



# Softly spoken or hard of hearing?

Is the latest crop of speech recognition software a real improvement, and what does it have to offer the translation industry?

by Michael Benis



## **Painless productivity but no magic**

The most frequent question I get asked about speech recognition is: "Does it really work?" The answer is a qualified yes. Ask anyone who has persevered long enough to deploy speech recognition successfully and they'll sing its praises. There are a number of reasons for this. Firstly, it quite simply makes life much more comfortable because you're not physically tied to your keyboard or screen. You can move around

**Although these systems are now very much easier to use, you shouldn't expect them to make any more of an immediate difference to your life than buying a car if you've never learnt how to drive.**

in your chair or even walk about while dictating your translations, correspondence or other documents, leaving you much less likely to feel tired after a hard day's work. What's more, because your hands are free, you can use them to hold your source text or any other documents you need to refer to, making it much less likely that you'll lose your place and lose time as a result.

When it comes to productivity, however, the big advantage is that most people can talk much faster than they can type. But that's not all. Speech recognition obviously frees you from your keyboard and mouse, leaving you much less exposed to the vagaries of repetitive strain injury while enabling sufferers to continue working despite its debilitating effects. That's not just good news for the competent technical translator who may be turning over many thousands of words per day, but anyone employing translators too, in terms both of getting the most out of their employees and protecting themselves against possible lawsuits.

It looks like a heavenly situation, doesn't it? Unfortunately, however, achieving heaven is, as always, just a little bit more difficult than describing it. Speech recognition software requires a powerful system to perform at its best (if you're buying new, don't go for anything less than a 1 GHz processor with 512 MB RAM). Fortunately, this sort of processing power is increasingly easy to come by. The bad news

is that speech recognition requires something else, which is becoming increasingly rare: patience. Although these systems are now very much easier to use, you shouldn't expect them to make any more of an immediate difference to your life than buying a car if you've never learnt how to drive. If you've never dictated before, you'll need to learn how to start thinking in whole sentences at a time, not words. You'll also need to learn how to speak fluently but clearly, enunciating every word without pausing between them; (just think of your favorite newscaster). As if that weren't bad enough, you'll need to experiment with microphone types and positioning as well as learn all the various different program commands and options before you start using any of the systems available on the market confidently and with any substantial benefits. Training can make an immense difference, helping to get you up and running both reliably and speedily, but it doesn't come cheap—often costing more than the price of the software itself. I know many a translation professional who has been highly skeptical of whether that investment is worthwhile only to subsequently abandon speech recognition as hopeless. On the other hand, I have yet to meet anyone who actually invested in training and regrets having done so. One

**One way or another, whichever choice you yourself take, you'll have to factor your own learning curve into the equation, in terms of both time and money.**

way or another, whichever choice you yourself take, you'll have to factor your own learning curve into the equation, in terms of both time and money.

### **All ears and no brain**

Half the battle of understanding how to use speech recognition lies in understanding how the programs themselves actually work. That in some ways is very easy, because despite their sophistication, these systems actually have very simple minds. As many a nervous translator has been delighted to discover about MT, the good news is that computers can't think. You can dictate what you like, but they won't understand a word. All they can do

is listen. They mainly use the sound of a word to recognize what it is, with varying results because they don't have the intelligence to filter out differences in pronunciation or the acoustic qualities of different voices precisely because they don't understand the meaning of the words. Consequently, whichever package you choose, you'll have to teach your system how to recognize your voice by

**As a result, they're much less likely to slip up when we tell them we rode our bikes to the boathouse where we rowed our boats.**

reading a passage it already knows. All the systems on the market require you to do this before you can use the program. They then build up a "speech file" of your own individual speech patterns to recognize the words you pronounce.

That, however, is not enough for a speech recognition program to achieve consistent results, partly because of mischievous homonyms ("Which witch is which?") and partly because we never enunciate every word perfectly or pronounce them exactly the same way every time. Speech recognition programs try to get round this by analyzing the statistics of our word usage, which is to say the probability of any given words being preceded and followed by any other given words. As a result, they're much less likely to slip up when we tell them we rode our bikes to the boathouse where we rowed our boats. All the packages now include a utility that allows you to feed in a large number of old document files for analysis so that they can rely upon good contextual information right from the very beginning, although recognition accuracy will continue to improve as the systems learn over months of use. This makes it very important for you to back up your user files on a regular basis. Once you've got used to the joys of successful speech recognition, the last thing you'll want is to have to start all over again from scratch.

### **Building correction into your life**

Even the best speech recognition systems make mistakes, just like the best typists. The only difference is that a spellchecker will never pick up any of the mistakes that a speech recognition program makes.



Every word will be spelt perfectly. It's just that the misrecognized words won't be the words you want. Once you've perfected the intonation and cadence required to get the best performance out of your system, you ought to be correcting no more than about one word in every twenty to thirty or more. Whatever you're dictating, you'll want to check your first draft, which is when you can check for misrecognition errors. Take care to correct them using the correct function, which helps the program to continually improve. Use its other functions to select text and dictate over it to make any improvements due to a change of mind. Translators will find that this correction process integrates seamlessly with the check they perform to monitor the completeness and terminological accuracy of their first draft. If speech recognition is used correctly, it should not intrude on the translation process.

### Where to put your money

If you've never used speech recognition before, the best place to put your money is in the hands of a specialist reseller who can check over your system components (including the vital soundcard) and teach you what it takes to get up and running effectively. There are currently only two packages on the market that are worth considering and in some ways learning how to use them correctly will have a greater effect on your results than which system you choose. Just think of my automotive analogy. If you've never driven

**If speech recognition is used correctly, it should not intrude on the translation process.**

before, it doesn't matter what car you buy: you won't get very far without good driving lessons, will you? The situation is, of course, a little less dramatic with speech recognition. After all, if you simply go ahead and experiment, there's no risk of your actually killing anyone, though you might try to strangle your computer. There are also some very helpful user groups on the Internet (mostly at [www.yahogroups.com](http://www.yahogroups.com)) that can help make the learning curve a little more comfortable albeit no less steep. Still, while there's no doubt that anyone who's moderately confident with computers can "eventually" learn how to get the best out of speech recognition, you should bear in mind that

"eventually" can prove to be a very long word. Effective training should mean that you're spending your time making money rather than spending time trying to save it.

### A two horse race

Not so long ago, you could choose from an array of four mainstream products, but now things are much easier. You've been left with the two market leaders: Dragon Systems and IBM, both pioneers in the field with experience second to none. We have tested the latest versions thoroughly to help guide you in making the most appropriate choice. Here is what we found.

### Dragon NaturallySpeaking

Once upon a time, speech recognition required you to talk like a robot, which didn't make it easier to learn. All that changed with the advent of Dragon NaturallySpeaking, the first "continuous speech" product, which finally allowed you to talk much more like a human being (or at least a newscaster). The current version is the sixth incarnation of the product and incorporates a series of significant improvements. Unfortunately, it also incorporates a series of irritating bugs. Worse still, a number of the latest improvements are not only of debatable value but also degrade performance in other areas. If you're a happy user of version 3.52, 4 or 5, you might well be happier sticking with your mature speech files and a system with which you're familiar while waiting until a bug-free version comes out.

But let's take a look at what you get for your money. NaturallySpeaking comes in a variety of different packages. Serious users should consider only two, known as *Preferred* and *Professional*. The crucial difference comes with *Preferred*, which allows you to play back your speech when you make a correction. This is important because when NaturallySpeaking makes a mistake it can, like any speech recognition program, come up with something that bears absolutely no resemblance to what you said. If you can hear what you said (just as if you were using a Dictaphone), you can make your correction immediately without losing precious time scratching your head while trying to remember. The *Professional* version adds a number of further functions, the most important of which is the ability to create complex control macros that can be very handy for integrating speech recognition with any of the major translation memory packages on the market. Note, however, that there is a significant jump in price be-

tween these two versions. Also, although you can't create speech macros in the *Preferred* version, you can speak keystroke shortcuts (e.g. by saying "Press control key down arrow"), which allows you to control most translation memory products although it can cause damage to the TRADOS tags in Microsoft Word, making the *Professional* version preferable for this product.

### Great out-of-the-box performance

NaturallySpeaking now comes with a cheap but respectable Plantronics microphone and a clear and comprehensive manual, alongside the usual installation CD. The installation process itself is organized in a series of easy-to-follow stages that help you position your microphone correctly, set your system's audio parameters as required and dictate a short passage that enables the system to familiarize itself with the way you speak. One of these passages, "Talking to your Computer", also provides useful information on how to get the best out of speech recognition. Version 6 includes a number of features that are designed to make the system easier to use. The correction window, which presents a selection of phrases as an alternative to that which has been misrecognized (one of which is normally right) now highlights the different words in the options available, making it easier for you to choose the

**Nothing But Speech has to be one of the most impressive examples of someone ignoring the old adage of "if it ain't broke, don't fix it".**

correct one quickly. You also get a Command Browser that reminds you of the commands available, while the professional version includes a Macro Recorder that makes it easier to create macros for automating your work and integrating speech recognition with other programs. Lastly, the various different functions available for improving accuracy and other tasks are grouped in logical families, which makes it easier for an inexperienced user to find just what they need.

Unfortunately, though, this is where the bugs begin. NaturallySpeaking features by far the best system for analyzing old work and building up a good vocabulary of all



the words you use and the context (statistics) in which they occur. Sadly, you're unlikely to find it unless you know it's there (the easiest solution is to simply say "build vocabulary"). The default tool provided in the Accuracy Centre group simply adds the words to your vocabulary without the contextual information, meaning it has a much less dramatic effect on improving recognition accuracy. Another irritating bug is that the automatic playback feature doesn't always work reliably and, indeed—unlike previous versions—wouldn't work at all in most translation memory programs. That's a major fault, since despite its failings NaturallySpeaking provides by far the most reliable performance for this purpose.

Perhaps the most useless "enhancement" in this particular version, however, is something called "Nothing But Speech", which is designed to filter out the funny little

**There's no doubt that for straight dictation with minimal formatting you can get good results from both systems, especially in their dedicated word processors.**

noises such as "ums" and "ahs" that we make while thinking out our more coherent utterances. Does it work? Frankly, I wouldn't know. Nothing But Speech has to be one of the most impressive examples of someone ignoring the old adage of "if it ain't broke, don't fix it". The bottom line is that NaturallySpeaking was one of the least problematic programs when it came to inserting strange words due to breathing or generally failing to keep one's mouth shut while thinking. On the other hand, it was a program that allowed you to dictate at a tremendous rate and would still keep up. That's no longer the case. You now have to dictate at a much more measured pace and that, to my mind, is very far from an improvement. Also, while you can still correct long sentences in one fell swoop (which you can't with ViaVoice), they have to be much shorter than with previous versions. But perhaps the most irritating bugs are those that have shamelessly survived from version 5, including the fact that it's impossible to turn underline off using speech commands in NaturallySpeaking's dedicated word processor, DragonPad, or that you can't interrupt

playback during correction to select the option you require by voice, which allowed one to make much faster progress in the past.

I will not go on listing all the various bugs and other "features" that make the latest version of NaturallySpeaking a very much less productive product than it could have been. The reason's simple: for all these extremely irritating failings, Dragon is still very much the top dog. It's just that this version should have been so much better. Let's hope the next one fulfils all its promise.

### IBM ViaVoice

Although IBM's manual lacks the clarity of Dragon's, which sets new standards, it's certainly comprehensive enough. The rest of the package simply can't be faulted, especially the Pro USB version. This not only comes with an excellent Andrea noise-canceling microphone, but a matching USB sound pod into the bargain, which ensures consistent performance irrespective of the quality of your soundcard or, indeed, whether you have one at all. This, apart from anything else, makes the system excellent value for money, especially if you consider that it comes with macro capabilities that are broadly comparable to those in the Professional version of NaturallySpeaking, which costs over three times more.

ViaVoice also has an easy-to-follow installation procedure and setup routine that takes you through all the necessary stages step by step. All in all, it's a neat and admirably complete package. Like NaturallySpeaking it not only allows you to dictate into other packages but also comes with its own dedicated speech recognition version of Microsoft's WordPad, in which it provides fast, sure-footed and accurate performance, although more training is required to bring it close to NaturallySpeaking's level of recognition accuracy. Where the program falls down is the ease and predictability (or the lack of it) with which other programs can be controlled or dictated into directly.

ViaVoice also has its quirks, including an "agent"—an animated talking pencil called Brocklee (don't ask me why, I know it sounds like a vegetable)—who pops up at regular intervals in a brave attempt to push every user's sanity to breaking point. The first thing you need to learn when using ViaVoice is how to switch him off; (go to "ViaVoice Options", click the "At Startup" tab and uncheck the "Display

agent" and "Display Recognition Wizard" checkboxes). Other unique features are much more useful, including an automatic spellchecker that stops you misspelling corrected words, but doesn't prevent you from adding any unknown foreign or "specialist" words to your vocabulary.

ViaVoice version 9 is not a giant step forwards from version 8, only really offering shorter enrolment (the initial training in which you read one or more passages to the computer) and full compatibility with Microsoft Windows XP. That said, it's a competent and admirably complete package which offers outstanding value for money. This is especially true of the Pro USB version that offers an excellent guarantee of compatibility with all systems—including laptops, which sometimes have cheap, interference-prone sound chips that can significantly degrade the performance of any speech recognition package.

### Which horse to back

I've deliberately used a betting analogy here, but not because anyone interested in speech recognition will be gambling on the products: in the end you'll be gambling on your own perseverance. There's no doubt that for straight dictation with minimal formatting you can get good results from both systems, especially in their dedicated word processors. NaturallySpeaking edges ahead here, delivering high accuracy without the need to read lots of training scripts. It also offers a wealth of navigation and editing features that significantly enhance productivity. Where NaturallySpeaking really gallops into the lead, though, is when it comes to integration with translation memory, delivering more accurate and reliable performance. Unfortunately the playback bug means that the best solution for this role continues to be version 5 of the product, which is luckily still available in the shops, where it is now generally offered with a very attractive reduced price tag. One word of warning, though: If you're already using Windows XP, either "downgrade" to Windows 2000 Professional or go for NaturallySpeaking version 6.

---

*Michael Benis is a freelance communications consultant, copywriter, journalist and translator. Visit [www.michaelbenis.com](http://www.michaelbenis.com) for more information.*