Yang, J. C. et al. (Eds.) (2018). Proceedings of the 26th International Conference on Computers in Education. Philippines: Asia-Pacific Society for Computers in Education

A Trial Study about the Effect of Hi-Speeded Educational Video Utilizing Synthetic Speech

Toru NAGAHAMA[a*], Masahiro MAKINO[b] & Yusuke MORITA[a]
[a]*Faculty of Human Sciences, Waseda University, Japan*
[b]*School of Human Sciences, Waseda University, Japan*
*tnagahama@aoni.waseda.jp

Abstract: 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 video dealing with declarative knowledge in 4 conditions: actual speed (1x) synthetic speech, double speed (2x) synthetic speech, actual speed (1x) normal speech, and double speed (2x) 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. An analysis of the interview results indicated that while learners found synthetic speech unnatural in terms of inflection and intonation when it was presented at actual speed, at high speed this unnatural impression was alleviated, and the speech became more acceptable to listeners.

Keywords: Educational video, Learning effect, Synthetic speech, High-speed presentation

1. Introduction

Attempts have been made to utilize synthetic speech when producing educational video (Kaburagi et al. 2003). There are existing studies that show that the most suitable synthetic speech presentation speed differs depending on the application. For example, 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).

By the way, Nagahama and Morita (2017) conducted experiments focusing on playback speeds of educational video, showing that differences between presentation at actual speed and at double (2x) speed had no impact on the learning effect. 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 aim of this study is to clarify the effect of high-speed presentation of educational video using synthetic speech.

2. Method

2.1 Experimental Procedure

Educational video was presented in four conditions: with synthetic speech at actual speed, with synthetic speech at double speed, with normal speech at actual speed, and with normal speech at double speed. The participants in the experiment were 40 students (24 male, 16 female; average age 21.4 [*SD*=0.9]).

In the experiment, 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, visual content in each of the four conditions was shown to each group. Next, a post-test with the same content as the pre-test was carried out to measure the effect on learning. Thereafter, participants were randomly shown visual content of the various conditions so that they all viewed all four presentation conditions. Finally, an interview survey was carried out.

2.2 Experimental Educational Video

The video presented in the normal speech condition (normal speech videos) was the same as that used by Nagahama and Morita (2017). Its theme 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, based on the Kasuya' (1992) research, we measured the speed of the speech in mora (a mora is a sound segment unit in phonics with a certain temporal length) and found that it was 327.9 morae per minute.

The video presented in the synthetic speech condition (synthetic speech videos) 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 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.

3. Results & Discussion

3.1 Comprehension Test Analysis

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

Table 1
Average increase in score on the 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*

As a result, there was no interaction regarding the growth in the score, F (1, 36) = 0.84, *n.s.*). When we investigated 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.

3.2 Interview Survey Analysis

When all of the comments obtained via the interviews were summarized and counted, there were 196 in total. All of the comments obtained were classified into the four categories of "comprehension," "intelligibility," "acceptability to the listener," and "concentration," according to their key words. In doing this, with reference to Kasuya (1992), we classified comments relating to "comprehension" in the "intelligibility" category and those relating to "naturalness" in the "acceptability to the listener" category. As a result, 17 comments were classified under "comprehension," 51 under "intelligibility," 80 under "acceptability to the listener," and 32 under "concentration." Please note that 16 comments did not fit into any of the 4 categories and were classified under "other." Table 2 summarizes the interview results relating to the double speed condition, mainly focusing on the acceptability to the listener category.

Among the comments relating to the ease of listening to the presentation speech, there were positive comments regarding normal speech at actual speed relating to the presence of inflection, a sense of friendliness, and the presence of familiarity. There were also negative comments relating to being bothered by slips of the tongue and fillers.

Meanwhile, there were negative comments regarding normal speech at double speed, to the effect that "there were too many fillers" and "sped-up fillers are jarring to hear." These comments point to the possibility, relating to the ease of listening, that the participants were positive about normal speech at actual speed, but fillers increased its unpleasantness when it was sped up. Elsewhere, there were positive comments regarding synthetic speech at actual speed relating to the absence of slips of the tongue and fillers. There were also negative comments relating to the absence of inflection, the unnatural intonation, and the monotonous rhythm. However, although there were negative comments relating to the double-speed synthetic speech to the effect that the monotonous or high tone was jarring, there were also positive comments to the effect that "the aspects of synthetic speech that are strange at actual speed are no longer annoying" and "the unnaturalness of the voice is less annoying than at actual speed and it becomes easier on the ear."

These comments suggest that when it comes to the ease of listening to the presentation speech, the participants found the synthetic speech unnatural in terms of inflection, intonation, and rhythm when the presentation was at actual speed, but that the sense of strangeness and unnaturalness was alleviated when the presentation was sped up and it became more acceptable to the listener.

Table 2
Main Comments in the Acceptability to the Listener Category

		Synthetic speech	Normal speech
Acceptability to the listener	Positive	The aspects of synthetic speech that are strange at actual speed are no longer annoying. / The unnaturalness of the voice was not as annoying as at actual speed, and it was easy on the ears. / The flat intonation and rhythm were much less annoying at double speed. / After I got used to it, it became easy on the ears.	After I got used to it, double speed was easier on the ears. / Because the fillers were no longer annoying, it was easier on the ears than actual speed.
	Negative	The monotonous tone was unpleasant at speed. / It really grated. / The shrillness was unpleasant.	There were too many fillers, which was annoying. / Speeded-up fillers are jarring to hear. / The sound got higher, so that it was shrill.

4. 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 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 nor the factor relating to the presentation speed had any impact on the learning effect. The interview survey results implied, with regard to the presentation's ease of listening, that the learners found the synthetic voice unnatural in terms of inflection, intonation, and rhythm when hearing it at actual speed, but that when it was presented to them at high speed, this perceived unnaturalness was alleviated and it became more acceptable to the listener. In addition, it seems that the absence of fillers and slips of the tongue as found in normal speech increased the ease of listening of the synthetic speech.

References

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*, 141-144.

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

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.

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.

This research was supported by JSPS KAKENHI Grant-in-Aid for Young Scientists (B) Number 50779270.