

Removing the Barriers Effective Speech Recognition for Special Needs Users

Dr Peter S Kelway

Introduction

Speech recognition technology has been commercially available in Britain since 1988. Many people have been able to use it effectively with minimal training. However, in the special needs area, failure rates have been high. The author believes this to have been more than 60% prior to 2000.

Underestimating the need for good user training and using proper equipment has devalued the perceived potential of the technology. As a result, there remains a reluctance amongst some special needs advisers and assessors to examine the potential benefits of voice recognition techniques. Many parents and teachers remain completely unaware of the benefits that can come from the intelligent implementation of the techniques.

This article describes the opportunities for exploring the use of speech recognition by those with special needs far more widely than is currently the case. New tools are now able to assist in assessment, in training and ultimately in the use of the technology in the special needs area.

All but a very small number of people with unusual voice characteristics will be able to use speech recognition. Even people who have significant speech production problems can become successful users of speech recognition. Those who cannot cope with continuous speech software can often use the discrete alternative where individual words are spoken rather than whole phrases or sentences. There should be no premature conclusions made without formal assessment.

Initial assessment

Professional assessment is vital where speech recognition is to be used to address special educational needs.

It is essential that an assessor is fully conversant with the equipment and all the factors likely to influence the success of applying the technology. Assessments by those with only a superficial understanding of the issues may result in inappropriate judgments.

For routine assessments it is important that a dedicated computer is used, so that direct comparisons can be

made for a candidate against previous results for other people. It is essential to remove any chance that the equipment itself is at fault. The assessor should verify that the correct specification of computer is available for the assessment. Processor type and speed, the amount of memory available and the type of audio system must be suitable.

All non-essential programs should be deactivated so that speech recognition performance is not compromised by background demands on processing power, such as is caused by virus checkers, screen savers and networking systems.

Once the equipment has been thoroughly checked, issues related to human factors need to be addressed. The user should be comfortably seated, in a relaxing environment with subdued lighting. Equipment should be made ready and checked before the candidate is asked to participate in the assessment.

The mere fact that a user does not achieve good results at an initial assessment should not preclude the technology from being considered an option. The assessor should therefore explore the possibility of the user requiring more extensive training than normally provided, rather than consider the candidate to be unsuitable.

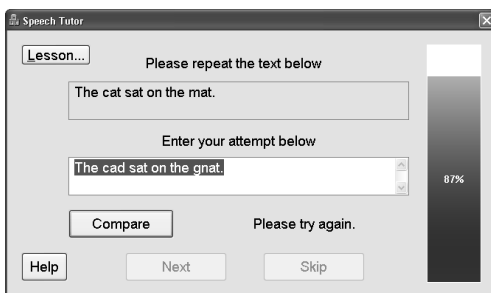
Conventional assessment techniques usually involve the candidate being asked to use the technology in a somewhat unscientific fashion. Initial enrolment entails the candidate reading from an on-screen script into a microphone. Once the computer has produced the necessary statistics, they are asked to dictate a number of sentences so that speech recognition performance can be assessed. Sessions are often of limited duration for practical reasons, especially where a peripatetic assessor is involved, carrying out a number of assessments en bloc.

The result of brief and unstructured testing has two potentially unsatisfactory outcomes. First, borderline candidates are at risk from being considered unsuitable to use the technology and may be rejected out of hand. Second, additional training needs may not be identified, resulting in over-expectation for the future.

There have been two recent developments which assist in assessment. First, the leading voice recognition

system **Dragon NaturallySpeaking** (DNS) now enables a novice to start using speech recognition without carrying out the enrolment process. While it is not recommended that enrolment is dispensed with by special needs users if they do subsequently use the technology, its omission can considerably speed up the assessment process.

In conjunction with this advance in recognition technology, **KeyStone Speech Tutor** has been introduced to enable an assessor to check how well the speech system responds to a candidate's voice. The assessment environment is isolated from any other activity on the computer, so there are no complications which could lead to poor recognition causing unexpected and confusing events.



KeyStone Speech Tutor display

A group of lessons is available within the program, each of which consists of a number of carefully prepared short sentences. The assessor selects an appropriate lesson to match the candidate's circumstances. For instance, if difficulties are expected with words starting with "th", a lesson would be selected containing several sentences such as This is thirsty work. Sentences are displayed successively and the integrated text-to-speech system is used to prompt the candidate, who repeats the sentence. The result is scored by a machine comparison of the resulting text with the original prompt. A perfect match produces a 100% result and the next sentence can then be selected.

After an unsatisfactory attempt (showing a score of less than 100%) the assessor can carry out one of three possible actions:

- If the attempt was reasonable, the speech recognition system can be instructed to accept it and to adjust the stored statistics relating to the user's voice. A new attempt can then be made to check whether the adjustment has been successful.
- The candidate can be asked to repeat the sentence to see whether improvement in enunciation is feasible.
- The example can be skipped and the next sentence is then displayed.

In this way, the assessor can quickly draw a conclusion as to whether the candidate has problems in enunciation or the equipment is operating less than effectively for the particular individual. In either case, a definite strategy can be drawn up to address the issues.

Experience to date indicates that this tool will be particularly valuable to those with special needs who do not have English as a first language. Many people with strong regional or foreign accents find difficulty when starting to use speech recognition systems. A few accents are catered for by allowing the selection of special voice files (Australian, SE Asian, United States and Indian English). However people with one of the large number of accents found in the UK will benefit from being able to identify particular problems which they need to overcome. Where appropriate, the recognition system can also be made to adapt to their voice characteristics.

Starting the training

Once a person has been considered suitable to use the technology, comprehensive training is essential for those who have special needs, especially where specific issues were identified at the assessment session. Before a training programme is embarked upon, the following requirements should be met.

- Trainers must be fully conversant with the technology, having used the equipment personally.
- Good technical backup must be available.
- As with assessments, the computer and audio equipment must meet the required minimum specification, environmental factors should be checked and issues related to human factors must be taken into account.
- Any specific problem areas which have been identified during the assessment should be addressed during the initial training period.

Ideally, the trainer should themselves have been trained by a qualified person, preferably brought in from an acknowledged centre of expertise. Failure frequently results where a teacher or classroom assistant with inadequate knowledge is expected to train pupils.

Once training starts, teachers and classroom assistants need to have ongoing technical assistance available to support them. In many cases training sessions are abandoned because minor technical issues prevent the equipment from functioning correctly. Where this continues, the resulting frustration in both teacher and pupil often leads to the subsequent abandonment of the technology.

Background noise levels are often blamed for failure. However, in over twenty years of providing training

services, the author has never identified a case where this has been a significant issue.

Many people have difficulty when starting to use voice recognition in speaking sufficiently naturally to ensure acceptable performance. Some voices are more susceptible to the effects of variations in speaking style than others, particularly those which are high pitched or have high aspiration levels. While many novices move from the introductory phase within a few hours and can then dictate efficiently and control their computers, others struggle in this early phase.

The author estimates that more than half of young users experience sufficient difficulty to produce a risk of unwarranted abandonment of the technology. Unless remedial action is taken at an early stage, frustration and lack of confidence may mean that the user abandons, perhaps for ever, attempts to use voice recognition. Teachers and trainers frequently become discouraged and come to the often mistaken conclusion that the technology is "inappropriate".

The usual training procedure adopted entails pursuing a strategy along the following lines. The user is first enrolled, when voice characteristics are collected automatically by the computer and stored. A word processor is then used in conjunction with the speech recognition system to allow the novice user to practise speaking and produce text on the screen.

When using this strategy, the user frequently has a poor understanding of both the word processor environment and the operating system. This deflects attention from the main task, especially when recognition errors cause unpredictable effects, taking the screen cursor to unexpected places and performing unwanted actions. Frustration results with a critical loss of concentration.

A novice user needs to become confident, at an early stage, that the computer is responding to their voice. The user should not be allowed to speak in an unconstrained manner before this stage, nor should they be allowed to become tongue-tied which can produce erratic and unnatural speech.

The emphasis should be on encouraging naturally flowing speech of a few words at a time, with a pause of several seconds between each phrase. This allows the user to get into the habit of delivering their phrases clearly and evenly, without hesitating or gabbling, both of which will have a negative effect on recognition rate.

Words should be used which have a reasonable chance of being successfully recognised. Very short words are particularly prone to misrecognition and should not be over-used. Although multi-syllabic words may seem inadvisable in the early stages, phrases that are easy to

pronounce such as elephants and kangaroos escape from zoos are better than go to work on an egg. The use of proper names should be very definitely avoided at the outset.

The user should not be asked to read out from a text book because, even if the user is capable of this, the additional burden may seriously reduce their performance.

The *KeyStone Speech Tutor* tool, which is primarily intended for assessment, is also ideal for taking a novice through the early stages of using speech recognition. The teacher can concentrate on improving the trainee's speaking style and correcting the speech recognition system as necessary. Different groups of sentences can be used in separate lessons, each concentrating on specific difficulties being experienced by the trainee.

Consolidating the training process

Once the mechanism for producing text is understood and the user has started to build confidence in this way, the process of learning to use voice recognition in earnest can begin. Only then should steps be taken to teach the user how to correct mistakes made by the speech recognition equipment.

For those with special needs a particular problem arises with spelling errors. Speech recognition is frequently put forward as a panacea for those who are not naturally good spellers because it is claimed that it is impossible to generate text with spelling errors in it. This is not strictly correct. Two issues arise which cannot be ignored, these being homophonic errors and problems with the correction process. These problems are often wrongly cited as the reason for speech recognition not being suitable for those with special needs.

It is impossible for a speech recognition system to predict exactly what is intended and incorrect words will need to be replaced: for instance the ship was named Grey Starling might be wrongly recognised as the ship was named Grace Darling. It is therefore common to find that users who consider themselves to have mastered the technology perfectly are disappointed to find that their work is riddled with homophonic errors, often with ludicrous results: e.g. John *rode* for the boat team in *grease*.

Speech recognition resolves many of the possible conflicts correctly without user intervention. However, they are rarely totally absent in a document of more than a few hundred words. Many teachers accept this limitation but the ambiguities should be identified and corrected where appropriate. Inadequate checking will lead to uncorrected mistakes such as: Thomas looked *four* apples in the garden. Writers frequently think that there will be no spelling errors, so minimal checking is

carried out and homophonic errors are missed. The use of suitable additional software is essential to give the user confidence that their work is free from these errors.

Secondly, correcting recognition errors requires words to be spelt accurately. If a misrecognition occurs, the correct word must be entered letter by letter into the correction field. For example, if a user says philosophy and the word fishhooks is produced, the user may be tempted to start spelling the word with f rather than p. If so, the word philosophy will not be produced and they may well produce flosfee as their attempt, spelling the word phonetically.

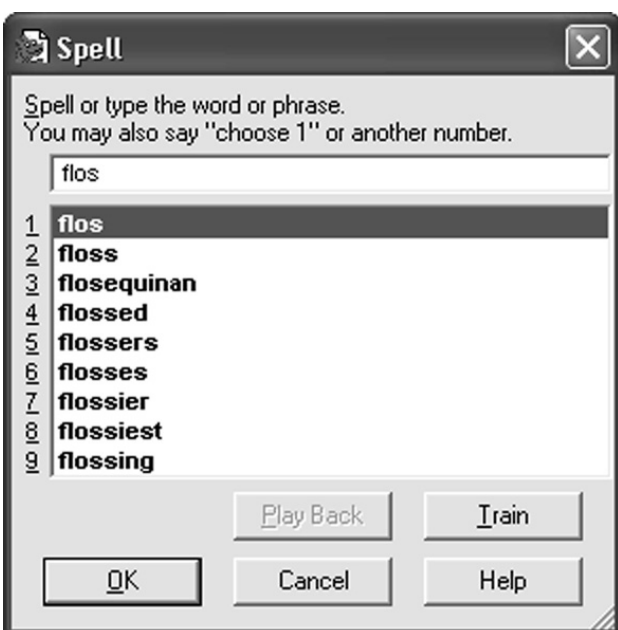
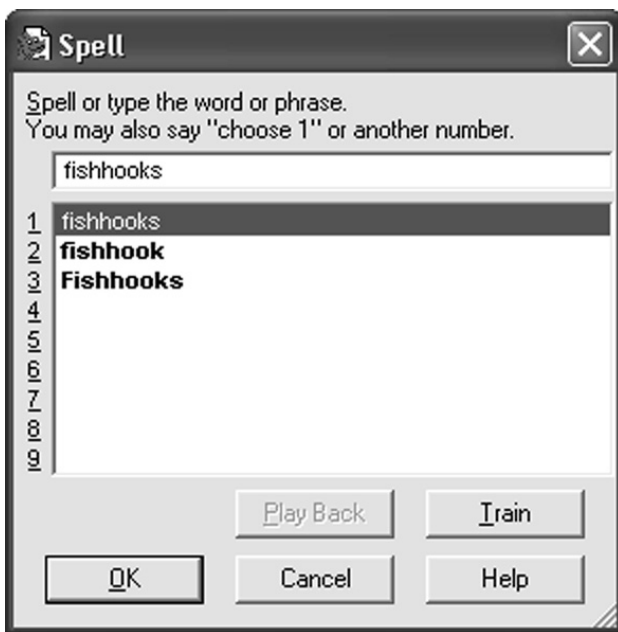


Example of a user embarking on a false trail in correcting an error.

If the user then accepts the spelling of the word as flosfee this is entered in the voice recognition vocabulary as a wrongly spelt word. The next time that the word is spoken it will be represented as that misspelled word. Phonetic attempts at spelling words will result in frustration or error unless additional software is available. It is vital that this software provides good support for people who tend to spell phonetically.

Special needs users may have an additional difficulty. Many are unable to read sufficiently well to check that the text displayed on the screen as a result of their dictation is what was intended. This can lead to uncorrected errors and cause frustration. Even if the error is detected, the user may then be unable to read the words produced in the selection list of alternatives during the correction process.

Text-to-speech systems such as **Read and Write** can assist users in checking work by allowing sections to be read back after completion. However, this remains a clumsy solution for many users who require immediate confirmation of what they have dictated. There are two products on the market which provide immediate confirmation of dictated text: **ClaroRead** and **KeyStone ScreenSpeaker**. Both echo the dictated text as soon as the user pauses. Some users prefer to have the text echoed on demand, rather than continuously which can be tedious and cause mental overload. KeyStone therefore has an option whereby the user can give a very simple command when an immediate echo is required after a particular sentence has been completed; at other times no echo is produced.



Summary

There continues to be poor awareness amongst parents, teachers and education authorities regarding the power of speech recognition when properly implemented.

A need remains for more assessors to be aware of the potential of the technology and for a greater understanding amongst teachers of the necessity for structured training.

Traditionally speech recognition systems have been under-used in providing support to the large number of people with special needs. The recent emergence of structured assessment tools and training systems for novices should give new hope for the future. Suitable technology is now available at affordable prices.

There is no longer an excuse for any person with special needs to be deprived of the opportunity to explore whether they can take advantage of this powerful technology.

Available software

Speech recognition systems

Dragon NaturallySpeaking (DNS): This is considered to be the only practical continuous speech solution for special needs users.

IBM ViaVoice, which has been available almost as long as DNS, is not recommended. The product has not been developed beyond version 10 which was released more than two years ago. Young users have found it difficult to enrol on the system and accuracy is considerably lower than DNS for users in general.

DragonDictate:. Although this product has not been upgraded for more than five years it remains a highly effective tool for users with speech production problems or learning difficulties. Dictate requires a distinct pause to be left between each word uttered (typically of about 0.2 seconds). It will cope with poor speech production provided that utterances are consistent. DNS can cater for pauses of any length between words and handles truly continuous speech but requires reasonably good delivery.

Text-to-speech systems with integrated homophone and spelling checkers

All the products listed comprise a good spelling checker and homophone checker, along with the capability to read out text which is displayed on the screen.

ClaroRead: In addition to providing conventional text-to-speech facilities, this product provides echoing of voice recognition utterances while a user dictates using a word processor. It also works well with Internet Explorer and provides a general screen reading capability.

Dolphin Tutor: Tutor has essentially the same

functionality as ClaroRead.

KeyStone ScreenSpeaker: KeyStone seamlessly integrates with Microsoft Word, WordPad, Notepad and with DNS and DragonDictate voice recognition systems. It is specifically designed to assist those with special needs to generate written work. For those wishing to extend text reading to general application programs and for browsing the internet, it is recommended that the product is used in conjunction with Read & Write.

Read & Write: This product was not designed to operate with voice recognition and is less effective than the other products. It does however provide an excellent text reading capability across a variety of application programs. It is best used by keyboard users and those browsing the internet.

Software for assessment and initial training

KeyStone Speech Tutor is a tool designed to assist special needs assessors and training of users in the initial stages of using speech recognition.

Complete solutions

Bundling speech recognition and text-to-speech software with suitable word processor equipment produces highly cost-effective solutions. The following comparisons show the range of alternatives, all of which should provide effective solutions for most special needs requirements. Significant discounts will be available for use by resource centres by negotiation with suppliers.

ClaroRead Plus with **DNS 9 Preferred** is £309.

Dolphin Tutor with **DNS 9 Preferred** is £349.

KeyStone SpeechMaster Solo with **DNS 8 Professional** is £249.

KeyStone SpeechMaster Solo with **DNS 9 Preferred** is £299.

It should be noted that DNS version 9 requires significantly greater machine resources than version 8. For this reason Words Worldwide has negotiated a special licence to keep the earlier version available to special needs users.

Dr Peter Kelway

Dr Peter Kelway is the Chairman of Words Worldwide Limited and has worked in the special needs arena for over twenty years. He has acted as a special needs assessor and trainer and designs software incorporating speech recognition and text-to-speech technology.

Useful web sites

following web sites can be browsed to find details of the software products described in this article:

ClaroSoft: www.clarosoftware.com (text-to-speech)

Dolphin: www.dolphinuk.co.uk/education/products (text-to-speech)

Nuance: www.nuance.co.uk (voice recognition)

Texthelp Systems: www.texthelp.com (text-to-speech)

Words Worldwide: www.keyspell.com (text-to-speech, voice recognition, voice assessment tools)