Does Text-to-speech Synthesis Fit the EFL Learners' Needs in EIL Context?

MARTRIWATI[a], Herri MULYONO[a]*, & Devi Nur VEBRIYANTI
[a]*University of Muhammadiyah PROF. DR. HAMKA, Indonesia*
*hmulyono@uhamka.ac.id

Abstract: The present study attempts to evaluate a TTS synthesis system from the perspective of EIL. 100 pre-service teachers are selected to evaluate speech produced by a TTS synthesis system with the respect of intelligibility, naturalness, prosody, and social impression. A questionnaire adopting the MOS X is developed to evaluate speech in three experiments.

Keywords: Text-to-speech (TTS), English as international language (EIL), English as foreign language (EFL)

1. Introduction

Many literature (e.g. Handley, 2009; Sha, 2010; Sobkowiak, 2003) have discussed potential benefits from utilizing text-to-speech (TTS) synthesis in English as foreign language (EFL) classroom. The TTS synthesis, in computer assisted language learning (CALL) contexts may be used as a reading machine, a pronunciation model and for conversational partner (Handley, 2009; Handley & Hamel, 2005). The use of English as international language (EIL) has emerged number of people who speak variety of English, known as *World Englishes* (see Jenkins, 2006; Jenkins, 2009). In this tenet, English may be contextualized with the respect of a local culture and context as well. Such contextualization, of course, will drive the scope of authentic materials beyond what it has previously been defined by some well-known authors (e.g. Brown, 2004; Harmer, 2007).

So far, the evaluation of TTS synthesis as CALL application has been evaluated in perspective of native speakers. In such evaluation, a good English model is still justified in reference with a model from its native speakers. This justification seems to overlook the wide use of the world Englishes in daily communication. Jenkins (2006, p. 175) argues that there is a need "to abandon the native speaker as the yardstick and to establish empirically some other means of defining an expert (and less expert) speaker of English." Thus, the present study attempts to evaluate a TTS synthesis system from the perspective of EIL. The research questions being addressed is: "Does text-to-speech synthesis fit the EFL learners' needs in EIL context?"

2. Methodology

2.1 Research design

The present study attempts to evaluate a TTS synthesis system from the perspective of EIL. For such a purpose, the study adopts an experimental design with two methods including survey and test. As shown in Table 1, the TTS synthesis system is evaluated with respect to four aspects, including intelligibility, naturalness, prosody and social impression (see section 2.4). A listening comprehension test will be performed to evaluate the pre-service teachers' listening comprehension before and after an intervention.

2.2 Participants

As the present study attempts to evaluate the use of TTS synthesis in CALL, Handley (2009) suggests that the participants should be the end-user of TTS synthesis in CALL context. For such a purpose, the present study selects 100 pre-service teachers to participate. At present, the 100 pre-service teachers are taking a course on instructional technology at the university of Muhammadiyah Prof. DR. HAMKA Jakarta Indonesia. One of the course modules taught to the pre-service teachers is using

technology to facilitate English teaching. In this module, the participants are introduced to TTS to support their further English teaching.

2.3 TTS synthesizer

In current study employs NaturalReader (NR) for the TTS synthesizer. We have sought that NR suggests benefits in ELT particularly the teaching of listening comprehension (Mulyono, 2014). In addition, NR has similar features as other TTS synthesizer as in Sha's (2010) experiment. Table 1 shows the features of NR used:

Table 1: The features of TTS synthesis used.

TTS synthesizer	Sex	Voice	English use	XML* Tag
NaturalReader	Female	Natural Voice Kate	American	Kate
NaturalReader	Male	Natural Voice Paul	American	Paul

2.4 Instruments

Three sets of questionnaire are employed to evaluate the pre-service teacher perceptions towards: 1) NS recorded materials, and 2) EIL recorded materials, spoken by NNS, 3) listening materials developed by using TTS synthesis. The three sets of questionnaire adopts 15 item number questionnaire with 7-point scale MOS (mean opinion scale) X suggested by Polkosky and Lewis (2003) for each factor including intelligibility, naturalness, prosody, and social impression. Polkosky and Lewis propose that such a questionnaire was "a theoretically-derived, [and] psychometrically sound factor structure" (p. 174). They also suggest that each factor deployed in the questionnaire was reliable for applicability evaluation, $\alpha > .70$.

In addition to questionnaire, a listening comprehension test will be developed to examine if the pre-service teachers' perceptions of TTS technology may affect their listening comprehensions. The test will be distributed prior to and after an intervention. A delayed posttest will also be distributed to the research participants to evaluate the effect in a longer term.

2.5 Experiment procedure

In the first experiment, the participants are given a listening comprehension task. The audio materials used for the listening materials are taken from webpages that use American English. The American English is selected as it is used during the instructional technology course. After the first experiment, the first questionnaire is distributed to gather information about the pre-service teachers' perception toward the NS recorded materials.

In the second experiment, the participants are given the second listening comprehension task that uses EIL recorded materials. This second listening comprehension task will be similar to the first one. The second questionnaire is distributed after the listening comprehension task to evaluate the participants' perception towards the EIL materials.

In the second experiment, the participants are asked to have the third listening comprehension task with similar content to the two previous ones. In the third listening task, the listening materials are developed by using TTS synthesis. The materials are designed in reference to context of the participants' daily life. The third set of questionnaire is distributed after the listening comprehension task to evaluate the pre-service teachers' perception towards the materials produced by the TTS synthesis.

2.6 Progress

Recently, we are at the preparation stage. In this stage, we are attempting to get consent for the use of materials from the webpage publisher. At the same time, we are now developing for the questionnaires used for the research instruments and the listening scripts.

Liu, C.-C. et al. (Eds.) (2014). Proceedings of the 22nd International Conference on Computers in Education. Japan: Asia-Pacific Society for Computers in Education

References
Brown, H. D. (2004). *Language assessment: Principles and classroom practices*. New York: Longman.
Handley, Z. (2009). Is text-to-speech synthesis ready for use in computer-assisted language learning? *Speech Communication, 51*(10), 906-919. doi: http://dx.doi.org/10.1016/j.specom.2008.12.004
Handley, Z., H, & Hamel, M.-J. (2005). Establishing a methodology for benchmarking speech synthesis for computer-assisted language learning (CALL). *Language Learning & Technology, 9*(3), 99-120.
Harmer, J. (2007). *The practice of English language teaching* (4th ed.). Essex: Pearson Education Limited.
Jenkins, J. (2006). Current perspectives on teaching world Englishes and English as a Lingua Franca. *TESOL Quarterly, 40*(1), 157-181. doi: 10.2307/40264515
Jenkins, J. (2009). *World Englishes: A resource book for students*. London: Routledge.
Mulyono, H. (2014). Creating native-like but comprehensible listening texts for EFL learners using NaturalReader. *TESL-EJ, 18*(1).
Polkosky, M., & Lewis, J. (2003). Expanding the MOS: Development and psychometric evaluation of the MOS-R and MOS-X. *International Journal of Speech Technology, 6*(2), 161-182. doi: 10.1023/A:1022390615396
Sha, G. (2010). Using TTS voices to develop audio materials for listening comprehension: A digital approach. *British Journal of Educational Technology, 41*(4), 632-641. doi: 10.1111/j.1467-8535.2009.01025.x
Sobkowiak, W. (2003). TTS in EFL CALL - Some pedagogical considerations. *Teaching English with Technology, 3*(4), 3-11.