International Journal of Social Robotics (2020) 12:187–199
https://doi.org/10.1007/s12369-019-00551-w

A Robot for Test Bed Aimed at Improving Telepresence System and Evasion from Discomfort Stimuli by Online Learning

Masahiko Osawa[1] · **Michita Imai**[1]

Accepted: 15 April 2019 / Published online: 25 April 2019

Abstract
This study contributes to improving the comfort of telepresence communication by adaptations to people. We developed a handheld telepresence robot comprised of a binaural microphone and a head mounted display as a test bed and conducted surveys on unpleasantness caused by special devices in use. We found that numerous people experienced an increase in unpleasant sound when the binaural microphone was being used. It was also found that types of unpleasant stimuli differed from person to person. Furthermore, we propose an automatic unpleasant stimuli avoidance system using online machine learning architecture constructed from echo state networks (ESNs) and Accumulator Based Arbitration Models (ABAMs) that can also flexibly adapt to remote users. Because the handheld telepresence robot can avoid unpleasant stimuli before remote users experience these stimuli, it provides them with a more comfortable communication environment while notifying people around the robot that uncomfortable stimulation is being avoided.

Keywords Telepresence robot · Machine learning · Online user adaptation · Unpleasant stimuli · Binaural microphone

1 Introduction

It is important to maintain a high social presence [1,2] in remote communication, i.e., a sense of dialogue that occurs with other people in face-to-face communication. It is becoming clear that social presence can be improved by using video along with audio in remote communication [2,3]. However, remote participants in a conference setting may be forgotten, even with video. This suggests that remote communication with video fails to sufficiently present the presence of remote participants [4,5]. Therefore, telepresence robots that are aimed at remote communication have been attracting attention.

It is important to design telepresence robots in a way that remote users gain realistic feelings as if they were in the same environment as local users. Therefore, much research has been conducted in attempts to enhance the sense of presence by using special devices [6–10]. However, these researches have not taken into consideration the disadvantages caused by using such devices. Even if a telepresence robot is used, remote users are (B-2) in weak positions. Local users need to care about remote users to communicate.

This paper presents a handheld telepresence robot that uses binaural microphones and head mounted displays (HMDs) that was introduced as a test bed. This robot is a design that is useful for verification in order to improve the remote user's weak position. Especially this research explains how we investigated the disadvantages for remote users of using special devices (Fig. 1). It turned out that there were three types of unpleasant auditory stimuli for many people when using binaural microphones. We found from the investigation that unpleasant stimuli differed depending on the individual remote user and situations. Therefore, we further propose a system that can be adapted to remote users online and that automatically avoids unpleasant stimuli. Since the proposed system can dynamically learn users' levels of discomfort, it can flexibly adapt to various users and situations. Evaluations from experiments indicated that it was possible for the system to respond flexibly to individual users and stimuli, and that there was less discomfort when using the function to automatically avoid unpleasant stimuli.

This paper makes three main contributions. First, it reveals the cons of enhancing the sense of immersion by using special devices. Second, it provides the first trial that was used to demonstrate the possibility of improving the level of comfort

✉ Masahiko Osawa
mosawa@ailab.ics.keio.ac.jp

[1] Keio University, Tokyo, Japan

Fig. 1 Motivation behind current work. The top photograph indicates the situation before learning has occurred for adaptation. The bottom photograph indicates the situation after learning has occurred for adaptation

of telepresence communication by online machine learning. Third, it explains how we created a telepresence system that dynamically adapts to operators and how we assessed the effectiveness of the system.

The structure of this paper is as follows. In Sect. 2, we introduce related works of this research. We describe a configuration of telepresence robot as a test bed in Sect. 3, and show case studies investigating the usability of the robot and analyzing unpleasant stimuli in Sect. 4. In Sect. 5, we explain the configuration and learning method of a detector of unpleasant stimuli that we developed on the basis of the experiment results in Sect. 4 and present a technique for autonomous avoidance behavior. Section 6 presents our evaluation of the proposed system to detect unpleasant stimuli and execute autonomous avoidance behavior. We conclude in Sect. 7 with a brief summary.

2 Related Works

2.1 Overview of the Telepresence Robots

Kristoffersson et al. administered a survey that overviewed telepresence robots [11]. A remote user is a person who is remotely connected to a robot via a computer interface. A local user is a user that is situated in the same physical location as the robot.

There are two types of telepresence robots. The first type is the mobile robotic telepresence (MRP) system. One of the typical example of the MRP is a system named PRoP. [12,13]. Many telepresence systems like PRoP have been proposed [14,15]. These systems have allowed remote users to move around in the local environment. The second type of telepresence robots has mobility functions that are replaceable. These telepresence robots were created to promote the sense that a remote user was in the same place as a local user through MRP [16–20]. This paper mainly deals with the latter type of telepresence robots.

2.2 Enhance Immersions

Past studies on telepresence robots have attempted to improve the sense of immersion and realism by devising methods of presenting sensor information to remote users to improve the sense of social presence. For example, TeleHead [6,7] enabled a more realistic sound experience by using dummy heads designed in the same shape as that of the remote user. It has been reported that a camera with a wide viewing angle, image distribution with a high frame rate [8], life-sized image display, and stereoscopic viewing [9] also improve the sense of social presence. Fernando et al.'s research has been advancing the development of a robot known as TELESAR that not only presents images and sounds but also tactile sensations and thermal sensations to remote users [10]. Takahashi et al. and Hayamizu et al. [21,22] have proposed a telepresence system that can calculate an appropriate volume from the noise and the distance to the speaker and can listen to the voice of an arbitrary local user by using a microphone array.

It is important to consider discomfort caused by special devices when using them. Virtual reality (VR) sickness caused by head-mounted displays (HMDs) is known to discomfort users of special devices. Measures against VR sickness are being considered, but the possibility of other discomfort situations arising from special devices has not been considered. Discomfort may particularly be caused by the situation with communication, and as it may differ for each individual, it is different from VR sickness that constantly becomes unpleasant during use. Further, its mechanism is dynamically acquired. For example, when binaural recording is used [6,7], it is necessary to speak at an appropriate volume and distance by taking into consideration the sounds presented by a local user to a remote user in the same space as the robot. However, there is currently no method of maintaining the level of comfort of remote users. Takahashi et al. and Hayamizu et al. [21,22] did not consider the discomfort of remote users in their research.

2.3 Discomfort in Telecommunications

Theofilis et al. pointed out the problem of image delay in telepresence systems and presented a method of reducing the perceived visual latency during remote robot teleoperation [23]. They did not visualize the stream of the camera directly into the user's VR headset like many other methods and proposed a stereoscopic technique of panorama reconstruction to compensate for head movements.

Hasegawa et al. dealt with the problem of utterance conflict when participating in a conversation using a telepresence system [24–27]. By devising the behavior of the robot, they constructed a system that made it hard to cause speech collision in advance.

Our research also dealt with discomfort levels when using a telepresence robot. We particularly focused on the discomfort experienced by remote users when improving the sense of presence by using the special device described in the previous section.

We designed a system to prevent VR sickness with sufficient consideration for what has already been demonstrated as a general solution: avoidance of low resolution and low frame rates. This can be done by utilizing the communication standard of the robot operating system (ROS). Then, by enabling the telepresence robot to be handheld, we checked whether uncomfortable feelings occurred under conditions where the remote user's image was increasingly disrupted due to local users' influence. We also examined the discomfort caused by binaural microphones since not much research has been done on this despite its effectiveness in telepresence communication.

2.4 Motion Capture

There have been various studies that have presented a sense of presence through robot behaviors. For example, Hasegawa et al. [27] reported that a multi-degree of freedom (multi-DOF) telepresence robot could express the preliminary motion of a remote user so that speech alternation could be efficiently performed. Matsui et al. aimed at facilitating remote communication by expressing information acquired by motion capture to a telepresence robot that had a very close appearance and mannerisms to a person [28]. Fernando et al.'s research [10] also used motion capture to realistically reproduce the movement of the remote user on the robot.

The movements of telepresence robots in the existing research [10,27,28] were designed to support telecommunication. The telepresence robots, on the other hand, did not offer any mechanism to let the local user know that the remote user was experiencing discomfort. This is problematic because it is important to express the internal state of the remote user, and if he/she is in an uncomfortable situation, it is important to tell the local user what the remote user felt uncomfortable about.

In addition, most telepresence systems using motion capture require robots to have multiple-DOFs. However, robots with multiple-DOFs entail high costs of creation and control. Therefore, we considered a method of conveying discomfort to remote users with only one-DOF robot in this research.

2.5 Simple and Effective Behavior

Funakoshi et al. reported that communication could be facilitated by expressing the internal state of a robot through an easily implementable behavior unique to the agent that did not imitate human behavior [29]. Their system was called artificial subtle expression (ASE).

Even if a telepresence robot has one-DOF, ASEs are highly likely to be effective when they are indicating that remote users are uncomfortable. Therefore, we actively utilized knowledge on ASEs in this research.

2.6 Semi-Autonomous Functions

Semi-autonomous functions are increasingly useful on MRP systems [30,31]. A typical use is to automatically move to the location pointed to by a remote user. There have been attempts to implement autonomous movement in a telepresence system using a humanoid robot called NAO [32].

Tanaka et al. implemented both remote and autonomous operation of a robot [33] and investigated when a local user felt that a remote user was remotely operating it. Choi et al. investigated the effect of a telepresence robot that mimicked the movements of a local user [34]. The authors are also working on the development of semi-autonomous systems adapted to remote users [35,36].

We propose new types of semi-autonomous functions in this paper. The remote user's level of discomfort is learned by online learning, and thus far the robot has performed partly automatic evasive actions that the user had to perform.

2.7 Machine Learning Method for Telepresence System

As the robot on the local user side in the telepresence system can receive the same sensor information as the remote user and can fully know the control information of the remote user, supervised learning can effectively be used.

As far as we know, there have been no cases where online machine learning has been introduced to cut off uncomfortable stimuli like that in the research reported in this paper.

2.8 User Adaptation Using Online Learning

Obo et al. proposed an on-line learning method based on spiking neurons for modeling human states in a support system for the elderly [37]. Tapus et al. [38] also proposed personal adaptation and a learning method in a social support robot system that could perform physical rehabilitation and cognitive stimulation to cope with the problem of an aging society. The system could adapt to individual disability levels.

Fukuda et al. [39] proposed a robot arm control system using biomedical signals. They used a novel statistical neural network called the log-linearized Gaussian mixture network (LLGMN) to distinguish biomedical signals and the system adapted to changes in biological signals such as individual differences and fatigue caused by online learning. Ghahramani et al. [40] proposed a method of adapting online learning by using a Bayesian classifier, by focusing on the fact that thermal comfort differs for people and changes due to climate change in an air conditioning management system.

We dealt with user adaptation in a telepresence system by using online machine learning technology in this research. We combined two machine learning techniques to achieve both robust learning of common discomfort experienced by many people and adaptation to individual users at high rates of speed.

2.9 (A-3) Related Learning Methods

Here, we give a brief explanation of machine learning methods related to this study. To realize a telepresence robot that can adapt to human, it requires a method that can finish training with the same time scale as human. We consider combining two machine learning methods to adapt to human and express behavior.

First is a machine learning method that can handle chronological data. Various methods that can handle chronological data have been suggest up to this point [41–46].

(A-5) From such, we use the echo state networks (ESNs) [42,43] in this study. An ESN has an intermediate layer composed of random bonds, and since only the coupling between the intermediate layer and the output layer is determined by linear regression, the parameters can be determined relatively quickly. The advantage of ENS is that it does not need to study the recursive part of its network making learning fast. Although there are reports on models with higher classification accuracy than ESN, this study did not need any advanced analysis such as speech recognition and only needed functionality to classify the stimuli given by the sound. Therefore, we believe that ESN was the most suited for the task. In long term, we would need a fast-enough model that can adapt to human. This makes the usage of fast learning model in this study ideal when considering future studies.

The second is a method that arbitrates the expression of behaviors based on stimuli classification output. For the arbitration method, we used an accumulator based arbitration method (ABAM) [47] suggested by the authors. When the cumulative evidence, A^t, in the ABAM represented by the following equation exceeds the threshold value, θ^t, the corresponding action is expressed:

$$A^t = \begin{cases} rA^{t-1} + y^t & (A^t < \theta^t) \\ 0 & (A^t \geq \theta^t) \end{cases} \quad (1)$$

where r is the discount rate and y^t is the output of the ESN. When A^t exceeds the threshold value, it is possible to prevent the same actions from consecutively being output by resetting A^t to 0. When $A^t > \theta^t$, it expresses the avoidance behavior that will be described in the next section.

We extended ABAM to allow online learning. Details will be given in 5.1.2.

3 Handheld Telepresence Robot

As was previously explained, it is important to focus on research that has a sense of discomfort in a telepresence environment, and it is particularly important to preemptively block discomfort stimuli and notify local users that the remote user is experiencing discomfort.

This paper presents a handheld telepresence robot as a test bed with a binaural microphone that enhances the immersion and realism experienced by the remote user and discusses how he/she can avoid discomfort.

Figure 2 is a photograph of the handheld telepresence robot. It was controlled by Raspberry Pi 3. Four modules were connected to Raspberry Pi 3: a camera module, a touch display module, a universal serial bus (USB) audio interface, and a servomotor via general purpose input/output (GPIO). The robot can display remote user's faces and present the stimuli touching the display to remote user using its touch display, but we didn't use these functions in this research in

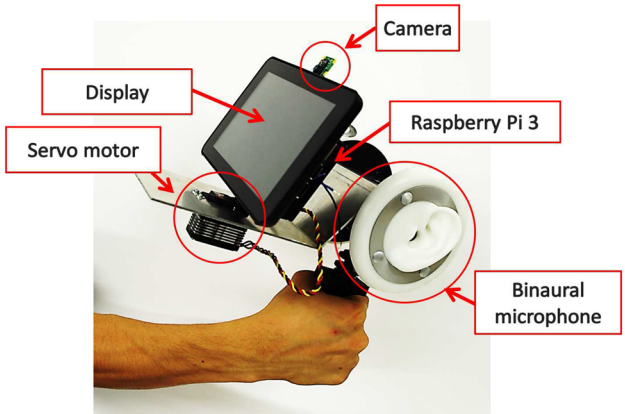


Fig. 2 Handheld telepresence robot

order to examine in detail the situation that simple, remote user is weak position.

A 3Dio Free Space binaural microphone was connected to the USB audio interface, which enabled the remote user to perceive the information acquired by the robot and promote an immersive and realistic feeling. One disadvantage of the conversational distance being controlled by the local user is that the remote user tends to experience an overly heightened sense of reality initiated by the local user. For example, if a local user shouts into the binaural microphone, the remote user will suffer a great deal of discomfort. Using a handheld telepresence robot makes it easy to evaluate discomfort caused by a high sense of presence.

The images and sounds acquired by the robot were transmitted to the remote user by wireless communication using Wi-Fi standardized in Raspberry Pi 3 or a wired network.

The robot did not have a mobile mechanism; rather, the local user who interacted with the robot gripped the handgrip of the handheld telepresence robot and carried it around while communicating. Since the local user dominated the distance between himself/herself and the robot, disadvantages related to an excessive feeling of realism originating from the local user were likely to occur. It was easier to assess the discomfort of stimuli caused by the local user by using the handheld structure. The device had one servomotor that could be used to express the internal state of the robot and remote user. The servomotor was connected to a movable stage equipped with a Raspberry Pi 3, a touch display, and a camera. Each could be rotated from left to right.

All robot software consisted of an asynchronous distributed system by using an ROS. Therefore, the Raspberry Pi 3 controlling the robot only acquired sensor information and expressed behaviors, and all calculation processing related to the machine learning that will be explained in the next section could be performed on the remote user's PC.

Pilot users had a control PC, an HMD (HMZ-T1) connected to the PC, and an earphone suitable for playback of binaurally recorded sound. This enabled them to experience the information transmitted from the robot.

We cautioned them beforehand not to create discomfort such as typical and familiar VR sickness. The video delay was sufficiently short and had high resolution and a high frame rate.

4 Case Studies

4.1 Study 1: Free Conversation

We conducted a case study to investigate what effects the sound and installation environments of the handheld telepresence robot had on the telepresence environment.

4.1.1 Method

Each participant conversed freely under four conditions: combinations of two types of installation environments (hand/desk) and two types of sound environments (binaural/monaural). In both installation environments, the distance between the robot and the local user, which is mainly involved, was about 50 cm. It was set so that information other than the sound quality such as volume of the two microphones was about the same. The handheld telepresence robot was installed in a room with one or more local users. There were two types of participants: the remote users of the telepresence robot and the local users around the robot. The remote users were not given any specific instructions. As there were no set limits on the time for the experiment, it ended when the participants wanted it to end. The average experimental execution time was about ten minutes to conduct all four conditions for each pair of participants.

4.1.2 Participants

Three men in their 20s conducted free remote communication using a handheld and a stationary telepresence robot.

(A-2) This experiment was performed without any consistency and done under a setting of natural conversation on purpose to raise some hypotheses of possible cons when using the newly developed telepresence environment. Therefore, it is difficult to find any statistically significant difference from the results of this experiment. Additional experiment and analysis will be done in Sect. 4.2 based on findings from this experiment so this experiment is limited to few participants.

4.1.3 Metrics

Subjective assessment was carried out by means of a questionnaire and interview after the experiment. Three items related to the system were evaluated on a Likert scale from one to seven and ten items concerning discomfort stimuli were evaluated on a Likert scale from one to three.

4.1.4 Results

The responses to the questionnaire administered to the remote user participants are given in Fig. 3. The binaural and handheld conditions were the highest in all items for the level of system satisfaction. However, the number of unpleasant feelings was also the highest under these binaural and handheld conditions. These findings were in line with our expectations.

Some unpleasant stimuli were reported by the remote users such as local users touching the microphone and using loud voices. This was why the number of unpleasant feelings under binaural conditions was higher than that under monau-

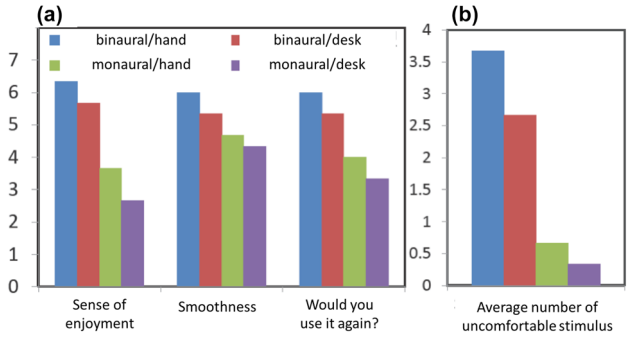


Fig. 3 (B-1) Results from subjective evaluation. **a** Satisfaction with system. **b** No. of uncomfortable stimuli felt by remote users

ral conditions. Although there were cases that it was simply uncomfortable due to the volume problem, there were more experiment participants who reported the influence of the sound quality such as "feeling creepy" or "ticklish feeling". While the presence of the remote users was emphasized under handheld conditions, and communication was clear, remote users felt uncomfortable with the amounts of trembling and large sways that accompanied the movements of local users.

4.2 Study 2: Recorded Videos

The case study discussed in Study 1 examined usability under free communication, and was not managed in a controlled environment. We therefore conducted an experiment using recorded video in Study 2 to evaluate the discomfort felt by remote users in a more controlled environment.

4.2.1 Method

We selected 15 irritating stimuli that the remote user was apt to find unpleasant that were related to sounds and images: ten items from the experiment in Study 1 and five new items selected after the interviews with the participants in the experiment. We created a approximately one-minute scenario (listed in Fig. 4) that included discomforting stimuli that were selected and asked two female speakers and one male speaker to perform the scenario under the same four conditions as those in Study 1. In both installation environments, the typical distance between the robot and the local user was about 50 cm (scenario 1, 2, 6). In scenario 4 and 5, the distances were between 3 and 10 cm, and about 4 m in scenario 3 . The volume of the two microphones was about the same. We then recorded it. We then selected the video of a male speaker who seemed to present the most unpleasant stimuli according to a certain preliminary experiment.

1. A simple greeting and explanation of binaural recording by a local user.
2. The local user speaks from the left and right.
3. The local user speaks from far away.
4. The local user speaks from a close distance (whispering in the ear).
5. Presentation of unpleasant stimuli by the local user:
— 5.1 Touching the microphone.
— 5.2 Covering the microphone with his/her hand.
— 5.3 Placing his/her finger on the microphone.
— 5.4 Breathing into the microphone.
— 5.5 Speaking loudly into the microphone.
6. A simple greeting by the local user.

Fig. 4 Scenario

One participant for each experiment watched a video through HMD under four conditions: a combination of two types of installation environments (hand/desk) and two types of sound environments (binaural/monaural).

4.2.2 Participants

Ten men and one woman in their 20s watched the recorded videos using HMDs and dedicated earphones for binaural recording. After the experiment was complete, they answered a subjective assessment that was administered by questionnaire. (A-2) The number of participants was decided based on number assumed to be needed to perform statistical analysis of significance.

4.2.3 Results

Figure 5a shows the results obtained from subjective evaluation when binaural and monaural conditions were compared.

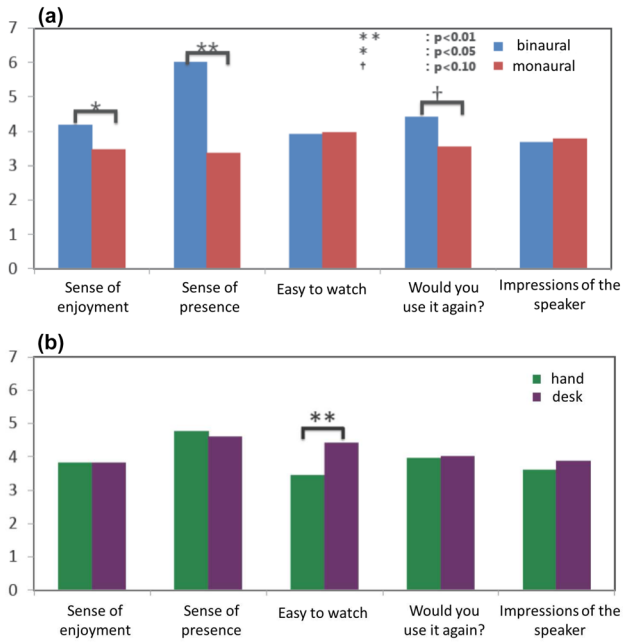


Fig. 5 (B-1) Results from subjective evaluation: **a** binaural versus monaural conditions and **b** hand type versus desk type

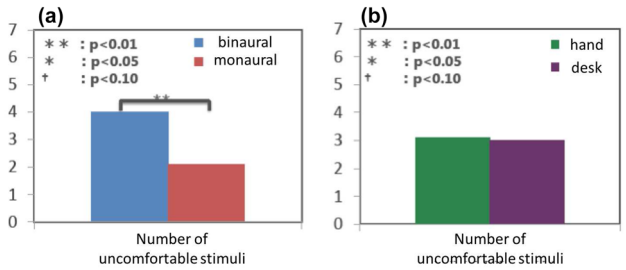


Fig. 6 (B-1) Comparison of number of stimuli remote users who felt uncomfortable: **a** binaural versus monaural conditions and **b** hand type versus desk type

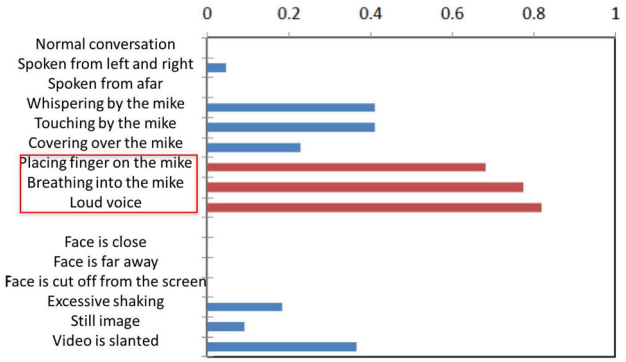


Fig. 7 Percentage at which remote users felt irritation was unpleasant

There was a significant difference in "sense of enjoyment" and "sense of presence", and a dominant tendency was observed with respect to the question of "would you use it again?". Figure 5b shows the results when the hand and the desk type conditions were compared. A significant difference was identified in terms of "easy to watch". However, when we conducted a test on the installation × sound interaction, there were no items that demonstrated any significant differences.

Next, Fig. 6 shows the results obtained on the numbers of stimuli that remote users felt were unpleasant. When comparing the binaural and monaural conditions, it seems there were many stimuli under the binaural conditions that were deemed unpleasant, which were similar to the results from Study 1. However, when comparing the hand and desk type conditions, there were no significant differences. There were also no significant differences in the installation × sound interaction. This suggests that the effect of binaural microphones was significant from the viewer's perspective.

Finally, Fig. 7 shows the rate at which remote users reported stimuli as unpleasant. The percentage of remote users reporting stimuli related to sounds is higher compared to that of installations. Notably, stimuli reported by a majority of participants as discomforting are "placing fingers on the microphone", "breathing into the microphone", and using a "loud voice".

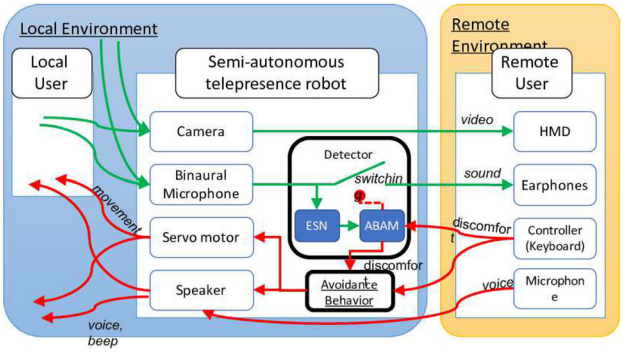


Fig. 8 The control/interaction framework and online learning architecture of the robot system. Local user faces the semi-autonomous telepresence robot under the same environment. The remote user remains at a different environment from both local and semi-autonomous telepresence robot and uses HMD, Earphones, Keyboard as controller, and Microphone to communicate with the remote user via the semi-autonomous telepresence robot. The green arrows convey information to the remote user. The red arrows convey information to the local user. The areas depicted by bolded line within the semi-autonomous telepresence robot indicates the online learning architecture. When the combined training model of ESN and ABAM detects unpleasant stimuli, the sound transmission to the remote user stops for a brief moment and express the avoidance behavior using the motor and speaker. (Color figure online)

4.3 Consideration

The experimental results obtained from the case studies suggest that although using binaural microphones improved user satisfaction, it came at the cost of many uncomfortable stimuli. We expected that irritating stimuli would increase even when the handheld device was used, but no significant differences were observed under the condition of watching the video, as had occurred under this experimental condition.

There were many experiment participants who reported the influence of the sound quality such as "feeling creepy" or "ticklish feeling" and it is a serious problem caused by emphasizing the presence feeling that has not been pointed out so far. It can not deal with the methodology that simply adjusts the volume dynamically, and it is necessary to detect the sound that the remote user feels unpleasant

We concluded that a system that could learn which sounds the remote user found uncomfortable and then automatically detected/avoided these sounds was effective, and decided to implement it. Discomfort to identify what should be learned dynamically by on-line learning and its adaptation to users were also important.

5 Automatic Avoidance

This section explains the detection of unpleasant sound stimulation and avoidance behavior that appeared when remote users detected unpleasant stimuli. The discomfort caused

by using the binaural microphone could be reduced by avoidance behavior, which is proposed in this section. The control/interaction framework and online learning architecture of the robot system are outlined in Fig. 8.

5.1 Learning of Unpleasant Stimulation

We found that there were three kinds of stimuli that were regarded as being unpleasant from the experimental results in Study 1 by the majority of remote users: "placing fingers on the microphone", "breathing into the microphone", and using a "loud voice". Here, we recorded these three stimuli presented in each of the left and right microphones for 1 min by two speakers to enable the machine to learn discomfort. We also read and recorded about 2.5 min of the beginning of "Run, Melos", which is a traditional Japanese novel, by two speakers to use it as a dataset of comfortable stimuli. Audio was collected at a sampling frequency of 16,000 Hz for one ear. From the 320 data points acquired every 10 ms, 256 data points from the front were Fourier transformed and converted into frequency spectra, and then the frequency bands that were evenly thinned into 32 dimensions were used as inputs to the learning device.

The detector we used consisted of the two modules described in Sect. 2.9.

5.1.1 The Echo State Network

The input was set to 32 dimensions, the intermediate layer was set to 300 dimensions, and the output was set to one dimension, which corresponded to the degree of discomfort. The coupling probability between the input layer and the intermediate layer was set to 3% and the coupling probability in the intermediate layer was set to 10%, and each combination was determined by a uniformly distributed random number between -0.5 and 0.5. All of the above parameters were decided appropriately by preliminary experiments

5.1.2 The Extended Model of Accumulator Based Arbitration Model

Here, we extended ABAM to allow online learning. In the setting in this paper, only threshold is learned by adapting to the user. It is necessary to decide the default value of threshold and the discount rate in advance.

$$\theta^0 = \theta_{default} \tag{2}$$

$$\theta^t = \begin{cases} \theta^{t-1} - lr_a * (\theta^{t-1} - \theta^{t-1}_{target}) & (disc^{t-1}_{op} = 1) \\ \theta^{t-1} - lr_b * (\theta^{t-1} - \theta_{default}) & (disc^{t-1}_{sys} = 0) \\ \theta^{t-1} & (\text{otherwise}), \end{cases} \tag{3}$$

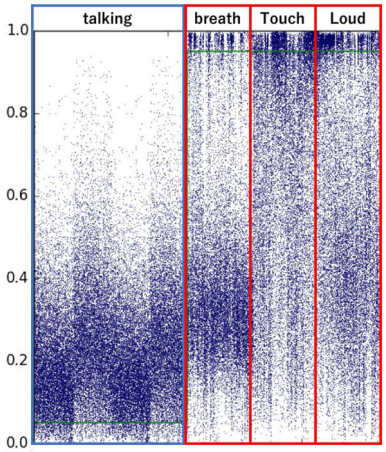


Fig. 9 Identification results of the echo state network. A horizontal axis represents time, and a vertical axis represents a scalar value indicating discomfort. Normal speech, breath sound, sounds touching microphone, loud voice are input in order

where lr_a and lr_b represent the learning rate, $disc^t_{op}$ is set to one if the operator feels discomfort, and $disc^t_{sys}$ is set to one if the system detects discomfort. In most cases, lr_a should be larger than lr_b.

5.1.3 (A-7) The Accuracy of Training Model

First, we will explain the accuracy of ESN by itself trained using the created data set. The data set consisted of chronological data containing 187,296 data points with the first 90% being used as training data for normal stimuli and the three unpleasant stimuli. The latter 10% was used as a test data. In total, 168,560 data points worth of training data set and 18,736 data points worth of test data set was created. Figure 9 shows identification results of the ESN. The average classification error against the test data was 0.275. When using ESN as a binary classifier determining any output above 0.5 to be unpleasant, the accuracy changed to 0.750. Both the average error and binary classification cannot be said to have extremely high accuracy. Two reasons can be brought up as to why the model cannot achieve extremely high accuracy against the data set. First, ESN has the tradeoff of having fast training speed but loses to other chronological machine learning models when it comes accuracy. In need of higher accuracy model, we think it can be achieved by using a different training model. Second, the data sets were created by the authors and have quite a few noises in both training and test data set. To avoid over fitting, the accuracy against the data seems to be low. To prove such point, the model has an average error rate of 0.272 and accuracy of 0.771 when used as binary classifier against the training data set.

It cannot be said that the ESN's accuracy is extremely high, but by using the Accumulator Based Arbitration

Model (ABAM), it is possible to properly express the action. For experiment in Sect. 6.2, the suggested algorithm (ESN + ABAM) was able to detect all unpleasant stimuli of all users (0% false negative). In addition, within the experiment, the robot never took the avoidance behavior when there were no unpleasant stimuli (0% false positive). Although 0% false positive and false negative may be a skeptical statement, a very important property of suggested algorithm plays a role in realizing it. ABAM monitors ESN outputs over an extended period of time and takes action based on accumulated evidences collected from the output stabilizing the robot's behavior even when the ESN have a spontaneous false output. Obviously, the false positive and negative does not always maintain 0%. Especially like in Sect. 6.1, before the system adapts, the false negative was at a high number and in Sect. 6.2, the experiment was done after fixing the parameter once it sufficiently adapted to human. Above reasons are thought to be why the robot did not express any "wrong" behaviors.

5.2 Expression of Avoidance Behavior

The proposed avoidance behavior simultaneously presented two kinds of behaviors. The first was a beep. Komatsu and Yamada reported that beeps make it easier for humans to understand that artifacts' are representing negative internal conditions [48]. We adopted one that uniformly changed its frequency from 256 to 6 Hz within the expression time of 418 ms. The second avoidance behavior was rotation of the display. The angle of rotation was 30 degrees, both negative and positive, relative to the state facing the front, and the time from start to stop was 1.5 s. Audio streaming playback to the remote user was stopped at the same time the avoidance behavior occurred.

6 Evaluation Expression

6.1 Evaluation 1: Adaptation to Pilot Users

This section describes how we investigated whether the proposed telepresence system could be adapted to remote users through online learning.

6.1.1 Method

The presented discomfort stimuli consisted of the three stimuli reported by the majority of remote users as unpleasant from the results of case studies. Each participant performed three experiments. They were presented an unpleasant stimulus in each experiment.

Participants could undertake avoidance behaviors by pushing buttons when they felt uncomfortable. The robot then learned based on the sensed stimuli and discomfort felt by the participants.

6.1.2 Participants

Two men in their 20s participated in the experiment. (A-2) There are only two participants in this experiment because its aim was to visualize and ensure that the suggested method was adapting to the participants. We were not comparing with other systems nor were we trying to find any statistically significant difference, we believe that enough results can be gathered with two participants in terms of visualizing how it adapts to different users.

6.1.3 Metrics

We made qualitative assessments associated with respect to time by visualizing four factors, i.e., recognized discomfort, accumulated discomfort, threshold to undertake avoidance behaviors, and the moment avoidance behaviors were automatically engaged.

6.1.4 Results

Figure 10 plot the results from our evaluation of the system.

Participant 1 in the experiment felt uncomfortable when the stimuli of " air blown on the microphone" and "finger on the microphone" were presented, so the threshold was lowered and the system sensitively took evasive action. In contrast, he did not feel uncomfortable with the stimulus of a "loud voice", so the system autonomously did not express avoidance behaviors. Meanwhile, participant 2 in the experiment felt sensitive discomfort to the "loud voice", so the threshold that was learned was low. As a result, the system sensitively took evasive action.

6.2 Evaluation 2: Quantification of Discomfort

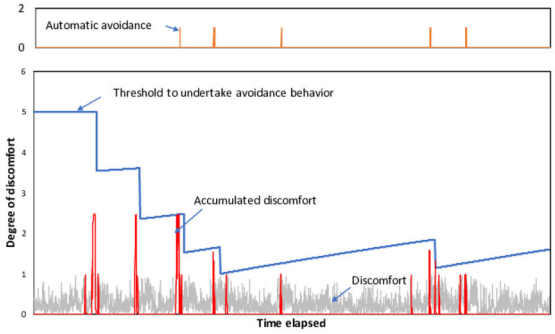
This section describes the evaluations we carried out to examine what effect the autonomous avoidance functions had on the discomfort felt by remote users.

6.2.1 Method

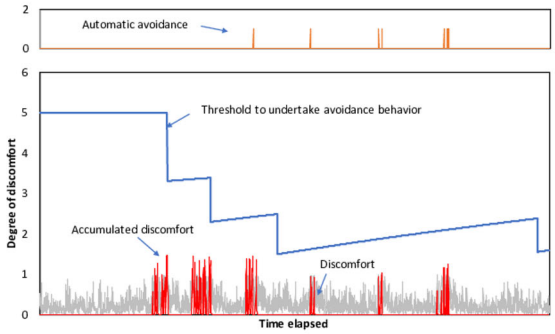
Three systems were used to evaluate the extent of discomfort felt by remote users when three kinds of unpleasant stimuli were presented:

Automatic expressed autonomous avoidance behaviors (proposed),

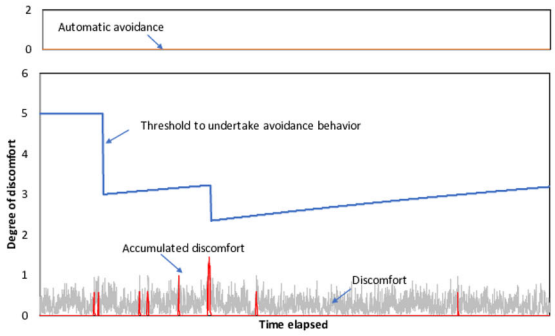
Manual in which remote users expressed avoidance behaviors themselves, and



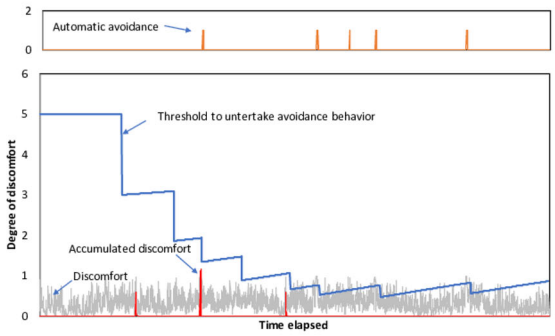
(a) Participant 1, Stimulus: Breathing into the microphone



(b) Participant 1, Stimulus: Placing finger on the microphone



(c) Participant 1, Stimulus: Loud voice



(d) Participant 2, Stimulus: Loud voice

Fig. 10 Results of online adaptation. The horizontal axis in each graph represents time. The waveform peaks in the top graphs indicate the timing at which automatic avoidance behavior was expressed. The bottom graphs indicate the adaptive accumulator threshold and cumulative evidence at each time, and the discomfort index calculated by ESN. The vertical axis is a scalar value that corresponds to the discomfort level

Immovable in which they did not express avoidance behaviors.

The presented discomfort stimuli consisted of the three stimuli reported by the majority of remote users as unpleasant from the results in Study 2. A telepresence robot was installed in the student room of a laboratory together with an experimenter, who participated as a local user. Participants in the experiment communicated in a separate room as remote users of the telepresence robot. The local user presented unpleasant stimuli while using each of the three systems. Both systems and stimuli were randomized for all participants in the experiment.

Hyper parameters were adjusted beforehand so that three unpleasant stimuli could be detected. The threshold was fixed at 1.5 in this system and the discount rate was set to 0.3.

6.2.2 Participants

Ten men and two women in their 20s participated in the experiment. (A-2) The number of participants was decided based on number assumed to be needed to perform statistical analysis of significance.

6.2.3 Metrics

All participants were administered questionnaires and were interviewed after the experiments. Evaluation items included a seven-level Likert scale for five items related to the system and a seven-level Likert scale for three items related to the uncomfortable stimuli.

6.2.4 Expected Results

We expected two phenomena to occur prior to the experiment.

- When avoidance behaviors were appropriate, the two systems that expressed avoidance behaviors would be more satisfactory and less uncomfortable than systems that did not express avoidance behaviors.
- When the accuracy of detection of unpleasant stimuli was sufficient, **automatic** and **manual** would obtain the same degree of precision.

6.2.5 Results

There were no false positive/false negative, namely, whenever unpleasant stimuli were presented, it showed evasive behavior and it never took evasive behavior even though it did not present an unpleasant stimulus.

Figure 11 shows the results obtained from our evaluation of the system. We used analysis of variance (ANOVA) to

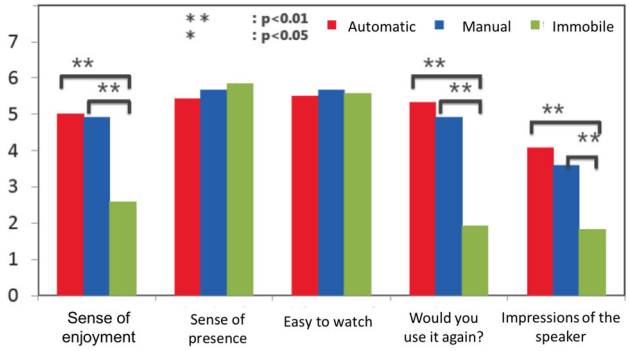


Fig. 11 Results from subjective evaluation

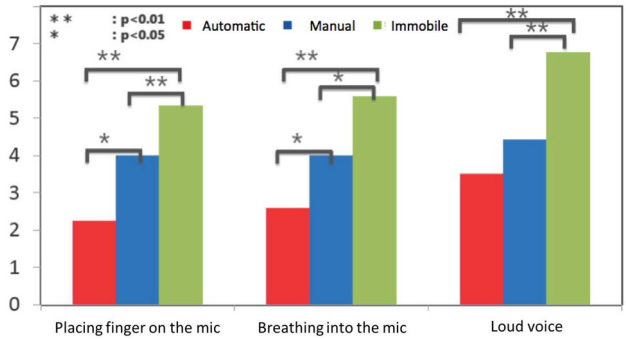


Fig. 12 Comparison of no. of stimuli remote users felt uncomfortable with

analyze the data and significant differences were observed at the 5% significance level in three items of levels of enjoyment, wanting to use it again, and impressions of the speaker. Multiple t-tests were then administered. The *t* test results indicated that significant differences could be observed at the 1% significance level in all groups with and without evasive behaviors.

Figure 12 shows the results on the level of discomfort due to uncomfortable stimuli. ANOVA indicated significant differences at the 5% significance level for all groups, and the multiple *t* test results indicated significant differences for the systems in the figure.

6.3 Discussion

The results in Fig. 11 indicate significant differences were observed between systems that engaged in avoidance behaviors and systems that did not. This demonstrates the significance of taking part in avoidance behaviors, which fits well with our hypothesis. The discomfort felt by the participants in the experiment when the telepresence robot automatically engaged in avoidance behaviors was a great deal lower than that of the manual avoidance action in two of the three items classified as uncomfortable stimuli, which was contrary to our initial hypothesis. It seems that one reason for the level of discomfort being lower when autonomous evasive behavior was used is that autonomous avoidance behavior preemptively took evasive action before the remote users felt uncomfortable stimuli, which is impossible in manual avoidance actions. In support of this claim, we found from the post-experiment interviews that multiple remote users were unaware that unpleasant stimuli had been presented to them.

There were no significant differences between **automatic** and **manual** for the "loud voice". There are two possible explanations for this. The first is that the presentation time was shorter than that for the other stimuli for "loud voice", so there were cases where the presentation of stimuli had been completed before avoidance behavior could be expressed. The second was that the learning accuracy of ESN for data other than prepared data set was somewhat less accurate than that of the other two stimuli.

A general question for this research is whether there is no problem even if the sound is completely erased. Since the robot takes avoidance action, communication is paused in most cases, and important information is not exchanged while the sound is stopped. Therefore it is considered that eliminating sound completely does not matter. On the other hand, with the approach to adjust the volume, it is difficult to solve the problem in this research. This is because the participants in the experiment are simply not concerned with the volume problem, but report the discomfort caused by the sound quality.

In this study, we focused on unpleasant stimulation given by local user. Therefore, we do not mention any unpleasant environmental sounds. If we focuse on environmental sounds, there is a high possibility that a different approach from this research will be effective. Because the approach of this study is considered to adapt local users themselves not to present unpleasant stimuli to remote users, but environmental sounds do not adapt to robots and remote users.

Experiment participants reported that making their evasive actions themselves was stressful as though they were rude to the opponent, but they did not feel stress when automatic avoidance behaviors were exposed. This is considered to be one of the effectiveness of this research approach.

This robot does not explain which stimulus was unpleasant. But local user can easily imagine which stimulus was unpleasant, because this robot reacts immediately after receiving an unpleasant stimulus.

This paper only focused on sound with reference to the results from the case studies. However, multimodal recognition including images and other sensor information may make it possible to handle discomfort more efficiently in the future.

The telepresence robot in this online experiment only learned the threshold of ABAM for the purpose of quickly adapting to the user. ESN's online learning in the future has the potential to be a system that can more flexibly adapt to users.

7 Conclusion

We created a system that could learn the unpleasant sound stimulation that arises when working with a telepresence robot and clarified by means of case studies the effect of autonomously expressed avoidance behaviors when such stimuli are presented. Experiments on evaluation demonstrated the significance of undertaking avoidance behaviors. In addition, the discomfort felt by remote users decreased when the robot autonomously expressed avoidance behaviors than when remotes themselves manually expressed them. Automatic avoidance behaviors could provide a telepresence environment with lower levels of discomfort, as was previously described. Our present version of the system dealt with categorized generic stimuli in advance.

(A-6) The engineering main contribution of this study is the suggestion of learning method for remotely controlled robot adaptable to human and its realization. A suggestion and realization of generic extensional functionality for remotely controlled robot is a major contribution since it can be applied to many other remotely controlled robots. The experimental results of sense of unpleasantness human feel under binaural microphone can be raised as the scientific contribution. We think that a quantification of unpleasant stimuli the user may experience when using remotely controlled robot with binaural microphone and suggestion of three major types of stimuli can be said to be a major scientific contribution.

(A-1) Finally, I will explain the limitations and future research.

As a premise, this research aims to develop a remote-controlled robot capable of rapidly adapting to human senses. To achieve such goal, we studied with focus to the easily implementable and evaluable "sudden appearance of unpleasant stimuli" first. Limitations are that it only discusses the suddenly appearing unpleasant sound stimuli and not on "how to encourage comfortable stimuli" nor "how to avoid continuous stimuli". Nevertheless, we believe that the topic of research, the suddenly appearing unpleasant stimuli, was a reasonable choice since despite the large effect that it has on the remote user, it can also be easily avoided by the suggested simple method in our study.

(A-1, A-2) In this study, we set the parameters on the premise that the response speed of the experiment participants is fast enough. Since all the experiment participants were in their twenties, we can not refer to the results when the elderly participate in the experiment. However, by applying appropriate parameter settings, there is a relatively high possibility that the proposed method can be adapted even if the experiment participants were elderly.

(A-1) We consider a development of similar system applicable to stimuli other than sudden unpleasantness as future research. Currently, we're thinking of realizing a method to naturally suggest the comfortable distance and volume for the remote user to the local user or to introduce more complex machine learning algorithm to response based on the local user's speech. All of them are a natural continuation of this study that puts emphasis on that the remote user is in a weaker position.

Acknowledgements This works was in part supported by JSPS KAKENHI (Grant No. 17J00580) and in part supported by MEXT KAKENHI (Grant No. 26118006).

Compliance with Ethical Standards

Conflict of interest The authors declare that they have no conflict of interest.

References

1. Nakanishi H, Murakami Y, Kato K (2009) In: Proceedings of the SIGCHI conference on human factors in computing systems. ACM, pp 433–442
2. De Greef P, Ijsselsteijn WA (2001) Social presence in a home tele-application. CyberPsychol Behav 4(2):307
3. Isaacs EA, Tang JC (1994) What video can and cannot do for collaboration: a case study. Multimed Syst 2(2):63
4. Jancke G, Venolia GD, Grudin J, Cadiz JJ, Gupta A (2011) In: Proceedings of the SIGCHI conference on Human factors in computing systems. ACM, pp 530–537
5. Fish RS, Kraut RE, Chalfonte BL (1990) In: Proceedings of the conference on computer-supported cooperative work. ACM, pp 1–11
6. Toshima I, Uematsu H, Hirahara T (2003) A steerable dummy head that tracks three-dimensional head movement: TeleHead. Acoust Sci Technol 24(5):327
7. Toshima I, Aoki S (2014) Perception of delay time of head movement in using an acoustical telepresence robot: TeleHead. Adv Robot 28(15):997
8. Jouppi NP (2002) In: Proceedings of the conference on computer supported cooperative work. ACM, pp 354–363
9. Prussog A, Mühlbach L, Böcker M (1994) In: Proceedings of the human factors and ergonomics society annual meeting, vol 38. SAGE Publications Sage CA, Los Angeles, CA, pp 180–184
10. Fernando CL, Furukawa M, Kurogi T, Kamuro S, Minamizawa K, Tachi S, et al (2012) In: IEEE international conference on intelligent robots and systems, pp 5112–5118
11. Kristoffersson A, Coradeschi S, Loutfi A (2013) A review of mobile robotic telepresence. Adv Hum Comput Interact 2013:3
12. Paulos E, Canny J (1998) In: IEEE international conference on robotics and automation, vol 4, pp 3173–3178
13. Paulos E, Canny J (2001) Social tele-embodiment: understanding presence. Auton Robots 11(1):87
14. Tsui KM, Desai M, Yanco HA, Uhlik C (2011) In: Proceedings of the international conference on Human–robot interaction. ACM, pp 11–18

15. Adalgeirsson SO, Breazeal C (2010) In: Proceedings of the international conference on Human–robot interaction. IEEE Press, Piscataway, pp 15–22
16. Sakamoto D, Kanda T, Ono T, Ishiguro H, Hagita N (2007) In: IEEE proceedings of the international conference on Human–robot interaction, pp 193–200
17. Ogawa K, Nishio S, Koda K, Balistreri G, Watanabe T, Ishiguro H (2011) Exploring the natural reaction of young and aged person with telenoid in a real world. JACIII 15(5):592
18. Yamazaki R, Nishio S, Ogawa K, Ishiguro H, Matsumura K, Koda K, Fujinami T (2012) In: Extended abstracts on human factors in computing systems. ACM, pp 351–366
19. Tobita H, Maruyama S, Kuzi T (2011) In: Extended abstracts on human factors in computing systems. ACM, pp 541–550
20. Lu JM, Lu C, Chen Y, Wang J, Hsu Y et al (2011) In: Proceedings of the Asia Pacific eCare and TeleCare Congress
21. Takahashi M, Ogata M, Imai M, Nakamura K, Nakadai K (2015) In: IEEE international symposium on robot and human interactive communication (RO-MAN), pp 517–522
22. Hayamizu A, Imai M, Nakamura K, Nakadai K (2014) In: Proceedings of the international conference on Human–agent interaction. ACM, pp 67–74
23. Theofilis K, Orlosky J, Nagai Y, Kiyokawa K (2016) In: IEEE international conference on humanoid robots (humanoids), pp 242–248
24. Hasegawa K, Nakauchi Y (2013) In: IEEE international symposium on robot and human interactive communication (RO-MAN), pp 350–351
25. Hasegawa K, Nakauchi Y (2014) In: Proceedings of the second international conference on human–agent interaction, pp 29–31
26. Hasegawa K, Nakauchi Y (2014) In: Proceedings of the international conference on human–agent interaction. ACM, pp 293–296
27. Hasegawa K, Nakauchi Y (2014) Telepresence robot that exaggerates non-verbal cues for taking turns in multi-party teleconferences. Trans Jpn Soc Mech Eng 80(819):DR0321 (in Japanese)
28. Matsui D, Minato T, MacDorman KF, Ishiguro H (2005) In: IEEE international conference on intelligent robots and systems, pp 3301–3308
29. Funakoshi K, Kobayashi K, Nakano M, Yamada S, Kitamura Y, Tsujino H (2008) In: Proceedings of the international conference on multimodal interfaces. ACM, pp 293–296
30. Coltin B, Biswas J, Pomerleau D, Veloso M (2011) In: Robot soccer world cup. Springer, pp 365–376
31. Tsui KM, Norton A, Brooks DJ, McCann E, Medvedev MS, Allspaw J, Suksawat S, Dalphond JM, Lunderville M, Yanco HA (2014) Iterative design of a semi-autonomous social telepresence robot research platform: a chronology. Intell Serv Robot 7(2):103
32. Fürler L, Nagrath V, Malik AS, Meriaudeau F (2013) In: IEEE international conference on communication systems and network technologies (CSNT), pp 262–267
33. Tanaka K, Yamashita N, Nakanishi H, Ishiguro H (2016) In: Proceedings of the international conference on human–robot interaction. IEEE Press, pp 133–140
34. Choi M, Kornfield R, Takayama L, Mutlu B (2017) In: Proceedings of the CHI conference on human factors in computing systems. ACM, pp 325–335
35. Okuoka K, Takimoto Y, Osawa M, Imai M (2018) In: Proceedings of the 6th international conference on human–agent interaction. ACM, HAI '18, pp 167–175
36. Seno T, Okuoka K, Osawa M, Imai M (2018) In: Proceedings of the 6th international conference on human–agent interaction. ACM, HAI '18, pp 377–379
37. Obo T, Sawayama T, Taniguchi K, Kubota N (2012) In: IEEE world automation congress (WAC), pp 1–6
38. Tapus A, Tapus C, Matarić M (2010) In: Field and service robotics. Springer, pp 389–398
39. Fukuda O, Tsuji T, Kaneko M, Otsuka A (2003) A human-assisting manipulator teleoperated by EMG signals and arm motions. IEEE Trans Robot Autom 19(2):210
40. Ghahramani A, Tang C, Becerik-Gerber B (2015) An online learning approach for quantifying personalized thermal comfort via adaptive stochastic modeling. Build Environ 92:86
41. Elman JL (1990) Finding structure in time. Cogn Sci 14(2):179
42. Jaeger H (2001) Bonn, Germany: German National Research Center for Information Technology GMD Technical Report 148(34):13
43. Lukoševičius M (2012) In: Neural networks: tricks of the trade. Springer, pp 659–686
44. Yamagishi Y, Osawa M, Hagiwara M (2015) In: International symposium on advanced intelligent systems
45. Gers FA, Schmidhuber J, Cummins F (2000) Learning to forget: continual prediction with LSTM. Neural Comput 12(10):2451
46. Osawa M, Yamakawa H, Imai M (2016) In: International conference on neural information processing. Springer, pp 342–350
47. Osawa M, Ashihara Y, Seno T, Imai M, Kurihara S (2017) In: The international conference on neural information processing
48. Komatsu T, Yamada S (2007) In: Extended abstracts on human factors in computing systems. ACM, pp 2519–2524

Masahiko Osawa is a PhD student of Graduate school of science and technology at Keio university and Research fellow (DC1) at The Japan Society of the Promotion of Science. He received his Master of Engineering degree from Keio Univ. In 2017. His research interests include machine learning, autonomous robots, human-agent interaction, cognitive science, biologically inspired cognitive architecture, and computational neuro science. He is a member of the the Japanese Society for Artificial Intelligence, Japanese Neural Network Society, Japanese Cognitive Science Society, Asia Pacific Neural Network Assembly, and ACM. His dream is making DORAEMON.

Michita Imai is a Professor of Faculty of science and technology at Keio university and a Researcher at ATR Intelligent Robot Laboratories. He received his Ph.D. degree in Computer Science from Keio Univ. in 2002. In 1994, he joined NTT Human Interface Laboratories. He joined the ATR Media Integration & Communications Research Laboratories in 1997. He was a visiting scholar of University of Chicago from 2009–2010. His research interests include autonomous robots, human-robot interaction, speech dialogue systems, humanoid, and spontaneous behaviors. He is a member of Information and Communication Engineers Japan (IEICE-J), the Information Processing Society of Japan, the Japanese Cognitive Science Society, the Japanese Society for Artificial Intelligence, Human Interface Society, IEEE, and ACM.