The 23rd IEEE International Symposium on Robot and
Human Interactive Communication
August 25-29, 2014. Edinburgh, Scotland, UK,

Robot Behavior Designed to Encourage Conversations Between Visitors in an Exhibition Space

Kohei Matsumura[1], Takumi Gompei[2] and Yasuyuki Sumi[3]

***Abstract*— In this paper, we propose a novel way to encourage visitors to share their experiences and interests in exhibition spaces. To acquire information about the experiences of the visitors, we used a robot that inhabits the exhibition space. The robot acquires the visitor experience data and then shares this data with other visitors. We believe that this information (i.e., the experiences of previous visitors) will enhance the exhibition experience for subsequent visitors because the information may include interesting topics for the new visitors. However, there are some difficulties in communication between robots and humans. We used "PhotoChat", which is an in-house photo communication software. PhotoChat is capable of communicating with others by taking photographs and adding annotations to each photograph. It also records the locations for coordination between the photographs and the statistical information contained in the annotations. We therefore used PhotoChat as the communication medium between the robot and the visitors. The robot acquires the visitor experiences and shares these experiences with other visitors through PhotoChat. The robot can also use bodily actions to express instructions to the visitors. We then developed a system that coordinates the photographs and their respective exhibition areas. We also implemented robot behavior (i.e., bodily actions and motions) that includes recommendations for photographs taken by others. We held workshops to perform data collection and manually classified the data into three content categories. We then performed experiments using the developed system to distribute the classified content. The results showed that the robot's behavior encouraged conversations between the visitors based on provided topics.**

I. INTRODUCTION

This paper describes our work towards realization of a museum guide robot that helps the visitors to share their experiences. Fig. 1 shows a typical scene where our system is working in a small exhibition space. Individual visitors can freely enjoy their tour of the exhibition while being connected with a system called PhotoChat, which supports casual conversation among the users by allowing them to take and share photographs and add hand-written notes to the photographs [1]. The robot located at the center of the exhibition space observed the individual visitors' touring histories and the PhotoChat-based conversations among them. The robot is then able to gather an overview of the exhibits, e.g., it can detect popular topics with regard to certain exhibits and observe relationships among the exhibits from the accumulated data. The robot can then provide new visitors with exhibition guidance based on the previous PhotoChat conversations.

In this paper, we investigate the effects of the robot's embodiment, i.e., its speech and gestures, on the visitors' acceptance of the robot and the interactions among the visitors.

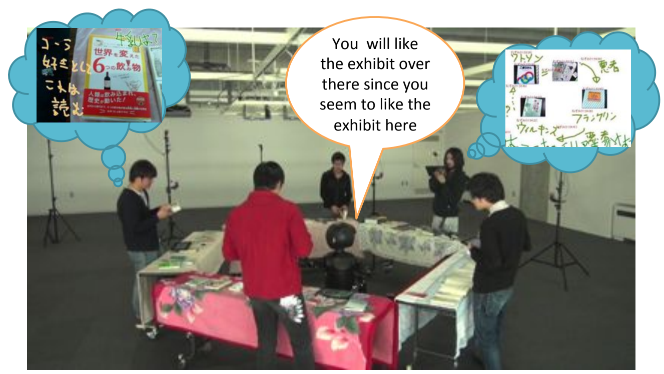

Fig. 1. Robot guidance based on the PhotoChat conversations among the visitors.

II. ROBOT GUIDANCE BASED ON PHOTOCHAT CONVERSATIONS AMONG VISITORS

The proposed robot guidance system supplies users with context-aware information provided by a robot that is networked with both environmental sensors and the hand-held systems that are used by the individual users.

We chose exhibition spaces to be our experimental field. The individual visitors carry hand-held systems (running PhotoChat) that enable them to share their interests and new findings instantly during exhibition tours with other visitors. Their touring histories (i.e. their current position and the time spent at each exhibit) are recorded using a location system, i.e., an optical motion capture system.

At the center of the exhibition space, we provide a guide robot, which acts on behalf of the human facilitators, and speaks to the visitors to suggest topics related to the individual exhibits and to recommend other exhibits to the visitors. The guidance timing and addressees are calculated based on the touring histories of the individual visitors. The robot sends a photograph to all PhotoChat terminals to serve as guidance content at times when it is speaking to a particular visitor. The PhotoChat photo used has been selected from stored photos taken by previous visitors; e.g., a photograph showing findings on and conversations with

[1]Kohei Matsumura was with Future University Hakodate. He is now with Ritsumeikan University, 1-1-1 Noji-higashi, Kusatsu, Shiga, 525-8577 Japan. matsumur@acm.org
[2]Takumi Gompei was with Future University Hakodate. He is now with Nomura Research Institute, Japan.
[3]Yasuyuki Sumi is with Future University Hakodate, 116-2 Kamedanakano-cho, Hakodate, Hokkaido, 041-8655 Japan. sumi@acm.org

978-1-4799-6765-0/14/$31.00 ©2014 IEEE

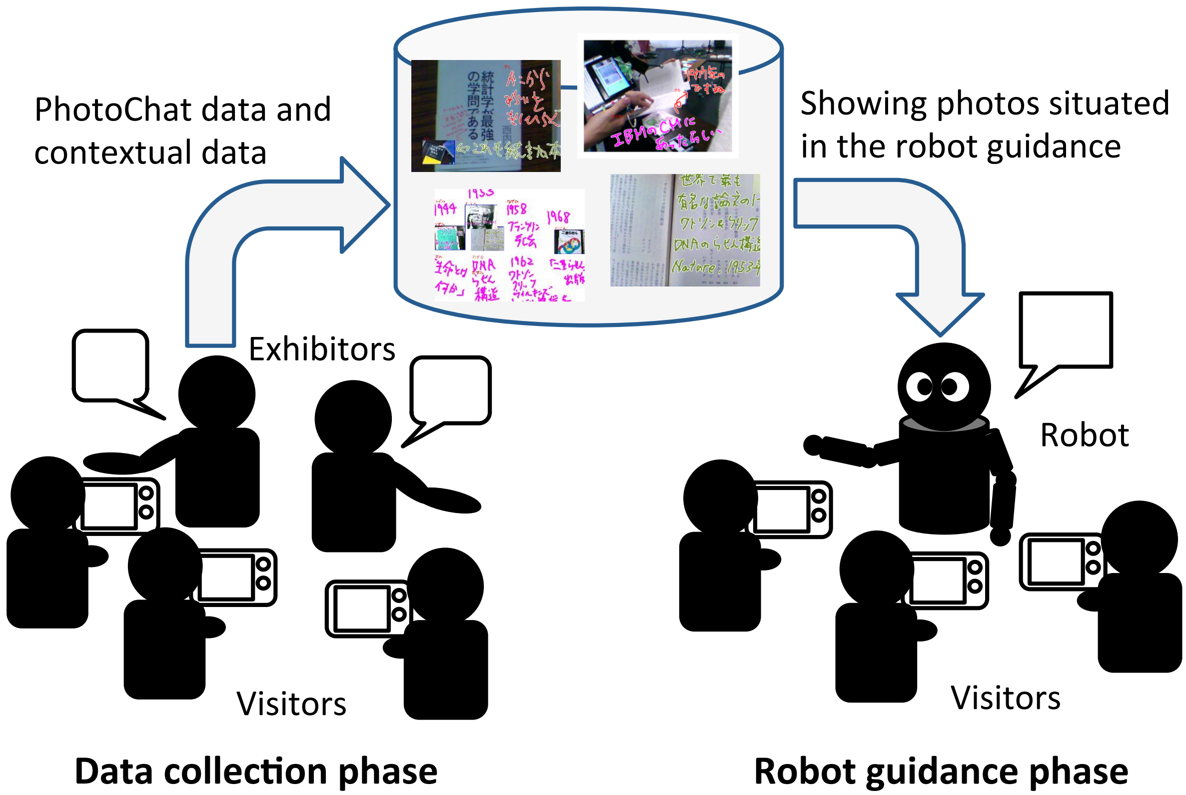

Fig. 2. Framework used for this study. The study consists of two phases. The first is the data collection phase, and the second is the robot guidance phase.

regard to certain exhibits by previous visitors is conveyed, based on the similarity of their contexts.

One important characteristic of our system is that we use knowledge from the experiences of the visitors (the system users), and not knowledge that was prepared by system designers beforehand. Another important characteristic is that our system embeds the guidance content within the physical context using a robot embodiment, and not simply via personal mobile systems. Our research interests also include the social effects of the guidance provided (i.e. acceptance of the guidance provided and initiation of interactions among the visitors) by the robot embodiment (including the effect of its approach, speech, head gestures, and hand gestures).

III. RELATED WORK

Building robots that can move autonomously based on an understanding of their environment and that can then assist people has been a long-standing goal for researchers in artificial intelligence and robotics. Many works on guide robots have been conducted at venues such as museums [2] and offices [3]. This pioneering research into guide robots focused mainly on autonomous movement based on a visual understanding of the surrounding environment and spoken dialogue with people.

Recently, many works have focused their research in another direction, i.e., on social interactions between people and robots. For example, Kanda et al. [4], [5] presented a robot that could act as a social partner for children, that had stayed for long periods and had communicated with children at a primary school and a science museum. The research group of Yamazaki and Kuzuoka [6], [7], [8] analyzed the effects of the nonverbal behavior of museum guide robots such as stance, the coordination of head movements with speech, and their paralanguage based on a sociological interaction analysis.

The robot guidance parts of these works was based on guidance content that was prepared beforehand by the researchers and the museum curators. In contrast, our system attempts to use the visitors' collaborative experiences and findings during their tours to realize a robot that can offer a medium to help the visitors to share their knowledge and findings.

One noteworthy attempt to collect real-world news by talking to people nearby was proposed by Harada et al. [9] to act as a robot journalist. However, it is still difficult to determine the significant scenes from visual processing alone. We use PhotoChat [1] to ease the exchange of personal interests and findings among the exhibition visitors. We expect our approach to raise an intuitive exchange of knowledge that is embedded in the exhibition context.

We previously proposed a method to encourage visitors to share their experiences by recording the social interactions among them and then virtually attaching first-person view videos and speech to the context in the exhibition space [10]. In this paper, we use PhotoChat to enhance the visitors' casual record-taking and allow them to exchange their interests and findings by taking photographs and making hand-written notes on them. This paper investigates the effects of the robot's embodiment, e.g., its positioning, head gestures, and hand gestures, on stimulation of social interactions among the robot and the visitors [11].

IV. IMPLEMENTATION

Fig. 2 shows the framework used for this study. The framework consists of two phases. The first is the data collection phase and the second is the robot guidance phase. We held an exhibition of books as a space for the data collection phase, where we collected the "conversations" between the visitors. The participants held "conversations" via the PhotoChat system. In the robot guidance phase, the robot distributed selections from the collected conversations as content for the guidance phase. We classified the collected conversations on the basis of statistical data, such as the number of text annotations and/or the structure of the conversation.

As a first step towards system implementation, we held two workshops and collected information on the participants' interests and some insights into the exhibits. Fig. 3 shows some snapshots taken of these workshops. There are five tables in the space and about ten books are on each table. The visitors freely picked up the books and had conversations about their interests and insights via PhotoChat.

Fig. 3. The overview of the two workshops (exhibitions). There are five exhibits in each exhibition (upper). Participants hold conversations via PhotoChat (bottom).

Fig. 4 shows an annotated screenshot taken from PhotoChat. PhotoChat is software that runs on mobile PCs that have a camera module and a pen tablet. Users can take photographs as they would with conventional digital cameras and write and draw on them via the pen interface. The photographs and the pen strokes are distributed to the other machines that are running PhotoChat by an ad-hoc wireless network in real time, so that the users can share photographs and chat about them.

As shown in Fig. 4, most of the PhotoChat display is laid out to show the camera viewfinder and the photograph browsing/notes. Thumbnails of all photographs taken by the connected users are listed on the left hand of the display, along with the times they were taken. Each thumbnail shows the user name and the time for each photograph. PhotoChat also has a unique function for hyperlinking of the photographs. The users can hyperlink between photographs by dragging thumbnails of the photographs and dropping them onto another photograph or a blank sheet. The users can then jump to the linked photographs by clicking on the thumbnails.

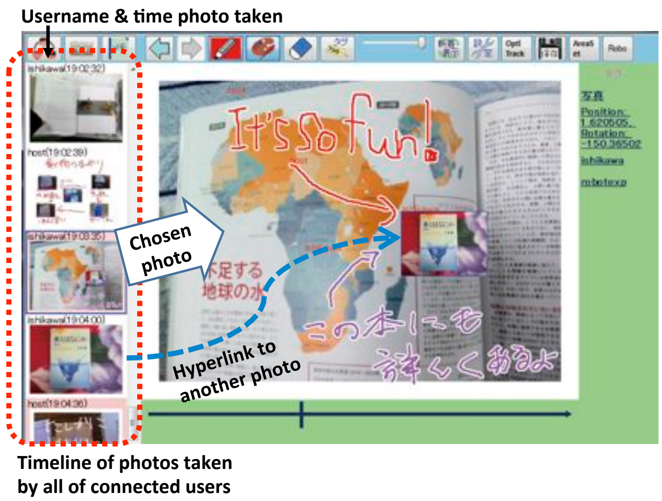

Fig. 4. An annotated screenshot of PhotoChat

We collected the PhotoChat data, including the photographs that were taken in the workshop and the handwritten annotations on the photographs. In addition to these data, we also collected the metadata (i.e., the context) for each photograph, such as user name, location and time taken. We can estimate the areas that got the most attention, the stimulated conversations, and typical structures for the conversation from the context data of the conversation on PhotoChat. To gather the locations (i.e. areas) of the participants, we used a motion capture system (Natural Point Inc., OptiTrack).

In this paper, we classified the annotated photographs on PhotoChat into three categories. The first is the type where the photograph includes a "short description" of the exhibit, e.g., a visitor abstracts the theme of the exhibit and annotates it on the photograph. The next category is the type where the photograph provides a detailed description of and/or new insight into the exhibit. Because it provides the other visitors with a new topic of conversation, we call this category "providing a topic". The third type is "relationship between exhibits", which includes link(s) between two or more of the exhibits and explains the relationship between these exhibits. The robot shows these contents in combination with his bodily actions.

We manually classified the annotated photographs on PhotoChat in the series of experiments and created the content for distribution; however we our feasibility study showed that the content classification process can be automated using a support vector machine (SVM) with an accuracy rate of 86%. In the feasibility study, we employed following three features for SVM parameters, the number of pen strokes, the number of hyperlinks and the number of people who annotated into the photograph. In order to evaluate an accuracy of SVM classification, we applied five class cross-validation to the 294 PhotoChat data (i.e., annotated photographs).

Fig. 5 shows an overview of robot's behavior. The robot finds a visitor who remains at an exhibit for several minutes, and provides him/her with content accompanied by bodily actions. We set the rules for providing content such that they depend on the visitor's behavior, as follows:

- **Sh** The robot provides a "short description" content if a visitor meets two conditions: they stay for more than 5 minutes at an exhibit and do not take photographs frequently (less than two photographs per 5 min). The robot states that "This is a short description of this exhibit" while simultaneously distributing "short description" content to PhotoChat.
- **Pr** The robot provides a new topic for a visitor if a visitor meets the following two conditions: they stay for over 9 minutes at an exhibit and take photographs frequently (more than three photographs per 9 min). The robot says "You will like this photo because you seem to like the exhibit here" while distributing the appropriate content to PhotoChat. The robot will then continuously provide new topics every 3 minutes if the visitor meets one condition: they have taken at least one photograph in the last 3 minutes.
- **An** The robot recommends another exhibit if a visitor meets two conditions: they have received a "short description" from the robot and have not taken any photographs in the last 3 minutes. The robot states that "There is another exhibit over there", while providing a pointing action and distributing appropriate content to PhotoChat.
  In addition, the robot also recommends another exhibit if there is no distributable content remaining, even if a visitor has continuously shown his/her interest in the exhibit. The robot states "You will like the exhibit over there because you seem to like the exhibit here", accompanied by a pointing action, while distributing appropriate content to PhotoChat.

Fig. 6 shows two examples of the robot's behavior. In the upper part of the figure, a visitor first receives a "short description" from the robot because he/she stayed for more than 5 minutes at an exhibit. Then, the visitor received a recommendation for another exhibit because he/she did not take any further action. In this case, the robot guessed that the visitor has no interest in this exhibit. In the lower part of the figure, a visitor continuously received new topics from the robot because the visitor took some pictures during his/her visit. The robot will continuously provide new topics every 3 minutes if a visitor shows his/her interest in the exhibit. In this example, the robot finally recommends another exhibit, because the visitor had not taken a picture in the last 3 minutes.

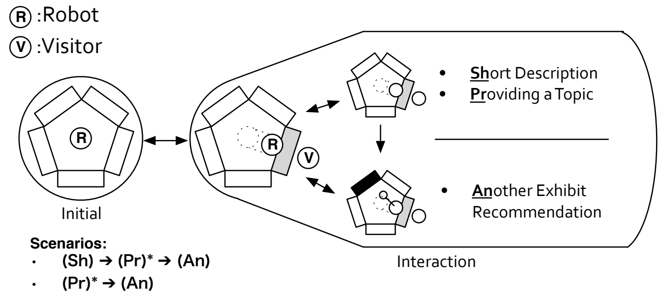

Fig. 5. Overview of the robot's behavior. The robot has three possible courses of action. The first is to provide a short description of the exhibit, the second is to provide a topic about the exhibit, and the third is to recommend another exhibit to the visitor.

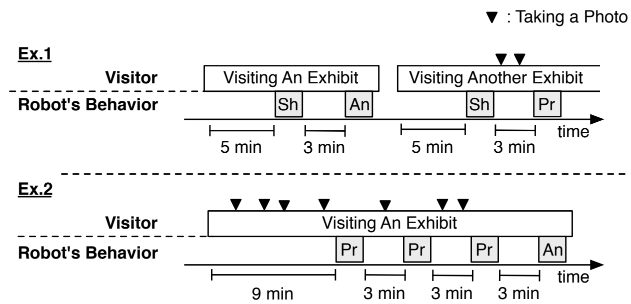

Fig. 6. Two examples of the robot's behavior. The robot varies its behavior depending on the visitor's actions.

V. Experiment

We performed two workshops for data collection and an experiment to investigate the interactions between the robot guide and the visitors. As shown in Fig. 3, we used a narrow table for each exhibit, with five tables in total. Each exhibit contains about ten books, which are prepared by each exhibitor in accordance with his/her interests (i.e., the theme of the exhibit) beforehand.

Five exhibitors participated in the first workshop. The workshop took approximately 50 minutes. Each participant explained his/her exhibit to the others and they visited each other's exhibits. This means that each participant talked as both an exhibitor and a visitor with the others. We asked all participants to take photographs and to add annotations to the photographs using PhotoChat. Each participant recorded his/her interests and insights into the individual books, their findings for each exhibit, and the relationships between the exhibits.

The five exhibitors who had participated in the first workshop and five new visitors participated in the second workshop. The workshop again took approximately 50 minutes. The exhibition setting was same as that of the first workshop. We asked the five visitors to use PhotoChat to take photographs and to add annotations to these photographs. We also asked the exhibitors to explain their respective exhibits. Thus, in this case, the exhibitors did not move away from their exhibits and did not have access to PhotoChat.

After the two workshops that were intended to collect data, we created the content for distribution. As explained earlier, we manually classified the PhotoChat data (i.e., the annotated photographs) for distribution. We also implemented the robot behavior, including physical actions such as moving forward and pointing at a distant object. However, because of technical limitations, we did not implement automated physical action functions for the robot. In this case, the movements of the robot were remotely controlled by one of the experimenters. Instead of implementing automated physical action, we took the Wizard of Oz approach [12]. We developed a notifier program that notifies both the timing of and the instructions for the robot's actions to the experimenter.

Eight participants participated in the series of experiments (three for the first session, and five for the second session). Each session again took about 50 minutes, for a total of 100 minutes. We used two video cameras to study the interactions among the robot and the participants.

VI. Results

We collected 207 and 87 photographs from the first and second workshops, respectively, giving 294 photographs in total. We manually classified them into three categories, as described above. In addition to the three categories, we dropped any photographs that had no annotation or seemed difficult to categorize. We categorized 52 photos into"short description", 22 photos into "providing a topic", and 12 photos into"relationship between exhibits". This means that 86 out of the 294 photographs are suitable for content distribution.

During the series of experiments, the visitors freely walked around the space of the exhibition and sometimes browsed a book, took a photograph and added some annotations to the photograph. The robot provided content to the visitors a total of four times. We will now describe some of our findings from our investigation of the interactions among the robot and the participants.

When the robot provided content, most of the participants that were spoken to by the robot looked at the robot (nine out of 11 people), and seven of them looked at their PhotoChat screen and confirmed the distributed content. In addition, five people took a book that was related to the distributed content and browsed the book for a little while, as shown in Fig. 7.

The robot behavior patterns were only designed to distribute content and to talk to a *single* visitor. However, interestingly, more than half (16 out of 26) of the people who were *not* spoken to by the robot also looked at the robot, and 12 of them confirmed the distributed content on PhotoChat. In addition, six out of 12 people moved to the exhibit for which the robot had provided content, and half of them selected a book to browse. This implies that our robot guidance system successfully affects the visitors' experiences of an exhibition by distribution of the content collected from visitors who had been there in the past.

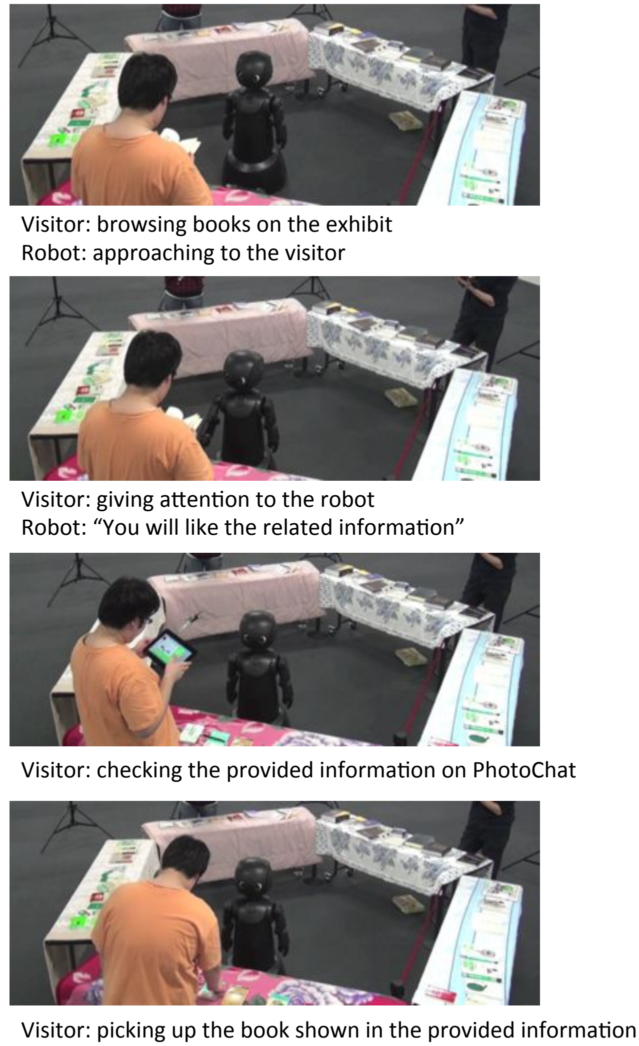

Fig. 7. Typical scene where the robot provided content to a visitor.

We also found an interesting situation where the robot's behavior triggered a real conversation (i.e., not a conversation via PhotoChat) among the visitors in the exhibition space. We observed that the robot's behavior sometimes drew a visitor who was at another exhibit. In this situation, most of the visitors (five out of six) started a conversation with the others about the exhibit. In contrast to the provision of *personalized* information, because a robot has a real body their presentations can be leaked. This kind of leaked information may then lead to interactions among the people that are present around the space. Fig. 8 shows a typical example of this situation.

VII. Conclusion

In this paper, we have proposed a novel way to share visitor experiences and interests in exhibition spaces and have implemented a system using a robot and PhotoChat. Specifically, we developed a system to collect conversations on PhotoChat and distribute them, accompanied by the robot's bodily actions, to other visitors. We found that our robot guidance system successfully affects the visitors' experiences at an exhibition by distributing content that was collected from the visitors who were there in the past. We also found an interesting situation where the robot's behavior triggered a real conversation among the visitors in the exhibition space.

In this study, we performed workshops and experiments in two phases, where the first was for data collection and the second was to investigate the interactions among the robot and the visitors in the exhibition space. We intend to merge these two phases into a system that will automatically classify the collected data and circulate this data from the past to the future.

Fig. 8. Case where the robot's behavior triggered a real conversation in the space.

ACKNOWLEDGEMENT

This work was supported by JSPS KAKENHI Grant Numbers 24118706, 22118505, 22300042, 25730123.

REFERENCES

[1] Y. Sumi, J. Ito, and T. Nishida, "Photochat: Communication support system based on sharing photos and notes," in *CHI 2008 Extended Abstracts*. ACM, 2008, pp. 3237–3242.

[2] W. Burgard, A. B. Cremers, D. Fox, D. Hähnel, G. Lakemeyery, D. Schulz, W. Steiner, and S. Thrun, "The interactive museum tour-guide robot," in *Proceedings of the Fifteenth National Conference on Artificial Intelligence (AAAI-98)*, 1998, pp. 11–18.

[3] H. Asoh, N. Vlassis, Y. Motomura, F. Asano, I. Hara, S. Hayamizu, K. Ito, T. Kurita, T. Matsui, R. Bunschoten, and B. Kröse, "Jijo-2: An office robot that communicates and learns," *IEEE Intelligent Systems*, vol. 16, no. 5, pp. 46–55, 2001.

[4] 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, pp. 61–84, 2004.

[5] M. Shiomi, T. Kanda, H. Ishiguro, and N. Hagita, "Interactive humanoid robots for a science museum," in *ACM/IEEE 1st Annual Conference on Human-Robot Interaction (HRI2006)*, 2006, pp. 305–312.

[6] Y. Kuno, K. Sadazuka, M. Kawashima, K. Yamazaki, A. Yamazaki, and H. Kuzuoka, "Museum guide robot based on sociological interaction analysis," in *Proceedings of CHI 2007*. ACM, 2007, pp. 1191–1194.

[7] 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 CHI 2008*. ACM, 2008, pp. 131–140.

[8] H. Kuzuoka, K. Pitsch, Y. Suzuki, I. Kawaguchi, K. Yamazaki, A. Yamazaki, Y. Kuno, P. Luff, and C. Heath, "Effect of restarts and pauses on achieving a state of mutual orientation between a human and a robot," in *Proceedings of CSCW'08*. ACM, 2008, pp. 201–204.

[9] R. Matsumoto, H. Nakayama, T. Harada, and Y. Kuniyoshi, "Journalist robot system: Robot system making news articles from real world," in *IEEE/RSJ Internatinal Conference on Intelligent Robotics and Systems (IROS) 2007*, 2007, pp. 1234–1241.

[10] Y. Sumi, K. Mase, C. Müller, S. Iwasawa, S. Ito, M. Takahashi, K. Kumagai, Y. Otaka, M. Tsuchikawa, Y. Katagiri, and T. Nishida, "Collage of video and sound for raising the awareness of situated conversations," in *Intelligent Media Technology for Communicative Intelligence (Second International Workshop, IMTCI 2004, Revised Selected Papers), Springer LNAI3490*, L. Bolc, Z. Michalewicz, and T. Nishida, Eds., 2005, pp. 185–194.

[11] Y. Koide, T. Kanda, Y. Sumi, K. Kogure, and H. Ishiguro, "An approach to integrating an interactive guide robot with ubiquitous sensors," in *2004 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS 2004)*, 2004, pp. 2500–2505.

[12] J. F. Kelley, "An iterative design methodology for user-friendly natural language office information applications," *ACM Trans. Inf. Syst.*, vol. 2, no. 1, pp. 26–41, Jan. 1984.