His master's

ILLUSTRATION JAN WHADCOCK

FOR MANY YEARS NOW, the Holy Grail of computing has been voice control. Thus far, it hasn't been much cop and we've had to stick to using keyboards or mice as input devices. However, if the current crop of speech recognition packages is anything to go by, their days are numbered.

A transformation in the state of speech recognition technology has taken place in the past year or so. Recognition software has continued to improve but the main reason for the rapid progress in speech recognition has been the enormous increase in processor power, courtesy of the Pentium III and the Athlon, coupled with a dramatic drop in memory prices. Today, PCs driven by 600MHz processors with 128MB of RAM are not uncommon. That type of specification may do precious little to the performance of Office 2000, but it does make a huge difference to speech recognition software, one of the few types of application capable of taking advantage of the Pentium III's SSE multimedia extensions. As a result, learning times have been slashed while accuracy has increased.

Current speech recognition programs let you control nearly every aspect of your computer without having to touch the keyboard. You can accurately dictate text at speeds approaching 150 words per minute, edit and manipulate documents, create spreadsheets and graphs, join chat rooms and surf the web – all with your voice.

Voice tests

In this feature, we tested the latest versions of the four market-leading speech recognition packages: Dragon Systems' NaturallySpeaking Preferred 4.0, Lernout & Hauspie's (L&H) VoiceXpress Professional 4.0, Philips' FreeSpeech 2000 and IBM's ViaVoice Millennium. At the time of testing, IBM couldn't supply ViaVoice Pro Millennium, which has a feature-set comparable with the other three products, and instead supplied the more basic ViaVoice Standard Millennium, which has the same recognition engine, but lacks most of the command and control functionality and direct dictation in to major applications. The products were tested on a 500MHz Athlon PC with 128MB of RAM running Windows 98 SE.

voice

'You say to-may-to, I say tomato...' Roger Gann makes his way to **the software side of speaker's corner** to judge the accuracy and performance of speech recognition packages.

Ease of installation

All the packages employ a wizard-led installation routine. A good, clean install is a key part of obtaining good recognition scores as so much hinges on optimising the audio input. All four packages lead you through a series of audio tests to ensure the microphone is set up correctly, while testing for level and background noise and, on the whole, this is handled in an easy-to-follow manner.

Both ViaVoice and VoiceXpress place great emphasis on getting the position of the microphone correct and even provide video clips to drive the point home. VoiceXpress also has excellent online help and diagnostics.

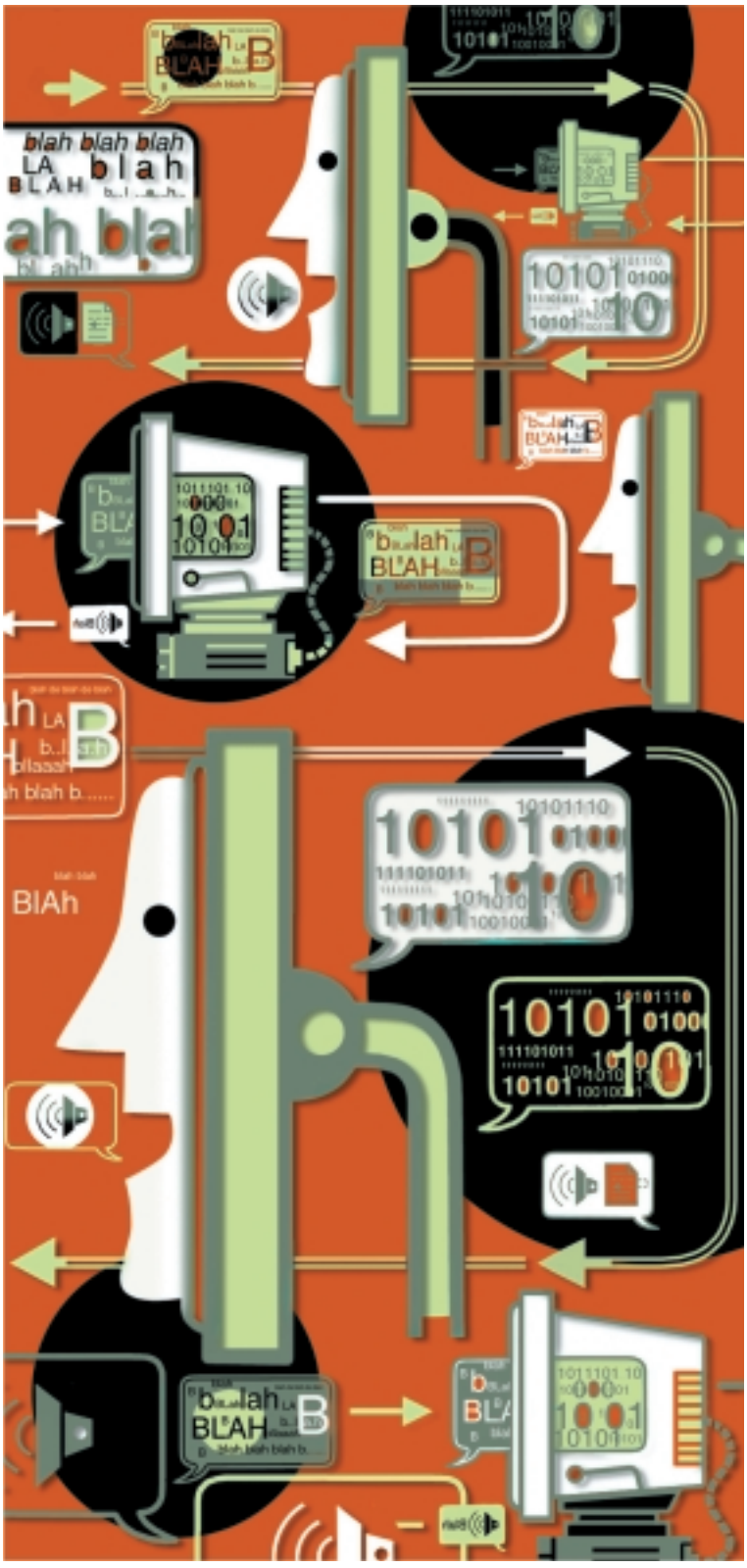
However, it's not all plain sailing. Most sound card manufacturers have at long last applied some common sense and are now colour coding the 3.5mm jack sockets on their cards – red for 'mic in', green for 'line out' and blue for 'line in'. It's a shame that the headsets provided made no attempt to match this colour scheme – the L&H Telex headset mic plug was blue while the IBM Andrea mic plug was green. How about red?

Training

The big news about this crop of speech products is the reduced enrolment times. Enrolment is a key element in the recognition process – the package has to learn how you speak in order to recognise your words. Previously this took a long time – perhaps 45 minutes to read out 100 sentences and then a further 20 minutes while this data was analysed and profiled. It was a very tedious process.

Today's ultra-fast AMD Athlon and Intel Pentium III processors only really justify their price tags when performing compute-intensive tasks, such as voice recognition. They make a significant difference, as all the speech packages are optimised for a range of processors from the Pentium III downwards. Both VoiceXpress and NaturallySpeaking now offer enrolment sessions that clock in at about eight minutes, with only two or three minutes of 'crunching' time.

These are remarkably brief times, especially when you consider the high levels of accuracy they then deliver. FreeSpeech 2000 offers a 15-minute enrolment lesson but these extra seven minutes aren't really a major burden. Curiously, ViaVoice seems to offer the same enrolment regime as



Vox pops for desktops

ViaVoice 98, with some enrolment sessions as long as 60 minutes. We chose a 15-minute lesson, of 88 sentences, but this was completed in just 10 minutes, with a four-minute data analysis period. All the packages recommended additional training, however.

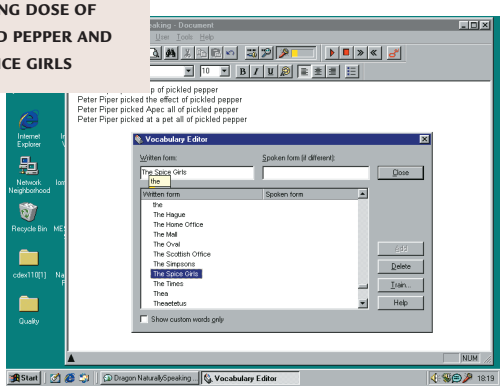
On top of all this, the four packages offer a document analysis facility for augmenting your dictation vocabularies. For example, ViaVoice has an 'Analyse Document' option, which searches your documents for unknown words, and a Topic feature, which loads specialised topics such as Computers or Chatter Jargon, depending on the current application.

Users and languages

The four packages supported multiple users – for example, other family members – and these were all easy to set up, not forgetting that every user needs to go through the enrolment procedure.

NaturallySpeaking 4.0 supports a wider range of language models, including children, teens, and senior citizens, making the software ideal for families with several users of different ages.

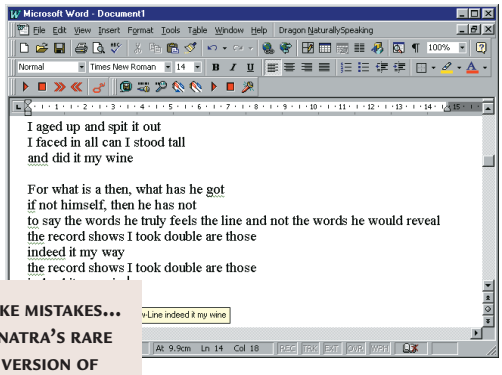
▼NATURALLYSPEAKING GETS A TONGUE-TWISTING DOSE OF PICKLED PEPPER AND THE SPICE GIRLS



Specific lexical groups, aimed at the medical and legal professions, for example, are available for VoiceXpress, while a legal vocabulary is a free bonus with ViaVoice Millennium. Surprisingly, only one of the four packages comes bundled with additional language models. With FreeSpeech 2000 you get no fewer than 13 European languages, not bad for an £80 package. All the other packages only supported one language, ie UK English and if you wanted, say, French speech recognition as well, then you have to buy the French version of that package.

Accuracy

All the packages we looked at delivered recognition accuracies that would have been astonishing a mere 12 months ago. With 'business-style' letters and reports, accuracy across all four products was remarkable, even with place names and surnames. Reading out a weighty business report of 160 words would typically result in about four errors or 97 per cent accuracy. All four did well but VoiceXpress seemed to have problems recognising the 'new line' command, thinking it was 'the line' instead, and FreeSpeech 2000 needed more training to correct mis-recognised words.



▶WE ALL MAKE MISTAKES... OR IS THIS SINATRA'S RARE PSYCHEDELIC VERSION OF HIS CLASSIC?

Our test of 'Peter Piper picked a peck of pickled pepper', which is difficult enough to say, let alone recognise, posed no significant problems with any of the packages either, something that wasn't true of their predecessors, which all made a pig's ear of

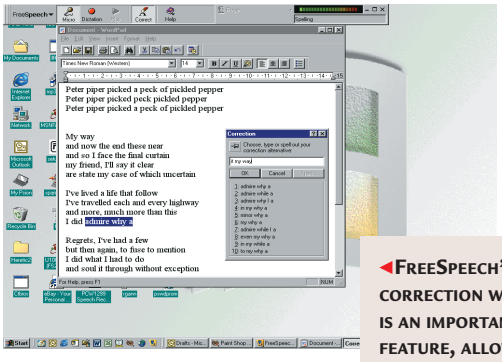
Speech therapy

Text-to-Speech is synthetic or computer-generated speech and has been around for 15 years. Human speech is notoriously difficult to artificially produce and, at the simplest level, a speech synthesiser must emulate the human vocal tract to have the clarity and naturalness of human speech.

Early attempts included the formant TTS engine, which created totally digitised or synthetic speech, with no human recordings being used. However, the sound results were poor, sounding very much like a cheap sci-fi robot.

A technique currently popular is to store actual segments of speech and build the voice from those. Phonemes are the smallest units of speech that distinguish one utterance from another. Smaller speech segments, known as diphones, are obtained from recordings of these phonemes. Diphones contain all co-articulation effects that occur for a particular language and are concatenated (or linked together) to produce words and sentences. The use of diphones, in combination with various synthesis techniques, produces speech that is intelligible and requires relatively little computing power.

The undoubted leader in this field is Lernout & Hauspie, with its RealSpeak engine, capable of generating speech almost indistinguishable from the real thing. It isn't available as a standalone product, but is used by manufacturers to add speech facilities to automated systems such as telephone directory enquiries. RealSpeak is based on concatenation algorithms, using 2MB of human voice segments. The drawback is it costs a lot to implement, as recording a RealSpeak voice requires the speaker to repeat the same text sample in a range of ways and styles.

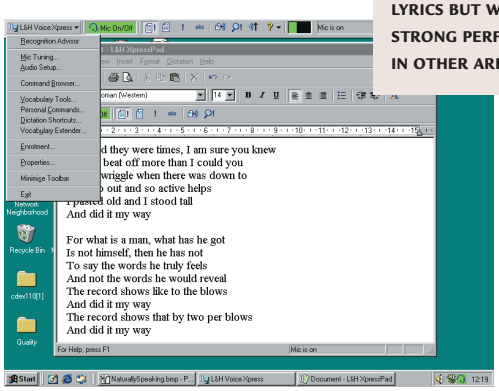


◀FREESPEECH'S CORRECTION WINDOW IS AN IMPORTANT FEATURE, ALLOWING YOU TO HIGHLIGHT AND CORRECT AS YOU GO

recognising the tongue-twister. Most had trouble with 'peck', but after training that word, they managed the tricky phrase with aplomb.

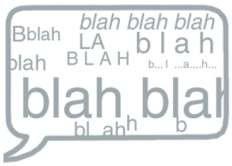
Their performance on less conventional material was more varied: we read out some song lyrics, Ol' Blue Eyes' *My Way*. ViaVoice scored the highest here, with VoiceXpress coming last, consistently thinking it was 'My wife'!

As initial recognition scores were so high, typically around 96 per cent, additional training didn't significantly improve on these – at best you'd get another two per cent boost in accuracy. A better strategy would be to add to your vocabulary, training new words as necessary.



▼VOICEXPRESS FELL DOWN IN THE DIFFICULT FIELD OF LYRICS BUT WAS A STRONG PERFORMER IN OTHER AREAS

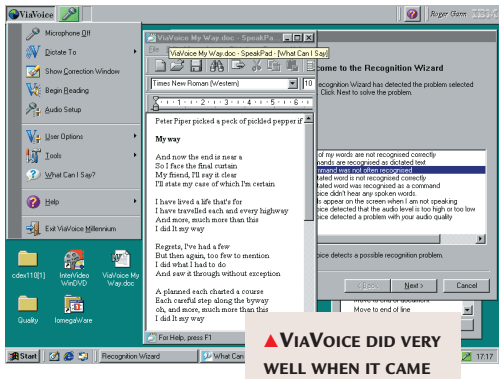
Correcting mistakes



It's particularly important to correct all mis-recognised words – if you don't, the package will assume it has got it right and will repeat the mistake. So, the ease with which corrections can be made is an important feature.

ViaVoice Millennium, FreeSpeech 2000 and VoiceXpress 4.0 all have a Correction window which you can keep open and simply highlight a wrong word and correct it immediately. This is fast and convenient, although VoiceXpress can't correct a pair of incorrect words eg recognising 'himself' as 'in self'.

The products' correction tools all worked in much the same way. It's probably easier to correct mistakes, as you make them, rather than waiting until the end. With the exception of VoiceXpress 4.0, the packages record what you say and can play it back, which is useful where a word is ambiguous. VoiceXpress 4.0 can actually read back your text to you, in a synthetic female voice via your loudspeakers, but this is no use when it comes to correcting text – suppose you said 're-evaluate' and it is recognised as 'Rio value weight' – VoiceXpress will just say 'Rio value weight' back to you.



▲VIAVOICE DID VERY WELL WHEN IT CAME TO RECOGNISING OL' BLUE EYES' WORDS

Command and control

All four packages support command and control, which allows you to control the Windows desktop or any Windows application simply by saying the menu names and menu choices,

Speech recognition hardware

All speech recognition products come with a microphone headset. Ordinary desktop microphones are just not good enough – in order to achieve high levels of accuracy, the microphone has to be close to the 'horse's mouth' to reduce the impact of ambient noise.

It is possible to use alternative input devices. Philips sells the SpeechMike, (£70) a combo trackball, speaker and mike that mimics dictation machine handsets of yester-year, while Plantronics <www.plantronics.com> sells a range of fancy stereo microphone headsets. It's also possible to obtain cordless radio headsets, very useful if you have to keep leaving your PC, but these tend to be quite pricey.

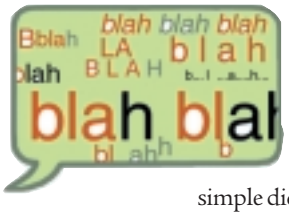
For those on the go, there are digital dictation devices that can record your pearls of wisdom and then input them in to a speech recognition package upon your return to the office.

L&H sells VoiceXpress Mobile Professional (£180) that includes the Olympus DS-150 digital voice recorder plus VoiceXpress itself. Dragon Systems has a similar deal with NaturallyMobile (£200) although this doesn't have as good a spec as the Olympus recorder.

Dragon also sells the £69 (ex VAT) NaturallyClear USB System H100. This is effectively a USB-based external sound card that can be used by any PC with a USB port. Dragon claims that it delivers the highest-quality speech input of any device tested by the company.

Telex has similar products on the way. These are microphones that digitise your speech and send it to the computer's USB port, bypassing the computer's sound system. Telex claims the new microphones will deliver a better-quality signal than microphones plugged into a standard sound card.

Vox pops for desktops



such as 'file menu' and 'export' to start an export action.

ViaVoice, VoiceXpress and NaturallySpeaking offer their own speech-enabled WordPad look-alike for simple dictation task (which is a good idea), but all four can input dictation directly into a range of major Windows applications. All the major players are supported, such as Microsoft Office 97 & 2000, Corel WordPerfect 8 & 9, Microsoft Outlook 97, 98 & 2000.

New this season is command and control of Internet Explorer – you can enter web addresses, navigate through pages and links, select check boxes and enter text into forms.

Generally, if you say what's on a menu, then that task will be actioned – both ViaVoice and VoiceXpress allow 'modeless' operation, that is you can dictate and issue commands in the same breath, so long as you insert a slight pause between them. Simply parroting the menu structure can make for a very stilted way of working and to this end, VoiceXpress, ViaVoice and NaturallySpeaking employ natural language commands, which let you issue commands in different ways.

For example, to change the font size in Word, you'd have to say 'Format, Font, Size' – with natural language, you highlight the word or letter and say 'make it larger' or 'increase the font size by one point.' Sadly, these natural language commands are largely confined to Office 97 and Office 2000 products. Of the three packages, VoiceXpress 4.0 was the clear winner when issuing natural language commands, offering a wider range of controls and customisation than its rivals.

PCW DETAILS

★★★★★
Dragon Systems NaturallySpeaking Preferred 4.0
Price *£130 (£110.64 ex VAT)*
Contact *Dragon Systems*
01628 894 150
www.dragonsys.com/

★★★★★
IBM ViaVoice Millennium Standard
Price *£40 (£34.04 ex VAT)*
Contact *IBM Speech Systems*
01705 492249
www-4.ibm.com/software/speech/

★★★★
L&H VoiceXpress Professional 4.0
Price *£120 (102.13 ex VAT)*
Contact *Lernout & Hauspie*
0800 056 0539
www.lhsl.com/

★★★
Philips FreeSpeech 2000
Price *£79.95 (with headset) (£68.04 ex VAT) £124.95 (with SpeechMike) (£106.34 ex VAT)*
Contact *Philips Speech Processing*
01206 755504
www.speech.philips.com

Top of the vox pops

Speech recognition software has come a long way in 12 months – the products we looked at are significantly better than their predecessors. Not only do they install and enrol quicker, but they're easier to use, while at the same time offering a higher level of accuracy. Command and control functionality has improved and it's quite possible, albeit with a little perseverance, to completely control a PC via the spoken word – it does help if you're using the Microsoft Office suite, though. On the downside, a lot of these gains depend on you having a PC with a fast processor and plenty of RAM: nothing less than a 300MHz Pentium II/Celeron/K6-3 with 128MB of RAM.

All four packages pretty much delivered what they promised. The cheapest package, FreeSpeech 2000 was marginally out-performed by its rivals but its USP is its language support and this, coupled with its low price, makes it a bargain.

ViaVoice Standard Millennium was just as accurate as VoiceXpress and NaturallySpeaking but this was not the 'full' version and so not strictly comparable. However, if you're not interested in command and control, at £40, it's remarkable. We await the Pro version with interest.

Of the remaining two, the top spot has to go to NaturallySpeaking Preferred 4.0, which offered the best combination of features, combined with accuracy and ease of use. But it's beginning to look stale and its interface needs a makeover. VoiceXpress offers marginally inferior accuracies but superior command and control functionality – if you want to dictate consider NaturallySpeaking Preferred, if you want to control your PC as well, go for VoiceXpress Professional 4.0. ❐

Voice control of applications

Voice control is the natural extension of dictation technologies. Not only does it allow people who can't type to use PCs, but future applications of the technology will turn up in computer hardware that doesn't have a keyboard. As device sizes shrink and handhelds become more powerful, speech will be the only practicable way to control them – there simply isn't room for a keyboard.

Starting with Windows 98, Microsoft has included the Speech API (SAPI) as part of the operating system. Previous attempts at voice control had essentially inserted interpreted commands as though they had originated from the keyboard.

This method was a bit of a kludge so in 1995 Microsoft developed the Speech API, currently in its fourth version. Based on the COM specification, SAPI provides an API abstraction layer between applications and speech technology engines, allowing multiple applications to share speech resources on a computer and avoid the need for writing specialised application code for a specific speech technology engine.

However, voice control isn't all it's cracked up to be. For a start, talking all day to a PC can be very tiring (if not tiresome). It's not suited to all tasks either – it will actually take longer to vocally command many 'one-click' tasks. It also makes for a noisier office environment and listening to someone at the next desk drone on and on to a computer will be tantamount to torture. The looming advent of voice synthesis will only serve to compound the problem.