Volume 6, Nomor 2, November 2016
ISSN: 2088-2025

SUPPORTING LANGUAGE LEARNING WITH TTS VOICE OUTPUT INTEGRATED INTO EPUB E-BOOKS

Achmad Suyono
Politeknik Negeri Malang
achmad.suyono@polinema.ac.id

Bambang Suryanto
Politeknik Negeri Malang
bambang.suryanto@polinema.ac.id

ABSTRACT

Text-to-speech (TTS) technology gains more popularity nowadays as computer and smart devices are becoming affordable and widely used. The adoption of this technology in language learning is also increasing. With the rise of mobile learning, an attempt to integrate this technology in EPub e-books is proposed. It is expected that the e-books incorporating TTS technology voice output can help teachers to efficiently produce and distribute learning materials, to reduce workload in dealing with individual students' learning problems, to implement necessary learning guidance, to provide distraction-free mobile language learning materials, and to be more productive with technology to promote a better language learning in the students' daily activities.

***Keywords**: TTS, EPub, foreign language*

INTRODUCTION

The popularity of Text-to-Speech (TTS) technology is widespread all over the world. Developed initially to assist people with visual disability, TTS technology are now present for various applications and environments. Car drivers may be aware of this technology through the GPS voice navigation feature in their cars. PC and laptop users may have used it as it is preinstalled in Windows, OS X, and Ubuntu (Linux) operating systems. Smartphone users may have considered it as invaluable as it enables them to listen to their audio books and check the pronunciations of words or phrases in their devices on a daily basis.

Though TTS technology has been developed and used for decades, it is the current availability and application in computer and mobile devices that have unwrapped more interests in its adoptions in the educational settings. Besides the possibilities to expand its usage for assisting learners with reading disabilities, the improvement of the voice quality has attracted its usage to support foreign language learning.

TEXT-TO-SPEECH TECHNOLOGY

The TTS technology may vary in its system complexity but basically it processes the text-based inputs to generate synthesized speech output. An illustration to this concept is shown in Speech-Over TTS technology, in which three main stages are involved (Tuval Software Industries, 2015). When a user feeds text-based information into the system, the speech engine will process the text input automatically. As the core of the system, the engine will first parse the input, then analyze

aspects like grammar, punctuation, and capitalization, and finally activate the selected voice to generate the audio format of the text. The selected voice depends on the configuration of the operating system (i.e. embedded or not embedded) or on the user personal choice (by adding free or commercial voices). The voices incorporated in the engine can be customized to meet the language, regional accent, and gender preference.

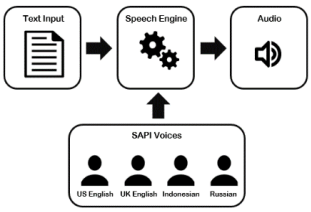


Figure 1: Text-to-Speech (TTS) Technology

TTS technology has developed from formant synthesis method into concatenation synthesis method. The main difference of the two methods lies on the use of real human speech database. Formant synthesis method does not use the speech database and relies entirely on simulating the acoustic properties of speech. On the other hand, concatenation synthesis method uses the speech database, where human voice is divided into sound segments and then concatenated or combined to produce words (Microsoft, 2003). In this way, the sound "fourteen" can be produced from "four" and "teen" segments. Whenever the word "teen" is needed, the same segment will be called, whether it is in "sixteen" or "two hundred nineteen". For the common user, the output of the two methods can be easily distinguished as the first one sounds robotic while the latter is perceived as more natural (Tuval Software Industries, 2015).

Nowadays there are many vendors providing TTS voices, whether free or commercial. Google, for instance, provides more than 15 language voices in its Android OS, with some languages having local variations (i.e. English, Spanish, and Chinese) and some others do not. Other companies such as Acapela Group, NeoSpeech, and Ivona Software provide commercial voices. To see how the quality among commercial voices compare, see the following chart (Figure 2) presenting the result of Text-to-Speech Accuracy Testing report published by ASRNews (2015). The accuracy test result reflects the overall capabilities of each product in dealing with foreign words, numbers, homographs, acronyms, abbreviations, names, and addresses. This report shows that Ivona Software achieves the highest score (97.5%), followed by Neospeech (95.5%) and Diotek (95.3%).

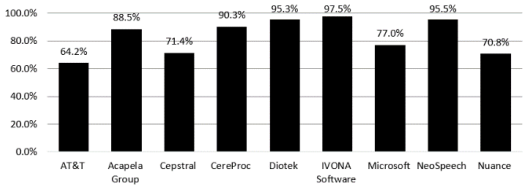


Figure 2: TTS Accuracy Test Result

Besides being installed in users' devices, TTS technology can be provided as an online service. The benefit of this service is not only that it does not require the need to install the system; it can also be accessed anytime and anywhere through various devices (e.g. smart phones, tablets, PCs). Some websites providing this service include *text2speech.org, ispeech.org* and *ttsreader.com*.

Text-to-Speech Technology in Language Learning

Much of the interest in the use of TTS technology in language learning is due to the improvement of TTS voice output quality. It is true that the quality has not achieved a stage to generate perfect natural human voices for every type of utterances, however, this technology has stepped far beyond monotonous machine voices. Since 1990s, a major progress in the quality of the voices has been acknowledged. Liberman (1995) wrote that the technology had achieved a high level of performance, with increasingly sophisticated models of linguistic structure, low error rates in text analysis, and high intelligibility in synthesis from phonemic input. In 2000s, research findings showed that the technology achieved higher naturalness for short utterances, and that it was difficult to differentiate the voice it produced from real human's voice (Schroeter, et al., 2002). In the past few years much

research was also in progress to find better system for higher quality voices. As a result, despite the enhancement on its accuracy, naturalness and expressivity which should continuously be improved, TTS technology has reached a level of readiness for language learning deployment (Kilickaya, 2006; Gelan, 2011).

Many researchers and educators do not only appreciate the improvement of the voice quality but also the effectiveness of TTS technology in supporting language learning. One notable advantage of using the technology is that it can help foreign language learners improve their pronunciation (González, 2007; Kılıçkaya, 2011) and spelling (Huang & Liao, 2015) abilities. Gonzales (2007) and Huang and Liao (2015) even confirmed that the technology fostered the students' independence in their learning. A different benefit of using the technology is in vocabulary learning. Rosa, Parent, and Eskenazi (2010) reported that using speech synthesis to produce the spoken versions of words benefitted non-native speakers during vocabulary learning lessons and their exposure to the spoken words led to increases in auditory vocabulary performance. A more interesting benefit of the technology is the provision of native-like voices to the students. In a situation where using voice talent is not possible or too costly, TTS technology can be the most feasible solution. Perhaps it is also the condition faced by Mulyono (2014) who implemented TTS voice output in listening learning materials and provided students with access to British and American accents.

TTS technology is also beneficial beyond pronunciation and vocabulary learning. Chong, Tosukhowong and Sakauchi (2002) suggested that the technology could be applied in web-based lectures and it helped audience with language understanding problems. Likewise, Parr (2013) agreed that the technology could circumvent frustration and reader withdrawal due to inadequate decoding and fluency and provide an increase in their motivation, confidence, and self-efficacy. Besides helping those with reading difficulties, the technology can motivate students to read more. When students listened to digital texts, whether from audio narration or read aloud by a TTS tool, they could increase their reading volume (Dalton & Grisham, 2011). In a classroom learning environment, the technology can be designed to support writing skill development. As Young and Stover (2013) reported, by using online service from *Voki.com* where the students wrote texts and the customizable Voki characters read aloud the texts, students were able to evaluate and revise their own writing.

EBOOKS: WHY EPUB FORMAT

E-books, as the name suggests, are books in electronic formats. Different from traditional books, e-books are presented in files, not in paper-based materials. E-books can be in different file formats, like Plain Text (.txt), Microsoft Word (.doc, .docx), Apple iBooks (.ibooks), Amazon Kindle (.azw3, .azw, .kf8), Mobipocket (.mobi), Portable Document Format (.pdf), EPub (.epub), and many others (Wikipedia, 2016a). Each format has its own strengths and limitations. Plain Text files, for example, can be read in any devices but contain only texts and do not have supports for features like tables and images. On the other hand, Microsoft Word, Apple iBooks, Amazon Kindle, Mobipocket, Portable Document Format, and EPub can support tables, images, and even more advanced features like audio and video, but may not be accessible in all platforms. In fact, despite its long-time reputation, Microsoft Word files are not natively supported by Amazon Kindle devices (Wikipedia, 2015).

EPub can be a good choice of e-book format due to several reasons, more significantly its popularity, content layout, standard, and global language support. EPub is said to be the most widely supported format as it can be opened across various platforms (Basu, 2015; DeLoatch, 2016). Since its initial release in 2007 by International Digital Publishing Forum, EPub has been a chosen standard of book publishing industries. Even Apple iBooks was built on the EPub standard (Bott, 2012). The layout of e-books in EPub can be designed to have fixed or re-flowable content layouts. PDF is an example of a format which only supports a fixed content layout,

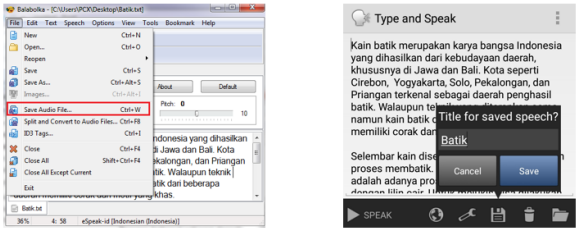
without the capability to adjust its content to suit various device screen sizes. On the other hand, EPub is capable of resizing its content so that users, regardless of the different device screen sizes they use, can have the best viewing experience. In terms of its standard, EPub is free and open (Wikipedia, 2016b). EPub is continuously supported, developed and maintained by interested parties, including publishers and developers. As for the global language support, EPub is capable of providing support for more diverse range of languages, writing modes, and styles (Makoto, 2014). Accordingly, authors and publishers from around the globe do not have issues regarding what languages and writing systems they will use (e.g. left-to-right or right-to-left).

TTS VOICE OUTPUT IN EPUB E-BOOKS: HOW AND WHY

As discussed above, TTS technology offers benefits in language learning, including in pronunciation, vocabulary, listening, reading and writing. Integrating this technology into an e-book format with advanced capabilities like EPub promises students more effective and efficient learning. Paper-based textbooks are essential parts of campus and student academic activities, however, neglecting the fact that students possess the tools which can be utilized to support their learning is definitely not a wise decision. Moreover, providing various learning materials in different formats is one of the strategies to address different learning styles students may have.

The biggest challenge now lies on the foreign language teachers' readiness to embrace the technology. How steep is the learning curve to generate TTS voice output? How difficult is it to integrate the output into EPub? The teachers should not worry about the technical issues as currently there are computer applications or online services which are user friendly and, sometimes, free.

The methods to generate TTS voice output can be as simple as activating the voice, writing the texts, and saving the audio file. By using Balabolka (*http://www.cross-plus-a.com/balabolka.htm*) in Windows OS, for instance, a user's texts can be saved into an audio file by selecting the Save Audio File in the File menu, just like saving a file in a word processor application. In Android OS, a user can use Type and Speak (https://play.google.com/store/apps/details?id=com.googamaphone.typeandspeak&hl=en) where the voice output can be generated by simply tapping the Save icon, an easy to recognize user interface similar to that in many other Android applications. An illustration of how each method is used, see Figure 3 below.



(a) (b)

Figure 3: Generating TTS Voice output in (a) Balabolka and (b) Type and Speak

What about integrating the voice output in EPub e-books? Nowadays, there are computer applications, whether commercial like Ultimate Ebook Creator (http://ultimateebookcreator.com) or free like Sigil (*https://sigil-ebook.com*), capable of performing the task with ease. If the teachers know how to operate computer and use a word processor, they will not find it a problem to perform the task as it is like inserting a picture in a document file. In Sigil, for example, integrating the voice output is done by simply selecting the Insert menu, choose the File submenu, and select the file to insert (see Figure 4). After the EPub e-book is completed, the e-book can be saved and distributed to the students.

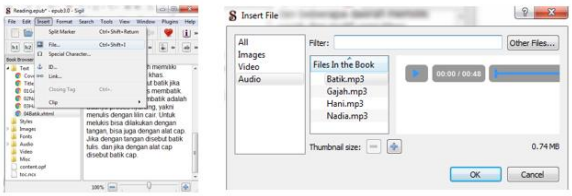


Figure 4: Adding audio file in Sigil

The teachers' capabilities in harnessing the potentials of both technologies open more opportunities which benefit both the teachers and their students. First, the technologies allow faster learning material

production and distribution. In a situation where voice talent narrations are not imperative, using TTS technology can be time saving. Also, as the learning material is in the form of a file, it can be distributed to the students almost instantly without significant device compatibility issue. Second, the e-book can reduce the teachers' workload in helping individual students to improve their, for instance, pronunciation. In fact, the teachers do not need to be present in all of the students' learning sessions since the students can work independently and use the feedback provided in the e-book when needed. Third, necessary learning guidance can also be implemented. Many devices have built-in TTS technologies which can read aloud the content of e-books and thus integrating the TTS voice output which will perform comparable tasks seems to be impractical. However, there are cases where the teachers need to guide the students so that they focus only on certain parts of the texts or where the teachers have to ensure the students listen to the pronunciation of words using a certain local dialect which the students' devices may not support. Fourth, the e-book can be designed for the sole purpose of helping the students learn. The Internet is a world of abundance where the students can find various language learning e-books for free. Unfortunately some of the e-books are free because it contains advertisements. The teachers' e-books are there to ensure that the students can learn without such distractions. Finally, various devices have been parts of teachers and students' lives. It is an opportunity for the teachers to use the technology more productively to promote a better language learning in the students' daily activities.

CONCLUSION

This article has presented a brief overview of TTS technology and EPub e-book and how the integration of TTS voice output EPub e-books can support language learning. Considering the benefits of the integration and the user friendly applications to integrate and publish the EPub e-books, teachers are encouraged to embrace the technologies and use them to help their students learn foreign languages more effectively and efficiently.

REFERENCES

ASRNews. (2015). *Text-to-Speech Accuracy Testing - 2015*. Retrieved September 7, 2016, from ASRNews: http://www.asrnews.com/TTS_acc_for website.pdf

Basu, S. (2015). *GT explains: What is the difference between EPUB, MOBI, AZW and PDF ebook formats?* Retrieved September 2, 2016, from Guiding Tech: http://www.guidingtech.com/9661/difference-between-epub-mobi-azw-pdf-ebook-formats/

Bott, E. (2012). *Some standards are more open than others*. Retrieved September 6, 2016, from ZDNet: http://www.zdnet.com/article/some-standards-are-more-open-than-others/

Chong, N. S., Tosukhowong, P., & Sakauchi, M. (2002). WhiteboardVCR: a web lecture production tool for combining human narration and text-to-speech synthesis. *Educational Technology & Society, 5*(4). Retrieved from http://www.ifets.info/journals/5_4/ng.pdf

Dalton, B., & Grisham, D. L. (2011). eVoc strategies: 10 ways to use technology to build vocabulary. *The Reading Teacher, 64*(5), 306-317. doi:10.1598/RT.64.5.1

DeLoatch, P. (2016). *Creating your book using the most popular ebook formats*. Retrieved September 6, 2016, from Edudemic: http://www.edudemic.com/most-popular-ebook-formats/

Gelan, A. (2011). Language and text-tospeech technologies for highly accessible language & culture learning. *iJET, 6*(2), 11-14. doi:10.3991/ijet.v6i2.1529

González, D. (2007). Text-to-speech applications used in EFL contexts to enhance pronunciation. *TESL-EJ, 11*(2). Retrieved September 7, 2016, from http://www.tesl-ej.org/ej42/int.pdf

Huang, Y.-C., & Liao, L.-C. (2015). A study of text-to-speech (TTS) in children's English

learning. *Teaching English with Technology, 15*(1), 14-30.

Kilickaya, F. (2006). Text-to-speech technology: What does it offer to foreign language learners? *CALL-EJ Online, 7*(2).

Kılıçkaya, F. (2011). Improving pronunciation via accent reduction and text-to-speech software. In M. Levy, F. Blin, C. B. Siskin, & O. Takeuchi, *WorldCALL: International Perspectives on Computer-Assisted Language Learning* (pp. 85-96). New York, NY: Routledge.

Liberman, M. (1995). Computer Speech Synthesis: Its Status and Prospects. *Proceedings of the National Academy of Sciences of the United States of America. 92 (22)*, pp. 9928-9931. National Academy of Sciences. Retrieved from http://www.jstor.org/stable/2368592

Makoto, M. (2014). *EPUB 3 and Global Language Support*. Retrieved September 2, 2016, from EPUBzone: http://epubzone.org/news/epub-3-and-global-language-support

Microsoft. (2003). *Text-to-speech and the Microsoft speech technologies platform*. Retrieved September 7, 2016, from Microsoft Developer Network: https://msdn.microsoft.com/en-us/library/ms994644.aspx#txt2spch_topic3

Mulyono, H. (2014). Creating native-like but comprehensible listening texts for EFL learners using NaturalReader. *TESL-EJ, 18*(1). Retrieved September 6, 2016, from http://tesl-ej.org/pdf/ej69/m1.pdf

Parr, M. (2013). Text-to-Speech Technology as Inclusive Reading. *LEARNing Landscapes, 6*(2), 303-322.

Rosa, K. D., Parent, G., & Eskenazi, M. (2010). Multimodal learning of words: A study on the use of speech synthesis to reinforce written text in L2 language learning. *Proceedings of the SLaTE Workshop on Speech and Language Technology in Education.* Tokyo: Waseda University. Retrieved from http://www.gavo.t.u-tokyo.ac.jp/L2WS2010/papers/L2WS2010_P2-10.pdf

Schroeter, J., Conkie, A., Syrdal, A., Beutnagel, M., Jilka, M., Strom, V., . . . Kapilow, D. (2002). A perspective on the next challenges for TTS research. *IEEE 2002 Workshop on Speech Synthesis*, (pp. 211- 214). doi:10.1109/WSS.2002.1224411

Tuval Software Industries. (2015). *Text to speech basics*. Retrieved September 7, 2016, from Speechover: http://www.speechover.com/wordpress/text-to-speech/

Wikipedia. (2015). *Amazon Kindle*. Retrieved September 7, 2016, from Wikipedia: https://en.wikipedia.org/wiki/Amazon_Kindle

Wikipedia. (2016a). *Comparison of e-book formats*. Retrieved September 2, 2016, from Wikipedia: https://en.wikipedia.org/wiki/Comparison_of_e-book_formats

Wikipedia. (2016b). *EPUB*. Retrieved September 3, from Wikipedia: https://en.wikipedia.org/wiki/EPUB

Young, C., & Stover, K. (2013). Look what I did! *Reading Teacher, 67*(4), 269-272. doi:10.1002/TRTR.1196