International Journal for Educational Media and Technology
2019, Vol.13, No. 1, pp.66-74

Effect Analysis of High-Speed Presentations of Educational Video Utilizing Synthetic Speech

Toru Nagahama
Waseda University, Japan
tnagahama@aoni.waseda.jp

Masahiro Makino
Waseda University, Japan
ma19941002@asagi.waseda.jp

Yusuke Morita
Waseda University, Japan
ymorita@waseda.jp

This study aims to clarify the effect of presenting educational video utilizing synthetic speech at a high speed. In the experiment, 40 university students were shown educational video dealing with declarative knowledge in 4 conditions: actual speed (1×) synthetic speech, double speed (2×) synthetic speech, actual speed (1×) normal speech, and double speed (2×) normal speech. An analysis of the comprehension test results showed no significant difference in the learning effect according to presentation condition, suggesting that speed and speech factors may have no impact on the learning effect. The results of a subjective questionnaire indicated that whereas the perception of normal speech as strange tends to be affected by the speed factor, the perception of synthetic speech as strange does not tend to be affected by that factor.

Keywords: Educational Video, High-Speed Presentation, Learning Effect, Synthetic Speech

Introduction

Background

Massive open online courses (MOOCs) have gained global prominence in recent years (Waldrop, 2013). Studies of the learning history of those taking MOOCs have shown that they spend longer viewing educational video than engaging in other learning activities related to MOOCs (Breslow et al. 2013; Kizilcec, Piech, & Schneider, 2013). Guo, Kim, and Rubin (2014) used the learning history of MOOC students to show that the rate of studying educational video in which the lecturer's delivery is fast is higher than in the past. Also, it was shown that the rate of studying was high for educational video lasting the relatively short time of 6–9 minutes.

Regarding the length of time spent viewing educational video, Aoyagi, Sato, Takada, Sugawara, and Onai (2005) showed that when studying themes that are relatively easy to understand, the same learning effect was obtained at an accelerated speed as at an ordinary speed. Also, Nagahama and Morita (2017) conducted experiments focusing on the variable speed function of MOOCs, in which educational video dealing with simple knowledge structures was presented at a fast speed. As a result, it was suggested that differences between presentation at actual speed and at double (2x) speed had no impact on the learning effect. This implies that with educational video dealing with learning themes that are comparatively easy to understand, the same learning effect is obtained from a high-speed presentation as from one at the original speed.

By the way, attempts have been made to utilize synthetic speech when producing educational video (Kaburagi, Uehashi, Asae, Kato, & Kang, 2003). Iwasaki and Ohashi (2015) prepared two versions of a narration to some educational video—one with the voice of the relevant lecturer (normal speech) and one with a synthetic voice—and conducted a comparative experiment. They found that evaluations of the synthetic speech were not positive because "it had no inflection and was monotonous" or the student "was distracted because the intonation and pronunciation felt strange."

Up to now, evaluations of speech simulation have been done from the viewpoint of understandability and naturalness (Watanabe, 1989; Kasuya, 1992). First, the understandability of synthetic speech was evaluated from the viewpoint of the degree of intelligibility at the level of phoneme, syllable, word, and sentence. It was reported that in highly understandable synthetic speech, the linguistic information was accurately conveyed (Pisoni, Nusbaum, & Green, 1985; Higuchi, Yamamoto, & Shimizu, 1989; Watanabe, 1989). The naturalness of synthetic speech can be evaluated from

three viewpoints in its synthesis, namely i) segmental characteristics: adherence to the rules on pronouncing vowels and consonants including abnormal pronunciation such as devocalization, lengthening, nasalization, and omission of vowels; ii) prosodic characteristics: duration of mora (a mora is a sound segment unit in phonics with a certain temporal length) and phonemes, accent, pauses, inflection, and loudness; and iii) voice quality: smoothness, "noise," and overall impression. It was reported that the higher the perceived naturalness of synthetic speech, the closer it was to normal speech (Hieda, 1988; Watanabe, 1991).

Kasuya, Morita, and Kumagami (1989) conducted interviews regarding the relationship between the understandability and naturalness of synthetic speech and its effectiveness in achieving a task. The interviews were with newspaper staff carrying out proofreading tasks using synthetic speech. Some participants expressed the opinion that "if the content is clear, when you get used to it, the unnaturalness of synthetic speech is no longer annoying." Besides, Kumagami and Kasuya (1991) asked participants executing a light task to answer questions posed by a normal voice and two synthetic voices of differing naturalness. They found that on the first-time hearing, the task was completed more slowly with the synthetic speech than with the normal speech, but that on the second and subsequent hearings, there was no significant difference in task speed between the synthetic speech and normal speech. In addition, they found no effect of differences in the naturalness of synthetic speech. These results suggest that the impact on task efficiency of inadequate naturalness in synthetic speech is smaller than the impact of understandability. In addition, it was suggested that an increase in the number of hearings allows acclimatization to synthetic speech.

Meanwhile, there are existing studies that show that the most suitable synthetic speech presentation speed differs depending on the application. For example, Kasuya and Morita (1991), in a proofreading task utilizing synthetic speech, asked participants to listen at 5 presentation speeds, from 340 to 680 mora per minute. They demonstrated that, when there was a high rate of inconsistency with the numbers read by the synthetic voice and the printed numbers, the high-speed speech production was poorly evaluated, and the slow-speech production was preferred. On the other hand, Shimahara (2000) sped up the presentation to visually impaired people unable to "speed read" of a synthetic voice at the part-of-speech level using syntactic information. He found that some users acquired the ability to "speed listen." In addition, Watanabe (2005) investigated the use of synthetic speech in screen readers for visually impaired people and found that many users set the presentation speed of their screen reader at the maximum (around 2x normal speed). Those findings imply that comprehension levels would not be affected if synthetic speech is presented at double speed.

Purpose

Kang, Kashiwagi, Treviranus, and Kaburagi (2009) pointed out that utilization of synthetic speech in education might be available approach to decrease the costs associated with the time and effort of creating educational materials. In addition, by using text-to-speech software embedded in a widely available computer, people can easily create educational videos with synthetic speech. Accordingly, the numbers of opportunity for students to learn with educational video using synthetic speech are expected to increase within the next decade. On the other hand, st udents use variable speed function and speed up the presentation rate while they learning online and viewing educational video (Brinton & Chiang, 2015; Shi, Fu, Chen, & Qu, 2015). Besides, the usefulness of high-speed presentation of educational video has been verified in a number of studies (Nagahama & Morita, 2017). However, so far there have been hardly any studies clarifying the effect of changing the presentation speed of educational video that uses synthetic speech. Thus, the goal of this study is to clarify the effect of high-speed presentation of educational video using synthetic speech. In order to achieve the goal, this study focused on the following research questions:

1. What is the difference in learning effects between when learning with educational videos using synthetic speech and when learning with ones using natural speech? (RQ 1)
2. How does video speed influence learning effects when educational videos using synthetic speech are played at double speed? (RQ 2)
3. How do student's opinions about educational videos using synthetic speech differ according to video speed? (RQ 3)

Method

Overview of the Experiment

Educational video was presented in four conditions based on the findings of Nagahama and Morita (2017): with synthetic speech at actual speed (1x), with synthetic speech at double speed (2x), with normal speech at actual speed (1x), and with normal speech at double speed (2x). The participants in the experiment were 40 students (24 male, 6 female; average age 21.4 [SD=0.9]) at a private university in the Tokyo area. Please note that to take account of the

impact of familiarity with synthetic speech, we confirmed with all participants that they had no prior experience of learning to utilize synthetic speech and did not utilize synthetic speech in their daily lives.

Figure 1 shows the experimental procedure. First, participants were sorted into four groups of equal size. Next, a pre-test was carried out to confirm the existing level of knowledge before the learning activity. Next, educational video in each of the four conditions was shown to each group (Table 1). They were asked not to pause or rewind the video. Thereafter, a post-activity test with the same content as the pre-test was carried out to measure the effect on learning. Finally, participants were randomly shown educational video of the various conditions so that they all viewed all four presentation conditions and then answered a subjective evaluation questionnaire.

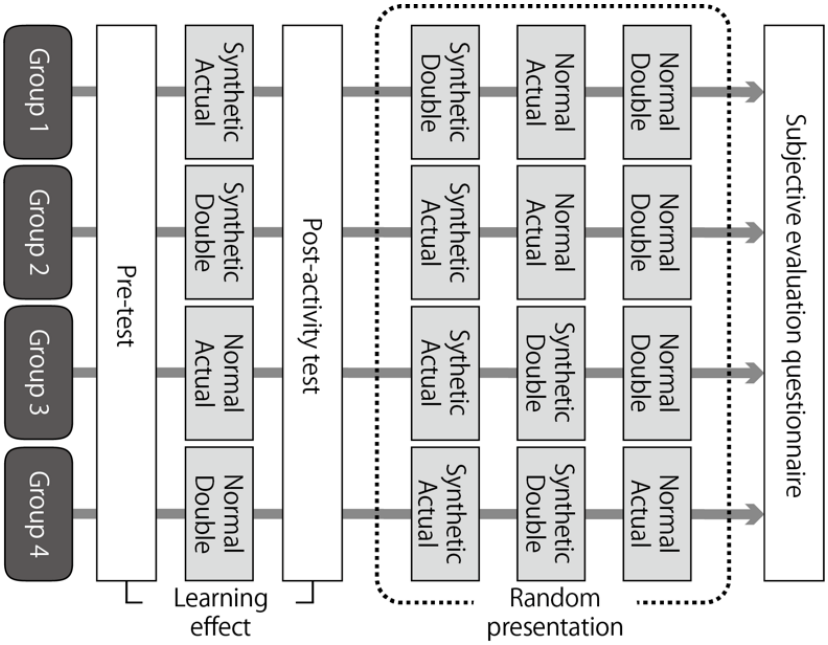

Figure 1. Experimental procedure

Table 1

Initial Educational Video Presentation Conditions by Group

Group		Age	Male	Female	Initial lecture video presentation condition	
					Speed	Speech
Group 1	(*n*=10)	21.6 (0.7)	5	5	Actual	Synthetic
Group 2	(*n*=10)	21.3 (0.7)	7	3	Double	Synthetic
Group 3	(*n*=10)	21.6 (1.3)	6	4	Actual	Normal
Group 4	(*n*=10)	21.1 (0.9)	6	4	Double	Normal

Overview of Experimental Videos

The educational video presented in the normal speech condition (normal speech videos) was the same as that used by Nagahama and Morita (2017). Its subject matter was network structure as taught in information studies at high school, and the lecturer was a currently practicing high school teacher of information studies at a private high school in Chiba Prefecture. In addition, with reference to Fukumori (2008), we measured the speed of the speech in mora and found that it was 327.9 mora per minute.

The educational video presented in the synthetic speech condition was produced based on that used by Nagahama and Morita (2017). For the synthetic speech, we utilized the text-to-speech function of an iMac (Retina 5K, 27 inch) produced by Apple to transfer data to speech. The script to be read was produced from teaching materials in a normal speech video, but without slips of the tongue, auxiliary words, or fillers. In addition, the reading speed and reading

voice were set using the default settings on the iMac. The voice data were edited to the match the voice production timing in the normal speech videos using Final Cut Pro X by Apple. Considering the findings of both Kasuya et al. (1989) and Kumagami and Kasuya (1991) that the impact on task efficiency of inadequate naturalness in synthetic speech is smaller than the impact of understandability, assumption was made that our audio content's intelligibility would be sufficiently guaranteed by creating it through a widely available computer like an iMac.

The presentation time of the normal speech and synthetic speech videos was 9 minutes and 12 seconds in the actual speed conditions and 4 minutes and 42 seconds in the double-speed conditions. In addition, six slides were created, excluding the introductory section. Table 2 summarizes the normal and synthetic speech information. The number of mora was higher for the synthetic speech than for the normal speech because fillers, auxiliary words, and slips of the tongue were excluded from the normal speech video's teaching content. In addition, the length of time that speech was presented (speech presentation time) was measured with a stop watch for both synthetic and normal speech; it was revealed that the synthetic speech presentation time was longer than the normal speech. Please note that the normal speech presentation time was measured with fillers and slips of the tongue in the teaching content excluded.

Table 2

Normal Speech and Synthetic Speech Overview

Slide 【subject matter】	Presentation time (seconds)	No. of mora		Speech presentation time (seconds)	
		Synthetic speech	Normal speech	Synthetic speech	Normal speech
Slide 1 【Networks】	89	455	466	60	60
Slide 2 【Components】	53	255	262	34	34
Slide 3 【Protocol】	140	781	792	104	98
Slide 4 【IP address】	112	616	631	83	77
Slide 5 【DNS server】	41	225	230	30	27
Slide 6 【URL creation】	79	411	428	54	54
Average (*SD*)	85.7 (33.6)	457.2 (194.5)	468.2 (196.7)	60.8 (26.1)	58.3 (24.2)

Comprehension Test

We used Nagahama and Morita's (2017) comprehension test to measure the learning effect, administering a pre-test and a post-activity test. The comprehension test consisted of 20 questions (11 recall questions and 9 applied knowledge questions). One mark was awarded for each correct answer, with 20 marks being the maximum score. The recall questions were presented in the form of an information recall test and were intended to measure the volume of information retained by the participants after viewing the educational video. The applied knowledge questions included one multiple choice question, five recognition questions, and three true or false questions, and were intended to measure the ability to apply knowledge learned from the educational video to new problems.

Subjective Evaluation Questionnaire

A paper question sheet was used in the subjective evaluation questionnaire. There were 25 items consisted of 1 item in Nagahama and Morita's (2017) "comprehension" category, 3 in their "lecturer" category, 3 in their "concentration" category, 1 in their "audibility" category, 2 in their "viewability" category, 2 in their "presentation speed, time taste" category, and 3 in their "content taste" category, to give a total of 15 items, plus 10 items added for this study. The questions were answered on a five-point scale, with five meaning "strongly agree," four meaning "somewhat agree," three meaning "neither agree nor disagree," two meaning "somewhat disagree," and one meaning "strongly disagree."

Results & Discussion

Confirmation of Homogeneity

To confirm homogeneity between the four groups, we carried out a one-way ANOVA on the scores in the pre-test. This showed no significant difference between the groups, F (3, 36) = 0.61, *n.s.* This confirmed that the existing knowledge of the teaching content prior to the learning activity was at the same level in all four groups.

Comprehension Test Analysis

We collected the overall scores on the comprehension test, the scores for the recall questions (recall score), and the scores for the applied knowledge questions (applied knowledge score), which are shown in Figure 2. We conducted a two-way ANOVA (Table 3) regarding the average rise in overall score, recall score, and applied knowledge score on the comprehension test, using the speech factor (relating to type of speech in the educational video) and the speed factor (relating to the speed of the presentation of the educational video).

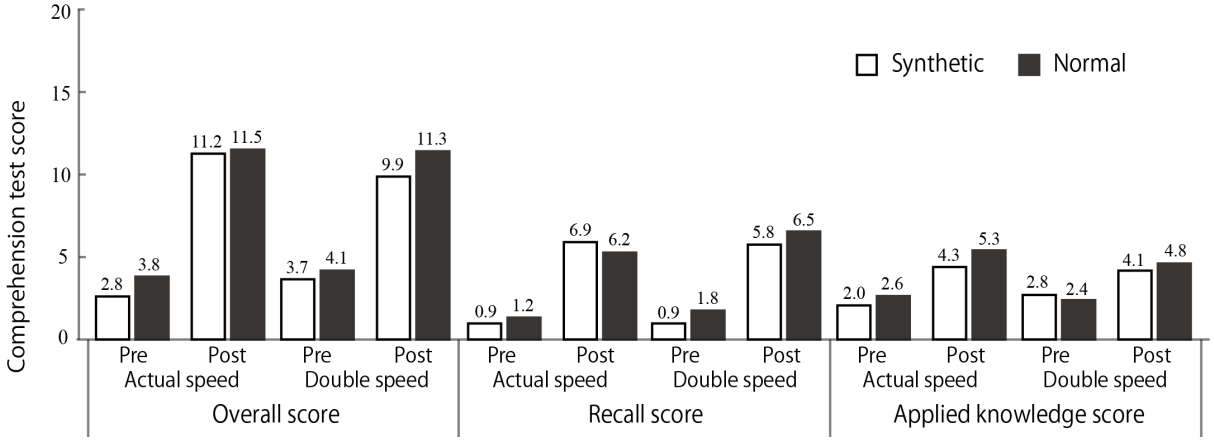

Figure 2. Comprehension test score

First, no interaction was visible regarding the growth in the overall score, F (1, 36) = 0.84, *n.s.* When we tested the main effect, we found no significant difference for the speech factor, F (1, 36) = 0.03, *n.s,* and no significant difference for the speed factor, F (1, 36) = 2.73, *n.s.* This clarifies that the speech and speed factors did not influence the rise in overall score.

Table 3

Average Increase in Score on the Comprehension Test

	Synthetic speech		Normal speech		*F* value		
	Actual speed	Double speed	Actual speed	Double speed	Speech factor	Speed factor	Interaction
Total score	8.3 (2.2)	6.2 (3.0)	7.7 (2.2)	7.1 (2.3)	0.0 *ns*	2.7 *ns*	0.8 *ns*
Recall score	6.0 (2.1)	4.9 (1.8)	5.0 (1.7)	4.7 (1.9)	0.9 *ns*	1.3 *ns*	0.4 *ns*
Applied knowledge score	3.0 (1.3)	1.8 (1.0)	4.4 (0.9)	2.0 (1.2)	15.8 *ns*	137.2 *ns*	13.5 *ns*

**: $p < .01$, *: $p < .05$, +: $p < .10$

Next, no interaction was visible regarding growth in the recall question score, F (1, 36) = 0.41, *n.s.* When we tested the main effect, we found no significant difference for the speed factor, F (1, 36) = 0.92, *n.s*, and no significant difference for the speed factor, F (1, 36) = 1.25, *n.s.* This clarifies that the speech and speed factors did not influence the rise in recall score.

Next, no interaction was visible regarding the growth in applied knowledge score, F (1, 36) = 0.32, *n.s.* When we tested the major effect, we found no significant difference for the speech factor, F (1, 36) = 1.48, *n.s*, and no significant

difference for the speed factor, F (1, 36) = 1.11, *n.s.* This shows that the speech and speed factors did not influence the rise in applied knowledge score.

In the overall, recall, and applied knowledge scores, there was no significant interaction between the speech and speed factors and no significant main effect. This suggests that under the conditions in this experiment, the speech and speed factors had no influence on the learning effect.

Analysis of Subjective Evaluation Questionnaire

Relating to the answers in the subjective evaluation questionnaire, we collected scores for each condition and calculated the average value for each item. To investigate the factorial structure, we conducted an exploratory factor analysis (maximum likelihood method, promax rotation). As a result, we identified four factors from the sharp decline in the scree plot. These factors having been indicated, we carried out a factor analysis (excluding items with a loading of less than 0.35: Item 8 "I focused on the spoken information"; Item 15 "I liked being able to see the lecturer's face"; The slides were clear) and obtained 4 factors and 22 items.

Table 4

Factor Analysis Table

			Factor 1	Factor 2	Factor 3	Factor 4
Usefulness of educational video (α = .86)	Q5	I was able to concentrate and listen	**.85**	.25	.08	-.15
	Q1	I understood the teaching content	**.84**	-.01	.11	-.34
	Q2	The explanation was clear	**.83**	-.17	.17	.07
	Q4	The explanation was well structured	**.66**	-.22	-.03	.11
	Q3	The lecturer's voice was intelligible	**.65**	-.09	-.02	.17
	Q21	The voice was easy on the ears	**.55**	.04	-.17	.13
	Q12	I would like to use this presentation condition to study again	**.52**	.13	.02	.19
	Q11	There were places where I wanted a slower explanation	**- .39**	.25	.03	.19
Perceived strangeness of presentation speech (α = .77)	Q22	The voice inflection was annoying	.03	**.98**	-.05	-.11
	Q23	The voice production intonation was annoying	.02	**.90**	.04	-.12
	Q20	There was a high volume of spoken information	.01	**.41**	.03	.20
Presentation information burden (α = .72)	Q7	The screen flicker was annoying	.20	.11	**.83**	-.07
	Q14	There were a lot of charts and tables on the slides	-.13	-.00	**.62**	.23
	Q6	My eyes became tired while watching	.01	-.01	**.58**	.03
	Q18	The presentation speed was appropriate	-.09	.13	**- .49**	-.44
	Q25	The timing of the voice production was appropriate	.34	-.09	**- .42**	.08
Form and understandability of the educational video (α = .66)	Q16	I liked being able to see the captions	.05	.10	-.01	**.70**
	Q24	The gaps between the voice production were unnatural	-.16	.17	.23	**- .49**
	Q19	The spoken information deepened my understanding	.29	.29	-.30	**.48**
	Q10	I followed the textual information with my eyes with difficulty	-.01	.43	.26	**.47**
	Q9	I focused on the visual information while viewing	.20	.16	-.13	**- .43**
	Q13	The text volume on the slides was low	.11	-.04	.32	**.42**
		Factor 1	—	-.44	-.25	.20
		Factor 2		—	.31	-.02
		Factor 3			—	.02
		Factor 4				—

Table 4 shows the factor analysis results. Factor 1 was comprised of items relating to the usefulness of the educational video (such as Item 5 "I was able to concentrate and listen," Item 1 "I understood the teaching content," Item 3 "The lecturer's voice was intelligible," and Item 21 "The voice was easy on the ears"), and so it was dubbed the "usefulness of educational video" factor. Factor 2 was comprised of items relating to perceptions of strangeness in the presentation speech of the educational video (such as Item 22 "The voice inflection was annoying" and Item 23 "The voice production intonation was annoying"), and so it was dubbed the "perceived strangeness of presentation speech" factor.

Factor 3 was comprised of items relating to the burden of the presentation's educational video (such as Item 7 "The screen flicker was annoying," Item 14 "There were a lot of charts and tables on the slides," and Item 6 "My eyes became tired while watching"), and so it was dubbed the "presentation information burden" factor. Factor 4 was comprised of items relating to the form of the educational video and its understandability (such as Item 16 "I liked being able to see the captions," Item 24 "The gaps between the voice production were unnatural," and Item 10 "I followed the textual information with my eyes with difficulty"), and so it was dubbed the "form and understandability of the educational video" factor.

We investigated the reliability of the criteria using *a* coefficients; that for "usefulness of educational video" was .86, that for the "perceived strangeness of presentation speech" was .77, that for "presentation information burden" was .72, and that for "form and understandability of the educational video" was .66. Please note that when investigating the reliability of the criteria using *a* coefficients, items where the loading showed a minus value were inverted for processing.

Analysis of sub-scale scores for each factor. We calculated the average value of the items in each factor for each presentation condition and made them the respective sub-scale's score. We also carried out a two-way ANOVA for the speech and speed factors (Table 5).

First, interaction was seen as relating to the sub-scale score for the "usefulness of educational video" factor, $F(1, 39) = 8.5$, $p < .01$. An investigation of the simple main effect revealed a significant difference in the actual speed condition, $F(1, 39) = 34.0$, $p < .01$, and the double-speed condition, $F(1, 39) = 5.5$, $p < .05$, regarding the speech factor. Meanwhile, regarding the speed factor, a significant difference was seen for the normal speech condition, $F(1, 39) = 103.7$, $p < .01$, and the synthetic speech condition, $F(1, 39) = 79.8$, $p < .01$. These results showed that the degree of usefulness of the educational video as felt by the participants differed according to the speech and speed factors.

Next, an interaction was seen relating to the sub-scale score of the "perceived strangeness of presentation speech" factor, $F(1, 39) = 14.5$, $p < .01$. An investigation of the simple main effect revealed a significant difference in the actual speed condition, $F(1, 39) = 54.3$, $p < .01$, and the double-speed condition, $F(1, 39) = 32.4$, $p < .05$, regarding the speech factor. Meanwhile, regarding the speed factor, a significant difference was seen for the normal speech condition, $F(1, 39) = 16.5$, $p<.01$, but no significant difference was seen for the synthetic speech condition, $F(1, 39) = 0.3$, *n.s.* These results showed that while the degree of "perceived strangeness of the presentation speech" as felt by the participants differed according to the speed factor in the normal speech condition, in the synthetic speech condition, it was the same regardless of the speed factor.

Table 5

Subscale Score for Each Factor

	Synthetic speech		Normal speech		*F* value		
	Actual speed	Double speed	Actual speed	Double speed	Speech factor	Speed factor	Interaction
Usefulness of educational video	3.1 (0.7)	1.8 (0.8)	4.2 (0.7)	2.2 (1.0)	32.3 **	170.2 **	8.5 **
Perceived strangeness of presentation speech	3.6 (0.9)	3.7 (1.1)	2.0 (0.8)	2.7 (0.9)	51.9 **	8.4 **	14.5 **
Presentation information burden	2.4 (0.6)	2.8 (0.7)	1.9 (0.6)	2.7 (0.7)	34.1 **	34.0 **	6.1 *
Form and understandability of the educational video	2.6 (0.7)	2.5 (0.6)	2.8 (0.6)	2.4 (0.6)	3.6 +	30.7 **	11.8 **

**: $p < .01$, *: $p < .05$, +: $p < .10$

Following on from this, an interaction was seen relating to the sub-scale score of the "presentation information burden" factor, $F(1, 39) = 6.1$, $p < .05$. An investigation of the simple main effect revealed a significant difference in the actual

speed condition, F (1, 39) = 30.9, $p < .01$, and the double-speed condition, F (1, 39) = 6.6, $p < .05$, regarding the speech factor. Meanwhile, regarding the speed factor, a significant difference was seen for the normal speech condition, F (1, 39) = 34.2, $p < .01$, and the synthetic speech condition, F (1, 39) = 20.2, $p < .01$. These results showed that the degree of "presentation information burden" felt by the participants differed according to the speech and speed factors.

Following on from that, an interaction was seen relating to the sub-scale score of the "form and understandability of the educational video" factor, $F(1, 39) = 11.8$, $p < .01$. An investigation of the simple main effect revealed a significant difference in the actual speed condition, F (1, 39) = 14.1, $p < .01$, but no significant difference in the double-speed condition, F (1, 39) = 0.4, *n.s*, regarding the speech factor. Meanwhile, regarding the speed factor, a significant difference was seen for the normal speech condition, F (1, 39) = 38.5, $p < .01$, and the synthetic speech condition, $F(1, 39) = 3.1$, $p < .10$. These results showed that while subjective evaluations of the "form and understandability of the educational video" differed according to the speech factor in the actual speed condition, they were the same regardless of the speech condition in the double-speed condition.

Conclusion

The aim of this study was to clarify the effect of a high-speed presentation of educational video using synthetic speech. In the experiment, 40 university students were presented with educational video dealing with declarative knowledge in 4 conditions (actual speed synthetic speech, double-speed synthetic speech, actual speed normal speech, and double-speed normal speech).

An analysis of the comprehension test results suggested that neither the factor relating to the speech (speech factor) nor the factor relating to the presentation speed (speed factor) had any impact on the learning effect. This result supports and extends Nagahama and Morita (2017)'s findings demonstrating that differences between presentation at actual speed and at double speed had no impact on the learning effect.

On the other hand, an analysis of the subjective evaluation questionnaire suggested that the subjective evaluations relating to the "usefulness of the educational video" and "the burden from the presentation information" differed according to the speech and speed factors. In addition, it was suggested that the subjective evaluations relating to the "perceived strangeness of the presentation speech" were not impacted by the speed factor in the synthetic speech condition. Furthermore, it was found that the subjective evaluations relating to the "form and understandability of the educational video" were not affected by the speech factor in the double-speed condition.

From the above, the main contributions of this study are as follows.

1. This study demonstrates that students can learn as much with educational videos using synthetic speech as with ones using normal speech (Answer to RQ 1).
2. This study demonstrates that students can learn with educational videos using synthetic speech 2 times as much within a set of period of times as they have been able to do in the past (Answer to RQ 2).
3. This study implies that unnaturalness relating to the inflection, intonation, and rhythm of synthetic speech can be alleviated by speeding up the presentation speed, and the acceptability to the listener and the intelligibility can be improved (Answer to RQ 3).

This study advances both the educational video creating literature and the synthetic speech literature by clarifying the learning effect of a high-speed presentation of synthetic speech. However, there was no thorough investigation of the naturalness of the inflection, pauses, and intonation of the synthetic speech used in this experiment. This was because the text-to-speech software used was one that is installed on a widely available computer. In addition, Kumagami and Kasuya (1991) indicated that users rapidly get used to synthetic speech. In this study, we only asked the participants to listen to synthetic speech in each condition once, and we did not get to the stage of thoroughly investigating the impact of increased familiarity with synthetic speech.

There were thus limitations to the evaluation in this study. Therefore, a more comprehensive discussion is required regarding the effect of high-speed presentations of educational video utilizing synthetic speech; this discussion should encompass the various elements relating to synthetic speech.

References

Aoyagi, S., Sato, K., Takada, T., Sugawara, T., & Onai, R. (2005). Evaluation of video skimming method to educational purpose movies. *Journal of Information Processing, 46*(5), 1927-1305.

Breslow, L., Pritchard, D. E., Deboer, J., Stump, G. S. S., Ho, A. D., & Seaton, D. T. (2013). Studying learning in the worldwide classroom: Research into edX's first MOOC. *Research and Practice in Assessment*, *8*, 13-25.

Brinton, C., G. & Chiang, M. (2015) Mining MOOC Clickstreams: Video-watching behavior vs. in-video quiz performance. *IEEE Transactions on signal processing*, *64*(14), 3677-3692

Guo, P. J., Kim, J., & Rubin, R. (2014). How video production affects student engagement: An empirical study of MOOC videos. *In Proceedings of the First ACM conference on Learning*, 41-50.

Hieda, I. (1988). Subjective indices for evaluation of synthesized voice. *Japan Ergonomics Society Research Journal*, *24*, 387-394.

Higuchi, N. , Yamamoto, S., & Shimizu, T. (1989) Evaluation of intelligibility and naturalness of the synthetic speech generated with a Japanese speech synthesizer by rule. *Journal of the Institute of Electronics, Information and Communication Engineers*, *D-Ⅱ* (J72-D-H), 1133-1140.

Iwazaki, K. & Ohashi, A. (2015). Active learning experiences in the flipped classroom. *Computer & Education*, *39*, 98-103.

Kaburagi, M., Uehashi, J., Asase, J., Kato, M., & Kang, M. (2003). Development of supporting system with speech engine for material creation and learning. *Japan Journal of Educational Technology*, *27*(Suppl.), 141-144.

Kang, M., Kashiwagi, H., Treviranus, J., & Kaburagi, M. (2008) Synthetic speech in foreign language learning: an evaluation by learners. *International Journal of Speech Technology*, *11*(2), 97-106

Kasuya, H. (1992). Assessment of speech synthesis technology. *The Journal of the Acoustical Society of Japan*, *48(1)*, 46-51

Kasuya, H., & Morita, K. (1991). Role of the speaking rate of synthetic speech produced by rule as an aid for proofreading. *The Journal of the Acoustical Society of Japan*, *47*, 96-98

Kasuya, H., Morita, K., & Kumagami, K. (1989). Investigation relating to evaluation of the quality of synthetic speech. *Report of a study funded by the Kakenhi Grant-in-Aid for Scientific Research in program (important areas, speech and language)*, PASL01-8-2.

Kizilcec, R. F., Piech, C., & Schneider, E. (2013). Deconstructing disengagement: Analyzing learner subpopulations in massive open online courses. *In Proceedings of the Third International Conference on Learning Analytics and Knowledge, ACM*, 170-179.

Kumagami, K., Kasuya, H. (1991). Objective evaluation of user's adaptation process to synthetic speech produced by rule. *The Journal of the Acoustical Society of Japan*, *47*, 243-249.

Nagahama, T., & Morita, Y., (2017). An analysis of the effects of learning with high-speed visual contents. *Japan Journal of Educational Technology, 40*(4), 291-300.

Nagahama, T., & Morita, Y. (2017). Effect analysis of playback speed for lecture video including instructor images. *International Journal for Educational Media and Technology*, *11*(1), 50-58.

Pisoni, D. B., Nusbaum, H. C., & Green, B. G. (1985). "Perception of synthetic speech generated by rule." *In Proceedings of the IEEE 73*: 1665-1676.

Shi, C., Fu, S., Chen, Q., & Qu, H. (2015) VisMOOC: Visualizing video clickstream data from Massive Open Online Courses. *In 2015 IEEE Pacific Visualization Symposium*, 159-166

Shimahara, S. (2000). 'Speed listening' – a fast reading system for visually impaired people using syntactic information. *Institute of Electronics, Information and Communication Engineers Technical Report, Fifth Well-being Information Technology Conference WIT00-28*.

Waldrop, M. M. (2013). Online learning: Campus 2.0. *Nature*, *495*, 160-163. http://dx.doi.org/10.1038/495169a (accessed 3.10.2018).

Watanabe, T. (2005). A study on voice settings of screen readers for visually-impaired PC users. *The IEICE Transactions on Information Systems Pt. 1, 88*(8), 1257-1260, 2005-08-01.

Watanabe, T. (1989). Investigation relating to methods of evaluating rule-based synthetic speech using degree of work comprehension. *Institute of Electronics, Information and Communication Engineers Research Journal*, *J72-A*, 1503-1509.

Watanabe, T. (1991). Investigation into evaluation of naturalness of rule-based synthetic speech. *Institute of Electronics, Information and Communication Engineers Research Journal*, *J74-A*, 599-609.