accessible IT

Choosing and Using Text-to-Speech Software

TTS SOFTWARE ISN'T PERFECT YET, BUT ITS TECHNOLOGY AND VOICES CONTINUE TO IMPROVE RAPIDLY.

by TOM PETERS and LORI BELL

Imagine choosing the voice you'd like to read aloud to you the morning news or a book as you drive to work. Imagine not being able to tell if it is a natural human voice or a synthetic, computer-generated voice. Believe it or not, the technology is available for you to have your choice of voice. It used to be that synthetic voices sounded like computers and were difficult to understand and annoying to listen to for extended periods of time. Now, there are a variety of high-quality voices, female and male, with different accents and pitches and speeds to choose from. Text-to-speech (TTS) software is ready for widespread use by libraries, other organizations, and individual users.

The Voder or Voice coder, the first electronic speech synthesis machine, was developed initially in the late 1920s by Bell Labs and demonstrated to the public at the New York World's Fair and at the San Francisco Golden Gate Exhibition in 1939. If we were to hear these voices today, perhaps we would chuckle. It would be difficult to understand what was being said. However, electronic speech synthesis opened up a whole world of literature to individuals with little or no sight.

Ray Kurzweil was the principal developer on the first print-to-reading machine for the blind. (He was also the principal developer for the first CCD flat-bed scanner, the first text-to-speech synthesizer, the first music synthesizer capable of re-creating the grand piano and other orchestral instruments, and the first commercially marketed large-vocabulary speech recognition application.) The first prototype for the Kurzweil reading machine was completed in 1975. Original units cost $30,000–$50,000. Although the price has dropped considerably since then, with stretched budgets and a small percentage of people who have need for this, most libraries cannot justify purchasing a Kurzweil machine.

What Text-to-Speech Can Do

TTS software offers the affordable ability to turn just about any electronic text that is not image-based into an artificially spoken communication. TTS can be used to create an audible substitute for—or complement to—visual reading. TTS software forms the basis of screen reader software that greatly improves the accessibility of electronic information for people who are blind or who experience temporary or permanent low vision. Interest in the use of TTS software is increasing among libraries and library patrons. If your library has not yet investigated and/or implemented TTS solutions, now is a great time to do so.

TTS software can be beneficial to more than just the blind and low-vision members

of the population that your library serves. Children and adults who are learning to read for the first time can benefit from hearing a book or passage read aloud, either in conjunction with visual reading or as a preparation or reinforcement for visual reading. Reluctant readers of all ages also can benefit from the ability to toggle TTS software on and off on demand to meet their particular needs. People learning a second language also can benefit, especially if the software is sufficiently sophisticated to provide a good vocalization of the sound of the second language when spoken, perhaps even in various dialects.

TTS software also can be used to make image-based information more accessible and informative to all. For example, the Illinois State Library has provided LSTA funds to the Alliance Library System to help teams of librarians involved in digital imaging projects across Illinois to develop and deploy TTS-based audio descriptions. The digital imaging teams are writing brief audio descriptions of selected images they have digitized. Then they'll use TTS software from NeoSpeech to create computer-generated audio renditions of these descriptions. The result will be an audio file that's smaller than a human-generated audio narration. It's also less expensive and faster to produce. In addition to having the audio available on the library's Web site and in ContentDM, we will have a central best practices site with the 10 to 12 images selected from each library's digital image collection with a link to the TTS-generated audio file.

If you are doing a large project that requires voice, using TTS might be more efficient than recording a human voice. For instance, we worked on an audio description project as part of a digital archives project (www.illinoisalive.info) where we created descriptions of historical photos and then recorded someone reading them. This was very time-intensive and involved two people, one to read and the other to record. These natural voice recordings were then converted to MP3 files and put on the Web. Human-narrated audio descriptions require a lot of disk space, because the files are large. Even over a broadband connection, visitors to the archive may not be willing to wait for these large files to download.

TTS software can prove useful in several different computing contexts:

1. As separate, stand-alone software, TTS can be used as production software to create synthetic audio versions of textual information. These audio versions can then be interwoven into a Web site, digital repository, online exhibit, or other digital resources. The software also can be used by library staff and patrons in-house to create on-demand audio renditions of many types of electronic documents.

2. TTS software also has been successfully embedded into a wide variety of applications. For example, for years the Microsoft Reader software, used primarily to read electronic books, has had an embedded TTS functionality that can highlight the visual text as it is being read aloud. Another example is the new version 5 of tcConference Web conferencing software from Talking Communities. The TTS functionality can be toggled on or off for various textual pieces of information that commonly occur during online events. For example, public text chat can be vocalized, as can private text chat. This obviously is beneficial to blind and low-vision participants in the online event, but it also could be beneficial to anyone who wants or needs to minimize the Web conferencing software in order to work on other tasks while attending an online event. Even with the online room minimized, if the user has TTS turned on for both public and private text chat, the user will hear not only the Voice-over-IP (VoIP) communication, but also all of the text chatting. Be forewarned, however, that sometimes the natural VoIP voice communication and the synthetic TTS-generated audible chat are delivered to the listener at the same time.

TTS software also can be embedded in operating systems (OS). The Wikipedia article on speech synthesis (http://en.wikipedia.org/wiki/Text-to-speech) credits Apple's 1984 MacInTalk system as the first TTS in an OS, followed closely by the AmigaOS in 1985. Although Microsoft includes several default TTS voices, such as Mary and Sam, in the Windows operating system, they sound very artificial. Microsoft's new Vista OS may contain the most advanced TTS system ever incorporated in an operating system for personal computers.

What to Consider Before You Invest In TTS Software

When exploring and testing available stand-alone TTS software packages, there are several factors to consider:

Voice quality: Some voices sound very artificial, but some of the newer software systems have incredibly natural-sounding artificial voices. When we did a read-aloud test project with PDF ebooks for the visually impaired, the default voices at the time were the Windows-provided Sam and Mary. Almost every tester preferred listening to the book in his or her own screen reader's voice over listening to Sam or Mary. Some visually impaired people who do a great deal of work at their computers spend the extra money to purchase a higher quality voice available from AT&T and other vendors.

Voice options: At the very least, the TTS software package you purchase or lease should have a male and female voice in the language or languages you need.

Custom pronunciation: You'd want the ability to fine-tune how the software

»

pronounces certain words or acronyms. For example, if you live in a part of the country where the word "creek" is pronounced with a long "e" rather than with a short "i," you will want to be able to customize the TTS pronunciation of the word to follow regional practices.

File output: If you are going to use the TTS software primarily as production software to create synthetic audio files, you will want to know what file format it puts out. If the file format you want or need to deliver is different than the one output by the software, you'll need to add a file conversion step to your production process. Also consider the size of file output. A general rule of thumb is the better the quality, the bigger the file. So weigh the quality of audio you want against the storage medium you have.

Please note: Intellectual property rights surrounding TTS-generated audio "performances" of copyright protected works remain under debate. In fact, there is debate about whether a TTS-generated audio file is indeed a performance. Libraries and library patrons should proceed with caution.

Size of the application files: Some vendors offer their software in different "footprints" so it can be loaded on a variety of computing devices from handhelds to servers.

Cost: High-quality synthetic voices are expensive. When you first listen to voices, you may think you would like the option of purchasing several to provide some variety. If you can afford it that is fine, but it is worth the time to listen to a variety of voices to determine what you like the best. Some vendors offer unlimited licenses for an unlimited time. You may pay more, but over time the cost is much lower than if you had to pay a license fee every year. Don't buy more than you think you will use.

Annual lease or maintenance fees: Sometimes it is less expensive in the long run to pay a little more up front for a lifetime license instead of paying each year. On the other hand, if you have any doubts about the voice you license, it makes sense to lease for 1 year.

Terms and conditions of use: Make sure you know your licensing options. If you create an audio file with your newsletter, can you purchase one license and put it on your Web site or does each person using it have to purchase a license?

Highlights of Some Products

There are several vendors of standalone TTS software programs. We can only describe a few briefly in this column.

TextAloud is a reasonably priced, versatile TTS system. For under $50 you can buy the software that converts text to speech and can convert the speech to an MP3 file that you can put on a small playback device. The text is inserted into the program, the user selects the voice he or she would like to listen to, and then the text is read aloud.

For an additional fee you can purchase the voice of your choice. Many new male and female voices are becoming available, including different nationalities. For a short time, the Mid-Illinois Talking Book Center used TextAloud to create MP3 files to put on its Web site as a digital audio version of its newsletter. The amount of time it takes to convert text to speech to an MP3 file is amazingly short and works very well for visually impaired readers to listen to documents of many types. TextAloud also offers News Aloud (which allows a person to listen to a custom newscast on his or her computer) as well as Stocks Aloud.

You must pay separately for a license for each voice. TextAloud has a large variety of voices to select from because it works with a number of TTS vendors such as AT&T Natural Voices, NeoSpeech voices, Cepstral Voices, Scansoft RealSpeak Voices, and Acapelo Voices. For a small TTS project, or for an individual, TextAloud costs are very reasonable. It can become pricey, however, to license the software for use on more than a few computers.

Loquendo is a TTS vendor that allows voice editors to use Voice XML to produce intuitive automated speech. Loquendo offers embedded text-to-speech and can be used in PDAs, tablet PCs, and other devices. If the population your library serves is particularly multilingual and Euro-centric, Loquendo offers male and female voices for British and American English, Dutch, French, German, Greek, Italian, Polish, Portuguese (both Brazil and Portugal), Spanish (both Spain and the Americas), Catalan, Swedish, and Chinese (the only Asian voice).

NeoSpeech is another TTS vendor. NeoSpeech voices are used by a wide variety of businesses (for their telephone systems), organizations, and individuals. NeoSpeech also is working with companies that provide massively multiuser virtual environments and online games, so participants can use synthetic audible voice instead of visual text chat to communicate with others. Oddcast, a company that produces bots (a picture of someone on your Web page who can communicate with users), is working with NeoSpeech to transform the bots' conversation from text to audio. NeoSpeech has a variety of high-quality voices, not only for American English but also for Chinese, Japanese, and Korean. NeoSpeech will also help a group develop a unique voice.

Cepstral offers synthetic voices in U.S. and U.K. English, Canadian French, Italian, German, and American Spanish. Cepstral seems to be especially strong in offering its TTS software on a variety of operating systems, including Mac OS X, MS Windows, i386-Linux, x86-64-Linux, Sparc-Solaris, and i386-Solaris.

AudioEye: There is a relatively new vendor on the market called AudioEye: Surf by Sound. AudioEye is simply a

Resources Discussed

TextAloud
www.textaloud.com

AudioEye
www.audioeye.com

NeoSpeech
www.neospeech.com/demo/demo_text.php

Loquendo
www.loquendo.com

Cepstral
www.cepstral.com

Microsoft Windows Vista
www.microsoft.com/enable/products/windowsvista

navigational audio tool that sits on top of the text-based visual version of a Web site. A Web designer could use AudioEye to convert the Web site through a small software download that plays a tone when visitors come to the site. This gives them the option of navigating the Web site via audio instead of the traditional text-based version.

People can use AudioEye in two ways. If you select the audio option for a site, the software automatically converts the text to audio for you. Or you can create MP3 files for greetings, information about a program, etc. This not only creates an accessible Web site but makes it more interactive for users.

Until products like AudioEye entered the market, it was the responsibility of each user to purchase screen reading software (which is sometimes quite costly) to retrieve information from the Web.

Challenges for TTS Software

Although the technology and the voices are rapidly improving, some text-to-speech software programs do not read a certain word or a phrase correctly, making it difficult to understand. Acronyms, abbreviations, numbers, and non-alphanumeric characters can be challenging for TTS software too. However, vendors continue to make improvements in these areas.

Advice for Librarians Considering Text-to-Speech

If you buy TTS software, will anyone use it? Some of these software packages are expensive, and although it is nice to offer them, you need to be sure it will be worth your investment. You may want to conduct a community survey to see if people would come to the library to use text-to-speech software. Also, do you have a public access computer to dedicate to this kind of use? Is the hardware strong enough to handle the demands of the software?

Making your staff members and patrons aware of the availability and usefulness of TTS is a key to reaching broad acceptance and usage. If you do purchase TTS software, be prepared to do a lot of marketing and to invest in staff time to help people learn to use it. Like any other software, it's important to train staff members on it before offering it to the public.

The TTS software designers are making great strides, and prices are dropping. There are many vendors and options available that fit libraries' budgets, and there are many potential ways to use TTS software in digital information systems that will improve the accessibility and usability of the information for everyone.

Lori Bell is director of innovation at the Alliance Library System in East Peoria, Ill., where she writes grants and coordinates special projects. Her email address is lbell927@yahoo.com. Tom Peters is the founder of TAP Information Services in Blue Springs, Mo., which provides project development, management, and evaluation services to libraries and other information-intensive organizations. His email address is tpeters@tapinformation.com.