International Journal of Education and Social Science Research

ISSN 2581-5148

Vol. 1, No. 01; 2018

PRODUCING ENGLISH SPEECH SOUNDS BY TEXT-TO-SPEECH TECHNOLOGY FOR ENGLISH AS A FOREIGN LANGUAGE EDUCATION IN JAPAN

Harumi Kataoka
Faculty of Economics, Kindai University, 3-4-1 Kowakae, Higashiosaka City, Osaka, 577-8502, Japan
Email Address: h_kataoka@kindai.ac.jp/h.kataoka2014@gmail.com

ABSTRACT

The purpose of this study is to describe Text-to-Speech (TTS) Technology: its history, mechanism, systems, and structures. This study explains how to produce artificially digital, synthesized English TTS speech sounds on personal laptop/desktop computers; further, it addresses the validity of TTS speech sounds for Japanese learners. Users can produce English speech sounds by using the downloaded TTS system on their personal computers. There are numerous English audio materials for English as a foreign language (EFL) education in Japan. However, almost all English language audio learning materials used in Japanese schools are developed by commercial publishing companies. As Japanese teachers of English (JTEs), we sometimes feel that such audio materials do not fit our students' English abilities. The result of five studies using English TTS sounds for Japanese EFL learners proved that none of the participants noticed that the sounds were artificially digital, synthesized sounds produced by personal computers. This finding implies that using TTS is one way for English language teachers—not only in Japan, but also in other EFL countries—to produce English audio materials that fit their students' English language abilities.

Key Words: Text-to-Speech (TTS) Technology, artificially digital synthesized English TTS speech sounds, personal laptop/desktop computer, English as a foreign language (EFL) education, teacher-produced English audio materials

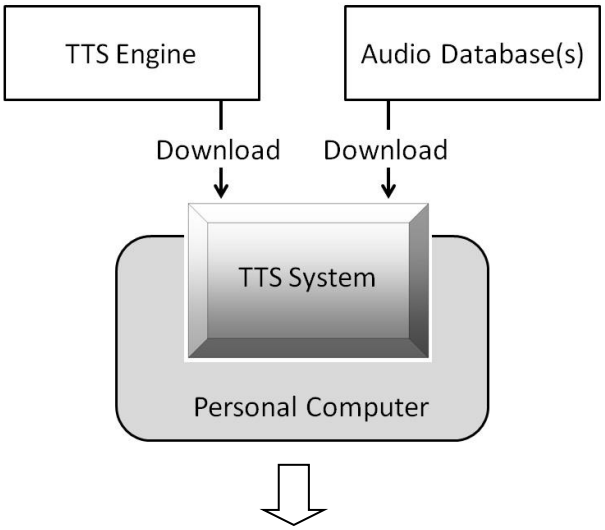
INTRODUCTION

Since 2000, Information and Communications Technology (ICT) has been developed and utilized in offices, schools, and homes in Japan (Okamoto, 2001; Yanagida, 2001). In 2002, the Japanese Ministry of Education, Culture, Sports, Science and Technology (MEXT) published a guideline recommending the use of ICT in the teaching of all subjects in elementary, junior, and senior high schools. In order to implement ICT in this manner, each school was required to install computers that could access the Internet. According to the results of a March 2007 survey by MEXT (2007), Internet access was available to 99.94% of students in public elementary schools, 99.96% in public junior high schools, and 99.98% in public senior high schools. However, the survey also revealed that teachers lacked ICT proficiency; the respondents who believed they were adequately capable of utilizing ICT during lessons when teaching on their own accounted for only 53.49% in public elementary schools, 50.40% in public junior high schools, and 53.30% in public senior high schools.

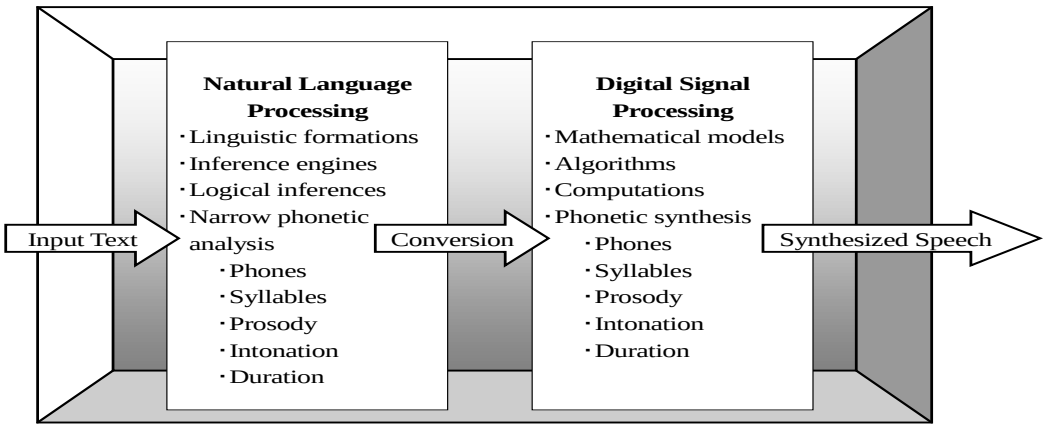
Text-to-Speech (TTS) technology is one of the forms of ICT that has been shown to be useful in language education. TTS sounds are produced by concatenating the recorded voices of native speakers (NSs) of English. TTS in this study is a digital speech synthesis technology composed of two items: (a) a TTS engine and (b) audio database(s) of originally recorded human voices (Dutoit, 1997). Users can produce synthesized digital speech sounds by typing words on the keyboard of their personal computer. Figure 1 shows the basic structure of a TTS system.

By using English audio databases downloaded into teachers' personal computers, teachers can easily create original English audio materials appropriate to their students' English proficiency level (see (1) to (3) in Figure 1). They do not need to secure any help from "live" native speakers of English at all, or to set a date and place to get together to record speech sounds. These are important advantages of TTS technology. However, although TTS is a useful ICT tool for English teachers in the Japanese school setting, many English teachers in Japan remain unfamiliar with TTS technology.

(1) How to create TTS system in personal computer: to download (a) a TTS engine and (b) audio database(s) into personal computer.



(2) After typing words/sentences (text data) on the keyboard of the personal computer and then clicking "Read Aloud Key" on the TTS system, TTS speech sounds are automatically produced.



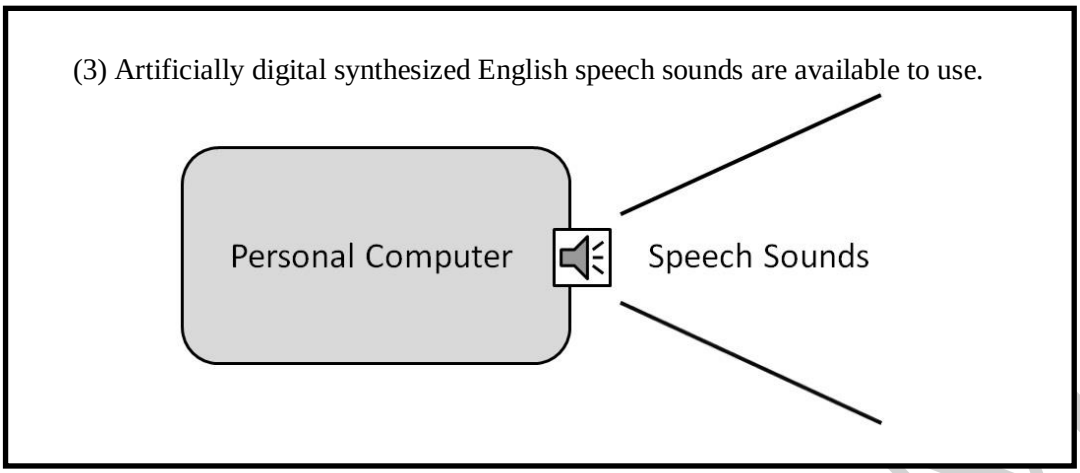


Figure 1.General functional diagram of a TTS system in this study.

To produce English speech sounds, there are a variety of free applications for smartphones—such as the iPhone's "Siri"—and/or free dictionary websites on the Internet that offer English audio pronunciations. However, Sato (2012) pointed out that there is ambiguity regarding copyright issues for speech produced as described above. Therefore, it is advantageous to have the TTS system downloaded on a personal computer; after obtaining the TTS selling agent's permission to use TTS speech sounds, they can be used for EFL education in schools.

History of TTS Systems

Humans have been trying to produce artificial voices for a long time; equally so, TTS systems also have a long history. According to Flanagan (1972, p. 1376), special long tubes were used to produce the voices of the gods in ancient Greece and Rome. In 1779, Christian Gottlieb Kratzenstein, in St. Petersburg, invented a "cavity resonator" machine that, with a vibrating reed, could produce five vowel sounds: ɑ, e, i, o, and u (Flanagan, 1972, p. 1380, Figure 9). This can be viewed as the first TTS system in history.

Since the 1950s, TTS has been pursued using electrical and electronic machines. A number of TTS systems have been developed by acousticians in universities or companies (Azuma, 2006, 2008; Beckman, 1997; Black &Tokuda, 2005; Campbell, 1997, 2002; Campbell & Higuchi, 1997; Chew, 2009; Flood, 2007; Harashima, 2006a, 2006b; Iwaki, Maki, Sonehara, Watanabe, &Kaneyasu, 2008; Kataoka, 2009; Klatt, 1987; Lemmetty, 1999; Matsui, Suzuki, Umeda, &Omura, 1968; Olive, 1997; Sagisaka, Campbell, & Higuchi, 1997; Schroeder, 1993; Taylor, 2009). Figure 21 lays out the modern history of TTS.

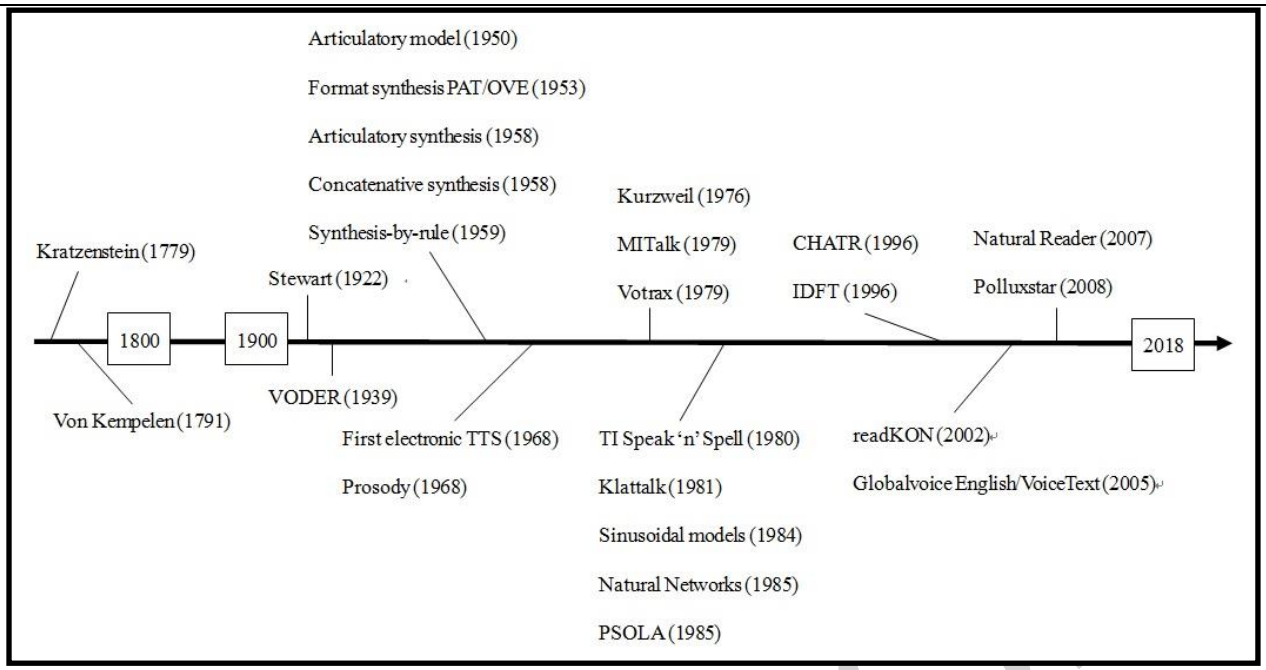


Figure 2. TTS chronology

Taylor (2009, p. 2) mentions that the first practical use of TTS was in reading systems for the visually impaired, where the system would process text from books and convert it into TTS sounds. In addition, TTS systems have come to be used as aids for people with speech disabilities. The system used by British physicist Stephen Hawking may be the most famous TTS system in existence.

In recent decades, computer-aided instruction using TTS sounds has increased in US schools (Reynolds & Jefferson, 1999, p. 174), leading to speech perception studies across a variety of American generations (Drager, Reichle, &Pinkoski, 2010; Koul& Hester, 2006; Pinkoski-Ball, Reichle, & Munson, 2012; Reynolds & Jefferson, 1999).

TTS systems are also utilized in telecommunications; for example, in the automatic responses of telephone services, automated reception desks in hospitals, ATMs in banks, announcements on public transportation facilities, and the assistive technology: Job Access with Speech (JAWS), which reads emails for visually impaired persons. Kevin Lenzo (2006), one of the founders of an American TTS company Cepstral, asserted that the key technology in speech science is TTS. He also mentioned that TTS technology is used for automobile navigation systems, weather forecasts, just-in-time broadcasts, podcasts, and spam radio.

In Japan, Matsui, Suzuki, Umeda, &Omura (1968) at the Electro technical Laboratory (Denshi gijutsusougoukenkyusho) produced 83 different TTS English speech sounds using an IBM 360 computer. Japanese-language TTS sounds have also been produced since the 1990s (Campbell, 1997; Campbell & Higuchi, 1997; Sagisaka, Campbell, & Higuchi, 1997). Iwaki, Maki, Sonehara, Watanabe, and Kaneyasu (2008) researched university lectures in a classroom setting that used Japanese TTS sounds loaded onto and played from a personal computer.

Since around 2000, some TTS systems produced by American companies have served as English teaching materials installed in Computer Assisted Language Learning (CALL) systems in Japan.

However, many English language teachers in Japan have been unable to use TTS technology in their lessons because CALL systems are very expensive. Thus, some stand-alone TTS software has been available for purchase since around 2005 (Azuma, 2006; Harashima, 2006a, 2006b; Sasaki, 2006). Users can install this software on their personal computers to, essentially, create their own TTS systems. The following is a list of available TTS systems as of February 22, 2018.

Table 1
Available TTS Systems

No.	TTS Systems	No.	TTS Systems	No.	TTS Systems
1.	acapela	16.	HADIFIX	31.	readKON
2.	AcuVoice	17.	IDFT	32.	ReadPlease 2003
3.	Apple Plain Talk	18.	Infovox	33.	Sanosse
4.	Balabolka	19.	Laureate	34.	SoftVoice
5.	Bell Labs Text-to-Speech	20.	Lernout &Hauspies	35.	Speechhify
6.	CHATR	21.	Listen2	36.	SPRUCE
7.	Cepstral	22.	Loquendo	37.	SVOX
8.	CNET PSOLA	23.	MBROLA	38.	SYNTE2 and SYNTE3
9.	CyberTalk	24.	Microsoft Reader	39.	TextAloud
10.	DECTalk	25.	ModelTalker	40.	TimehouseMikropuhe
11.	eLite	26	Natural Reader	41.	Ultra Hal Reader
12.	ETI Eloquence	27.	Natural Voices	42.	VoiceText
13	Eurovocs	28.	NeuroTalker	43.	WavePad
14.	Festival	29.	ORATOR	44.	Whistler
15.	Globalvoice English	30.	Polluxstar		

Note. This list shows TTS systems available as of February 22, 2018.

The quality of the synthesized digital speech sounds created by each TTS system varies (Mokhtari& Campbell, 2003); as such, Azuma (2010) and O'ki (2010) recommended that teachers pay close attention to these differences in sound quality as they develop TTS audio materials for their English lessons.

A Variety of TTS Languages

TTS systems are available in many languages; Table 2 shows a list of 54 sets of TTS sounds covering dialects of 33 languages, sold by the online ordering service NextUp.com2 as of February 22, 2018.

Table 2
54 Sets of TTS sounds in 33 Languages

No.	TTS Languages	No.	TTS Languages	No.	TTS Languages
1.	Arabic		English (Wales)	25.	Portuguese (Portugal)
2.	Basque	9.	Faroese		Portuguese (Brazil)
3.	Catalan	10.	Finnish	26.	Romanian
4.	Chinese (Canton)	11.	French (France)	27.	Russian
	Chinese (Mandarin)		French (Canada)	28.	Scanian
	Chinese (Taiwan)	12.	Galician	29.	Slovak
5.	Czech	13.	German	30.	Spanish
6.	Danish	14.	Greek		Spanish (Argentina)
7.	Dutch	15.	Hebrew		Spanish (Castilian)
	Dutch (Belgium)	16.	Hindi		Spanish (Colombia)
	Dutch (Netherlands)	17.	Hungarian		Spanish (Latin American)
8.	English (Australia)	18.	Icelandic		Spanish (Mexico)
	English (Great Britain)	19.	Indonesian		Spanish (United States)
	English (India)	20.	Italian	31.	Swedish (Sweden)
	English (Ireland)	21.	Japanese		Swedish (Finland)
	English (Scotland)	22.	Korean		Swedish (Gothenburg)
	English (South Africa)	23.	Norwegian	32	Thai
	English (United States)	24.	Polish	33.	Turkish

Note. Place names in parentheses show the areas in which people speak the language dialect.

Advantages of TTS Synthesis Speech Sounds as English Teaching Materials

The results of previous TTS studies (Azuma, 2008; Chew, 2009; Harashima, 2006a, 2006b; Hirai &O'ki, 2011; Jones, Berry, & Stevens, 2007; Matsuda, 2012, 2013; Sasaki, 2006; Yoshida, 2008) have shown that TTS sounds have great potential for English language education. Using TTS sounds as audio materials has the following six major advantages over using a native speaker's voice:

(a) By using a TTS audio database downloaded onto their personal computer, English language teachers—not only native speakers of English, but also native speakers of Japanese or other languages—can produce audio materials on their own, and do not need to ask for assistance from native speakers of English.3

(b) Teachers do not have to be concerned with scheduling or paying for recording sessions with native speakers.

(c) TTS systems in this study can control the speed, pauses, pitch, or volume of the voice, and can even change "speakers" just by clicking on a voice selection bar (see Figure 3). It is an easy way for teachers to produce English audio materials that fit their students' English proficiency and/or learning content (i.e., what the students hope to study) in their classroom lessons.

(d) An audio database downloaded into TTS systems can be used to synthesize a native English speaker's voice originally recorded onto a computer, creating speech sounds of stable acoustic quality.

(e) It is easy for teachers to conduct their English lessons using TTS audio materials suitable for their target learners.

(f) The teacher does not need to stand in front of the classroom4 in order to read aloud each English vocabulary/sentence of teaching materials; rather, the TTS software or system can create speech sounds to read aloud the lesson materials. If teachers insert enough pauses for repetitions by adding XML tags in the TTS program (see Figure 3), the speech sounds will include adequate pauses to

allow for repetitions. Therefore, teachers do not need to push a "pause button" to create these pauses for repetitions during read aloud (RA) activities. Consequently, it is easier for teachers to check/inspect students' learning situations by walking between desks while their students are doing RA activities, rather than having to stand in front of the class to read the lesson materials.
Investigation into the use of TTS systems to create audio materials for Japanese EFL students should consider these specific advantages, which may lead to additional ways to further exploit this technology.

Material and Method

Dutoit (1997) stated that TTS systems are composed of two items: (a) a TTS engine and (b) one or more TTS audio databases. To make the results presented here as broadly generalizable as possible, the author chose an economical5 and widely available TTS system.

As the TTS engine, Barabolka (Morozov, 2007), was selected. Because it is charge-free software, users do not need to pay any money, however it can produce high-quality TTS sounds. Users can also insert pauses as short as one millisecond (ms) by easy typing XML tags on the Barabolka interface. Other TTS systems such as Natural Reader and Globalvoice English, which are used in English lessons in Japanese schools, do not allow users to set pauses as sensitively as 1 ms apart. By placing pauses just 1 ms apart, teachers can easily produce TTS audio with more naturally synthesized speech sounds than in the case of non-controlled TTS sounds. These two advantages—no cost and the ability to set pauses in 1 ms units—are why Barabolka (Ver. 1.9.0.241) was used in this study.

For the TTS audio database, the author installed both a male and female TTS voice, both in American English, on a personal computer (Kataoka, 2009). The TTS audio databases6 used were Kate & Paul 16kHz Voices (Ver. 1.0) by NeoSpeech7 and Natural Voice 16kHz Crystal & Mike (Ver. 1.4) by AT&T. These were American English TTS audio databases, purchased on NextUp.com. Only the audio database needs to be purchased, not Barabolka itself.8 Figure 3 shows a screenshot of the TTS system used in this study.

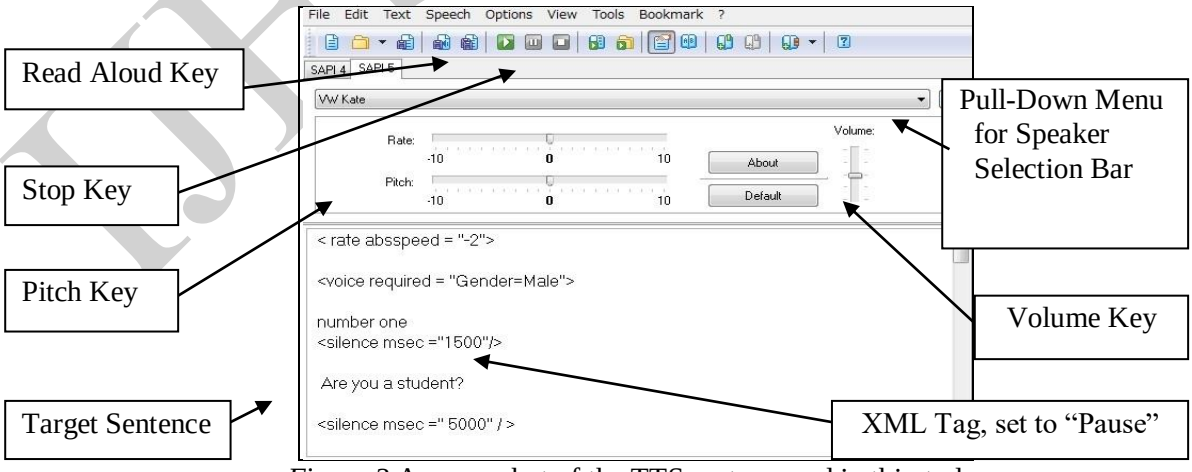


*Figure 3.*A screenshot of the TTS system used in thisstudy.

The author conducted five studies (Kataoka, 2008, 2009, 2013, 2014, 2015) using this TTS system.

The TTS audio materials used in these studies were produced in only eight steps (see appendix for details). Participants in the studies were asked to listen to English speech sounds produced by the TTS system and played on commercial stereo CD players in regular classrooms in Japanese schools. Because the regular classrooms used in these studies were outdated and not equipped for the use of computers, CD decks were employed instead of computers.

To investigate the validity of TTS speech sounds produced by this system, the five previous TTS studies for which this system was utilized were referenced to determine whether the participants in these studies noticed that English speech sounds were artificially synthesized and produced by this TTS system.
The research question (RQ) follows:
Did the participants in the five studies notice that English speech sounds were artificially synthesized sounds produced by this TTS system?

RESULTS and DISCUSSION
Table 3 shows the results of five previous TTS studies. All participants in these studies were Japanese.

Table 3
A Total Number of Participants in Five Studies Used This TTS System

No.	Study	Participants	Did the Participants Notice that English Speech Sounds were Artificially Synthesized Sounds Produced by the TTS System?
1	Kataoka (2008)	85 high school students, 10 JTEs in high schools and universities	×
2	Kataoka (2009)	Nine university students and 10 university graduate	×
3	Kataoka (2013)	151 high school students	×
4	Kataoka (2014)	31 high school students	×
5	Kataoka (2015)	49 high school students	×

Notes. "JTE" means Japanese Teacher of English. "×" mark shows "none".

A total of 345 participants (316 Japanese high school students, nine Japanese university students, 10 Japanese university graduates, and 10 JTEs in high schools and universities) in the previous studies did not notice that English speech sounds were artificially synthesized and produced by the TTS system introduced in this study.

This result shows that the speech quality of TTS is satisfactory for Japanese EFL learners studying English in Japanese schools.

CONCLUSION
The RQ in this study was "Did the participants in the five studies notice that English speech sounds

were artificially synthesized sounds produced by this TTS system?"
As mentioned previously, the findings revealed that none of the participants in the five studies noticed that the English speech sounds were artificially synthesized and produced by this TTS system.

The TTS audio materials used in the five studies were produced in only eight steps (see Appendix for details). Although TTS technology is a form of ICT, the procedures to create English speech sounds using users' personal computers are quite simple. Therefore, teachers will find it rather easy to produce English audio materials for EFL education during their lessons. It is important that the procedure be very simple to help address concerns of time management and technical skill levels among English teachers in Japanese schools who may think that using a TTS system would be too difficult or time-consuming for them. Knowledge on these areas is somewhat lacking among English language teachers in Japanese schools.

It is the hope of the author that the findings of this study encourage more teachers of English—not only in Japan, but also in other EFL countries—to produce TTS English audio materials suited to their students' English language abilities

Notes
1. The TTS chronology in Figure 2 is a reorganized version referring to a figure from Lemmetty (1999, p. 9, Figure 2.4).
2. Anyone can purchase TTS audio databases produced by four companies—AT&T, Acapela, Ivone, and NUANCE—on the website NextUp.com, as of February 22, 2018. These TTS audio databases can be used not only with personal computers but also with iPhones/iPads via the TTS app TextAloud for iOS, which can also be purchased on NextUp.com.
3. Professor Robert Maran, who is a native speaker of English and the author's coworker and supervisor at Osaka Shoin Women's University, told her that he usually produces TTS sounds with the Globalvoice English system, and uses them as listening materials for a Moodle-based learning management system (LMS). He explained that there were two reasons why he uses TTS sounds: (1) it would be difficult to find two native speakers of English (e.g., a male and a female) to record their voices every time he wanted to make audio materials, and (2) he can produce TTS materials on his laptop computer on the train while commuting to the university.
4. English audio materials to be used in Reading Aloud (RA) (Kataoka, 2013, 2014, 2015) need to have enough pauses between each phrase, clause, or sentence so that participants can repeat the item using chorus reading. Without appropriate pauses, teachers have to stand by the audio player and manually pause the audio materials for repetition each time. With a TTS system, teachers can simply insert a pause, of a length fit for their students' English proficiency, between items by using XML tags (Figure 3, and Table 4 in Appendix) on the TTS system interface. This is less time-consuming than using English audio materials such as audio CDs published by commercial publishing companies or English voices recorded directly by native English teachers working at Japanese schools. In those instances, teachers may have to record the English speech sounds using audio editing software (e.g., Audacity) first, in order to insert a pause between each item by listening to all items in turn, since the audio data cannot be visualized as text onscreen. Although some waveforms are shown on a screen in Audacity, teachers have to listen to each waveform one-by-one by clicking a start/stop bar on the screen in order to recognize which waveform is which English sentence. Of

course, some commercial English audio CDs have pauses sufficient for RA, but sometimes the pauses do not fit students' English proficiency.

5. The five previous studies (Kataoka, 2008, 2009, 2013, 2014, 2015) were conducted in Japanese high schools and a university. The budget that allows the purchase of software programs for English language teaching in a year is limited and varied across each school; therefore, the author decided to choose one of the more economical ways to use a personal computer as a TTS system.

6. TTS sounds used in this study were checked by the author, 11 JTEs, and four native speakers of English. They were teachers in high schools/universities. They listened to the TTS sounds and judged them on audio quality for use as English audio materials during regular high school/university English lessons in Japan.

7. The TTS audio databases, Kate & Paul, were also installed in TTS system, Globalvoice English, as of 2010.

8. The author bought a TTS audio database CD in April 2007 for $50, including postage charges. She received permission through the selling agent for this audio database to create TTS CD-ROMs for her studies. Although she downloaded the TTS software application Balabolka as a TTS engine onto her personal computer, it was free, so the total cost for her to have a TTS system installed was $50.

REFERENCES

Azuma, J. (2006). CALL oyobi eLearning kankyoudeno TTS eigogouseionsei no katsuyoukenkyuu (Practical use research of the TTS English synthesized speech sounds in CALL and/or eLearning environment). Proceedings of the 46th Annual Conference of the Japan Association for Language Education and Technology, 358–366.

Azuma, J. (2008). Applying TTS technology to foreign language teaching.In F. Zhang & B. Barber (Eds.), Handbook of research on computer-enhanced language acquisition and learning (pp. 497–506). New York: Information Science Reference.

Azuma, J. (2010). TTS gouseionseidehirogarugaikokugomaruchimediakyouzaikaihatsunoaratanachihei (Impact of TTS technology on foreign language teaching: New horizons of multimedia teaching material development). Journal of the Center for Research and Development in Higher Education, University of Marketing and Distribution Sciences, 6, 1–11. Retrieved from https://www.umds.ac.jp/facility/ksc/bullet/documents/kiyo_vol6_azuma.pdf

Beckman, M. E. (1997). Speech models and speech synthesis. In J.P.H. Santen, R. W. Sprout, J. P. Olive, & J. Hirschberg (Eds.), Progress in speech synthesis (pp. 185–209). New York: Springer.

Black, A. W., &Tokuda, K. (2005). The Blizzard Challenge – 2005: Evaluating corpus-based speech synthesis on common datasets.
Retrieved from https://www.cs.cmu.edu/~awb/papers/is2005/IS051946.PDF

Campbell, N. (1997). CHATR: Onseigouseidatabeseshorinituite (Processing a Speech Corpus for Synthesis with CHATR). Information Processing Society of Japan, SIG Notes, 97 (66), 109–114.

Campbell, N. (2002). Onseigouseinokantenkaramitagengoonseinotokucyou (Characteristics of speech: from the viewpoint of speech synthesis). GekkanGengo 2002nen10gatsugou, 52–61.

Campbell, N., & Higuchi, N. (1997).Maruchiwashamaruchigengoonseigouseisisutemu CHATR: Okiniirinokoewo computer de gousei (Multi speaker multi language voice synthesis system , CHATR: With favorite voice synthesized by the computer). ATR Journal, 26, 8–9.
Retrieved from http://results.atr.jp/atrj/ATRJ_26/08/abstract.cgi#top

Chew, L. C. (2009) Promises and challenges of e-assessments: A case of multimedia listening testing. LET Kansai Chapter Collected Papers, 12, 1–19.
Drager, D. R. K., Reichle, J., &Pinkoski. C. (2010). Synthesized speech output and children: A scoping review. American Journal of Speech-Language Pathology, 19, 259–273.
Dutoit, T. (1997).An introduction to text-to-speech synthesis. The Netherlands: Kluwer Academic.
Flanagan, J. F. (1972). Voice of men and machines. Journal of the Acoustical Society of America, 51, 1375–1387.
Flood, J. (2007).NaturalReader: A new generation text reader. Developmental Disabilities Bulletin, 35, 44–55. Retrieved from http://files.eric.ed.gov/fulltext/EJ812645.pdf
Harashima, H. (2006a). Review of "VoiceText." Electronic Journal of Foreign Language Teaching, 3 (1), 131–135. Retrieved from http://e-flt.nus.edu.sg/v3n12006/rev_harashima.pdf
Harashima, H. (2006b). Onseigouseiniyorueigorisuningusozai no sakusei (Creating English listening materials using speech synthesis). Proceedings of the 22nd Annual Conference of Japan Society for Educational Technology, 789–790.
Hirai, A., &O'ki, T. (2011).Comprehensibility and naturalness of Text-To-Speech synthetic materials for EFL listeners. JACET Journal, 53, 1–17.
Iwaki, T., Maki, I., Sonehara, R., Watanabe, S., &Kaneyasu, T. (2008).Intekishitadaigakukyoujyuniyorugouseionseiwotukattakougi no houkoku to kousatsu (A report and consideration regarding lectures and dialogues using synthesis voice).Proceedings of Human Interface 2008, 1334.
Jones, C., Berry, L., & Stevens, C. (2007). Synthesized speech intelligibility and persuasion: Speech rate and non-native listeners. Computer Speech and Language, 21, 641–651.
Kataoka, H. (2008). The use of Text-To-Speech synthesis technology for foreign language education in a Japanese school.Proceeding of CLaSIC 2008, The Third International Conference, 318–326.
Kataoka, H. (2009). Text-To-Speech (TTS) synthesis technology wokatsuyoushitaeigokyouikukyouzai no kaihatsu to nihonjin no onseininshiki (The use of Text-To-Speech (TTS) synthesis technology for English education: Speech recognition of Japanese EFL learners). Journal of Kansai University Graduate School of Foreign Language Education and Research, 7, 1–33.
Kataoka, H., & Ito, M. (2013). A comparative study on reading aloud: Instruction by Text-To-Speech synthesis sounds and a high school Japanese English teacher. THE JASEC BULLETIN, 22 (1), 39–54.
Kataoka, H. (2014). Oral repetitions of English vocabulary using Text-To-Speech synthesis sounds: Preparation for university entrance examinations. THE JASEC BULLETIN, 23 (1), 71–86.
Kataoka, H., Ito, M. & Yamane, S. (2015). Retention of English sentences learned by reading aloud using Text-To-Speech (TTS) speech sounds: A longitudinal study in a Japanese high school. International Journal of Research Studies in Educational Technology, 5 (1), 29–47. DOI: 10.5861/ijrset.2015.1331
Klatt, D. H. (1987). Review of text-to-speech conversion for English. The Journal of the Acoustical Society of America, 82 (3), 737–793.
Koul, R., & Hester, K. (2006).Effects of repeated listening experiences on the recognition of synthetic speech by individuals with severe intellectual disabilities. Journal of Speech, Language, and Hearing Research, 49 (1), 47–57.
Lemmetty, S. (1999). Review of Speech Synthesis Technology (Master's thesis). Helsinki University of Technology Department of Electrical and Communications Engineering. Retrieved

from http://www.acoustics.hut.fi/publications/files/theses/lemmetty_mst/thesis.pdf
Lenzo, K. (Speaker) (2006, January 25). Text-to-speech: Make it talk. Recorded in O'Reilly Media Emerging Telephony Conference, San Francisco, USA. [Review of the web site, IT Conversation]. Retrieved from http://www.itconversations.com/shows/detail1660.html
Matsuda, N. (2012). Effects of auditory word repetition on speech processing of Japanese EFL learners. Language Education & Technology, 49, 143–172.
Matsuda, N. (2013). Second-language speech processing: Auditory word priming in Japanese EFL learners and native English speakers. Journal of the Japan Society for Speech Sciences, 14, 43–62.
Matsui, E., Suzuki, T., Umeda, N., &Omura, H. (1968).Synthesis of fairy tales using an analog vocal tract.Proceedings of the 6th International Congress on Acoustics, B-159–B-162.
Ministry of Education, Culture, Sports, Science and Technology (MEXT). (2002). Jouhoukyouiku no jissen to gakkou no jouhouka: Shin "jouhoukyouikunikansurutebiki" (Doing lessons of information education and computerization in the schools: Newly "a guide about the information education").
Retrieved from http://www.mext.go.jp/a_menu/shotou/zyouhou/020706 d. pdf
Ministry of Education, Culture, Sports, Science and Technology (MEXT). (2007). Heisei 18 nendogakkouniokerukyouiku no jyouhouka no jittaitounikansurucyousakekka (A summary of the results of a survey about the actual situation of information technology/computerization in the schools for the year of 2006 (Heisei 18 fiscal year).
Retrieved from http://www.mext.go.jp/a_menu/shotou/zyouhou/08092208.htm
Mokhtari, P., & Campbell, N. (2003).Quasi-syllabic and quasi-articulatory-gestural units for concatenative speech synthesis. Proceedings of the 15th International Congress of Phonetic Sciences, 1–4. Retrieved from http://www.speech-data.jp/nick/pubs/0986anav.pdf
Morozov, I. (2007). Barabolka. Retrieved from http://www.cross-plus-a.com/jp/balabolka.htm
NextUp.com. (2018). Get the best voice available for TextAloud 3 on your PC. Retrieved from http://nextup.com/index.html
Okamoto, T. (2001).Jyouhoukomyunikeisyongijyutu (ICT) to gakusyuu (New learning envirnments with information/communication tencnology).Japan Journal of educational technology, 25 (2), 59–61.
O'ki, T. (2010). Gousei onsei no EFL risuningutesutohenotekiyoukanouseinitsuite (Applicability of synthesized speech for EFL listening tests). The Hakuoh University Journal, 25 (1), 195–209.
Olive, J. P. (1997). Section introduction: Concatenative synthesis. In J. P. H. van Santen, R. W. Sprout, J. P. Olive and J. Hirschberg (Eds.) Progress in Speech Synthesis (pp. 261–262). New York: Springer.
Pinkoski-Ball, C. L, Reichle, J., & Munson, B. (2012). Synthesized speech intelligibility and early preschool-age children: Comparing accuracy for single-word repetition with repeated exposure. American Journal of Speech-Language Pathology, 21, 293–301.
Reynolds, M. E., & Jefferson, L. (1999). Natural and synthetic speech comprehension: Comparison of children from two age groups. Augmentative and Alternative Communication, 15, 174–182.
Sagisaka, Y., Campbell, N., & Higuchi, N. (Eds.) (1997). Computing prosody: computational models for processing spontaneous speech. N.Y.:Springer
Sasaki, A. (2006). Goi no jugyonimochiirudijitarufurasshukaadonionseigoseigijutsu (TTS) wotorikomukokoromi (Use of the TTS system for digital flashcards in vocabulary lessons). Proceedings of the 46th Annual Conference of the Japan Association for Language Education and Technology, 451–454.

Sato, H. (2012). Siri no siyouwokumikondaeigohatsuonsidounokouka: Pairottosutadhii (Effects of pronunciation training combined with the use of Siri: Pilot study). JACET Kansai Chapter 2012 Fall conference, 1. Retrieved from http://www.jacet-kansai.org/file/2012f-2.pdf
Schroeder, M. (1993).A Brief history of synthetic speech. Speech Communication, 13, 231–237.
Taylor, P. (2009) Text-to-Speech synthesis. Cambridge, U.K.: Cambridge University Press.
Yanagida, S. (2001). Denshi ticyounojitugenwomezasite: Toukyoutono IT kasuisinnotorikumi (Toward electronic Tokyo Metropolitan Government: propulsion effort to IT). Japan Society of Information and Knowledge, 11 (3), 57–62.
Yoshida, S. (2008).Development and practice of an electronic phrasal verb wordbook with GIF animations.Proceedings of the 3rd International Conference of World CALL 2008, 84.

Appendix: How to Produce TTS Audio Materials for EFL Education

Step 1: Users have to download two items into their personal computers: (a) a TTS engine—here, the free TTS software application "Balabolka" (Ver. 1.9.0.241); and (b) one or more TTS audio database(s).

Step 2: Users type the English sentences or vocabulary they want to make TTS audio materials out of into the keyboard of their personal computer. In case they would like to make same English sentences or vocabulary again and again, instead of directly typing them on the screen of Balabolka, they can type them on a memo pad integrated with the system and keep it. They can also write XML commands to control features like speed, pause, and voice type, and use these commands when they make TTS speech sounds.

Step 3: Users open the memo pad. Figure 4 is a screenshot[1] of the TTS system used in this system.

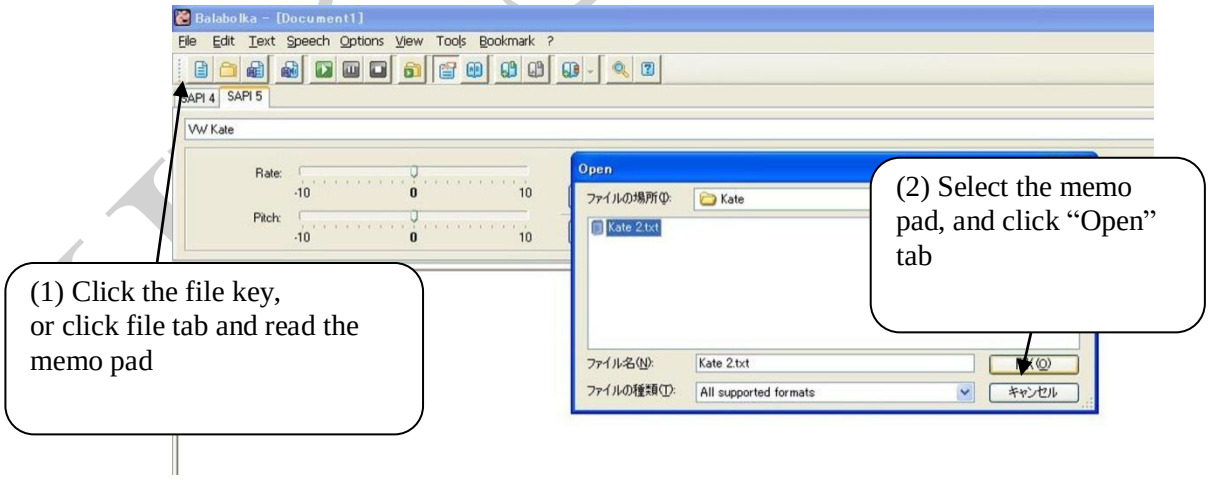


Figure 4.A screenshot of the TTS system: Selecting and reading a memo pad.

Step 4:Users can listen to TTS speech sounds produced by the TTS system. First of all, they can move a cursor to the beginning of the English text on the screen of the TTS system. Then, they can click the "Read Aloud" key and listen to check the TTS speech sounds.

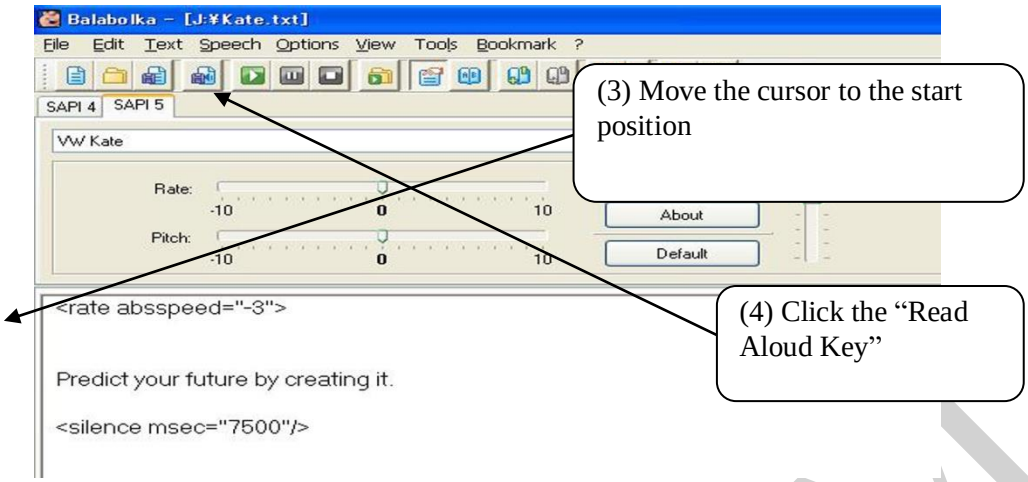


Figure 5.A screenshot of the TTS system: ListeningTTS speech sounds.

Table 4
Voice Generation MethodUsing a TTS System

Item	English Sentence/Vocabulary; XML Tag
An English Sentence	Predict your future by creating it.
Speed; Speech Rate	<rate absspeed="-3">
Pause	<silence msec="7500"/>

Step 5:Users can save TTS speech sounds. They move the cursor back to the start position,thenclick the "File" tab and select "Save Audio File."

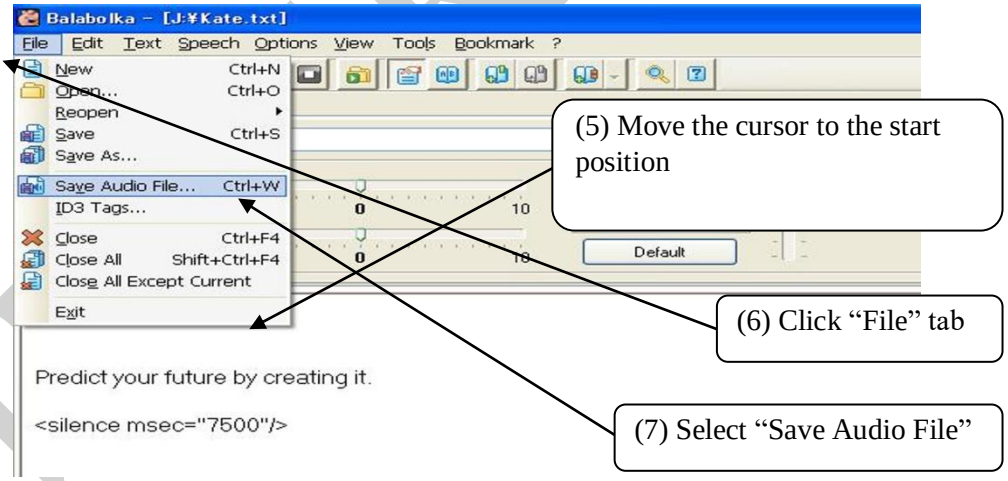


Figure6.A screenshot of the TTS system: Saving an audio file.

Step 6:Users write a "File Name" and select a "File Type," and they finally click the "Keep" key.

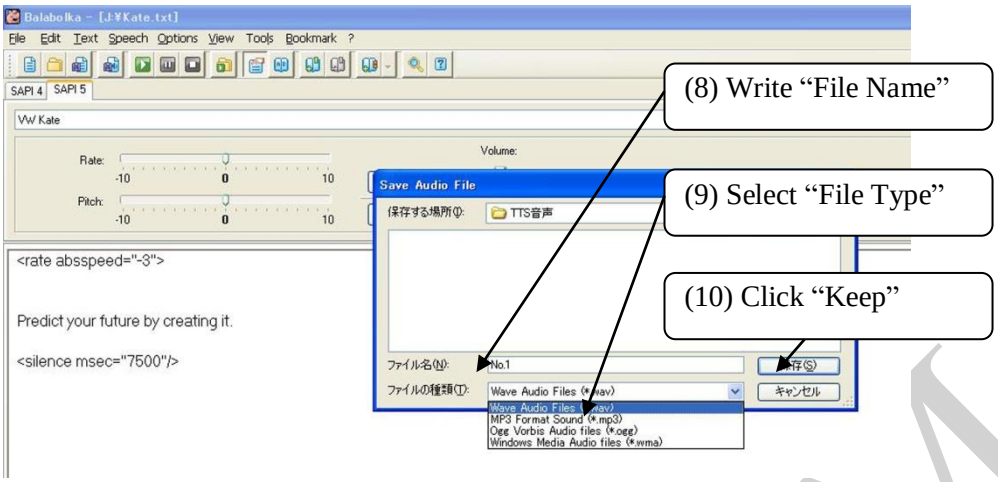


Figure7. Screenshot of TTS system: Selecting an audio file to save it.

Step 7:Users have to wait for a while to save the audio file. The author kept her TTS audio files on a USB.

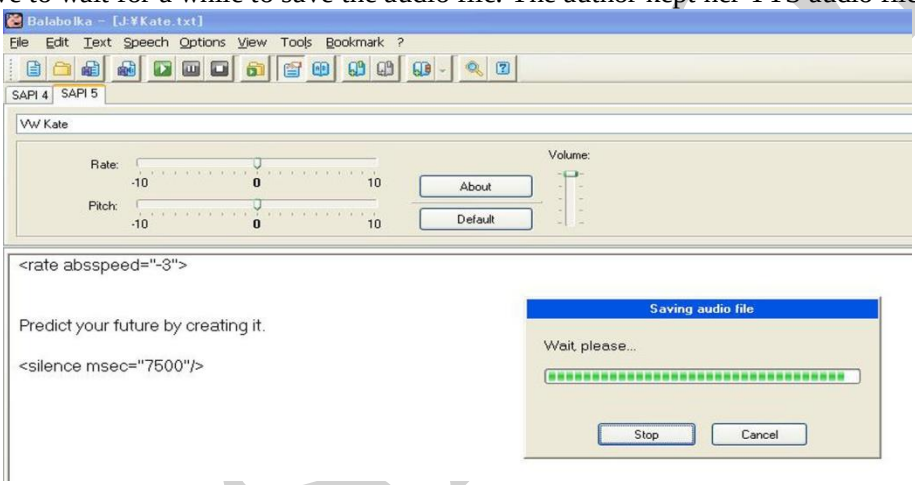


Figure 8.A screenshot of the TTS system: Converting an audio file.

Step 8:Record audio files onto a CD-ROM; the author used iTunes (see Figure 9).[2]The TTS CDswere played using a stereo CD player, and given to the participants in the studies.

Figure 9.A screenshot of iTunes.

Notes

1. The screenshots of the TTS system from Figures 5 to 8 were recorded in August 2008.
2. The screenshot of iTunes in Figure 9 was recorded on March 5, 2014.