Abstract—The study of the telepresence robot as a tool for telecommunication from a remote location is attracting considerable attention. However, it does not consider varying conditions in the environment, such as noise. In addition, when talking with several people in remote location, the user would like to be able to change the speaker volume freely according to the situation. In a previous study, a telepresence robot was proposed that has a function that automatically regulates the volume of the user’s utterance. However, the manner in which the user exploits this function in a practical situation needs to be investigated. We propose a telepresence conversation robot system called “TeleCoBot.” TeleCoBot includes an operator’s user interface, through which the volume of the user’s utterance can be automatically regulated according to the distance between the robot and the conversation partner and the noise level in the remote environment. We conducted a case study, in which the participants played a game using TeleCoBot’s interface. The results of the study reveal the manner in which the participants used TeleCoBot and the additional factors that the system requires.

I. INTRODUCTION

The demand for video conference systems, such as Skype and Polycom, that allow users to communicate with people in remote locations has grown. These commercial systems enable the user to see the remote partners’ face and hear their voice. However, they do not give the user a sense of conversation partner’s existence or presence. To solve this problem, researchers have developed telepresence systems that include robot. As part of a remote communication system, a telepresence robot, which the user controls remotely, can convey a sense of the user’s presence to the people with whom s/he is communicating. Some examples of a telepresence robot have been implemented, such as Texai and Giraffe [10, 11].

Existing telepresence systems have not, however, addressed the problem related to the auditory scene in telecommunications that the noise in a locality has to be compensated. BeamPro [1], a commercial telepresence robot, has six microphone arrays and moderates the influence of echo and noise. BeamPro can output the user’s voice at a constant quality. However, it cannot deal satisfactorily with the following two cases. The first case involves telecommunication between locations in different time zones where day and night do not occur at the same time, such as between Japan and Los Angeles. In this case, if the telepresence system transmits the user’s voice without adapting its volume, her/his voice will be too loud/quiet for her/his partner, and this inappropriate volume may cause uncomfortable communication. The second case is where people are present near the conversation partner. In this case, the user may want to be able to change the volume of her/his own voice freely according to the content of the conversation, such as a secret subject or an important message.

To resolve these problems, we propose a telepresence robot system named TeleCoBot(Telepresence Conversation robot). TeleCoBot has a function that automatically regulates the volume of the user’s utterance according to the robot’s distance from the conversation partner and the noise level in the partner’s location. TeleCoBot’s interface allows the user to control the volume of her/his own voice (Figure 1). Figure 1 shows the example of using TeleCoBot.

![Figure 1. TeleCoBot](image)

(a) Comfortable mode

(b) Secret mode

We conducted a case study in an actual environment to investigate TeleCoBot’s usability. In the case study, we observed under what conditions a user changes the volume using the above function of TeleCoBot.

This paper is organized as follows. In Section II, we explain related work to clarify the target of TeleCoBot. In Sections III and IV, we describe the structure and the functions of TeleCoBot. In Section V, a case study to investigate the usability of TeleCoBot is presented, and Section VI concludes the study and describes future work.
II. RELATED WORK

Papcke et al. [2] focused on the mediation by a telepresence robot of the user’s speaking volume in telecommunication and applied a sidetone to the telepresence system. The sidetone constitutes the user’s voice fed back to her/his own ear. The sidetone causes a user to pay attention to her/his own speaking volume, and s/he naturally reduces her/his speaking volume. However, this system cannot guarantee that the adjusted voice volume is appropriate for the conversation partner’s location. Therefore, it does not sufficiently support voice volume control.

Martinson et al. [3] implemented a volume control module in a mobile robot, and conducted experiments to investigate the performance of the method. The mobile robot was capable of varying its speaking volume according to the surrounding noise level and its distance from a conversation partner. Their method has this feature in common with our study. However, whereas Martinson et al.’s method focused on changing the sound volume of a mobile robot, our research focuses on a telepresence robot. In the case of a telepresence robot, the user has to set parameters for changing the volume, such as the choice of conversation partner and the speaking volume. The system should set up an operation on its UI. In addition, the mobile robot cannot behave appropriately according to the type of conversation, for example, a confidential one, because the volume adjustment of the mobile robot is designed for maintaining speech intelligibility.

Hayamizu et al. [4] proposed a telepresence robot named LombaBot in which an optimal volume control system is incorporated. To obtain the optimal volume model, in their research two types of sound volume were investigated, a comfortable volume that does not cause the conversation partner to feel uncomfortable and a secret volume that is equal to the minimum volume that a conversation partner can hear. These volumes were investigated experimentally by changing the noise level and the distance between a person and the loudspeaker. A model formulated according to these results allows automatic volume control. In this research, however, only the performance of the volume control was evaluated, and not its usability in telecommunication. In this paper, we investigate how people use a volume control interface in telecommunication.

III. TeleCoBot

A. Proposed System

In this paper, we propose a telepresence system, “TeleCoBot,” that automatically changes the user’s speaking volume in the locality of the conversation partner. We used the robot audition software HARK [5] for audio input/output and volume control.

In this study, we used the following two modes for supporting automatic volume control of TeleCoBot. One is a mode that allows the conversation partner to feel comfortable; we call this the comfortable mode. TeleCoBot’s volume parameter is determined on the basis of the noise volume and the distance between the robot and the conversation partner. In a previous study, an experiment was conducted to measure the optimal speaker volume for various conditions of distances and noise levels [4], and the optimal volume model was found. We applied this model in TeleCoBot. The second mode is one that supports user’s wish to use a volume that only the conversation partner can hear; we call this the secret mode. In the previous study, minimum speaker volume that allows a listener to understand words under various conditions of distance and noise level was measured [4]. We used this volume as the secret volume.

In TeleCoBot, a graphical user interface (GUI) for the two modes is implemented, which allows automatic volume control. The TeleCoBot robot is equipped with a Kinect sensor [6], and therefore, it can detect people. The operator can determine the conversation partner by choosing from the detected people displayed on the GUI. The operator can switch between the two modes, comfortable and secret, on the GUI.

TeleCoBot ideally enables a remote user to change her/his speaking volume as if s/he were talking with the partner in the same location. For example, when the user attempts to give important messages, s/he has to switch to comfortable mode. Then, the user’s voice reaches not only the conversation partner but also others. When the remote user wants to talk with the conversation partner in a whisper, the user has to switch to the secret mode. Then, the user’s voice reaches only the partner, and others cannot hear it. In this research, we investigated whether in fact the user changes the speaking volume by using TeleCoBot’s interface in the same manner as when talking with a partner in the same location.

B. Overview of TeleCoBot

Figure 2 shows a system overview of TeleCoBot. Remote means a robot user or a remote place where the user is located, and Local means a conversation partner or a place where the robot and the conversation partner are located. Sound signal processing is implemented by HARK. We use the Robot Operating System (ROS) [7] for data acquisition from Kinect and for data communication. The user can select the volume mode (comfortable or secret mode) using the GUI. The noise level in the remote place is measured using microphones installed on robot. Using Kinect sensor with which the robot is equipped, the distance between the robot and the conversation partner is measured. TeleCoBot applies the selected conversation mode, the noise level, and the distance to the optimal volume model, and then changes the user’s speaking volume.
IV. IMPLEMENTATION OF TELECOBOT

A. Devices for TeleCoBot

The TeleCoBot operator uses a laptop computer for interacting with the user, a camera set on a work desk for streaming video in the remote place, a desk-top USB microphone for recording the user’s voice, and headphones for hearing local sound. As shown in Figure 3, a TurtleBot [12] is installed; we call it the robotic avatar of TeleCoBot. The devices with which it is equipped are: an Xbox 360 Kinect for detecting persons and measuring the distances to the detected persons, a microphone for recording sound and measuring the noise level in the conversation partner’s location, a display for video streaming from the remote location, a loudspeaker for outputting the remote user’s voice, and a switch for mode selection. The person who presses this switch, that is, the person nearest to the robot, can select the comfortable or the secret mode and the volume gain is calculated according to distance between her/him and the robot. The display attached to the robot also shows the current volume mode (comfortable or secret). The person can perceive the current mode by watching the display.

![Figure 3. Robotic Avatar of TeleCoBot](image)

B. Robot Operating System

We built the programs for TeleCoBot over ROS. One basic function provided by ROS is data communication among nodes. We developed nodes related to the GUI, sound processing, and image processing; these nodes can communicate with each other through topic publish/subscribe. In addition, ROS prepares libraries corresponding to various software programs. We used open sources, such as HARK and OpenNI. ROS prepares its own libraries for these software programs, and thus, we could easily implement data communication over ROS.

C. Optimal volume control

The user’s voice output from the robot is determined by the selected volume mode. As mentioned previously, two volume modes are provided by TeleCoBot; a comfortable volume, such that the conversation partner can hear comfortably, and a secret volume such that the user’s voice can be understood only within the distance between the conversation partner and the robot. The parameters required for determining the speaking volume are the noise level in the location of the conversation partner and the robot, and the distance between the robot and the partner. In a previous study, the optimal volumes for the comfortable and secret mode were determined experimentally [4]. In the experiment, the participants listened to voices of various volumes from a speaker under conditions of various noise levels and distances from the speaker. After the experiment, the participants answered questions as to whether the volume was comfortable, and whether they understand the words from the speaker. The results showed the optimal voice volume according to distance and noise level. We used the data, shown in Figure 4, as the optimal volume model. We use minimum volume shown in figure 4 as a secret volume. We employed the sound volume models proposed in [4].

![Figure 4. Optimal speaking volume per distance and noise](image)

D. Graphical User interface

Figure 5 shows the graphical interface. On the GUI, a black circular frame is shown at the position of the people detected by the Kinect. When the user clicks with the mouse button on the frame, TeleCoBot begins to perform volume control according to the distance between the robot and the person corresponding to the selected frame. If the user clicks the left button, the robot’s speaking volume is set to the comfortable mode and the user’s voice reaches throughout. If the user clicks the right button, the robot’s speaking volume is set to the secret mode and people at a distance from the robot and the partner cannot understand the user’s talk. The circular frame’s color is changed to red by a left click and yellow by a right click. The user easily perceives the current mode by watching the circular color.

We used a Kinect sensor for human detection and OpenNI [8] for controlling it. When people are detected by the Kinect sensor, their coordinates are sent to the GUI. In addition, the Kinect sensor sends an image, and therefore, the GUI can draw circular frames around the people on streaming data in the conversation partners’ location, and the user can select a specific conversation partner.
E. Volume Control Module

Figure 6 shows an overview of the volume control module based on HARK. The volume gain of the user’s voice output from the robot speaker is determined by three parameters: the distance between the robot and the conversation partner, the noise level in the locality, and the selected conversation mode. VAD stands for voice activity detection. After the user’s voice is detected by the VAD module, its power is normalized. This process reduces the influence of individual differences in speaking volume. If the user does not select a conversation partner, TeleCoBot transmits the voice without controlling its volume. In this case, the louder the user’s voice becomes, the louder are the robot speaker’s outputs.

![Overview of volume control module](image)

Figure 6. Overview of volume control module

V. Case Study

To investigate the usability of TeleCoBot in an actual situation using a telepresence system, we recruited participants to use TeleCoBot when playing a game. In this case study, we investigated what situation the participant used mode selection, and whether the participant selected the appropriate sound volume mode. The participants were three male graduate students aged in their twenties and novices of TeleCoBot. One participant operated TeleCoBot from the GUI, and the other two participants were located with the robot separated from the operator.

A. Study Environment

As shown in Figure 7, the two participants sat 0.5 m, and 3.0 meter away from the robot, respectively. We emitted a noise at 55dB (A) from the loudspeakers located in each of the four corners of the room. This arrangement reduced the influence on the measured noise level even if the robot microphone caught the robot’s speech and the participants’ voices. As the ambient noise, we employed a sound recorded in department store [9]. The remote user operated the robot from another room.

![Study environment](image)

Figure 7. Study environment

B. Details of game

We prepared a word guessing game that involved graphical drawing. This game is called Pictionary. Participants of Pictionary are divided into teams. The role of one team is to choose a theme and draw pictures in accordance with the theme. The members of the other teams try to identify the specific words. In our case study, the participants were required to play this game while situated in separate locations using TeleCoBot. The roles of the three participants were: choosing a word from a theme list and telling the word to the participant whose role it was to draw, drawing a picture in accordance with the word, and guessing the word from the drawing. We called the above three roles questioner, drawer, and answerer, respectively.

We established rules for this game, as follows. 1. Do not write words as part of the drawing. 2. Do not say a correct answer or keywords out loud. 3. The time limit is two minutes per theme. 4. The answerer has to give the correct word, not a synonym. 5. The answerer has to continue giving answer until the allotted time has expired or he has answered correctly. If the answerer guesses the correct word, the questioner tells him “Your answer is correct.” If the answerer cannot answer the correct word, the questioner tells the selected theme to the answerer.

Two patterns of the Pictionary game were held in which the roles were rotated. The assignments of the first pattern were as follows: A remote user was the questioner, a participant sitting near the robot (Participant-A) was the drawer, and the participant sitting far from the robot (Participant-B) was the answerer. In the first pattern, the
remote user told the theme to Participant-A using TeleCoBot, then Participant-A drew pictures of the theme on a white board, and Participant-B looked at the drawing and gave answers. When Participant-B guessed the word correctly or the allotted time had expired, the remote user had to inform Participant-B.

The assignments of the second pattern were as follow. The remote user was the drawer, Participant-A was the questioner, and Participant-B was the answerer. Participant-A communicated the theme to the remote user by whispering it to the robot microphone, and then the remote user drew pictures of it on paper by hand. The drawing activity was streamed from the camera located on the work desk, and the robot displayed the streaming video. Participant-B watched video on the robot’s display, and gave answers. When Participant-B guessed the word correctly or the allotted time had expired, Participant-A had to inform Participant-B.

C. Study Results

After the game, the participants answered questionnaires related to the usability of TeleCoBot and their thoughts about remote communication using TeleCoBot. The experiment was recorded by cameras installed near the robot, near the Participant-B, and behind the remote user. From these data sources, we obtained the results of the study.

First, we evaluate the performance of volume control. Participant-B’s response about the first pattern was that he (as the answerer) was sometimes aware of a sound when the user (as the questioner) said “Now, I will tell you a theme”, however, he could not identify the sound. At that time, Participant-A could hear the user’s message and draw the picture. Therefore, the volume of the secret mode was adequately controlled. The sounds of the beginning of the game in the first pattern can be heard in the video taken by the camera near the Participant-B. Therefore, the volume of the comfortable mode was sufficient to reach a long distance from the robot. From these results, we conclude that the volume control function was effective.

Next, we evaluate the usage of TeleCoBot. In the first pattern, where the remote user selected the words, we observed the following actions when he operated the GUI: 1: whether he switched from comfortable mode to secret mode when telling the theme, 2: whether he switched back from secret mode to comfortable mode. We observed the first action; however, when the user said the correct word or that the allotted time had expired he did not change back to the comfortable mode, and then Participant-A told the correct word to Participant-B. This situation resembled “Chinese whispers.” Therefore, we could not observe the second action.

In the second pattern, where Participant-A selected a word, we expected the remote user to switch to the secret mode, because even if he said the word and gave hints unconsciously or accidentally, his voice could be heard only by Participant-A and to do so was therefore all right. However, the remote user did not change the mode. We also expected Participant-A to push the switch installed on the robot and change to the secret mode. If Participant-A had switched to the secret mode when the remote user was drawing, an accident, whereby Participant-B heard the correct answer, may not have occurred. However, Participant-A did not change the mode.

VI. DISCUSSION

In the first pattern, the remote user continued to play the game leaving the speaking volume in secret mode. This problem indicates that the user of TeleCoBot did not understand the effect of changing the volume. The user forgot to change his voice volume. However, similar problems occur not only in remote communications, but also in face-to-face communications. If people are absorbed in a face-to-face conversation, their voices will become louder in any situation. When they become angry or, excited or when their conversation becomes livelier, their voices may also become louder. However, when people play Pictionary in the same place, they do not tell the correct answer in a whisper. Therefore, we think this problem is due to that the fact that it is difficult for the user to comprehend an auditory scene in another place. In other words, the user could not understand how loud his voice was in the other place where the conversation partners and the robot were located.

The TeleCoBot operating UI suffers from the problem that the user cannot change her/his voice volume using an intuitive method. When people using a personal computer want to change the volume, they slide the bar of the volume mixer. In face-to-face conversation, people vibrate their vocal cords strongly or weakly according to conversation situations. In the operation of TeleCoBot, the user has to switch the volume of the robot speaker by clicking the left or right mouse button; however, this operation is unusual. This inexpedient volume operation made the user think that they could play the game while remaining in the secret mode.

In the second pattern, the user and participant sitting near the robot did not change the volume. This means they could not suppose the user said the correct answer unintentionally and there is a possibility that they did not understand the Pictionary game well.

There are several problems that should be addressed in future studies. The first concerns the improvement of the GUI. If the system provides a slide bar for changing a volume, the user might be able to adjust the volume more intuitively. We feel the telepresence system should improve the sense of presence for not only the people talking with the robot but also the user who control the robot. For example, the GUI shows a sound pressure map overlapping or aligning with a video streaming. If the GUI shows the map to indicate where people have a conversation, the map prompts the user to change the volume mode according to the conversational situation.

The second problem concerns the selection of conversation partner. A user can select the conversation partner from people who is some distance from the robot; however s/he cannot select a partner from people who are situated side by side because TeleCoBot only serves the function of volume control. To resolve this problem, the robot avatar for TeleCoBot should be equipped with a directional speaker or a movement mechanism to keep a distance for a conversation.
VII. CONCLUSION

In this paper, we proposed TeleCoBot and described a case study in which the usability of the UI of a proposed telepresence system in which an automatic volume control function was applied was investigated. A TeleCoBot user can use her/his voice in the comfortable or secret mode without being concerned about the auditory environment in the conversation partner’s location. The case study involved a guessing word game using TeleCoBot. The volume control functioned well. However, the user could not change the volume as in face-to-face communication during the interaction. The results of the case study indicate that the TeleCoBot UI should be improved based on the case study.

REFERENCES