

TECHNOLOGY IN PEDAGOGY

Series



If you can't Say it, Voice it: Using Text-to-speech in Presentations

By A/P Stéphane Bressan

Technology in Pedagogy, No. 7, February 2012

Written by Kiruthika Ragupathi (kiruthika@nus.edu.sg)

Text-to-speech (TTS) is not anymore the nasal and metallic robot-like voice with which it is often associated. Although text-to-speech technology may not yet have reached the quality level at which it can be used to read very long texts, it is definitely effective enough for use in dynamic and interactive presentations.

Whether you are tired of recording the same sentence for the umpteenth time because you keep stumbling over the same tongue twister, are generally feeling unsure about your diction, or simply want to introduce some variety in your presentations, if your budget does not allow hiring professional voice actors, instructors should look into should use text-to-speech, says **A/P Stéphane Bressan** from the School of Computing, National University of Singapore.

Text-to-speech narrations are set to overtake professional voice narrations for use in eLearning presentations. With the variety of many good quality voices available for different genders, accents and languages, one can easily author presentations with professional sounding narration. Simulating multiple commentators and discussions with such presentations is also easily achievable with this technology. In this session, A/P Bressan highlighted on why and how he uses Text-to-speech in his presentations, and showcased on how he integrates it into PowerPoint and Breeze presentations.

Setting the framework

A/P Bressan started the session with an example comparing the quality of various TTS voices from early days to the most recent and explained how technology has moved away from using narrations that sound robotic to the more professional voices. He also showcased other examples to illustrate the possibilities of using TTS to engage student learners. This helped set the framework on what could be achieved using the TTS technology. For instance, how it can be used to create a take away list of vocabulary and be made available to students as podcasts, iPods, MP3 audio, enabling students to listen to such vocabulary them while on the move.

It is not just about creating these TTS voices but more importantly on how these can be integrated into the presentations. Therefore, it is also important to see how these voices can be integrated and combined to be used in PowerPoint presentations. When integrating TTS with PowerPoint presentations, there is an issue of coordination as it is not easy to have the sound and the display be synchronized. In an ideal situation, as the text is being typed, it would be good to sync the audio accordingly. He explained on how this would be difficult to achieve as a lot would invariably depend on the hardware used. Hence, A/P Bressan explored using TTS in other applications like the Adobe Presenter (Breeze).

Text-to-speech (TTS) technology

TTS or Text-to-speech system is the name given to a system that can synthesize speech from text, those that can read text in the natural language and translate it to a sound in a wav file or mp3 file. The technology that is being used in TTS is SAPI (Speech Application Programming Interface). SAPI is the Microsoft standard interface to both speech recognition and speech synthesis applications. Since 2000 computers using Windows have SAPI 5, and this interface is built into the windows OS for all voice applications – voice recognition or speech synthesis. Text-to-speech tools on Windows also use SAPI to integrate with other software such as Microsoft Office like PowerPoint.

A/P Bressan also pointed how the technology is sophisticated enough to determine the pronunciation of specific based on the part of speech of each word in the text. In particular, the system is able to determine parts of speech for each word based on the context and can expand known abbreviations. Again, he illustrated how the technology was able resolve ambiguities with ease by using example.

Examples of TTS systems

A TTS system comes in two parts – The software to generate the sound files from the text and importantly, the voices. Listed below are some of the available TTS systems:

- TextAloud 3.0 (www.nextup.com)
- CoolSpeech 5.0 (<http://www.bytecool.com>)
- TextSound 2.0 (<http://www.bytecool.com>)
- VoiceText (<http://www.neospeech.com>)
- 2nd Speech Center (<http://www.zero2000.com>)

However, A/P Bressan's choice is TextAloud, as it offers a simple and easy interface. It supports both SAPI4 and SAPI5 voices and allows for changing pitch, tone, volume, rate speed, and emphasis in a voice. The application exports sounds to mp3 and wav file formats. It also offers the possibility to change voice within a text and to develop your own vocabulary particularly when special characters and acronyms are used. Interestingly the software also allows for batch conversion – the capability to create voice files for the corresponding text files compiled in a folder.

The interface is simple with a window for entering text along with 2 buttons to Play and Save. Text from a PDF document can be imported over to the window, and then the application converts the text to speech. To make changes to the reading, you can use XML.

XML TTS could be used to provide instructions to the TTS engine in order to control what is happening and uses tags to provide such information to the system. Based on the context, the XML will be able to indicate to the program on how to speak (e.g., how to pronounce 'record' as a noun, verb etc.).

It supports a variety of tags – to control the state of the current voice like Volume, Rate, Pitch, Emphasis, and Spell; to insert special items like Silence and Pronunciation directly; to provide context to the voice; and to provide variations to the language. For example when special acronyms or names need to be used, then special instructions in the form XML tags can be used. It would be painful to learn to write these tags, especially for non-programmers. However the application interface allows for easy insertion of tags using the menu bar.

Regardless of the software tool that is used for conversion, it is important to have good voices and therefore investing in voice packages (which is not free) is necessary. The free voices from Microsoft are the typical robot-like monotone voices, but the voices available from companies like AT&T, Cepstral, NeoSpeech, Acapela and others are of premium quality.

Pedagogical advantages that TTS technology offers

How can it be used?

- One simple way is to use Insert sounds into the PowerPoint slides.
- Another way is to use to record voice narrations for eLearning lessons, for e.g., in Adobe presenter / Breeze where “import audio” feature from the Presenter Tab can be used to easily embed audio narrations on a slide-by-slide basis.

How A/P Bressan uses the TTS technology in his lectures:

1. Idea is to have a mixed voice lecture, as very often you, as an instructor, wonder why you would need to read the slides aloud. It is obvious that students can read the slides much faster than the instructor. However, for courses that use symbols, the symbols would need to be read out aloud to students. Therefore, it would be nice if someone could read the slides for you so as to wake the students up, and help them to follow, and follow along with the slides.

And particularly when it involves long texts that need to be read, then it becomes rather boring for the students. As students are not very familiar with symbols and languages when they are new to such concepts, it becomes imperative for instructors to read these out to the students. Using TTS voice narrations instead of the instructor reading it aloud, captures the student interest very easily and also breaks the monotony while also giving the instructor the break – thus making it advantageous to both the instructor and the student.

2. Having two voices promotes the idea of discussion – sort of triggers the argumentative part of the mind. For example, when he teaches database he needs to explain to his students on how to translate request queries into a machine language for the computer (e.g., SQL). Based on the question displayed to the students during the presentation, instead of just displaying the program snippets to students, it makes it interesting to use TTS voices to read out the program snippet in class – capturing students’ interest and attention.
3. Change of voice means of change of attention on the students, and implicitly signals to the student that there is a change in the mode of the lecture or a change in emphasis. Voices can be changed for different kind of notations. Thus a series of voices or two voices can be used to read theoretical/conceptual aspect and technical aspect of the lecture. His earlier Breeze presentations had 2 voices: one reading the theories another reading the technical aspect. Different voices can also be used for different notations – choosing from a series of voices.
4. Ethnic voices can be introduced to bring in diversity and to keep the attention of students.

Other Uses/Advantages

- Keeps narrated presentations continuously up-to-date (it's too time consuming/expensive to re-record human narration). Therefore, using TTS makes it easy to keep the material current and accurate as it is easy to re-record.
- Prepare scripts, particularly the research papers and articles. Usually authors read aloud what has been written to check if it sounds okay, and the same concept would work here. You could also advise your students to practice the same when working on written assignments.
- Provide audible feedback for student work.
- Use to proofread for helping the user catch typing errors (e.g., in Excel) missed by the usual proofreading.
- Easy-to-listen news: News information can be converted to audio allowing you to listen to the news from your favourite newspaper or magazine.

Summary of Feedback/ Suggestions from the Discussion

The participants felt that the session provided a good overview of the existing TTS technologies. They considered it useful for eLearning, enhancing presentations during lectures, presenting materials more clearly and in particular useful to put across technical concepts. There were others who felt that this technology could be useful for teaching in a foreign language, but not sure how useful it would be in a native language.

Listed below are some questions from the subsequent Q & A session.

Q: What are other possible applications for using TTS in teaching?

- EM:**
- Creating podcasts making using of the TTS voice recordings will be useful for short phrases, definitions, short lessons, FAQs, or for situations that can be easily explained vocally.
 - I have also used it when I had visitors from New Zealand and had some of their sessions recorded. Then we could introduce questions and then combine with the audio recordings of the visitors.
 - Could be used for providing technical definitions, pre-laboratory demonstrations/instructions or extra content.

Q: How is it useful as a podcast?

- EM:** Podcast is particularly powerful for providing take away points for students so that they can be used by students during revision. Podcasts integrates very well into the mobile applications. However, he pointed out that when students listen to such podcasts using earphones, there tends to be noise that makes it difficult to listen for a longer time.

One of the participants offered a solution to the problem indicating that light background music combined with these voice recordings would make it better. A/P Bressan indicated that Audacity can be used to edit and prepare such audio recordings.

Q: Are there any unusual applications of this technology where students are using it?

EM: In one of my classes, I get my students to do video presentations as a part of the assessment. Students (non-native English speakers) who are shy to use their own voices use the TTS technology rather than recording their own voices for the videos.

For developing mobile applications for iPad, iPhone or other smartphones, this technology integrates very easily and well into Flash. It can be used with animation applications (e.g., ToonBoom) where it can allow for full animation along with character mouth movements.

Q: Is it possible for the system to learn from my voice?

EM: Technically it is possible to create a voice similar to yours but this currently requires huge efforts by a team of specialists and is prohibitive.