EUROCALL 2023: CALL for all Languages
15-18 August 2023, University of Iceland, Reykjavik
DOI: https://doi.org/10.4995/EuroCALL2023.2023.16916

Effects of text-to-speech synthesized speech on learners' presentation anxiety and self-efficacy: A comparison of two models

Takatoyo Umemoto[a], Shinnosuke Takamichi[b], Yuta Matsunaga[c], Yusuke Yoshikawa[d], Kikuko Yui[e], Kishio Sakamoto[f], Shigeo Fujiwara[g], and Yasushige Ishikawa[h]

[a]Institute for Liberal Arts and Science, Kyoto University of Foreign Studies, , t_umemoto@kufs.ac.jp; [b]The University of Tokyo, , shinnosuke_takamichi@ipc.i.u-tokyo.ac.jp; [c]The University of Tokyo, , matsunaga-yuta339@g.ecc.u-tokyo.ac.jp; [d]Kyoto University of Foreign Studies, , y_yoshikawa@kufs.ac.jp; [e]Kyoto University of Foreign Studies, , k_yui@kufs.ac.jp; [f]Kyoto University of Foreign Studies, , k_sakamoto@kufs.ac.jp; [g]Uchida Yoko Co., Ltd., , shigeo.f@uchida.co.jp and [h]Kyoto University of Foreign Studies, , y_ishikawa@kufs.ac.jp

How to cite: Umemoto, T.; Takamichi, S.; Matsunaga, Y.; Yoshikawa, Y.; Yui, K.; Sakamoto, K.; Fujiwara, S.; Ishikawa, Y. (2023). Effects of text-to-speech synthesized speech on learners' presentation anxiety and self-efficacy: A comparison of two models. In *CALL for all Languages - EUROCALL 2023 Short Papers.* 15-18 August 2023, University of Iceland, Reykjavik. https://doi.org/10.4995/EuroCALL2023.2023.16916

Abstract

This paper reports on the effects of two Text-To-Speech (TTS) synthesized speech models, one based on English utterances by a native English speaker and the other based on English utterances by a Japanese non-native speaker of English, on presentation anxiety and self-efficacy of Japanese English as a Foreign Language learners. We hypothesized that learners' presentation anxiety would decrease and their self-efficacy would increase when using the Japanese non-native English speaker model compared with the native English speaker model. 55 first-year university students (upper level: 33; lower level: 22) voluntarily participated in the study and were divided into experimental and control groups. To measure the participants' presentation anxiety and self-efficacy, the scale developed by Ishikawa et al. (2021) was used. A mixed-design three-factor ANOVA with the group, class level, and period as independent variables showed an interaction between class level and period. A simple main effect test indicated a significant increase in self-efficacy in the upper-level students. These results reveal that, regardless of the model used, the use of TTS-synthesized speech significantly increases the self-efficacy of the upper-level students. The paper concluded that further research on technology use and learner affect needs to be conducted.

Keywords: *presentation anxiety, self-efficacy, text-to-speech, synthesized speech.*

1. Introduction

Studies have demonstrated the usefulness of Text-To-Speech (TTS) in language-learning settings. Handley and Hamel (2005) reported positive results for TTS use and recommended the technology for English as a Foreign Language (EFL) learners' listening and speaking practice. TTS synthesis, which generates speech from text input, offered means of providing spoken language input to learners in Computer-Assisted Language Learning (CALL) environments (Handley, 2009). A trainee teacher in Turkey utilized a web-based TTS tool outside class in order to improve her English pronunciation, and it was found that her accent started being perceived as native, indicating that the online TTS tool may be effective as a self-study tool for improving the trainees' English pronunciation (Ekşi, & Yeşilçınar, 2016). Liakin et al. (2017) examined the impact of the pedagogical use of mobile TTS on the L2 acquisition of French liaisons. The results showed that the mobile TTS technology complemented and enhanced L2 pronunciation teaching. In addition, Chiang (2019) compared traditional teacher-led dictation and dictation with TTS among EFL learners' vocabulary performance and found a significant difference. Even in an environment where English is the first language, TTS technology has been used for English language lessons. Parr (2013) conducted an eight-month survey with 28 grade five students whose first language was English, and revealed that the TTS technology promoted an inclusive reading practice that facilitated language learning for students with different reading abilities.

Regarding learners' perceptions of pedagogical TTS use, Bione et al. (2016) found that participants had positive attitudes toward it. According to Papin and Cardoso (2022), learners viewed the self-directed learning experience with TTS and the automatic speech recognition features of Google Translate as positive. Moon (2020) conducted a questionnaire survey on using self-generated listening materials based on TTS and found that most students felt that TTS-based listening materials reduced their listening anxiety and increased their confidence.

Although previous studies have investigated the usefulness of TTS in language-learning settings, no study has compared different models of TTS-synthesized speech and their effects on the emotional aspects of learner engagement. Therefore, in this study, we developed two TTS-synthesized speech models and an interface that allows learners to use them. We hypothesized that learners' presentation anxiety would decrease and their self-efficacy would increase when using the Japanese non-native English speaker model compared to when using the native English speaker model.

2. Method

2.1. Development

Two TTS-synthesized speech models were created: one based on English utterances by a native English speaker and the other based on English utterances by a Japanese non-native speaker of English. The speech of both models was synthesized by a deep learning model 'FastSpeech 2: Fast and High-Quality End-to-End Text-to-Speech' (https://arxiv.org/abs/2006.04558v8) using data in which text and speech were paired.

The interface to allow learners to use the two above-mentioned models was developed as a web-based application that runs on students' mobile devices, such as laptops, tablets, and smartphones. After logging into the application, students can listen to the TTS-synthesized speech by the following three steps: (1) Click on the 'Create New' button, and when the dialog box appears, select the document you want to convert to a TTS-synthesized speech, and then, click the 'Start' button. (2) Click the 'Convert to Audio' button at the bottom left of the screen to convert it to a TTS-synthesized speech. (3) When the conversion is completed, a playback bar appears in the lower-left corner of the screen, which can be pressed to listen to the TTS-synthesized speech (see Ishikawa et al., 2021 for more details).

2.2. Participants

Two class levels of first-year students (upper:33; lower:22; total:55) from a Japanese university volunteered to participate in this study and were randomly divided into two groups (A:26; B:29). They were divided into two classes at their university according to their scores on the Test of English for International Communication (TOEIC) for Listening and Reading. The participants' mean TOEIC scores were 537.12 (SD = 88.43) and 456.67 (SD = 92.92) in the upper and lower group, respectively.

2.3. Instrument

Ishikawa et al.'s (2021) questionnaire was used to measure the participants' presentation anxiety and self-efficacy. The questionnaire included ten items (five items each for presentation anxiety and self-efficacy), such as “I feel anxious when I give a presentation in English” and “I am confident in my ability to give presentations in English.” Participants rated their responses on a five-point Likert scale (1 = disagree to 5 = strongly agree).

2.4. Procedure

A pre-model survey using the above-mentioned questionnaire was administered in late November 2022. The participants then practiced their presentations for approximately three weeks outside the classroom. The participants in Group A used the non-native Japanese speaker of the English model, whereas those in Group B used the native English speaker model. In mid-January 2023, a post-model survey was conducted using the same questionnaire after the participants delivered their presentations in class. A mixed-design two-factor Analysis of Variance (ANOVA) was conducted with the groups and periods as independent variables. In addition, a mixed-design three-factor ANOVA was conducted with groups, periods, and class levels as independent variables.

3. Results

Cronbach’s alpha coefficients for presentation anxiety and self-efficacy were calculated. The survey results were sufficiently reliable (presentation anxiety: pre: α = .88, post: α = .84; self-efficacy: pre: α = .79, post: α = .83).

The two-factor ANOVA results showed no differences between the groups and periods. Furthermore, no interaction effect was observed, implying that our hypothesis was invalid. However, the three-factor ANOVA showed an interaction effect between class level and period (F (1, 51) = 5.21, p < .05, η_p^2 = .07; see Table 1). Thus, a simple main-effect test (the Holm method) was conducted, and a significant increase in self-efficacy was found among upper-level students. This finding indicates that the self-efficacy of upper-level students increased significantly when using TTS-synthesized speech, regardless of the model used.

Table 1. Mean and standard deviation for each variable by group, class level, and period.

	Group	Class level	Pre		Post	
			Mean	*SD*	*Mean*	*SD*
Presentation anxiety	Group A	Upper	3.55	0.84	3.54	0.57
		Lower	3.42	0.74	3.38	1.33
	Group B	Upper	3.55	1.10	3.41	0.85
		Lower	3.46	0.67	3.26	0.94
Self-efficacy	Group A	Upper	2.93	0.52	2.96	0.61
		Lower	3.04	0.47	2.89	0.96
	Group B	Upper	2.55	0.58	2.89	0.54
		Lower	2.95	0.58	2.85	0.69

4. Discussion

Our study hypothesized that learners’ presentation anxiety would decrease and their self-efficacy would increase when using the Japanese non-native English speaker model compared to the native speaker model. However, this hypothesis was not supported. Nevertheless, the three-factor ANOVA and a simple main test revealed that upper-level students’ self-efficacy increased significantly when using TTS-synthesized speech, regardless of the model used. This outcome implies that presentation practice using TTS-synthesized speech was closely associated with participants' English proficiency. By practicing their presentations using TTS-synthesized speech, the participants in the upper-level group felt that their presentations in class would proceed smoothly. However, the participants in the lower-level group did not have sufficient English language skills to attain such confidence. This presentation practice was an out-of-classroom self-study (participants decided whether they wanted to practice their presentations). Therefore, participants in the lower-level group might have been less motivated to engage in presentation practice than those in the upper-level group.

The study's limitations are as follows: 1) the intervention used a one-shot, three-week design and 2) the presentation practice using TTS-synthesized speech was participant-directed.

Based on the results and limitations described above, there are two directions for future research. The intervention period using TTS-synthesized speech should be expanded from three weeks to one semester and incorporated into a course designed to help students improve their speaking skills and give presentations in English, offering them opportunities to practice their presentations with their instructors. Moreover, TTS-synthesized speech factors that affect learners' presentation anxiety, self-efficacy, and motivation should be assessed qualitatively, similar to Teng and Wang's (2021) research on Chinese EFL learners.

Acknowledgements

While conducting the research described in this paper, we were deeply saddened by the sudden loss of our co-researcher, Dr. Masao Aikawa, on December 18, 2022. He was a Professor at the Department of British and American Studies and Vice President at Kyoto University of Foreign Studies, Japan. He contributed significantly to TESOL and foreign language education policy research in Asia. We offer sincere condolences to his family.

This research was supported by the Intra-University Research of Kyoto University of Foreign Studies, Japan, and Grants-in-Aid for Scientific Research provided by the Japan Society for the Promotion of Science (#18K00763). The data presented, statements made, and views expressed are solely the responsibility of the authors.

References

Bione, T., Grimshaw, J., & Cardoso, W. (2016). An evaluation of text-to-speech synthesizers in the foreign language classroom: Learners' perceptions. In S. Papadima-Sophocleous, L. Bradley, & S. Thouësny (Eds.), *CALL communities and culture—Short papers from EUROCALL 2016* (pp. 50–54). Research-publishing.net. https://doi.org/10.14705/ rpnet.2016.eurocall2016.537

Chiang, H.-H. (2019). A comparison between teacher-led and online text-to-speech dictation for students' vocabulary performance. *English Language Teaching*, *12*(3), 77–93. https://doi.org/10.5539/elt.v12n3p77

Ekşi, G. Y., & Yeşilçınar, S. (2016). An Investigation of the effectiveness of online text-to-speech tools in improving EFL teacher trainees' pronunciation. *English Language Teaching*, *9*(2), 205–214. https://www.ccsenet.org/journal/index.php/elt/article/view/56606

Handley, Z. (2009). Is Text-to-speech synthesis ready for use in computer-assisted language learning? *Speech Communication*, *51*(10), 906–919. https://doi.org/10.1016/j.specom.2008.12.004

Handley, Z., & Hamel, M. J. (2005). Establishing a methodology for benchmarking speech synthesis for computer-assisted language learning (CALL). *Language Learning & Technology*, *9*(3), 99–120. http://dx.doi.org/10125/44034

Ishikawa, Y., Takamichi, S., Umemoto, T., Aikawa, M., Sakamoto, K., Yui, K., Fujiwara, S., Suto, A., & Nishiyama, K. (2021). Japanese EFL learners' speaking practice utilizing text-to-speech technology within a team-based flipped learning framework. In P. Zaphiris & A. Ioannou (Eds.), *Learning and collaboration technologies: New challenges and learning experiences* (pp. 283–291). Springer. https://doi.org/10.1007/978-3-030-77889-7_19

Liakin, D., Cardoso, W., & Liakina, N. (2017). The pedagogical use of mobile speech synthesis (TTS): Focus on French liaison. *Computer Assisted Language Learning*, *30*(3-4), 325–342. https://doi.org/10.1080/09588221.2017.1312463

Moon, D. (2020). Learner-generated digital listening materials using text-to-speech for self-directed listening practice. *International Journal of Internet, Broadcasting and Communication*, *12*(4) 148–155. http://dx.doi.org/10.7236/IJIBC.2020.12.4.148

Papin, K., & Cardoso, W. (2022). Pronunciation practice in Google Translate: Focus on French liaison. In B. Arnbjörnsdóttir, B. Bédi, L. Bradley, K. Friðriksdóttir, H. Garðarsdóttir, S. Thouësny, & M. J. Whelpton (Eds.), *Intelligent CALL, granular systems, and learner data: Short papers from EUROCALL 2022* (pp. 322–327). Research-publishing.net. https://doi.org/10.14705/rpnet.2022.61.1478

Parr, M. (2013). Text-to-speech technology as inclusive reading practice: Changing perspectives, overcoming barriers. LEARNing Landscapes Journal, *6*(2), 303–322. https://doi.org/10.36510/learnland.v6i2.618

Teng, Y., & Wang, X. (2021). The effect of two educational technology tools on student engagement in Chinese EFL courses. *International Journal of Educational Technology in Higher Education*, *18*, 27. https://doi.org/10.1186/s41239-021-00263-0