Title: **Universal Access to Text Using Speech Recognition**
Author: Gerry Kennedy © March 2009
Software: Speech Recognition Software
Category: Creating text using voice in Speech Recognition Programs

1. Introduction

One of the most sought after inclusive technology products for students is actually a technology that was primarily designed for other purposes. "One of the most notable domains for the commercial application of speech recognition in the United States has been health care and in particular the work of the medical transcriptionist. According to industry experts, at its inception, speech recognition (SR) was sold as a way to *completely eliminate transcription* rather than make the transcription process more efficient, hence it was not accepted. It was also the case that SR at that time was often technically deficient. Additionally, to be used effectively, it required changes to the ways physicians worked and documented clinical encounters, which many if not all were reluctant to do.

The biggest limitation to speech recognition automating transcription, however, is seen as the software. The nature of narrative dictation is highly interpretive and often requires judgment that may be provided by a real human but not yet by an automated system. Another limitation has been the extensive amount of time required by the user and/or system provider to train the software.

Speech recognition (also known as automatic speech recognition or computer speech recognition) converts spoken words to machine-readable input (for example, to key presses, using the binary code for a string of character codes). The term "voice recognition" is sometimes incorrectly used to refer to speech recognition, when actually referring to speaker recognition, which attempts to identify the person speaking, as opposed to what is being said. Confusingly, journalists and manufacturers of devices that use speech recognition for control commonly use the term Voice Recognition when they mean Speech Recognition."
[Source: http://en.wikipedia.org/wiki/Speech_recognition]

2. Background to Speech Recognition Software

People with a range of different disabilities often benefit from using speech recognition programs. It is especially useful for people who have difficulty with or are unable to use their hands. This may have resulted from physical conditions from birth or an acquired injury or trauma (e.g. a stroke), as well as mild repetitive stress injuries. It can also involve disabilities that require alternative input for support in accessing the computer.

In fact, people who used the keyboard continually over a period of time and developed RSI became an early market for speech recognition. Speech recognition is commonly used in deaf telephony, such as voicemail to text, relay services, and captioned telephone services. Many industries now rely on Voice and/or Speech Recognition for a range of automated tasks, including communications, artificial intelligence and security systems

3. Catering to Disability

Users with learning disabilities who have problems with thought-to-paper communication (they formulate ideas but are processed incorrectly causing them to end up differently on paper) can benefit from the software as it helps to overlap that weakness. There are many students who would rather speak than write or type. Others have more control over the language used or words expressed verbally than written. Some students with poor motor function benefit from being able to speak into a microphone, at their own rate, rather than be restricted by poor spelling, dyslexia, minimal or approximated typing skills or lack of confidence. Users who have autism and other syndromes often find starting and initiating writing difficult if not impossible.

Being able to speak clearly, with some fluency and control, using consistent speech patterns with sufficient volume and clarity will be necessary for some degree of success. Users vary with their performance. There are some technical computer hardware and

sound card, microphone and configuration issues that may need be resolved as well as practise, continued exposure, a knowledge and application of key voice commands and appropriate working environments. Background noise, vibrations, unexpected sounds and other extraneous noise issues may interfere. With noise cancelling microphones, and high quality USB to analogue converters these previous constraints have been solved to a degree.

There are numerous factors that must be taken into account. Not all people are suited to this technology genre ad the software does always deliver the results that users may anticipate. It is not always an "out of the box" solution. There are certain prerequisite skills and understandings as well as age considerations, ability to plan speak coherently and an ability to formulate ideas as well as the ability to read text. Many students rehearse the text to be "learnt" by the software over a period of minutes, hours and sometimes even many sessions over days. Some of these issues can be circumvented or overcome. These important issues must at least be realised, considered and tackled.

Naïve users in the past have given this technology a poor name. Many students and even educators have unsuccessfully tried and failed. Mitigating circumstances were often the major cause, with poor research. Minimal preparation, sub-standard equipment and a lack of understanding and appreciation of how the technology performs in certain conditions and the poor assessment of the candidate's ability to learn to control and access the software all contributed to a frustrating experience.

The software genre has matured over the years with current leading products now fulfilling promises made by vendors quite a few years ago. The speed, performance and memory constraints have largely been resolved and cost is no longer a major issues. Hardware, both desktop and portable computers (namely Notebooks and even Ultra Lights) can be purchased and configured to perform with more than satisfactory results for many users. Versions are available for MAC OS and MS Windows XP and Vista operating systems.

4. Key Issues

There are certainly impediments to successful application of SR. A great deal has been documented around the world, with some excellent material available in Australia. Many agencies provide services and support. People who have moderate to significant needs can find and secure support in each state and territory. Some issues may need to be resolved:

- ➢ The Australian accent has been an issue ion the past – but this has been resolved in newer versions of SR software
- ➢ Schools, training centre and tertiary institution classrooms are often noisy and not conducive to SR
- ➢ The number of different environments impact upon consistency and integrity
- ➢ Poorly configured computers, slow processors and minimal RAM memory issues need to be addressed
- ➢ Placement and use of headset, handheld, lapel or desktop microphone positioning requirements
- ➢ USB based microphones vs the older style analogue 3.5mm jack connectors
- ➢ The sound card technology and capability (separate card or integrated motherboard device)
- ➢ Young boys who train the software, and then experience a change when their 'voice breaks' require SR re-training
- ➢ The ability of the user to use consistent speech patterns
- ➢ The ability of the user to pronounce words consistently
- ➢ Breathing control
- ➢ Issues of fatigue
- ➢ Posture and body/head control
- ➢ Ergonomics, height of screen display, placement of the computer for visual amenity
- ➢ Capacity of the user to speak fluently and/or speak consistently over time
- ➢ The ability to learn, master and control voiced commands ad editing functions
- ➢ The ability and capacity to correct errors as SR is being used
- ➢ Periodically backing up voice files (to a memory stick or external hard drive/server) for data integrity
- ➢ The skills to transfer relevant and critical user data from one software version or computer to the next

4. Performance of speech recognition systems

The performance of speech recognition systems is usually specified in terms of accuracy and speed. Accuracy may be measured in terms of performance accuracy that is usually rated with word error rate (WER), whereas speed is measured with the real time factor. Other measures of accuracy include Single Word Error Rate (SWER) and Command Success Rate (CSR).

Most speech recognition users would tend to agree that dictation machines could achieve very high performance in *controlled* conditions. There is some confusion, however, over the interchange ability of the terms "speech recognition" and "dictation".

Commercially available speaker-dependent dictation systems usually require only a short period of training (sometimes also called `enrolment') and may successfully capture continuous speech with a large vocabulary at normal pace with a very high accuracy. Most commercial companies claim that recognition software can achieve between 98% to 99% accuracy if operated under optimal conditions. Optimal conditions usually assume that users:

- ➢ Have speech characteristics which match the training data
- ➢ Can achieve proper speaker adaptation, and
- ➢ Work in a clean noise environment (e.g. quiet office or laboratory space).

This explains why some users, especially those whose speech is heavily accented, might achieve recognition rates much lower than expected. Speech recognition in video has become a popular search technology used by several video search companies.

Limited vocabulary systems, requiring no training, can recognize a small number of words (for instance, the ten digits) as spoken by most speakers. Such systems are popular for routing incoming phone calls to their destinations in large organizations.
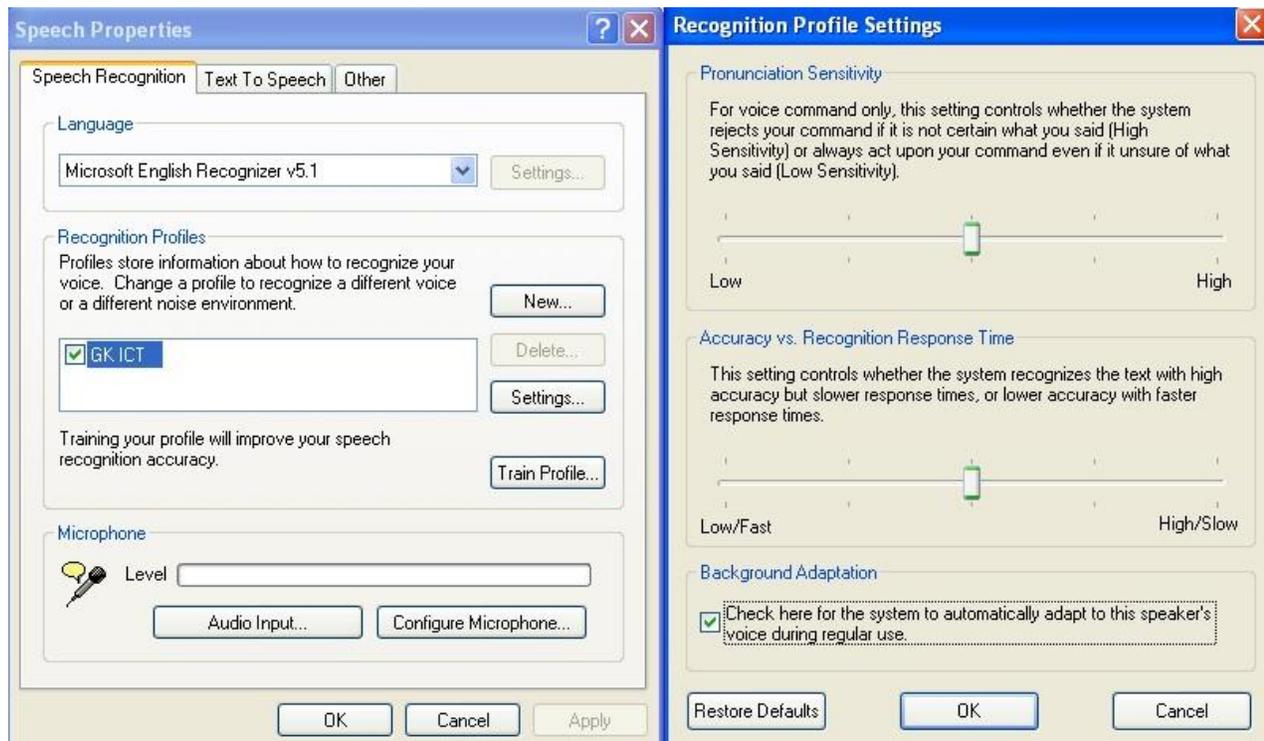

4.

MS Windows Option

Voice recognition has been a feature in both the *Windows XP* and *Vista* versions, yet most people are unaware that it is packaged in the standard operating system. These versions have some merit, and provide some opportunities but are really entry-level

technologies. These inbuilt programs cater for users but are not necessarily a complete or total solution for everyone. They usually do not cater to certain disability groups or users with some learning disabilities. MS Windows based voice recognition is part of the *Language Bar*. The *Speech* option in the *Control Panels* in MS Windows XP, for example is used to access this technology.

Windows Speech Recognition in *Windows Vista* allows users to interact with their computers by voice. It was designed for people who want to significantly limit their use of the mouse and keyboard while maintaining or increasing their overall productivity.
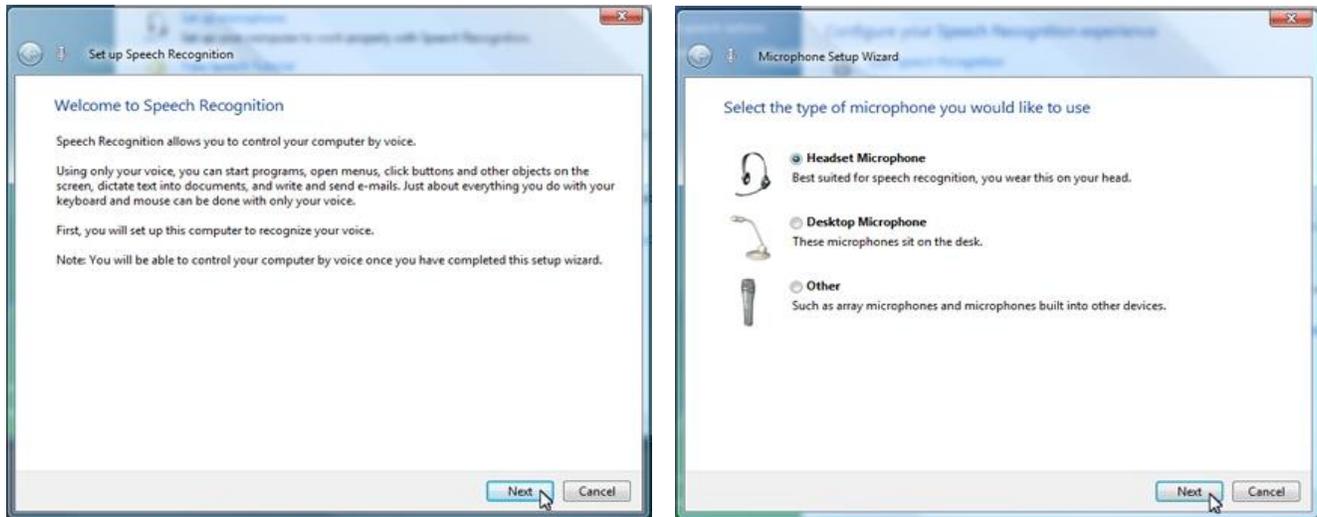
Students and educators alike can dictate documents and emails in mainstream applications use voice commands to start and switch between applications, control the operating system and fill out forms on the Internet.

It is a new feature in *Windows Vista*, built using the latest Microsoft speech technologies and it provides worthwhile recognition accuracy that improves with continued exposure and use as it adapts to the user's speaking style and vocabulary. Speech Recognition is available in English (U.S.), English (U.K.), German (Germany), French (France), Spanish (Spain), Japanese, Chinese (Traditional), and Chinese (Simplified).

The speech recognition capabilities in Windows XP used with Office 2003 productivity software can enhance computing in such areas as gaming, data entry and text editing. Other software vendors are taking the speech recognition capabilities of a PC into new areas such as home automation and telephony.

Speech recognition requires three elements with *Windows XP Service Pack 1* installed, then Microsoft Speech Recognition Engine v5.0 and an application that will allow for speech input. Typical programs include MS Notepad, MS WordPad, Outlook Express, and other programs that come packaged with in Windows XP. There is no Speech Recognition Engine (SRE) built directly into the Windows XP operating system as such. Users have to install a compatible engine from one of two sources.

➢ If individual students or users, schools or training centres have Office XP, or one of these programs on a computer, then access to the SRE is most likely.

➢ Another option is available for advanced users. The SRE is provided for free as part of the *Microsoft Speech Software Development Kit 5.1*. Microsoft provides no technical support for this software and it is not generally recommended for beginner or end users.

The Speech Recognition is quite basic and elementary as compared to the fully featured programs offered in the market. It can be an interesting exercise for some users to experiment and 'play' with this technology. To all intent and purposes, to avoid disappointment and frustration, it is best to use a well-regarded program as it will provide far more accurate results and be easier to use and master.

The MS Vista SR engine is superior to the XP option and offers more features and accuracy. It may suit some students and be sufficiently enabling for some tasks.

People with disabilities and impairments definitely need to consider more robust and long-term options and seek to secure more viable solutions in order to realise sustained results. Every user has different needs, abilities, and capabilities and he or she may require individual strategies, training and support. Time and practise are fundamental requirements, though.
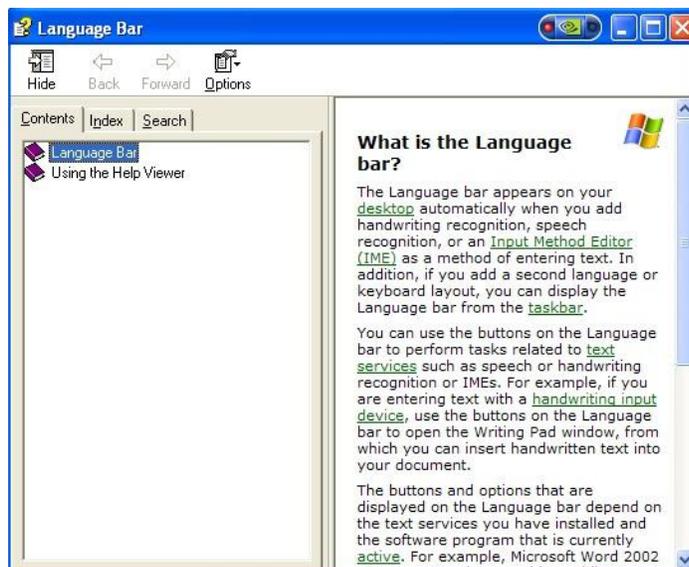
5. Speech Recognition Programs

a. Commercial Speech Recognition Software:

Dragon Naturally Speaking V10   www.nuance.com/naturallyspeaking/
Standard/Preferred/Professional   www.edsoft.com.au/DNS
                                         www.spectronicsinoz.com/catalogue/22536
                                         http://www.novitatech.org.au/product.asp?p=247&id=2492
                                         http://www.voiceperfect.com/ and other dealers around Australia

Support for *Dragon Naturally Speaking* is available from a number of disability support services, agencies and groups. It is best to locate expert users, companies and individuals who can guide and direct students to the most beneficial options. Increasingly, support on campus is available to students from Students Services or DLOs in the Tertiary sector.

Note: Registered teachers in Australia can go online to http://australia.nuance.com/naturallyspeaking/education/terms.html and purchase DNS V10 Preferred for $39.95

Speak-Q   *http://www.wordq.com/speakqenglish.html*
SpeakQ is plug-in software that enhances WordQ Version 2 with simple, speech-to-text functionality. At any time users have the choice of typing with the keyboard, using word prediction, or speaking straight into their text. Speech recognition and word prediction are integrated to enhance the effectiveness of each other. Users can train SpeakQ to recognize their speech using texts provided that match their reading level or educators and trainers, or users themselves can write their own training texts.

Say-MAGic   www.ngtvoice.com/products/software/tandt/say-magic/
                                 www.spectronicsinoz.com/product/speakq
                                 www.novitatech.org.au/product.asp?p=247&id=1873
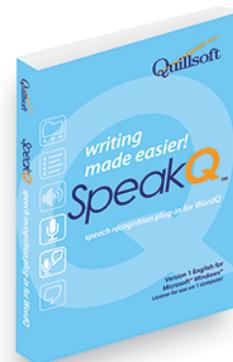
For some time it has been possible to talk to the computer using a natural voice while at the same time receiving speech feedback from it, allowing a blind person to control the computer without using the keyboard. But low vision users (or people with dyslexia and language challenges who need enhanced visual display support) have not been able to benefit from this

technology. Linking *Dragon NaturallySpeaking* from *Nuance* and *MAGic* from *Freedom Scientific* together, *Say-MAGic* provides a range of facilities for users of display management technology who wish to control the computer with the voice.

MacSpeech Dictate V1.5   www.macspeech.co.uk/product_info.php?products_id=978
                                           www.macsense.com.au/
MacSpeech Dictate is a speech recognition solution for the Macintosh. MacSpeech Dictate's accuracy and capabilities make the ubiquitous MAC OS productive and intuitive to use. The program recognizes and understands   13
English language variations - nine with U.S. spelling and four with U.K. spelling.

b. Freeware Speech Recognition Software

MS Windows Speech Recogniser V5.1 Part of MS Windows XP Home and Professional Versions (with SP1) & Office XP
MS Windows Vista SR   Part of MS Windows Vista Versions

Dspeech   http://dimio.altervista.org/eng/
This uses the engine in the prevailing Windows version. It has limitations but it offers an interesting experience. Dspeech is also a Portable application on *EduApps/AccessApps (www.eduapps.org)*
DSpeech is a TTS (Text To Speech) program with integrated functionality of ASR (Automatic Speech Recognition). It is able to read aloud any typed text and it chooses the sentences to be pronounced, based upon the vocal answers of the user.

SpeechVibe V2.0.3   www.tucows.com/preview/510373#MoreInfo
It offers universal mouse-control through automated hot spots and a mouse-grid to perform more complex operations such as drag-and-drops. Further allows dictation anywhere with formatting flexibility and a quick alternate replacement user-experience. It also includes text-to-speech, application launching voice commands and a speech-enabled Internet browser page invocation environment. It is *Shareware*, so users can try it before they commit any money.

CoolInfo 1.0 Voice Recognition   www.coolsoftllc.com/coolinfo/register/
CoolInfo 1.0 Voice Recognition is a free speech recognition program distributed by *CoolSoft, LLC.* It lets users acquire information from the Internet by speech including news, weather, horoscopes and other websites of common interest. Users can search Google, Yahoo, eBay and MSN entirely by speech. Other features such as User Commands and Calculator allow users to create their own speech commands and use the MS Windows Calculator with speech commands. To make it possible for CoolInfo Voice Recognition to provide it freely, CoolInfo displays a banner space, speaks messages from sponsors and requires registration (which is also free). It offers another 'interesting experience'.

6. Portable Solutions

Versions 9 and 10 on DNS provide users with the capability of recording voice into Digital Voice Recorders and importing text into the program for speech recognition outcomes. The *Preferred* and *Pro* version of this software can be used with various models of handheld or digital voice recorders (DVR). Some models work far better than others, and advice should be sought before purchase. The quality of sound varies as does the internal memory storage, the overall operation, battery life and controls.

Cost is an indicator but size, dimensions weight and particularly the appropriate functions and features that will be required will have significant bearing. Once again, research and choosing a device carefully will result in successful speech recognition. Most dealers will provide a list of their preferred models that have proven to work well with the matching DNS software versions.

7. Universal Access using Speech Recognition Software

Being able to speak to a computer is liberating for many students. The ability to use SR well, though, can be a complicated and frustrating experience though if it is hurriedly introduced, without proper research. Once skills and understandings are established and the computer system configured to maximise performance, the results can be life changing. It is not a technology to every student, and it does not necessarily cater to everyone's needs. There are many factors to consider.

Resources

An interesting article in 2008 on *DNS V10 Preferred:* http://www.zdnet.com.au/reviews/software/applications/soa/Dragon-NaturallySpeaking-10-Preferred/0,2000065797,339291627,00.htm.

Thirteen useful FAQs on *DNS V10* at http://www.spectronicsinoz.com/library/dragon-naturallyspeaking-faqs and seventeen videos at http://www.spectronicsinoz.com/product/dragon-naturallyspeaking-preferred-10 are great resources.

A practical Fact Sheet can be located at http://www.adcet.edu.au/Oao/view.aspx?id=3912 on the ADCET site. Some more practical advice is available at http://www.talking.co.uk/spee.htm and a well-written article on DNS is also at http://www.talking.co.uk/nspf.htm (although the choice of colours used on this site are quite unusual)!

Read a PDF article at the following web address on Voice Recognition software as it has some interesting information: www.abilitynet.org.uk/content/factsheets/pdfs/Voice%20Recognition%20Software%20-%20An%20Introduction.pdf

Some 'Tips and Tricks' for DNS software: http://www.nuance.com/naturallyspeaking/customer-portal/tips-tricks.asp whilst a *Voice Recognition Forum* offers some more insights at http://forums.voicerecognition.net.au/.

NB: It was interesting to 'Google' *Dragon Naturally Speaking* – as there were 1,9200,00 hits and for Voice Recognition, 20,000,00 and Speech Recognition 5,680,000.

Microphones are critically important. It is better to purchase a high quality, well-regarded USB model straight away if the student or user is going to require speech recognition as a long-term solution. Placement, positioning and care of the device will provide better results than an inexpensive model that is not looked after or 'knocked about'.

Note: This is by no means a definitive article on speech recognition; it is merely an overview and introduction to the many facets and intricacies of an extremely powerful software genre. The prevalence and significance of SR is gathering momentum and the quantum leaps the IT industry in the delivery of lower cost, more efficient computer power, speed and memory performance are all contributing to more robust systems that, in time, will deliver increased performance and reliability.

It is quickly becoming a mainstream solution and companies are promoting it as a viable if not more practical input method than the keyboard and mouse. When accommodating people with difference, though, great care is required to research all available options. Choosing and acquiring a suitable program and setting up the system, with support advice from people who know the technology well, will be key to successful introduction and implantation. Some students may need direction and assistance from Speech Pathologists or other practitioners as well.

Author: Gerry Kennedy ICT Consultancy © May 2009