

Research Article

An Ethnomethodological Study of a Museum Guide Robot's Attempt at Engagement and Disengagement

Madhumita Ghosh and Hideaki Kuzuoka

Department of Intelligent Interaction Technologies, Graduate School of Systems and Information Engineering, University of Tsukuba, 1-1-1 Tennodai, Tsukuba, Ibaraki 305-0006, Japan

Correspondence should be addressed to Madhumita Ghosh; madhumita2ghosh@gmail.com

Received 25 May 2014; Revised 13 September 2014; Accepted 17 September 2014; Published 19 October 2014

Academic Editor: Giovanni Muscato

Copyright © 2014 M. Ghosh and H. Kuzuoka. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

We conducted a study of a museum guide robot's attempt at engaging and disengaging the audience at predetermined points in time during a guided tour. We used "ethnomethodology" as a tool for our study and analysis. In this paper, we describe how we developed, tested, and analyzed a museum guide robot system that borrows cues from social scientists to manage an audience. We have described how we began our study, the previous studies that we referred to, the initial attempts to test our concept, the development of the system, the real-world experiments, and the analysis of the data that we collected. We have described the tools of engagement and disengagement that the robot has used and presented the results of our statistical analysis of the experimental data. Most prominently, we found that a verbal gesture called "summative assessment" and a nonverbal gesture called the "lean-back" gesture are very effective as tools of disengagement. These tools help a robot guide to manage the audience in the same way as a human guide. Moreover, we found that a combination of the aforementioned two gestures is more effective than employing them separately.

1. Introduction

Our research seeks to make a contribution to the field of human-robot interaction (HRI). In this field, scientists study the reactions of human participants who engage in a social interaction with robots. Different research teams have studied human-robot interactions in different contexts. Some have chosen the classroom [1] or the school corridors [2], and some have chosen a shopping mall [3], while others have chosen a train station [4] environment and so on. There are a large number of studies conducted in different environments to understand what human users expect from their robot companions or guides. These studies usually target different areas of improvement in machine learning, artificial intelligence (AI), and robot behavior programming. Such improvements make the robot not only aesthetically more pleasing, approachable, friendly, and easy-to-use, but also a future commercial success.

Our lab has been studying human-robot interaction (HRI) in museum environments. Over several years, different

researchers in our lab have studied different aspects of interaction and communication in a museum environment and tried to improve visitor approval, enjoyment, and entertainment level. While some have attempted to draw attention by employing a common human communication impairment [5], others have tried to understand the effect of a robot's physical orientation on its audience [6]. At the core of such interactions lies the endeavor to pass information or educate the audience on the various exhibits. Even social scientists have studied such context-specific communication in great detail [7, 8].

Over the years, there have been many interesting researches with robot museum guides [9–12] and human-robot interaction [13, 14] in general. Our research uses a simpler robot designed specifically to study human-robot interaction. Our main focus of study is the reactions of the human participants to the robots speech and physical gestures. Our robot platform is described in Section 2.7.

We have observed during these previous studies that when people interact with robots, they do not always feel

compelled to stay till the end of the interaction. For example, if the robot tries to deliver a lengthy explanation of an exhibit, the audience might move on without staying till the end of the explanation. As a result, we learned that people do not feel compelled to show the same degree of politeness and respect to a robot guide as they do in case of a human guide.

To address the situation described above, we decided to attempt to make a robot not only engage the audience in an interesting explanation of the exhibits so that the audience feels intrigued and feels compelled to stay till the end of the explanation, but also try to equip the robot such that it can disengage the audience on time so that the current group would move on and a fresh set of people can listen to the explanation. If the robot is stationary [15], this is more difficult to achieve since people sometimes feel curious about the robot's external appearance and want to touch it or try to further interact with it. In case of a normal guided tour, it is only natural for the guide to take its audience through a predetermined course and leave them at the end of the course so that the audience can move on to the next room and the guide can take on the next set of audience members. Therefore, in our current study, the robot guide is mobile and takes the visitors on a short tour of three particular exhibits.

We employed ethnomethodology as our method of study. In this method, human behavior is observed closely to identify nonverbal cues about the subject's thoughts at a given instance. This method is often employed by behavioral psychologists to study human-human communication. In this method, the changes in the subject's behavior are observed at definite time intervals to understand how changes in the immediate environment can cause changes in the subject's behavior. A detailed explanation of ethnomethodological ideas used in this research is given in Section 2.1 of this paper. We collaborated with social scientists to understand how ethnomethodology may be applied to our study. Our collaborators employ ethnomethodology to study the interaction between people and machines or technology in general. This inspired us to use ethnomethodology to study the interaction between our museum guide robot and visitors.

We have analyzed engagement through a questionnaire administered to all the subjects at the end of the experiment. We have also analyzed disengagement by measuring the time taken by each subject to disengage from each exhibit. We employed two specific tools of disengagement: a verbal cue, "summative assessment," and a nonverbal cue, "lean-back gesture." We found that a combination of these two tools is more effective than using them separately.

This paper is organized in the following manner: we have described the flow of our research, the tools that we used, and the experimental setup in the Materials and Methods section; we have presented the results and their implications in the Results and Discussion section; we have presented our conclusions and contribution to the HRI field in the Conclusion section; we have expressed our thanks and gratitude to everybody involved with this research in the Acknowledgments section; and finally, we have appended the complete bibliography for our research in the References section. Our research comprised the following sequence of activities: conducting an ethnomethodological study (Section 2.1),

conducting a field study (Section 2.2), analyzing the results of the field study (Section 2.3), developing the script and speech for the robot (Section 2.4), conducting controlled lab experiments (Section 2.5), analyzing the results of the controlled lab experiments (Section 2.6), and finally the real-world experiment (Section 2.7) that is the focus of this paper.

2. Materials and Methods

2.1. Ethnomethodological Study. Garfinkel has described ethnomethodology as follows in his book [16] on the subject: "*the following studies seek to treat practical activities, practical circumstances, and practical sociological reasoning as topics of empirical study, and by paying to the most commonplace activities of daily life the attention usually accorded extraordinary events, seek to learn about them as phenomena in their own right.*"

We decided to start by studying how human museum guides go about their work. Surely, through years of experience, guides develop their techniques to attract and hold the visitors' attention while conducting a tour. For this purpose, we collaborated with a team of researchers from England.

To conduct our ethnographic study, we collaborated with a group of social scientists who have been studying interaction between people and machines for decades. Through our collaboration, we arrived at a wealth of literature that describes human reactions to technology and how they prefer to interact with it in different contexts [17, 18]. Through our exchanges with our collaborators, we discovered different elements of human-human communication that can be applied to human-robot interaction. We used this wealth of knowledge as the base of our current study.

Our collaborators have done considerable work on studying humans at work [8]. Particularly, one team member has devoted her research efforts to studying human museum guides [19]. Through our exchanges with this team, we learnt that guides employ various verbal and nonverbal gestures to manage and guide the audience during a tour. Different gestures are used to accomplish different purposes like, a topic shift, a focus shift, or a subject shift. A subject shift is required when the guide has to explain a different facet of the same exhibit. A focus shift happens when the guide tries to draw the audience's attention to a different part of the same exhibit. An object shift happens when the guide introduces a completely different exhibit. A few examples of verbal gestures are as follows: a guide may pause for a second and then say "now..." in order to create a focus shift, the guide may pause and restart a sentence to draw extra attention by saying "This...this over here...", or a guide might put extra emphasis on startling pieces of numerical data in order to draw a gasp of admiration by saying something like "This chandelier is made up of thirteen thousand pieces..." saying the words "thirteen thousand" very slowly and with great emphasis and so on. Nonverbal gestures might include the following: the guide might step back after revealing an interesting piece of information so that the audience can lean forward and take a closer look at the exhibit; the guide usually starts stepping backwards and moving towards the next exhibit

even before he finishes the last sentence while wrapping up his talk on the current exhibit. This last nonverbal gesture signals the audience to finish looking at the current exhibit and start moving towards the next exhibit. Generally, human guides follow a particular flow of events during a tour. The guide begins the explanation of an exhibit by mentioning an interesting fact about the exhibit to draw the audience's interest. The initial statement always involves the exhibit as a whole. This is followed by a set of specific details about the exhibit, or objective statements. The explanation ends with a summative assessment that is also a subjective statement. Summative assessments are accompanied by summative gestures like circling the whole exhibit with the hand and so on. Then before the last sentence is over the guide leans back and begins orienting towards the next exhibit. If the next exhibit is too far from the current location, the guide uses metatalk to guide the audience and clears up the path to create a clear audience trajectory to the next exhibit in the tour.

From the above ethnographic study, we learned about the way human guides use various verbal and nonverbal gestures to control the engagement and disengagement of the audience with an exhibit. We developed our robot's disengagement behavior based on the above study. After scouring through many verbal and nonverbal gestures, we decided to incorporate the verbal gesture, "summative assessment" and the nonverbal gesture, "lean-back," as our robot's disengagement tools. These tools appealed to us because of their simplicity and because it is possible to easily incorporate them in our robot platform.

We conducted the ethnomethodological study for this research by analyzing the subjects' body language second-by-second to understand whether the subjects are engaged or disengaged. This required that we study the video recording of the experiment by combing through it repeatedly to look for any signs of loss of interest or abandonment behavior. As described in this paper, we considered eye contact and physical orientation as the main signs of engagement and disengagement. After conducting the ethnomethodological study we found that the subjects showed sustained interest in the interactions with the robot and displayed disengagement behaviors at very distinct points in the interaction. The engagement was proven by sustained eye contact for the duration of the robot's explanation of the exhibits. The disengagement behavior was observed during different robot body movements and mobility. The ethnomethodological idea used in our research is inspired by the second-by-second analysis method used to study the work of tour guides in museums and art galleries [19]. A sample of our ethnomethodological study of human museum guides is given in Section 2.3 of this paper.

2.2. Field Study. After we gained some initial experience, we launched a field study at the Science Museum in Tokyo. We had a professional museum guide conduct a tour of three selected exhibits. We selected an exhibit hall displaying the history of bicycles starting from the earliest bicycles ever built. We conducted four trials of the tour.

We requested the guide to prepare a short explanation based on the findings of our English collaborators. Based on our instructions, the guide prepared the explanation of each exhibit as follows: starting statement—3 objective statements—summative assessment. We captured the trials on video.

2.3. Results of Field Study. We analyzed the video data and performed a sample ethnomethodological analysis. We present here a part of the sample:

00:00 Guide points to the first exhibit and launches into the explanation of the first exhibit. His gaze is directed at the audience. "This is a bicycle built 194 years ago."

00:04 Guide withdraws his hand

00:06 Guide looks at the exhibit and turns his gaze back at the audience.

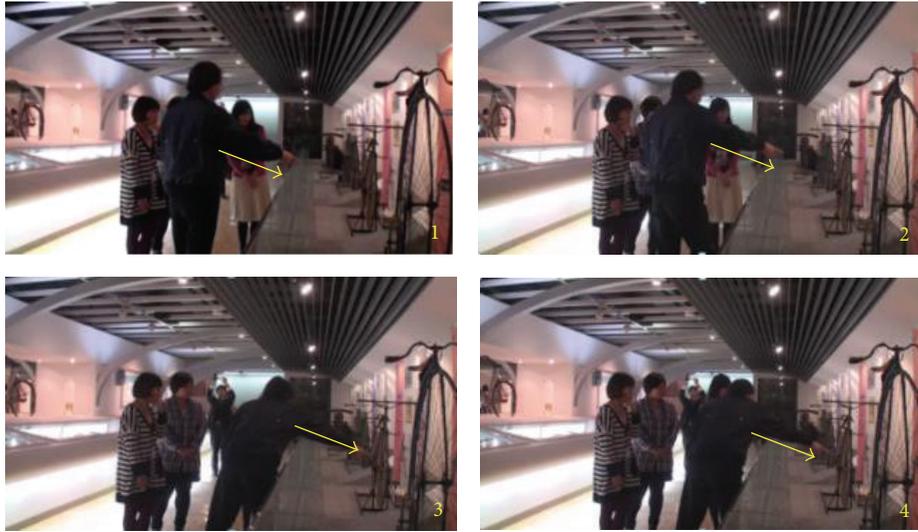
00:07 Guide points to the exhibit again and maintains eye contact with the audience while continuing with the subjective explanation.

00:08 Guide takes a step towards the exhibit and makes a square gesture with his right hand to indicate the whole bicycle. "The point that differs from modern bicycles..."

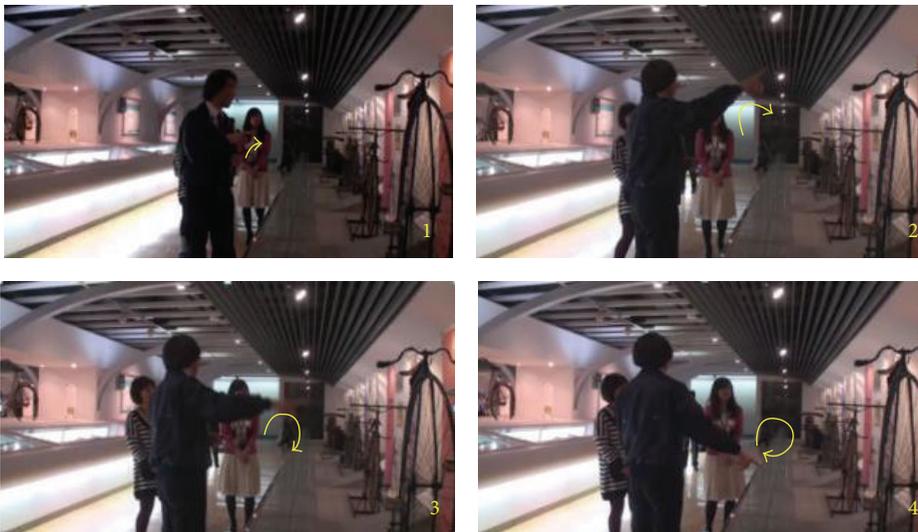
The above portion of the sample analysis was taken from the explanation of the first exhibit during the third trial. Here we analyzed the order of actions performed by the professional guide. Similar ethnomethodological analysis can be conducted for the visitors as well. For our study, we analyzed the ethnomethodological sequence of events among the subjects to understand whether they were engaged with the explanations and the exhibits themselves.

From our field study, we found that human museum guides use a number of physical gestures while explaining an exhibit (Figure 1). For example, the pointing gesture is used to draw attention to a particular area of an exhibit. The circling gesture is used to indicate the exhibit as a whole. The height-indicating gesture is implemented by pointing with the index finger, starting at the top and drawing it down to the end at the bottom. The enactment gesture is used to act out a specific incident or occurrence. The counting gesture is implemented by showing a number of fingers, making them appear at a time. This gesture is usually an indicator of a small numerical value, for example, "...three years..." and the like. The pause pose is employed to allow the audience some time to lean in and inspect the exhibit in greater detail, especially if the guide has just finished explaining a finer detail about a certain (small) portion of the exhibit. The beckoning gesture is used to guide or cajole the audience in a particular direction, urging them to start moving. Our study reconfirmed the findings from the teams of Lehn et al. [7] and Luff et al. [8, 18].

Based on the gestures observed during the field study, we programmed the robot to use the same gestures (Figure 2) as a human guide when delivering the explanation about an exhibit.



(a) Pointing gesture

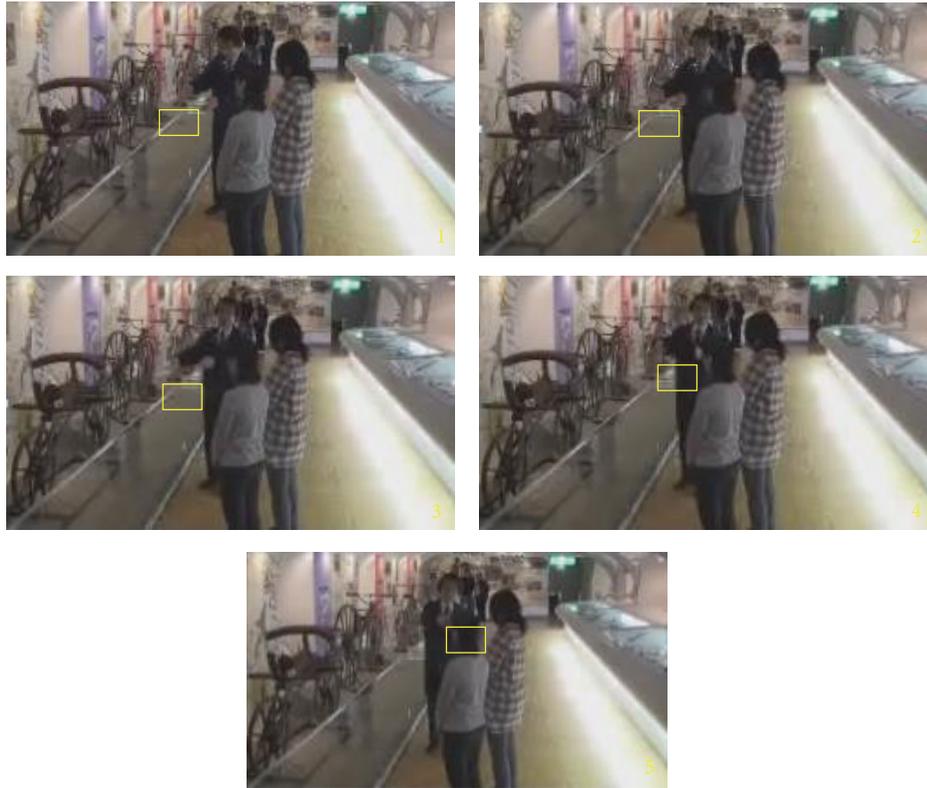


(b) Circling gesture



(c) Height-indicating gesture

FIGURE 1: Continued.



(d) Enactment gesture



(e) Counting gesture



(f) Pause pose



(g) Beckoning gesture

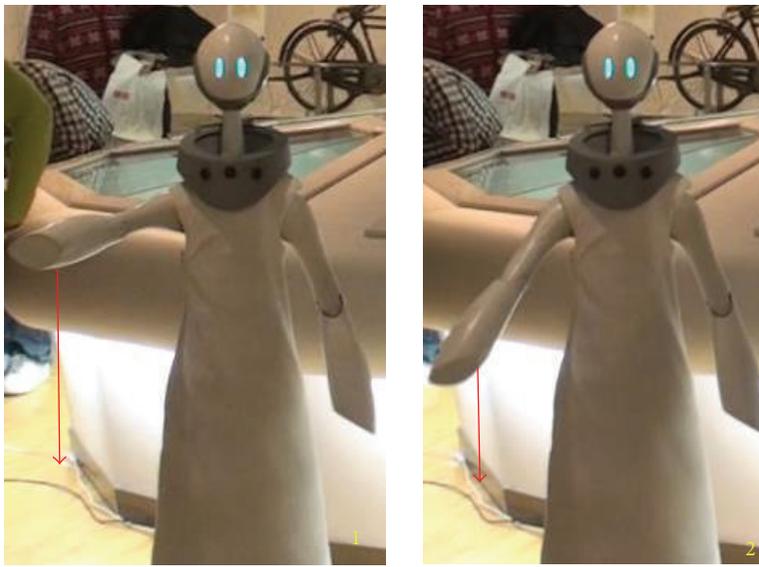
FIGURE 1: Gestures used by the museum curator to explain the exhibits to visitors during our field study.

2.4. *Development of the Script and Speech.* We used an application named AquesTalk2 (AQUEST Corporation) to develop the intonations of the robot's speech. This is a Japanese language speech development tool that allows for more natural pronunciations. Based on our field study, we developed a script for our robot. We made sure that the script was concise and contained humor, in addition to

being interesting overall. The explanation for each exhibit started with an introductory statement, followed by three objective statements, and ending with either the summative assessment statement or another objective statement of the same duration. After we developed the script, we created the robot's voice using AquesTalk2 text-to-speech tool. Then we developed the intonations of speech to make the robot's



(a) Pointing gesture



(b) Height-indicating gesture

FIGURE 2: Continued.



(c) Circling gesture



(d) Beckoning gesture

FIGURE 2: Robot's gestures based on our field study.

speech sound more natural to native Japanese speakers. We tested our robot's speech on several Japanese university students and collected their feedback through a questionnaire to improve the robot's speech as much as possible. We used a voice that was midway between the voice of a woman and the voice of a child. We did this on purpose to make the robot appear more friendly to the normal Japanese public.

2.5. Controlled Lab Experiments. After the ethnomethodological study, we began conducting a few pilot experiments. In the first set of pilot experiments, we tested the elements of metatalk, summative assessment combined with summative gesture and body orientation. We arranged for three laptop computers displaying three different exhibits, Exhibit A, Exhibit B, and Exhibit C (Figure 3), and placed them on the vertices of a triangle with each side measuring one and half meters (Figure 3(a)). We decided on this distance through trial and error. When the robot is placed beside one of the exhibits and the subject is positioned at a distance of about one meter from the robot (Figure 3(b)), the subject is still able to clearly see the other two exhibits without moving from the current location. We prepared two different sets of exhibits to counter the effect of the attractiveness of the exhibit. One set of exhibits displayed three rare plants while the other set displayed three different planets. We divided the subjects into two groups, Group A and Group B, each consisting of three subjects. We recruited subjects from among graduate school students at the university. All subjects were familiar with robotics and the particular robot platform used in the experiment. There were two conditions for each set of the experiment. In the first set of the experiment, in the first condition the robot delivers a summative assessment and metatalk at the end of the explanation of the current exhibit (Exhibit A) while in the second condition the robot does not include either of the two. In both conditions, the robot orients its body towards the midway point between Exhibit B and Exhibit C at the end of its talk so that the subject could not tell which exhibit the robot would move to next just from the robot's body orientation. In the second set of the experiment, in one condition the robot is programmed to deliver metatalk and correctly orient its body towards the next intended exhibit (Exhibit B) while in the second condition the robot delivers metatalk intended for one of the other two exhibits (Exhibit C) but orients its body toward the other exhibit (Exhibit B). We referred to [15] to design the robots pointing gestures and summative gesture. The robot's speech was adjusted so that in both conditions of the first experiment, the speech lasted for one minute and five seconds, whereas in the second experiment it was programmed to be fifty five seconds in both conditions. All of the six subjects participated in both experiments on consecutive days. The order in which the subjects were exposed to the conditions was reversed in Group B for both experiments to counter for the effect of the presence or absence of elements in the robot's explanation.

We placed position labels for both the robot and the subject near all three exhibits to indicate where each is supposed to be positioned initially. We employed a humanoid robot, Talk Torque 1 (TT1) with a 3 DOF neck, and 4

DOF arms but no mobility features. We briefed the subjects before we began the experiments, asking them to stay on the label marked position only for the purpose of video camera settings and move around freely after that. Also, they were requested to imagine a regular museum situation where they were visitors and a robot guide was explaining the exhibits. The experiment was conducted with one subject at a time. After each experiment, we requested the subjects to fill out a questionnaire in which we asked them to (1) rate, on a five-point scale, how well they understood at what point the robot was wrapping up the explanation for the current exhibit, (2) out of Exhibit B and Exhibit C which did they think was the next exhibit that will be explained, and (3) whether they thought the robot's speech and (4) gesture were executed at a comfortable-to-understand speed (also rated on a five-point scale).

2.6. Results of Controlled Lab Experiments. After analyzing the questionnaire data, for understanding the point of summative assessment, we found that almost all subjects in both the experiments across all conditions were able to gauge the next exhibit correctly except for one instance. One subject in Group B, during the second experiment, in the condition where the robot delivers the metatalk for Exhibit B but orients its body towards Exhibit C, the subject trusted the robot's body orientation more than the metatalk. Other subjects mentioned that during that part of the experiment they assumed that the robot was executing the incorrect body movement due to some technical trouble and so they decided to trust the metatalk than the body orientation. An analysis of the video data showed that, during the very first trial for all the six subjects, they tried to physically transfer to the next exhibit. One subject took two steps towards the next exhibit, while the other five subjects took one step towards the next exhibit, shifting their body weight in a pose to get ready to transfer. This is an interesting finding because all the subjects knew that the robot was immobile and therefore could not possibly move to the next exhibit to conduct the next explanation. The subjects may have expected that the experimenter will carry the robot to the next position. An analysis of subjects' body orientation and gaze direction revealed that both body orientation and gaze were directed at Exhibit A when robot's body was oriented towards Exhibit A. When the robot oriented towards the subject, the subject also reciprocated equally by turning towards the robot. When the robot was oriented towards the mid-way mark between Exhibit B and Exhibit C, the subjects orientated and looked at the exhibit that they thought was the next. In the last condition where the robot was oriented towards the wrong exhibit, five subjects were oriented towards the exhibit that was referred to in the metatalk but their gaze darted between the two candidate exhibits. In this condition, only one subject oriented towards the exhibit that corresponded with the robot's body orientation.

The above experiments indicated that subjects might be more likely to trust the content of the robot's speech, rather than the robot's body orientation. In the experimental conditions described above, these results might have been

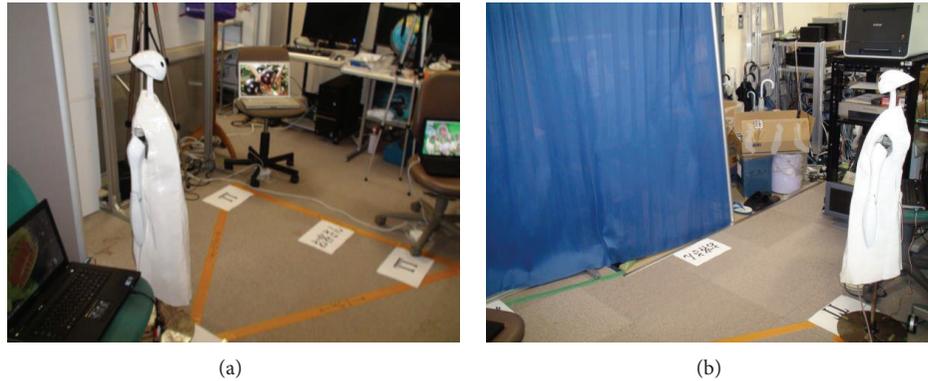


FIGURE 3: (a) Robot placed beside the first exhibit. (b) The robot placed on its label-marked position, waiting for the subject to step in front of it.

because the subjects were engineering students and expected the robot to possibly make an erroneous movement, due to some technical bug, but thought it to be unlikely for the robot to make a mistake in its speech, since it is preprogrammed and not generated in real time. From these results, it seemed that for subjects who are from the robotics field, a robot's verbal gestures might be more effective than its physical gestures. When we conducted the updated version of this experiment with subjects from the general public, we did not observe a significant difference in the responses to verbal and nonverbal gestures as discussed later in this paper (section: Results and Discussion, Statistical Analysis).

The controlled lab experiments indicated that subjects are more likely to trust a robot's speech rather than its physical orientation. Therefore, in our real-world experiment, we did not use the robot's physical orientation as a tool of disengagement. We used the lean-back gesture as a nonverbal indication of the end of the explanation of an exhibit. The robot used in the real-world experiment was a mobile robot, unlike the static robot used in our controlled lab experiment. During the real-world experiment, the robot would physically orient away from the exhibit that it had just finished explaining and orient towards the direction in which it will move to get to the next stopping point in the tour. This was an explicit indication to the subject that the explanation of the current exhibit is over and the robot guide is ready to move on to the next exhibit or the finishing point of the tour. Our real-world experiment showed that a verbal indication of the robot's intentions was unnecessary in case of the mobile robot. When the robot moved away, the subjects were forced to disengage from the current exhibit and move on with the robot guide. Our controlled lab experiment showed that in case the immobile robot is correctly oriented towards its next destination, a verbal indication of the same is rendered unnecessary.

2.7. The Experiment

2.7.1. Experimental Setup. Our robot platform, Talk Torque 2 (TT2), is a humanoid robot developed in our lab specifically

to study human-robot interaction. It has a 2-DOF neck, 3-DOF arms (2-DOF shoulder and 1-DOF elbow), and 1-DOF torso. It has two speakers, 3 cameras, and blue LED eyes. It has four sets of omni wheels for mobility. It has a placeholder at its base for a laptop computer. For our current research we attached a webcam to the laptop that was used by the robot to spot ARToolkit markers near the floor. The on-robot battery supplied power to its joint motors and wheels. The robot was connected to the laptop via USB.

The software architecture of the robot for the experiment described in the following pages includes device drivers and a simple user interface that can be used to launch a batch of commands to make the robot move and speak in a predetermined manner. The on-robot laptop can connect to a wireless local network. Another laptop connected to the same network can remotely control the on-robot laptop and hence the robot. The ARToolkit marker data captured through the webcam was processed and the corrective wheel movements were calculated and executed from inside the same module as the one that controlled the omni wheels. (Other versions of the robot software which include modules for other features that were not used in this research are not being described here).

For the real-world experiments, we programmed our robot platform, Talk Torque 2 (TT2), to deliver an explanation of three exhibits at the Tokyo Science Museum. The three exhibits were selected based on several different criteria, for example, location, relevance to each other, and entertainment value. We prepared TT2 to greet visitors after they consented to participating in our experiment and agreed to be videotaped. The subjects were guided to a starting point. Here, the robot begins by explaining the course of the tour; then it guides the subjects to the first exhibit and delivers the explanation of the first exhibit, followed by the second and the third exhibit. TT2 ends the guided tour by leading the subjects to the end point where it informs the group that the tour is over and bids them goodbye. The starting point was about one meter away from the first exhibit. The first and second exhibits were about five meters away from each other (with another exhibit between them). The second and third exhibits were about two meters away from each other. The

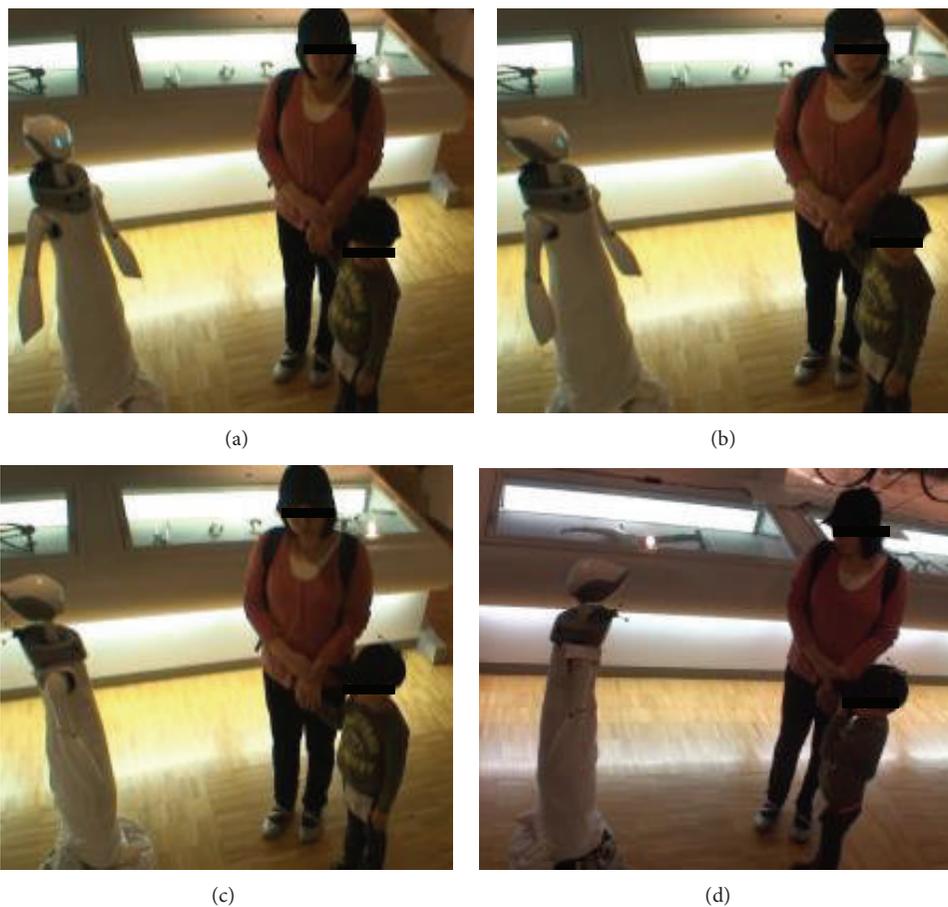


FIGURE 4: (a) Orientation when explaining an exhibit, (b) orientation when making eye contact and changing upper body orientation, (c) orientation when delivering summative assessment, and (d) orientation when moving on to the next exhibit.

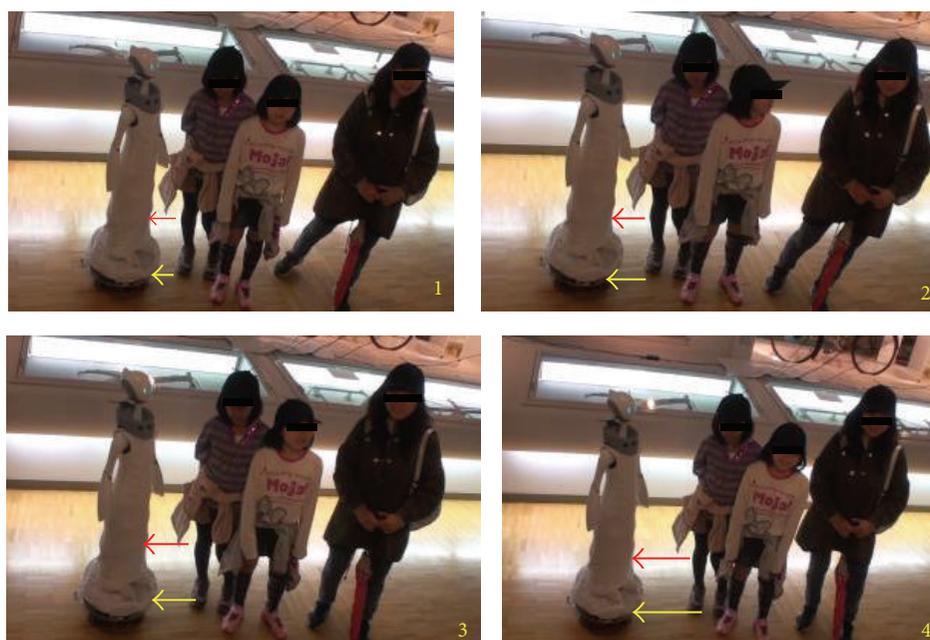
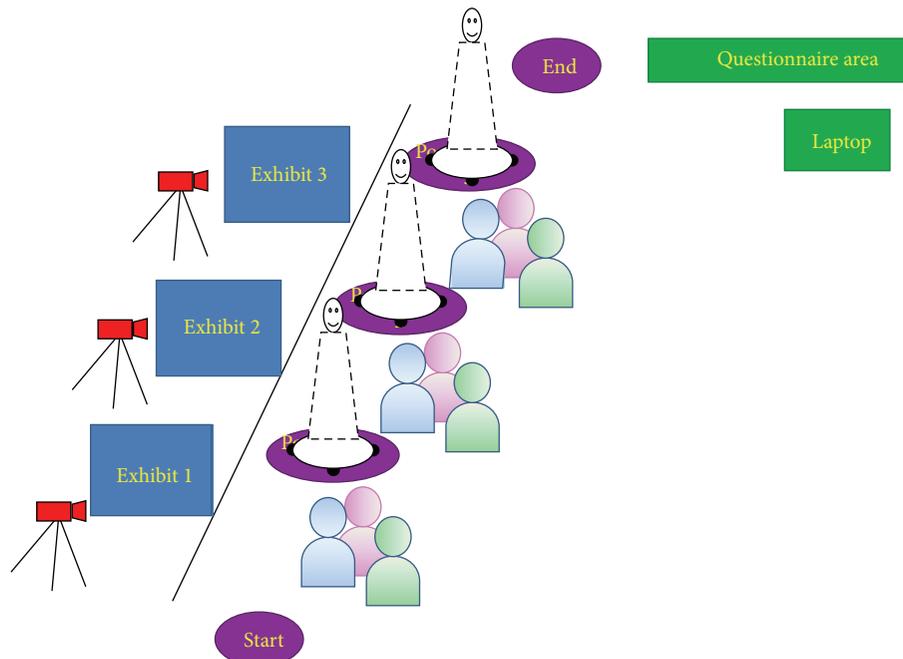


FIGURE 5: The lean back gesture.



(a) A schematic diagram of the experimental setup



(b) The starting point and trajectory of the robot

FIGURE 6: The experimental setup, the actual scene of the experiment.

end point was about one meter away from the third exhibit. We conducted our experiments in a room that displayed old bicycles. Some of the bicycles were over 200 years old. The explanation of the exhibits consisted of an account of the history of the bicycles. The robot moved along a more or less straight trajectory. The subjects usually followed the robot by walking behind it and arranged themselves on the

left of the robot when the robot was explaining the exhibits (Figure 4). The robot would face the exhibits at an angle of 45 degrees to the exhibit when delivering the explanation of the exhibit (Figure 4(a)) and make eye contact with the subjects at transition relevance places (TRPs) (Figure 4(b)). The robot would face the subjects directly during the delivery of the summative assessment (SA), thus being at an angle of 90

degrees to the exhibit (Figure 4(c)). The robot would move in a direction that would be 270 degrees with the exhibit during the lean back gesture (Figure 5). The robot would turn on its wheels, 180 degrees after the completion of each explanation and move on to the next stopping point (Figure 4(d)). We placed a camera behind each exhibit and also one just behind the end point. The cameras behind each exhibit captured the subjects' reactions and head and eye movements. The fourth camera focused on the whole trajectory and captured the entire experiment from start to finish.

2.7.2. Experiment History. We conducted the controlled lab experiments in the beginning of our study, as explained before. This was followed by several practice experiments with university students as subjects. Finally, we prepared the robot to conduct the first trial at the Tokyo Science Museum (Figure 6). After the first trial, we spent some time upgrading the hardware of the robot (e.g., controllers, etc.) and developed a locomotion guidance system for the robot using the free software library, ARToolkit. This helped us to ensure that there were very few instances when the robot strayed from its predetermined trajectory. Thereafter, we conducted the second trial of the experiment at the same venue. All parameters, space and time measurements, and protocol were kept exactly the same as in the first trial. The instances, where the robot did not move along the predetermined trajectory or did not stop at the predetermined stopping points or an experimenter had to manually adjust the robot's position, were excluded from our data analysis.

2.7.3. ARToolkit-Aided Guidance System. We used the freely available software library, ARToolkit (augmented reality toolkit), to develop an augmented-reality-marker based guidance system for the robot's locomotion. We decided to develop this system because we wanted to minimize the number of times the robot would stray from its predetermined trajectory due to slippage and other factors. Since we could not include the instances where the robot did not move in the preplanned manner in our analysis, we wanted to minimize the occurrence of such instances. Our guidance system receives distance and orientation feedback from ARToolkit markers (Figure 7) planted along the robot's desired trajectory and adjusts the robot's movements accordingly so that the robot moves in a more or less straight line. We fixed a web camera near the feet (base) of the robot, facing the side. This ensured that the camera's view was focused on the walls in front of the exhibits that prevented visitors from getting too close to the exhibits. Therefore, as the robot moved in a straight line from exhibit 1 to exhibit 2 to exhibit 3, the web camera recognized the markers on the walls on the side and sent the information to the robot's processors. This feedback helped the robot stay orientated directly in front and move along a straight path. At the end of each round of the robot-guided tour, the subjects were asked to fill up a questionnaire. Through this questionnaire, we collected data about the subjects' age, previous experience with robots, their opinion and impression of our museum guide robot, and



FIGURE 7: ARToolkit markers that can be spotted by the robot's on-board web camera.

historical facts about the exhibit to find out whether they were paying attention and carefully listening to the robot's speech.

2.7.4. Subject Profile. The subjects in both the first trial and the second trial consisted of mostly families who visited the Tokyo Science Museum over the weekends of our experiments. A typical subject group consisted of two parents and two children. There were some participants who came alone or came with their friends. Almost 90% of the subject groups were families of three or four people. They consisted of parents, typically a mother in her thirties and a father in his forties and two elementary school children (aged between 7 to 12 years). Less than 5% of the participants were above 60 years of age.

2.7.5. Tools of Engagement. The different tools of engagement used in our experiment were as follows: (a) eye contact at transition-relevance places (TRPs) [20], (b) upper body orientation, and (c) humorous content.

People make eye contact with their conversation partner(s) at certain points during a natural interpersonal conversation. These points are usually the places in a sentence where we put punctuation marks or pause during natural speech, for example, the end of a sentence where we put a period or a question mark. The points at which the turn to talk naturally passes from one participant in a conversation to another is termed as a transition relevance place (TRP). A natural unit of speech during which a speaker expects to retain the turn to speak is called turn construction unit (TCU) in conversation analysis. A speaker may naturally allow/expect the other participant to take the turn to start speaking at the end of a TCU or the speaker may continue to speak by starting a new TCU. Previous research by Yamazaki et al. [20] has shown that when a robot is engaged in an interaction with a person, the coordination between the robot's head movement (at TRPs) and its utterances engages the person more deeply in the interaction.

Our robot was programmed to make eye contact with the subjects at TRPs to mimic natural conversation flow. The robot would usually make eye contact with the subjects towards the end of each sentence (where we would normally insert a period in written materials). We accomplished this by



FIGURE 8: The subjects' reactions to the engagement tools.

programming the robot to look in the general direction of the anticipated location of the subjects. This happened to be at a 90-degree angle from the exhibit. We observed this behavior during our field study. This gave the subjects the impression of attention from the robot and resulted in engagement with the robot.

The robot was also programmed to torque its upper body to orient towards the audience at an angle of zero degrees with the subject (90 degrees with the exhibit) every time it made eye contact with the subject. This was also designed to draw the audience into the explanations [4]. This also helped to engage the audience.

Based on our field study, we prepared the robot's script making sure that we infuse some humor into the explanations. Since the historical information about the bicycles that

were the exhibits in our experiment had many anecdotes, we were able to prepare interesting content for the robot's speech.

2.7.6. The Conditions. We conducted our experiments with four different conditions. The first condition was the *SALB condition* where both summative assessment and the lean back gesture were used. The second condition was the *SA only condition* where only the summative assessment was used but there was no lean back gesture. The robot simply turned around after finishing the summative assessment and moved on to the next exhibit. The third condition was the *LB only condition* where there was a lean back gesture at the end of the last statement which was a similar objective statement like the three preceding it. The fourth condition was the

control condition, the Neither condition, where neither the summative assessment nor the lean back gesture was used. At the end of the explanation the robot would deliver another objective statement of the same duration as the summative assessment and simply turn around and move on to the next exhibit. The duration of all the explanations was exactly the same for all the exhibits during all the conditions. The lengths of the summative assessment and the last alternative objective statement were also of the same duration for all the exhibits during all the conditions.

2.7.7. Hypothesis

- (i) *Our first hypothesis was that a combination of the summative assessment and the lean back gesture will temporally be more effective than the use of these elements separately.*
- (ii) *Our second hypothesis was that there would be a lingering effect (failure to disengage) in all the conditions except for the SALB condition.*

3. Results and Discussion

Since all the parameters in both the first and the second trials were exactly the same, we have combined the numerical data from the two trials to present the results described in this section.

3.1. Engagement. We analyzed the audience reaction to all the engagement tools employed by the robot (Figure 8). The audience responded to the eye contact and upper body orientation of the robot by making eye contact with the robot (Figure 8(a)) and orientating towards the robot (Figure 8(b)) at the appropriate points. We found that the subjects responded to the humorous content by laughing (Figure 8(c)) at the appropriate places. We also asked the subjects some simple historical facts about the exhibits through the questionnaires that the subjects filled up at the end of each experiment. All the subjects answered these questions correctly, showing that they were engaged in the robot's explanation of the exhibits and listened carefully.

3.2. Point of Disengagement. We defined disengagement as the following ethnomethodological indicator: "the act of turning the head away from the current exhibit and not looking back at it." In other words, for the purpose of this experiment, we equate "not looking" with disengagement. For example, if a subject listens to the summative assessment being delivered by the robot and during the 5th second after the beginning of the summative assessment looks away from the current exhibit, looks at the next exhibit, and then looks back at the current exhibit and only when the robot starts moving towards the next exhibit, the subject starts walking with the robot but keeps gazing at the current exhibit and looks away during the 12th second and never looks back at the current exhibit again, then the subject will be considered to have disengaged from the current exhibit during the 12th second after the commencement of the summative

assessment. Figure 9 shows how we analyzed the point of disengagement from the video slices.

3.3. Independent Raters. We recruited two independent raters to look at a random sample of the video clips and provide their opinion about the disengagement status of the subjects to eliminate any bias on the part of the experimenters in determining the point of disengagement. We hired two graduate students from our research lab who were not involved with this research in any way and did not know what kind of data trends we were looking for. We asked these independent raters to analyze whether the subject was disengaged or not based on the definition of disengagement that we provided to the raters. To accomplish this, we created video clips of all the units of interaction, focusing exclusively on the part where the robot guide delivers the summative assessment (in case of the SALB and SA only conditions) or the last statement (in case of the LB only and Neither conditions) and moves on to the next exhibit. Each video clip started precisely at the point in time when the robot starts delivering the summative statement/last statement and ends exactly at the point when all the people in the current subject group have moved out of the camera's focus area. On an average, each video clip was approximately between 20 to 25 seconds long. This was because each subject took a different amount of time to move on to the next exhibit. We divided each video clip into one-second slices and labeled each slice with a number beginning from "1" and ending with the total number of seconds of the video clip. We asked the independent raters to identify the number in the video that corresponded with the exact point at which they thought that the subject had disengaged from the current exhibit.

We had two independent raters analyze all the videos of the first trial and a different pair of independent raters analyzed a random sample of around 20 video clips from our second trial.

We applied Cohen's Kappa to find out whether the experimenters' opinion concurred with the independent raters' opinion. Cohen's Kappa is a conservative measure of interrater agreement. It is used in research where a qualitative analysis by multiple raters is important to understand the outcome of experimental data. It takes cases where interrater agreements occur by chance into consideration and hence is considered more robust than a simple percentage calculation of interrater agreement. Although not universally accepted, it is widely considered that a magnitude of 0.41 to 0.6 indicates a moderate agreement; a magnitude of 0.61 to 0.8 indicates a substantial agreement, while a magnitude of 0.81 to 1 indicates perfect agreement. Only the cases where the Kappa coefficient was 0.7 or higher were considered for our data analysis.

3.4. Results of Disengagement Analysis. We considered a single subject's interaction with the robot with respect to a particular exhibit as a single unit of interaction [21]. We separated those units of interaction where the robot conducted the tour perfectly from those instances where there was some error or interference. When an entire tour was conducted perfectly, it resulted in the number of subjects

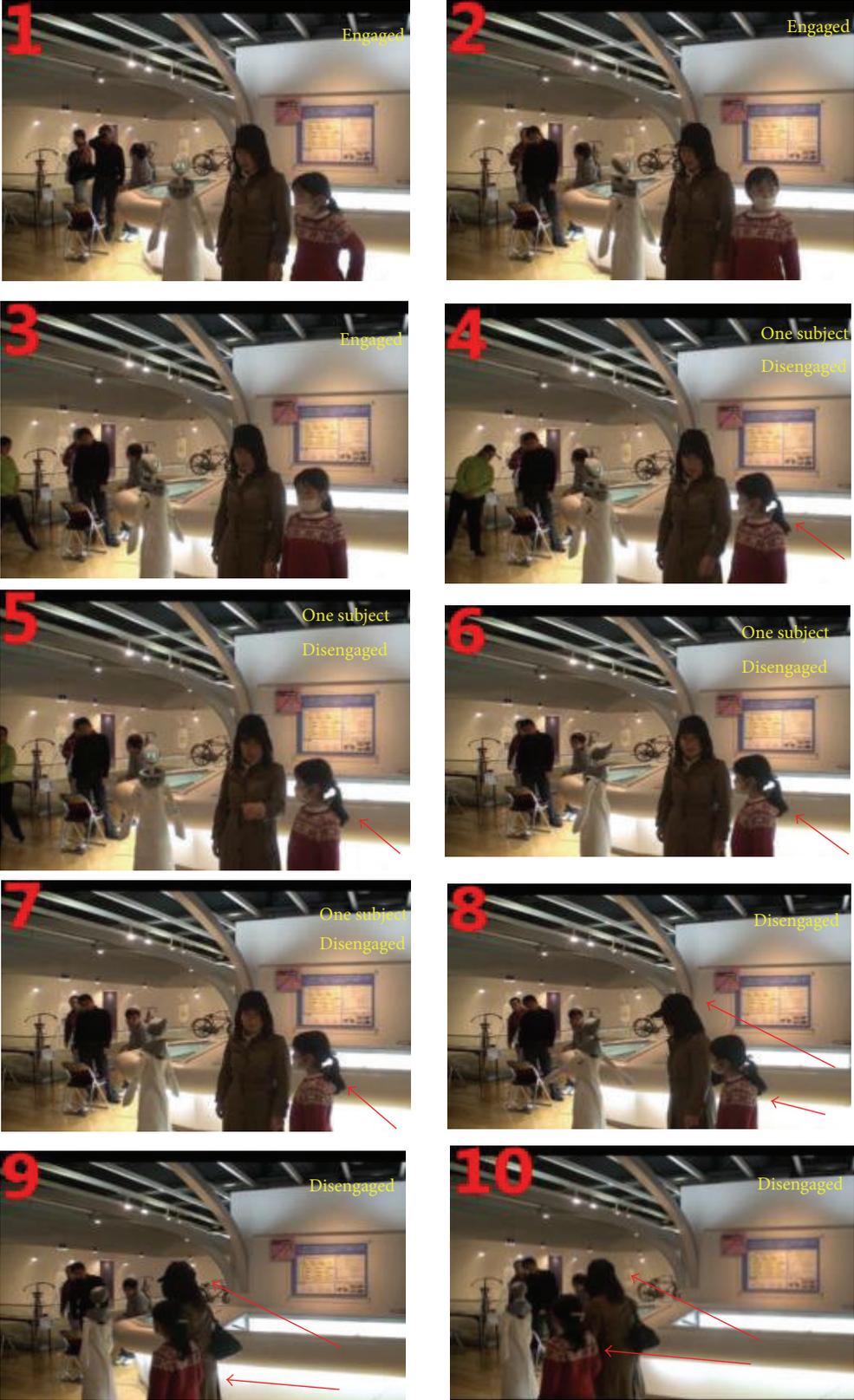


FIGURE 9: The process of disengagement as seen in sliced videos.

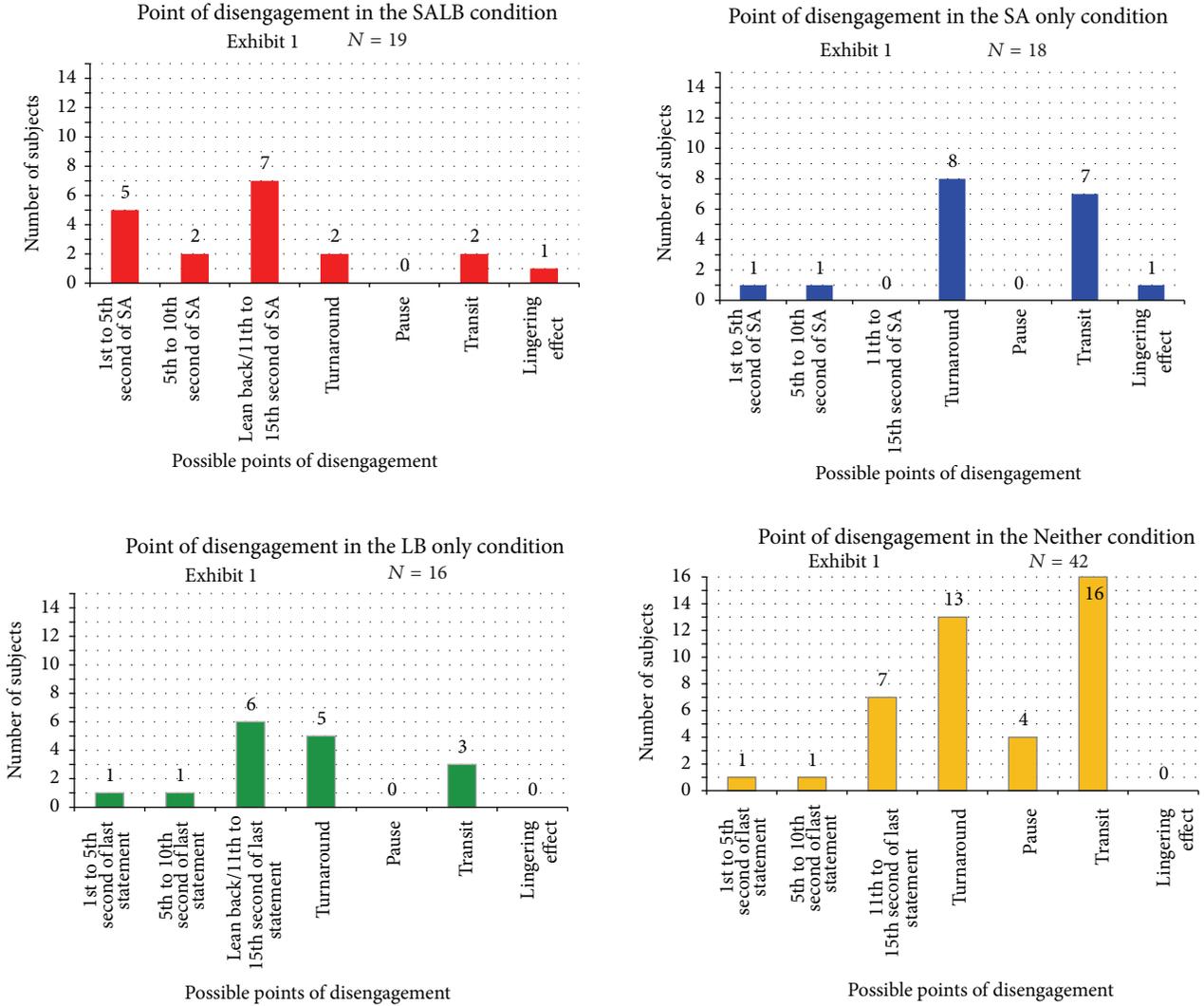


FIGURE 10: Points of disengagement in the four conditions at exhibit 1.

(n) multiplied by 3 (exhibits), that is, $3n$ units of interaction. When only a particular exhibit was explained without any interference, it resulted in n units of interaction. We included in our analysis only those instances where the explanation of an exhibit went smoothly and there was no interference to the subject’s engagement or disengagement with the exhibits. We eliminated any instances where either the robot did not function properly or an experimenter interfered with the scene. We analyzed a total of 232 units of interaction.

On an average, 34.41% of subjects disengaged from the exhibits during the first 10 seconds of the summative assessment in the SALB condition. An average of 22.92% of subjects disengaged during the lean back gesture in this condition. The disengagement results for the SALB condition are given in Table 1.

Among the subjects who participated in the SA only condition of the experiment, 15.19% disengaged from the exhibits during the first 10 seconds of the summative assessment. None of the subjects disengaged during the last 5

TABLE 1: Disengagement of the subjects from the exhibits in the SALB condition.

SALB condition	1st exhibit	2nd exhibit	3rd exhibit
1st to 10th second of SA	36.84%	42.86%	23.53%
Lean back/11th to 15th second of SA	36.84%	14.28%	17.65%

seconds of the summative assessment in this condition. The disengagement results for the SA only condition are given in Table 2.

About 6.73% of the subjects in the LB only condition disengaged from the exhibits during the first 10 seconds after the commencement of the last statement in the explanation of an exhibit. An average of 14.72% of subjects disengaged during the lean back gesture in this condition. The disengagement results for the LB only condition are given in Table 3.

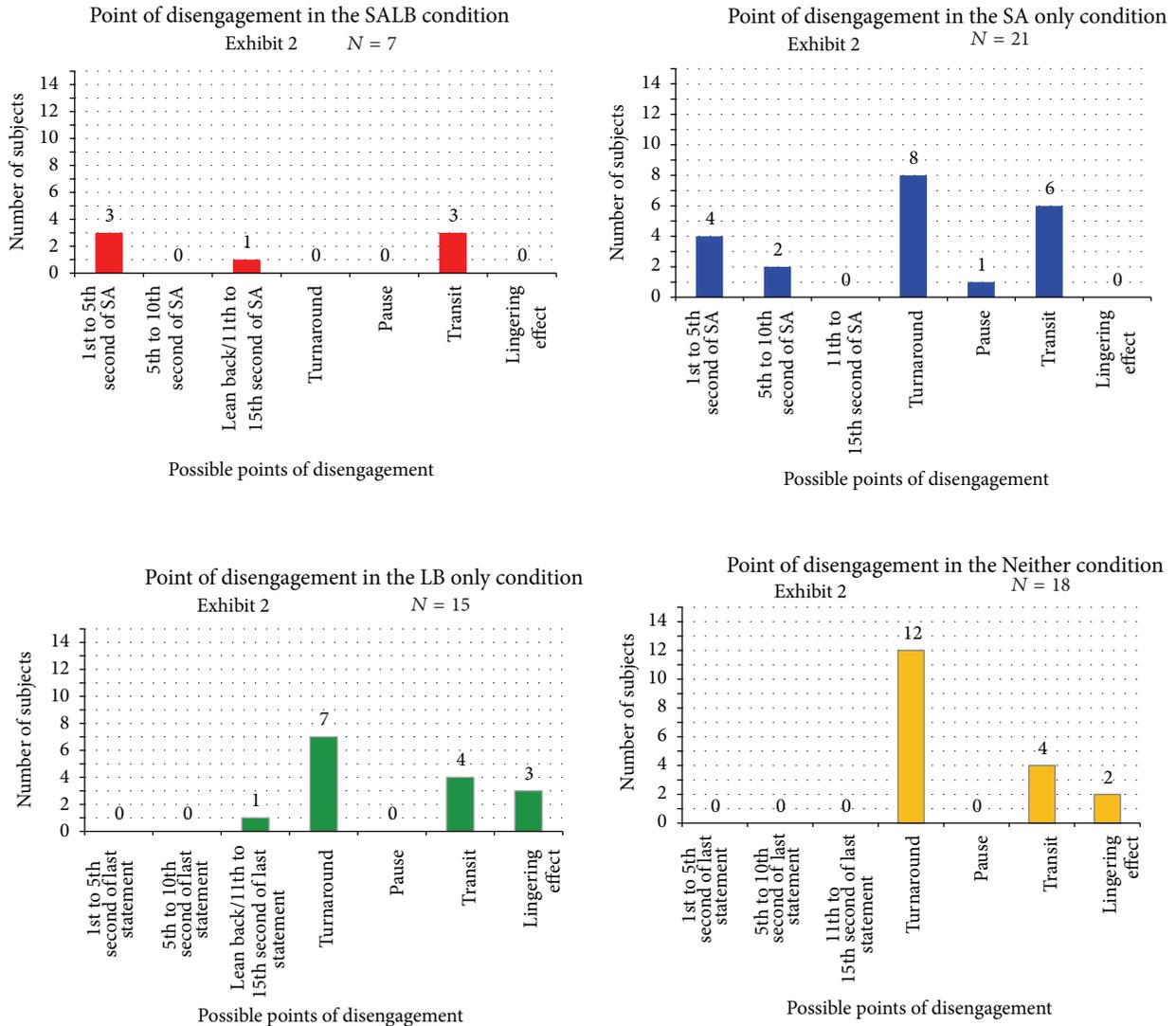


FIGURE 11: Points of disengagement in the four conditions at exhibit 2.

TABLE 2: Disengagement of the subjects from the exhibits in the SA only condition.

SA only condition	1st exhibit	2nd exhibit	3rd exhibit
1st to 10th second of SA	11.11%	28.57%	5.88%
11th to 15th second of SA	0%	0%	0%

The average figure for disengagement of the subjects from the exhibits during the first 10 seconds after the commencement of the last statement in the Neither condition is 4.76%. An average of 6.71% of subjects disengaged during the last 5 seconds of the last statement in the explanation of an exhibit in this condition. The disengagement results for the Neither condition are given in Table 4.

Figures 10, 11, and 12 show the trends in disengagement of the subjects from the exhibits in the different conditions.

TABLE 3: Disengagement of the subjects from the exhibits in the LB only condition.

LB only condition	1st exhibit	2nd exhibit	3rd exhibit
1st to 10th second of last statement	12.5%	0%	7.69%
Lean back gesture (11th to 15th second of last statement)	37.5%	6.67%	0%

3.5. Statistical Analysis. We performed an analysis of variance (ANOVA) on the experimental data. ANOVA is a test for statistical significance for the means of three or more groups. A single factor ANOVA may be performed to compare the means of two or more groups of numerical data. A multiple factor ANOVA can be performed to test the effect of two independent variables on one dependent variable. We performed a single factor ANOVA to find the statistical

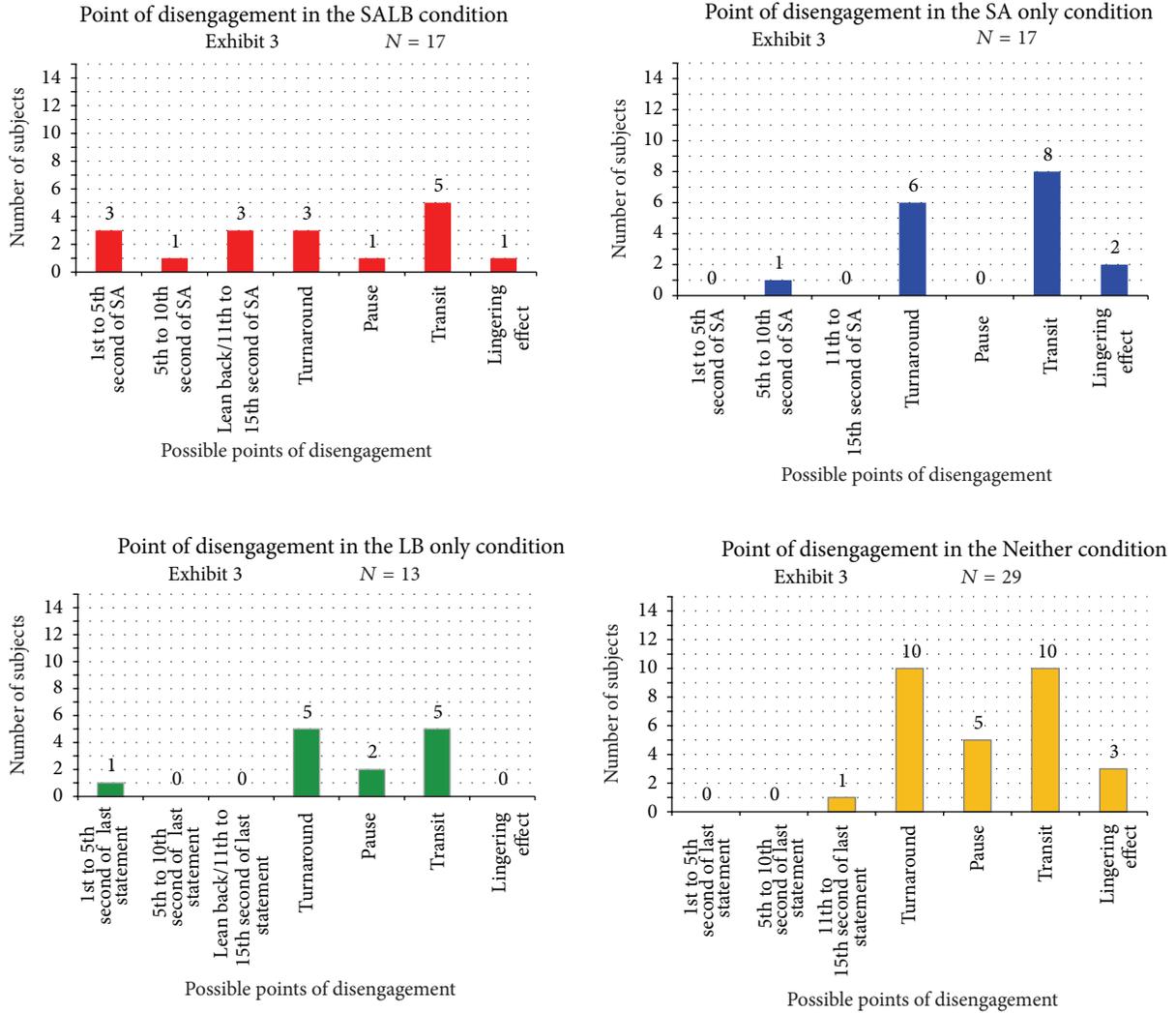


FIGURE 12: Points of disengagement in the four conditions at exhibit 3.

TABLE 4: Disengagement of the subjects from the exhibits in the Neither condition.

Neither condition	1st exhibit	2nd exhibit	3rd exhibit
1st to 10th second of the last statement	4.76%	0%	0%
11th to 15th second of the last statement	16.67%	0%	3.45%

significance of the time taken to disengage in the four conditions of our experiment for each exhibit.

We performed the statistical analysis on a total of 232 units of successful interactions. The result of single factor ANOVA for exhibit 1 was $P = 0.0028$. For exhibit 2, P was equal to 0.0004 and for exhibit 3 $P = 0.0184$. A nonparametric analysis showed that there is a significant difference between the conditions SALB and Neither/None (0.013) in the case of exhibit 1. In the case of exhibit 2, significant differences were observed between the conditions SALB and Neither/None (0.027), SALB and SA only (0.020),

and SALB and LB only (0.014). A significant difference was not observed in the case of exhibit 3.

3.6. Discussion. The results showed that on an average, 57.33% (highest) of the subjects disengaged from the exhibits during the delivery of the summative assessment statement in the SALB condition. In the SA only condition, an average of 15.19% and, in the LB only condition, 21.45% of the subjects disengaged during the first 15 seconds after the commencement of the disengagement efforts by the robot. In the Neither condition, an average of 8.29% of the subjects disengaged from the exhibit during the 15 seconds of the delivery of the last statement. From these results, we concluded that the SALB condition is more effective in disengaging the audience, if time efficiency is the most important consideration.

We found that although a large number of subjects disengaged from the exhibits during the robot's movements like turning around or moving on to the next exhibit during almost all the conditions, an exceptionally high number of subjects disengaged during the delivery of the summative

assessment in the SALB condition. This leads us to conclude that if time is a major consideration [22] and early disengagement is desirable, then using a combination of summative assessment and the lean back gesture (SALB condition) is a more effective tool of disengagement than using these elements separately or not at all.

We observe from the Figures 10, 11, and 12 that an unusually large number of subjects disengaged during the robot's turnaround movement in the Neither condition: 30.95% in case of exhibit 1, 66.67% in case of exhibit 2, and 34.48% in case of exhibit 3. This indicates that, in the absence of any cue to the subject about when the explanation of an exhibit is going to end, the subject realizes that the explanation has ended only when the robot turns around in preparation to move to the next exhibit, thereby delaying disengagement. This cluelessness might also be the reason why quite a few subjects disengaged during the pause after the turnaround in the Neither condition in case of the 1st exhibit (9.52%) and the 3rd exhibit (17.24%), even though subject disengagement during the pause has been negligible in all the other cases.

With the exception of exhibit 2, the lack of cues, like summative assessment and the lean back gesture, seems to delay the disengagement till the last moment, as is evident from the exceptionally high number of subjects who disengaged from the exhibit during the transit of the robot to the next destination in the Neither condition: (38.09%) in case of the 1st exhibit and (34.48%) in case of the 3rd exhibit.

Although, in 5 out of 12 cases above, the lingering effect was not observed, it did appear in a small number of subjects, being mostly limited to 1 or 2 subjects per exhibit per condition. Only in case of exhibit 2, in the SA only condition, and exhibit 3 in the Neither condition, the lingering effect rises to 3 subjects: 20% and 10.34%, respectively. Yet, the lack of cues did not always produce the lingering effect, in contrary to what we had predicted. We can see from exhibit 1 that the Neither condition does not show any lingering effect.

From our results, we can see that hypothesis (i) is supported. There seems to be an advantage in using a combination of the summative assessment and the lean back gesture when we want the audience to disengage quickly and prepare to move on to the next exhibit. The results also indicate that hypothesis (ii) is not supported. There seems to be a lingering effect in certain cases (exhibits 1 and 3) during the SALB condition, whereas there is no lingering effect in the Neither condition at exhibit 1. There are varying degrees of the lingering effect in the other conditions but not always.

4. Conclusion

We conducted a detailed study over a period of three years about borrowing well-established social cues used in human-human communication and apply them in human-robot communication. We analyzed whether such methods can be successful and whether it has the potential to make human-robot interaction smoother. Our results show that, within the context described in this paper, the use of such social cues can be beneficial in human-robot interaction. In particular, summative assessment and the lean back gesture

show great promise as tools of disengagement. We presented a detailed account of our journey in this paper and hope that the scientific community will benefit from our findings. We plan to address nonlinear robot trajectories and collision avoidance or obstacle avoidance in robot-guided museum tours in our future work.

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

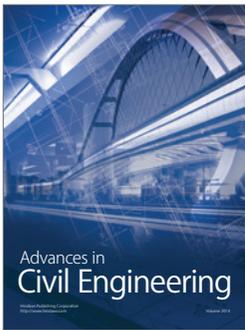
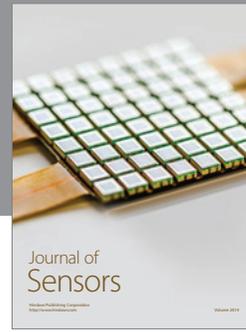
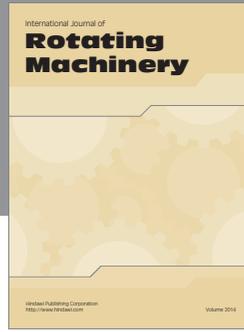
Acknowledgments

The authors would like to thank their collaborators, Professor Paul Luff and Dr. Katie-Ann Best from the Work, Interaction, and Technology (WIT) research group, King's College, London, for sharing valuable insights about the work and behavior of human museum guides and the interaction between people and technology in general. They would also like to thank Yūki Akamine, Deokyoung Shin, Takumi Itō, and Sosuke Takase for helping them with the logistics of conducting the experiments at the Tokyo Science Museum and for assisting them in the fieldwork.

References

- [1] M. Ghosh and F. Tanaka, "The impact of different competence levels of care-receiving robot on children," in *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems: Celebrating 50 Years of Robotics (IROS '11)*, pp. 2409–2415, San Francisco, Calif, USA, September 2011.
- [2] T. Kanda, T. Hirano, D. Eaton, and H. Ishiguro, "Interactive robots as social partners and peer tutors for children: a field trial," *Human-Computer Interaction*, vol. 19, no. 1-2, pp. 61–84, 2004.
- [3] T. Kanda, M. Shiomi, Z. Miyashita, H. Ishiguro, and N. Hagita, "An affective guide robot in a shopping mall," in *Proceedings of the 4th ACM/IEEE International Conference on Human-Robot Interaction (HRI '09)*, pp. 173–180, San Diego, Calif, USA, February 2009.
- [4] M. Shiomi, D. Sakamoto, T. Kanda, C. T. Ishi, H. Ishiguro, and N. Hagita, "A semi-autonomous communication robot—a field trial at a train station," in *Proceedings of the 3rd ACM/IEEE International Conference on Human-Robot Interaction (HRI '08)*, pp. 303–310, Amsterdam, The Netherlands, March 2008.
- [5] H. Kuzuoka, K. Pitsch, Y. Suzuki et al., "Effect of restarts and pauses on achieving a state of mutual orientation between a human and a robot," in *Proceedings of the ACM Conference on Computer Supported Cooperative Work (CSCW '08)*, pp. 201–204, San Diego, Calif, USA, November 2008.
- [6] H. Kuzuoka, Y. Suzuki, J. Yamashita, and K. Yamazaki, "Reconfiguring spatial formation arrangement by robot body orientation," in *Proceeding 5th ACM/IEEE International Conference on Human-Robot Interaction (HRI '10)*, pp. 285–292, Piscataway, NJ, USA, March 2010.
- [7] D. vom Lehn, C. Heath, and J. Hindmarsh, "Exhibiting interaction: conduct and collaboration in museums and galleries," *Symbolic Interaction*, vol. 24, no. 2, pp. 189–216, 2001.

- [8] P. Luff, J. Hindmarsh, and C. Heath, *Workplace Studies, Recovering Work Practice and Information System Design*, Cambridge University Press, 2000.
- [9] W. Burgard, A. B. Cremers, D. Fox et al., “Experiences with an interactive museum tour-guide robot,” *Artificial Intelligence*, vol. 114, no. 1-2, pp. 3–55, 1999.
- [10] I. R. Nourbakhsh, J. Bobenage, S. Grange, R. Lutz, R. Meyer, and A. Soto, “An affective mobile robot educator with a full-time job,” *Artificial Intelligence*, vol. 114, no. 1-2, pp. 95–124, 1999.
- [11] S. Thrun, M. Beetz, M. Bennewitz et al., “Probabilistic algorithms and the interactive museum tour-guide robot Minerva,” *The International Journal of Robotics Research*, vol. 19, no. 11, pp. 972–999, 2000.
- [12] A. Chella and I. Macaluso, “The perception loop in CiceRobot, a museum guide robot,” *Neurocomputing*, vol. 72, no. 4–6, pp. 760–766, 2009.
- [13] C. Breazeal, A. Takanishi, and T. Kobayashi, “Social robots that interact with people,” in *Springer Handbook of Robotics*, pp. 1349–1369, Springer, New York, NY, USA, 2008.
- [14] M. A. Goodrich and A. C. Schultz, “Human-robot interaction: a survey,” *Foundations and Trends in Human-Computer Interaction*, vol. 1, no. 3, pp. 203–275, 2007.
- [15] Y. Suzuki, H. Kuzuoka, J. Yamashita et al., “Study on drawing attention by a speechless robot,” in *Proceedings of the Human Interface Symposium*, pp. 615–618, 2007, (Japanese).
- [16] H. Garfinkel, *Studies in Ethnomethodology*, Prentice-Hall, 1967.
- [17] A. Kendon, *Conducting Interaction, Patterns of Behavior in Focused Encounters*, Studies in Interactional Sociolinguistics 7, Cambridge University Press, Cambridge, Mass, USA, 1990.
- [18] C. Heath, P. Luff, J. Hindmarsh, and D. vom Lehn, “Work, Interaction & Technology Research Group,” Selected Papers I, II, and III, King’s College London, University of London, 2003.
- [19] K.-A. Best, *The interactional organization of the guided tour: the work of tour guides in museums, galleries and historic houses [Ph.D. thesis]*, King’s College London, London, UK, 2008.
- [20] A. Yamazaki, K. Yamazaki, Y. Kuno, M. Burdelski, M. Kawashima, and H. Kuzuoka, “Precision timing in human-robot interaction: coordination of head movement and utterance,” in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI ’08)*, pp. 131–139, Florence, Italy, April 2008.
- [21] M. K. Lee, S. Kiesler, and J. Forlizzi, “Receptionist or information kiosk: how do people talk with a robot?” in *Proceedings of the ACM Conference on Computer Supported Cooperative Work (CSCW ’10)*, pp. 31–40, Savannah, Ga, USA, February 2010.
- [22] M. Ghosh and H. Kuzuoka, “A trial attempt by a museum guide robot to engage and disengage the audience on time,” in *Proceedings of the Winter International Conference on Engineering and Technology (AASRI-WIET ’13)*, American Applied Sciences Research Institute, Saipan, Northern Mariana Islands, USA, December 2013.



Hindawi

Submit your manuscripts at
<http://www.hindawi.com>

