Toyota and Terai *Research and Practice in Technology Enhanced Learning* (2025) 20:40
https://doi.org/10.58459/rptel.2025.20040

Research and Practice in Technology Enhanced Learning

RESEARCH Free and Open Access

Proposal and evaluation of on-demand lecture video using avatars, humor, and questions to avoid boredom

Satoshi Toyota [1] and Asuka Terai *[2]

*Correspondence:
aterai@fun.ac.jp
Faculty of Systems Information Science,
Future University Hakodate,
116-2 Kamedanakano, Hakodate,
Hokkaido 041-8655, Japan
Full list of author information is available at the end of the article

Abstract

On-demand lectures have become widespread due to COVID-19, and while there are advantages to being able to watch lectures whenever students want, disadvantages such as not being able to concentrate on them have been pointed out. Then, it is necessary to devise a more preparatory method that is less boring. Thus, a method was proposed to generate humor and questions about important words in lecture scenarios by the lecturers' specification of the words, and create lecture videos for demonstration by two avatars to attenuate the problem of "boredom with on-demand lectures". Experiments were conducted to validate the proposed method, and since previous research suggested that learners' attitudes (preferences) toward an avatar affect their learning motivation, the influence of viewers' impressions of avatar lectures on the effectiveness of the proposed method was examined. The experimental results suggested that including humor or question in on-demand lectures affects resistance to boredom. Furthermore, results of the confirmation test conducted after viewing the videos indicated that the more natural the viewers perceived the lectures given by the avatars, the greater the positive effect on recognition of targeted words. The effects of the proposed method were shown especially for viewers who were familiar with avatar's lecture.

Keywords: On-demand lecture, Humor, Question, Avatar

Introduction

Since COVID-19, the need for on-demand lectures has increased, with on-demand videos also being utilized in flip-style lectures post COVID-19. However, it is necessary to devise a more preparatory method that is less boring, because some students find it difficult to concentrate on lectures when attending them from home. According to Osugi (2021), on-demand lectures can guarantee knowledge gain, but they are not as effective in achieving

independent, interactive, and in-depth learning, or in changing perceptions and thoughts. In addition to problems related to educational effectiveness, the burden on teachers to prepare on-demand class materials has also been pointed out. Matsushima (2020) noted that on-demand lectures require significant time for teachers to prepare, indicating an increased burden on faculty.

A previous study proposed a system in which a robot gave lectures at a university on behalf of lecturers (Hiyori et al., 2018). One of the problems faced in this scenario is the mechanical nature of the robots, inducing boredom in students. To avoid student boredom, the insertion of questions into a robot lecture scenario was proposed (Yashio & Araki, 2019). Watanabe et al. (2020) further suggested that an avatar of the learner's own liking may generate interest in the content and motivate the learner to take the next class.

In addition, a previous study (Yamashiro et al., 2020) on incorporating humor into educational settings reported that adding toilet humor to kanji drills can positively change students' grades and concentration and suggested humor has a positive impact on learning. The study found that students who used question books incorporating humor had a higher percentage of correct answers compared with students who used traditional question books. In addition, when concentration was measured using fixed-point camera imaging, students who used humor-infused materials were found to concentrate from an earlier stage. Furthermore, Suzuki et al. (2021) found that incorporating a comic dialogue-like script structure into a presentation improved audience awareness, empathy, and familiarity.

As such, previous studies (Suzuki et al. 2021; Yamashiro et al. 2020; Yashio & Araki, 2019) indicate that incorporating humor and interaction within on-demand lectures may be effective in reducing boredom. However, adding humor or questions as interactive elements to the regular lecture scripts increases the burden on the faculty preparing for on-demand lectures. This paper proposes a method to generate automatically humor and questions for important words in lecture scenarios by the lecturers' specification of the words, and create lecture videos for demonstration by two avatars to attenuate the problem of "boredom with on-demand lectures." A lecture scenario incorporating humor is generated by adding "Boke [jokes]" and "Tsukkomi [corrections]" regarding important words to the lecture scenario. The lecture scenarios follow a conversational format in which two avatars proceed with humor to soften the strict atmosphere. In addition, questions to the viewers regarding important words are added for relief from loneliness and more active learning. The avatar asks students questions and gives them time to think, thus creating an active learning environment that makes the lectures less boring. Experiments were conducted to validate the proposed method. Since previous research (Watanabe et al., 2020) suggests that learners' attitudes (preferences) toward an avatar affect their learning motivation, the influence of viewers' impressions of avatar lectures on the effectiveness of the proposed method was examined.

Method

We proposed a method to generate automatically humor consisting of "Boke [joke]" and "Tsukkomi [correction]," and questions regarding important words in lecture scenarios by lecturers' selection of the words. Two avatars representing a teacher and a student ("teacher avatar" and "student avatar") perform these scenarios incorporating humor and questions to prevent student (viewer) boredom with the lecture videos. The teacher avatar presents the lecture, during which the student avatar makes a joke as a "Boke [joke]." Immediately afterwards, the teacher avatar makes a correction to the joke as a "Tsukkomi [correction]." The student avatar also asks questions directly to the viewers regarding important words from the lecture. Furthermore, the student avatar produces "Aizuchi [reaction]" to the teacher avatar's talk to make their dialogue more natural.

The lecture avatars are placed at both ends of the lecture video screen, as shown in Figure 1. Vroid Studio was used to generate the two avatars, with the teacher on the right side and the student on the left (see Figure 1). In addition, the lecture script was read aloud using Coe Font by avatars with an avatar lip-sync (the avatar's mouth moving in time with the voice) using 3tene. Open Broadcaster Software (OBS) was used to overlay a lecture slide and avatars to create a single lecture video. Subtitles were added using Adobe Premiere Pro (Figure1) in case it was difficult to hear the text.

Generation of "Aizuchi [reaction]"

The templates "Ho-ho," "So-Nanda," "Naruhodo," and "Hee" (Japanese "Aizuchi [reaction]" expressing an agreement like "I agree") were created, and one of them was randomly selected to make each "Aizuchi [reaction]." According to previous research (Kamiya et al., 2010), there is a 27.5% probability that "Aizuchi [reaction]" is given after auxiliary verbs ("Desu," "Masu," etc.) in a dialogue. Thus, assuming that the end of a sentence in a lecture scenario is almost always an auxiliary verb, an "Aizuchi [reaction]"

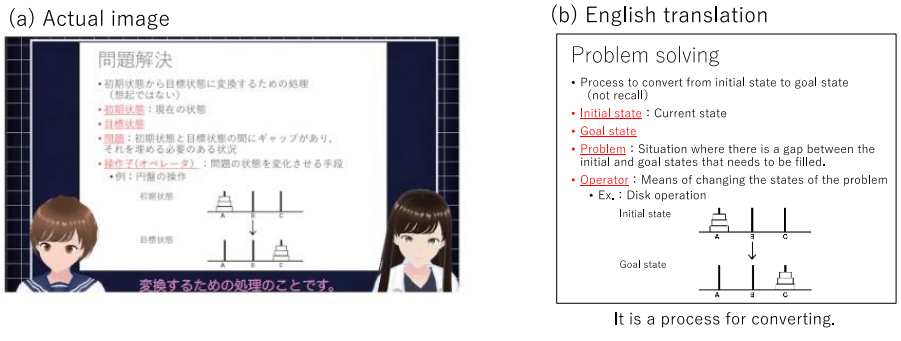

Fig. 1 Lecture video by two avatars. (a)Actual image in Japanese, (b) English translation

was inserted after the end of the sentence in the lecture avatar's talk at 27.5% probability to make lecture by two avatars more natural.

Generation of "Humor"

The proposed method aims to make the lecture scenario humorous based on the Japanese comedy style "Manzai [comic dialogue]," in which two comedians have a humorous conversation. In "Manzai [comic dialogue]," one speaker says a "Boke [joke]" and the other speaker corrects the joke by "Tsukkomi [correction]." Thus, "Boke [joke]" and "Tsukkomi [correction]" are generated for important words in the lecture scenarios.

Generation of "Boke [joke]"

A previous study on the automatic generation of comic dialogue (Yoshida & Hagiwara, 2012) proposed three types of "Boke [joke]": "Kotoba-Asobi-Boke [word game]," "Kajo Boke [excessive joke]," and "Mujun Boke [conflicting joke]." This study adopted the "Kotoba-Asobi-Boke [word game]," which can be easily applied to important words in a lecture.

In "Kotoba-Asobi-Boke (word game)," the Japanese target word is Romanized, and one consonant is replaced with another, as shown in Figure 2. The generated "Boke [joke]" must be a real word that is commonly heard in daily life, to ensure its meaning can be understood. Therefore, a textbook vocabulary corpus (Kyokasho Corpus, Goihyo) was used, containing words used in elementary, junior high, and high school textbooks, with each word written in Japanese in Chinese character and Romanization, along with its frequency of use. The consonants of the target word are randomly converted and compared with the words in the textbook corpus, and those that match the reading of the word in the textbook corpus with the highest frequency of use are used as "Boke [joke]." Words such as "death" that do not easily evoke laughter are removed from the "Boke [joke]" candidates. Words consisting of multiple morphemes (e.g., "Shusoku teki shikoh [thinking of convergence]") are split into their individual morphemes by MeCab (Kudo et al., 2004); one morpheme (e.g., "Shusoku [convergence]") other than a suffix (e.g., "teki [of]") become the target, and the consonant of the target morpheme is changed randomly to create the "Boke [joke]" (e.g., "Shutoku teki shikoh [thinking of learning])."

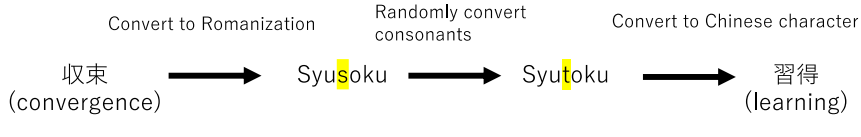

Fig. 2 Generation method of "Boke [joke]"

Generation of "Tsukkomi [correction]"

As a "Tsukkomi" against the generated "Boke [joke]," we adopted the "Setsumei Tsukkomi (explanation correction)" used by Yoshida and Hagiwara (Yoshida & Hagiwara, 2012). The "Setsumei Tsukkomi" provides negations and explanations for the "Boke [joke]." To generate explanations for "Boke [joke]," the meaning of the concept word used as the "Boke [joke]" is searched in the Japanese WordNet, a Japanese dictionary in which the relationships between concepts and definitions of words corresponding to concepts are registered. For words that consists of multiple morphemes, only words that are generated as a "Boke [joke]" for target morphemes are searched. For example, when "Shutoku [learning]" is generated as a "Boke [joke]" for "Shusoku [convergence]" in "Shusoku teki shikoh [thinking of convergence]," "Shutoku [learning]" is searched for in the Japanese WordNet to generate an explanation.

"Setsumei Tsukkomi" consisting of negations and explanations is generated by randomly combining one of three types of negations ("Why are you [Boke [joke]]?", "No [Boke [joke]]," and "[not saying anything]"), and one of three types of explanations ("[Boke [joke]] is [explanation]," "That is [explanation]," and "That is [explanation], isn't it?") (e.g., "No Shutoku [learning]. Shutoku [learning] is the acquisition of knowledge or ability."). These are based on the canned sentences in Yoshida and Hagiwara (2012) and were modified as "Tsukkomi" appropriate for the teacher avatar.

Question generation

Questions are generated based on the work of Yashio and Araki (2019), who noted that extracted questions have nouns similar to the important words in the lecture from the Corpus of Japanese classroom Lecture speech Contents (CJLC) (Tsuchiya et al., 2008). As such, questions can be generated by replacing the nouns in the questions with the important words in the lectures (Figure 3). Instead of the CJLC, this study used the Nagoya University Conversation Corpus (Fujimura et al., 2012) and the Corpus of Japanese Teacher Speech (Nihongo Kyoshi Hatsuwa Corpus), which is a corpus of transcribed conversations, mainly by graduate students, that can be used to generate more comprehensible questions for students. Moreover, the Corpus of Japanese Teacher Speech contains questions actually asked in lectures, and is thus more realistic. Questions that have nouns similar to the important words in the lecture are selected from these two corpora, and these nouns are replaced with the important lecture words. For example, for the important word "Shutoku teki shikoh [thinking of convergence]," the question "Shutoku teki shikoh suru koto arimasuka? [Do you ever do thinking of convergence?]" is generated.

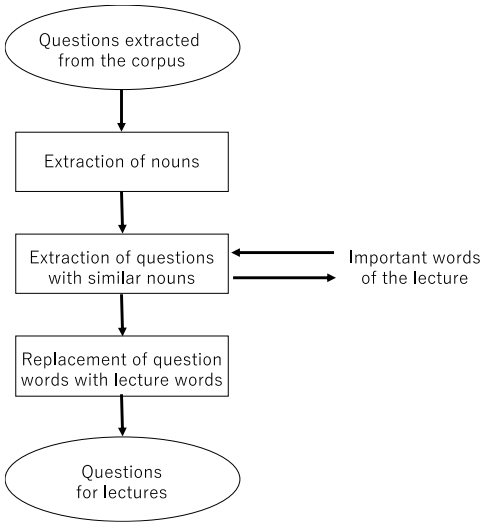

Fig. 3 Question generation method

Evaluation experiments

Two experiments were conducted to validate the proposed method.

Experiment1

The purpose of this experiment was to verify the effect of humor, consisting of "Boke [joke]" and "Tsukkomi [correction]," for "comprehensibility" and "boredom resistance."

Method

A total of 304 paid cloud workers recruited by Lancers participated in the experiment through a web-based survey (males =187, females = 111, No Answer = 6, Average Age = 42.29, Age SD = 9.26). Three types of lecture videos were used: (1) a lecture video in which a teacher avatar reads a lecture scenario without a student avatar (Teacher), (2) a lecture video in which a teacher avatar reads a lecture scenario and a student avatar only gives an "Aizuchi [reaction]" (Teacher+Student), and (3) a lecture video in which a teacher avatar reads a lecture scenario and "Tsukkomi [correction]," and a student avatar gives an "Aizuchi [reaction]" and "Boke [joke]" (Teacher+Student with Humor). In this study, the participants were not shown the original lectures because it was difficult to compare lectures given by avatars and actual teachers. Three types of existing scenarios and slides (lectures 1, 2, and 3) used in a cognitive science lecture related to creativity at Future University Hakodate were used. At the end of the questionnaire, the participants were asked to rate their learning experience of cognitive science on a seven-point scale. The

Table 1 Video watching conditions (Experiment1)

Lecture Type	Group 1	Group 2	Group 3
Lecture 1	Teacher	Teacher+Student	Teacher+Student with Humor
Lecture 2	Teacher+Student	Teacher+Student with Humor	Teacher
Lecture 3	Teacher+Student with Humor	Teacher	Teacher+Student

mean learning experience of cognitive science was 1.36, and the standard deviation was 0.89. The lecturer, who generated the scenarios and slides and was one of the authors, selected two important words for each scenario. To avoid an order effect, the participants were divided into three groups (Group 1: 93, Group 2: 105, and Group 3: 106; the number of participants in each group), and they watched the videos in the order of lectures 1, 2, and 3 under different conditions depending on the group (Table 1). The length of each lecture video was about 1.5 to 2 minutes, and viewers had time to answer questions after each video.

Immediately after watching each video, they were asked to rate how easy they found the videos to understand (comprehensibility) and how resistant they were to boredom (boredom resistance) on a 7-point scale using Google Forms. After watching all the videos, the participants were asked to rate the naturalness of the avatar's lecture videos on a 7-point scale, and to answer a four-choice confirmation test regarding the two important words in each video. One of the important words was a target word for "Boke [joke]" under the Teacher+Student with Humor condition.

Results

The naturalness ratings of the avatar lecture videos are shown in Figure 4(a). The participants were divided into two groups based on their naturalness ratings; participants who rated less than 4 (neutral) (low familiar group) can be regarded as unfamiliar with the avatar's lecture, and the other participants (high familiar group) can be regarded as being familiar with the avatar's lectures. Then, using R, a three-way ANOVA (group × video-watching condition × familiar group) with Greenhouse–Geisser epsilon adjustment was conducted to assess comprehensibility. A main effect of familiar group ($F(1,298)=70.7$, $p<.01$, $\eta^2=.12$; high familiar > low familiar) an interaction between group and video-watching condition was observed at the 1% level ($F(3.92,584.54)=40.1$, $p<.01$, $\eta^2=.07$), and multiple comparisons using Holm's method revealed significant differences (Figure 4(b),Total). The results showed significantly lower average ratings for the videos that were viewed last regardless of the condition (5%). Similarly, three-way ANOVA was conducted for boredom resistance. A main effect of the familiar group ($F(1,298)=49.8$,

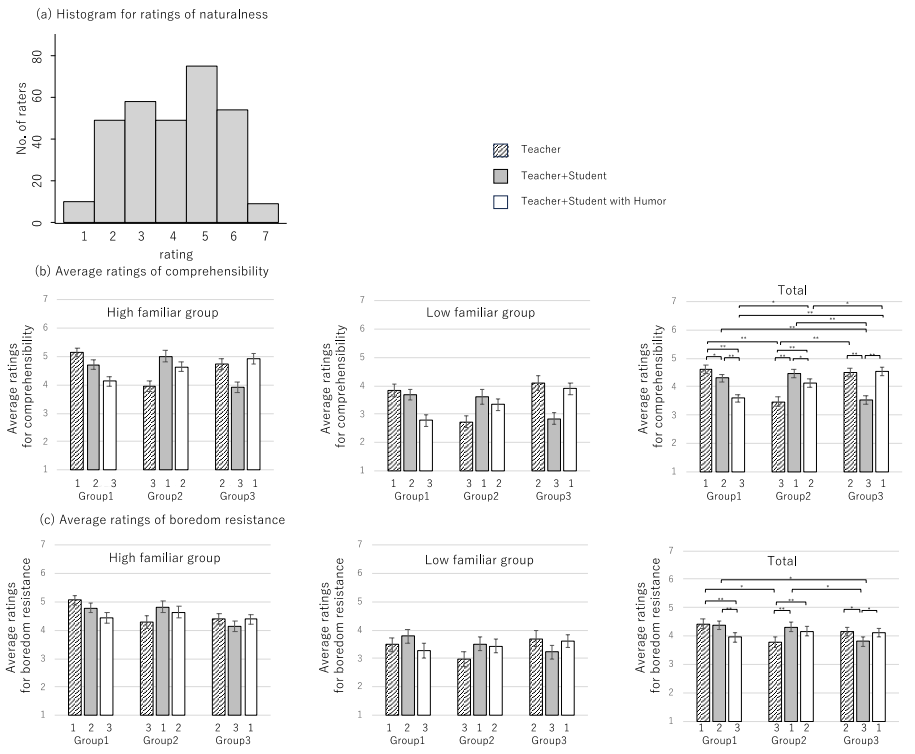

Fig. 4 Ratings of naturalness (a), average ratings of comprehensibility (b) and boredom resistance (c) with Teacher+Student + Humor condition

(** p<.01, * p<.05 · The numbers below the bars indicate the order of viewing the lectures)

p<.01, η^2=.11; high familiar > low familiar) and an interaction between group and video-watching condition were observed at the 1% level (F(4,596)=9.59, p<.01, η^2=.01), and multiple comparisons using the Holm method revealed significant differences (Figure 4(c),Total). The average rating was higher for the first video viewed than for the last video in the Teacher condition, and the average rating was significantly lower for the last video viewed than for the others in the Teacher+Student condition (5%). However, no simple main effect was observed in the Teacher+Student with Humor condition.

To compare correct response rates for questions about the important word with “Boke [joke]” and the other without “Boke [joke],” a McNemar test with Bonferroni correction was conducted using the results of the confirmation test under Teacher+Student with Humor condition. The results are shown in Table 2. The rates of correct responses to questions regarding important words with and without “Boke [joke]” were 74.3% and 66.3% in the high familiar group, and 58.2% and 53.0% in the low familiar group, respectively, which was significant trend at the 10% level in the high familiar group ($\chi^2(1)$=5.00, p<. 1, g=.17), but not in the low familiar group ($\chi^2(1)$=2.25, n.s., g=.19).

Table 2 Results of confirmation test regarding important words with and without "Boke [joke]" under the Teacher+Student with Humor condition

High familiar group				Low familiar group			
		With "Boke [joke]"				With "Boke [joke]"	
		Correct	Incorrect			Correct	Incorrect
Without "Boke [joke]"	Correct	109	15	Without "Boke [joke]"	Correct	57	5
	Incorrect	30	33		Incorrect	11	44

Discussion

The comprehensibility results show the lowest average ratings for the videos that were viewed last, regardless of the condition. Simple main effects were observed among the groups in all conditions. Similarly, the boredom resistance results indicate the effect of the viewing order. The ratings for comprehensibility and boredom resistance were significantly lower for the last video viewed than for the first video viewed in each group, regardless of the condition. Meanwhile, simple main effects were observed among groups in the Teacher and Teacher+Student conditions, but not in the Teacher+Student with Humor condition. These results suggest that the inclusion of humor in a lecture video prevents the viewer from being influenced by the viewing order and may make them concentrate on the lecture until the end. Both comprehensibility and boredom resistance showed significantly higher average ratings in the high familiar group than in the low familiar group. However, effects of the viewing order on comprehensibility and boredom resistance were observed in both groups.

The results of the confirmation test under the Teacher+Student with Humor condition showed that the rate of correct responses for important words with "Boke [joke]" was significantly higher than without "Boke [joke]" in the high familiar group. Although the rate with "Boke [joke]," as compared to without "Boke [joke]," was also higher in the low familiar group, there is no significant difference. It can be said that the inclusion of "Boke [joke]" for important words may have an effect on word recognition, especially when viewers perceive the avatar's lecture as natural.

The experiment was conducted as a web-based survey, and viewers (participants) came from diverse backgrounds. An experiment was conducted as a pilot study for Experiment1 using the same method with 24 undergraduate and graduate students at an information technology university, who were divided into three groups of the same number of participants. The results of these experiments are presented in Figure 5. A two-way ANOVA (group × condition) revealed no main effect or interaction, owing to limitations in the number of participants. However, the ratings of comprehensibility and boredom resistance for humorous videos were relatively high, regardless of the order. Given that students belonging to an information technology university were more likely to be familiar

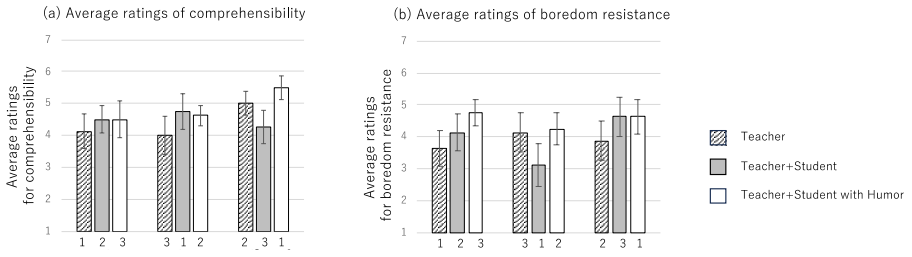

Fig. 5 Average ratings of comprehensibility (a) and boredom resistance (b) in the pilot study of Experiment1

with avatars, it can be said that familiarity with avatars may enhance the effects of humor and questions on viewers in on-demand lecture videos.

Experiment2

The purpose of this experiment was to verify the effect of questions for "comprehensibility" and "boredom resistance."

Method

A total of 641 paid cloud workers recruited by Lancers participated in the web-based survey experiment. Of these, only 562 participants' data were used (male = 308, female = 253, No Answer = 1, Average Age = 42.25, Age SD = 10.72), as 79 participants had also participated in Experiment1, and were hence not included. As in Experiment1, the target lecture was a cognitive science course on creativity. The mean learning experience in cognitive science was 1.39, and the standard deviation was 0.95. Four types of lecture videos were used: (1) a teacher avatar reads a lecture scenario without a student avatar (Teacher), (2) a teacher avatar reads a lecture scenario and a student avatar asks a question about an important word and gives an "Aizuchi [reaction]" (Teacher+Student with Q), (3) a teacher avatar reads a lecture scenario and a student avatar asks two questions about important words and gives an "Aizuchi [reaction]" (Teacher+Student with 2Qs), and (4) a teacher avatar reads a lecture scenario and "Tsukkomi [correction]," and a student avatar gives an "Aizuchi [reaction]," "Boke [joke]," and question (Teacher+Student with Q+Humor). As in Experiment1, three types of scenarios and slides were used (lectures 1, 2, and 3), and the participants were not shown the original lecture.

To avoid order effect, the participants were divided into six groups (Group 4: 97, Group 5: 92, Group 6: 96, Group 7: 95, Group 8: 89, and Group 9: 93; the number of participants in each group), and they watched the videos in the order of lecture 1, 2, and 3 under different conditions depending on the group (Table 3). It was assumed that the

Table 3 Video watching conditions (Experiment2)

With Q condition			
Lecture Type	Group 4	Group 5	Group 6
Lecture 1	Teacher	Teacher+Student with Q	Teacher+Student with Q+Humor
Lecture 2	Teacher+Student with Q	Teacher+Student with Q+Humor	Teacher
Lecture 3	Teacher+Student with Q+Humor	Teacher	Teacher+Student with Q

With 2Qs condition			
Lecture Type	Group 7	Group 8	Group 9
Lecture 1	Teacher	Teacher+Student with 2Qs	Teacher+Student with Q+Humor
Lecture 2	Teacher+Student with 2Qs	Teacher+Student with Q+Humor	Teacher
Lecture 3	Teacher+Student with Q+Humor	Teacher	Teacher+Student with 2Qs

questions were selected by the lecturer (user); thus, each question used in Experiment2 was selected from 10 generated candidate questions according to the results of a pilot experiment in which seven undergraduate and graduate students were asked to rate the questions on a 5-point scale. As in Experiment1, the length of each video was about 1.5 to 2.0 minutes, and viewers were asked to rate their naturalness and answer four-choice confirmation test after watching all the videos.

Results

The results of the naturalness of the avatar lecture videos are shown in Figure 6(a) using groups with the Teacher+Student + Q condition (Groups 4, 5, and 6). Similarly, the participants were divided into two groups based on their naturalness rating. Using the statistical software R, a three-way ANOVA (group × condition × familiarity group) was conducted for comprehensibility. A main effects of group ($F(2,279)=5.42$, $p<.01$,$\eta^2=.02$; Group4, Group6 > Group5) and familiar group ($F(1,279)=74.9$, $p<.01$, $\eta^2=.13$; high familiar > low familiar) and an interaction between group and condition ($F(3.95,551.37)=25.5$, $p<.01$, $\eta^2=.06$) was observed at the 1% level, and multiple comparisons using Holm's method revealed significant differences (Figure 6(b), Total). Similarly, three-way ANOVA was conducted for boredom resistance. A main effect of familiar group ($F(1,279)=58.0$, $p<.01$, $\eta^2=.12$; high familiar > low familiar) an interaction between group and condition ($F(4,557.38)=14.14$, $p<.01$, $\eta^2=.03$) was observed at the 1% level, and multiple comparisons using the Holm method revealed significant differences (Figure 6(c),Total).

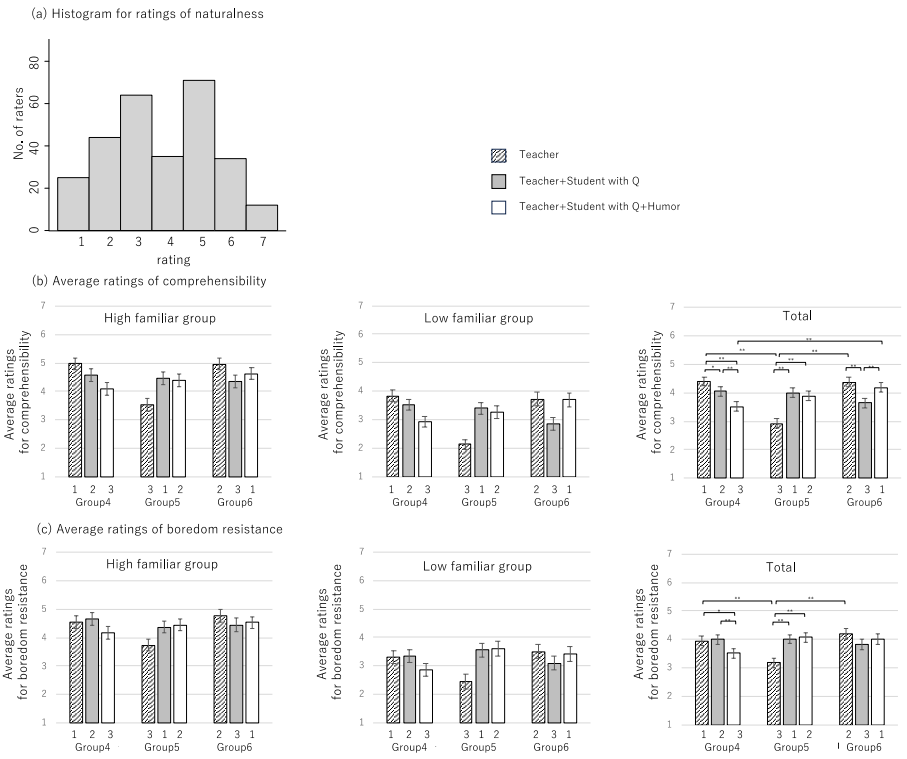

Fig. 6 Ratings of naturalness (a), average ratings of comprehensibility (b) and boredom resistance (c) of groups with Teacher+Student + Q condition

(** p<.01, * p<.05, The numbers below the bars indicate the order of viewing)

In addition, the results of the naturalness of the avatar lecture videos are shown in Figure 7(a) using groups with the Teacher+Student + 2Qs condition (Groups 7, 8, and 9), and the participants were divided into two groups based on the ratings. A three-way ANOVA was performed using the groups to assess comprehensibility. A main effect of the familiar group ($F(1,271)=45.5$, $p<.01$, $\eta^2=.09$; high familiar > low familiar) and an interaction between group and condition ($F(3.99,540.21)=12.4$, $p<.01$, $\eta^2=.03$) was observed at the 1% level, and multiple comparisons using Holm's method revealed significant differences (Figure 7(b), Total). A three-way ANOVA was conducted for resistance boredom. A main effect of the familiar group ($F(1,271)=26.4$, $p<.01$, $\eta^2=.06$; high familiar > low familiar) and an interaction between group and condition ($F(3.86,522.66)=6.15$, $p<.01$, $\eta^2=.01$) was observed at the 1% level, and multiple comparisons using Holm's method revealed significant differences (Figure 7(c), Total).

To compare correct response rates for questions about the important words with and without a question, McNemar's test was conducted using the results of the confirmation test under the Teacher+Student with Q condition using Bonferroni correction (Groups 4, 5, and 6). The number of participants for each question is presented in Table 4. The rates of

correct responses to questions regarding important words with and without question were 71.1% and 55.3% in the high familiar group, and 57.1% and 48.9% in the low familiar group, respectively, which were significantly different at the 1% level in the high familiar group ($\chi^2(1)=12.5$, p<. 01, g=.26), but significant trend at the 10% level in the low familiar group ($\chi^2(1)=4.17$, p<.1, g=.19).

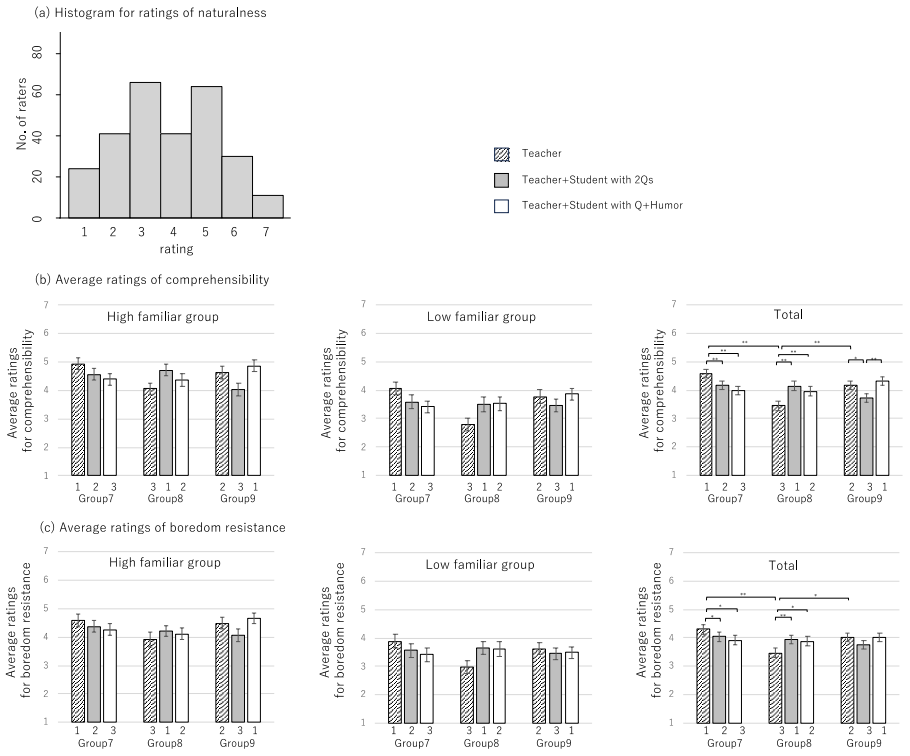

Fig. 7 Ratings of naturalness (a), average ratings of comprehensibility (b) and boredom resistance (c) of groups with Teacher+Student + 2Qs condition

(** p<.01, * p<.05, The numbers below the bars indicate the order of viewing)

Table 4 Results of confirmation test regarding important words with and without a question under Teacher+Student with Q condition

High familiar group				Low familiar group			
		With question				With question	
		Correct	Incorrect			Correct	Incorrect
Without question	Correct	73	11	Without question	Correct	56	9
	Incorrect	35	33		Incorrect	20	48

To compare correct response rates for questions about important words with a question and with a "Boke [joke]," a χ^2 test with Bonferroni correction was conducted using results of the confirmation test under the Teacher+Student with Humor condition (Groups 1, 2, and 3) and Teacher+Student with Q condition (Groups 4, 5, and 6). No significant differences in both familiar groups were observed at the 5% level (high: $\chi^2(1)=0.31$, n.s., $\emptyset=.23$, low: $\chi^2(1)=0.00$, n.s., $\emptyset=.01$).

Furthermore, to compare correct response rates for questions about important words with a question and with a "Boke [joke]" in the same video, McNemar's test with Bonferroni correction was conducted using results of the confirmation test under the Teacher+Student with Q+Humor condition. The number of participants who responded to the questions is presented in Table 5. The rates of correct responses to questions about important words with question and with "Boke [joke]" were 57.0% and 65.1% in the high familiar group, and 58.0% and 60.6% in the low familiar group, respectively, which were significantly different at the 5% level in the high familiar group ($\chi^2(1)=6.86$, $p<.05$, $g=.14$), but not in the low familiar group ($\chi^2(1)=0.83$, n.s., $g=.06$).

To compare effects of questions and "Boke [joke]" when a second question was included in the same lecture video, a McNemar's test with Bonferroni correction was conducted using results of the confirmation test under the Teacher+Student with Q+Humor condition in Teacher+Student with 2Qs condition (Groups 7, 8, and 9). The rates of correct responses to the questions about important words with a question and a "Boke [joke]" were 61.0% and 71.2% in the high familiar group ($\chi^2(1)=3.81$, n.s., $g=.13$), and 60.3% and 64.1% in the low familiar group ($\chi^2(1)=0.38$, n.s., $g=.04$), respectively (Table 6), with no significant difference between both groups.

Table 5 Results of confirmation test for important words with a question and a "Boke [joke]" in the same lecture video under Teacher+Student with Q+Humor condition

High familiar group				Low familiar group			
		With question				With question	
		Correct	Incorrect			Correct	Incorrect
With "Boke [joke]"	Correct	140	54	With "Boke [joke]"	Correct	127	33
	Incorrect	30	74		Incorrect	26	78

Table 6 Results of confirmation test for important words with a question and a "Boke [joke]" when another question was included in the same lecture video under Teacher+Student with Q+Humor condition and Teacher+Student with 2Qs condition

High familiar group				Low familiar group			
		With question				With question	
		Correct	Incorrect			Correct	Incorrect
With "Boke [joke]"	Correct	67	37	With "Boke [joke]"	Correct	49	35
	Incorrect	22	20		Incorrect	30	17

Discussion

The comprehensibility results indicated the lowest average ratings for the lecture viewed last in the viewing order, regardless of the condition. Simple main effects among groups in each condition were observed under the Teacher and Teacher+Student with Q+Humor conditions, but not under the Teacher+Student with Q or Teacher+Student with 2Qs conditions. Similarly, the results for boredom resistance indicate the influence of the viewing order. However, the results of Group 6 (Teacher+Student with Q+Humor, Teacher, and Teacher+Student with Q, in that order) and Group 9 (Teacher+Student with Q+Humor, Teacher, and Teacher+Student with 2Qs, in that order) do not show an order effect. A simple main effect between groups in each condition was observed under the Teacher condition, but there was no significant difference among groups under the other conditions. This suggests that the inclusion of a question directed at the participants regarding important words may make it difficult for the viewer to be influenced by the order in which they watch the videos, and may allow them to concentrate on the lecture until the end. When the viewers watched the first video that incorporated question and humor, and the second video with neither humor nor question, they were less bored with the last video with question(s). In watching the three videos, it was suggested that the timing of the questions and humor presented could affect overall boredom resistance.

Both comprehensibility and boredom resistance showed significantly higher average ratings in the high familiar group than in the low familiar group. However, the effects of the lecture viewing order on comprehensibility and boredom resistance were observed in both groups.

The results of the confirmation test regarding important words with and without a question showed that the rate of correct responses for important words with a question was significantly higher than those without a question in the high familiar group. The rate of correct responses with question was higher than without question in the low familiar group, but there was no significant difference. The inclusion of questions about important words may affect word recognition, especially when viewers feel that the avatar lectures are delivered naturally.

Furthermore, the results of the comparison between correct response rates for questions about the important words with a question and with a “Boke [joke]” in the same video show that the rate of correct responses for the important words with “Boke [joke]” was significantly higher than with question in the high familiar group. The comparison between the effects of a question and “Boke [joke]” when another question was included in the same lecture video shows that a “Boke [joke]” is more effective than a question. There was no significant difference between both groups, but the effect size in the high familiar group was higher than that in the low familiar group. The inclusion of “Boke [joke]” with a question may affect the high familiar group more than the low familiar group.

Conclusion

This paper proposed a method to generate humor and questions for important words in lecture scenarios, to address the problem of "boredom with on-demand lectures." Two experiments, both presented using two avatars, were conducted to validate the proposed method. It was found that the addition of either a "Boke [joke]" or question made the viewers less bored, and that the inclusion of a "Boke [joke]" and a question for important words may have an effect on word recognition. In particular, the more natural viewers felt the avatar's lecture to be, the more positive was the effect on recognition of important words in the lecture. In addition, the results of the pilot study of Experiment1, where participants were students at an information technology university show the possibility that familiarity with avatar may also enhance the effects of humor and questions on viewers in on-demand lecture videos.

This study has various limitations. First, since the content of the videos used in the experiments was only on cognitive science, it is necessary to examine whether the proposed method is also effective for learning other content, such as mathematics. Furthermore, since the lecture video in this experiment lasted from only 1.5 to 2 minutes, it is unclear whether the method will work for a long, on-demand lecture that can be practical. Therefore, it is necessary to conduct experiments using lecture videos similar in length to actual on-demand lectures. In addition, the results suggest that the timing of questions and humor presented could affect boredom resistance; when questions and sets of "Boke [joke]" and "Tsukkomi [correction]" are added to lecture videos that are similar length to actually used on-demand lectures, their timing and frequency need to be discussed.

Furthermore, the evaluation of boredom resistance is subjective, and an objective evaluation is not yet possible. It is necessary to verify the use of physiological indices, such as the number of blinks. In addition, because including many sets of "Boke [joke]" and "Tsukkomi [correction]" increases the time of the lecture video and is considered impractical, we did not set a double humor condition in which two sets of "Boke [joke]" and "Tsukkomi [correction]" were included in the experiments. Hence, it is necessary to conduct additional experiments with a double humor condition to compare between effects of question and "Boke [joke]" when another "Boke [joke]" is included in the same lecture video.

Abbreviations

OBS: Open Broadcaster Software; CJLC: Corpus of Japanese classroom Lecture speech Contents, ANOVA: Analysis of Variance.

Authors' contributions

This paper was extended from a conference paper Toyota and Terai (2023) presented at the 31st International Conference on Computers in Education. ST conducted the experiments. ST and AT analyzed and interpreted the data and were contributors in writing the manuscript. All authors read and approved the final manuscript.

Authors' information
Satoshi Toyota was a graduate student at Future University Hakodate when this paper was produced.

Asuka Terai is a professor at Future University Hakodate.

Funding
Not applicable.

Availability of data and materials
The data cannot be shared because we informed the participants that their data will be kept confidential.

Declarations

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

Author details
[1] Graduate School of Systems Information Science, Future University Hakodate (Current affiliation: NTT Communications Corporation)

[2] Faculty of Systems Information Science, Future University Hakodate

Received: 31 January 2024 Accepted: 10 October 2024
Published online: 7 February 2025

References

3tene Studio. https://3tene.com/studio/

CoeFont. https://CoeFont.cloud

Fujimura, I., Chiba, S., & Ohso, M. (2012). Lexical and grammatical features of spoken and written Japanese in contrast: Exploring a lexical profiling approach to comparing spoken and written corpora. In H. Mello, M. Pettorino & T. Raso (Eds.), *Proceedings of the VIIth GSCP International Conference: Speech and Corpora* (pp. 393–398). Firenze University Press.

Hiyori, K., Araki, K., Hasegawa, D., & Yoshio, S. (2018). Performance evaluation of automatic gesture generation system using bi-directional LSTM on humanoid robot (in Japanese). In *Proceedings of the 32nd Annual Conference of the Japanese Society for Artificial Intelligence*. The Japanese Society for Artificial Intelligence. https://doi.org/10.11517/pjsai.JSAI2018.0_2C401

Kamiya, Y., Ohno, T., Matsubara, S., & Kashioka, H. (2010) Docho teki taiwa system ni okeru Aizuchi sounyuu timing (in Japanese). In *Proceedings of the 16th Association for Natural Language Processing* (pp. 395–398).

Kudo, T., Yamamoto, K., & Matsumoto, Y. (2004). Applying conditional random fields to Japanese morphological analysis. In D. Lin & D. Wu (Eds.), *Proceedings of the 2004 Conference on Empirical Methods in Natural Language Processing (EMNLP-2004)* (pp. 230–237). Association for Computational Linguistics.

Kyokasho Corpus, Goihyo. https://clrd.ninjal.ac.jp/bccwj/freq-list.html

Lancers. https://www.lancers.jp/mypage

Matsushima, Y. (2020). Practical report regarding to an on-demand class with Google Classroom (in Japanese). *Senshu University Joho Kagaku Kenkyujo Shohou*, *96*, 17–20. https://doi.org/10.34360/00011597

Open Broadcaster Software. https://obsproject.com

Osugi, N. (2021). Status quo of knowledge and understanding of college students regarding special support education (3): Comparison of in-person and on-demand classes (in Japanese). *Journal of the Faculty of Education KOGAKKAN UNIVERSITY*, (4), 21–30.

Suzuki, H., Funaoi, H., Kubuta, Y., & Kato, H. (2021). Enhancing dialogic communication with audience in planning of presentation-talk through "Manzai" script method (in Japanese). In *Proceedings of the 38th Annual Meeting of Japanese Cognitive Science Society* (pp. 687–692). Japanese Cognitive Science Society.

Tatebe, F., & Fujita, Y. *Nihongo Kyoshi Hatsuwa Corpus*. http://www.corpus-ft.com/terms

Toyota, S., & Terai, A. (2023). On-demand lectures with humor and questions using avatars. In J. L. Shih (Eds.), *Proceedings of the 31st International Conference on Computers in Education* (pp.372–381). Asia-Pacific Society for Computers in Education.

Tsuchiya, M., Kogure, S., Nishizaki, H., Ohta, K., & Nakagawa, S. (2008). Developing corpus of Japanese classroom lecture speech contents. In N. Calzolari, K. Choukri, B. Maegaard, J. Mariani, J. Odijk, S. Piperidis & D. Tapias (Eds.), *Proceedings of the Sixth International Conference on Language Resources and Evaluation (LREC'08)* (pp.2061–2065). European Language Resources Association.

Vroid Studio. https://vroid.com/studio

Watanabe, K., Nakagawa, K., Ando, M., & Yukawa, T. (2020). Verifying the effects of changing the appearance of the lecture to a favorite character in lecture (in Japanese). *Japanese Society for information and Systems in Education Hokushinetsu Student Research Presentation*, A12 (pp. 47–48).

Yamashiro, K., Yagishita, H., Aihara, K., Yamamoto, S., & Ikegaya, Y. (2020). Humorous sentences enhance memory and induce beta and gamma power in the human frontal cortex. *Journal of Neuronet*, *1028*, 1–13.

Yashio, S., & Araki, K. (2019). Generation of question sentences using conversation corpus for automatic generation of lecture scenario on lecture substitution system (in Japanese). In *Proceedings of the 33rd Annual Conference of the Japanese Society for Artificial Intelligence*. Japanese Society for Artificial Intelligence. https://doi.org/10.11517/pjsai.JSAI2019.0_4Rin105

Yoshida, Y., & Hagiwara, M. (2012). An automatic manzai-dialogue creating system (in Japanese). *Transactions of Japan Society of Kansei Engineering*, *11*(2), 265–272. https://doi.org/10.5057/jjske.11.265

Publisher's Note