

# Speech Recognition Systems



**CALL Information Sheet 15**

**Revised January 2005**



Communication Aids for Language and Learning  
The University of Edinburgh  
Paterson's Land, Holyrood Road  
Edinburgh EH8 8AQ  
Tel: 0131 651 6235  
Fax: 0131 651 6234  
Email: [callcentre@ed.ac.uk](mailto:callcentre@ed.ac.uk)  
<http://www.callcentrescotland.org.uk>

# Speech Recognition Systems

## What is Speech Recognition?

**Speech recognition (SR) systems** allow people to control a computer by speaking to it through a microphone, either entering text, or issuing commands to the computer, e.g. to load a particular program, or to print a document.

SR systems have been around for over twenty years, but the early systems were very expensive and required powerful computers to run. During recent years, manufacturers have released basic versions of their programs selling for less than £50 and have also reduced the prices of the full versions. The technology behind speech output has also changed. Early systems used **discrete** speech, i.e. the user had to speak one word at a time, with a short pause between words. *DragonDictate* is the only discrete speech system still available commercially. Over the past few years most systems have used **continuous** speech, allowing the user to speak in a more natural way. The main continuous speech systems currently available for the PC are *Dragon NaturallySpeaking* and *IBM ViaVoice*. Microsoft have included their own speech recognition system within recent versions of Windows. There is now a version of *IBM ViaVoice* for recent Apple Mac computers.

## Hardware Issues

At the same time as the software has been changing, computer prices have continued to drop and the power of a typical desktop machine has grown to such an extent that SR software will run on almost any standard PC manufactured in the past couple of years, provided that it has sufficient memory. Nevertheless, it is important to pay attention to the recommended specification for a particular program. The latest versions of the continuous speech programs have been designed to take advantage of the increased processing power of Pentium IV computers and their equivalents. If your computer does not meet this specification, it may be possible to use an older version of the software, unless there are particular features of the latest versions that you need. If the processor in your computer only meets the minimum requirement, then the best way to improve performance is to increase the memory (RAM) of your computer. As a general rule, if you are wanting to get a computer that will run speech recognition software it is always best to get a computer with the highest specification that you can afford.

In addition to the basic processor / memory specification of the computer, the performance of a speech recognition system also depends on the soundcard, microphone and general build of the hardware.

**Table 1: System Requirements for Speech Recognition Systems**

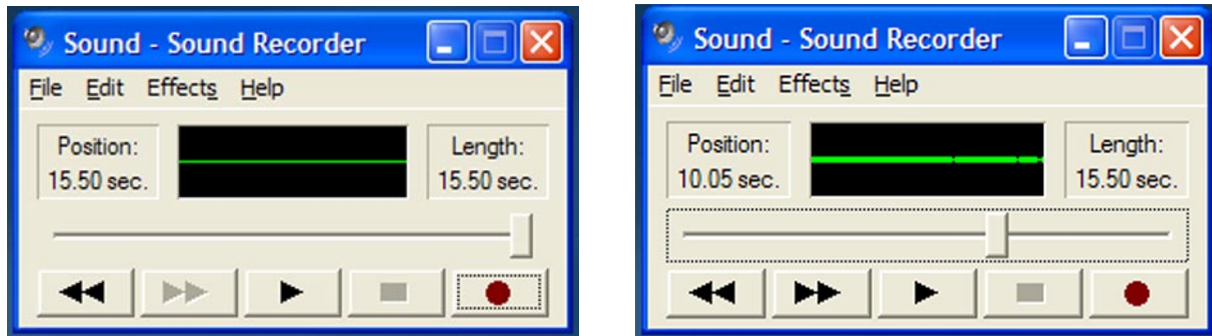
Program	Minimum processor / RAM / Hard disk space for 1 user	Recommended processor / RAM / Hard disk space for 1 user
<i>DragonDictate Power Edition</i>	486 / 20 MB / 43 MB	Pentium / 32 MB
<i>NaturallySpeaking 8 (all versions)</i>	500 MHz Pentium III / 256 MB / 300 MB	Pentium IV / 512 MB / 300 MB
<i>ViaVoice Pro v 10</i>	300 MHz Pentium III / 64 MB Win 98 (192 MB, Win XP) / 510 MB	Pentium IV / 256 MB / 510 MB
<i>ViaVoice Standard v 10</i>	266 MHz Pentium III / 64 MB Win 98 (192 MB, Win XP) / 500 MB	Pentium IV / 256 MB / 510 MB
<i>ViaVoice for Mac (OS 10)</i>	300 MHz G3 / 192 MB / 600 MB	600 MHz G4 / 256 MB / 600 MB

## Soundcards

Sound functions within a PC are generally controlled by a **soundcard**. At one time, the quality of sound input signal produced by some cards was not good enough for speech recognition, but this is rarely a problem in desktop computers built over the past couple of years. There can still be problems with some laptops where components are closely packed, occasionally leading to electrical interference.

It is possible to get an indication of the quality of sound input of a PC by using the Windows *Sound Recorder* utility. In *Windows 98 and later versions* this is found by selecting *Programs, Accessories, and Entertainment*. Try recording and playing back a few seconds of 'silence'. The 'playback' should also be silent, indicated by a thin straight horizontal line in the graph that measures the recording. If the recording has a lot of 'hiss', shown by a thick or jagged line in the graph, there may be a problem with the soundcard. It is possible that the interference is being created by other electrical equipment close by so you may have to experiment by turning equipment off, or re-positioning it. A poor signal may also be due to a faulty microphone, so, if possible, try a different microphone. If it is not possible to get an acceptable recording of silence, then the problem almost certainly lies within the soundcard. If this is the case, and you cannot get an acceptable level of success when running speech recognition software then it may be necessary to get a new soundcard.

#### Recording 'silence' using good quality (left) and poor quality sound cards



A 'poor' soundcard will record a lot of background 'hiss' and will produce a 'fuzzy' signal.

In general, it is best to go for a fairly 'standard' card, e.g. a Creative Labs *Soundblaster - The Soundblaster Live*, costing around £30 would be a good option at present. Alternatively, particularly in the case of a laptop computer where there is no room for an additional sound card, it is possible to get a USB Microphone Adaptor Pod. This connects to the USB interface of a modern PC and has its own sound circuitry, which eliminates the problems that arise from a poor soundcard, or components being placed too close together in a laptop. Any microphone with a standard 3.5mm jack plug can be connected to your computer via the pod. The *Andrea USB Microphone Adaptor Pod* costs £32.95 from iANSYST. If you go down this route, try to avoid having too many other USB devices plugged in at the same time - this can weaken the signal coming from the microphone.

### Apple Mac Hardware Issues

It can be difficult to get a good enough signal with the microphone originally supplied with *ViaVoice*, which was plugged into the microphone socket of the iMac and other Mac computers. IBM are now distributing a USB microphone with *ViaVoice*. The *Andrea USB Microphone Adaptor Pod* mentioned above can also be used with the Mac, allowing people to choose an alternative microphone if, for some reason, they do not wish to use the microphone supplied by IBM. The *Griffin iMic* (£29.90 from amazon.co.uk) is an alternative to the Andrea pod for the Mac.

### Microphones

A good microphone is an essential component of a SR system. Adequate microphones are generally provided with the software package, but there are times when you may wish to consider other options. If a system is not performing particularly well, changing to a better quality microphone can make the difference between success and failure. An important factor is comfort – the standard microphones can be uncomfortable to wear and do not fit well on "small" heads. Here are some other options worth considering:

- *Talkmic* microphone (£49 from iANSYST) – this is a high-quality 'pressure gradient' microphone, which picks up close sound better than background noise and is attached to the head by means of a loop that fits over the user's ear. It is generally very stable and comfortable.
- *Andrea NC61* (£27 from iANSYST) - a good quality microphone which uses 'active noise cancelling' to reduce background noise. Less stable to wear than the *Talkmic*.

**Andrea NC61 (left) and Talkmic (right) microphones**

## Discrete Speech

The early speech recognition programs all used **discrete** speech, which forced the user to have a short pause between words as they were dictated. Most systems now use *continuous* speech, where the user is expected to dictate complete phrases or sentences in one go. Discrete speech systems were particularly useful for people who had difficulty in forming complete phrases in one utterance and were generally more tolerant of non-standard speech. The focus on one word at a time was also useful for people with a learning difficulty. They did, however, force people to speak in an un-natural manner, which could cause voice strain. The only discrete system still available for the PC is *DragonDictate*, available from Words Worldwide. For a small number of people, e.g. those with very dysarthric speech, it may still be the best option.

## Continuous Speech Systems

Over the past few years the field of speech recognition has become dominated by the use of **continuous speech** systems – there is no longer a need to pause between words. It must be emphasised that continuous speech is **NOT** the same as **natural speech**. Success with speech recognition requires the adoption of a precise, clear style of speaking, avoiding slurring and running words together. There is also a need to dictate punctuation, and, possibly, formatting commands.

Continuous speech systems have been designed primarily for use by businessmen and professionals, rather than children in education, or people with disabilities. Although they have the potential to be very useful tools, they contain a number of ‘barriers’ that can make them difficult to use successfully.

The main systems are *Dragon NaturallySpeaking* and *IBM ViaVoice*, both now under the control of Scansoft Inc. *NaturallySpeaking* is only available for PCs, but there are versions of *ViaVoice* for both PC and Mac. Microsoft have built a speech recognition system into their Windows XP operating system, but it is currently designed only for US English speech. Apple have also been working on incorporating speech recognition into the Mac OS, to the extent that the new OS 10.4, due in 2005, could have a very sophisticated interface for controlling the computer by voice. MacSpeech Inc have developed *iListen* for the Mac. It is now available in both US and UK English versions. It is not possible to say definitively that any system is ‘better’ than another in terms of recognition, though *NaturallySpeaking* and *ViaVoice* are more sophisticated than the systems included within the computer’s operating system. Each system is based on a different voice model, so if a user’s voice is close to the model of a particular system, then they are likely to achieve a better level of recognition with that system. There are, however, a number of features related to training and using different systems that make them easier or more difficult to use. *NaturallySpeaking* and *ViaVoice* are available in different editions, varying considerably in price and features provided. It is important to take care to get the most appropriate version for your needs.

### Dragon NaturallySpeaking

At the time of writing, *NaturallySpeaking Version 8* has just been released in the UK. There are three basic editions of the program: *Standard*, *Preferred* and *Professional*. All three allow text to be dictated into most computer applications, and allow control over menus and dialogue boxes by voice. The *Preferred* and *Professional* editions include text to speech playback facilities which are useful for proof-reading. We generally recommend the *Preferred* edition – speech playback is very important in a school situation – where the program is used by a single person using a laptop, or a desktop in a single location, e.g. a learning support base. The *Professional* edition is significantly more expensive, but has additional features which can be useful depending on where the system is being used. The Roaming User facility allows the program to be used over a network, i.e. a pupil could access his/her personal voice file from different networked computers in different classrooms in the school. The *Professional* edition also allows a recording of what the pupil actually dictated to be saved, which makes it easier for the text to be corrected later by another person. Deferred correction is also possible with the *Preferred* edition, but only when the text was dictated into a digital recorder. The differences between the programs are summarised below. Further comparisons are available from <http://www.scansoft.com/naturallyspeaking/matrix/>.

NaturallySpeaking is only available for a PC and requires Windows XP, Millenium or 2000. Mac users will need to use *ViaVoice* or *iListen*.

Different Editions of Dragon NaturallySpeaking 8 Compared		
Program	Cost	Features
<i>NaturallySpeaking Standard</i>	£68	Can dictate, edit and format text in most applications; control menus and dialogue boxes by voice; move the mouse pointer by voice.
<i>NaturallySpeaking Preferred</i>	£120	As above, but with the ability to playback a recording of what was said (while the file is in use) and to have the computer read back the text as it appears on the screen; can be used with digital recorder.
<i>NaturallySpeaking Professional</i>	£399	As for Preferred, but with Roaming User facility for use on a network; can save audio recording of dictated text; can use macro tools to create voice commands for frequently used sequences of commands.
(Prices exclude VAT.)		

### IBM ViaVoice

*ViaVoice* was developed and originally marketed by IBM, but exclusive distribution rights have now been sold to ScanSoft. The program has been updated less regularly than *NaturallySpeaking* in recent years and there is some uncertainty over its long-term future. The latest version is *ViaVoice Version 10*, released in early 2004. As is the case with *NaturallySpeaking*, there are different editions of the program available: *Pro USB* and *Standard*. (More editions are available in the USA). Some of the differences relate to functionality, but others relate to versions of operating system and application program with which it can be used. The *Pro USB* edition can be used with pretty well any application and any version of *Windows*, while the *Standard* edition cannot be used with *Windows XP Professional* or *Windows 2000*. We would generally recommend using the *Pro USB* edition, unless you are using a fairly old version of *Windows* and a restricted set of application programs. If this is the case, it is important to check the detailed comparison of the various editions available at <http://www.scansoft.com/viavoice/matrix/>, to make sure that the program can be used with your software.

There is also an edition for *Mac OS X: Mac OS X Version 3*, which can be used with most applications and allows control of the Finder by voice.



Different Editions of IBM ViaVoice Compared		
Program	Cost	Features
<i>ViaVoice Standard</i>	£30	Can be used to dictate, edit and format text in its own Speak-Pad word processor; can read back dictation and text from document on screen; can dictate directly into limited number of Microsoft applications.
<i>ViaVoice Pro USB</i>	£80	Can be used to dictate, edit and format text in almost all Windows applications; provides command and control over Windows operating system and applications; can be used to transcribe text from digital recorder.
<i>ViaVoice Mac OS X</i>	£80	Can be used to dictate edit and format text in its own Speak-Pad word processor and in text fields of most Mac applications; can navigate Finder by voice.

(Prices exclude VAT.)

### *iListen*

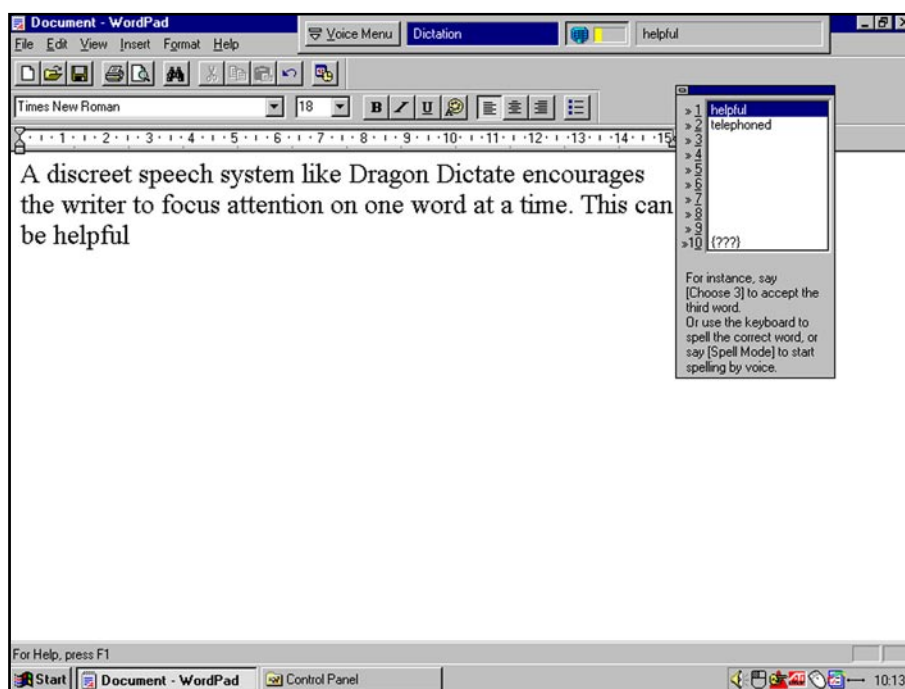
*iListen* is the only current alternative to *ViaVoice* for Mac users. A UK English version can be downloaded from <http://www.macspeech.com> for \$99. Versions are available for Mac OS 9 and for Mac OS X. *iListen* can be used to dictate into most Mac applications and provides reasonable control over the operating system. It can also be used to transcribe words dictated into a digital recorder into text.

### Discrete Speech versus Continuous Speech

Continuous speech recognition systems now dominate, with *DragonDictate* the only discrete speech system still available – and it is hard to find! The argument between the systems is therefore largely academic now, but discrete systems, which require the user to speak with a short pause between each word, can still be more useful for some people than continuous speech systems:

- Discrete speech systems are generally better at recognising non-standard speech patterns, such as dysarthric speech. People with speech and language difficulties can find it very difficult to produce continuous speech.
- Some people with writing difficulties benefit from the focus on a single word at a time that is implicit in the use of a discrete system. This is particularly noticeable when correcting a mis-recognised word when the focus is on one word in a discrete system, but a phrase in a continuous system.

### *Dragon Dictate*



- Continuous speech needs to be *continuous!* People need to be able to speak at least a few words, without pausing, to get a good level of recognition. Changing your mind about a word, half-way through a sentence can cause problems.
- Young, inexperienced writers, particularly children in school, need to learn to distinguish between spoken English and written English. This is easier to do when attention is concentrated on individual words, as in a discrete system, than with continuous speech. Anecdotal evidence suggests that students who have started to use continuous speech systems to alleviate RSI, for example, may develop a much chattier, less formal, style of writing.

Nevertheless, discrete speech is an un-natural way of speaking, which some people find difficult, whereas continuous speech systems allow the user to speak with a natural flow at a normal conversational speed.

## Using Speech recognition Systems


### Setting Up

Once the program has been installed – this is generally straight forward – there is usually an enrolment procedure that has to be followed to allow anyone to use the system. This is usually done in four stages: setting up the microphone; creating a voice file by reading set text into the computer; analysing any existing text that the user has created so that the system can learn his / her writing style; watching an on-line tutorial.

The first thing that has to be done is to set up the microphone sound levels. All of the programs have a short set-up program for doing this, involving the speaking of one or two small samples of text into the system. At this stage, if the person using the system has difficulty in reading the text, it is possible for the user to say something else, e.g. repeat a few lines from a nursery rhyme – all that is necessary is for the user to provide a sample of continuous speech. At the end of this process, you will be given a score measuring the sound input quality. If you are using *ViaVoice*, you should be looking for this to be ‘Excellent’, with *NaturallySpeaking* a score less than ‘15’ will give poor recognition - you should be looking for a score greater than ‘20’. If the sound quality is low, you should be able to improve it by adjusting the positioning of the microphone.

It is good practice to check the microphone setup from time to time before you use your system, particularly if the computer is used for a number of different tasks. Other programs which make use of sound will often change volume settings on the PC's sound card which can be carried over into the SR system. If the level of recognition at the start of a session is unexpectedly low, there could be a problem with the microphone setup so it would be worth running the setup procedure to check this.

**Microphone Setup – Measuring Voice Level**



There is no need to use the text provided, which can be difficult for a poor reader, when the microphone is measuring the volume of the user's voice.

### Training the System

Training the system to develop a good voice file has always been a major issue with speech recognition systems. In the past, enrolment could involve reading up to 20 pages of text into the computer – an arduous task for most people that was almost impossible for people with reading difficulties. There were limited options for the training text, none of which were particularly easy to read.

Recent versions of the main programs involve much shorter training texts, with more choice so that training is less arduous than in the past. Some of the available texts are aimed specifically at children, but there can still be difficulties. Text tends to be offered in chunks of several lines, which can be difficult to read. For people with reading difficulties, there are two main options: to have somebody sitting beside the computer to provide “whispered coaching” of difficult words or phrases; or, in the case of *NaturallySpeaking*, to use the *Keystone ScreenSpeaker* program (see later) to read out the sentences. Another possibility, requiring some basic computer knowledge, would be to print out the training text in a ‘user friendly’ font to allow the user to practice the text beforehand. In the case of *NaturallySpeaking*, it is even possible to edit the training text, breaking large paragraphs up into smaller chunks of text, or even creating an entirely new text that is easier for the user. In a standard installation of *NaturallySpeaking 8*, the existing texts can be found in data files in `.\Documents and Settings\All Users\Application Data\ScanSoft\NaturallySpeaking8\Data\Training\Enx`. We would not recommend that non-expert computer users go down this route.

*ViaVoice* allows the user to pause part way through the training, save progress and come back to it later, if the training is becoming tiring. *NaturallySpeaking* forces the user to complete the whole process in one go. This is less of an issue than it was in the past given that it now typically takes a person with average reading ability less than five minutes to read the training text, but it can still be a major issue for somebody who needs a lot of support. *NaturallySpeaking* is generally more forgiving of slurred speech, mistakes and other problems than *ViaVoice*, which is much more likely to require the user to repeat phrases. *ViaVoice* has an option to magnify the training text, which is very helpful, not just for people with poor eyesight.

### **Analysing Existing Text**

When you have created your voice file, you will have an option to analyse your existing documents. This will pick out phrases that you use regularly and give you an opportunity to add any specialist vocabulary that you use to the system. It is certainly worth doing this if you have a number of existing document files. **NB** If you choose for *NaturallySpeaking* to analyse your documents, the program will automatically analyse text documents in your ‘My Documents’ folder, without giving you a chance to select documents located elsewhere, or de-select files in ‘My Documents’ that you do not wish to use. Make sure the text you wish to use is in My Documents before you start. *ViaVoice* lets you select the documents yourself. Both programs allow you to analyse documents later and to select additional documents.

### **Learning to use the System**

Like any other computer program, there are various commands and operating procedures that the user has to take time to learn. Each of the systems has an on-line tutorial and help facility, that can be used as an aid to learning the system. Most programs provide a help card with a summary of commands. It is not necessary to learn all the commands in one go – it is possible to make effective use of any of the systems with knowledge of only a handful of commands.

In addition to learning the commands needed to operate the program, the user must learn how to make best use of the system, by monitoring how the system is responding to his / her voice. This can only be done on an individual basis. The user should be prepared to experiment with such factors as the volume, tone and rate of speaking until the best level of recognition is achieved.

The CALL Centre produced two books, *Speech Recognition in Schools: Using NaturallySpeaking* and *Speech Recognition in Schools: ViaVoice* in 2002. These include sets of lesson plans and resources which can be used to help introduce speech recognition. The books were designed to be used with *NaturallySpeaking* Version 5 and *ViaVoice* Version 9, but the information contained in them is still very relevant for people with later versions of the programs.

### **Correcting and Revising Text**

Once the initial enrolment has been completed, the speech recognition systems will continue to adapt to the speech of the user, provided that any mis-recognised words are corrected. Mis-recognised words **MUST** be corrected if the user is to maintain and improve the accuracy of the system. It is **VITAL** to distinguish between correcting and revising text. – the former involves the correction of mis-recognised words, the latter involves any changes of word order, vocabulary, etc. typical in most editing tasks.

For example, suppose a person dictated the sentence “I went to the bus stop”, but the sentence “I went to



the bust top.” appeared. Changing the phrase “bust top” to “bus stop” would be a correction. If the user then decided to change the word “went” to “walked”, giving a final sentence of “I walked to the bus stop”, this would be a revision of the text. It is important to distinguish between correction and revision, otherwise, in this case, the computer might offer the word “walked” the next time “went” is dictated.

### How to Speak

Continuous speech is not quite the same as natural speech! It is important to speak clearly, without slurring words, otherwise, a phrase like “the stuff he knows” might be interpreted as “the stuffy nose”. Pairs of small words like ‘the’ and ‘a’, or ‘that’ and ‘but’ are most likely to be misrecognised, particularly in a phrase like “with the book”. It is important to enunciate each word clearly, rather than running them together as people tend to do in natural speech. The user should try to be aware of words that are being mis-recognised. It is possible to give the system additional training in the pronunciation of individual words and it may be that the user might want to slightly alter their pronunciation of some short words if they are consistently mis-recognised.

It is important to try to stay relaxed when using a SR system, particularly when things start to go wrong. It is very easy to become tense and frustrated when the system fails to recognise words. The inevitable change of voice only leads to a further decrease in recognition rates.

### Background Noise

Background noise can be a problem with speech recognition systems, causing words to be misrecognised, or phantom words to appear in your text. It is best to use the system in a quiet room – a typical classroom is not necessarily a good location! As software and hardware has improved over the past few years, it has become less of a problem. If the user does experience a problem, then the use of a good microphone, e.g. the TalkMic or Andrea NC61, both of which have built-in technology to cancel out background noise, should reduce the problem.

### Additional Support - Keystone ScreenSpeaker

*Keystone ScreenSpeaker* is a screen-reading program that has been specially adapted for use with *NaturallySpeaking*. It provides additional support for people with a visual impairment or who have reading difficulties in a number of ways:

- it provides additional, easy training texts, with text presented in short sentences
- training text can be read out, phrase by phrase, allowing the user to repeat the text after hearing it
- when text is dictated into a document, it can be echoed back providing instant feedback as to whether or not the text has been recognised accurately.

### Using a Digital Recorder

It is now possible to make a high quality recording of dictated speech using a portable digital recorder, e.g. the Olympus DM1 or the Sony ICD-MS525. Both *NaturallySpeaking* and *ViaVoice* can be used to transcribe .wav files created by using a digital recorder to produce a text document. This allows the user greater flexibility in terms of where text is created and can lead to time savings where a pupil with writing difficulties is used to dictating work into a tape recorder for transcription by a parent or classroom assistant. *NaturallySpeaking* requires the creation of a special user file for use with a digital recorder, which requires a minimum of 15 minutes of dictation from one of the longer training texts into the digital recorder. It is possible to pause the recorder occasionally to take a break, but this is still an arduous task. Using *NaturallySpeaking* with a digital recorder it is possible to achieve accuracy levels close to those found by dictating directly into the computer. We have not been able to test this using *ViaVoice*.

### Language Issues

Using dictation requires reasonable sentence building skills which might be a difficulty for users with language disorder. This is particularly the case with continuous speech systems. In any case, learning to dictate written, rather than spoken language is a skill that has to be learned, particularly by children. Although the user is speaking to the system, the idea is still to produce written text. This is probably the biggest barrier to the successful use of speech recognition. The technology is now at a level where most people can achieve

high accuracy levels when reading text into the computer, but creating new text is a very different task.

Many people use longer words when dictating than they would if they were typing. Two factors are involved in this: firstly, longer words are easier for the system to recognise, thus a person may say “demonstrate”, rather than “show” because it is more likely to be recognised correctly; secondly, particularly for people who have difficulties with spelling or typing, there can be a sense of liberation in being able to produce a long word that they actually want just as easily as a short word that they might previously have used as a substitute.

## Who can Benefit from the Use of Speech Recognition Systems?

The commercial SR systems were generally designed for lawyers, medical staff and other professionals who wanted to be able to enter text into a computer at speed without having to learn to type. As the systems have become cheaper and more reliable, they have increasingly been useful for many, but by no means all, people with disabilities.

### People with Physical Access Difficulties

People with conditions such as RSI (Repetitive Strain Injury), arthritis, high spinal injury can usually benefit from SR systems, particularly where they have past experience of using computers. The choice of system will depend on the level of disability and the nature of the tasks involved. We would generally recommend *ViaVoice Pro*, or *NaturallySpeaking Preferred* in these cases. Where a person finds it very difficult to use a mouse and keyboard and is likely to be performing regular routine tasks, involving particular sets of commands, the Macro facilities in *NaturallySpeaking Professional* might become necessary.

### People with Specific Learning Difficulties (Dyslexia)

Since a SR system uses words from its own dictionary, spelling will generally be correct. Unfortunately, a word guessed by the SR system will not necessarily be the same as the one spoken by the user. In the case of a discrete system, such as *DragonDictate*, the user needs to have sufficient word recognition skills to be able to choose the desired word from the list of choices or enter it him/herself. The dynamic prediction feature is very useful since the desired word will usually appear quickly provided that the user correctly guesses the first 2 or 3 letters. The text reader can be used to confirm that the final text is correct.

Continuous speech systems, like *NaturallySpeaking* and *ViaVoice* all feature word prediction as part of the correction system, but the programs encourage the user to correct a phrase, rather than a single word. It can be more difficult to identify incorrect words as the focus of the system is on phrases, rather than individual words. The ability of most systems to read back the words that appear on screen can be helpful in these cases. Some systems will also make a recording of what the user has actually said, which can be helpful as a reminder of what was originally dictated. This feature is also useful if a piece of text is being proofread by a person different from its originator.

Although a speech recognition system will help a person who has difficulty in spelling, it cannot help with the organisational skills needed to ensure that a piece of writing has a logical flow. Nevertheless, knowing that the final document will generally be correct can free a person from worrying too much about spelling, enabling them to concentrate more on the organisational tasks.

## What Factors can Influence Success?

BECTA carried out a major research project from 1998 until 2000 into the use of Speech Recognition to assist pupils with Special Educational Needs and concluded that good results were dependent on ‘The Three Ts’: Time, Training and Technology. Speech Recognition cannot be seen as a ‘quick fix’ - it takes time to decide the most appropriate software to be used, and the circumstances in which it will be used. Training is vital to create a good voice file and to learn and adapt to the system. It is important to have the right technology - a five year old computer with 64 MB of RAM won’t do!

Some people can use a SR system and get good results more or less straight away, others need to complete the training procedure and spend many hours using the system, painstakingly correcting misrecognised words, before a satisfactory level of recognition can be achieved. Even where the best technology is in use by a person who has a clear voice, is well motivated and has all the skills apparently necessary, speech recognition can be unsuccessful. A number of other factors can influence the likelihood of success:

## Speech Consistency

Consistency of speech is much more important than voice quality. Many people with quite dysarthric speech are able to use discrete speech recognition systems provided that they are consistent in their speech. The training period will generally be longer and it may be necessary to train words individually to match the speaker's pronunciation. The most recent versions of *NaturallySpeaking* and *ViaVoice* claim to have better support for people with disabilities and non-standard speech.

Speech recognition systems are generally able to cope with regional accents, although initial results may be disappointing and the training period will again be longer.

## Literacy Skills

Given the frequent need to choose the desired word from a list of choices, speech recognition will be most useful for users with reasonably reliable word recognition skills. Discrete speech systems are easier to use for this than continuous ones since attention is focused on one word at a time. SR systems, by themselves, are not suitable for non-readers, though support from *Keystone ScreenSpeaker* can open doors for people with severe reading difficulties.

## Cognitive Skills

The cognitive load involved in using speech recognition systems can be quite high. The person using it must not only think about what they want to say, but also how to say it; they must monitor whether the words they used have been recognised accurately, if not, they must decide on an appropriate strategy to correct them. That is not to mention the possibility of formatting text as it is being dictated. It is not surprising that people using speech recognition often find it tiring! People with conditions such as ME who find prolonged keyboard use tiring should bear in mind the likelihood that the use of speech recognition could be equally tiring - though in a different way. It would be a good strategy to switch between the two input methods.

It is not possible to give some minimum level of cognitive ability below which SR will not be suitable – so many other factors such as motivation and support must also be taken into account.

## Visual Skills

A wide range of information is presented visually from the screen: the text that has been entered; choices for an unrecognised word or phrase; information on how the program is running; even basic information as to whether or not the microphone has been switched on. Nonetheless, people with a visual impairment can successfully use speech recognition systems; the best results have been obtained in combination with, an appropriate screen reader. An example of this is the use of *NaturallySpeaking* with *Keystone ScreenSpeaker*. There is an inevitable increase in the cognitive load, with the user having to dictate to the computer, while simultaneously monitoring a synthesised speech playback of what the system thinks he/she has just said.

## Support

The level of support required will vary from user to user. An experienced computer user who has developed RSI, for example, should require much less support than a young child with complex disabilities. Nevertheless, even the former would benefit from advice from an experienced user as to whether initially poor recognition levels might be due to a microphone problem, or some other factor. A young child would need a lot of support, particularly in the training process, and it is arguable whether any but the most able child under 11 could manage a SR system, even with good support.

## Motivation

This is probably the most important factor for most people who try to use a speech recognition system. Initial results are often disappointing, particularly in comparison with the manufacturers' claims. There will also be occasions when levels of recognition will seem to drop for no apparent reason (usually a change in the microphone setup, although this will not be obvious). Unless the user is well motivated to use the system, working to overcome the difficulties that arise, it is very easy to give up before the system has had a chance to adapt to the user's voice. It is often the case that speech recognition systems are used most successfully by people with a desire to write, which may well have been frustrated by years of failure in the past.

## Health Issues

A number of people have reported having voice strain and other problems arising from the use of discrete

speech SR systems. Speaking in a monotone with a short pause between each word for lengthy periods, usually sitting down without moving, is unnatural and can lead to voice strain. There is less of a problem with continuous speech systems, but people making extensive use of speech recognition should take frequent short breaks, drink plenty of water and vary their tone of speech. If a problem does arise, users should seek professional advice from a speech and language therapist.

## Further Information

Because of the speed of change with regard to the pricing and capabilities of speech recognition systems, the best sources for up to date information are on the world wide web. Here are some of the sites that are worth visiting:

### General Sites

*AbilityNet* (<http://www.abilitynet.co.uk>) Site includes several downloadable factsheets covering different aspects of the use of speech recognition.

*ACE Centre* (<http://www.ace-centre.org.uk/html/resources/vrpart2/vrec1.html>) Update to Mick Donagan's excellent book on using speech recognition in education (see below).

*BECTA SEN Speech Recognition Project* ([http://www.becta.org.uk/teachers/display.cfm?section=1\\_7\\_4](http://www.becta.org.uk/teachers/display.cfm?section=1_7_4)) Description of the use of speech recognition systems in education, with particular reference to project carried out in a number of schools in England and Wales.

*CALL Centre Speech Recognition in Schools Project* (<http://www.callcentrescotland.org.uk/speechrecognition>) Detailed report on project in Scottish schools, and many other resources, including detailed lesson plans for the introduction of speech recognition in schools and an electronic copy of this document.

*Comp.Speech FAQ* (<http://svr-www.eng.cam.ac.uk/comp.speech/index.html>) This is a site with answers to frequently asked questions regarding all aspects of computers and speech. Unfortunately, it no longer appears to be maintained. Chapter 6 is devoted to speech recognition. Some of the information is quite technical, but the site contains a comprehensive guide to the many different packages available, with links to manufacturers web sites, and a lot of useful background material, including material on health issues.

*Computing Outloud* (<http://www.out-loud.com/>) Excellent site, maintained by Susan Fulton. This site brings together the thoughts of a number of speech recognition users on different programs, with useful information and tips.

*Typing Injury FAQ: Speech Recognition* (<http://www.tifaq.com/speech.html>) Detailed reviews of most of the systems available and lots of links to other sites. People should be able to track down any information they want from this, or the Comp.Speech FAQ site.

*SNOW* (<http://snow.utoronto.ca>) Excellent Canadian site with a special educational needs focus. Look under Best Practice and Special Programs for details of speech recognition, including lesson plans for use of *DragonDictate*.

*Speaking to Write* (<http://www.edc.org/spk2wrt/>) This was an American project exploring the use of speech recognition technology to support secondary age students with disabilities. The main focus of the resources is on the use of *DragonDictate*, but there is also some useful general material. The project finished in 2001 and resources have not been updated, but the excellent discussion group is still well used and has moved on to discuss the use of continuous speech systems in education.

### Manufacturers and Suppliers

*ScanSoft Inc.* (<http://www.scansoft.com/>) ScanSoft have taken over development and distribution of *NaturallySpeaking*, *ViaVoice* and other speech-related programs. This is the place to come for technical information about the programs, updates, etc..

*iANSYST Ltd.* (<http://www.dyslexic.com>) iANSYST are suppliers of most of the speech recognition systems currently available and provide detailed unbiased comparisons of the different systems.

*MacSpeech Inc.* (<http://www.macspeech.com/>) Developers of *iListen* program for Apple Macintosh

computers.

Words Worldwide Ltd. (<http://www.keyspell.com>). Suppliers of *NaturallySpeaking* and *ViaVoice* and developers of *Keystone Screenspeaker*.

## Books and Articles

Personal Computer World and other computer magazines regularly feature articles on speech recognition systems.

Baumgarten, J. Alan, Barksdale, Karl & Rutter, Michael (2000) *Dragon NaturallySpeaking QuickTorial* Pub. South Western Educational Publishing.

Coleman, Colette & Meyers, Lawrence S. (1991) *Computer Recognition of the Speech of Adults with Cerebral Palsy and Dysarthria AAC* Vol. 7 (1) pp 34 42.

Donegan, Mick (2000) *Voice Recognition Technology in Education* Pub. ACE Centre.

Fulton, Susan (1998) *The Essential Simply Speaking Gold* Pub. Science and Humanities Press.

Fulton, Susan (1999) *Begin Dictation Using ViaVoice Gold* Pub. Science and Humanities Press.

Gandhi, Parmod et al (2000) *Dragon NaturallySpeaking Complete* Pub. InSync Software Ltd.

Higgins, Eleanor L. & Raskind, Marshall H. (2000) *Speaking to Read: The Effects of Continuous vs Discrete Speech Recognition on the Reading and Spelling of Children with Learning Disabilities*. Frostig Centre, Pasadena, California.

Kambeyanda, Dona, Singer, Lois & Cronk, Stan (1997) *Potential Problems Associated with the Use of Speech Recognition Products Assistive Technology* Vol 9.2 pp 95 101.

Kelway, Peter (1996) *Speaking to Text Interactive* June 1996 pp 34 37.

Kotler, Ava-Lee & Thomas-Stonell, Nancy (1997) *Effects of Speech Training on the Accuracy of Speech Recognition for an Individual with a Speech Impairment AAC* Vol. 13 (2) pp 71 80.

McCall, Karen (2004) *A Client Based Approach to Using Voice Recognition with Screen Readers*, Center On Disabilities Technology And Persons With Disabilities Conference 2004

Miles, Martin, Martin, Di & Owen, Jim (1998) *A Pilot Study into the Effects of Using Voice Dictation Software with Secondary Dyslexic Pupils* Devon LEA.

Nisbet, Paul (2003) *Speech Recognition for Students with Disabilities*, HITmesse Conference, Denmark (<http://www.callcentrescotland.org.uk/speechrecognition>).

Nisbet, Paul with Wilson, Allan, (2002) *Speech Recognition in Schools: Using NaturallySpeaking*, CALL Centre, Edinburgh.

Nisbet, Paul with Wilson, Allan, (2002) *Speech Recognition in Schools: Using ViaVoice*, CALL Centre, Edinburgh.

Rahamim, Lesley (1997) *Talking with Computers Special!* Autumn 1997 pp 57 59.

Rothstein, Norman (2004) *Helpful Steps to Use Voice Recognition with Difficult Voices Closing the Gap* Vol. 23.2.

Tam, Cynthia & Kotler, Ava-Lee (1995) *Assessment Criteria for Using Voice Recognition Systems as Writing Aids Closing the Gap* December 1995 pp 14 17.

Wetzel, Keith (1996) *Speech-Recognizing Computers: A Written Communication Tool for Students with Learning Disabilities? Journal of Learning Disabilities*, Vol. 29 (4), pp 371 380.

(1997) *Voice Recognition Systems for Windows - Frequently Asked Questions* CENMAC Information Paper 1997. (from CENMAC, Eltham Green Complex, 1a Middle Park Avenue, London SE9 5HL. Tel: 0181 850 9229).

(1997) *Using Low Cost Speech Recognition Systems with Children* CENMAC Information Paper 1997.



(2000) *DfEE/BECTA SEN Speech Recognition Project – Final Report* Pub. BECTA

### **Where to get a Speech Recognition System**

Systems are readily obtainable in most computer shops, e.g. Dixons, PC World, Electronic Boutique. Dealers with particular expertise who will be more likely to be able to answer questions and provide basic support and training, if required, include:

*IANSYST Ltd*, The White House, 72 Fen Road, Cambridge, CB4 1UN, UK. Tel: 01223 420101 Fax: 01223 426644. (<http://www.dyslexic.com/>) The web site has lots of very useful, and generally unbiased, comparative information about different products.

*Words Worldwide Ltd.*, Ash House, Belle Villas, Ponteland, Newcastle-upon-Tyne NE20 9BE Tel: 01661 860999 Fax: 01661 822777. (<http://www.keyspell.com>)