Jiang, B. et al. (Eds.) (2025). Proceedings of the 33rd International Conference on Computers in Education. Asia-Pacific Society for Computers in Education

Audience Robot Control System Adapted to Presenter's Skills for Presentation Practice

Yuya KISHIMOTO[a*] & Tomoko KOJIRI[b]
[a] *Graduate School of Science and Engineering, Kansai University,* Japan
[b] *Faculty of Engineering Science, Kansai University,* Japan
*k504134@kansai-u.ac.jp

Abstract: During a presentation, a presenter must adapt his/her style by observing audience reactions. To help develop this skill, we have developed a system with an audience robot that reacts according to the presenter's non-verbal expressions. However, since there are multiple non-verbal expression, it is difficult for a presenter to modify all of them at once. The objective of this research is to develop a system that allows presenters to practice multiple non-verbal expressions step-by-step, by recommending specific expressions tailored to his/her skill level. The system uses the robot's sensors to assess the presenter's proficiency with each non-verbal expression. It then assigns target expressions for practice and configures the robot to react solely to those expressions, ensuring a focused training session.
Keywords: Presentation support, communication robot, non-verbal expression, presentation skill, training

1. Introduction

A presentation is an activity in which a presenter conveys his/her point of view to an audience. The presenter effectively communicates with audiences by utilizing verbal and non-verbal expressions. Since an understandable presentation differs depending on the audience and the environment, it is not always possible to give a good presentation. Therefore, it is necessary to observe the audience's reactions and modify one's own presentation according to the audience's reaction (this action is referred to "modify behavior").

Researches exist that support modifying the presenter's behavior during the presentation. (Kopf, Schön, Guthier, Rietsche, & Effelsberg, 2015) (Schneider, Börner, Rosmalen, & Specht, 2015). They evaluate whether the presenter's behaviors are in keeping with predefined good verbal and non-verbal expressions and provide direct feedback whether presenter's behaviors are appropriate in real time. Such researches encourage presenters to have predefined behavior, but it is not possible to cultivate the modify behavior. During the modify behavior, the presenter observes the audiences and modifies behavior dynamically according to the audience's reactions. The verbal and non-verbal expressions that audiences perceive as effective vary among individuals. Therefore, it is difficult to define these factors in predefined.

Palmas et al. deployed a virtual audience in a VR space that provides real-time feedback in reaction to the presenter's non-verbal expressions (Palmas, Cichor, Plecher, & Klinker, 2019). The necessity of special equipment to utilize VR space makes this system difficult to use casually. In addition, there are possibility of inducing VR sickness.

In order to solve such problems, we have developed an environment in which presenters can practice their presentations anytime, anywhere by developing an audience robot (Kato, Kunori, & Kojiri, 2024). This audience robot changes its internal state, which expresses emotion, in reaction to non-verbal expressions acquired by its sensors, and expresses its behavior in reaction to its internal state. Using this robot, it is possible to practice the activity of modifying non-verbal expressions in the audience robot's reactions. However, since there are multiple non-verbal expressions, it may be difficult for the presenter to modify all the non-verbal expressions at once. Therefore, the objective of this research is to develop the system, which we name audience robot control system, that determines non-verbal expressions to be practiced according to the presenter's skill in each non-verbal expression. The target non-verbal expressions and their order are determined so that the presenter can start practicing with one non-verbal expression and increase the number of non-verbal expressions to be considered step by step until he/she finally becomes master of all the non-verbal expressions. In addition, this research extends the audience robot so as to adjust its reactions based on the determined non-verbal expressions for each practice session.

In the field of the intelligent tutoring systems, learning materials are determined based on the learner's understanding state and skills (Wenger, Brown, & Greeno, 1987). The system has materials, which are the contents of the learning, and the system selects content from the learning materials based on a learner model that represents the learner's state of skills. The audience robot control system we develop is regarded as an intelligent tutoring system.

2. Approach

Figure 1 shows the configuration of the audience robot control system and the audience robot.

We have developed an audience robot that reacts according to the quality of the presenter's non-verbal expressions (Kato, Kunori, & Kojiri, 2024). The robot uses various sensors to detect non-verbal expressions with its non-verbal expression detection function. Emotion generation function updates the emotion model of the audience robot based on the non-verbal expressions detected by the non-verbal expression detection function, and the reaction corresponding to the internal model is expressed by the reaction generation function.

In order for the audience robot to react to different non-verbal expressions depending on the presenter's skill, the emotion generation function should be able to update its model only from specific non-verbal expressions. Therefore, we introduce evaluation method determination function that dynamically changes the evaluation function of the non-verbal expressions used in the emotion generation function. In addition, sensor setting function is introduced to adjust sensor list to detect only the non-verbal expression of the target practice.

The evaluation function is set for each non-verbal expression and has a list of threshold values that represent the maximum and minimum of the range of having positive emotions, respectively. A positive emotion indicates when the value of the presenter's non-verbal expression is between the maximum and minimum values of the threshold, and a negative emotion shows the value is greater than the maximum or less than the minimum value. The audience robot gives positive or negative behaviors according to its emotion. In order to practice modify behavior, audience robot needs to give negative reaction for the target non-verbal expression. Therefore, the evaluation method determination function adjusts the range of maximum and minimum values according to the presenter’s usual non-verbal expression, which we call baseline behavior. That is, the range does not include the baseline behavior.

On the other hand, the audience robot control system maintains practice contents that consists of set of non-verbal expressions to be practiced and their order relations. It also has a user model that represents presenter’s skills. The user model is developed by evaluating the practice history with audience robot using user model development function.

The audience robot control function determines the non-verbal expressions targeted by the audience robot and sends the maximum or the minimum value values. The audience robot control function has two phases. First, a preparation phase that operates at the beginning of practice. Second, a practice phase that operates during the practice. The preparation phase is for gathering the baseline behavior (item 1) and for checking whether the presenter can perform the modifier behavior (item 2). For item 1, it makes presenter practice with audience robot which does not give any reaction. For item 2, it asks presenter to practice with audience robot with sensors of all non-verbal expressions and give reactions to all non-verbal expressions. In practice phase, it controls the audience robot by referring to the user model and the practice contents.

Currently, eye contact, speaking speed and volume of voice are the target non-verbal expressions.

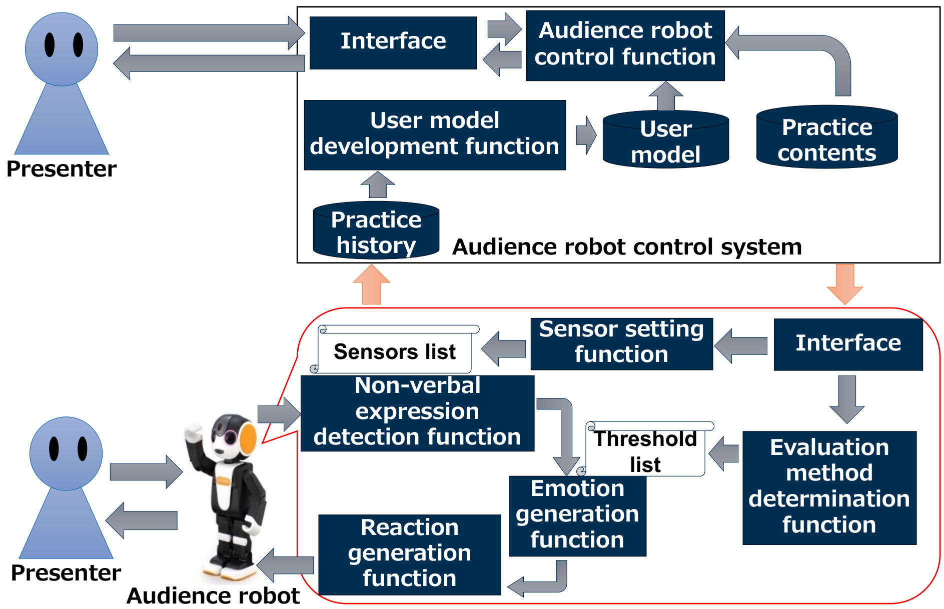

Figure 1. System Configuration of Audience Robot Control System

3. User Model Development Function

The user model development function identifies whether the presenter could control the non-verbal expressions based on the presenter's behavior during the presentation with the audience robot. In order to identify the well-controlled non-verbal expressions, it is necessary to grasp whether they are trying to modify the non-verbal expressions when the audience's reaction is negative, or whether they are trying to keep the non-verbal expressions when the audience's reaction is positive.

If the presenter tries to modify the inappropriate behavior or keeps the good behavior for most of the time, the presenter is regarded to perform the modify behavior. The user model development function calculates the average and standard deviation of the behavior in the previous 30 seconds and sets the range of the baseline behavior as *average* – *standard deviation* to *average* + *standard deviation*. Then, it detects the behavior as different from the baseline behavior if it is outside the range of the average and standard deviation. For the first 30 seconds, the average and standard deviation of the preparation phase behavior are considered as the baseline behavior.

On the other hand, if the presenter is trying to keep his/her own behavior, the good behavior will be kept.

The presenter usually does not always monitor the audiences and change his/her behavior, so if the presenter controls non-verbal expression, such as keeps or modifies the target non-verbal expression, more than 50 % of the presentation, the presenter is succeeded in modifying behavior. When there are more than one target non-verbal expression, if more than one non-verbal expression is well-controlled, the presenter is regarded to master the modify behavior of the target non-verbal expressions.

4. Practice Contents

The practice contents consists of a set of non-verbal expressions to be practiced (*contents*) by presenter and their relations. Since the final goal of the presenter is to control non-verbal expressions in the modify behavior, practice with all non-verbal expressions is the last contents to be practiced. Practicing with targeting only one non-verbal expression is the easiest, and difficulty level increases as the number of non-verbal expressions increases. Based on these considerations, all possible sets of non-verbal expressions are defined as contents and their inclusive relation was attached between them.

Figure 2 shows an example of the practice contents when eye contact, volume of voice, and speaking speed were selected as target non-verbal expressions.

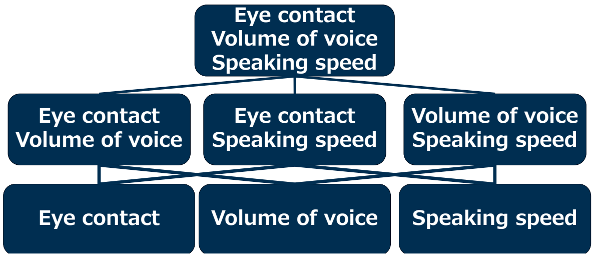

Figure 2. Example of Practice Contents

5. Audience Robot Control Function

Audience robot control function identifies a sequence of contents to be practiced by the presenter from the materials based on the user model.

5.1 User Model

The User model maintains the non-verbal expressions that presenter could control, his baseline behavior identified in the preparation phase such as their average and standard deviations of non-verbal expressions. In addition, the model maintains the contents used for practice and their achievement status.

5.2 How to Decide Which Non-verbal Expression to Practice

Figure 3 shows a flowchart for determining the practice contents. If the presenter could control all non-verbal expressions to modify behavior, the practice ends. If there are non-verbal expressions which the presenter is not able to control, the content consisting only of that non-verbal expression is first selected as practice content. If there are ancestor nodes for each practice content, they are selected as the practice content. Since the parent nodes of different non-verbal expressions may be the same node, the content is selected only once. The selected contents are provided, starting from the lower layer to the upper.

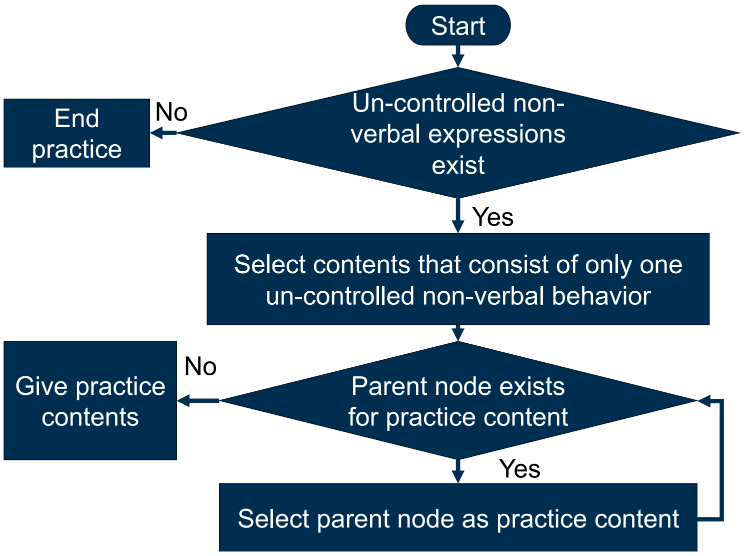

Figure 3. Practice Contents Selection Algorithm

6. Evaluation Experiment

6.1 Overview

An experiment was conducted to evaluate the effectiveness of the presentation practice environment developed in this research. 10 undergraduate and graduate students (A~J) participated in the experiment.

The experimental steps were as follows.

1. Participants give a presentation in the preparation phase (pre-presentation)
2. Participants were divided into two groups, WITH audience robot control function (*Group E*) and WITHOUT audience robot control function (*Group C*), and each group practiced 7 times.
3. Participants give a presentation in the preparation phase (post-presentation)

In this experiment, the participants were asked to give a presentation to an audience robot. Slides and dialogue for the presentation were prepared by authors. The presentation themes and the number of slides used in the experiment are shown in Table 1. The lengths of the slides were about 3 to 4 minutes. Different slides were used for the pre-presentation, practice, and post-presentation to avoid the influence of the theme. Slides and dialogue were given to the participants before the presentation, but they were allowed to change the dialogue. During the presentation, we explained that the audience robot would evaluate the presentation and express its reaction according to the values of the non-verbal expressions displayed on the interface, and instructed the participants to change their non-verbal expressions when they felt that the audience robot's reaction was negative so as to improve the audience robot's reaction.

In step 1, the presentation is performed in the preparation phase. At the end of Step 1, the audience robot control system identifies the presenter's skill using the data from Step 1. In Step 2, the number of participants in the group WITH and WITHOUT the audience robot control function was equal. For both groups, participants were asked to give presentation for seven times. For participants without the audience robot control function, the three combinations of non-verbal expressions were presented every time. In Step 3, the presentation was conducted again in preparation phase to identify the presenter's skills.

Table 1. *Presentation Theme and the Number of Slides*

Presentation theme	The number of slides
1: Introduction of Kansai University	7
2: What is Chat GPT?	6
3: Introduction of Kojiri Laboratory	5

6.2 Result

Let's evaluate the effectiveness of the audience robot control function in item 1. Table 2 shows the percentage of participants who were able to perform the modify behavior in the pre-presentation and the post-presentation. A *t-test* was conducted between the pre-condition and the post-condition. The result was $p>0.05$ ($p=0.768$), indicating that there was not a significant difference. Table 3 shows the practice contents of *Group E* in all seven practice sessions, where "E" is eye contact, "V" is volume of voice, and "S" is speaking speed. All experimental participants in *Group E* could not reach to the final practice contents in seven sessions. Therefore, when the number of presentation practice is limited, the participants were not always able to perform all practice contents and, thus, were not guaranteed to improve their modify behavior.

Next, we evaluate the load on practice using the audience robot control function. If the practice is appropriate for the participants, the percentage of participants who could perform modify behavior increases. are controlled should maintain or improve with each practice session. For this purpose, we investigated the percentage of improvement in the target non-verbal expressions between exercises of the same practice contents in *Group E*. Table 4 shows the number of practice that participants continued the same practice contents and the ratio of achieving modify behavior during these practice. The *t-test* was conducted to examine whether there was a significant difference between *Group E* and *Group C*. The result was $p=0.047$ ($p<0.05$), indicating a statistically significant difference. This indicates that the load of modify behavior was lowered in *Group E* by practicing step-by-step.

The participant in *Group E* commented, "By observing the feedback from the audience robot control system and the reactions of the audience robot, I was able to notice that the speaking speed was changing unconsciously," and "By practicing in a step-by-step practice, it became easier to control the non-verbal expression and modify behavior." These results suggest that the audience robot control function is effective in acquiring modify behavior.

Table 2. *Percentage of Participants Who were Able to Perform Modify Behavior*

Group	Participants	Pre-presentation	Post-presentation
E	A	0.667	0.800
	B	0.833	1.000
	C	0.300	0.333

	D	0.846	0.600
	E	0.571	0.700
C	F	0.750	0.929
	G	0.600	0.625
	H	0.286	0.667
	I	0.667	0.300
	J	0.909	0.667

Table 3. *Practice Contents of Group E*

	1	2	3	4	5	6	7
A	E	E	E	V	S	S	S
B	E	E	S	S	S	S	S
C	E	V	V	V	V	V	V
D	E	S	S	S	S	EV	EV
E	V	V	V	V	S	S	S

Table 4. *The Number of Times Contents are Repeated and Rate of Improvement*

Group	Participants	The number of times practice contents are repeated	Improvement rate
E	A	4	0.75
	B	5	0.60
	C	5	1.00
	D	3	0.75
	E	5	0.80
C	F	6	0.67
	G	6	0.50
	H	6	0.50
	I	6	0.50
	J	6	0.75

7. Conclusion

In this research, we proposed a framework that determines the combination of non-verbal expressions to practice according to the presenter's non-verbal expression skills and adjusts the non-verbal expressions that the audience robot reacts in a presentation practice environment. In this research, the goal of learning is to present with controlled of all non-verbal expressions. In order to do so, the timing and the type of the non-verbal expressions to use important. It is considered that appropriate non-verbal expression to use depends on the intention to be conveyed by the non-verbal expressions. In the future, we need to analyze the contents that can be conveyed by each set of non-verbal expressions and expand the training environment to have the audience robot select the appropriate non-verbal expressions based on the type of reaction.

References

Kato, S., Kunori, K., & Kojiri,T. (2024). Development of Audience Robot for Supporting Presentation Rehearsal of Nonverbal Expression. Proceedings of EdMedia + Innovate Learning, 799-808.

Kopf, S., Schön, D., Guthier, B., Rietsche, R., & Effelsberg, W. (2015). A Real-time Feedback System for Presentation Skills. *Proceedings of In EdMedia+ Innovate Leaning: Association for the Advancement of Computing in Education*, 1686-1693.

Palmas, F., Cichor,J.E., Plecher,D. A., & Klinker, G., (2019). Acceptance and Effectiveness of a Virtual Reality Public Speaking Training. 2019 IEEE International Symposium on Mixed and Augmented Reality (ISMAR), 363-371.

Schneider, J., Börner, D., Rosmalen, P. v., & Specht, M. (2015). Presentation Trainer, your Public Speaking Multimodal Coach. *Proceedings of the 2015 ACM on International Conference on Multimodal Interaction*, 539-546.

Wenger, E., Brown, J. S., & Greeno, J. (1987). *ARTIFICIAL INTELLIGENCE and TUTORING SYSTEMS Computational and Cognitive Approaches to the Communication of Knowledge.* Morgan Kaufmann Publishers, Inc.