Maintaining Presence of Remote Speaker on Telepresence Robots by Visual Morphing to Reduce Loneliness

Hirokazu Yoshida[1] and Fumihide Tanaka[2]

***Abstract*— Telepresence robots can provide the sense that a remote person is physically present. However, this sense disappears suddenly once a call ends. Therefore, the speaker may feel loneliness. This study explores "tele-vestige" in which speakers can perceive the weak social presence of a remote speaker after a call ended. To provide this tele-vestige, the remote speaker's face is morphed into a telepresence robot's face after the call ends. We conducted an experiment to investigate the effect of this morphing technique on loneliness.**

I. Introduction

With advances in telecommunication technology, such as videotelephony, communication between different locations has become easier. However, technology-based remote communication is not considered equivalent to face-to-face communication. The use of telepresence robots has attracted increasing attention as a way to improve remote communication. By displaying a remote user's face or gestures, a telepresence robot can create the sense that the remote speakers are in the same physical space [1][2][3][4].

However, when a call ends, the social presence disappears abruptly. While high-quality communication can be provided easily in telepresence robots, speakers' loneliness become intense when the social presence disappears. Especially, people with high social skills will be affected because there is a relationship between the frequency of using a cellphone and loneliness [5]. McLuhan suggested that a telephone creates an intense feeling of loneliness because it is a participant form that demands a partner [6]. He also suggested that telephone offers a very poor auditory image [6]. Because telepresence robots use other senses such as visual images in addition to auditory images, speakers will feel more intense loneliness than the case with using telephones.

This study proposes a method to reduce loneliness by presenting a "tele-vestige" of a speaker when a call is terminated (Fig.1). In the proposed method, features that identify the remote user are mixed with the features of the robot and presented to the local user. We focus on the face as an element that can identify the remote speaker, and an image of the remote speaker's face is mixed with the robot's face. To verify that tele-vestige is effective in alleviating loneliness, we conducted an experiment. This experiment focused on a change of loneliness after a call ends. To maintain the presence of a remote speaker, we employed morphing as a switching method.

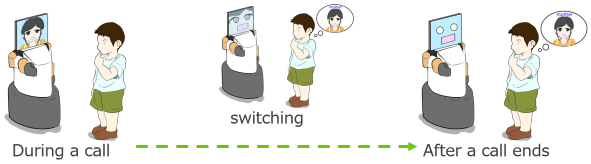

Fig. 1. The concept of tele-vestige

II. Related works

Several studies have addressed problems that occur when a remote speaker disappears temporarily in telepresence systems. Most telepresence systems [7][8][9][10] use a flat display; consequently, the remote speaker can only be seen when the local user is directly in front of the flat display [11][12][13]. This problem can be solved by changing the shape of the display. For example, SphereAvatar [14] displays the remote speaker's face on a spherical display such that they can be observed from any position. However, this study uses a virtual avatar, and did not investigate whether a social presence occurs when displaying a human face. To enhance the presence of a remote speaker, Misawa et al. developed LiveMask [15], a telepresence system that displays the face-shaped screen. The face-shaped screen is formed by a 3D printer based on data obtained by a 3D scanner, and elements that identify the remote speaker can continue to be displayed on the screen after the call has ended. However, that research did not investigate the social presence after the end of the call ends.

Studies have also investigated the social presence based on physicality. For example, ChameleonMask is a telepresence system in which a surrogate wears a helmet comprising an iPad Air and a head-mounted display. This system presents a social presence using a human who has similar physique with a remote person as a surrogate and displaying the remote person's face to the ChameleonMask's iPad Air. It was reported that the surrogate wearing the ChameleonMask felt as though they were the remote user [16]. However, when the surrogate removes the ChameleonMask, the social presence disappears abruptly.

Similar to this study, methods to sustain the social presence of remote users are also being investigated. For example, Tanaka et al. clarified that the gap between the telepresence mode and the autonomous mode becomes a

*Research supported by JSPS KAKENHI 17K19993.

[1]Hirokazu Yoshida is with Graduate School of Systems and Information Engineering, University of Tsukuba, Tsukuba, Japan. yoshida@ftl.iit.tsukuba.ac.jp

[2]Fumihide Tanaka is with Faculty of Engineering, Information and Systems, University of Tsukuba, Tsukuba, Japan. tanaka@iit.tsukuba.ac.jp

reduction in the presence, and they report that this gap is reduced when the autonomous system acts as a remote person [17]. However, for the speaker, it was difficult to identify whether the controller is the robot or the remote speaker.

III. Switching

To present the proposed tele-vestige, switching from telepresence mode to autonomous mode is performed gradually. During the switching process, elements that can identify the remote person and the robot are blended. In this study, the face is considered an element that can identify a remote speaker. We employ morphing as a switching method to mix the two faces. Morphing is a deformation technique used in video technology in which pixel values are changed to alter the shape of an image based on feature points [18]. The morphing procedure is described in the following.

A. Obtaining facial image

In this study, the remote person's face, which will be displayed on the robot's head, is acquired using a Microsoft LifeCam Studio webcam installed at the remote location. The Open Source Computer Vision library is used to obtain the facial image from the camera.

B. Obtaining feature points

Feature points are acquired to determine corresponding points between the remote user's facial image. We employ a face detector in the dlib toolkit for C++ that uses a histogram of oriented gradients HOG and using an ensemble of regression trees to obtain human facial feature points (maximum: 68) [19][20]. Note that the feature points of the robot's facial image are determine manually.

C. Normalization

In a two-dimensional image, the size of the face changes as the distance from the camera changes. Therefore, based on the acquired feature points, the height and width of the remote person's face are normalized to the height and width of the robot's face.

D. Morphing

In the morphing process, the face is divided into triangles using Delaunay division. There triangles are then deformed. In the transformation from coordinate x_1 of a feature point in image 1 to coordinate x_2 of a feature point in image 2, coordinate x_M of the feature point in the morphing image is obtained as follows:

$$x_M = (1-\alpha)x_1 + \alpha x_2 \tag{1}$$

where α is a mixed ratio $(0 \leq \alpha \leq 1)$. Y coordinate is calculated as follows.

$$y_M = (1-\alpha)y_1 + \alpha y_2 \tag{2}$$

Based on the moved feature points, the entire face is deformed by affine transformation of the triangles. Simultaneously, the pixel value mixing ratio is also changed. This method is called an alpha blending [21]. Here, pixel value $I_{M(x,y)}$ of the point (x, y) at the mixing ratio α is obtained as follows.

$$I_{M(x,y)} = (1-\alpha)I_{1(x,y)} + \alpha I_{2(x,y)} \tag{3}$$

The image obtained by these processes is then displayed as the robot's face. Fig.2 shows images of a morphed face.

Fig. 2. Images of morphed face

IV. Experiment

An experiment was conducted to verify that the morphing method is effective at alleviating loneliness. In this experiment, the experimenter acted as the remote person, and the participant acted as the speaker. The participants were seven students (5 males and 2 females, all in their twenties) recruited at University. We wanted to analyze an amount change of loneliness after the call ends. Therefore, we recruited participants who have seen the robot previously. They knew the concept of tele-vestige and the object of this experiment, but not for sure.

A. Robot

For the experiment, we use the Pepper robot from SoftBank Robotics. Pepper's display is mounted on its chest rather than head; thus, in this experiment, two faces were utilized to make the participants feel strange. To achieve this, an exterior display was mounted onto the head of Pepper Fig.3 shows the robot's appearance. The additional face was displayed using a Surface. Note that attaching the exterior display made it impossible to use Pepper's camera. As a result, the experimenter could not observe the state of the laboratory. Therefore, a USB camera (Microsoft LifeCam Studio) was attached to the top of Pepper's head. The facial image was transmitted and received by UDP communication, and voice communication was transmitted and received using Skype. We had to separate the participant from experimenter. Thus, we used a wi-fi router(SoftBank, Pocket WiFi 501HW). To provide the participant with instructions using the robot, an utterance was generated using the "say" module of Choregraphe 2.5.5, and the reproduced utterance was transmitted by Skype. Unfortunately, because of the

network issue and the balance of the robot's body, we could not control the body of the robot.

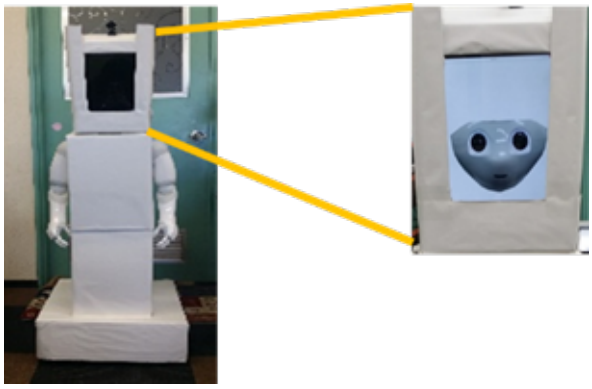

Fig. 3. Robot's appearance

B. Experimental conditions

The experimental conditions were as follows.

A) Cutting condition: After the telepresence, the human face immediately switched to the robot's face.
B) Fading condition: After the telepresence, the human face gradually became thin, and the robot's face appeared gradually (no deformation))
C) Morphing condition: After telepresence, the human face was morphed into the robot's face.

Here, condition A was considered the baseline condition, and Condition B was prepared to confirm that the effect of Condition C does not exclusively come from the alpha blending (formula (3)). Switching images in Conditions B and Condition C required approximately 20 seconds. Note that this experiment followed within-subject design. In consideration of the order effect, the order of the conditions was varied for each participant.

Switching images in Conditions B and Condition C required approximately 20 seconds.However, switching time sometimes fluctuated because of the performance of the computer. Note that this experiment followed within-subject design. In consideration of the order effect, the order of the conditions was varied for each participant.

C. Experimental procedure

The experimental procedure is described as follows.

1. A participant entered the laboratory (Room1) with the experimenter and sat on a chair. The experimenter explained participants to the participants that the participant would talk to remote speakers in this experiment.
2. The participant answered a pre-task questionnaire. After answering this questionnaire, the participant was told to wait for a while, and the experimenter collected the pre-task questionnaire. The flipped post-task questionnaire was placed on the laboratory desk. After that, the experimenter left the laboratory and moved to another room (Room2).
3. Approximately one minute after the experimenter left the laboratory, the robot spoke, "Denwa ga kitami-taidesu" (the equivalent of "You have a phone call") and switched to the telepresence mode. Here, the screen immediately switched to the remote person's face.
4. The participant then played a Yes/No game with the remote experimenter. The rules of the game are described as follows.
   - The number of players was two.
   - One person was a questioner (participant) and the other was the respondent (experimenter). The questioner gives a question, and the respondent thinks the answer.
   - The question options were prepared in advance and included illustrations and names. The questioner determined a single answer among the options.
   - After the questioner determined the answer, the respondent attempted to guess it correctly. An appropriate "Yes" or "No" response was then given. In the case of "No," the questioner provided a hint.
   - The answer did not have to be the same as the illustration; however, questioner was not permitted to give an incorrect answer.
   - The respondents answered after repeating this interaction three times. The respondent won if the answer was correct (and vice versa).

   Note that the experimenter spoke monotonously in their responses.
5. After completing the task, the experimenter informed the participants that they will return to the laboratory and terminated the call. Upon termination of the call, the robot performed one of the switching methods (Condition A, B, or C). Note that only the display of experimenter's facial image and the audio were terminated at this point. The experimenter could continue to receive the camera image and audio from the participant.
6. Approximately 30 seconds after switching was completed, the robot spoke "Anketo youshi ni kinyuu wo onegaishimasu" (the equivalent of "Please fill out the questionnaire"). The participant then answered a post-task questionnaire and informed the robot when they were finished. After confirming that the post-task questionnaire had been completed, the experimenter returned to the laboratory.
7. Procedures from 1 to 6 were also performed for Condition A, B, and C.

Fig.4 shows the experimental environment. For five participants, the camera was positioned to the right of the participant. For the other two, the camera was positioned at the front left of the participant to analysis participants' behavior from different viewpoints. The face of the experimenter displayed on the surface was obtained from the camera on the PC in Room2. As explained above, a USB camera was attached to the top of the robot. Therefore, the experimenter was able to observe participants through this PC.

The Temporary Mood Scale (TMS) [22] was used in

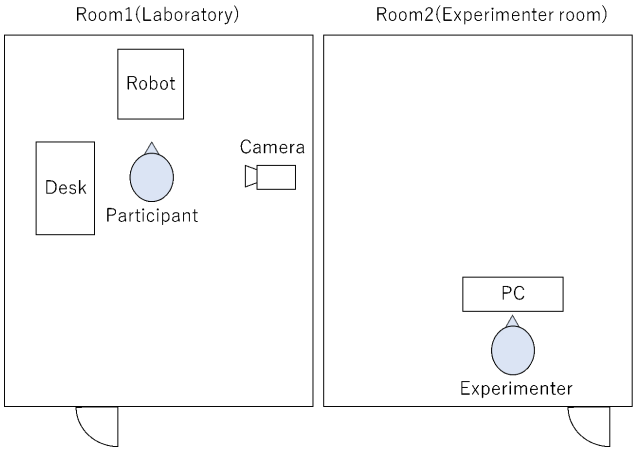

Fig. 4. The setting of the experiment

the questionnaire to investigate changes in the participant's moods under each condition. The TMS evaluates mood changes according to six subscales, i.e. vigor, fatigue, anger, depression, tension, and confusion on a scale of 1 (strongly disagree) to 5 (strongly agree). The pre-task questionnaire included only the TMS, and the post-task questionnaire included the TMS and free descriptions. Here, the total of the three measures represents the total score, and the amount of change between the pre- and post-task questionnaire results was evaluated. If the post-task questionnaire was administered immediately after the call had ended, it is possible that the task and the impression of the image would not be influenced by the different switching methods. Therefore, a wait time of 30 seconds was implemented after the screen had switched.

The hypotheses tested in this experiment were as follows.

- Hypothesis1: An increase in depression in the cutting condition is greater than that in the morphing condition. (cutting condition $>$ morphing condition)
- Hypothesis2: An increase in depression in the fading condition is greater than or equal to the one in the morphing condition. (fading condition $\geq$ morphing condition)

V. Result

The number of samples was insufficient in this study. Therefore, only a numerical comparison was performed (no analysis of variance was conducted). From the results of the TMS questionnaire, the average amount of the change in moods before and after the call is shown in Fig.5. We observed an increase in a amount of vigor for all conditions, and a decrease in a amount of fatigue, anger and depression. However, we observed an increase in tension in the fading condition. Furthermore,we did not observe any change in confusion for the cutting and fading condition. Fig.6 shows the average change in each question item relative to depression. The questions related to them are as follows [22].

- Q10. To feel that one has no hope.
- Q11. To feel sad and lonely.
- Q12. To feel dark and gloomy.

The average change for Q11 and Q12 were similar for the cutting condition and morphing condition.

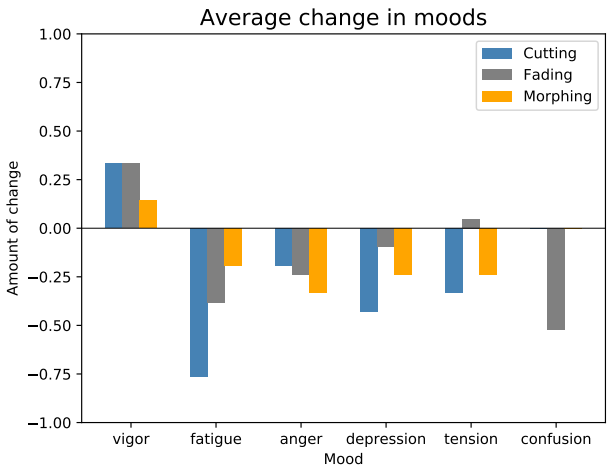

Fig. 5. Average change in moods

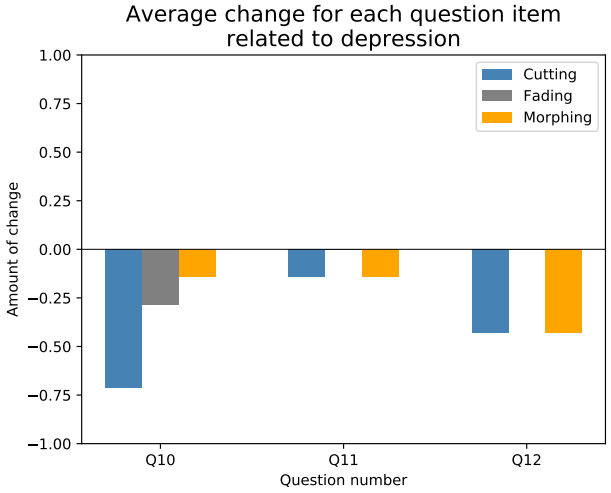

Fig. 6. Average change for each question item related to depression

VI. Discussion

"For depression", the average change for each condition were ordered as follows: cutting condition $<$ morphing condition $<$ fading condition, which indicates that hypotheses 1 was rejected. However, according to the questionnaire results, depression was reduced under all conditions after the call had ended, and all question items related to depression reduced. Furthermore, the average amount of change for cutting and morphing condition in Q11 and Q12 were same. This means that instead of the difference in conditions, there is a possibility that the participants could not feel lonely. Indeed, most participants chose 1 (strongly disagree) or 2 (disagree) for both questionnaires. This might have occurred because it may have been difficult to feel loneliness after

the call had ended. In usual remote communication, after a call ends, the people communicating generally do not meet for a while. However, in this experiment, the participants understood that they would meet the remote experimenter again shortly after the call ended. Therefore, we consider that it was difficult to feel loneliness.

Relative to vigor, fatigue, and confusion, the fading condition showed slightly better results than the morphing condition. Here, eeriness due to morphing can be considered. In a post-experiment interview, some of the participants said that "I did not know the specific factors, but I felt that the morphing condition was eerie compared to the fading condition" and "In the fading condition, it seemed like taking a time to switching a screen because there was a gap between robot's face and remote speaker's face. However, I felt that I wanted to leave this place early because it was eerie under the morphing condition." MacDorman et al. reported that displaying an image of an intermediate face between an android and a humanoid robot increased the level of eeriness [23]. A similar observation was made during the dynamic morphing in our experiment. The main cause of the eeriness in our experiment was considered to be the eyes in the morphed face. The morphed face resembled the human face except the eyes part [24]. This may have caused a creepy impression to participants, which could have influenced vigor, fatigue, and confusion. However, the difference between the conditions was not significant, and it is considered that no significant difference would be observed even if the number of samples were increased.

This may be attributed to the fact that morphing was performed only after the call had ended because we wanted to evaluate the presence of a remote person after a call ends. Therefore, the participants observed the morphing for the first time when the call ended. Despite that there was no change in the past, the robot's appearance suddenly changed after the call ends. As a potential improvement, morphing can be performed at both the begining and the end of the call. Some participants indicated that "It might be better to morph during calls." However, as described above, it is possible that changes in the face due to morphing may make the speaker feel strange. Therefore, as future work, we will consider the gaze direction of the speaker. When the speaker is looking at the robot, the robot's face does not change, and it does morph slightly when the speaker is not looking at the robot. This suggests that morphing during a call may not cause the speaker to feel strange.

We consider the evaluation by questionnaire to be inappropriate. Note that the process of answering a questionnaire may change the mood of a participant. In an experiment [25] that investigated the effect of hugs during a call, an evaluation was performed by measuring the degree of change in stress hormones before and after a call. It is appropriate to conduct an evaluation using such physiological changes. According to IJzerman et al., body temperature decreases due to loneliness [26]. Therefore, in future, we plan to evaluate loneliness by measuring changes in body temperature before and after a call.

There were problems to be fixed in this experiment. However, morphing condition showed some possibility to present a tele-vestige. The comment that "in fading condition, it seemed like taking a time to switching a screen because there was a gap between robot's face and remote speaker's face" means changing just pixel value is not enough effect to maintain a social presence. On the other hand, we observed that five out of seven participants had a longer gaze time than the other two conditions at the morphing condition. Among them, two participant did not move their gaze direction after the call had ended under this condition[1] . This suggests that the morphing condition caught the attention of participants. In addition, a decrease in the amount of anger under the morphing condition was the largest of the three conditions. This means that stress due to switching using morphing was smaller than other two switching condition. In summary, these results suggest that the morphing condition was the most natural switch of the three conditions. As a result, despite the speaker can identify the controller, the social presence may not be reduced because the gap between autonomous and telepresence mode become gradually small. Therefore, the morphing condition has a possibility to maintain a social presence.

VII. Future work

Based on these results, future experiments with improved procedures are planned. The most problematic part of this experiment was that participants could not feel loneliness. Hence, it will be important to create a situation in which the participants cannot do physically encounter the remote speaker immediately after the call has ended.

Therefore, we will consider separating the experimenter and remote person roles. Furthermore, if the number of calls is too high, the participant's loneliness may gradually weaken along with the number of calls. Therefore, we intend to compare only two conditions. In this experiment, we will focus on the morphing condition. One condition will employ morphing only at the beginning and the end of a call, and the other condition will employ morphing at the beginning and the end, as well as during the call.

Moreover, changes in body temperature will be investigated. It is possible that it is difficult for a participant to feel loneliness in the 30 second wait time employed in this study. Therefore, in reference to an experiment performed by IJzerman et al., this wait time should be greater than 300 seconds [26].

VIII. Conclusion

To reduce loneliness caused by the rapid disappearance of the presence of a remote speaker shown by a telepresence robot, we have proposed the switching method to maintain the presence of the remote speaker. In this study, an experiment was conducted in which the face of a remote person displayed on the face of a robot was morphed into the face of the robot after a call ended. In this experiment, the cutting

[1]The camera for a video observation was positioned on the right of these two participants

and fading conditions were also evaluated. Questionnaires were administered before and after the call to investigate changes in mood under each condition. However, no significant difference between conditions was observed. There were two reasons for this. First, the experiment was unsuitable to evaluate change in loneliness. Second, it was difficult to present a sustained social presence of a remote person by simply morphing the images. In addition, the robot made the speaker feel a social eeriness.

However, the gaze time and the decrease in the amount of anger suggested that the morphing condition was the most natural switch of the three conditions. Therefore, the morphing condition has a possibility to maintain the social presence of the remote speaker.

ACKNOWLEDGEMENT

The work was supported by JSPS KAKENHI Grant Number 17K19993.

REFERENCES

[1] S. Tachi, N. Kawakami, H. Nii, K. Watanabe, and K. Minamizawa, "Telesarphone: Mutual telexistence master-slave communication system based on retroreflective projection technology," *SICE Journal of Control, Measurement, and System Integration*, vol. 1, no. 5, pp. 335–344, 2008.

[2] D. Sakamoto, T. Kanda, T. Ono, H. Ishiguro, and N. Hagita, "Android as a telecommunication medium with a human-like presence," in *Human-Robot Interaction (HRI), 2007 2nd ACM/IEEE International Conference on*. IEEE, 2007, pp. 193–200.

[3] I. Rae, B. Mutlu, and L. Takayama, "Bodies in motion: mobility, presence, and task awareness in telepresence," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 2014, pp. 2153–2162.

[4] D. Sirkin and W. Ju, "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*. ACM, 2012, pp. 57–64.

[5] B. Jin and N. Park, "Mobile voice communication and loneliness: Cell phone use and the social skills deficit hypothesis," *New Media & Society*, vol. 15, no. 7, pp. 1094–1111, 2013.

[6] M. McLuhan, *Understanding media: The extensions of man*. MIT press, 1994.

[7] H. Nakanishi, K. Tanaka, and Y. Wada, "Remote handshaking: touch enhances video-mediated social telepresence," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 2014, pp. 2143–2152.

[8] A. Kadomura, A. Matsuda, and J. Rekimoto, "Casper: A haptic enhanced telepresence exercise system for elderly people," in *Proceedings of the 7th Augmented Human International Conference 2016*. ACM, 2016, p. 2.

[9] M. Shiomi, K. Abe, Y. Pei, T. Zhang, N. Ikeda, and T. Nagai, "Chicaro: Tele-presence robot for interacting with babies and toddlers," in *Proceedings of the Fourth International Conference on Human Agent Interaction*. ACM, 2016, pp. 349–351.

[10] N. Nakazato, S. Yoshida, S. Sakurai, T. Narumi, T. Tanikawa, and M. Hirose, "Smart face: enhancing creativity during video conferences using real-time facial deformation," in *Proceedings of the 17th ACM conference on Computer supported cooperative work & social computing*. ACM, 2014, pp. 75–83.

[11] K. Otsuka, S. Kumano, R. Ishii, M. Zbogar, and J. Yamato, "Mm+ space: nx 4 degree-of-freedom kinetic display for recreating multiparty conversation spaces," in *Proceedings of the 15th ACM on International conference on multimodal interaction*. ACM, 2013, pp. 389–396.

[12] R. Ishii, S. Ozawa, T. Mukouchi, and N. Matsuura, "Mopaco: Pseudo 3d video communication system," in *Symposium on Human Interface*. Springer, 2011, pp. 131–140.

[13] O. Morikawa and T. Maesako, "Hypermirror: toward pleasant-to-use video mediated communication system," in *Proceedings of the 1998 ACM conference on Computer supported cooperative work*. ACM, 1998, pp. 149–158.

[14] O. Oyekoya, W. Steptoe, and A. Steed, "Sphereavatar: a situated display to represent a remote collaborator," in *Proceedings of the SIGCHI conference on human factors in computing systems*. ACM, 2012, pp. 2551–2560.

[15] K. Misawa, Y. Ishiguro, and J. Rekimoto, "Livemask: A telepresence surrogate system with a face-shaped screen for supporting nonverbal communication," *Information and Media Technologies*, vol. 8, no. 2, pp. 617–625, 2013.

[16] K. Misawa and J. Rekimoto, "Chameleonmask: Embodied physical and social telepresence using human surrogates," in *Proceedings of the 33rd Annual ACM Conference Extended Abstracts on Human Factors in Computing Systems*. ACM, 2015, pp. 401–411.

[17] K. Tanaka, N. Yamashita, H. Nakanishi, and H. Ishiguro, "Teleoperated or autonomous?: How to produce a robot operator's pseudo presence in hri," in *The Eleventh ACM/IEEE International Conference on Human Robot Interaction*. IEEE Press, 2016, pp. 133–140.

[18] G. Wolberg, "Image morphing: a survey," *The visual computer*, vol. 14, no. 8, pp. 360–372, 1998.

[19] "dlib C++ Library," http://dlib.net/.

[20] V. Kazemi and S. Josephine, "One millisecond face alignment with an ensemble of regression trees," in *27th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2014, Columbus, United States, 23 June 2014 through 28 June 2014*. IEEE Computer Society, 2014, pp. 1867–1874.

[21] B. Shen, I. K. Sethi, and V. Bhaskaran, "Dct domain alpha blending," in *Image Processing, 1998. ICIP 98. Proceedings. 1998 International Conference on*, vol. 1. IEEE, 1998, pp. 857–861.

[22] R. Hayashi and S. Kato, "Psychological effects of physical embodiment in artificial pet therapy," *Artificial Life and Robotics*, vol. 22, no. 1, pp. 58–63, 2017.

[23] K. F. MacDorman and H. Ishiguro, "The uncanny advantage of using androids in cognitive and social science research," *Interaction Studies*, vol. 7, no. 3, pp. 297–337, 2006.

[24] R. S. Nagayama, "The uncanny valley: Effect of realism on the impression of artificial human faces," 2007.

[25] H. Sumioka, A. Nakae, R. Kanai, and H. Ishiguro, "Huggable communication medium decreases cortisol levels," *Scientific reports*, vol. 3, p. 3034, 2013.

[26] H. IJzerman, M. Gallucci, W. T. Pouw, S. C. Weiβgerber, N. J. Van Doesum, and K. D. Williams, "Cold-blooded loneliness: social exclusion leads to lower skin temperatures," *Acta psychologica*, vol. 140, no. 3, pp. 283–288, 2012.