Text-to-Speech Technology: What Does It Offer to Foreign Language Learners?

Ferit Kilickaya (kilickay@metu.edu.tr)
Middle East Technical University, Turkey

Abstract

Language teaching is a difficult process that requires careful work. Educators try to find ways to make this difficult process enjoyable for language learners. As technological developments came into use, language learning became more attractive. 'Text to Speech' (Speech synthesis) technology, one of these technological developments, is the process of synthesizing natural-sounding speech from any text via special computer programs. These programs can create the most realistic, human-sounding synthetic speech which is available today. This paper explores the possible uses, advantages, and disadvantages of this technology considering EFL learners.

Keywords: Language learning, text-to-speech technology, speech synthesis, listening, and technology in the classroom

Introduction

Language teaching is a rather difficult and complicated process that requires careful and diligent work. Educators in the field of language teaching always try hard to find ways to make language learning enjoyable and attractive for the learners. Different activities, games, and interesting stories helped language teachers to achieve this aim through many years and they still do.

Today, we have access to many CALL programs that are currently used and tested in language classrooms for teaching grammar, speaking and other skills. Text-to-speech technology is a common feature of almost any CALL application. 'Text-to-Speech' (Speech Synthesis) technology is the ability of a computer to produce 'spoken words'. Computer speech can be produced either by "splicing" prerecorded words together or, with much more difficulty, by having the computer produce the sounds that make up spoken words (Microsoft Encarta Encyclopedia Deluxe, 2004)

In other words, text-to-speech is the conversion of text to speech through special computer applications, which is often referred to as Text To Speech software (TTS). Text-to-speech software is invaluable for blind computer users as it enables them to "read" from the screen. This technology was first introduced as Texas Instruments Speak and Spell handheld electronic learning aid in 1978. Language learners and teachers need to be informed about this technology, its possible uses, advantages, and limitations since this technology is new to them.

Language Learning and Technology: A brief review of the history

As Warschauer and Meskill (2000) suggest, every type of language teaching uses its techniques to help learners. With the introduction of the Grammar-translation method, the blackboard came into use in language classrooms. Later it was replaced by an overhead projector. Following them, computer software was used to provide students with drill-and-practice exercises.

The first use of computers by institutions related to teaching and learning coincided with the introduction of second-generation computers towards the 1950s. Large universities started to use computers for administrative processes and student record keeping. At the same time computers were used for instructional teaching and research. PLATO (Programmed Logic for Automatic Teaching Operations), the very first project related to using computers in educational research, began in 1960 at the University of Illinois to design a large computer-based system for instruction. The PLATO system included a mainframe machine supporting hundreds of terminals which have high capacity comparing to that age. Many courses in many disciplines were developed, designed, and delivered on PLATO systems (Alessi & Trollip, 1985; Warschauer, 1996; Levy, 1997; Culley, 1992). Later, new versions of PLATO came into use with new changes to provide interactive and self-paced instruction.

During the 1960s and 1970s, the use of computer-assisted instruction expanded in public schools with the introduction of the next generation of computers and microchips which were cheaper (Bullough & Beatty, 1991). In 1971, another important project, TICCIT (Time-shared, Interactive, Computer Controlled Information Television) was initiated at Brigham Young University (Levy, 1977). The system combined television technology with the computer to deliver instruction to the learners.

During the 1980s, microcomputers started to be adopted by the schools, and new developments such as CD-ROM, speech-based software, and interactive videos appeared. Also, experiments were done in the integration of computers into the curriculum. In the 1990s and 2000s, with the introduction of fast, affordable processors, new software, wide-scale and fast access to the Internet made computers available in almost all public and private schools as well as homes for personal and educational use.

Meanwhile, what went unnoticed was the 'text-to-speech' technology designed for visually impaired people. Speech synthesis is the conversion of text to speech through special computer applications, which is often referred to as Text To Speech software (TTS). Text-to-speech software is considered invaluable for the blind since it enables them to read from the computer screens. However, it didn't take much attention from language learners and teachers. This might be attributed to the views on this new technology as Higgins (as cited in Ehsani & Knodt, 1998, p.46) states "Because speech technology isn't perfect, it is of no use at all. If it cannot account for the full complexity of human language', why even bother modeling more constrained aspects of language use."

Although what Higgins said cannot be confronted since speech technology is not perfect in terms of the complexity of human language, it is important to note that it has some possible uses in language teaching and learning. We also should take into account that technology improves day by day and it is no doubt that what is good today will be better tomorrow as Ehsani and Knodt (1998) and Sobkowiak (2003) stated that text-to-speech technology will be a common feature of any CALL application and human language technologies will improve the current software of foreign language teaching. (Please refer to Audio5 for the early TTS sound technology).

'Text to Speech' Computer Applications

Currently, there are three computer applications available to home users who want to benefit from this technology, namely, *Natural Voice Reader*, *ReadPlease Plus 2003*, and *TextAloud MP3*. All these programs are aimed to produce the most realistic human-sounding voices. However, Garrett (1998, p. 81) states "This technology isn't at a stage where it can reliably render a target language accent authentic enough for language use." Before dealing with the performance issue, the first question to be answered is what these applications can do.

What these applications can do

In general, these applications using 'Text to Speech' technology can

- Read any text on the computer (web pages, word documents, rich texts, e-mails, news articles, online books, etc.)
- Give the option of reading any text and saving it to a file in the form of .wav or .mp3 files, which allows listening to them later in your MP3 or CD player.
- Read any text at any speed and any speaking quality.
- Read any text using the voice or any accent (male, female, British English American English, etc.)

Performance consideration

The most important consideration perhaps is whether this technology can create authentic speech. In other words, will the speech produced be authentic enough? To understand this, Natural Voice Reader Enterprise Edition having AT&T Mike and Crystal American English Voices were tested and used to create human-sounding versions of a mini dialogue and a long text (see Appendices) from TOEFL Test Preparation Kit published by Educational Testing Service (ETS) in 1995 (Audio2 and Audio3). The original speech files are Audio1 and Audio4). When compared, it was noticed that the resulting voices were satisfactory in terms of pronunciation and clearness, however; some limitations were noted.

Taking language learners into consideration, the following list can be made regarding the uses of these programs:

Advantages

- You can listen to any text and any topic (Most EFL listening materials cover a limited range of topics and some of them are rather expensive.
- You can adjust the speed of reading according to your own needs.
- You can create audio versions from any text (.wav or mp3 files).
- You can create pronunciation exercises for yourself (A single word can also be read.)
- You can create mini dialogues (changing speakers at run times is possible).

Limitations

- Although these programs can create realistic, human-sounding voices, there is always a difference in terms of intonation and stress. In other words, it still lacks the complexity of naturally occurring speech, resulting in 'dead' sound having no emotions. This is easily identified while these programs are reading rather longer sentences (compare Audio1 with Audio2 for a mini dialogue and compare Audio3 and Audio4 for a long text). It should be noted that there is no limit to the technological advances and acquiring the complexity of naturally occurring speech may be possible in the near future.
- These programs require newer and faster computers and enough hard disk space to run. (Operation system: Windows 98/Me/NT/2000/XP, Processor: 500 MHz, Memory: 128 MB memory, Disk Space required: 500MB +600 MB for each voice.). However, today new computers are highly capable of what is required.)

Examples of how this technology can be used

1. Creating a list of frequently-mispronounced words

Language teachers and learners can create a list of frequently-mispronounced words and save this list as a ".wav" for later use. Learners can listen to these words and repeat them while the file is placed. Below is an example of a possible list (Audio5):
Foreign
Interesting
Determine
Occurrence
Preface
Comparable
Compare
Comparison
Carriage
Marriage
Natural
Mature
Iron
Capable
Business
Major
Subtle
Impotent
Suitable
Support

2. Writing short sentences and listening

Language learners can short sentences to the extent that their imagination allows and listen to these dialogues. In this way, this process can be made enjoyable and fascinating. Below are possible short sentences that can be created (Audio6)
Excuse me, Where is the nearest post office, please?
What kind of books do you read?
What kind of music do you like?
What do you do when you are bored?

3. Creating short dialogues

Language learners can also write dialogues while changing the speakers of the programs. Below is a possible dialogue that can be created (Audio7)

Excuse me, Where is the nearest lost property office, please?
I'm sorry, I don't know.
Thank you anyway.
Not at all.

Reading and listening to newspapers online

Language learners can also read newspapers online and save them as .wav files for later use. Below is an online article which was saved as a sound file (Audio8)
"A 28-year-old South Korean man has died after playing an online computer game for almost 50 hours non-stop. The man, known only by his family name of Lee, started playing the popular battle simulation game *Starcraft* on August 3 and was fixed to his seat for over two days. His marathon gaming session was broken only with the occasional toilet break or five-minute nap. Reuters News Agency reports police sources saying the man died from cardiac arrest "stemming from exhaustion".
Lee was on a mission to become a professional gamer. This is an increasingly attractive and well-paid profession in South Korea. Top players can earn substantial amounts of money each year. Lee had recently been fired from his job because of absences due to his obsession with gaming. The dangers of being addicted to fantasy games are resulting in many social problems. In particular, MMORPGs, or massively multiplayer online role-playing games, keep thousands of players glued to their screens for many hours."

Conclusion

'Text to Speech' (Speech Synthesis) technology has improved a lot and it is ready to be deployed in language learning provided that its limitations are taken into consideration. If instructors are trying to expose students to natural language audio input and 'comprehensible input' (Krashen, 1985) as much as possible, this technology can provide a valuable way of doing it provided that its limitations are fully understood and as Ehsani & Knodt stated "it is used in ways that work around these limitations.

Recommendations

Based on the discussions made in this paper, the following recommendations have been made:

- 'Text to Speech' (Speech Synthesis) technology is ready to be deployed in second language education and instructors should be willing to explore possible uses of this technology having its limitations into considerations.
- Experimental studies are needed to fully understand the possible uses/effects of this technology in language learning situations. Also, language learners' views and needs on the use of this technology will be beneficial in directing the future development of this technology.

References

Alessi, S. M. & Trollip, S. R. (1985). *Computer-based instruction: Methods and development.* New Jersey: Prentice-Hall.

Bax, S. (2003). *CALL-past, present, and future*. System, 31(1), 13-28.

Bullough, R. V., & Beatty, L. F. (1991). *Classroom applications of microcomputers*. Republic of Singapore: Macmillan Publishing Company.

Culley, G. R. (1992). From syntax to semantics in foreign language CAI. In J. H. Larkin and R. W. Chabay (Eds.), (pp. 47-72). *Computer-assisted instruction and intelligent tutoring systems: Shared goals and complementary approaches*. USA: Lawrence Erlbaum Associates, Inc.

Cunningham, D. (1998). 25 years of technology in language teaching: A personal experience. Babel: *Journal of the Australian Federation of Modern Language* Teachers' Associations, 33(1), 4-7, 35.

Educational Testing Service (ETS). (1995). *TOEFL Test Preparation Kit*. USA: ETS.

Ehsani, F., & Knodt, E. (1998). Speech technology in computer-aided language learning: Strengths, and limitations of a new call paradigm. *Language Learning and Technology*, 2(1), 45-60.

Garret, N. (1991). Technology in the service of language learning: Trends and issues. *The Modern Language Journal*, 75(1), 74-101.

Krashen, S. (1985). *The input hypothesis: Issues and implications*. London: Longman.

Levy, M. (1997). *Computer-assisted language learning*. USA: Oxford University Press.

Microsoft Encarta Encyclopedia Deluxe (Version 2004). [CDROM]. Microsoft Corporation.

Natural Voice Reader (Version 3.5.1) [Computer software]. USA.

ReadPlease Plus 2003. (Version 2003.1.10) [Computer software] Canada: ReadPlease Corporation.

Sobkowiak, W. (2003). TTS in EFL CALL- some pedagogical considerations. *Teaching English with Technology: A journal for Teachers of English*, 3 (4).

TextAloud MP3. (Version 1.60) [Computer software] USA: NextUp Technologies.

Warchauer, M., & Meskill, C. (2000). Technology and second language learning. In J. Rosenthal (Ed.), *Handbook of undergraduate second language education* (pp. 303-318). Mahwah, New Jersey: Lawrence Erlbaum.

Warschauer, M. (1996). Computer-assisted language learning: An introduction. In S. Fotos, (Ed.), *Multimedia language teaching* (pp.3-20). Tokyo: Logos International.

Sites of interest to readers

http://www.naturalreaders.com

(Natural Voice Text-to-speech Reader software. You can have your computer read documents aloud, using high-quality Natural voice. With Build-in web browser, you can view any web news on the Internet, and have the computer to read any part of the news, weathercast, charting messages, and emails. The application can read word documents, rich text files, and PDF files.)

http://www.readplease.com

(TextAloud MP3 lets you listen to text you copy to the clipboard. It uses 'Text to Speech' technology which synthesizes human-sounding speech from ordinary text.)

http://www.nextup.com

(ReadPlease Plus 2003 will read any text you see on your screen. This can be from your Browser, Email, Word processor, Spreadsheet, or any program which displays text.)

http://vlc.polyu.edu.hk

TTS resources at the Virtual Language Centre of the Polytechnic University of Hong Kong.

http://www.gutenberg.net

(Project Gutenberg, the brainchild of Michael Hart, is an excellent source of a lot of famous and important texts which are in plain text format. The computer applications above can read these texts.)

http://www.pcww.com

Winspeech: a TTS program.

http://www.freedomscientific.com

JAWS for Windows: A remarkable TTS tool.

http://elsap1.unicaen.fr/KaliDemo.htm

KALI, a demonstration TTS package (French) from the University of Caen.

http://www.rhetorical.com

a TTS interactive demo (male and female voices speaking with American, British, Scottish, and Australian, etc).

APPENDICES

Appendix A

Script for Audio1 & Audio2

Woman: Do you know anyone who can translate this document?
Man: What about the new secretary? I heard he's bilingual

Appendix B

Script for Audio3 and Audio4

(Man) Before we begin our tour, I'd like to give you some background information on the painter Grant Wood-we'll be seeing much of his work today.

Wood was born in 1881 in Iowa farm country and became interested in art very early in life. Although he studied art in both Minneapolis and at the Art Institute of Chicago, the strongest influences on his art were European. He spent time in both Germany and France and his study there helped shape his own stylized form of realism.

When he returned to Iowa, Wood applied the stylistic realism he had learned in Europe to the rural life he saw around him and that he remembered from his childhood around the turn of the century. His portraits of farm families imitate the static formalism of photographs of early settlers posed in front of their homes. His paintings of farmers at work, and of their tools and animals, demonstrate serious respect for the life of the Midwestern United States. By the 1930s, Wood was a leading figure of the school of art called "American regionalism."

In an effort to sustain a strong Midwestern artistic movement, Wood established an institute of Midwestern art in his home state. Although the institute failed, the paintings you are about to see preserve Wood's vision of pioneer farmers.