasks were conducted separately, and the touch panel and tablet were not placed simultaneously.

Procedure

For each condition, we asked the participant to observe the stimuli (i.e., Looker gazing at a target), estimate the position of the gaze target either on the vertical touch panel (task 1) or on the horizontal tablet (task 2), and indicate the estimation by touching the position with a stylus pen (in both tasks). The touched positions were automatically logged in the system. We did not set a time limit; nonetheless, the participant was asked to respond reasonably fast.

In the FLT and TE conditions, the display image was blank for 3 seconds to eliminate any effect of the previous stimulus. In the F2F condition, the participant was asked to close his/her eyes between the stimuli to avoid observing the Looker's eye movement while the target was changed.

The targets were aligned in a grid at intervals of 10 cm, five grid points in the x-direction and three grid points in the y- (for task 2) or z-directions (for task 1), for a total of 15 points. For each gaze presentation condition, all 15 targets were presented twice in a random order (30 targets for each condition). Using optical filters, it was ensured that these points were not visible to the participant, and we confirmed that all participants could not guess that the target points were regularly arranged in a post experiment interview.

For each task, the participants undertook six sessions (three gaze presentation conditions × two observation directions). To become familiar with the system, each participant had a practice session before each session. To eliminate the order effect, the orders of the three gaze presentation conditions, two observation direction conditions, and two tasks were counter balanced.

A total of 12 university students (12 males of age 22–25 years) with normal or corrected-to-normal visual acuity participated in both tasks.

RESULTS

We separated the error vectors (vectors from the Looker's target position to the participant's estimated position) into (ev_x, ev_z) for task 1, and (eh_x, eh_y) for task 2, because we expected that the Mona Lisa effect affects mainly on x-component. Here, ev_x and ev_z denote the x-component and z-component of the error vectors, respectively, in the vertical plane. Similarly, eh_x and eh_y denote the x-component and y-component of the error vectors, respectively, in the horizontal plane. We analyzed the size of ev_x, ev_z, eh_x, and eh_y components (i.e., $|ev_x|$, $|ev_z|$, $|eh_x|$, and $|eh_y|$) of the error vectors by two-way factorial repeated-measures ANOVA (three gaze presentations and two observation directions). The error bars in the following graphs show standard deviations.

In task 1 (vertical plane), Figure 5 shows the average of $|ev_x|$ and $|ev_z|$. For $|ev_z|$, we did not observe any significant difference for the conditions. For $|ev_x|$, there was a significant main effect for both gaze presentation (F(2, 22) = 9.4, $p < 0.01$) and observation direction (F(1, 11) = 21.3, $p < 0.001$) and their interaction was significant (F(2, 22) = 9.2, $p < 0.01$). A simple main effect test between two observation direction conditions at each gaze presentation condition revealed that in the FLT condition, $|ev_x|$ was significantly larger for the oblique condition than for the front condition ($p < 0.001$). For the other conditions, a simple main effect test showed no significant differences. A simple main effect test between the three conditions for each observation direction condition revealed that for $|ev_x|$, there was significant difference for the oblique condition ($p < 0.001$). Tukey's post-hoc test revealed that $|ev_x|$ for the FLT condition was significantly larger than that for TE and F2F conditions ($p < 0.001$); however, no significant difference was observed between TE and F2F conditions.

For task 2 (horizontal plane), Figure 6 shows the average of $|eh_x|$ and $|eh_z|$. For the average of $|eh_x|$, there was a significant main effect for only the observation direction (F(1, 11) = 10.5, $p < 0.01$). Furthermore, the interaction of gaze presentation and observation direction was significant

(F(2, 22) = 5.7, p < 0.05). A simple main effect test between two observation direction conditions at each gaze presentation condition revealed that for the FLT condition, $|eh_x|$ was significantly larger for the oblique condition than that for the front condition (p < 0.001). Furthermore, a simple main effect test between the three conditions for each observation direction condition revealed that for $|eh_x|$ in the oblique condition, there was significant difference between the gaze presentation conditions (p < 0.01). In the oblique condition, Tukey's post-hoc test revealed that $|eh_x|$ for the FLT conditions was significantly larger than that for the TE and F2F conditions (p < 0.001); however, there was no significant difference between the TE and F2F conditions. $|eh_y|$ showed a significant main effect for only gaze presentation (F(2, 22) = 40.0, p < 0.001), but there was no significant interaction between the gaze presentation and observation direction. A post-hoc test (Bonferroni correction) for the three gaze presentation conditions revealed that there were significant differences between them (FLT vs TE: p < 0.001; FLT vs F2F: p < 0.001; TE vs F2F: p < 0.05).

DISCUSSION

The result shows that the error was greatest when the flat display was observed from an oblique angle; however, ThirdEye could reduce the error size. To understand how the estimations were shifted from the target, we calculated the averages of ev_x and eh_x for the oblique condition. Results showed that ev_x was -7.0 cm for FLT, 1.8 cm for TE, and -2.6 cm for F2F conditions; eh_x was -7.6 cm for FLT, 2.1 cm for TE, and -3.1 cm for F2F conditions. These results clearly indicate that the FLT was inevitably affected by the Mona Lisa effect, but TE was not.

To better understand the characteristics of the errors for the horizontal plane, we drew a bubble graph that depicts the average end points of the error vectors of each target for the front condition (Figure 7). The width and height of each ellipse indicates standard deviations of error vectors. Interestingly, as shown by the blue circles in the graph, the Looker's gazes in the FLT condition were estimated to be much closer to the Looker compared to the TE and F2F conditions. We assume that the three-dimensional cues that were lost in the FLT condition significantly degraded the estimation accuracy. In fact, three participants mentioned that ThirdEye was better for estimating the target position than the flat display because ThirdEye is three-dimensional.

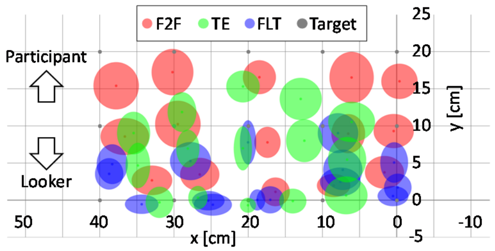


Figure 7. End point of error vectors for front condition for horizontal plane (width and height of each ellipse indicates standard deviation of error vectors).

In the FLT condition, the Looker's face was captured when he/she was focusing on designated targets in the same environment as that in the F2F condition. However, in actual video communication, the remote Looker's gaze direction in the FLT condition is determined by a combination of the field-of-view and position of the local camera, size of remote Looker's display, distance between the remote Looker and the remote display, and the relative position of the remote camera in relation to the remote display. Thus, the Looker's gaze direction as estimated from the Looker's face in the local display is normally inconsistent with the Looker's actual gaze target. In this experiment, we assumed the best case for flat display. Therefore, we can expect that ThirdEye will afford more advantages than indicated by our experiment.

Our study has some limitations. As shown in Figure 6, $|eh_y|$ for the TE condition was significantly larger than that for the F2F condition. The errors of ThirdEye (Figure 7, green circles) indicate that the Looker's gaze tended to be estimated closer to the Looker as the targets got further from the Looker. In the post-experiment interview, four participants mentioned that, in the F2F condition, they used the directions of the Looker's face or a nose as cues for the estimation. The absence of such cues may be one of the reasons why ThirdEye was still inferior to face-to-face presentation. However, similar to Mayers' method [16], we expect that this limitation can be alleviated by calibrating ThirdEye's pupil position so as to orient it a little more in the outward direction than the calculated position.

Another limitation is that we have still not tested the effect of ThirdEye under the expected typical system configuration (Figure 2). When ThirdEye and the Looker's face coexist in a display, it is possible that an observer tends to pay more attention to the Looker's face than the ThirdEye. Further study is required to determine the appropriate face size in a display for ThirdEye to be most effective.

CONCLUSION

To alleviate the problem of inaccurate gaze awareness in a video mediated communication system, we developed ThirdEye, an add-on eye display that represents a remote participant's gaze direction. Our experiment proved that ThirdEye eliminates the Mona Lisa effect and improves the gaze estimation accuracy compared to the case in which the remote participant's face is shown on a flat display.

We expect that ThirdEye will be most effective when the participants need to refer to objects in a remote environment during remote collaboration [6, 7, 19]. In addition, since ThirdEye is quite responsive to quick eye movements, correctly mediating quick glances of a remote participant may improve video communication. Thus, in future work, we will develop the system shown in Figure 2 and investigate whether it is really effective for such tasks.

REFERENCES

1. Sigurdur O. Adalgeirsson, and Cynthia Breazeal. 2010. MeBot: a robotic platform for socially embodied presence. In *Proceedings of the 5th ACM/IEEE international conference on Human-robot interaction* (HRI '10). IEEE Press, Piscataway, NJ, USA, 15-22. DOI: 10.1145/1734454.1734467
2. Eric Brockmeyer, Ivan Poupyrev, and Scott Hudson. 2013. PAPILLON: designing curved display surfaces with printed optics. In *Proceedings of the 26th annual ACM symposium on User interface software and technology* (UIST '13). ACM, New York, NY, USA, 457-462. DOI: http://dx.doi.org/10.1145/2501988.2502027
3. Frédéric Delaunay, Joachim de Greeff, and Tony Belpaeme. 2010. A study of a retro-projected robotic face and its effectiveness for gaze reading by humans. In *Proceedings of the 5th ACM/IEEE international conference on Human-robot interaction* (HRI '10). IEEE Press, Piscataway, NJ, USA, 39-44. DOI: 10.1145/1734454.1734471
4. Jens Edlund, Samer Al Moubayed, and Jonas Beskow. 2011. The Mona Lisa gaze effect as an objective metric for perceived cospatiality. In *Proceedings of the 10th international conference on Intelligent virtual agents* (IVA'11), Hannes Högni Vilhjálmsson, Stefan Kopp, Stacy Marsella, and Kristinn R. Thórisson (Eds.). Springer-Verlag, Berlin, Heidelberg, 439-440. DOI: 10.1007/978-3-642-23974-8_52
5. John V. Forrester, Andrew D. Dick, Paul G. McMenamin, Fiona Roberts, and Eric Pearlman. 2015. *The eye: Basic sciences in practice* (4th ed.). Saunders Ltd.
6. Susan R. Fussell, Leslie D. Setlock, and Elizabeth M. Parker. 2003. Where do helpers look?: gaze targets during collaborative physical tasks. In *CHI '03 Extended Abstracts on Human Factors in Computing Systems* (CHI EA '03). ACM, New York, NY, USA, 768-769. DOI:http://dx.doi.org/10.1145/765891.765980
7. Steffen Gauglitz, Benjamin Nuernberger, Matthew Turk, and Tobias Höllerer. 2014. World-stabilized annotations and virtual scene navigation for remote collaboration. In *Proceedings of the 27th annual ACM symposium on User interface software and technology* (UIST '14). ACM, New York, NY, USA, 449-459. DOI=http://dx.doi.org/10.1145/2642918.2647372
8. Charles Goodwin. 1994. Professional vision, *American anthropologist*, Vol. 96, No. 3, 606-633. DOI: 10.1525/aa.1994.96.3.02a00100
9. Christian Heath and Paul Luff. 1991. Disembodied conduct: communication through video in a multi-media office environment. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (CHI '91), Scott P. Robertson, Gary M. Olson, and Judith S. Olson (Eds.). ACM, New York, NY, USA, 99-103. DOI: http://dx.doi.org/10.1145/108844.108859
10. Keita Higuchi, Ryo Yonetani, and Yoichi Sato. 2016. Can Eye Help You?: Effects of Visualizing Eye Fixations on Remote Collaboration Scenarios for Physical Tasks. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems* (CHI '16). ACM, New York, NY, USA, 5180-5190. DOI: https://doi.org/10.1145/2858036.2858438:
11. Tomoko Imai, Dairoku Sekiguchi, Masahiko Inami, Naoki Kawakami, and Susumu Tachi. 2006. Measuring gaze direction perception capability of humans to design human centered communication systems. Presence, Vol. 15, No.2 (Apr. 2006), 123-138. doi: 10.1162/pres.2006.15.2.123
12. Hiroshi Ishii, Minoru Kobayashi, and Jonathan Grudin. 1993. Integration of interpersonal space and shared workspace: ClearBoard design and experiments. ACM Trans. Inf. Syst. 11, 4 (October 1993), 349-375. DOI=http://dx.doi.org/10.1145/159764.159762
13. Ikkaku Kawaguchi, Hideaki Kuzuoka, and Yusuke Suzuki. 2015. Study on Gaze Direction Perception of Face Image Displayed on Rotatable Flat Display. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems* (CHI '15). ACM, New York, NY, USA, 1729-1737. DOI: http://dx.doi.org/10.1145/2702123.2702369
14. Hiroaki Kawanobe, Yoshifumi Aosaki, Hideaki Kuzuoka, and Yusuke Suzuki. 2013. iRIS: a remote surrogate for mutual reference. In Proceedings of the 8th ACM/IEEE international conference on Human-robot interaction (HRI '13). IEEE Press, Piscataway, NJ, USA, 403-404. DOI: 10.1109/HRI.2013.6483618
15. Adam Kendon. 1967. Some functions of gaze-direction in social interaction, Acta Psychologica 26, 22-63. DOI: 10.1016/0001-6918(67)90005-4
16. Sven Mayer, Katrin Wolf, Stefan Schneegass, and Niels Henze. 2015. Modeling Distant Pointing for Compensating Systematic Displacements. In Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems (CHI '15). ACM, New York, NY, USA, 4165-4168. DOI: http://dx.doi.org/10.1145/2702123.2702332
17. Kana Misawa, Yoshio Ishiguro, and Jun Rekimoto. 2012. LiveMask: a telepresence surrogate system with a face-shaped screen for supporting nonverbal communication. In *Proceedings of the International Working Conference on Advanced Visual Interfaces* (AVI '12), Genny Tortora, Stefano Levialdi, and Maurizio Tucci (Eds.). ACM, New York, NY, USA, 394-397. DOI=http://dx.doi.org/10.1145/2254556.2254632

18. Moore, C.; Dunham, P (1995). *Joint Attention: Its Origins and Role in Development*. Lawrence Erlbaum Associates.
19. Bonnie A. Nardi, Heinrich Schwarz, Allan Kuchinsky, Robert Leichner, Steve Whittaker, and Robert Sclabassi. 1993. Turning away from talking heads: the use of video-as-data in neurosurgery. In *Proceedings of the INTERACT '93 and CHI '93 Conference on Human Factors in Computing Systems* (CHI '93). ACM, New York, NY, USA, 327-334. DOI=http://dx.doi.org/10.1145/169059.169261
20. Hirotaka Osawa, Ren Ohmura, and Michita Imai. 2009. Using attachable humanoid parts for realizing imaginary intention and body image. International Journal of Social Robotics, Vol. 1, No. 1 (Jan. 2009), 109-123. doi:10.1007/s12369-008-0004-0
21. David Sirkin and Wendy Ju. 2012. Consistency in physical and on-screen action improves perceptions of telepresence robots. In *Proceedings of the seventh annual ACM/IEEE international conference on Human-Robot Interaction* (HRI '12). ACM, New York, NY, USA, 57-64. DOI=http://dx.doi.org/10.1145/2157689.2157699
22. Roel Vertegaal, Robert Slagter, Gerrit van der Veer, and Anton Nijholt. 2001. Eye gaze patterns in conversations: there is more to conversational agents than meets the eyes. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (CHI '01). ACM, New York, NY, USA, 301-308. DOI=http://dx.doi.org/10.1145/365024.365119