Ishino *et al. RPTEL* (2022) 17:1
https://doi.org/10.1186/s41039-021-00176-6

Research and Practice in Technology Enhanced Learning

RESEARCH Open Access

Robot lecture for enhancing presentation in lecture

Tatsuya Ishino[1], Mitsuhiro Goto[2] and Akihiro Kashihara[1*]

*Correspondence: akihiro.kashihara@inf.uec.ac.jp
[1] The University of Electro-Communications, 1-5-1, Chofugaoka, Chofu, Tokyo 182-8585, Japan
Full list of author information is available at the end of the article

Abstract

In lectures with presentation slides such as an e-learning lecture on video, it is important for lecturers to control their non-verbal behavior involving gaze, gesture, and paralanguage. However, it is not so easy even for well-experienced lecturers to properly use non-verbal behavior in their lecture to promote learners' understanding. This paper proposes robot lecture, in which a robot substitutes for human lecturers, and reconstructs their non-verbal behavior to enhance their lecture. Towards such reconstruction, we have designed a model of non-verbal behavior in lecture. This paper also demonstrates a robot lecture system that appropriately reproduces non-verbal behavior of human lecturers with reconstructed one. In addition, this paper reports a case study involving 36 participants with the system, whose purpose was to ascertain whether robot lecture with reconstruction could be more effective for controlling learners' attention and more beneficial for understanding the lecture contents than video lecture by human and robot lecture with simple reproduction. The results of the case study with the system suggest the effect of promoting learners' understanding of lecture contents, the necessity of reconstructing non-verbal behavior, and the validity of the non-verbal behavior model.

Keywords: Devices for learning, Human–Robot interaction, Non-verbal behavior, Robot presentation, Robot lecture

Introduction

Recently, small communication robots such as Sota (Vstone Co. Ltd., 2010), Robohon (Sharp Corporation, 2016), NAO (Softbank Robotics Co. Ltd., 2018), and PALRO (FUJISOFT Inc., 2010) have become widespread in various contexts such as nursing care, education, and guidance service. There has been also an increasing interest in utilizing these robots, especially in the field of education. In this paper, we focus on using communication robots for small class lectures and e-learning lectures on video.

In a lecture, it is generally important to present the lecture contents as slides with oral explanation so that learners' understanding could be promoted. This requires lecturers to control the attention of learners to slides and oral explanation by means of gaze, gesture, paralanguage, etc., which are viewed as non-verbal behavior (Collins, 2004). If lecturers want to attract learners' attention to an important point in a slide, for example, they should direct their face to it, and point it out with pointing gesture in concurrence

Springer Open

© The Author(s), 2022. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

with its oral explanation. On the other hand, excessive and unnecessary non-verbal behavior would prevent learners from keeping attention to understanding the lecture contents. It is accordingly indispensable to properly use non-verbal behavior in lecture presentation (called lecture behavior) (Ishino et al., 2018).

However, it is not so easy even for well-experienced lecturers to continue making proper use of lecture behavior during their lecture presentation. If lecturers are inexperienced, in addition, they tend to focus on oral explanation prepared in advance without any non-verbal behavior. Learners would accordingly have difficulties in keeping their concentration, and finish the lecture with incomplete understanding.

Towards this issue, this paper proposes robot lecture, in which a communication robot substitutes for human lecturers. The main purpose of robot lecture is to reproduce their own lecture behavior as appropriately as possible with their lecture contents, and to reconstruct their improper and insufficient behavior for enhancing their lecture presentation. In order to make it possible, we have also designed a model of how lecturers should conduct lecture behavior to promote learners' interest and understanding (Ishino et al., 2018). As for lecture behavior, it is important for lecturers to conduct it not at random but according to their intention (Arima, 2014). The lecture behavior model represents the relationships between lecture intentions and non-verbal behavior to be used for controlling learners' attention and promoting their understanding.

We have also developed a robot lecture system so far, which deals with face direction, and gesture (without paralanguage) as lecture behavior. This system records the presentation made by human lecturers to detect and reconstruct inappropriate or insufficient behavior by following the lecture behavior model (Ishino et al., 2018). The robot reproduces the reconstructed presentation, which could appropriately convey the lecture contents, control learners' attention, and promote their understanding. In our previous work (Ishino et al., 2018), we confirmed that the system could keep learners' attention more effectively. We conducted a case study with the system. The results obtained from questionnaires suggest that gaze with face direction, and pointing gesture reconstructed by the robot are more acceptable and understandable in terms of keeping and guiding attention than non-verbal behavior in video lecture by human. Most of the participants also felt that their concentration on the lecture contents could be promoted with eye contact by the robot.

In this paper, we refine the robot lecture system, which can reconstruct lecture behavior including paralanguage. This paper also reports another case study whose purpose was to confirm whether the robot lecture promotes learners' understanding. In this study, we compared three conditions: video lecture conducted by human, robot lecture simply reproducing the original one, and robot lecture reconstructing the original one. The results suggest that the reconstructed robot lecture significantly promotes learners' understanding of the lecture contents more than the video lecture and the simply reproduced robot lecture.

This paper is organized as follows. Section "Presentation in lecture" outlines presentation in lecture. Section "Robot lecture" describes robot lecture involving the model of lecture behavior. The robot lecture system is minutely described in Sect. "Robot lecture system." Section "Case study" and "Discussion" describe about the case study with the system. Conclusions and suggestions for future work are presented in Sect. "Conclusion."

Presentation in lecture

Non-verbal behavior in lecture

Small class lectures in university or e-learning lectures on video are often conducted with presentation slides, which represent the contents lecturers intend to present. These slides include illustrations, graphs, and keywords. Lecturers explain not only the slides, but also the contents that are not explicitly represented in the slides. This suggests that the lecture contents consist of lecture slides and oral explanation.

In making presentation, it is important for lecturers to attract learners' attention to either slides or oral explanation to promote their understanding by utilizing non-verbal behavior such as gaze, pointing, and pitch/volume of paralanguage. For example, it is possible for lecturers to hold eye contact with learners to attract their attention to oral explanation. It is also possible to face to, point at, and intensively explain an important point in a slide to control learners' attention to it.

On the other hand, it is not beneficial to confuse learners or disrupt their concentration using excessive and unnecessary non-verbal behavior. It is accordingly important to properly use lecture behavior. Melinger and Levelt (2005) confirmed that a speaker often used a hand gesture according to his/her intention. They argued that the speaker intended to complement his/her oral contents with it. Arima (2014) found that skillful teachers conducted more intentional gaze behavior in their class than novices. Goldin-Meadow and Alibali (2013) found that speakers often utilized gesture in communication so that they could promote communication partner's understanding. Such related work claims the proper use of lecture behavior, and also suggests the necessity of intentionally conducting lecture behavior.

Problems

It is not necessarily easy for inexperienced lecturers to intentionally use non-verbal behavior to control learners' attention in lecture presentation. In addition, it would not be easy even for experienced lecturers to continue properly conducting lecture behavior during their presentation. There are also some lecturers who tend to fix their eyes on PCs without any gaze behavior. In such cases, learners could not keep their concentration and interest in lecture. As a result, they would finish the lecture with incomplete understanding.

Related work

There is a lot of work on non-verbal behavior in interaction between human users and robot, whose main intentions are to attract their attention and to promote their understanding (Witt et al., 2004). Huang et al. (2014), Liles et al. (2017) and Admoni et al. (2016) confirmed that robot gestures contributed to understandability and recall performance. These results suggest that gestures are effective for understanding and retaining lecture content. Tanaka et al. (2017) also suggest that when driving a car with a robot as a navigator, attention control by robot gestures is more effective than only voice or screen agent. Sauppé et al. (2014) confirmed that the robot successfully directed the attention of the collaborators to the object using a pointing gesture. In addition, Kamide et al. (2014) confirmed that a humanoid robot could attract audiences' attention to particular

position using non-verbal behavior. These results suggest that non-verbal behavior of robot is effective in gathering audiences' attention in presentation. Mutlu et al. (2007) also proposed a gaze model for storytelling robot, and evaluated the effectiveness of the model-based behavior of the robot for telling a fairy tale to audiences. According to this result, the audiences tended to recall all of the tale contents when the number of having eye contacts with the robot was moderate. On the other hand, they tended to have difficulties in recalling the contents when they had a lot of eye contacts.

These findings from the above related work suggest that it is necessary to appropriately control non-verbal behavior to prevent learners from their incomplete understanding of the lecture contents (Belpaeme et al., 2018). In this paper, we aim to substitute a robot for human lecturers in actual lectures.

Robot lecture

Lecture behavior model

We have introduced robot lecture whose purpose is to enhance lecture behavior of human lecturers with a communication robot. In robot lecture, they are required to prepare oral and slide contents to make lecture presentation. Their lecture behavior is then enhanced by the robot. It needs making it clear how to use lecture behavior. We have accordingly designed a model of lecture behavior with reference to related work on non-verbal behavior (Kamide, 2014, Mutlu, 2007, McNeill, 1994, Goto & Kashihara, 2016).

It can be useful for lecturers to consider aligning their non-verbal behavior with their intentions in lecture, which could be determined with learning states of learners. In this work, we divide the states into the following four:

Learning states

- *State 1*: Not listening to lecture presentation,
- *State 2*: Listening to lecture presentation,
- *State 3*: Noticing important points of the lecture contents, and
- *State 4*: Understanding the lecture contents.

Lecturers would intend to change learning states from state 1 to 4. We accordingly define lecture intention as changing learning states, and classify it into three as follows (Ishino et al., 2018):

Lecture intentions

- *Intention 1* (from *state 1* to *2*): Encouraging learners to get interested in lecture presentation,
- *Intention 2* (from *state 2* to *3*): Encouraging learners to pay attention to and get an understanding of important points in lecture contents, and
- *Intention 3* (from *state 3* to *4*): Encouraging learners to understand the details of the lecture contents.

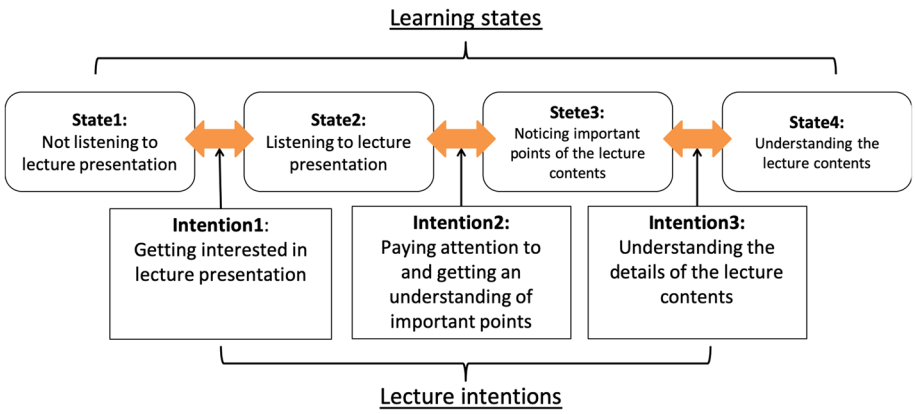


Fig.1 Relationships between learning states and lecture intentions

Figure 1 shows the relationships between learning states and lecture intentions. Lecturers intend to conduct lecture behavior to change learning states from state 1 to 4. There are two contexts in determining lecture intention. First, lecturers dynamically determine their intention depending on learning states, which could also change during lecture presentation. Second, they assume learning states in advance when they prepare their presentation. In video lecture, lecture intention is usually determined in the second context, which we presume in this work.

Let us explain lecture behavior corresponding to each lecture intention in the following.

Lecture behavior for intention 1 In order to help learners get interested in lecture presentation, it is necessary to give them an impression that lecturers talk to them, and to attract their attention to the presentation. For example, making eye contact with learners increases the impression. It is also possible to attract learners' attention by means of multimedia such as sound effects or visual effects and lecturers' over-actions.

Lecture behavior for intention 2 According to the findings from related work mentioned in Sect. "Related work", a communication robot can use gesture and gaze to control learners' attention to an important point in lecture contents that it wants them to concentrate on and understand it. As shown in Fig. 2, for example, a lecturer could induce learners to pay their attention to the slide by gazing at it, and also induces them to focus on his/her oral explanation and gesture by gazing at them.

Lecture behavior for intention 3 In order to help learners understand the details of lecture contents, lecturers need to explain and convey important points of the contents. In this case, it is effective to utilize gestures for making them conspicuous. McNeil (1994) have classified such gestures often used for communication into the following three:

- *Deictic* Gestures expressing important points such as pointing.
- *Metaphoric* Gestures expressing order or magnitude such as counting on fingers and moving hands up to down.

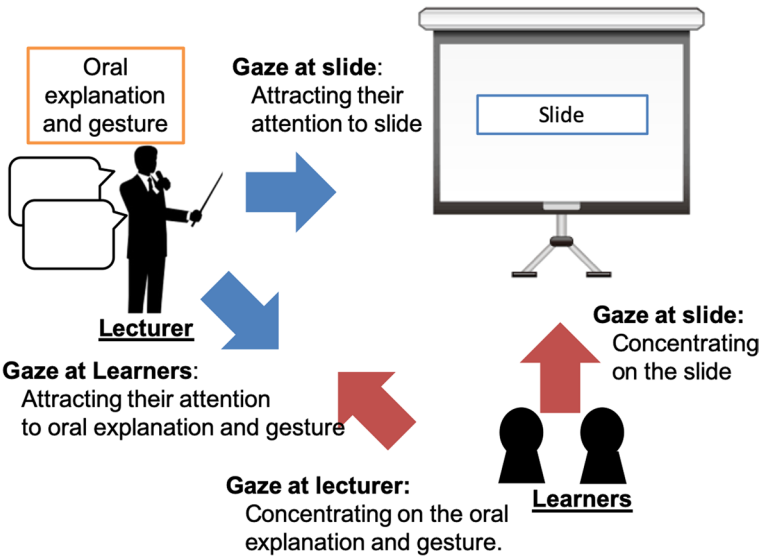


Fig. 2 Examples of attention control

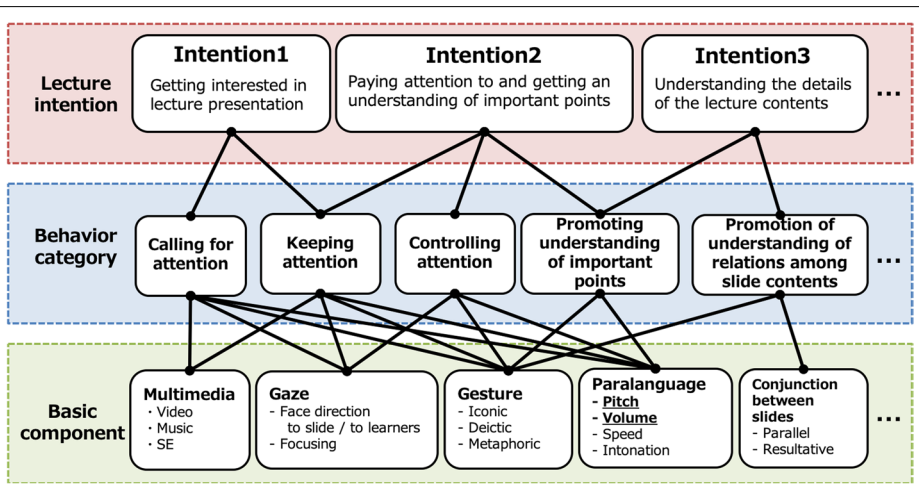


Fig. 3 Lecture behavior model

- *Iconic* Gestures expressing size and length such as drawing shape with both hands.

In this work, lecturers are expected to use these gestures classified by McNeil during their lecture presentation, when they want to convey important points of the slide contents.

Referring to these lecture behaviors, we have designed a model of lecture behavior for reconstructing inappropriate or insufficient lecture behavior conducted by human lecturers as shown in Fig. 3. The model is composed of three layers, which are lecture intention, behavior category, and basic components of lecture behavior. It derives lecture behavior appropriate to each lecture intention from the relationships among them.

When lecturers have the intention 2, for example, the model suggests the necessity of non-verbal behavior for *keeping attention*, *controlling attention*, or *promoting understanding of important points* as behavior category. If they select *controlling attention*, the model induces them to select and combine the corresponding basic components to conduct behavior such as facing to the slide with deictic pointing gesture.

Model-based presentation reconstruction

The robot lecture aims to appropriately reproduce lecturers' non-verbal behavior in their presentation. Related work has been taking two approaches towards reproduction of non-verbal presentation behavior with robot. One is to manually tag their own non-verbal behavior, which is used for the robot to reproduce (Vstone Co. Ltd., 2018, Nozawa et al., 2004). The other is to follow oral explanation to automatically tag non-verbal behavior by means of machine learning methods (Nakano et al., 2004, Ng-Thow-Hing et al., 2010, Le & Pelachaud, 2011). However, the manual tagging is not so easy for lecturers. In addition, non-verbal behavior tagged is similarly reproduced by robot even if the corresponding non-verbal behavior conducted by individual presenters are slightly different.

In the robot lecture, on the other hand, the robot attempts to reproduce non-verbal behavior of lecturers by keeping their presentation individuality (timing and duration) as much as possible, and then to reconstruct their inappropriate or insufficient behavior with appropriate one to be derived from the lecture behavior model. The presentation reproduction with reconstruction is done as follows.

Lecturers are first expected to set their own lecture intention according to a learning state to be assumed when they prepare lecture presentation. We currently assume video lecture in which lecturers have learners in the learning state 2 with the lecture intention 2. The learning state and lecture intention are also supposed to be unchanged during lecture presentation. Second, the robot records lecturers' presentation including lecture behavior and slide/oral contents, and detects important points which they want to emphasize in their slide/oral contents.

The robot next analyzes whether their lecture behavior for conveying the important points detected is included within the one the lecture behavior model can combine for the lecture intention set by them, and whether it is appropriately conducted in accomplishing the intention. This means diagnosing the sufficiency and appropriateness of lecture behavior conducted by the lecturers. If their lecture behavior is not included within the model, it is diagnosed as insufficient. In this case, the robot reconstructs it with appropriate behavior to be derived from the model. If the lecture behavior is inappropriate as for arm angle, face direction, etc., it is reconstructed with desirable angle or direction. The detail of the diagnosis procedure is described in the next section.

The robot then reproduces the lecture presentation with reconstructed behavior. The robot has fewer joints than human, and its movement is limited. In order to appropriately reproduce lecture behavior, we have accordingly converted human lecture behavior into robot one.

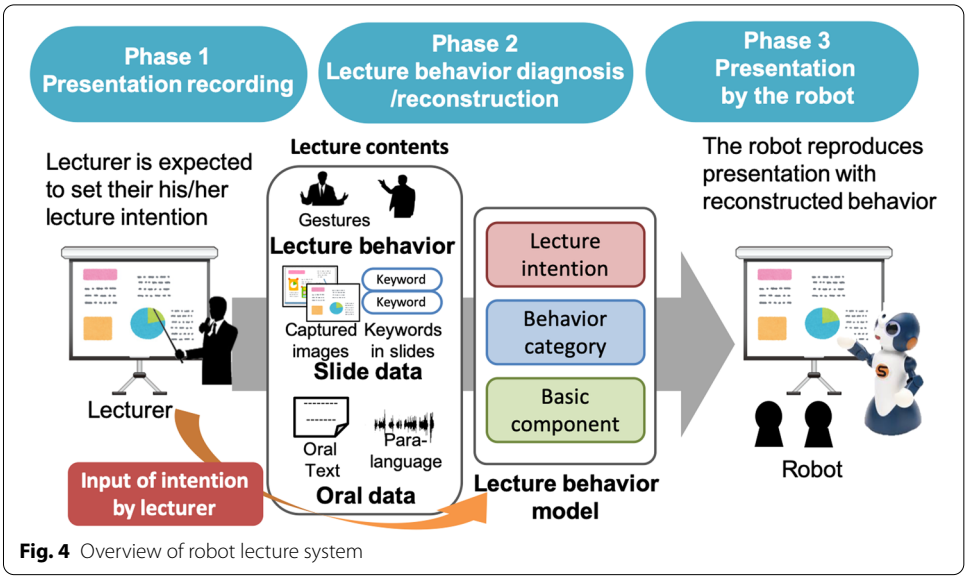


Fig. 4 Overview of robot lecture system

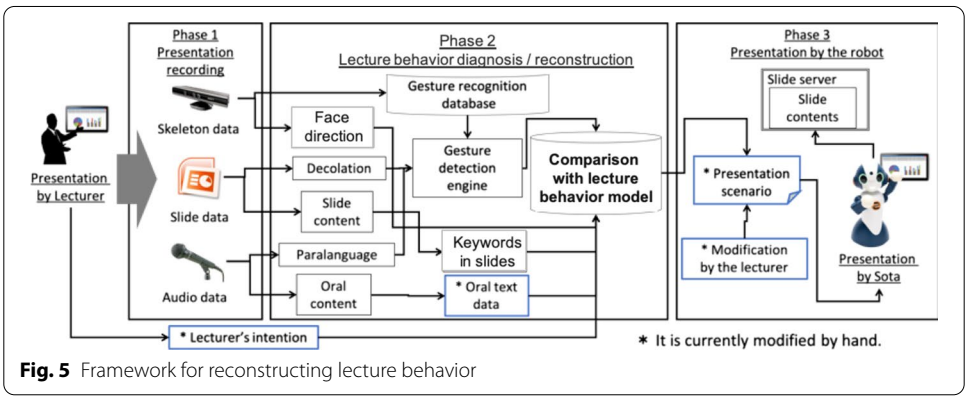


Fig. 5 Framework for reconstructing lecture behavior

Robot lecture system

Overview

In order to reconstruct lecture behavior conducted by human, we have developed the robot lecture system. Figure 4 shows an overview of the system. The system records gestures as skeleton data, slide images, and oral explanation as audio data. The system next detects the important points in the lecture contents to diagnose the lecture behavior by following the lecture behavior model, and reconstructs insufficient or inappropriate behavior diagnosed. The system then reproduces the lecture presentation with the recorded lecture contents and the reconstructed lecture behavior. We currently use Sota as the robot, which is produced by Vstone Co.,Ltd.

Framework for reconstructing lecture behavior

As shown in Fig. 5, this system implements the substitution of lecture presentation by the robot through the following three phases.

- *Phase 1* Presentation recording.
- *Phase 2* Lecture behavior diagnosis/reconstruction.
- *Phase 3* Presentation by the robot.

In phase 1, slide data, slide transition timing, and oral explanation (audio) data are recorded. Gestures of human lecturers during presentation are also recorded as skeleton data using Kinect (Microsoft Corporation). In phase 2, the system analyzes slide data and audio data to detect the important points. Using the results, gestures obtained from the skeleton data are then diagnosed. We currently deal with face direction, pointing gesture, and paralanguage as lecture behavior to be diagnosed, which are necessary to keep/control attention and convey important points. If the system diagnoses lecture behavior as insufficient or inappropriate one, it is reconstructed with appropriate one. In phase 3, the robot reproduces synchronously the presentation with the reconstructed behavior, captured images of the slides, and oral explanation. The oral explanation is obtained from the speaking audio data, to which Text-To-Speech engine converts the text recognized from the recorded audio data.

Presentation recording

In the presentation recording phase, the system records the lecturers' skeleton data including face direction and gesture via Kinect, and records the audio data via an external microphone at the same time. The system also obtains captured images of the slide data and transition timing of each slide via PowerPoint API, and extracts slide text data, decoration data such as character color and size from the slide data. The captured images are uploaded to the slide server that the robot can connect via the Internet, and the robot presents them to learners as the lecture slides. Since all of the recorded data are retained with timestamps, the robot can reproduce presentation behavior and oral explanation that are synchronized at the timing of the captured images presented.

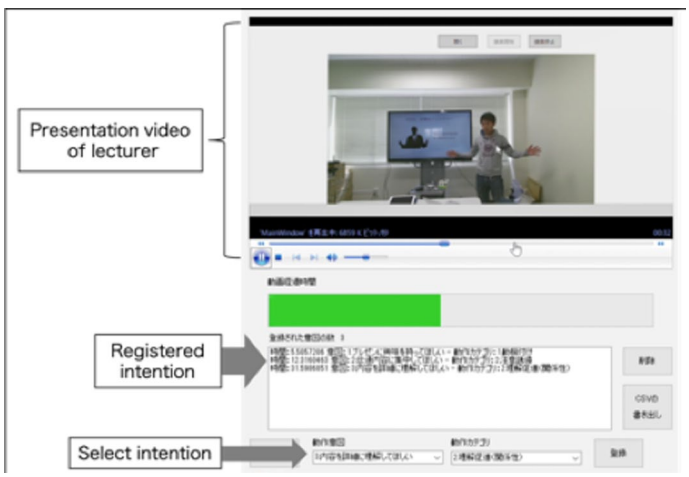


Fig. 6 User interface for intention setting

Presentation behavior diagnosis/reconstruction

As shown in Fig. 6, lecturers first set their intention and behavior category while watching the presentation video by themselves. Let us here describe the procedure for analyzing lecture behavior from the recorded data to reconstruct it as appropriate one.

(a) *Behavior analysis*

In this system, skeleton data of lecturers are recorded in time series. Then, gestures conducted by the lecturers are recognized from the skeleton data using Visual Gesture Builder (Microsoft). Visual gesture builder is a tool to create a gesture recognition database by means of machine learning. This tool allows the system to detect specific gestures from time series of the recorded skeleton data with the recognition database. In constructing the database, we select the section of gesture that we want to recognize, and tag it. The tag represents the gesture in the recorded skeleton data. For example, we can construct a recognition database for pointing gesture by selecting the sections from time series of the recorded skeleton data corresponding to pointing, and by tagging as pointing. According to the gestures classified by McNeill (1994), we currently constructed recognition databases for pointing gesture, expressing of counting gesture, expressing of size gesture, and face direction. We particularly constructed 10 databases which include three for pointing gesture (high, middle, low), two for counting gesture (1st, 2nd, 3rd), two for size gesture (big, small), and two for face direction (at slide, at learner) In order to construct these databases, we prepared 10 short presentation data including skeleton data in our laboratory, and selected 100 sections per each gesture, which we wanted to recognize.

The system compares the intentions set by lecturers with the gestures recognized with these databases. If the system does not recognize the gestures corresponding to the intentions, it detects them as insufficient gestures. If the system recognized the corresponding gestures with inappropriate direction, it detects them as inappropriate gestures.

(b) *Slide analysis*

The system extracts text and decorated data such as character color/form from the slide data, and detects the important points. As shown in Table 1, the system weights the decoration data in four degrees from 0 to 3. As for weights of font color, it is necessary to change depending on the slide theme and lecturer preference. If

Table 1 Weighting of importance

Text decoration			
Font color	**Importance**	**Form**	**Importance**
Red	3	Underline	2
Blue	3	Italic	2
Black	0	Bold	2
Other	1		

the weight of the decorated data exceeds 3, it is regarded as an important point. In the case of multiple decorations of the data, the system sums up the weight of each decoration. If the weight exceeds 3, it becomes an important point.

(c) *Audio data analysis*

The recorded audio data is converted into text by voice recognition. The recognition rate is about 50% in speech recognition. It is difficult to completely transform lecturers' oral explanation (audio) data into text by speech recognition. We accordingly modify the transformation results by hand. In addition, we use Praat (Boersma & Weenink, 2018), which is a free software for speech analysis in order to obtain paralanguage. It can obtain the values of pitch (voice high and low) and volume (strength of voice) with its timestamp. Currently, the system obtains the values of pitch and volume of each sentence in each slide, and calculates the maximum value of pitch and intensity in each slide. It also detects the sentence whose value of pitch and volume exceeds 80% of the maximum value as emphasized point.

(d) *Diagnosis/reconstruction*

The system diagnoses insufficient or inappropriate points, and reconstructs the behavior while comparing the keywords in the slide contents and the ones in the oral contents to detect the corresponding ones as important points in the lecture contents.

Here are some examples of reconstructing lecture behavior. When lecturers explain an important point in a slide detected in the slide analysis, they should use gazing, paralanguage or pointing to the slide to attract learners' attention to it. If they do not conduct the non-verbal behavior in this case, the system reconstructs it with face direction or pointing behavior. At the same time, paralanguage is also reconstructed by increasing the value of pitch and volume of oral explanation in the important point. When lecturers explain with the oral contents, they should also gaze at learners. If they have shifty eyes or looks at PC display in this case, the system reconstructs their behavior with the one for facing to them. In this way, it is possible to appropriately convey lecture contents to learners with reconstructed behavior, even if lecture behavior conducted by lecturers is insufficient or inappropriate.

Presentation by robot

In this phase, the system controls Sota and the display connected to the slide server by means of presentation scenario generated through the two phases of presentation recording and presentation behavior diagnosis/reconstruction. The reconstructed behavior, slide number, and oral explanation data are managed by time in the presentation scenario. It includes the behavior data (basic components recognized, start timing and duration), the text data for oral explanation (contents of explanation and paralanguage parameters) and slide number data. According to time, for example, the robot performs a gesture of pointing downward if the gesture is "Pointing at low". When the face direction is "To Learner", the robot faces towards the learner, and in the case of "To Slide" it turns to the slide.

Sota has a total of 8 freedom degrees of joint rotary (body 1 axis, arm 2 axis, shoulder 2 axis, neck 3 axis). Joints of Sota are fewer than joints of human lecturers. Sota also has no fingers. Therefore, we convert human behavior for Sota. As shown in (a) and (b) of

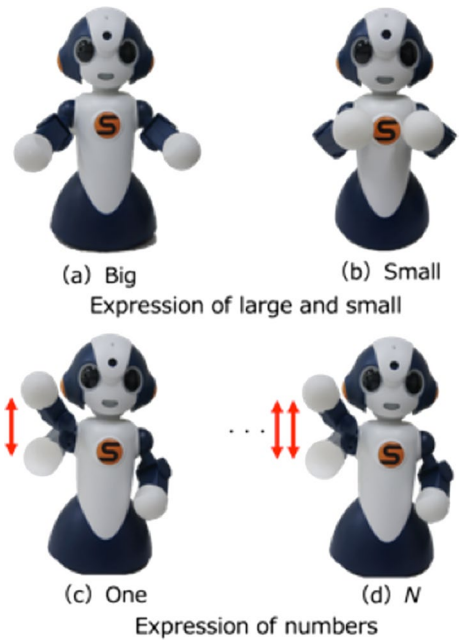


Fig. 7 Expressions of lecture behavior by Sota

Fig. 7, for example, Sota represents big and small as iconic gesture by opening and closing the arm in front of its body.

In conducting the robot lecture in an actual lecture, the system sends the behavior and oral explanation data to Sota, and slide number data to the slide server. Sota reproduces the presentation with the behavior, and the slide server presents the captured image corresponding to the oral explanation synchronously. The reconstructed behavior is converted into the behavior to be reproduced within the joints of Sota. Sota's oral explanation is also converted from the text data via NTT's Text-To-Speech engine.

Case study

Design

As this work assumes e-learning video lectures and small class lectures to be attended by a few students, we conducted a case study whose purpose was to ascertain whether robot lecture with reconstruction could be more effective for controlling learners' attention and more beneficial for understanding the lecture contents than video lecture by human and robot lecture with simple reproduction. By comparing the robot lecture between reconstruction and simple reproduction, it is possible to confirm the validity of reconstruction using the lecture behavior model. By comparing the reconstructed robot lecture with the video lecture, we can also confirm the advantages of robot lecture regardless of lecturer appearance.

Preparation

Participants were 36 graduate and undergraduate students. As shown in Table 2, we prepared three video lectures whose topics were *learning model*, *social learning*, and *learning technology*, which were recorded from lectures by the same lecturer who was one

Table 2 Details of video lectures

Lecture topic	# of slides	Presentation time
Learning model	11	5 min 12 s
Social learning	12	5 min 59 s
Learning technology	12	5 min 55 s

of the authors. These lectures had the almost same numbers of slides and were about 5 to 6 min. We also prepared three robot lectures that reconstructed the corresponding lectures by following the lecture behavior model, and three robot lectures that simply reproduced the corresponding lectures without reconstruction. The reconstructed lecture behaviors were gestures, face orientation, and paralanguage. We set three conditions:

(a) Robot-Reconstruction condition,

Lecture by robot involving reconstruction

(b) Robot-Reproduction condition, and.

Lecture by robot involving simple reproduction.

(iii) Video condition.

Video lecture by human lecturer

In the following, we describe the details of reconstructed lecture behavior in the Robot-Reconstruction condition. Table 3 shows the numbers of gestures reconstructed, which included face direction and pointing gesture. In the lecture topic of *Learning model*, the system added one new gesture, and modified three gestures recorded. In *Social learning*, the system added no gesture, and modified three gestures recorded. In *Learning technology*, the system added no gesture, and modified four gestures recorded. The system also deleted no gesture in all of the lecture topics. Since the gesture and voice recognition of the system are not perfect, we manually checked each lecture presentation to add new gestures after gesture reconstruction by the system. In this manual checking, we looked into each slide to identify important points embedded in figures/illustrations that were not covered by the current system, and added pointing gestures to them according to the lecture behavior model. It took about 15 min for each lecture. As for appropriate gestures that was not reconstructed, there were two gestures in *Learning*

Table 3 Numbers of gestures reconstructed

Lecture topic	Appropriate gesture	Adding gesture	Modifying gesture	Deleting gesture	Author-added gesture
Learning model	2	1	3	0	3
Social learning	7	0	3	0	4
Learning technology	2	0	4	0	2

Table 4 Average numbers of paralanguage for emphasis per slide

Lecture topic	In Robot-Reproduction condition	In Robot-Reconstruction condition
Learning model	2.7	1.7
Social learning	2.5	2.5
Learning technology	2.5	1.6

model and *Learning technology, and* seven gestures in *Social learning*. Since the timing of the reconstructed gesture is not synchronized, we made modifications by hand.

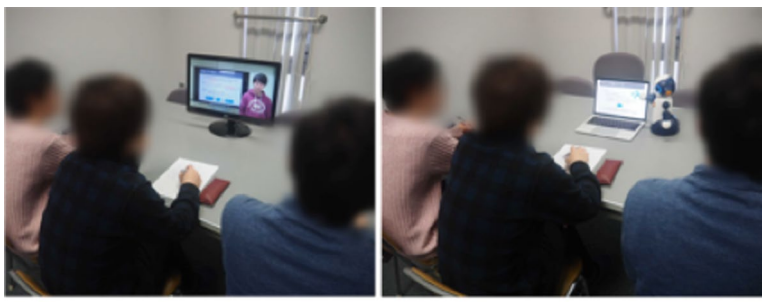
Table 4 shows the details of paralanguage for emphasis in the Robot-Reproduction and Robot-Reconstruction conditions. The values in Table 4 represent the average numbers of paralanguage for emphasis per slide. In *Learning model* and *Learning technology*, the numbers in the reconstruction condition were less than in the reproduction condition because the system deleted the inappropriate paralanguage. In *Social learning,* the number in the reconstruction condition was the same in the reproduction condition because there was no inappropriate paralanguage.

Procedure

As within-participant design, each participant took the three lectures under the three conditions. In order to counterbalance the order effects of the conditions, we randomly

6 participants in each group (Total: 36 participants)	Lecture topic		
	Learning model	Social learning	Learning technology
Group 1	Robot-Reconstruction	Robot-Reproduction	Video
Group 2	Robot-Reconstruction	Video	Robot-Reproduction
Group 3	Video	Robot-Reconstruction	Robot-Reproduction
Group 4	Robot-Reproduction	Robot-Reconstruction	Video
Group 5	Video	Robot-Reproduction	Robot-Reconstruction
Group 6	Robot-Reproduction	Video	Robot-Reconstruction

Fig. 8 Procedure for taking lectures



(a) Taking video lecture (b) Taking robot lecture

Fig. 9 Examples of taking a lecture by participants

assigned 36 participants to six groups as shown in Fig. 8. For example, Group 3 first took the lecture of *learning model* under the Video condition, then took the lecture of *social learning* under the Robot-Reconstruction condition, and took the lecture of *learning technology* under the Robot-Reproduction condition. Figure 9 shows how the participants took the video lecture and robot lecture.

After taking each lecture, the participants were required to have an understanding test as objective evaluation including three in-slide questions and three between-slides questions. Since lecture behavior for encouraging learners to pay attention to important points is conducted for individual slides, it is expected to have a direct effect on understanding the slides. We accordingly used the in-slide questions to evaluate the direct effect, which were about the contents to be answered from individual slides. In addition, we assumed that the lecture behavior would also indirectly have an effect on understanding the relationships between slides. This indirect effect seems to play a crucial role in understanding the whole of lecture contents. In order to evaluate it, we also used the between-slides questions, which were about the contents to be answered from the relationship between multiple slides. Each question was scored one point (The perfect score of the test was six points). An answer consisting of multiple elements was scored as partial points by dividing one point. For example, if it consisted of two elements, each had 0.5 points.

After the understanding test, the participants were required to answer a 7 Likert scale questionnaire as subjective evaluation that asked the following 11 questions from 4 viewpoints (Table 5). Q1 to Q4 were about understandability. Q5 to Q7 were about concentration. Q8 and Q9 were about gazing. Q10 and Q11 were about motivation. The participants were required to answer on a scale of 1 to 7 (1: Extremely disagree < 4: Neither agree nor disagree < 7: Extremely agree) in each question. They were also required to write the reason why they selected in Q1, Q5, Q7, Q8 and Q9.

The hypotheses we set up in this study were as follows:

> **H1** Robot lecture involving reconstruction promotes understanding of the lecture contents including the slide contents and the relationships between slides more than robot lecture involving simple reproduction, and.

Table 5 Details of question items

Q1	How easy was it to understand this lecture overall?
Q2	How easy was it to notice timing that you should pay attention to in this lecture?
Q3	How easy was it to notice the contents that you should pay attention to?
Q4	How easy was it to notice the important points that you should pay attention to?
Q5	How easy was it to concentrate on the presentation under this condition?
Q6	How much was it not to distract your concentration by means of face direction and pointing gesture?
Q7	How easy was it to keep taking the lecture for a long time?
Q8	How much was it to perceive that the lecturer was speaking to you?
Q9	How much was it to perceive that the lecturer made eye contact with you?
Q10	How much was it to want to take the lecture again?
Q11	How much was it to get interested in the lecture?

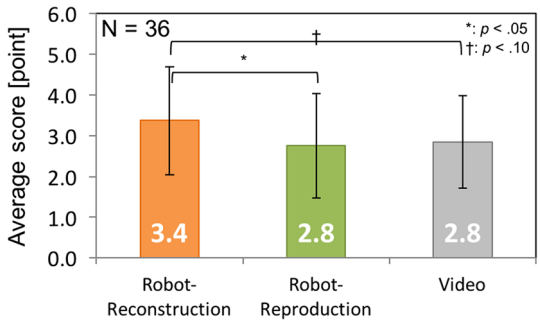


Fig. 10 Average scores of understanding test

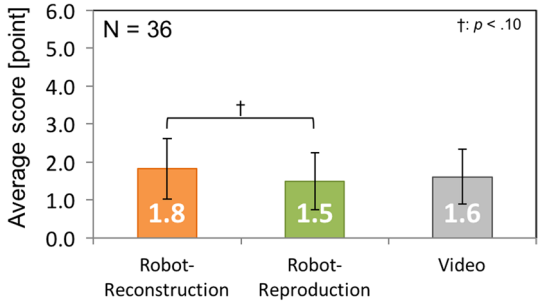


Fig. 11 Average scores of in-slide question

H2 Robot lecture involving reconstruction promotes understanding of the lecture contents including the slide contents and the relationships between slides more than lecture video.

Results and considerations

Statistical analyses of the three conditions were performed using one-way analysis of variance (ANOVA), with the Tukey–Kramer test for post hoc comparisons when significance was determined by ANOVA.

Objective evaluation: understanding test

Figure 10 shows the average scores of the understanding tests under each condition. The results of ANOVA revealed a statistically significant difference in the conditions ($F(2, 35) = 3.855$, $p < 0.05$). From post hoc comparisons, there was a significant difference between the Robot-Reconstruction condition and the Robot-Reproduction condition ($p < 0.05$), and was a marginally significant difference between the Robot-Reconstruction condition and the Video condition ($p < 0.10$), and was no significant difference between the Robot-Reproduction condition and the Video condition ($p = 0.93$).

Figure 11 shows the average scores of in-slide questions, and Fig. 12 shows the average scores of between-slides questions. As for in-slide questions, there was

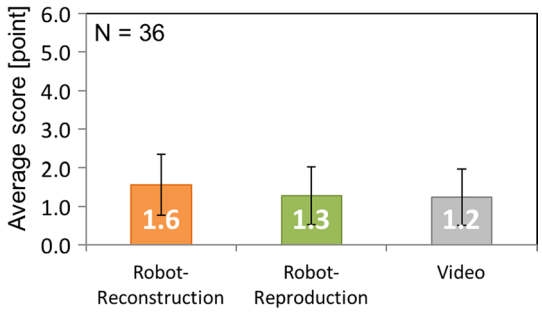


Fig. 12 Average scores of between-slides question

Table 6 Effect sizes (Cohen's d) between two conditions

Type of questions	Reconstruction – Reproduction	Reconstruction – Video	Reproduction – Video
Total	0.5 (Medium)	0.4 (Small)	0.1 (Very small)
In-slide	0.4 (Small)	0.3 (Small)	0.2 (Small)
Between-slides	0.4 (Small)	0.4 (Small)	0.0 (Very small)

a marginally significant difference between the Robot-Reconstruction condition and the Robot-Reproduction condition ($p < 0.10$) and were no significant difference between the Robot-Reconstruction condition and the Video condition ($p = 0.37$), and between the Robot-Reproduction condition and the Video condition ($p = 0.72$). As for between-slides questions, there were no significant differences between each condition (Robot-Reconstruction—Robot-Reproduction: $p = 0.22$, Robot-Reconstruction—Video: $p = 0.20$, Robot-Reproduction—Video: $p = 0.98$).

Table 6 shows the effect sizes (Cohen's d (Jacob, 1998)) between two conditions. The texts in parentheses represent interpretation for magnitudes of d defined by Cohen. As for between the Robot-Reconstruction and other conditions, the effect sizes d were 0.3 and more. On the other hand, the effect sizes between the Robot-reproduction and the Video were 0.2 and less.

From these results, the robot lecture involving reconstruction promotes understanding of the lecture contents more than the video lecture and the robot lecture involving simple reproduction, which overall supports H1 and H2. The results also suggest the necessity and importance of reconstructing lecture behavior with robot since the simple reproduction with robot did not significantly promote understanding compared to the video lecture. In addition, the results shown in Figs. 11 and 12 suggest that the lecture behavior reconstruction promotes understanding of the contents within slides rather than the relation between the slides. The current robot lecture system mainly deals with the lecture behavior for emphasizing the contents of each slide, not for emphasizing the relation embedded in multiple slides. These results show the validity of attention control and understanding promotion by means of the lecture behavior model.

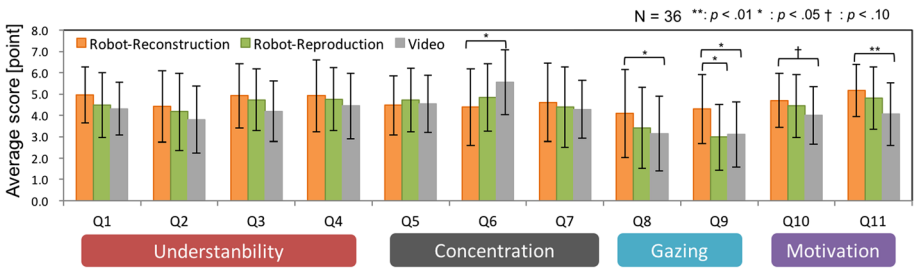


Fig. 13 Average scores of questionnaire

Questionnaire

Figure 13 shows the results of all questions in the questionnaire. The vertical axis represents the average scores, the horizontal axis represents each question item, and the error bars represent the standard errors of mean values. The Robot-Reconstruction condition tended to be better than other conditions in 9 of 11 questions. In the remaining 2 questions (Q5: Robot-Reproduction tended to be better, Q6: Video tended to be better), the other conditions tended to be better. From the results of the ANOVA, there were significant differences or marginally significant differences in Q3 ($F(2, 35) = 1.45$, $p < 0.10$), Q6 ($F(2, 35) = 6.67$, $p < 0.01$), Q8 ($F(2, 35) = 3.85$, $p < 0.05$), Q9 ($F(2, 35) = 5.04$, $p < 0.01$), Q10 ($F(2, 35) = 2.56$, $p < 0.10$) and Q11($F(2, 35) = 6.19$, $p < 0.01$). As for these question items in which there were significant differences, we also conducted Tukey–Kramer test as post hoc one. The results of the test reveal significant differences in Q6, Q8, Q9 and Q11, and a marginally significant one in Q10 as shown in Fig. 13.

There were no significant differences from Q1 to Q4. These questions were about understandability of gestures or the lecture contents. Most participants commented on paralanguage in Q1. There were a few positive comments that “It was easier to get an explanation under the Robot conditions than the Video condition because voice of the robot was clearer and more fluent.” On the other hand, we obtained many negative comments that “It was harder to get explanation under the Robot conditions than the Video condition because voice of the robot had no strength, intonation and rhythm.”

From these comments, many participants suggested that the robot could not emphasize and explain important points with paralanguage even if the robot’s voices were emphasized at the important points that were detected under the Robot-Reconstruction condition. As a result, there were no significant differences in terms of understandability of gestures or the lecture contents. Meanwhile, some participants commented on lecture behavior in Q1. There were some positive comments that "Robot often attracted my attention since the number of lecture behavior under the Robot-Reconstruction condition were more than under the Video condition." These comments suggest that reconstructing the lecture behavior is effective for gathering attention.

Q5 to Q7 were about concentration, and there were no significant differences in these questions except Q6. As for Q5, one participant commented "I concentrated because the robot moved the face and spoke smoothly." As for Q6, the participant

commented "I sometimes felt that Robot's motor sound was noisy during lecture. It distracted me from understanding the lecture contents."

Q8 and Q9 were about gazing, such as face direction and eye contact, and there were significant differences in these questions. In these questions, the Robot conditions obtained more scores than the Video condition. Some participants commented "We met eye to eye a lot since the robot moved its face direction to me." and "the lecturer in the video fixed his face direction but the robot tried to make eye contact." These comments suggest that the robot is addressing to the participants. From these results, it is suggested that the robot contributes to gathering attention and concentration using face direction.

Q10 and Q11 were about motivation, and there were also significant differences in these questions. The robot lecture contributed to keeping motivation of learners due to the novelty and presence of the robot.

Discussion

Let us here discuss the functional restrictions and related considerations of robot lecture in comparison to human lecture. First, human components such as gestures, paralanguage, etc. necessary for conducting lecture behavior are obviously superior to robot ones. Although such components allow human lecturers to more precisely conduct lecture behavior, it is so difficult for them to properly use the components. They often fail in keeping and controlling learners' attention in their lecture. On the other hand, robot has much difficulty in conducting precise lecture behavior due to limited components, but its behavior tends to be discriminating and recognizable.

In case of pointing gesture by human lecturers, for example, it is required to point to precise places. But, if it is imprecise, learners would be concerned about it and prevented from directing their attention. Pointing gesture by robot is apt to be rough by nature. Learners might be accordingly unconcerned and would be induced to give their attention to the rough direction. It could not be an obstacle for them to focus on the points.

In spite of limited robot components for lecture behavior, we need to consider how to complement lecture behavior by Sota. In order to complement its pointing gesture, for example, we can attach a laser pointer to Sota or add visual effect to the slide presented such as highlighting keywords that are synchronized with Sota's gestures.

Second, the current robot lecture system uses the gesture recognition databases to identify specific non-verbal behavior conducted by human lecturers. It is time-consuming to prepare such databases even if we can utilize machine learning techniques to tag lecture behavior. However, these are indispensable for conducting the robot lecture although we need to construct them from scratch.

Third, the robot lecture system presents lecture contents in the direction from Sota to learners, which could bring about their boredom during lecture presentation. In order not to get them bored, Sota accordingly needs to recognize learning states and to change lecture behavior depending on the states. For example, if there are learners who feel lecture presentation is difficult, Sota should present repeatedly with different non-verbal behavior.

Finally, the results of the case study with Sota suggest that the robot lecture promotes understanding of lecture contents, and that learners' impression of the robot

lecture are almost positive. These positive results might be brought about by a novelty effect provided that using Sota as a lecturer is novel for learners and they feel fascinated by Sota. The short-time lectures used in the case study might also have an influence on the effects of robot lecture. On the other hand, there are no significant differences between the robot lecture with simple reproduction and the video lecture, and there are significant differences between the robot lecture with reconstruction and the one with simple reproduction. These suggest that the positive results of the case study are not necessarily brought about by the novelty effect. As for the influence of lecture length on promotion of understanding a lecture, we need to ascertain if the robot lecture in a long time could bring about the same effects as the ones in the case study. As another option, nevertheless, we can consider a hybrid of robot and human lecture where robot gives an introduction in a short time and then human gives the remaining in each part of the lecture, since the robot lecture has positive effects in a short-time lecture.

The result of the questionnaire Q6, in addition, suggests that the video lecture is significantly better for concentration than the robot lectures. However, some learners often seem to be concerned with lecture behavior by Sota involving face direction and pointing gesture, and with its motor noise. Its lecture behavior is certainly conspicuous due to its embodiment, which would cause learners to distract their attention to lecture contents. If they get accustomed to Sota's behavior, such distraction could be ignored. We accordingly need to re-evaluate the robot lecture after learners get accustomed to lecture behavior by Sota. It is also necessary to make the motor noise smaller. This requires the motion as to lecture behavior to be smaller or slower. Another approach may be to cover the noise with Sota's oral explanation or to add sounds to motions for distraction from noise. By giving sounds to motions, it is also possible to contribute to calling attention and keeping attention. We still have to consider these points as future work.

Conclusion

In this work, we have proposed robot lecture, and demonstrated a robot lecture system, which augments lecture by human, and which reconstructs lecture behavior conducted in the lecture. Towards such reconstruction, we have designed a model of lecture behavior.

In addition, we have conducted the case study that examined the effect of understanding promotion with the robot. The participants attended different lecture contents under the three conditions of video lecture, robot lecture with simple reproduction, and robot lecture with reconstruction. According to the results of understanding test, the robot lecture with reconstruction promoted learner's understanding of slide contents more than the video lecture and the robot lecture with simple reproduction. There was also no significant difference between the robot lecture with simple reproduction and the video lecture. According to the results of the questionnaire, the robot lecture with reconstruction also contributed to keeping and controlling learners' attention. In addition, it became clear the importance of paralanguage for promoting understanding of the lecture contents. These results suggest the necessity and importance of the reconstruction of lecture behavior.

In future, we will consider how to more effectively present the lecture contents with lecture behavior of Sota. We will also aim to detect learning states to dynamically change lecture behavior for interactive lecture, although the current robot lecture system conveys the lecture contents to learners one-sidedly.

Acknowledgements
Not applicable.

Authors' contributions
TI developed theoretical framework and system for robot lecture and conducted case study. MG modeled of robot lecture and developed robot lecture system. AK managed overall and made theory of robot lecture. The authors read and approved the final manuscript.

Funding
The work is supported in part by JSPS KAKENHI Grant Numbers 18K19836 and 20H04294.

Availability of data and materials
None.

Declarations

Competing interests
The authors declare that they have no competing interests.

Author details
[1]The University of Electro-Communications, 1-5-1, Chofugaoka, Chofu, Tokyo 182-8585, Japan. [2]NTT Human Informatics Laboratories, 1-1, Hikari-no-oka, Yokosuka, Kanagawa 249-0847, Japan.

Received: 2 November 2020 Accepted: 28 November 2021
Published online: 05 January 2022

References

Admoni, H., Weng, T., Hayes, B., & Scassellati, B. (2016). Robot nonverbal behavior improves task performance in difficult collaborations. In *Proceedings of 11th ACM/IEEE international conference on human-robot interaction (HRI2016)* (pp. 51–58). https://doi.org/10.1109/HRI.2016.7451733.

Arima, M. (2014). An examination of the teachers' gaze and self reflection during classroom instruction: comparison of a veteran teacher and a novice teacher. *Bulletin of the Graduate School of Education, Hiroshima University, 63*(9–17), 2014. (in Japanese).

Belpaeme, T., Kennedy, J., Ramachandran, A., Scassellati, B., & Tanaka, F. (2018). Social robots for education: A review. *Science Robotics, 3*, 21. https://doi.org/10.1126/scirobotics.aat5954

Boersma, P., & Weenink, D. (2018). Praat: doing phonetics by computer [Computer program]. Version 6.0. 37, Retrieved Oct 26, 2020 from https://www.fon.hum.uva.nl/praat/.

Collins, J. (2004). Education techniques for lifelong learning: Giving a PowerPoint presentation: The art of communicating effectively. *Radiographics, 24*(4), 1185–1192. https://doi.org/10.1148/rg.244035179

FUJISOFT Inc. (2010). PALRO is A robot who cares. Retrieved Oct 26, 2020. https://palro.jp/en/

Goldin-Meadow, S., & Alibali, M. W. (2013). Gesture's role in speaking, learning, and creating language. *Annual Review of Psychology, 64*, 257–283. https://doi.org/10.1146/annurev-psych-113011-143802

Goto, M., & Kashihara, A. (2016). Understanding presentation document with visualization of connections between presentation slides. *Procedia Computer Science, 96*, 1285–1293. https://doi.org/10.1016/j.procs.2016.08.173

Huang, C. M., & Mutlu, B. (2014). Multivariate evaluation of interactive robot systems. *Autonomous Robots, 37*, 335–349. https://doi.org/10.1007/s10514-014-9415-y

Ishino, T., Goto, M., & Kashihara, A. (2018). A robot for reconstructing presentation behavior in lecture. In *Proceedings of the 6th international conference on human-agent interaction (HAI2018)* (pp. 67–75). https://doi.org/10.1145/3284432.3284460.

Jacob, C. (1998). *Statistical power analysis for the behavioral sciences* (2nd ed.). Routledge.

Kamide, H., Kawabe, K., Shigemi, S., & Arai, T. (2014). Nonverbal behaviors toward an audience and a screen for a presentation by a humanoid robot. *Artificial Intelligence Research, 3*(2), 57–66. https://doi.org/10.5430/air.v3n2p57

Le, Q., & Pelachaud, C. (2011). Generating co-speech gestures for the humanoid robot NAO through BML. *Gesture and Sign Language in Human-Computer Interaction and Embodied Communication*. https://doi.org/10.1007/978-3-642-34182-3_21

Liles, K. R., Perry, C. D., Craig, S. D., & Beer, J. M. (2017). Student perceptions: The test of spatial contiguity and gestures for robot instructors. In *Proceedings of the companion of the 2017 ACM/IEEE international conference on human-robot interaction (HRI2017)* (pp. 185–186). https://doi.org/10.1145/3029798.3038297

Mcneill, D. (1994). Hand and mind: What gestures reveal about thought. *Bibliovault OAI Repository, the University of Chicago Press*. https://doi.org/10.2307/1576015

Melinger, A., & Levelt, W. (2005). Gesture and the communicative intention of the speaker. *Gesture, 4*, 119–141.

Mutlu, B., Forlizzi, J., & Hodgins, J. (2007). A storytelling robot: modeling and evaluation of human-like gaze behavior. In *Proceedings of the 2006 6th IEEE-RAS international conference on humanoid robots, HUMANOIDS* (pp. 518–523). https://doi.org/10.1109/ICHR.2006.321322.
Nakano, Y., Okamoto, M., Kawahara, D., Li, Q., & Nishida, T. (2004). Converting text into agent animations: Assigning gestures to text. In *Proceedings of HLT-NAACL 2004: Short Papers* (pp. 153–56).
Ng-Thow-Hing, V., Luo, P., & Okita, S. (2010). Synchronized gesture and speech production for humanoid robots. *IEEE/RSJ International Conference on Intelligent Robots and Systems*. https://doi.org/10.1109/IROS.2010.5654322
Nozawa, Y., Dohi, H., Iba, H., & Ishizuka, M. (2004). Humanoid robot presentation controlled by multimodal presentation markup language MPML. In: *Proceedings of the 13th IEEE international workshop on robot and human interactive communication* (pp. 153–158). https://doi.org/10.1109/ROMAN.2004.1374747.
Sauppé, A., & Mutlu, B. (2014). Robot deictics: How gesture and context shape referential communication. In *Proceedings of the 9th ACM/IEEE international conference on human-robot interaction (HRI2014)*, 342–349.
Sharp Corporation. (2016). Robohon. Retrieved Oct 26, 2020 from https://robohon.com/global/.
Softbank Robotics Co. Ltd. (2018). NAO the humanoid and programmable robot. Retrieved Oct 26, 2020 from https://www.softbankrobotics.com/emea/en/nao/.
Tanaka, T., Fujikake, K., Takashi, Y., Yamagishi, M., Inagami, M., Kinoshita, F., Aoki, H., & Kanamori, H. (2017). Driver agent for encouraging safe driving behavior for the elderly. In *Proceedings of the 5th international conference on human agent interaction* (pp. 71–79). https://doi.org/10.1145/3125739.3125743.
Vstone Co. Ltd. (2010). Social communication robot Sota. Retrieved Oct 26, 2020 from https://www.vstone.co.jp/products/sota/.
Vstone Co. Ltd. (2018). Presentation Sota. Retrieved Oct 26, 2020 from https://sota.vstone.co.jp/home/presentation_sota/.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Submit your manuscript to a SpringerOpen journal and benefit from:

- ► Convenient online submission
- ► Rigorous peer review
- ► Open access: articles freely available online
- ► High visibility within the field
- ► Retaining the copyright to your article

Submit your next manuscript at ► springeropen.com