PAPER

Designing and Evaluating Presentation Avatar for Promoting Self-Review

Keisuke INAZAWA†, *Nonmember and* Akihiro KASHIHARA†a), *Senior Member*

SUMMARY Self-review is essential to improving presentation, particularly for novice/unskilled researchers. In general, they could record a video of their presentation, and then check it out for self-review. However, they would be quite uncomfortable due to their appearance and voice in the video. They also struggle with in-depth self-review. To address these issues, we designed a presentation avatar that reproduces presentation made by researchers. The presentation avatar intends to increase self-awareness through self-reviewing. We also designed a checklist to aid in a detailed self-review, which includes points to be reviewed. This paper also demonstrates presentation avatar systems that use a virtual character and a robot, to allow novice/unskilled researchers as learners to self-review their own presentation using the checklist. The results of case studies with the systems indicate that the presentation avatar systems have the potential to promote self-review. In particular, we found that robot avatar promoted engagement in self-reviewing presentation.
key words: *self-review, presentation skill, avatar, robot, engagement*

1. Introduction

For researchers, presentation is a must-do activity. They must consider not only what to present but also how to present with non-verbal behavior (called presentation behavior) involving gesture, gaze, and paralanguage. Presentation behavior is particularly critical for effectively communicating the presentation contents.

The presentation also requires that researchers practice their presentation to improve it before the actual presentation. There are two types of rehearsal: peer rehearsal and self-rehearsal. During peer rehearsal, researchers could get feedback on their presentations from peers, including more skilled lab members [1]. Self-rehearsal, on the other hand, involves researchers reviewing their presentations by themselves to identify points that need to be modified.

In this work, we look at how to help novice or unskilled researchers as learners self-review their presentations, particularly presentation behavior [2]–[4]. Although the most common way for them to self-review is to make a presentation to themselves on their PC, they must also review. On the other hand, there is another option: learners could record a video of their presentation and then review it. Although it allows them to direct more efforts to review, they would be quite uncomfortable due to their appearance particularly face [5], [6] and voice in the video [7]. This uncomfortableness prevents the learners from self-reviewing their presentation. In our previous work [2]–[4], in addition, we have ascertained that novice/unskilled researchers had difficulties in self-reviewing due to their insufficient knowledge about what to review.

This paper proposes a presentation avatar (P-Avatar for short) to address these issues. P-Avatar stands in for learners, and reproduces their presentation, including their presentation behavior. On a computer interface, it appears as a virtual character. We also use a communication robot with a head, hand, and body as a P-Avatar to more apparently embody presentation behavior. P-Avatar allows learners to self-review their own presentation behavior without feeling uncomfortableness because the avatar's appearance and voice can be changed from their original [2]. Such discomfort-free self-reviewing allows them to promote their awareness in self-reviewing. We also used a model of research presentation behavior that we designed to prepare a checklist with points to review [8], which allows for model-based self-review of presentation behavior.

This paper describes two P-Avatar systems that we developed using a virtual character and a robot as P-Avatar. These systems enable learners to use the checklist to self-review their own presentation behavior. In addition, this paper reports two case studies (study I and study II). The purpose of study I was to ascertain whether a P-Avatar using a virtual character could promote model-based self-review over self-review with a presentation video. The purpose of study II was to determine the difference of effectiveness between virtual character and robot as P-Avatar.

2. Self-Review

2.1 Self-Review in Presentation Rehearsal

In this work, we look at presentation rehearsal as a cyclic process with three phases: preliminary presentation, review, and modification. During the preliminary presentation phase, learners practice their presentation using a presentation document (P-document for short), such as PowerPoint/Keynote file. During the review phase, they receive peer reviews or check their presentation out on their own to determine which points need to be modified. During the modification phase, the learners use the review results to modify the P-document, oral explanation, and non-verbal behavior (gesture, face direction, paralanguage, etc.). Before the actual presentation, the learners can improve their

Manuscript received October 3, 2021.
Manuscript revised March 4, 2022.
Manuscript publicized May 26, 2022.
†The authors are with The University of Electro-Communications, Chofu-shi, 182–8585 Japan.
a) E-mail: akihiro.kashihara@inf.uec.ac.jp
DOI: 10.1587/transinf.2021EDP7210

presentation by repeating these phases.

Among these phases, we focus on self-review. Although peer review is an effective way to improve presentation, conducting self-review before peer review has the following advantages in both reviews. First, peers could direct their attention to more important aspects of the presentation because learners are expected to identify minor flaws in advance through self-review. This helps to improve the quality of peer review. Second, learners could benefit from a comparison of the results of self-review and peer review. This causes them to become aware of points that need to be modified, which they had overlooked during self-review. It also enables them to discover the shortcomings of their self-review and unknown review points. Learning from the comparison accordingly contributes to improving skills in self-review. Such skill improvement enables learners to better modify their presentation on their own. As a result, self-review is essential, particularly for novice/unskilled learners [2].

Learners are also expected to review their own presentation from the perspective of the audience during self-review. They must specifically review non-verbal behavior such as gaze, gesture, paralanguage etc., because it is critical to present the research contents as slides with oral explanation in research presentations to engage the audience's attention and promote their understanding. This requires the proper application of such non-verbal behavior [8]. It is referred to as presentation behavior.

The common way for learners to self-review is to make a presentation to themselves with their PC. However, it is difficult to review their presentation because they must make and review it at the same time. On the other hand, if learners record a video of their presentation and then review it, they may be able to devote more efforts to review. However, their appearance (particularly face) and voice in the video would cause them uncomfortable feelings [2], which would also result from a discrepancy between the expected and recorded appearance/voice [5]–[7]. It is difficult for them to look back on their own presentation with such uncomfortableness.

Even if they were able to overcome uncomfortable feelings, there would be a limit to reviewing themselves objectively, i.e., from the perspective of the audience. According to Grez et al., assessment of oral presentations by peer learners had a positive relationship with assessment by teachers, but self-assessment of oral presentations by learners was significantly different from assessment by teachers [9]. Furthermore, in our previous case study, learners were unable to self-review their presentation in detail due to a lack of knowledge about what to review [2]. It is difficult for learners to identify points for modification in self-reviewing as a result of these issues.

With a focus on presentation behavior, we designed P-Avatar to reduce uncomfortable feelings to promote objective review [2]. We also designed a checklist based on a model of research presentation behavior that indicates what learners should look for in their presentation behavior.

2.2 Related Work

Related work on supporting presentation self-rehearsal has mainly addressed how to automatically analyze nonverbal behavior and oral explanation from motion capture and presentation voice data. Kopf et al. [10], Schneider et al. [11] and Chollet et al. [12] proposed systems that analyze nonverbal behavior and present the results as feedback during their presentation. Trinh et al. [13] and Melinger et al. [14] also proposed systems that recognize gesture or oral explanation to provide verbal feedback via robot after their presentation. Furthermore, Chollet et al. proposed a system that uses an audience agent to enhance learners' engagement in presentation [12].

The information recognized by these systems appears to be instructive for presentation enhancement. However, these systems do not explicitly address how learners should self-review and how they should improve their self-reviewing skills. Zhao et al. proposed a system that provides feedback based on statistical data of non-verbal behavior analyzed after the presentation and allows learners to view the presentation video with the analyzed information. However, they would still have difficulty gaining awareness of points to be modified because the system provides no support for reducing uncomfortable feelings caused by looking at their appearance and listening to their voice during the self-review process.

The P-Avatar proposed in this paper is not intended to present data of non-verbal behavior recognized as feedback, but rather to allow learners to review their presentation in a self-directed and model-based manner to gain their awareness of points that need to be modified.

3. Self-Review with Presentation Avatar

3.1 Presentation Avatar

P-Avatar was designed to address issues with self-reviewing presentation behavior [2]. P-Avatar is an avatar that reproduces learners' presentation as if it were made by another person. Such reproduction enables them to self-review their presentation behavior without experiencing any uncomfortable feelings. This helps to increase their awareness of points that need to be modified in their presentation. P-Avatar is currently being considered as virtual character running on a computer interface or communication robot. P-Avatar has some requirements for encouraging self-review. First, the appearance and voice of P-Avatar need to be completely different from learners' ones so that they do not receive any uncomfortable feelings. Second, P-Avatar must maintain their utterance and reproduce their non-verbal behavior as precisely as possible for learners to properly self-review.

In contrast to a virtual character such as a P-avatar, the robot has physical embodiment, which allows it to reproduce presentation behavior using its body. Such reproduc-

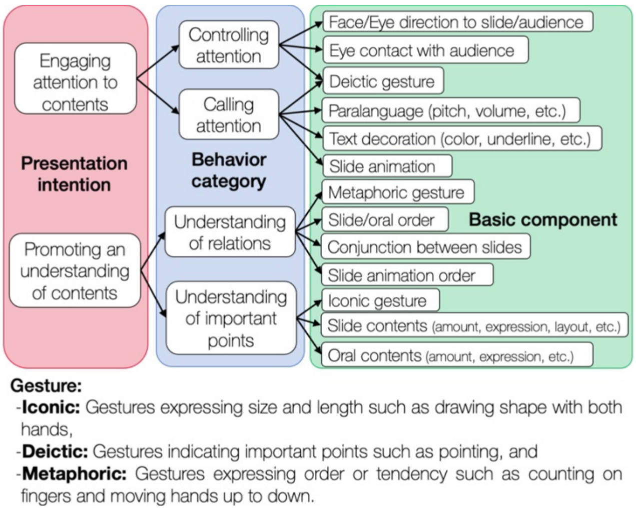

Fig. 1 Model of research presentation behavior.

tion allows for a more visible expression of face direction and pointing gesture. According to related work on comparing robots and virtual characters, robots provide a more acceptable embodiment of humans than virtual characters. Cory D et al. pointed out that a robot was more engaging, credible, and informative in interaction with a human than a virtual character because of its physical embodiment [15]. Tanaka et al. reported that the robot was highly visible and that its motion with a pointing gesture would be effective in directing human attention [16]. According to Goto et al., the robot expressed clearer face direction and a stronger impression of watching a presentation [8]. Based on these findings, we can say that the physical embodiment of a robot promotes engagement in self-reviewing. Another related work on engagement in learning suggests that engagement motivates learners to acquire knowledge [17], [18]. In addition, motivation related to engagement influences the accuracy of self-assessment [19].

Regarding the motion of the P-avatar, the virtual character is capable of reproducing continuous motion of gestures made by learners. As a result, it could provide more accurate reproduction. In the case of the robot, however, continuous gesture motion is segmented and transformed into robot gestures. But, the original motion could be made more discriminatory.

3.2 Model of Research Presentation Behavior

In order to encourage self-review, we provide learners with a checklist that indicates what they should review. To prepare the checklist, we use the model of research presentation behavior with slides as shown in Fig. 1, which we have designed by consulting a number of related works on presentation [8], [14], [20]–[22].

In general, non-verbal behavior, such as gaze, gesture, face direction, or paralanguage, is used during presentation to explain and supplement slide/oral contents. Such presentation behavior should also be conducted in accordance with the presentation intention [20]. It is particularly important in a research presentation to draw the audience's attention to and promote an understanding of important points that presenters like to make.

In the presentation behavior model, we classify presentation intention for research into two: engaging attention to contents and promoting an understanding of contents. The model, as shown in Fig. 1, is composed of three layers: the intention of presentation behavior, the behavior category, and the basic components of presentation behavior. The model shows how presentation intention can be achieved through presentation behavior involving a variety of their components. For example, if a learner wants to draw the audience's attention to the slide contents, he/she could use non-verbal behavior for calling attention, which includes several components such as highlighting the important parts of the slide as text decoration, pointing to them as deictic gesture, increasing voice volume for them as paralanguage, and so on. This non-verbal behavior for achieving these presentation intentions has been theoretically validated in related work on presentation. In our related work on robot lecture, we established a model of non-verbal behavior for lecture that includes the same presentation intentions as research, which allows the robot to conduct its lecture behavior and control learners' attention [23].

Following the model of research presentation behavior, we have prepared a checklist for self-review as shown in Table 1, which includes points to be reviewed. Using this checklist during self-review, learners can determine whether they use presentation behavior to control attention to slide

Table 1 Checklist for self-review.

Presentation intention	Behavior Category	Basic Component	Review points
Engaging attention to contents	Controlling attention	Face/Eye direction to slide contents or to audience	Directing face/eye to slide or to audience
		Eye contact with audience	Making eye contact
		Deictic gesture	Pointing to important points in slide
	Calling attention	Deictic gesture	Pointing to important points in slide
		Paralanguage	Emphasizing important points by voice volume/pose/speaking speed
		Text decoration	Emphasizing important points in slide with color/bold/underline
		Slide Animation	Emphasizing importance points by animation
Promoting an understanding of contents	Promoting an understanding for relations	Metaphoric gesture	Expressing metaphonic gesture
		Slide/Oral order	Adequate order of slides/oral explanation
		Conjunction between slides	Conjunctive explanation at slide transition
		Slide animation	Adequate and expected movement, and suitable timing in animation
	Promoting an understanding for important points	Iconic gesture	Expressing shape or size by gesture
		Slide contents	Proper/correct expression, sufficient information, and proper use of figure/table/illustration Appropriate position/adequate amount of information, suitable text decoration (size, font, etc.), suitable color, and unified use of words
		Oral contents	Proper/correct/sufficient explanation. Appropriate expression, explanation of technical terms, and unified use of words

contents, behavior to promote an understanding of important points in slides, and so on. They can also determine the suitability of the slide/oral contents based on their intended use. This allows them to raise their awareness of points that need to be modified in their presentation.

4. Presentation Avatar System

4.1 Framework for Self-Review with P-Avatar

Figure 2 shows the framework for self-review with P-Avatar. The main purpose of P-Avatar is to allow learners to conduct model-based self-review with the checklist. P-Avatar first records and captures learners' presentation and then replicates their presentation behavior within the basic components of the presentation behavior model. They are expected to become aware of points that need to be changed in their presentation behavior.

We have developed two systems, which promote self-review with virtual character (VC self-review for short) and the one with robot (Robot self-review for short). We use Sota [24] of Vstone Co., Ltd. as the robot. Let us explain these systems in the following.

4.2 VC Self-Review System

Unity Technologies are used to implement the VC self-review system [25]. The system records and captures the presentation learners make using Kinect [26] as input device, as shown in Fig. 3. The current system uses a CG model of Sota as a virtual character (VC) in reproducing the presentation with the captured data, shown in Fig. 2. We

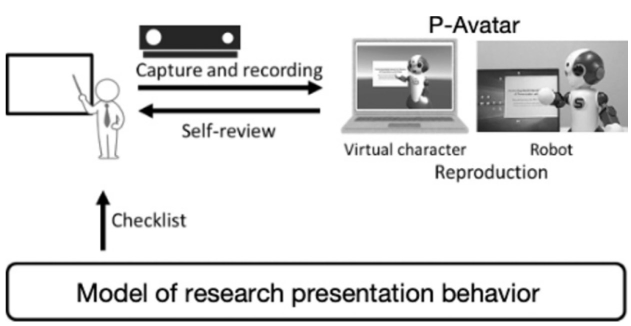

Fig. 2 Framework for self-review with P-Avatar

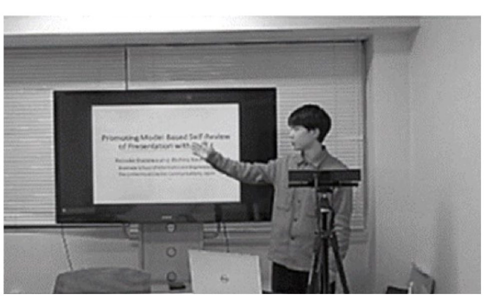

Fig. 3 Presentation

use the virtual Sota to compare with Sota as a robot. It has 8 joints, which are the same as Sota.

Let us show how to reproduce the presentation using the VC. The system displays the slides in the user interface using a PPT file. The system displays a slide when learners begin the presentation as self-rehearsal. When they change the slide to the next or previous one using the keyboard, the system captures the timing. When they demonstrate animation in a slide, the system also captures the timing of the animation. Kinect records their movement including hand gestures, face direction except for facial expression, and body direction with position coordinates of 25 joints they have. The frame rate of Kinect is 30 fps. It also records their oral explanation.

When the learners have completed the presentation, the system generates a video using the recorded and captured data, in which the VC reproduces the presentation. In the presentation reproduction, the system extracts motion data of the 8 joints from their movement recorded by Kinect, and operates each joint within the motion range that is wider than Sota. In terms of oral explanation, the system also transforms voice tone by adjusting the fundamental frequency and formant frequency values. Voice height is determined by the fundamental frequency. The formant frequency also indicates resonance determined by the shape of a vocal tract or an oral cavity, and characterizes individual voices.

When the learners replay the video, their captured movement is projected as the VC's movement, as shown in Fig. 2. For movement projection, the Kinect Examples with MS-SDK tool [27] is used. Due to the limitation of the current CG model, the VC attempts to reproduce the captured gesture, face direction, and body direction as precisely as possible. The presentation slides and animation are demonstrated in the captured timing.

4.3 Robot Self-Review System

Robot self-review system is implemented with C#. It records and captures learners' presentation with a PPT file in the same way that the VC self-review system does. Robot reproduces their presentation behavior using the recorded and captured data, reproducing the captured gesture, face direction, body direction, and utterance as precisely as possible, and changing the voice tone. Since Sota does not have fingers, in addition, counting gesture as metaphoric one is not reproduced.

Sota has 8 joints in total, including the neck (yaw, pitch and roll), shoulder, elbow, and hip. From the data of these joints recorded and captured by Kinect, the system generates data for robot movement as a text file in real-time during presentation recording. Each joint angle is included in the data for robot movement. The captured face direction is used by the system to calculate the angle of Sota's neck. From the captured position coordinates of shoulder, elbow, and hip, the system calculates the angle of shoulders. The angle of its elbows is calculated using the captured position coordinates of shoulder, elbow, and wrist. It also calculates the body direction from the captured shoulder position coordinates. The temporal change in the angle calculated for each joint is used as the velocity of joint motion. The data of angle and velocity is rounded within Sota's motion range, which is transformed to the data used by servomotor of each joint Sota has. The Kinect-equipped microphone also records learners' oral explanation. The system transforms the voice tone in the same way that the VC self-review system does.

When the learners finish their presentation, the system generates the recorded data of slide and animation transition timing, robot movement, and voice data. It is used by the robot to reproduce the presentation.

In order to allow Sota to reproduce presentation properly, it is necessary to maintain the positional relationship between learners and a display showing presentation slides. We accordingly calibrate display position, Sota's position, face direction, and pointing gesture before presentation by learners is recorded. We first ask learners to turn their face to slide or to audience and to point to the upper/middle/lower part of displayed slide. These face direction and pointing gestures are projected on Sota in real time. Checking the projected face direction, we calibrate the positional relationships between the display and Sota. Then, checking the pointing gestures, we calibrate the distance between the display and Sota, and the size of displayed slide.

When the learners self-review the presentation, the recorded gestures are projected as robot movement simultaneously with the transformed voice. Sota reproduces gesture, face direction, and body direction as precisely as possible within the constraints of Sota's movement. The constrained movement helps them detect points to be modified, but they could not always recognize the detected points as defects of their own presentation behavior due to the shape of Sota and its movement. In order to prevent this as much as possible, the learners are asked to identify the movement of Sota as their avatar conducting their own presentation behavior before self-review with Sota. The presentation slides are displayed on a windowed version of PowerPoint, which receives data about slide and animation transition timing from Sota via WiFi and performs the transition. During self-reviewing, learners can pause and replay the reproduction using a browser-based controller.

5. Case Study

There were two case studies. In case study I, we compared VC self-review and self-review with presentation video. We also compared VC self-review and Robot self-review in case study II. In each study, the P-Avatar used for VC self-review was the same.

On the same day, all participants took part in both case studies. Study II was conducted first, followed by study I. There are two reasons for this order. To begin with, modified points to be noticed by participants would be less in the second study than the first because they could improve presentation in the second study after experiencing model-based self-review in the first study even though the presentation contents in the two studies are different. Second, the difference in modified points between two conditions in each study is expected to be smaller in study II than in study I because the difference between the two conditions in study II is the presence or absence of P-Avatar's embodiment. We conducted study II before study I to make the difference between VC self-review and Robot self-review as clear as possible, though we report in the order of study I and then study II in the following. Figure 4 shows the procedure of each case study.

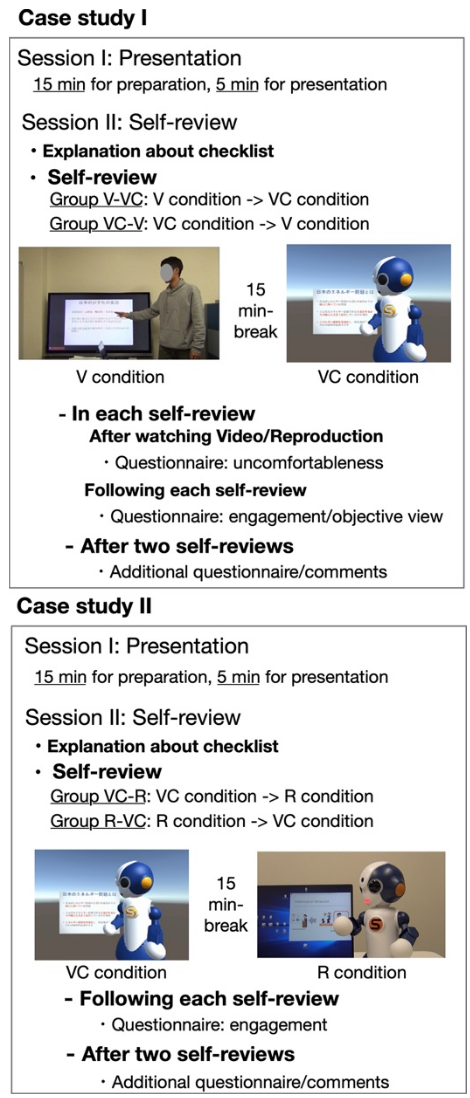

Fig. 4 Procedures of case studies

5.1 Comparing VC Self-Review and Self-Review with Presentation Video

(1) Purpose

We investigated whether a VC as P-Avatar could promote model-based self-review of presentation behavior rather than presentation video in this study. We also used three criteria: awareness of points to be modified, uncomfortable feelings, and engagement in self-reviewing.

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

H1: P-Avatar increases more awareness of points to be modified.
H2: P-Avatar decreases uncomfortable feelings and promotes an objective view of presentation behavior.
H3: P-Avatar promotes more engagement in self-reviewing.

(2) Method

The participants were 16 informatics and engineering graduate and undergraduate students. They could be regarded as novice/unskilled researchers since they had research experience within three years in their own laboratory. We set two conditions: self-review with presentation video (V condition), and VC self-review (VC condition). Under these two conditions, each participant performed self-review twice using the checklist including the review points shown in Table 1 as within-subject design. The participants were divided into two groups: group V-VC and group VC-V. Each group received eight participants at random. Group V-VC self-reviewed first under the V condition and then under the VC condition. The VC-V group self-reviewed in the opposite order of the two conditions.

Compared to the V condition, we analyzed how many points to be modified could be found in the VC condition. We also used questionnaires to analyze uncomfortable feelings and objective views of presentation behavior. The questionnaire for analyzing uncomfortable feelings had 7-items, each of which was a 5-point Likert-scale question. One 5-point Likert-scale question was included in the questionnaire for analyzing an objective view. In addition, we analyzed engagement with a questionnaire. The questionnaire was based on the User Engagement Scale (UES) [28], and it consisted of 17 items, each of which was a 5-point Likert-scale question. UES originally consists of 31 items and comprises 4 engagement dimensions: aesthetic appeal, focused attention, perceived usability, and reward. We used focused attention and reward as dimensions of engagement in this study and translated them into Japanese. As an additional questionnaire, we prepared a questionnaire that asked which condition allowed the participants to improve their concentration on self-review.

We prepared a P-document with 5 slides about Japanese energy problems. It could be regarded as a research P-document, since the contents included problems to be addressed, their factors, and solutions proposed. Each slide also contained a sample speech script. Participants in this case study were required to use the prepared P-document for presentation. They were also asked to self-review not the P-document but their non-verbal behavior about oral explanation and gesture involving face direction. In other words, they were required to review the points except the ones related to the basic components that are text decoration, slide animation, slide order, and slide contents in Table 1. In addition, we prepared a printed checklist.

As shown in Fig. 4, this study included 2 sessions referred as Session I (presentation), and Session II (self-review). In Session I, they were given explanation about the model of research presentation behavior. They were then given 15 min to prepare for the presentation. They were permitted to modify/change the example speech scripts, but not to edit the contents of the slide. They were then required to assume their own presentation intentions to make a presentation in front of Kinect and a video camera using the prepared P-document projected (without speech script) on-screen in 5 min. The presentation was recorded with Kinect and the video camera.

The checklist was first explained to the participants in Session II. They were then required to follow it and review their presentation behavior twice with the presentation video and P-Avatar, as well as identify points that need to be modified. P-Avatar reproduced the presentation using the same data as the presentation video. We gave each participant a copy of the P-document. Participants were asked to annotate the corresponding slide in the handout with the modified points they identified. We had a 15-min break between two self-reviews. They were required to watch the presentation video or reproduction with a VC prior to each self-review and then answer the questionnaire to measure uncomfortable feelings.

Following each self-review, participants were asked to complete a questionnaire to assess their engagement and objective view of their presentation behavior. After the two self-reviews, they were required to complete an additional questionnaire and make any comments on self-reviews under the two conditions.

(3) Results

Table 2 shows the average numbers of modified points found for presentation behavior about gesture and oral explanation under each condition in each group, as well as the total average numbers obtained under each condition in both groups. The one-sided t-test revealed a tendency of significant difference in the total average number of found points as for gesture under each condition ($t(15) = -1.596$, †$p < 0.10$). There were also significant differences in the total average numbers of found points as for oral explanation, and total under each condition (oral explanation: $t(15) = -1.954$, $^{*}p < 0.05$; total: $t(15) = -2.305$, $^{*}p < 0.05$). In group V-VC, in addition, there were significant differences between the average numbers of found points as for gesture,

Table 2 Average numbers of modified points found under V and VC conditions

	V condition			VC condition		
	Gesture	Oral	Total	Gesture	Oral	Total
Average (V-VC)	2.25	2.50	4.75	4.25**	3.75†	8.00*
Average (VC-V)	4.13	3.88	8.00	3.63	4.75	8.38
Total average	3.19	3.19	6.38	3.94†	4.25*	8.19*

**: p<.01, *: p<.05, †: p<.10

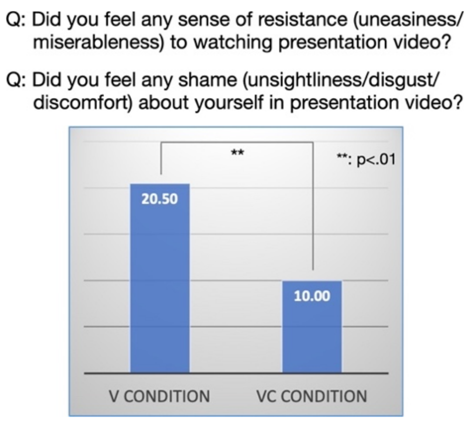

Fig. 5 Uncomfortable feeling scores

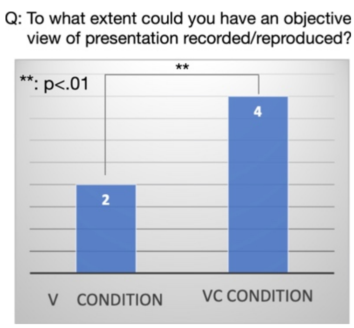

Fig. 6 Scores of questionnaire for objective view.

and total under each condition (gesture: $t(7) = -3.347$, $^{**}p < 0.01$; total: $t(7) = -2.728$, $^{*}p < 0.05$). There was also a tendency of significant difference in the average numbers of found points as for oral explanation (oral explanation: $t(7) = -1.722$, $\dagger p < 0.10$). These results support H1.

Figure 5 shows the average scores in the questionnaire for analyzing uncomfortable feelings, which were calculated by summing the participants' answers to all the seven questions. The perfect score was 35. From the result of the one-sided t-test, there was a significant difference between the average scores in V and VC conditions ($t(15) = 6.501$, $^{**}p < 0.01$). Figure 6 also shows the median values in the questionnaire results for analyzing objective view. The Wilcoxon signed-rank test revealed a significant difference between the median values in the V and VC conditions ($T = 6.5$, $^{**}p < 0.01$). These results support H2.

Figure 7 shows the average engagement scores. The scores were also calculated by summing the participants' responses to all the 17 questions. The highest possible score was 85. From the result of the one-sided t-test, there was a significant difference between the average scores in the V and VC conditions ($t(15) = -3.954$, $^{**}p < 0.01$). This result supports H3.

Figure 8 shows the results of the additional questionnaire. The one-sided accurate binomial test revealed a tendency of significant difference between the V and VC conditions ($p = 0.0768$, $\dagger p < 0.10$).

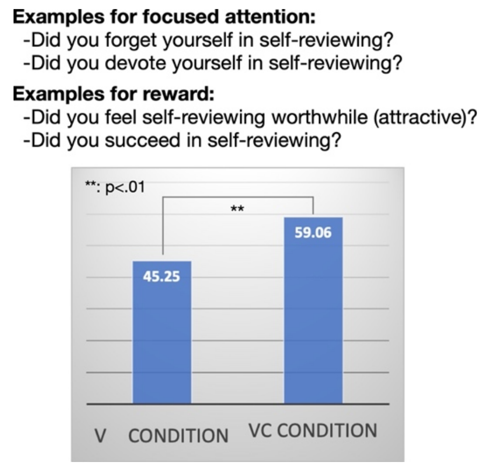

Fig. 7 Engagement scores.

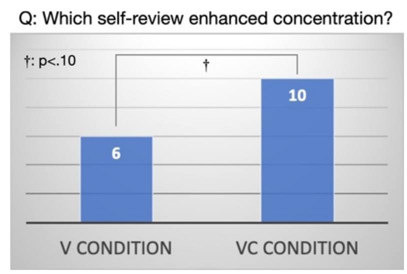

Fig. 8 Results of additional questionnaire.

Table 3 Average numbers of modified points commonly found under V and VC conditions

	Common		
	Gesture	Oral	Total
Average (V-VC)	0.75	1.38	2.13
Average (VC-V)	1.13	0.50	1.63
Total average	0.94	0.94	1.88

(4) Discussion

First, we discuss the effectiveness of self-review with P-Avatar. Table 2 suggests that using a VC as P-Avatar aids in identifying more points to modify for presentation behavior related to both of gestures and oral explanation. However, the average numbers of modified points for oral explanation in the VC-V group decreased from the VC condition (4.75) to the V condition (3.88). We think some modified points found with P-Avatar seem to learners not serious when they watch the presentation video. This suggests that the participants might have overly self-reviewed their oral explanation with P-Avatar. Although the over self-review should be avoided, it could be fail-safe for novice/unskilled learners because they often miss finding points to be modified.

Table 3 shows the average numbers of modified points commonly found under both conditions in each group. From the results of Tables 2 and 3, there were modified points found under only one condition. This suggests P-Avatar (or the presentation video) allows learners to become aware of

Table 4 Average numbers of modified points found for behavior category

	V condition			VC condition		
	Gesture	Oral	Total	Gesture	Oral	Total
Controlling attention	1.88	n/a	1.88	2.19	n/a	2.19
Calling attention	1.00	0.94	1.94	1.00	1.63	2.63
Understanding relations	0.31	0.88	1.19	0.38	0.94	1.31
Understanding of important points	0.06	1.31	1.38	0.38	1.69	2.06

points to be modified, which they overlook or they do not think serious with the presentation video (or P-Avatar).

In addition, we classified the modified points found under the V and VC conditions into each behavior category in Fig. 1, which is shown in Table 4. This suggests that learners tend to detect more modified points for face direction and deictic gesture such as pointing gesture rather than metaphoric and iconic gestures in self-reviewing. P-Avatar reinforces this tendency. It also promotes self-review of oral explanation.

In terms of uncomfortable feelings and an objective viewpoint, Fig. 5 suggests that P-Avatar could decrease uncomfortable feelings in self-review. According to Fig. 6, P-Avatar enables learners to take an objective view in self-reviewing. Eight participants stated that they felt uncomfortable in the V condition, but could objectively self-review in the VC condition. According to Figs. 7 and 8, P-Avatar could promote engagement and concentration in self-reviewing. From these results, we believe that P-Avatar can promote self-review of presentation behavior more effectively than presentation video.

Consider the following examples involving two participants, A and B, who obtained more points to be modified in the V condition. In the case of participant A, the uncomfortable feeling was not a major issue during self-reviewing. He stated that he did not feel uncomfortable in the V condition and that he could observe presentation behavior in detail, despite the fact that the uncomfortable feeling score was higher in the V condition. As for participant B, the uncomfortable feeling score and the engagement score were higher in the VC condition. In the additional questionnaire, he answered P-Avatar promoted his concentration. He also mentioned how easy it was to observe presentation behavior in the VC condition. Thus, his subjective impression was better in the VC condition. However, he found one more point that needed to be modified in the V condition. He seems to be able to self-review in both conditions.

Finally, consider participant C. Although he did not feel uncomfortable in the V condition, he obtained more points to be modified in the VC condition. He stated that he felt self-reviewing consciously in the VC condition. P-Avatar may provide learners with a more conducive environment for them to self-review.

In this study, we confirmed that P-Avatar increased awareness of points that needed to be modified. However, we were unable to determine whether P-Avatar helps to improving presentation. We need to conduct case studies to ascertain this. In addition, we used Sota as a VC. However, gender or learner's preference of its appearance or voice would influence self-reviewing. We accordingly need to investigate the more suitable appearance or the voice of P-Avatar.

5.2 Comparing VC Self-Review and Robot Self-Review

(1) Purpose

In this study, we investigated whether Robot self-review could be more effective than VC self-review. We accordingly used two criteria: awareness of points to be modified in self-reviewing and engagement in self-reviewing.

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

H1: Robot self-review increases more awareness of points to be modified.
H2: Robot promotes more engagement than VC.

(2) Method

The participants were 16 students who participated in study I. We set two conditions: VC self-review (VC condition), and Robot self-review (R condition). As a within-subject design, each participant performed self-review twice with the checklist under these two conditions. The participants were divided into two groups: group VC-R and group R-VC. We randomly assigned eight participants to each group. Group VC-R self-reviewed first under the VC condition and then under the R condition. Group R-VC self-reviewed in the opposite order of the two conditions.

We analyzed how many points to be modified could be found in the VC and R conditions. We also analyzed engagement with questionnaires used in study I. We prepared an additional questionnaire with two questions that asked which self-review made presentation behavior more apparent and enhanced concentration on self-review.

We prepared a P-document with 5 slides about Japanese declining birthrate problems. It could be regarded as a research P-document in the same way as the one used in study I. Each slide contained a sample speech script. The participants in this study were required to present using the prepared P-document. They were also required to self-review not the P-document but their presentation behavior as for gesture and oral explanation. We also prepared a printed checklist.

As shown in Fig. 4, this study included 2 sessions: Session I (presentation), and Session II (self-review). In Session I, their presentation was captured using Kinect in the same manner as in study I. The checklist was first explained to the participants in Session II. They were then required to review their presentation twice with VC self-review and Robot self-review in the same way as study I. Following each self-review, they were required to complete a questionnaire to assess their engagement. After the two self-reviews, they were also required to answer the additional questionnaire

Table 5 Average numbers of modified points found in self-review.

	VC condition			R condition		
	Gesture	Oral	Total	Gesture	Oral	Total
Average (VC-R)	7.13	4.50	11.63	7.25	5.25	12.50
Average (R-VC)	4.88	6.88	11.75	5.75	7.75	13.50*
Total average	6.00	5.69	11.69	6.50	6.50	13.00*

*: p<.05

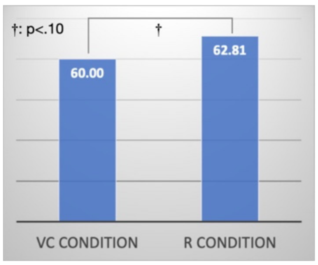

Fig. 9 Engagement scores

and to make any comments on self-reviews under the two conditions.

(3) Results

Table 5 shows the average numbers of modified points found for presentation behavior about gesture and oral explanation under each condition in each group, as well as the total average numbers obtained under each condition in both groups. According to the two-sided t-test results, there was a significant difference between the total average numbers under each condition ($t(15) = -2.406$, $^{*}p < 0.05$). There was also a significant difference between the total average numbers in group R-VC under each condition ($t(7) = -2.497$, $^{*}p < 0.05$). These results support H1.

Figure 9 shows the average engagement scores. According to the result of the one-sided t-test for the scores, there was a tendency of significant difference between the average scores in VC condition and R condition ($t(15) = -1.442$, $\dagger p < 0.10$). This result supports H2.

Figure 10 shows the results of the additional questionnaire, which included Q1 and Q2. As for Q1, the result of the two-sided accurate binomial test showed a tendency of significant difference between the VC and R conditions ($p = 0.077$, $p < 0.10$). As for Q2, the result of the two-sided accurate binomial test showed no difference between the VC and R conditions, but more participants selected Robot self-review.

(4) Discussion

First, we compare the effectiveness of Robot self-review to that of VC self-review. Table 5 indicates that Robot self-review aids in finding points for modification. VC self-review, on the other hand, tended to make presentation behavior more apparent, as shown in the results of Q1 in Fig. 10. Six participants commented that the VC's reproduction of presentation behavior was more accurate and continuous. Furthermore, four participants stated that it was simple to view the slide contents and presentation behavior in VC self-review because they were presented in the same display.

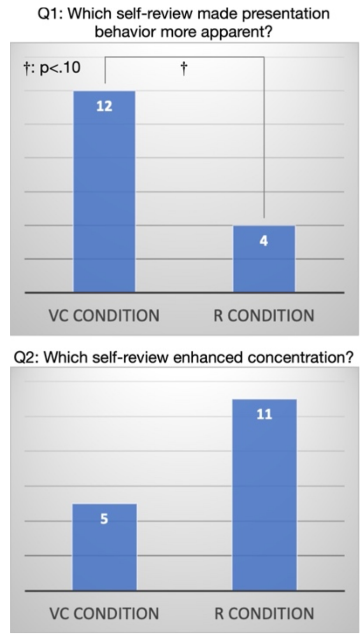

Fig. 10 Results of additional questionnaire.

Table 6 Average numbers of modified points commonly found under VC and R conditions

	Common		
	Gesture	Oral	Total
Average (VC-R)	3.25	1.63	4.88
Average (R-VC)	1.88	1.75	3.63
Total average	2.57	1.69	4.26

Based on the above discussion, the VC can provide an accurate reproduction of presentation behavior. From the results of Table 5, on the other hand, we believe that a robot is more suitable for promoting self-review. We think that the robot could make presentation behavior less accurate but more discrete and discriminable, causing learners to become more awareness of points that need to be modified.

As described in Sect. 5.1 (4), in addition, there is some possibility of overly self-reviewing in finding modified points under both conditions, which may result from unnatural behavior conducted by P-Avatar. It is important to avoid the over self-review to make the modified points more appropriate. On the other hand, we think it allows learners to take more care to check their own presentation behavior.

Table 6 shows the average numbers of modified points commonly found under both conditions in each group. From the results of Tables 5 and 6, there were modified points found under only one condition. This brings about the same suggestion as the one obtained from Tables 2 and 3.

We also classified the modified points found under the VC and R conditions into each behavior category, which is shown in Table 7. This suggests that learners tend to detect

Table 7 Average numbers of modified points found for behavior category

	VC condition			R condition		
	Gesture	Oral	Total	Gesture	Oral	Total
Controlling attention	3.69	n/a	3.69	4.19	n/a	4.19
Calling attention	1.81	2.00	3.81	1.56	2.63	4.19
Understanding relations	0.44	1.44	1.88	0.56	1.56	2.13
Understanding of important points	0.13	2.19	2.31	0.13	2.38	2.50

more modified points for face direction and deictic gesture rather than metaphoric and iconic gestures in self-reviewing. The robot overall reinforces this tendency. It also promotes self-review of oral explanation rather than the VC.

Next, we discuss about engagement. Figure 9 suggests that Robot self-review promotes engagement in self-reviewing. Four participants commented that the presence of a presenter in the R condition caused them to focus more on robot. Two other participants also expressed interest in the robot as well. Following the result in study I that a VC promoted self-review, there is a possibility that a robot as P-Avatar promotes engagement in self-reviewing more strongly. The engagement may contribute not only to awareness of points that need to be modified but also to accuracy of self-review as mentioned in 3.1.

However, it remains unclear whether Robot self-review is better than self-review with presentation video. We accordingly intend to conduct another case study to compare Robot self-review and self-review with presentation video in our future work.

In this study, we used Sota as a robot. Sota currently has some shortcomings, such as motor noise and restricted motion. We must either resolve these issues, or consider using a different robot.

6. Conclusion

In this paper, we proposed P-Avatar and developed P-Avatar systems using a VC and a robot to promote awareness in self-reviewing. We also followed the model of research presentation behavior to design the checklist for self-review, allowing for a detailed self-review of the presentation.

This paper also reported the case studies in which we confirmed the possibility that P-Avatar promotes self-review and the difference of effectiveness between a VC and a robot as P-Avatar. The results of the case study suggest that P-Avatar could improve awareness of points to be modified in self-reviewing, and robot avatar promotes awareness of points to be modified and engagement more than VC avatar.

In the future, we intend to conduct a case study to ascertain whether self-review with P-Avatar is appropriate in comparison to review by peers or skilled researchers. We must also consider the appearance and voice of P-Avatar, as well as what type of robot is more suitable for self-review.

We also intend to use the model to diagnose the reproduced presentation to automatically identify points that need to be modified, which could help learners in self-review and diagnose their self-review results.

Acknowledgments

This paper was supported in part JSPS KAKENHI Grant Number 18K19836 and 20H04294.

References

[1] R. Okamoto and A. Kashihara, "Construction of Presentation Review Environment with Realtime-created Hyper Video Technique," IEICE Technical Report Order Information, Educational Technology, vol.106, no.583, pp.133–138, 2007.
[2] K. Inazawa and A. Kashihara, "A Presentation Avatar for Self-Review," The 25th International Conference on Computers in Education, pp.345–354, 2017.
[3] K. Inazawa and A. Kashihara, "A Presentation Robot for Promoting Model-based Self-Review," The 26th International Conference on Computers in Education, pp.358–360, 2018.
[4] K. Inazawa and A. Kashihara, "Promoting Engagement in Self-Reviewing Presentation with Robot," 6th International Conference on Human-Agent Interaction, pp.383–384, 2018.
[5] T.H. Mita, M. Dermer, and J. Knight, "Reversed facial images and the mere-exposure hypothesis," Journal of Personality and Social Psychology, vol.35, no.8, pp.597–601, 1977.
[6] M. Low, "The Psychology Behind Video-Phobia," https://www.linkedin.com/pulse/psychology-behind-video-phobia-matthew-low.
[7] P.S. Holzman and C. Rousey, "The voice as a percept," Journal of Personality and Social Psychology, vol.4, no.1, pp.79–86, 1966.
[8] M. Goto, T. Ishino, K. Inazawa, M. Matsumura, T. Nunobiki, and A. Kashihara, "A Presentation System with Robot for Recomposing Presenter Behavior," Japanese Society for Information and Systems in Education 2017 fifth time workshop, vol.32, no.5, pp.121–128, 2017.
[9] L. Grez, M. Valcke, and I. Roozen, "How effective are self- and peer assessment of oral presentation skills compared with teachers' assessments?," Active Learning in Higher Education vol.13, no.2, pp.129–142, 2012.
[10] S. Kopf, D. Schön, B. Guthier, R. Rietsche, and W. Effelsberg, "A Real-time Feedback System for Presentation Skills," Proc. EdMedia 2015, pp.1686–1693, 2015.
[11] J. Schneider, D. Börner, P. Rosmalen, and M. Specht, "Can you help me with my pitch? Studying a tool for real-time automated feedback," IEEE Transactions on Learning Technologies, vol.9, no.4, pp.318–327, 2016.
[12] M. Chollet, T. Wörtwein, L.-P. Morency, A. Shapiro, and S. Scherer, "Exploring feedback strategies to improve public speaking: an interactive virtual audience framework," Proc. 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing, pp.1143–1154, 2015.
[13] H. Trinh, R. Asadi, D. Edge, and T. Bickmore, "RoboCOP: A Robotic Coach for Oral Presentations," Proc. ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies, vol.1, no.2, article no.27, pp.1–24, 2017.
[14] A. Melinger and W.J.M. Levelt, "Gesture and the communicative intention of the speaker," Gesture, vol.4, no.2, pp.119–141, 2004.
[15] C.D. Kidd and C. Breazeal, "Effect of a robot on user perceptions," 2004 IEEE/RSJ International Conference on Intelligent Robots and Systems, vol.4, pp.3559–3564, 2004.
[16] T. Tanaka, K. Fujikake, T. Yonekawa, M. Yamagish, M. Inagami, F. Kinoshita, H. Aoki, and H. Kanamori, "Driver Agent for Encouraging Safe Driving Behavior for the Elderly," HAI 2017 5th International Conference on Human-Agent Interaction, pp.71–79, 2017.

[17] H.L. O'Brien, "Antecedents and learning outcomes of online news engagement," Journal of the Association for Information Science and Technology, vol.68, no.12, pp.2809–2820, 2017.
[18] F. Redzuan, A.M. Lokman, Z.A. Othman, and S. Abdullah, "Kansei design model for engagement in online learning: a proposed model," International Conference on Informatics Engineering and Information Science, vol.251, pp.64–78, 2011.
[19] F. Dochy, M. Segers, and D. Sluijsmans, "The use of self-, peer and co-assessment in higher education: A review," Studies in Higher education, vol.24, no.3, pp.331–350, 1999.
[20] D. McNeill, "Hand and Mind," The University of Chicago Press, USA, 1995.
[21] D. Nancy, Slide:Ology: The Art and Science of Creating Great Presentations, Oreilly & Associates Inc., USA, 2008.
[22] H. Kamide, K. Kawabe, S. Shigemi, and T. Arai, "Non-verbal behaviors toward an audience and a screen for a presentation by a humanoid robot," Artificial Intelligence Research, vol.3, no.2, p.57, 2014.
[23] T. Ishino, M. Goto, and A. Kashihara, "A Robot for Reconstructing Presentation Behavior in Lecture," The 6th International Conference on Human-Agent Intraction (HAI2018), pp.67–75, 2018.
[24] VStone, "Sota," https://sota.vstone.co.jp/home/.
[25] Unity Technologies, "Products - Unity," https://unity3d.com/jp/unity.
[26] Microsoft Corporation, "Kinect for Xbox One | Xbox," https://www.xbox.com/en-US/xbox-one/accessories/kinect.
[27] Rumen F, "Kinect v2 Examples with MS-SDK and Nuitrack SDK," https://rfilkov.com/2014/08/01/kinect-v2-with-ms-sdk/.
[28] H.L. O'Brien, P. Cairns, and M. Hall, "A practical approach to measuring user engagement with the refined user engagement scale (UES) and new UES short form," International Journal of Human-Computer Studies, vol.112, pp.28–39, 2018.

Keisuke Inazawa received the B.S. and M.S. degrees in Informatics and Engineering from The University of Electro-Communications in 2017 and 2019 respectively. His research interests lie in designing robot systems for learning and education.

Akihiro Kashihara received the B.S. and M.S. degrees in Computer Science and Systems Engineering from Tokushima University in 1987 and 1989 respectively, and the Ph.D. in Information and Computer Science from Osaka University in 1992. He was an assistant/associate professor of the Institute of Scientific and Industrial Research in Osaka University from 1992 to 2004. He is now a professor of Graduate School of Informatics and Engineering in The University of Electro-Communications. His main research interests lie in designing of learning models, robot in learning, meta-cognition in hyperspace, knowledge externalization, cognitive loading, and scaffolding with cognitive tools.