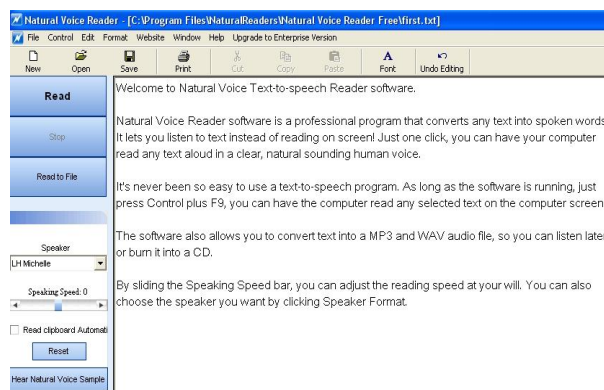


Title: **Universal Access using Text-to-Speech**
 Author: Gerry Kennedy © February 2009
 Software: Text-to-Speech Programs
 Category: Access to text using synthesized speech output



1. Introduction

Text-to-Speech (TTS) has been much maligned and misunderstood. This enabling and liberating technology has been available since the advent of personal computers. Numerous TTS programs offer features and have capabilities that provide access to electronic text, scanned text (using OCR), onscreen menus, dialogue boxes (screen reading such as *JAWS* or *Thunder*) as well as support in word processing software. Some programs will speak the contents of the clipboard (*Cliptalk* or *Deskbot*), others require text to be copied then pasted into a window (*ReadPlease 2003*, *Helpread* or *DSpeech*), whilst others run and function as a floating (e.g. *Natural Reader*) or in a specific toolbar within an application (e.g. *Wordtalk*).

Literacy support tools such as *textHELP Read & Write* and *ClaroRead for PC 2008* have multiple TTS features that provide universal support across all computers applications and tasks. Programs designed for younger students include graphic and symbol supports (e.g. *Clicker 5*, *Textease*, and *SymWriter*) combined with multi-media support and tailored customisable toolbars. Yet others also include multiple access technologies including switch and external programmable keyboard input (e.g. *IntelliTools Classroom Suite*).

2. Background to Text-to-Speech Technology

‘Speech synthesis is the artificial production of human speech. A computer system used for this purpose is called a speech synthesizer, and can be implemented in software or hardware. A text-to-speech system converts normal language text into speech; other systems render symbolic linguistic representations like phonetic transcriptions into speech.’

Synthesized speech can be created by concatenating pieces of recorded speech that are stored in a database. Systems differ in the size of the stored speech units; a system that stores phones or diaphones provides the largest output range, but may lack clarity. For specific usage domains, the storage of entire words or sentences allows for high-quality output. Alternatively, a synthesizer can incorporate a model of the vocal tract and other human voice characteristics to create a completely "synthetic" voice output. The quality of a speech synthesizer is judged by its similarity to the human voice and by its ability to be understood.’

How it All Works

‘A text-to-speech system (or "engine") is composed of four parts: a front-end and a back-end. The front-end has two major tasks. First, it converts raw text containing symbols like numbers and abbreviations into the equivalent of written-out words. This process is often called text normalization, pre-processing, or tokenization. The front-end then assigns phonetic transcriptions to each word, and divides and marks the text into prosodic units, like phrases, clauses, and sentences. The process of assigning phonetic transcriptions to words is called text-to-phoneme or grapheme-to-phoneme conversion. Phonetic transcriptions and prosody information together make up the symbolic linguistic representation that is output by the front-end. The back-end, often referred to as the synthesizer, then converts the symbolic linguistic representation into sound.’
 [Source: Wikipedia - http://en.wikipedia.org/wiki/Speech_synthesis]

Text-to-Speech (TTS) software has existed ever since the early micro computers appeared in homes and schools in the early 1980's. Special voice cards or external synthesizers (e.g. *Echo II* on *Apple II* computers) provided robotic, synthesized male voices. Programs that offered voice feedback often just used the small low-quality computer speaker to speak aloud the key words, commands and phrases. The voices were approximated and of low quality and volume – but critically provided access to computers for blind and vision impaired users. Choice was limited, yet users who required voiced instructions,

directions and feedback were grateful for the pioneers of early TTS technologies. Educational software, quite elementary in design compared to today's sophisticated software, was basically text based with 16 bit or 64 bit colour graphics.

As computers became more powerful and processors increased in speed and power, software companies added higher quality voice and TTS to programs. Often voice was a human voice speaking menu items, dialogue box instructions or explaining the program's navigation. These small amounts of voice took up valuable program storage space and so were limited due to the prevailing 5.25" and 3.5" disk technology, especially pre hard drive configurations in PCs.

Early TTS word processing and text editing software (e.g. *Dr Pete's TalkWriter* and *IntelliTalk V1*) began to appear along with small utility applications (e.g. *MacTalk* on Macintosh computers). *MacTalk* was a revelation as the speech quality was more acceptable and the range and variety of voices were more advanced than program on other platforms (e.g. BBC, Amiga, Atari and early IBM DOS). Funny or amusing voices such as parrots, robots 'silly' voices were fun for younger children and eventually some female voices began to appear.

The TTS programmers designing and improving this technology were usually working in the southern states of America. The voices had an American sound, were still quite robotic, nasally and had an 'accent'. Some educators found this challenging, yet students of all ages worked with the technology as it provided opportunities to interact with software, listen to instructions, word process with auditory feedback and for blind and vision impaired students, provide access to computers!

3. Implications for Users who Struggle with Text

The written word on paper or on a screen can liberate or disadvantage a student. Students, who have had poor or minimal experience with a language such as English, often fail in reading, writing, comprehension - or all three. School attendance may suffer, students rebel or simply give up. The number of students who disguise their inability to read and write is quite common and prevalent. They continue to progress through school with significant areas of need. Learners of all ages use other strategies in order to stay 'under the radar'. It is not due to a lack of intelligence but more of a lack of aptitude due to continued failure. If the curriculum being offered does not cater to students with different learning styles, then content and delivery may be regarded as irrelevant. The disengaged student may fall further behind and ultimately leave school.

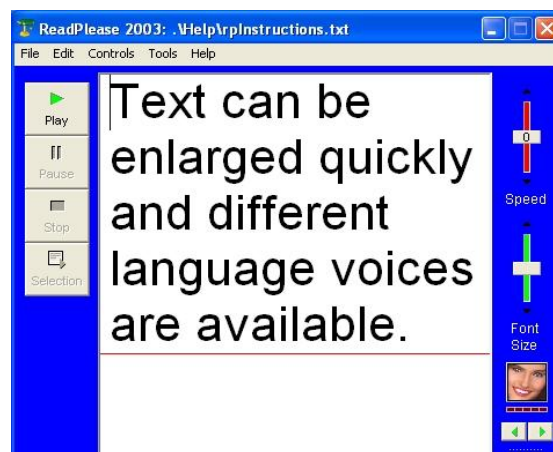
Reluctance to type a thought, phrase or sentence, often results in continued failure in other subjects and further disenfranchisement when working with text across the curriculum. Some users have minimal or poor spelling skills, limited vocabulary, and limited command of a language (e.g. new arrivals and students in ESL) or are print disabled. Dyslexia poses a raft of other problems.

Reading and writing text can be perceived as being the enemy and something to avoid at all costs. This can result in limited advancement in school, reduces the likelihood of engagement and causes behaviours that are unacceptable or damaging to students and peers.

4. Benefits of TTS for Students

TTS technology allows users to work the sense of hearing. Each individual student has different needs and abilities. Multiple approaches to the creation of text are therefore required. An important adjunct to accessing text is in listening to prose or poetry created by the user or text sourced from other sources. These sources may include:

- ✓ Newspapers, journals and magazines
- ✓ Books, maps and other reference materials (scanned as electronic text)
- ✓ MS Word or word processing documents, recipes, instructions, fact sheets or directions
- ✓ Encyclopaedias and atlases online, or on CD or DVD



- ✓ Emails and online chat
- ✓ FaceBook, MySpace, Bebo and other social networking sites
- ✓ Other web content on any topic or theme
- ✓ Notes or text files
- ✓ In fact, *any text* that can be selected or highlighted

Students can listen to their own work so as to proof read and check for errors. Editing using TTS allows users to identify and amend various mistakes. Students can listen to errors and check for:

- ✓ Missing words
- ✓ Added words or repeated unnecessary words (e.g. 'the the')
- ✓ Misspelt words that don't "sound right"
- ✓ Overly long sentences
- ✓ Sentences that contain too many conjunctions
- ✓ Short sentences
- ✓ Sentences that do not adequately convey meaning
- ✓ Clumsy or ill constructed thoughts and ideas
- ✓ Documents that require additional formatting (e.g. paragraphs)
- ✓ Words that are voiced in unexpected ways
- ✓ Words that inappropriate or used in an incorrect context

When used for reading third-party text, students may enjoy:

- ✓ Listening to new text with a male or female voice
- ✓ Slowing the voice or speeding it up
- ✓ Listening with highlighted text
- ✓ Listening with a choice of different text and background colour combinations
- ✓ Listening through headphones (for private writing or text that is not for public consumption)
- ✓ Re-reading passages for study purposes

Text is used for many purposes. If students find reading boring, tedious, arduous or difficult, then TTS programs provide a degree of independence. Students need not rely wholly on family members, peers or educators to read or clarify part of or whole documents. They can begin to become more self-sufficient. Increased confidence in creating and reading text invites additional exposure and access to text and the enjoyment of various genres.

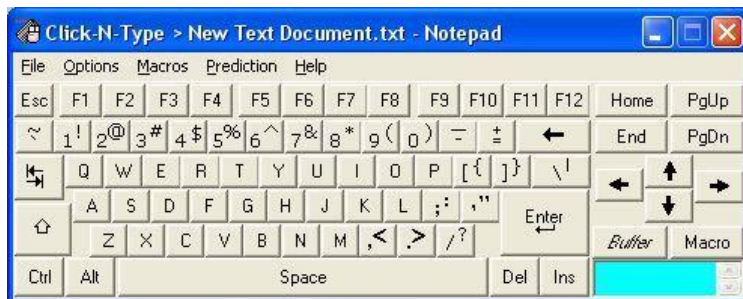
Listening to text can be accomplished on a computer, or using more portable devices such as mobile phones, iPods and other MP3 players. Text can be typed by the student or accessed as:

- ✓ eBooks (online or downloaded from sites such as Gutenberg (www.gutenberg.org))
- ✓ Online stories, Blogs and Wikis
- ✓ Twitter feeds and other Web 2.0 content
- ✓ RSS feeds
- ✓ PDF and MS Word files downloaded as study guides, instructions or course content
- ✓ Documents and files in HTML from web sites or course material on CD or DVD
- ✓ Recipes and procedures
- ✓ Maps and directions (e.g. Google Earth)
- ✓ Directories and information (e.g. White and Yellow Pages)

Many more web resources are also available and all of this is made possible for users of all ages. Elderly citizens who are losing their sight or who find reading a computer screen fatiguing and tiring enjoy reading and listening to the London Times, a race form guide or bowls newsletter. Young children can participate in contests, in accessing and learning new games and navigating through unfamiliar content.

A Means of Communication

In some instances, TTS can also be used for some students as a means of communication. They may have quite approximated speech, be non-verbal or choose not to speak. The computer program can voice their typed text as they create words and sentences “on the fly”. They may be fast proficient typists or they may use word prediction software (e.g. *Click ‘n’ Type* - <http://www.lakefolks.org/cnt/> or *Co:Writer*) to develop text more quickly.



Using abbreviation and expansion (e.g. in *MS Word*) they can type more quickly and conduct a conversation with the computer’s male or female voice as their own. This may prove to be a good short-term strategy or it may work well whilst the student is using a computer and not a dedicated communication device. It depends on a number of factors and the voice being used would have to be sufficiently clear for others, so that the text could be clearly heard and words distinguished and identified easily.

5. Voicing Text using Different File Formats

The most commonly accessed file formats that students will work with text include:

Text	Pure text (without graphics or tables etc) that can be opened in Notepad or WordPad or other text editing programs
MS Word	Edited in versions of <i>MS Word</i> , <i>MS Works</i> , <i>Open Office Org</i> or other word processing programs
PDF	This can be a useful format but has some restrictions and impediments in <i>Adobe Reader</i>
HTML	Ready to edit in a variety of programs (<i>MS Front Page</i>) and FTP’d to web sites and Intranets
Proprietary Formats	As used in editing, document layout and desktop publishing programs or art and design software (e.g. <i>MS Publisher</i>)
Daisy Reader	A popular format for reading books <ul style="list-style-type: none"> • Amis Daisy Reader (http://daisy.org/projects/amis/) or • Read:Outloud Bookshare Edition www.bookshare.org/readingTools#downloadROLF • Victor Reader Soft Bookshare Edition- www.bookshare.org/readingTools#downloadVRSOFT

6. Commercial and Free Software for Text-to-Speech

In this genre of software, there is a great deal of choice. Generic programs cater to all age groups and abilities. Other software utilities and programs suites cater more specifically to younger students in pre-school, primary or special settings. More comprehensive and expansive programs deliver multiple options and features and therefore suit older or more capable users. The following categories include software that will predominantly cater to and accommodate users who have different learning, access or communication needs.

Commercial Software with Text-to-Speech – Younger Students:

Clicker 5 ANZ	www.cricksoft.com (uses onscreen grids and templates with graphics & photos)
Textease	www.textease.com (“click and write” anywhere WP - part of Textease Studio CT)
IntelliTalk	www.intellitools.com (talking WP - part of the IntelliTools Classroom Suite V4)
Write:Outloud V3	www.donjohntson.com (talking WP - part of the SOLO suite of programs)
Max’s Toolbox	http://www.maxstoolbox.com/products/maxwrite - works with MS Office
Draftbuilder	www.donjohntson.com – predominantly text based organisational and planning with templates for structured writing scaffolding and supports

Kidspiration

www.inspiration.com – mind mapping using multiple templates and webs. Ideal for planning and organisation with extensive graphic library, thesaurus and TTS

Writing with Symbols 2000 V2.6

<http://www.widgit.com/> (older program but still used widely in special schools)

Communicate: SymWriter

www.widgit.com (updated version of WWS with additional features & PCS symbols)

Communicate: Webwide

<http://www.widgit.com/products/communicate/index.htm> - symbol based web browser with speech

Boardmaker 6 and Boardmaker Plus

<http://www.mayer-johnson.com/> – a program that has DTP and WP functions with over 9000 PCS symbols



Commercial Software with Text-to-Speech – Senior Students and Adults:

textHELP Read & Write V9;

www.texthelp.comn or www.spectronicsinoz.com

ClaroRead for PC 2008

<http://www.clarosoftware.com> or www.spectronicsinoz.com

Kurzweil 3000

<http://www.kurzweiledu.com>

Wynn Wizard from Scientific software

<http://www.freedomscientific.com/LSG/products/WYNN.asp> or locally from <http://www.quantech.com.au/wynn>)

BrowseAloud

http://www.browsealoud.com/page.asp?pg_id=80004 – will voice web sites that are BrowseAloud friendly. It works with MS Internet Explorer and FireFox.

WordQ

<http://www.wordq.com/> - it has natural sounding text-to-speech, in-context prediction for corrections with usage examples for confusing words and the predictions are based on creative spelling

These software programs all provide quick and elegant access to print materials together with a number of different learning supports, potentially including spell checking, word prediction, thesaurus and dictionaries, text to speech with human sounding voices, text-to-audio conversion, other visual and auditory features and organisational and planning tools.

Freeware- Text to Speech programs and Utilities:

Natural Reader V7

www.naturalreaders.com – is a very useful free program that will use SAPI or SAPI 5 voices. It runs as a floating toolbar as well as full screen. Will voice MS Internet Explorer.

ReadPlease 2003

www.readplease.com – It runs in a small window but has multiple language support, SAPI 4 voices and easy to use font resize sidebar. *Freeware.*

Deskbot

<http://www.bellcraft.com/deskbot/> - DeskBot is a freeware, multi-featured Clipboard Reader, Text Reader, Time Announcer, Desktop Application featuring Microsoft Agent Animated and a range of talking characters. *Freeware*

Cliptalk

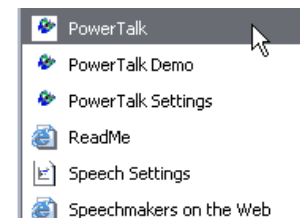
<http://fullmeasure.co.uk/cliptalk/> - automatically speaks text that is copied to the Windows clipboard (e.g. with **Right Click-Copy** or **Ctrl + C**) *Freeware*

TextAloud V2.2

<http://www.nextuptech.com/> - create MP3 or WMA files for use on portable devices like iPods, Pocket PCs, and CD players Uses AT&T, Acapela and/or RealSpeak voices) Free trial, then *Shareware.*

WordTalk V4.2

<http://www.wordtalk.org.uk/Home/> WordTalk is a free text-to-speech plug-in developed for all versions of Microsoft Word (from Word 97 onwards). It will speak the text of the document and will highlight text as it goes. It contains a talking dictionary to help decide which word spelling is most appropriate. Also converts text to audio in MP3 or WAV formats. *Open Source/Freeware*



PowerTalk V1.2.1

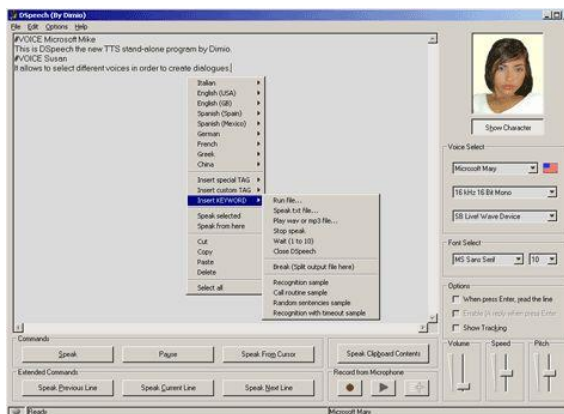
<http://fullmeasure.co.uk/powertalk/> - PowerTalk is a program that automatically speaks any presentation or slide show running in MS PowerPoint for Windows. It uses SAPI 4 or 5 voices. *Freeware*.

DSpeech V1.55

<http://dimio.altervista.org/eng/> - Allows the user to save the output as a .WAV, .MP3 or OGG file and quickly select different voices, even combine them, or juxtapose them in order to create dialogues between different voices.

Many other free programs exist. Users have their favourites as it depends on what the user wishes to achieve. Many companies now sell and distribute voice technologies. The trick is to purchase a program that has the voice(s) that are required and use them in other programs as well. The two popular Australia voices are Lee and Karen. These are packed with most textHELP products and Claro Software programs. These are licensed to these companies for inclusion in their literacy support programs.

The SAPI 4 versions of Sam, Mike and Mary are the most likely the voices that will be included in most of the Freeware TTS programs.



7. Portable TTS Solutions Documents

DSpeech can also be used as a *Portable App*. It comes packaged in the *AccessApps* USB suite. It has many powerful features and can be used in a variety of ways.

TopOCR, is another free portable OCR application in *AccessApps*. It runs from a USB memory stick or thumb drive. Using a camera or mobile phone, users take a photo that has text embedded in the frame. The photo is then sent to a computer (via USB or Bluetooth), then start *AccessApps* and load *TopOCR* software. It is then a matter of opening the photo and *TopOCR* will not only convert the text but also read it out aloud using built in TTS.

Note: *AccessApps* is freely available to download and use: <http://www.rsc-ne-scotland.ac.uk/accessapps/>.

8. Converting Text to Audio File Formats

Text can be converted to a sound format such as MP3 using commercial or free programs (e.g. *TextAloud*, *DSpeech*). MP3 or Audio format files can be created to playback on a computer, Personal Digital Assistant (e.g. Palm or HP device), MP3 player, mobile phone or other music capable device. Students are used to listening to devices and so this becomes a very socially acceptable modality.

Conversion software is required and most of the commercial literacy support tools have this capability built in.

In Conclusion

This article is a brief discussion into some of the issues raised with reference to speech feedback using synthesised speech on computers and other related digital devices. Students and users of all ages and abilities can access and gainfully use this technology. It certainly liberates those people who are print disabled. Students with dyslexia may need TTS and other tools in order to read or create text.

Speech recognition may also be an issue. The combination of these two technologies can afford great advances in access for some individuals. There are always advantages and disadvantages so some research and trial and error may be required.

The quality of voice will need to be considered for some users who are:

- ✓ Deaf
- ✓ Have a short or long term hearing deficit
- ✓ Blind or vision impaired
- ✓ Very young and require high quality voices (for voice modelling)
- ✓ Intellectually impaired and who might struggle with some voices
- ✓ Fussy!

It is a matter of using one or more programs and experimenting. Students of all ages enjoy having text spoken back from emails and web sites as they can listen, read or listen and read at the computer or on a mobile device of their choosing. It promotes increased independence and provides opportunities for taking more risks with text. Reading can be a very negative experience for students. TTS offers a new way to work with data and text. It's worth a try!

Resources:

www.gateway2at.eu/ - Guidance for Assistive Technology in Education

www.tucows.com – a very useful website for locating Freeware, Shareware and Open Source programs

www.oatsoft.org/ - OATSoft is dedicated to improving Assistive Technology and computer accessibility through the power of Open Source development techniques. OATSoft makes the best Open Source Assistive Technology Software (OATS) easy to find. Open Source Software is free and the 'source code' that makes the software is freely available. It is developed by international communities operating on-line. Assistive Technology Software allows people with disabilities to overcome some of the disabling effects of society and technology, including computer and web accessibility.

www.freewarefiles.com/cat_9_105_Text2Speech.html - a list of free text-to-speech programs

www.research.att.com/~ttsweb/tts/demo.php - online demonstration of various AT&T voices in English and other languages

www.texttospeechblog.com/2008/01/axistive-article-on-talking-books.html Useful text-to-speech Blog

