An evaluation of text-to-speech synthesizers in the foreign language classroom: learners' perceptions

Tiago Bione[1], Jennica Grimshaw[2], and Walcir Cardoso[3]

Abstract. As stated in Cardoso, Smith, and Garcia Fuentes (2015), second language researchers and practitioners have explored the pedagogical capabilities of Text-To-Speech synthesizers (TTS) for their potential to enhance the acquisition of writing (e.g. Kirstein, 2006), vocabulary and reading (e.g. Proctor, Dalton, & Grisham, 2007), and pronunciation (e.g. Cardoso, Collins, & White, 2012). Despite their demonstrated effectiveness, there is a need for up-to-date formal evaluations of TTS systems, specifically for their potential to promote the ideal conditions under which languages are acquired, particularly in an English as a Foreign Language (EFL) environment, as suggested by Cardoso, Smith, and Garcia Fuentes (2015). This study evaluated a modern English TTS system in an EFL context in Brazil, at a number of levels, including speech quality, opportunity to focus on form, and learners' cognitive processing of TTS-generated texts. Fifteen Brazilian EFL learners participated in the study in which they listened to both human and TTS-produced speech samples while performing the abovementioned tasks. Semi-structured interviews were used to collect data about participants' perceptions of the technology. We report an analysis of these interviews, which indicate that EFL learners have overall positive attitudes towards the pedagogical use of TTS, and that they would like to use the technology as a learning tool.

Keywords: text-to-speech synthesis, TTS, L2 pronunciation, English as a foreign language, Brazil.

[1] Concordia University, Montréal, Canada; tiagobione@gmail.com
[2] Concordia University, Montréal, Canada; jennica.grimshaw@gmail.com
[3] Concordia University, Montréal, Canada; walcir.cardoso@concordia.ca

How to cite this article: 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

1. Introduction

Second/foreign language (L2) learners need to be exposed to a significant amount of input in the L2 in order to acquire it. The literature suggests that the learning process is boosted when learners have access to comprehensible (Krashen, 1985) and acoustically variable input (Barcroft & Sommers, 2005) in an environment that is learner-centered and that provides multiple opportunities for self-regulation (Chapelle, 2001). However, EFL students often lack input exposure outside the classroom as they do not usually have access to proficient speakers. One way to overcome this limitation is via the pedagogical use of TTS in L2 education, which has been demonstrated to be beneficial for the development of writing (Kirstein, 2006), vocabulary and reading (Proctor et al., 2007), and pronunciation skills (Cardoso et al., 2012).

Despite their demonstrated effectiveness, there is a need for up-to-date formal evaluations of TTS systems, specifically for their potential to promote the ideal conditions under which languages are learned (e.g. via enhanced input environments – Chapelle, 2001). Recently, Cardoso et al. (2015) investigated whether TTS systems provided university-level English as a second language speakers in Canada with the opportunity to complete focus-on-form tasks, detecting English regular past tense marking (i.e. /t/, /d/ or /Id/) in forms such as *walked*, *played,* and *wanted* in both human and TTS-generated texts. Findings revealed that participants were able to identify the target forms correctly, regardless of the type of voice heard. They also examined how students perceived the quality of TTS speech in comparison with human speech, and found less favorable ratings for TTS-generated texts. They acknowledged, however, that their findings could not be generalized to ‘foreign’ language contexts, as their participants were native-like English speakers with vast experience with the language. The current study, therefore, addresses the following research question: What are learners’ perceptions of the pedagogical use of TTS in an EFL context?

2. Method

2.1. Participants and design

Fifteen Brazilian EFL learners (university students or professionals from a variety of educational backgrounds and proficiency levels, between the ages of 16 and 32) were recruited to complete four tasks in order to evaluate the quality of TTS-generated texts: they rated speech quality, answered comprehension questions, participated in a dictation, and were asked to identify past -ed forms in what they heard. For all tasks, participants listened to speech samples alternately produced by TTS and a human. The TTS voice was Julie (by NeoSpeech), a female North American speaker, and the human was a female native-speaker of the same dialect with similar speech properties. At the end of the one hour session (one-shot design), participants were interviewed about their insights on the quality of the TTS-generated voices. This paper reports findings from the analysis of the interview data.

2.2. Analysis

Participants' interviews were analyzed and categorized into four themes: TTS system vs. human voice, speech accuracy, comprehensibility, and potential to be used as a learning tool. Under each theme, students' opinions were labelled as positive or negative to depict their overall perception and acceptance of TTS for pedagogical purposes.

3. Results

3.1. TTS system versus human voice

Only four participants were able to identify the electronic/human voice opposition without prompting: "The first voice was like the Google voice, when you try to translate"; "Clearly, there was a computer voice and other that was not". Eight students only realized differences between the two types of voices when they were explicitly told that one was machine-made: "Now that you've mentioned it, yes. Actually, I hadn't stopped to think about it". The remaining three students were not able to differentiate TTS and human voices even when they were told about them: "It completely fooled me", declared a participant.

3.2. Accuracy

At the segmental level, TTS was perceived to be as good as native voices: "For me, both were speaking as native speakers, both were super correct". However, suprasegmentals were problematic as TTS was seen as unnatural and lacking native rhythm in phrasal stress, intonation and pauses: "[the computer voice] helped my comprehension, but sometimes words have a different intonation

[…] words that should give away phrasal tone [stress]". One student noticed, however, that TTS is efficient for questions since it accurately reproduces the rising tone in interrogative clauses.

3.3. Comprehensibility

Inaccuracy in suprasegmental production made most participants consider TTS harder to understand when compared to the human voice. Nine students stated that the pace of the TTS and consequently its intonation patterns and connected speech phenomena made it more difficult to understand: "The 'tempo' between words [in TTS] draws too much attention, because in spoken language you feel more fluidity due to different timing between word connections, grammatical constructions […]. With computers, everything seems to have the same timing, so it creates a different perception". Five learners favored TTS because it was slower and made word boundaries more salient as it speaks "word by word".

3.4. Potential as a learning tool

11 participants agreed that TTS technology could and should be used as a tool for language learning: they perceived that morphological features such as past -ed markings were more salient in TTS, and that it could be a source for extra input outside the classroom. Three participants suggested that TTS should only be used with beginners as more advanced learners need to experience human voice characteristics: "People will [need to] get used [to a human voice] because when they travel, they won't hear paused, word by word speech". Seven participants stated that TTS should be used for all levels as long as the human voice is not excluded from the learning process. For two students, however, TTS should not be used as a learning tool: "I don't think it is a good idea […] it's always better when a native teacher speaks".

4. Conclusions

This study investigated learners' perceptions of the pedagogical use of TTS in an EFL context. The findings show that participants had an overall positive impression of TTS-generated voices. TTS has evolved significantly in terms of quality in recent years to the point that most participants could not tell the difference between human voice and TTS until prompted. However, the TTS voice was still rated less favorably in terms of comprehensibility when compared to the human voice. These findings corroborate previous studies (e.g. Cardoso et al., 2015; Handley, 2009). On the other hand, while the low ratings for comprehensibility may appear negative, this had little impact on participants' perception of TTS as a pedagogical tool. Almost all participants recognized that TTS could and should be used for teaching purposes, and most said that it should be used regardless of students' proficiency levels. This contrasts to Cardoso et al.'s (2015) findings from a similar study conducted in a 'second' language context, wherein participants showed lower acceptance towards TTS. One reason for this high acceptance of TTS as a pedagogical tool in the current study may be due to the fact that EFL environments lack naturally occurring L2 input and access to native or proficient speakers in the target L2 outside of online environments. With TTS, language learners can gain additional exposure and have access to enhanced input environments to increase their learning opportunities (Chapelle, 2001). These findings suggest that EFL students appear to be ready to adopt TTS systems as pedagogical tools in L2 education.

References

Barcroft, J., & Sommers, M. S. (2005). Effects of acoustic variability on second language vocabulary learning. *Studies in Second Language Acquisition, 27*(3), 387-414. https://doi.org/10.1017/s0272263105050175

Cardoso, W., Collins, L., & White, J. (2012). Phonological input enhancement via text-to-speech synthesizers. *Paper presented at the AAAL Conference, Boston, U.S.A*.

Cardoso, W., Smith, G., & Garcia Fuentes, C. (2015). Evaluating text-to-speech synthesis. In F. Helm, L. Bradley, M. Guarda, & S. Thouësny (Eds), *Critical CALL – Proceedings of the 2015 EUROCALL Conference, Padova, Italy* (pp. 108-113). Dublin Ireland: Research-publishing.net. https://doi.org/10.14705/rpnet.2015.000318

Chapelle, C. (2001). *Computer applications in second language acquisition*. Cambridge, UK: Cambridge University Press. https://doi.org/10.1017/CBO9781139524681

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

Kirstein, M. (2006). *Universalizing universal design: applying text-to-speech technology to English language learners' process writing*. Doctoral dissertation. University of Massachusetts, U.S.A.

Krashen, S. (1985). *The input hypothesis: issues and implications*. New York: Longman.

Proctor, C. P., Dalton, B., & Grisham, D. (2007). Scaffolding English language learners and struggling readers in a universal literacy environment with embedded strategy instruction and vocabulary support. *Journal of Literacy Research, 39*(1), 71-9.

Published by Research-publishing.net, not-for-profit association
Dublin, Ireland; Voillans, France, info@research-publishing.net

CALL communities and culture – short papers from EUROCALL 2016
Edited by Salomi Papadima-Sophocleous, Linda Bradley, and Sylvie Thouësny

Disclaimer: Research-publishing.net does not take any responsibility for the content of the pages written by the authors of this book. The authors have recognised that the work described was not published before, or that it is not under consideration for publication elsewhere. While the information in this book are believed to be true and accurate on the date of its going to press, neither the editorial team, nor the publisher can accept any legal responsibility for any errors or omissions that may be made. The publisher makes no warranty, expressed or implied, with respect to the material contained herein. While Research-publishing.net is committed to publishing works of integrity, the words are the authors' alone.

Trademark notice: product or corporate names may be trademarks or registered trademarks, and are used only for identification and explanation without intent to infringe.

Copyrighted material: every effort has been made by the editorial team to trace copyright holders and to obtain their permission for the use of copyrighted material in this book. In the event of errors or omissions, please notify the publisher of any corrections that will need to be incorporated in future editions of this book.

Typeset by Research-publishing.net
Cover design by © Easy Conferences, info@easyconferences.eu, www.easyconferences.eu
Cover layout by © Raphaël Savina (raphael@savina.net)
Photo "bridge" on cover by © Andriy Markov/Shutterstock
Photo "frog" on cover by © Fany Savina (fany.savina@gmail.com)
Fonts used are licensed under a SIL Open Font License

ISBN13: 978-1-908416-43-8 (Paperback - Print on demand, black and white)
Print on demand technology is a high-quality, innovative and ecological printing method; with which the book is never 'out of stock' or 'out of print'.

ISBN13: 978-1-908416-44-5 (Ebook, PDF, colour)
ISBN13: 978-1-908416-45-2 (Ebook, EPUB, colour)

Legal deposit, Ireland: The National Library of Ireland, The Library of Trinity College, The Library of the University of Limerick, The Library of Dublin City University, The Library of NUI Cork, The Library of NUI Maynooth, The Library of University College Dublin, The Library of NUI Galway.

Legal deposit, United Kingdom: The British Library.
British Library Cataloguing-in-Publication Data.
A cataloguing record for this book is available from the British Library.

Legal deposit, France: Bibliothèque Nationale de France - Dépôt légal: décembre 2016.