Voice Recognition – all it’s ‘suppozd’ to be?

Andrew E. Fluck, University of Tasmania
Andrew.Fluck@utas.edu.au         http://www.educ.utas.edu.au/users/afluck

Abstract

So you can talk to your computer? And it can talk back? But do you make sense to each other???

This paper presents findings from a project run in the UK on the use of speech recognition in special education, and compares these with the use of speech recognition products by students and staff in the Faculty of Education at the University of Tasmania. In each case, there are particular expectations about the benefits that speech recognition will bring, some of which have been met, and others that require specialist help to achieve. The specialist can advise on the folly of watching the screen, of going too slow, and show where faults arise from using incompatible hardware or software.

In special education, there is a clear case for some students to be given access to this new kind of facility, unlocking their potential in ways that expose their real abilities. But as the price of this speech recognition technology drops, and availability increases, other students can expect to get a go at dictating. What happens to spelling when you can print a version of what you just said? Does grammar get ‘mangled’ if your write what you say?

There have been some wonder-stories, and this paper will present some of them. It’s almost a revolution in style, and certainly already becoming a small industry of its own.

Background

Greg Vanderheiden from the TRACE Centre in Wisconsin, USA, has a habit of commenting that technological improvements for the disabled often lead to developments that benefit the whole of society. He is continuing this crusade by developing guidelines for making web-pages more universally accessible, and suggesting ways our domestic devices can be improved with the capacity of speech. This wider benefit can be seen in a whole range of human activities, from the sock-sorter for the blind that eliminates the plethora of single socks most people get, to computer aids such as the concept keyboard. Views of the proportion of the population which is disabled change according to the particular activity under scrutiny, but many find themselves in this category when faced with a QWERTY keyboard.

This keyboard is the standard input device for most computer users. Many school students are offered the opportunity to undertake considerable study of its use, with proficiency being widely required. Yet popular myth holds that the 1870 invention of Christopher Sholes was purposely designed to be unergonomic, to prevent mechanical typewriters from jamming (Hartman, 1997). Alternative designs have been proposed, notably the 1930s Dvorak keyboard, which places common letters in the English language close to the home, or resting, keys.

However, for many people, the keyboard whatever the key layout, represents a major obstacle when considering computer use. This may be due to unfamiliarity, and lack of training to establish confidence, or due to some physical difficulty. In the arenas of special education and adaptive devices, alternatives to the keyboard are common. These include the expanded keyboard, mini keyboards, keyboard emulators, soft keyboards and so on. A most useful adaptation is a keyguard, which rests over the keyboard, and supports the hand whilst the correct hole is located to depress
the key below. The range of keyboard emulators is huge, but these devices are excellent for those who have limited reliable physical movement. Activated by switches, or even by eye-gaze, the keyboard emulator sends signals to the computer identical to those that would be sent by a conventional keyboard.

Such is the need for modifications to the standard input device, many operating system manufacturers include options to alter keyboard characteristics. Microsoft Windows from version '95 onwards, includes the Accessibility Options control panel (START > Settings > Control Panel > Accessibility Options) which allows the use of Stickykeys, Filterkeys and Togglekeys. These facilitate one-handed operation, or give sound cues to augment feedback on keyboard status (such as caps-lock etc).

Speech recognition has long held the promise of eliminating the need for keyboards. It is therefore not surprising that special educators should seek to exploit it at an early stage. This report gives a summary of the findings of a British project that deployed a large number of speech recognition systems into schools. Another project, 'Speak to Write' in the United States of America is also noted. The experience of the Faculty of Education at the University of Tasmania leads to an analysis of the implications for the wider school student population.

**SEN Speech Recognition Project**

This project in England was started in 1997 by the British Educational Communications and Technology agency (BECTa, 1999) and ran in 12 centres with the aim of investigating 15 research questions. The research questions aimed to discover how useful speech recognition products were to disabled students, looking at ease-of-use, effects upon motivation and subsequent literacy achievements. Here is the full list of the research questions:

1. Can the software respond efficiently to a wide range of voices and accents used by learners in a broad age range?
2. Can learners with SEN acquire the skills needed to use the software?
3. How might the software support a range of learners who have difficulty in recording their work using the more traditional method of pen and paper?
4. Can the software offer specific opportunities for the development of literacy skills?
5. Are there other skills that can be developed using the software, such as the speech and syntax of dyslexics or those for whom English is their second language?
6. What range of support is needed to enable learners to use the software effectively?
7. What is the range of hardware best suited to using the systems for particular purposes?
8. In what environments do the systems function best?
9. Are the systems easy to install?
10. Is an enrolment period necessary and is enrolment easy for, for example, blind learners or poor readers?
11. Does the potential for inaccuracies demotivate users?
12. Are other solutions better (switches and predictive word processors, for example) for people with physical disabilities?
13. What do the systems offer to visually impaired learners?
14. Is it possible to effectively control applications other than word processing or give access to the NGfL (National Grid for Learning) using speech?
15. Can speech recognition be used to control a person’s environment?

Each centre examined a mixture of the research questions. A summary conference in November 1999 looked at general findings from the project. Several speech recognition products had been used:
• Dragon NaturallySpeaking (DNS)
• Dragon Dictate (DD)
• IBM Via Voice (VV)
• IBM Simply Speaking (SS)

One of the main topics of discussion was the comparison between discrete and continuous speech recognition. The centres had found that for some students there were advantages in using the discrete systems, where each word had to be enunciated separately. Shorter training times and more accurate recognition were the desirable features quoted for the discrete systems. One solution to this dichotomy was a technological advance called Keystone 2000, which allowed a continuous speech recognition system to automatically switch to discrete mode. Other users of more recently released software (such as DNS version 4) described the way in which technological improvements had narrowed the gaps between the different recognition methodologies.

Comparing manufacturer's specifications with actual performance, the centres found that the lowest hardware specification was inadequate for good and prompt performance. Some participants found that different combinations of peripheral devices suited different students, so choosing the right combination of software, software settings, sound card, and microphone was a critical process. Systems, which were only capable of detecting discrete speech, worked on hardware with slower processors and half the memory of equipment offering continuous recognition. Training was long-term, requiring considerable support for an extended period if the system was to be used successfully. Separate duplex soundcards, especially those from Turtlebeach, performed better than those integrated on the motherboard, and this made the use of most laptop computers problematic (except the Compaq Armada 1700 worked with DNS). Universal serial bus (USB) microphones that by-passed the sound card were also very good. Active noise cancellation microphones from Andrea were recommended.

Trials with blind students showed that speech recognition and screen-reading software, such as TextHelp 98, or JAWS for Windows developed by Henter-Joyce, which interpreted the computer display as a spoken voice, worked well together. Systems were surprisingly resilient to background noise. When offered a pitch option on enrolment, the low pitch choice generally suited boy students above the age of 14.

Staff commented on the effects of using the speech recognition systems: "There was a clear improvement in user's speech. It was slower and more distinct." This was not an isolated comment, but pointed towards some resolution of the fifth research question. Students soon learned that getting angry at poor recognition was counter-productive, and moved on to practice using a calm intonation and mood. A student said: "It was very easy to learn how to use it, in the future I will use it in class and for homework." In terms of other effects, it was noted that students using the systems never saw a misspelt word on the screen, so no bad habits were reinforced and some spellings were learned through frequent exposure. One school measured recognition accuracy rates, getting 67% with a hearing impaired student, and 99% at best with another student.

Moving on from technical issues, staff in the projects where discrete recognition systems were adopted, expressed worries about the potential effects upon literacy. Whilst overall the project had found positive outcomes in most areas of literacy development, they felt that there was the potential for the stilted form of speech used in dictation to become an impediment. It was interesting that no such difficulty was reported, but several centres commented on improvements in student articulation.

One centre found the reading age of the enrolment script for VV too high, and was invited by the manufacturer to write a new one. The resulting product has a reading age of 10 years, and is now marketed by the manufacturer. Public examination boards were the subject of some concern. These bodies generally appear to accept the use of augmentative and adaptive aids to writing where a medical case can be made for it, but do not accept the argument that some students require such assistance in order to match an average writing speed.
Speaking to Write
In Northern America, the Speaking to Write project was initiated in 1997 at the Education Development Centre and the Communication Enhancement Centre at Children's Hospital, Boston. It aims to investigate commercial speech recognition products by secondary students with disabilities, and to produce materials and adaptations to improve their functionality. The project is Federally funded through the U.S. Department of Education, National Institute on Disability and Rehabilitation Research (NIDRR, 1998). The project runs both a web-site and an associated listserv.

Since continuous speech recognition technology was not commercially available when this project was initiated, the researchers produced an update to accommodate the new products. Bearing in mind that the focus was upon students with disabilities, it is salutary to read:

“based on our explorations into continuous speech with a few students and our own impressions after years of watching individuals with disabilities use discrete speech recognition software, we have some serious questions about the utility of continuous speech recognition for many individuals with disabilities.” (Speak to Write, 1999)

This echoes many of the feelings expressed by the English researchers. The point is made that speech models used within the software are based upon adult fluent speakers, and these are inappropriate for understanding disabled students. However, it should not be drawn from this that the same would apply to most students in a regular classroom. The project advises that speech recognition can be used to facilitate various stages of the writing process, from brainstorming, to drafting, revising, and editing.

Reference is made to some initial studies upon the effectiveness of these systems in improving academic performance. One study (Higgins & Zvi, 1995) looked at the performance of learning disabled college students using speech recognition technology in a written proficiency exam. Learning disabled students using speech recognition systems achieved the same distribution of scores on the exam as their non disabled peers. Their score distribution fell below that of their non-disabled peers when a human transcriber, or no assistance at all, was provided.

University of Tasmania, Faculty of Education
Since 1996 the Bachelor of Education course at the University of Tasmania has included a mandatory unit on Personal and Professional computer skills for intending teachers. Much of the material presented in this course is aligned to the teaching profession, and aims to give a broad base of understanding with this vocation in mind. The opportunity to extend student knowledge of alternative input devices was combined with the practical realisation that for many of them the speech recognition route would prove personally appropriate. The latter has indeed been the case, with several anecdotal comments from students that they regularly use similar software on their home computer for first drafts of essays.

The course unit demonstrates the use of Dragon Naturally Speaking version 3 to students. Generally, one in each tutorial group undertakes the training session, then reads a randomly selected text into the simple editor supplied. Discussions following are generally amusing, focusing on the mistakes in recognition. However, this also allows some reflection on the conditions for success: good equipment, similar background noises, repeatable intonation and so on. It also gives a chance to speculate on the viability of the methods for personal productivity, for literacy lessons in the classroom, and for equity issues for disabled persons. Since the equipment is permanently installed on one computer in each 24 hour teaching laboratory, the students are able to follow up by recording their personal speech template to be stored in their fileserv account for future use.

The Dean of the Faculty, seeing the success of this undertaking, supported the purchase of further sets for staff experimentation. The Philosophy lecturer was an early adopter, and within a short time was able to dictate material for an academic paper. At the current time, approximately one quarter of the academic and general staff in the Faculty have the facility installed on their workstations.

Meanwhile, in a separate development within the same University, a student had her learning interrupted. No longer able to make extensive use of a QWERTY keyboard, she convinced the vice-chancellor to provide a speech recognition system. She successfully installed and used a system on a university computer, and was subsequently employed to install further systems and train peers to
use them. Since graduation, she has begun work in a consultancy that provides specialist support for
voice recognition systems, with more than half the work coming from business clients rather than
disabled users.

She has provided some training for the staff of the Faculty, and several interesting uses are being
made of the systems. Several staff are using the facility to improve the process of transcribing
research interviews. Both direct input using mini-disc recorders, and also indirect methods (where
the transcriber listens to a tape and speaks the content into speech recognition system they have
previously trained) are being tried. Another member of staff is using the system to interface with a
widely used e-mail package, because this represents a high proportion of his computer use. Office
staff are not yet using the system extensively, and this is the next area for a break-through.

The Future

Speech recognition products are developing fast, and many major software companies are
positioning themselves to take advantage of this. IBM now bundles VV with many personal
computer systems, and have made it available for Macintosh computers. Microsoft has formed an
alliance with Dutch firm Lernout & Hauspie (L&H, 2000), which is based in the Flanders
Language Valley, a high-tech industrial estate initiated by L&H and the Flemish Government.
Work is proceeding by several research institutes on the automatic translation of one spoken
language into another. "The SLT [Spoken Language Translator] is currently capable of translating
spoken English queries in the domain of air travel planning into either Swedish or French, using a
vocabulary of about 1200 words" (Rayner, 1997). The Janus Project of Carnegie Mellon University
and Universitat Karlsruhe aims to translate "spoken language, much like a human interpreter. It
operates on a number of limited domains such as appointment scheduling, hotel reservation, or
tavel planning" (Woszczyna, 1999).

As teachers, we should give thought to the effect that these technological developments might have
upon general education for the majority of our children. It has been suggested that these new
systems might best be deployed in situations where pre-pubescent boys with low literacy skills are
causing disruptions in schools. Even if the systems assist teachers in the education of such students,
this gives little idea of the true potential for less demanding students.

Staying with a focus upon literacy, what would the development process be like if the skill of
writing were transformed into one primarily of diction? Since newer systems deliver very high
standards of accuracy, and that they normally display words drawn from a dictionary (giving the lie
to this paper's title), spelling as a skill could be transformed or even eliminated. For students of a
non-phonetic language like English, this would be a big change.

Perhaps the easiest way to apply such a facility in the classroom would be for the daily journals of
Year 1-3 children, or the creative writing for older Primary school children. A common set of
problems the younger children have is that of letter reversal, where a b is written instead of a d.
Another common problem is the exchange of capital letters for lower-case. Given good accuracy
dictation software inputting text into a common word-processing program, students would only see
correctly spelled text, and grammatical errors would be gently highlighted. Since most software can
read text back, the issues would change from letter reversals to editing techniques for correcting bad
grammar, a higher-level issue.

Likewise, the older children typically construct creative stories with the infinitely concatenated
sequence of events:
... the bus went on the road and the boy jumped on and the parcel moved and ...

The challenge for the teacher would be to find a way in which dictation systems could lead onto a
way to teach stylistic patterns, using very short sentences for action sequences, and longer more
descriptive ones to set the scene. Text-to-speech would probably be a great aid here, especially if
used to read out comparative versions of the same material.

Moving to the curricula offered in many high schools, keyboarding is a popular option. The
introductory courses at least, would be eliminated if speech recognition were installed and
accessible throughout the school. With many manufacturers bundling the software with computer
systems, this might be achievable very soon. This author’s understanding is that computer capable
of running the Windows 2000 operating system will necessarily be of a class that can easily handle
the processing and memory requirements of speech recognition and text-to-speech software. At this
stage however, it does not appear that the operating system itself will be speech recognition-able,
which might be the next logical jump. The reader may have a personal opinion on the integration of
a web-client into desktop computer operating systems, and it may well be that integration of speech
will be equally litigious. Yet this is perhaps the kind of development that would be needed for the
user to carry out a working conversation with a workstation. Perhaps something like this:

Me: Open e-mail package
Computer: e-mail package is open
Me: Read headings for all highest priority messages
Computer: Jane Hay has sent a message about a dinner date tonight.
Me: Reply saying yes, attach the photograph called Jane’s cat, and copy Jane’s message to
my palmtop.

Given the impending production of speech language translation software, what would this do to
Languages Other than English? Whilst there would still be a great deal to be learned from the study
of the culture of a foreign country, there are already opportunities to speak to indigenes via desk-top
video. If the translation projects are as successful as current textual ones, participants could each
speak and hear their own native tongues.

A significant point has been reached with any new technology, when it causes radical change in
school curricula. Computers have just begun to introduce elements into the classroom which were
not possible prior to their development. From the evidence presented in this paper, it appears that
speech recognition technology is on the brink of causing a similar change.

References

BECTA (1999) Speech Recognition Project, Coventry, England and on-line at


NIDRR (1998) Speaking to Write project web-site, on-line at http://www.edc.org/spk2wrt/ on 24th
January 2000.


Speak to Write (1999) Update on Continuous Speech, on-line at

electronic devices: expanding the market for voice products, on-line at