Revealing Gauguin: Engaging Visitors in Robot Guide's Explanation in an Art Museum

Keiichi Yamazaki*, Akiko Yamazaki, Mai Okada*, Yoshinori Kuno*, Yoshinori Kobayashi*, Yosuke Hoshi*, Karola Pitsch***, Paul Luff****, Dirk vom Lehn**** , Christian Heath****

* Saitama University
255 Shimo-Okubo, Sakura-ku, Saitama, Japan
{yamakei, s08cs002, kuno}@mail.saitama-u.ac.jp

** Tokyo University of Technology
1401-1 Katakura-cho, Hachioji, Tokyo, Japan
ayamazaki@media.teu.ac.jp

ABSTRACT

Designing technologies that support the explanation of museum exhibits is a challenging domain. In this paper we develop an innovative approach – providing a robot guide with resources to engage visitors in an interaction about an art exhibit. We draw upon ethnographical fieldwork in an art museum, focusing on how tour guides interrelate talk and visual conduct, specifically how they ask questions of different kinds to engage and involve visitors in lengthy explanations of an exhibit. From this analysis we have developed a robot guide that can coordinate its utterances and body movement to monitor the responses of visitors to these. Detailed analysis of the interaction between the robot and visitors in an art museum suggests that such simple devices derived from the study of human interaction might be useful in engaging visitors in explanations of complex artifacts.

Author Keywords

Interaction analysis, conversation analysis, museum, guide robot, human-robot interaction, computer vision.

ACM Classification Keywords

H5.2. Information interfaces and presentation (e.g., HCI): User Interfaces – interaction styles.

INTRODUCTION

In recent years there have been a variety of initiatives involving the introduction of new technologies into museums. In part these have been motivated by the requirements of museums, having greater numbers of visitors and wishing to open up galleries, science centers and exhibition to a greater diversity of visitors. This has placed demands on one of the traditional ways of providing tailored information to visitors – the tour guide. Technologies, of different kinds, have been deployed to address this need, including 'interactive' displays and mobile technologies such as PDAs. However, there is an increasing recognition that the information that is to be provided needs to be tied not only to the local ecology of the environment but shaped to the ongoing participation and engagement of the visitor. In this paper we consider how a technology can monitor aspects of a visitor's conduct and then produce actions to display that sensitivity and thereby produce engaging explanations.

Our analysis draws on ethnomethodology and conversation analysis and the growing corpus of research that has come to be known as workplace studies. We focus in particular on the interactional and sequential organization of the participants' actions with regard to the conduct of the robot. Our analysis involved the detailed transcription of a significant number of extracts and quantitative summary of different kinds of responses. We are particularly interested in the interrelationship between the speech and the visible conduct of the participants, such as head movements, hand and arm gestures and bodily and facial orientation [2] [5]. Interaction in museums, galleries and science centers [3] have been of interest, particularly as here talk and visual conduct are tied to the local circumstances of the setting.

Perhaps curiously, there has been an increasing interest in museum settings in the field of human robot interaction (HRI). This may be because it is a location in which it is possible to deploy technology and conduct experiments. It may also be because it places demands on the design of the robot. Although the setting can be tightly circumscribed, the robot 'tour guide' needs to interact with a range of participants and needs to be sensitive to changes in how they are participating: how and where they are oriented to and how their conduct is related to objects in the local environment. While most effort in this area has tended to focus on the autonomy of robots enabling them to navigate safely through a museum [1] [15], another line of research has begun to explore the interaction between robot and user [11] [13] [14], gaze during talk [10] or the use of particular communicational patterns such as 'pauses and restarts' as a means to gain a visitor's attention [7].

In our own research, we have shown the effect that a tour

CHI 2009, April 4–9, 2009, Boston, MA, USA.

*** Applied Informatics & CoR-Lab Bielefeld University, 33501 Bielefeld, Germany, karola.pitsch@uni-bielefeld.de
**** WIT Research Centre, King's College, London, UK SE1 9NH, {paul.luff, dirk. vom_lehn, Christian.Heath}@kcl.ac.uk

guide robot - by precise timing of head and body movement at systematic places in the talk (derived from studies of human interaction) - can systematically guide the visitor's attention between the guide and the exhibit. These studies have been carried out as laboratory experiments [6] [17].

In this paper, we have moved from the laboratory to a real museum setting, the Ohara Museum of Arts (Kurashiki, Japan). This change sets particular challenges: visitors come spontaneously and can walk away at any time; also the context of the arts museum (as opposed to a science museum where visitors are likely to have an interest in technology [11] [13]) might be more prone to not engage with the technology. Furthermore, a focus on content and the successful provision of information becomes increasingly important, and ways need to be found for controlling this, such as visitors nodding at relevant places, verbal utterances or whether they indeed listen through the entire explanation.

In what follows we present our findings from a series of studies, in which we have (a) undertaken extensive ethnographic fieldwork on the interaction of human guides and visitors during gallery talks. This has been the occasion to both verify the interactional patterns, which we originally found in the laboratory studies, in a real museum and to discover a range of new features of conduct in the human guide's behavior – such as the use of 'involvement questions' – to help engage visitors in the presentation. (b) These findings (i.e. the systematic coordination of head/body orientation and talk during the 'involvement questions') have been implemented (in addition to our previous results on the precision timing) in our museum guide robot explaining a painting by Gauguin ("Te Nave Nave Fenua"), and we have conducted an evaluation experiment of the system again in the same museum with real visitors.

For this, we have three concerns when analyzing and evaluating our guide robot in the arts museum.

1) Do visitors, who spontaneously participate, listen to the robot's explanation until the end or not?
2) Do visitors respond to the robot's explanation positively and interactively appropriate moments?
3) Do visitors display that they have indeed received some information from the robot or not?

To answer these questions, we first examined how human curators explain paintings to visitors. We then designed the robot and assessed it in the art museum.

ANALYSIS OF INTERACTION BETWEEN HUMAN GUIDE AND VISITORS AT OHARA MUSEUM OF ART

In order to develop our museum guide robot, we have undertaken – in addition to our previous semi-experimental studies [6] [17] – extensive ethnographic fieldwork at the Ohara Museum of Art (Kurashiki, Japan). As part of this, we have videotaped 15 sessions (about 15 minutes each), in which human guides (two experts and two volunteer guides) as part of their normal work explains a painting either to an individual visitor or to a small group of visitors.

Qualitative analysis of these data shows the similarity to the original experimental setting. In the arts museum, the guides use the same head and body coordination when talking about an exhibit as in our previous experimental studies: Human guides coordinate their talk and non-verbal action, most frequently around Transition Relevance Places (TRPs).Guides accompany their verbal deixis and keywords with pointing gestures to the paintings and gaze towards the visitors.

Transcript 1 (Pablo Picasso "Still life with a skull"):

The 1st line represents the speaker's gaze, the 2nd line is the original Japanese wording and the 3rd line gives the English translation.

```
,,,:Gaze movement
 (.):Short pause, (0.9):Pause 0.9sec.

      P--------------,,,,,,,,M1----------------
01 GE1:kono sakuhin paburo pikasono zugaikotuno

      P-------------------------,,,,,,,,,,V,,,,,
02 GE1:aru seibutu toiu sakuhin nan desukeredomo

This work is called Pablo Picasso's "Still Life with a Skull"

      ,,,,,,,,,,,,,P----M1-----------------------
03 GE1:Puburo Pikaso : (.) wa gozonji::deshoka( .)
      Do you know Pablo Picasso?

       M1------------------
04 GE1: dokkade kiita kotowa
      Have you heard of him

       GE1-
05 GE1: Hai
      Yes((nods))
```

From our analysis it emerges that there are ways that guides systematically produce explanations: Particularly at the beginning of an explanation, guides regularly attempt to actively involve listeners by asking a question which tries either (1) to relate to any prior knowledge that visitors might have about the artist, or (2) by hinting at any particular feature that can be seen in the painting under consideration. Such 'involvement questions' can also be found at other places in the guide's talk, e.g. as a means to emphasize a particular aspect in the explanation. With regard to the coordination of head and body orientation guides do indeed face the visitor when asking such questions.

In what follows we will present an in-depth analysis, focusing on the close interrelationship of the guide's head and body movement and the recipient's responses to it.

Transcript (1) shows a typical fragment of such human guide behavior in the arts museum, and more specifically the first type of 'involvement questions' in which the guide asks about the visitor's prior knowledge about the artist. With regard to the organization of gaze, we can see the following:

First, we can identify the same 'precision timing' of the guide's head movement at the TRP to the visitor that we have found in the experimental setting: In the first line, the guide (GE1) explains whilst looking at the painting (P) and towards the end of this sentence (i.e. at the TRP) she turns her head towards the visitor (M1). Second, the guide – when asking a question about the visitor's prior knowledge (here: "Do you know Pablo Picasso", line 02-03) – directs her gaze at the visitor. And the visitor, when answering (here: "yes", line 04), turns his head to the guide.

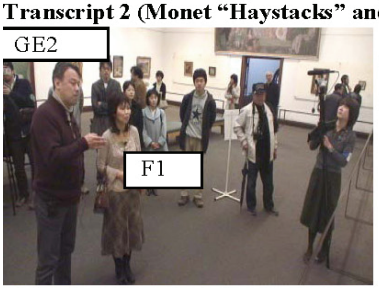
However a further aspect arises in this case. After the fragment, the guide asked the visitor "What country do you think of first?" He answered "Italy." The answer turns out to be wrong. Picasso was not born in Italy, but in Spain. Whilst it is easy for a human guide to simply understand this utterance and then with a range of communicational resources at hand deal with an incorrect answer and in a friendly and face-saving manner, this sets a range of further challenges for a robot. Not only would robust voice recognition be required, but also a large knowledge base as opposed to a mere pre-recorded explanation.

In the other type of 'involvement questions', in which the guide asks about elements that can be found in the painting itself (type 2), the coordination of gaze and talk is slightly different. As can be seen in Transcript 2, the guide while asking the question ("Which of the two paintings is painted with a wider landscape?" line 04) looks at the painting, and at the end of this turns his head towards the visitor. The visitor not only looks at the painting while the guide is looking at it, but also his gaze remains fixed on the painting when the guide turns his head to the visitor at the TRP. This shows that the visitor when answering tries to find resources in the picture for providing an answer. Also the guide looks at the painting when he finally gives himself the correct answer; and the visitor does not turn his head to the guide, but he keeps looking at the painting inspecting what the guide is talking about.

Our analysis shows that not only do guides systematically look at different places while giving the answer to the two types of 'involvement questions', but also so do the visitors.

When the question refers to the painting (type 2), visitors answer by pointing to it or gazing at it. In the case of questions about prior knowledge of the painting (type 1), visitors answer the question by gazing at the guide.

Transcript 2 (Monet "Haystacks" and "Water-lilies"):



P_1:Right painting,

P_2:Left painting,

```
                                                F1:P1-----
                                               GE2:,,,,,,,
01 GE2:((Waves hand twice towards picture))Ikinari
                                                   I'll

 F1:,,,,,,,,,,,P2------,,,,,,,,,P1-
GE2:F1-------------------------,,
02  ijiwaruna situmon  simasune=
    start with a "mean"(=tough)question.

 F1:,,,,,P2
GE2:P----
03  (1.0)

F1:P-----------------------------------,,,P2----
GE2:,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,F1---
04  =Docchinohoga((puts hands out))hiro::i basho
    =Which of the two (paintings) is painted with

 F1:P2-P1-------P2----
GE2:F1--------------
05  ga kaite arimasu
    a wider place (=landscape)?

 F1:,,P1----
GE2:,,,,,P-
06  (0.9)
                                      F1:P1------
                                     GE2:F1------
07  F1:((Points to the right picture))Kochira
                                         This one.
                                     F1:P1,,,,,,,,
                                     GE2:P1-------
08  GE2:((Points to the right picture))Kochira?
                                         This one?


                        F1:P2----------
                       GE2:P2----------
09  F1:((While nodding))[Hai
                            [Yes
10  GE2:                [((Nods))
```

From these findings about guide-visitor interaction in museums, we can derive five interaction patterns for an advanced design of a museum guide robot:

1) As suggested in our previous research, the robot should turn its head at the sentence end (TRPs) from the painting to the visitor. (TRP is the first possible completion of a first turn-construction unit [12] and is not a synonym for sentence end. However, since explanation in museums is basically composed of sentences, we consider each sentence end as TRP.)

2) In addition, especially to engage visitors in lengthy explanations, the robot could make use of 'involvement questions'.

3) When making an 'involvement question', the robot should turn its head from the painting to the visitor at the end of the question (TRP), which is similar to declarative sentences.

4) Type 1 questions (about prior knowledge): While asking the question, the robot should gaze at the visitor. During the visitor's answer it should keep looking at the visitor. When the robot gives the answer, it should look at the visitor, and expect that the visitor looks at the guide. Currently, however, due to unforeseen and potentially wrong answers that visitors might give, robots cannot handle this kind of question. Hence we don't use Type 1 questions.

5) Type 2 questions (about the painting): While asking the question, the robot should gaze at the painting and at the end turn its head to the visitor. During the visitors answer the robot should keep looking at the visitor. When giving itself the answer, the robot should look at the visitor and accept that the visitor looks at the painting. This is what we have implemented in the robot.

As can be seen from the framegrabs inserted with the above transcripts, these analyses currently refer to situations in which one guide interacts with one visitor – a configuration which happens frequently at least in Japanese museums. However, explanations – whether by a museum guide or some other form of teacher or information source – have often multiple recipients. Currently, we are doing analysis of situations with multiple visitors as well, and preliminary analysis of such cases shows that the guide's gaze has an additional function: it is also relevant for selecting which visitor might answer the question. One function of gaze is to monitor the visitors' behavior with regard to questions and responses (similar to one-to-one). A second function is to select the next speaker as reported by [12] and [8]. In this case, most frequently it is the last visitor who receives the guide's gaze who answers the question. With this mechanism, a similar situation as in the one-to-one setting is created, and the regularities presented above are deployed.

MUSEUM GUIDE ROBOT SYSTEM

When deploying a guide robot in an art museum the system designers need to be sensitive to the specific circumstances in which the technology will be situated. Not only has the robot to be robust and reliable but also features like laser range sensors that could damage valuable paintings have to be avoided. Furthermore, the robot's behavior in the gallery has to be susceptible to people's conduct in the museum.

System Overview

Figure 1 gives an Overview of our guide robot system. We use a humanoid robot Robovie-R Ver.2 (ATR) which is developed as a research platform for human-robot communication. The robot is 120 cm high and it is able to move by wheels installed in the bottom. It can move its head and arms by controlling its joints (29 degree of freedom in total including fingers). Its head including eye camera and ear microphone moves along three axes (Yaw, Roll and Pitch) like a human head. Arms with fingers can be controlled in a wide area so that the robot can perform pointing gestures. Inside its body, there are two general purpose PCs (Windows XP) and all joints are connected to the main PC by serial communication. Three USB cameras (Logicool Inc. Qcam) are attached on the top of the pole installed on the back of the robot to acquire scenery images around the robot by 320x240 pixel size, irrespective of the movement of the robot's head. The PCs inside the robot are connected with a wired network and a wireless network (wireless LAN Draft IEEE802.11n) to communicate with outside PCs. The inside PCs can be monitored and controlled by the outside PCs while the robot gives an explanation.

Our system consists of two software units: one is the head detection and tracking unit and the other is the robot control unit. Each of two PCs runs each unit independently. The PC which runs head detection and tracking unit employs a frontal USB camera to detect a human's head and track its positions and directions. The result of detection or tracking is sent to the robot control unit and is used if necessary. Both units communicate with each other by using Socket with asynchronous mode. The robot control unit controls all joints by serial commands and employs USB camera (right or left, depends on the position related to the painting) to localize the position itself.

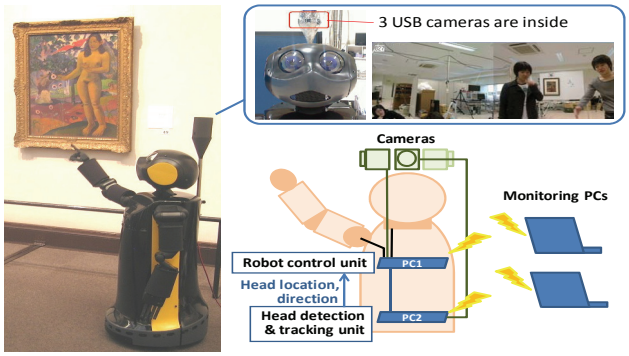


Figure 1. Overview of our guide robot system.

Sensing Techniques based on Camera Images

All of these functions, detecting visitor's face, tracking visitor's head position and direction, localizing robot's position, are performed based on the images acquired by the cameras on the top of the pole. Here, we describe each sensing technique.

Detecting Visitor's Face

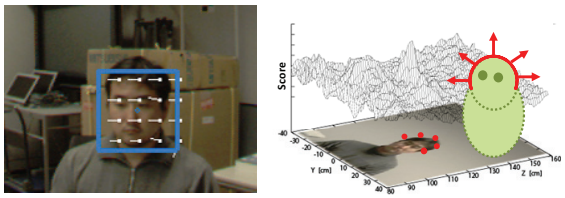
Our robot starts to give an explanation to the visitor when s/he expresses an interest in the robot. Therefore the head detection and tracking unit tries to find a listener by detecting front of the faces of visitors. The cascaded classifiers based on AdaBoost and Haar-like features proposed by Viola et al. [16], which are implemented as the

face detector in the OpenCV library, are used to detect a visitor's face. When multiple faces are found in the image, the system selects the largest face as the listener assuming that the nearest person may be most interested in the robot. Very small faces will not be selected even if the detected face is the only one.

Tracking Visitor's Head

Our robot explains a painting while addressing its gaze towards the listener. Therefore the listener's head has to be tracked continuously during the explanation. A particle filter [4] is used for the listener's head tracking. It is difficult to find the listener's head by using the face detector when his/her head is not facing the camera. In case of using multiple detectors to detect multiple directions of the head yields other difficulties such as calculation cost or establishing correspondence between frames. Thus we employed a tracking framework.

Here we describe only key points of our tracking method, because it is beyond the scope of this paper. In the particle filter framework, a prediction step and an evaluation step are important for the tracking performance. We employed the optical flow for estimating the head position in the next time step. We computed optical flow based on the block matching algorithm for 16 small regions, each of which is divided from the listener's face region. We employed the average vector of 16 flow vectors as the prediction model (Figure 2(a)). We evaluated hypotheses based on the contour orientation similarity [9]. We calculated the average of the inner products of contour gradients and normal vectors at 5 points on the contour of the upper head model. The distribution of evaluation scores is shown in Figure 2(b). Figure 2(b) shows that our evaluation method has the peak around the ground truth of a head position.



(a) (b)

Figure 2. Head tracking based on particle filter, (a) optical flow for motion estimation (b) evaluation score based on contour orientation similarity.

Observing Visitor's Head Direction

Our system measures the listener's head direction during an explanation to terminate the explanation based on the information whether the listener is looking at the painting or not. The system extracts the skin color pixels and hair color pixels and estimates the rough direction of the listener's head based on the distributions of those pixels.

First, we assume that the hair color of the listener is dark. The color space of the image is changed into HSV color space and the image region is extracted a little larger than the tracking region to ensure that the entire head image is indeed captured. Then, we project the skin color pixels and dark color pixels onto the horizontal axis. This projected pixel distribution is used for estimating the head direction. The system previously learned the distributions of 8 head directions. While tracking the listener's head, the system extracts the tracking region a bit larger and computes the distribution. Bhattacharyya distance is used to compare between the previously learned distribution and the distribution of the tracked region. The nearest direction is used as listener's head direction. The distribution of skin color pixels and dark color pixels is shown in Figure 3.

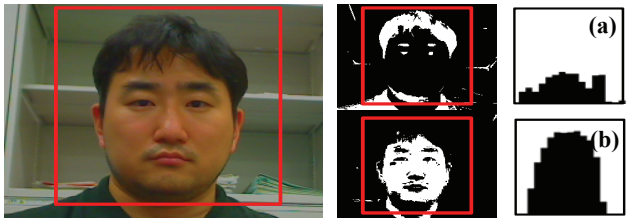


Figure 3. Pixel projections for measuring a head direction, (a) dark color pixels (b) skin color pixels.

Localizing Self-position

A self localization technique is required when the robot moves to welcome a listener. The relative location of the robot from a painting (here we assume a painting is the target exhibit) can be measured by observing the 4 corners of the painting detected by using a template matching algorithm.

Robot's Behavior

Here, we describe the robot's behavior in an explanation procedure. Our system has a set of three voice samples which consist of a female voice, male voice and synthesized voice. The robot can switch between them by changing its setting.

The motions of the robot turning its gaze toward the listener or pointing at the painting are previously recorded with their exact timings. So the robot simply plays back the prerecorded data. However, the gaze direction when the robot turns toward the listener is adaptively tuned. A relationship between the head location in the image and the gaze direction is previously learned. Therefore the robot can turn its gaze toward the listener's head direction precisely even if the listener moves. This function is effective to attract the listener. When the robot starts to give an explanation, it turns its head precisely toward the listener while saying "May I explain this painting to you?" This may attract the listener to hear an explanation.

Our robot continuously measures the listener's head direction during the explanation so that the robot can change its behavior depending on the situation. For instance, after the robot has emitted an utterance like "May I explain this painting to you?" the listener's gaze is toward the painting or camera (this means robot's direction): the system evaluates that the listener is highly interested in the painting or the robot. Then, the robot moves toward the listener to welcome him/her and move back to its original position to start the explanation. On the other hand, if the

listener's gaze is not directed to the painting or camera, the system evaluates that the listener is not interested in the painting or the robot. Then the robot gives up explaining and waits for another listener to come. Or, when the listener's gaze is not toward the painting while saying "This woman is probably Eve" with pointing gesture, the robot makes an utterance again like "This woman". In this way our robot is able to behave these actions.

EXPERIMENT AT OHARA MUSEUM OF ART

We have conducted an experiment with our museum guide robot on November 17, 2007 at the Ohara Museum of Art, well known for its collection on European masterpieces and attracting more than one million visitors every year. In this experiment, the robot explains a painting by Gauguin, "Te Nave Nave Fenua".

The robot is set up as follows: The robot shows its availability for communication by slowly turning from left to right in front of the painting. When the robot detects the face of an approaching visitor, it turns its head towards him/her, saying, "May I explain this painting to you?" gesturing towards the exhibit. Then, if the robot finds that the visitor's face is still facing either towards the painting or the robot, the robot starts the explanation. After finishing the explanation, the robot returns to its waiting mode, again slowly turning from left to right.

For this experiment, we disabled the robot's locomotion function to ensure the safety of participants. In addition, we programmed the robot to complete an explanation once it had been started even though the visitor might leave during the explanation. We did this to obtain various data under the same condition.

Body Movement of Robot's Explanation

In this section, we describe the action and utterances that the robot was programmed to perform/play during the experiments. These are based on the actual curator's talk and body movements when explaining this painting to visitors. For this, we videotaped the curator explaining this particular painting and determined how and when the robot would move its head, body and arms in relation to the talk.

The Script gives the details of the robot's verbal utterances, and in relation with the robot's bodily behavior (120 sec. in total). TRPs 1 to 6 are declarative sentences, in which the robot talks about the painting. TRP 7 is the 'involvement question' and TRPs 8 and 9 are the answers given to it by the robot. TRPs 10 and 11 again are declarative sentences, in which the robot continues to explain about the painting. As the robot at the current state of development is not able to detect whether the visitor indeed produces an answer or not (after TRP 7) or what this answer might be, we could only introduce a pause which should be long enough so that some kind of visitor's response could occur in it. Also, the concrete pauses in the robot's talk have been determined with regard to the length and complexity of the previous sentence, in relation to the preliminary user tests and technical constrains of the robot's hand/arm movements.

Script: Robot's verbal and bodily behavior for explaining Paul Gauguin "Te Nave Nave Fenua".

```
This is a famous work of Gauguin. (TRP1)

(3.2 sec. pause)

Gauguin, who was at a loss with Western
civilization, went to Tahiti in 1891. (TRP2)

(2.1 sec. pause)

He painted pictures that sought to answer
 philosophical questions, such as 'What is human?'
 'What is civilization?' (TRP3)

(1.1 sec. pause)

One of those pictures is this painting, 'Te Nave
Nave Fenua' (TRP4)

(5.0 sec. pause)

Gauguin has said this "When Eve in the garden of
Eden timidly plucked an evil flower, the wings of
the monster bird Chimera fluttered and hit her on
the temples."(TRP5)

(7.2 sec. pause)

This woman is probably Eve; right? (TRP6).

(1.0 sec. pause)

Do you know which is the evil flower and which the
monster bird Chimera? (TRP7: question)

(4.0 sec. pause)

The evil flower is the orange-ish one that Eve is
plucking; right? (TRP8: Answer 1)

(1.0 sec. pause)

And the monster bird Chimera is this creature with
red wings; right? (TRP9: Answer 2)

(4.0 sec. pause)

In Tahiti, Gauguin came into contact with the
tropical sun, the people who lived there, and
various myths of the spiritual world. (TRP10)

(1.1 sec. pause)

Don't you think Gauguin's feelings -- as if he was
in a daze and used a lot of strength -- remain in
a pure state in this painting? (TRP11)
```

With regard to the coordination of talk and body movements, the single underline shows the moment when the robot starts raising its hand for pointing and then brings it back to its previous position. The double underline indicates that the arm starts rising, but not to perform a pointing action. Figure 4 gives an impression of what this looks like when being implemented with the robot.

Furthermore, the robot turns its head (gaze) from the painting to the visitor shortly before each TRP (TRP 1 to 11). The robot's head stops facing the visitor at the TRP. When the next utterance starts, the robot again turns its

Figure 4. Pointing gesture of the robot.

head towards the painting. This movement can be seen in Figure 5 (Transcript 3).

Transcript 3: A part of Paul Gauguin (Te Nave Nave Fenua):

P: painting, V: robot gaze towards visitor

---: gaze towards the exhibit ,,,,: transfer of gaze

1: the robot keeps its gaze towards the painting. 2: the robot turns its head towards the visitors. 3: the robot keeps looking at the visitor. 4: the robot turns its head to the painting. 5: the robot keeps its gaze towards the painting.

1 2 3 4 5

P----------,,,,,,,,,,,,,,,,,V------,,,,,,,,,,,,,,,,,,,,,,,,P----

Gauguin no issaku desu（silence）
Work of Gauguin

Figure 5. TRP1: Robot turns its head towards the listener while saying "Issakudesu" / "work".

Analyzing Guide Robot Experiment (1): Does the Audience Listen to the Robot's Explanation?

In our experiment we made the robot perform its explanation of the Gauguin painting about 95 times. In these explanations, about 500 visitors looked at the performance, the average number of participants in the audience being 4.8 persons. Some of these explanations were not carried out under good conditions due to the tuning of the robot set up, interruptions by staff or media interviews. Therefore, we have 63 cases with well conditioned data, in all of which the visitors join the explanation spontaneously.

From these 63 cases, 38 are "Full Cases" in which the listener attends to the robot's explanation from the beginning to the end. There are 6 "Combined Full Cases", in which the listener joins the robot's explanation in the middle of an ongoing explanation and stays until the middle of the next explanation to have an entire run through the information. In 6 cases the robot fails to start the explanation; and in some cases the robot detects a visitor who remains beyond the end of the explanation and starts anew. In a few other cases the robot recognizes a person who is just passing by in front of the robot as a visitor. There are 4 cases, in which the explanation is not completed because the listener is interrupted by someone who calls the listener. The remaining 9 cases are failures to complete the explanation without any external factors.

Omitting the cases of the robot's misrecognition and external factors, 53 cases remain that can be included into our detailed analysis. In these, 71.7% of the listeners spontaneously attend the explanation as "Full Case", and 83% of listeners attend the robot's explanation as "Full Case" or "Combined Full Case". Considering the duration of the explanation (120 seconds), this result suggests that our guide robot is able to successfully engage visitors – who are not particularly interested in technology – to pay attention to a robot's explanation.

Analyzing Guide Robot Experiment (2): Audience Responses

We have analyzed the visitors' responses when the robot turns its gaze at the TRPs. In this, we have focused on the recipients' behavior such as the visitor looking at the robot (R), nodding (N), smiling, gazing toward the painting or pointing at the painting (P), or emitting some kind of utterance (U). (We count the responses of the visitors from the robot head turning towards the visitors at a TRP to the robot head turning back to the painting.)

The percentage of each of these responses occurring at the TRPs is shown in Table 1. The heading "R+N+P+U" means that the listener's behavior includes a combination of multiple responses at the same time. We can see from the Table 1 that more than 50% of listeners in average respond to the robot's gaze motion.

	R	N	P	U	R+N+P+U
Average	30.3%	21.0%	11.2%	13.1%	53.1%

Table 1. Rate of each response occurring at TRPs.

The number of responses occurring at the TRPs in all examples of "Full Case" is shown in Figure 6. We can see from this that there is a great variety in the amount of responses that occur. In fact, there are 7 cases in which the listener responds only one or two times. In contrast, there are 20 cases (it is more than a half of the valid explanations) that the listener responds more than 6 times from a total of 11 TRPs.

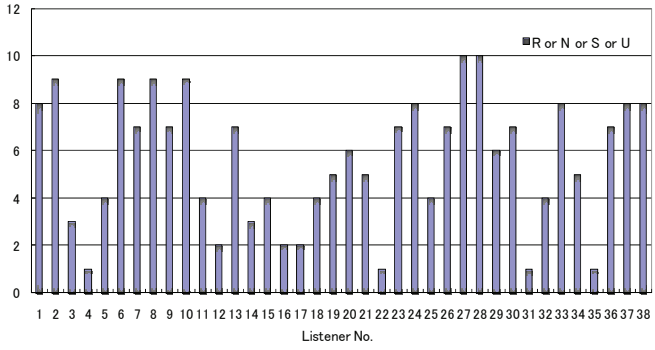


Figure 6. Number of responses occurring at TRPs for each listener.

Furthermore, we have focused on those cases in which the listener rarely responds. Close inspection shows that this is due to the following reasons: the robot turns its gaze towards the listener imprecisely (No.31); the listener leaves the robot's field of view (No.17); during the explanation the robot recognizes another visitor as recipient instead of the

original target listener (No.4, No.35); and the listener is concentrating on the painting (No.12, No.22).

The participants of all 38 cases consist of 12 male visitors, 22 female visitors, 2 boys and 2 girls. In the case of the visitor responding up to 2 times, 3 males and 4 females responded. In contrast, in the case of the listener responding more than 6 times, 4 males, 13 females, 1 girl and 2 boys can be found. From this, we can see the trend that female visitors seem to respond to the robot's motion more than male visitors. On the other hand, we cannot find any difference between cases in which the robot uses human voice vs. synthesized voice.

Analyzing Guide Robot Experiment (3): General Engagement of Audience

We have analyzed the details of 31 cases (excluding those cases in which the listener responds less than 2 times) with regard to the visitor's response at each TRP. Figure 7 shows the ratio of the listener responding at each TRP. We can see from this that about 60% of listeners respond to the robot at every single TRP. This result supports our previous results in the laboratory environment [6] [17].

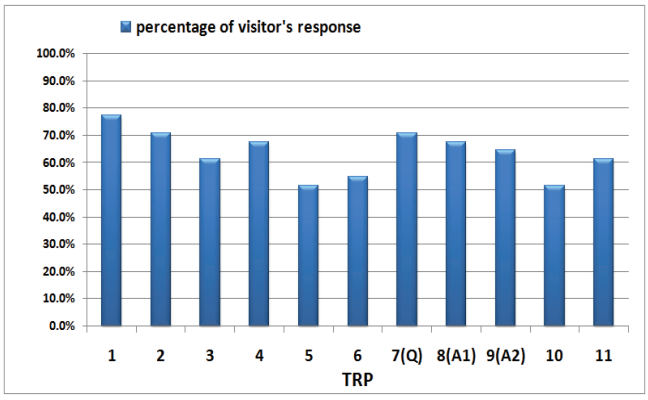


Figure 7. Rate of responses occurring at TRPs.

Figure 8 shows the occurrence ratio of four responses in which the listener does a particular motion, such as looking at the robot (R), nodding (N), smiling, gazing toward the painting or pointing at the painting (P), or making some kind of utterance (U). We can see from Figure 8 how the visitor performs the response in relation to the TRP.

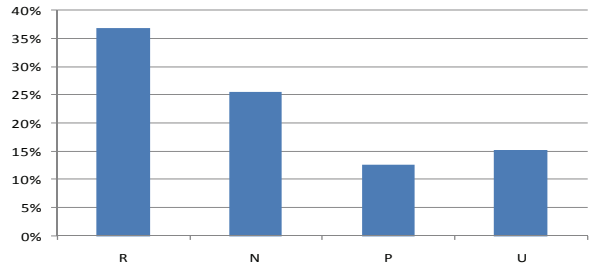


Figure 8. Occurrence ratio of four responses.

The 'nodding' response (N) that occurs at the TRPs can be observed often in both environments. The 'utterance' response occurring at TRPs can be observed more often in comparison to the laboratory environment. Recipients' behavior such as a smiling, gazing towards the painting or pointing at the painting can also often be observed. This way, we can see that the recipient's active engagement – measured in terms of visible behavior such as nodding, pointing, smiling or doing some kind of utterance – is a consequential effect of the robot's gaze movement at TRPs. We can conclude from this that our museum guide robot is able to successfully engage visitors by using appropriate and natural gaze behavior.

Analyzing Guide Robot Experiment (4): Effectiveness of Question and Answer Sequence

Figure 9 presents the four different categories of the visitors' responses towards the robot (R = visitor looks to robot; N = visitor nods towards robot; P = visitors' positive response like smiling, leaning forward to the painting, pointing at the painting; U = visitor emitting some kind of utterance).

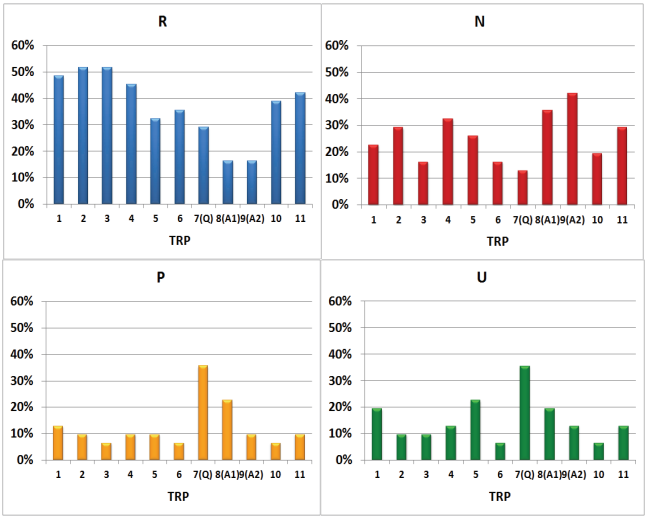


Figure 9. Percentage of each response occurring at the TRPs.

Of particular interest to us is the visitors' behavior with regard to the 'involvement question' (TRP 7), where the robot asks "Do you know which is the evil flower and which the monster bird Chimera?"

A) Visitors' responses following the question (TRP7)

Figure 9 shows that the visitors' utterances (U), pointing gestures etc. (P) increase at TRP7. Figure 10 lists the visitors' bodily behavior and utterances following the robot's question (TRP7).

No. 09 ((points to the painting)) Is it so?
No. 10((shakes his head))
No. 11 ((talks to her mother while pointing at the painting)) I don't know
No. 24 I don't know ((laughing))
No. 26 ((nod))
No. 27 Yes, I do.
No. 28 ((talks to her companion while pointing at the painting))
No. 30 I don't know which is which
No. 33Yes ((nodding))
No. 36 There? ((shaking his head))
No. 37((shakes his head))
No. 38 ((nods))

Figure 10. Visitors' answers during 4.0 sec. silence after TRP.

This close examination shows that visitors respond to the robot's question appropriately and interactively. We can conclude from this that the robot's action – asking an 'involvement question' and at the same time inviting the visitor to look at the painting by appropriately coordinated head and body movement – engenders appropriate, interactive and attentive responses from the visitors.

B) Visitors' response after robot's correct answer (TRP 8 and TRP9)

At TRP8 (Answer 1) and TRP9 (Answer 2) the visitors' nodding increases (Figure 9). Since the robot gives two answers – "evil flower" (TRP8) and "chimera" (TRP9), visitors frequently nod at both places. Figure 11 shows the relationship between the visitors' head turning and gazing towards robot. +R+N signifies that the visitor nods and gazes towards the robot. (-R)+N signifies that the visitor nods without gazing towards the robot.

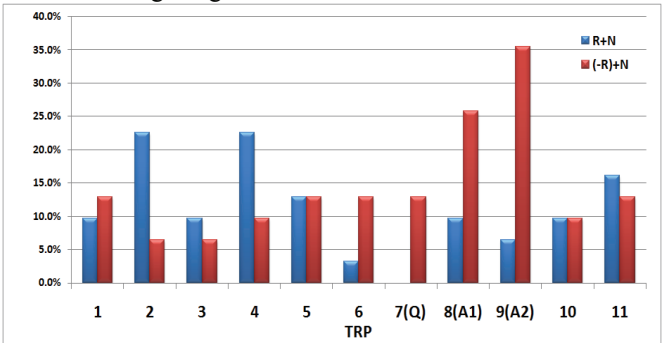


Figure 11. Percentage of each response occurring at the TRPs.

The ratio of the visitors' nods at TRP 8 (A1) and TRP9 (A2) is distinctive in Figure 11. At other TRPs, visitors' nods are related to the visitor's directing their gaze towards the robot as in the laboratory experiment. At TRP8 (A1) and TRP9 (A2), however, visitors nod without turning their gaze towards to the robot (-R+N). Visitors' nods at TRP8 and TRP9 do not show the sign of visitors listening to the robot's explanation, but this means that visitors display their understanding for the robot's correct answers when they can link it to the corresponding place in the painting. This shows that our robot meets our third standard for evaluation.

DISCUSSION

Based on previous semi-experimental studies and new ethnographical studies in an arts museum, we can derive a range of interaction patterns for designing guide robot:

1) As suggested in our previous research, the robot should turn its head at the sentence end (TRPs) from the painting to the visitor.

2) In addition, especially to engage visitors in lengthy explanations, the robot could make use of 'involvement questions'.

3) When making an 'involvement question', the robot should turn its head from the painting to the visitor at the end of the question (TRP), which is the same as with declarative sentences.

4) From the two types of 'involvement question' currently only one 'question about painting' can be implemented into the robot: While asking the question, the robot should gaze at the painting and at the end turn its head to the visitor. While the visitors answer the question, the robot should keep looking at the visitor. When giving itself the answer, the robot should look at visitor and accept that the visitor looks at the painting.

The implementation of these interaction patterns in our research robot and its exposure to an extensive trial in an arts museum shows the following results:

1) We have been able to confirm the results from our previous laboratory studies in a real arts museum: Similar to the laboratory findings, visitors indeed respond to a robot's explanation by nodding at TRPs.

2) Beyond this, we have found that in 83 % of the cases (n= 53, omitting all external factors), visitors stay from the beginning to the end of the robot's explanation. This seems to show that the current robot is an effective means to engage visitors and successful in delivering a museum guide presentation.

3) With regard to the newly introduced questioning sequence, visitors respond to the robot by not only nodding but also giving verbal answers. Although currently being designed to model one-to-one interaction, the robot performed as well robustly with multiple participants. In such circumstances, it is regularly the visitor being looked at by the robot (at this stage, the robot cannot differentiate between different visitors yet), who respond to the robot's question.

Despite this success, we would like to point out the following limitations of the technology which have come to light in the experiments and which need addressing in future work:

1) The robot occasionally repeated its explanation to the same person if he/she stayed after the end of the talk. In such cases, the visitor soon went away. This problem arises because the robot's actions depend on the recognition of the existence and direction of human faces. This problem could be solved by adding face recognition to the system.

2) The robot occasionally failed to gaze (turn its head) at the chosen visitor when s/he moved fast or some other person blocked this visitor from the robot's view. In such cases, the visitors' responses decreased. To solve this problem, we need to improve the face-tracking module. In addition, we need to further consider multiple visitor cases as they become an issue in the real museum setting.

To systematically deal with multiple visitors is our next research target. In the current study, the robot is set up to continuously look at the first visitor who is tracked and originally starts the explanation. In our field experiments, however, more than one visitor gathered around the robot. We currently have begun to study human guide behavior with regard to multiple visitors, and preliminary analysis shows that guides seem to have two different ways of gaze behavior: (1) gazing at a particular visitor and (2)

continuously changing their gaze between different participants. In our current robot system we have implemented the first behavior, which – as the study here presented shows – is an appropriate means to engage visitors in a rather lengthy (120 sec.) explanation. Human expert guides, however, use both gazing patterns depending on the particular circumstances of the situation. Our next steps consist in deepening and systematizing our analysis of human guide behavior with regard to multiparty settings. Based on this, we will extend our robot system to robustly handle multi-party interaction in the museum setting.

CONCLUSION

The action of 'explaining' is an important aspect of communication in various contexts, ranging from everyday life through informal learning situations to educational settings. Amongst the range of new technologies developed over the last years to support this need, robots offer a particularly promising way as they can not only tie the information that is provided to the local ecology of the environment, but also they can monitor the visitors' actual conduct and then shape their presentation with regard to his/her ongoing participation and engagement. In this paper, we have presented findings of a design of robot that makes use of precisely this advanced communicational potential: A robot which – by closely coordinating its verbal and non-verbal actions – is able to explain a painting in complex form to visitors in an arts museum. During this, as the analysis shows, the robot is not only able to interact with the human visitors, but also to engage them over a longer period of explanation displaying recipiency and understanding at relevant moments.

On a methodological level, we have deployed a method for fine-grained analysis of the sequential interplay of visual and verbal conduct in interactional settings as a basis to derive patterns of conduct from human interaction as a model for designing robots. With this, our robot presents a platform for researchers to experimentally investigate interaction.

ACKNOWLEDGMENTS

This work was supported in part by Japan-UK Bilateral Joint Projects (JSPS, The British Academy), the Ministry of Internal Affairs and Communications under SCOPE, Grant-in-aid for Scientific Research (KAKENHI 19203025, 19300055, 20700152) and JSPS New Research Initiates for Humanities and Social Sciences. We thank the Ohara Museum of Art and Hideaki Kuzuoka and the anonymous reviewers.

REFERENCES

1. Bennewitz, M., Faber, F., Joho, D., Schreiber, M. and Behnke, S. Towards a humanoid museum guide robot that interacts with multiple persons. In *Proc. HUMANOIDS 2005*, (2005), pp.418-423.
2. Goodwin, C. Action and embodiment within situated human interaction. *Journal of Pragmatics*, 32, (2000), 1489-1522.
3. Heath, C. and vom Lehn, D. Configuring reception: (Dis-) regarding the 'Spectator' in museums and galleries. *Theory, Culture & Society*, 21(6), (2004), 43-65.
4. Isard, M. and Blake, A. Condensation - conditional density propagation for visual tracking. *Computer Vision*, 29(1), (1998), 5-28.
5. Kendon, A. *Gesture: Visible action as utterance,* Cambridge University Press, (2004).
6. Kuno, Y., Sadazuka, K., Kawashima, M., Yamazaki, K., Yamazaki, A. and Kuzuoka, H. Museum guide robot based on sociological interaction analysis. In *Proc. CHI 2007*, ACM press (2007), 1191-1194.
7. Kuzuoka, H., Pitsch, K., Suzuki, Y., Kawaguchi, I., Yamazaki, K., Yamazaki, A., Kuno, Y., Luff, P. and Heath, C. Effect of pauses and restarts on achieving a state of mutual orientation between a human and a robot. In *Proc. CSCW 2008*, (2008), 201-204.
8. Lerner, G.H. Selecting next speaker: the context-sensitive operation of a context-free organization. *Language in Society*, 32(2), (2003), 177-201.
9. Matsumoto, Y., Kato, T. and Wada, T. An occlusion robust likelihood integration method for multi-camera people head tracking. In *Proc. INSS 2007*, (2007), 235-242.
10. Mutlu, B., Hodgins, J.K. and Forlizzi, J. A storytelling robot: modeling and evaluation of human-like gaze behavior. In *Proc. HUMANOIDS 2006*, (2006), 518-523.
11. Nomura, T., Tasaki, T., Kanda, T., Shiomi, M., Ishiguro, H. and Hagita, N. Questionnaire-based social research on opinions of Japanese visitors for communication robots at an exhibition. *AI and Society*, 21(1), (2006), 167-183.
12. Sacks, H., Schegloff, E. and Jefferson, G. A simplest systematics for the organization of turn-taking in conversation. *Language*, 50(4), (1974), 696-735.
13. Shiomi, M., Kanda, T., Koizumi, S., Ishiguro, H. and Hagita, N. Group attention control for communication robots with wizard of OZ approach. In *Proc. HRI 2007*, (2007), 121-128.
14. Sidner, C.L., Lee, C., Kidd, C.D. and Rich, C. Explorations in engagement for humans and robots. *Artificial Intelligence*, 166(1), (2005), 140-164.
15. Thrun, S., Bennewitz, M., Burgard, W., Cremers, A.B., Dellaert, F., Fox, D., Hhnel, D., Lakemeyer, G., Rosenberg, C., Roy, N., Schulte, J., Schulz, D. and Steiner, W. Experiences with two deployed interactive tour-guide robots. In *Proc. FSR 1999*, (1999).
16. Viola, P. and Jones, M. Rapid object detection using a boosted cascade of simple features. In *Proc. CVPR 2001*, 1, (2001), 511-518.
17. Yamazaki, A., Yamazaki, K., Kuno, Y., Burdelski, M., Kawashima, M. and Kuzuoka, H. Precision timing in human-robot interaction: coordination of head movement and utterance. In *Proc. CHI 2008*, ACM press (2008), 131-140.