

Speech Recognition Finds Its Voice



Wells H. Anderson

President, Active Practice LLC

Helping lawyers leverage technology

<http://www.activepractice.com> Tel: 800.575.0007

Toronto LegalTech 2001 – November 14, 2001

Note: Much of the advice in this article about working with speech recognition software remains valid; however, the specific product information is obsolete. Dragon NaturallySpeaking continues its hold on first place in this product category.

Wells H. Anderson – January 2006

Introduction

Session Description

Simply talking to your computer holds the promise of being so much easier than fussing with a mouse and a keyboard. Indeed, recent innovations like continuous speech recognition are bringing us closer to a practical, voice-based interaction. But much work remains to be done before the only input device you will need is a microphone. This session will give you a thorough grounding in the state of the art of voice recognition.

Does It Work?

The biggest question most people have about voice recognition is: "Does it really work?" Most of us are skeptical, figuring, sure, it works for experts and for carefully trained company representatives, but what about for regular people? Won't it be more trouble than it is worth?

My response sounds all too familiar: It depends. The good news is that it depends on a lot less than it used to. The bad news is that it still takes a significant investment of time and money. Those who invest wisely will receive far higher returns than the stock market ever offers.



Voice recognition technology seems deceptively simple: You talk; it types. What that simple view omits are the many differences between the creation of oral and written communications. In order for you to determine whether voice recognition may work for *you* or for *your staff*, you need to know more about the nuts and bolts of writing with a microphone, software and a computer.

Here we consider the use of voice recognition from a number of different perspectives. How do the competing products differ? What hardware do you need? What are the different skills you need to have or learn? What features of the voice recognition software will you need to use? What factors affect how efficient you can become?

Competing Products

Dragon NaturallySpeaking by Dragon Systems pioneered affordable continuous speech recognition. <http://www.lhsl.com/naturallyspeaking> It competes with IBM ViaVoice by IBM Corporation. **[IBM ViaVoice is now obsolete. Jan. 2006]** We have counted VoiceXpress out of the running since its parent, Lernout & Hauspie, now owns and promotes NaturallySpeaking.

In the fast-changing world of technology, we expect improvement every year. IBM ViaVoice did some catching up in 2000. Release 9 now works with the Windows XP operating system. That upgrade is important because Windows XP and Windows 2000 work much more reliably than Windows 95, 98, or ME. It is also important because of the questionable quality of the speech recognition technology included in Windows XP.

Unfortunately, many annoyances still plague ViaVoice. Some of them have survived upgrade after upgrade. We find that disappointing. ViaVoice achieves the high accuracy essential to success. But effective voice recognition software must be more than accurate; it must be smart and helpful in punctuating, formatting and correcting your documents. Imagine the difference between working with an assistant who anticipates what you want and one who repeatedly makes small errors that you must correct yourself.

IBM has done a nice job of extending the feature set of ViaVoice. Unlike NaturallySpeaking, ViaVoice works not only on various Windows versions, but also on Apple and Linux computers. Despite these advances, we rank ViaVoice behind NaturallySpeaking because of annoyances that interfere with the productive creation of documents.

IBM ViaVoice

Obsolete. ViaVoice is no longer in development. Jan. 2006

Installation and Setup

ViaVoice Software

Installing IBM ViaVoice software took ten minutes and went smoothly. It required the usual steps: Enter your name and company, fill in a serial number, choose an installation folder, and register the product. Like other IBM software

products, ViaVoice loaded an enormous number of files, 802, and took up 170 MB on the hard disk. Thankfully it added only three entries to the Windows registry. A reboot was required to activate ViaVoice.

Hardware Setup

When you first open ViaVoice, a User Wizard and begins a multi-step process for setting up the program and creating “ voice model” that it uses to recognize what you say. First you choose sound input and output devices from those installed on your system. On ours, a USB microphone was the correct input device. Wizard windows walk you through setting up the microphone, including positioning it, testing for room noise and testing the recording level.

Up to this point in the materials, I have been attempting to dictate into Microsoft Word 2000. Unfortunately, ViaVoice repeatedly drops letters from dictated words. This problem seems to be specific to Microsoft Word 2000. After spending an hour on the phone with a committed but the exasperated technical support person, she stated that the only solution for the problem would be to do further voice training to improve the voices recognition accuracy. That explanation was incorrect. Voice recognition software is designed so that it can only spell words they way they are spelled in its dictionary. for example, it cannot spell the word "receive" without the letter "i." But that is exactly the sort of error that ViaVoice was making.

Creating A Voice Model

With voice recognition software needs to adjust to your pronunciation and dialect. Surprisingly, it can manage wide individual variations. You need to 10 to 15 minutes to read a story while the ViaVoice records your voice. ViaVoice then needs five to 20 minutes to process the results. Reading a second story for 25 to 35 minutes improves ViaVoice's accuracy. It needs 20 to 45 minutes to process the second story after you finish reading it aloud. In my view, IBM made excellent story choices, Treasure and ghost story by Mark Twain. Part of the process involves reading some strange, fun senses to assist ViaVoice and identifying related sounds. For example: “Loyal Lloyd employs alloys for exploit, and uses a scale to measure his treasure.”

The User Wizard halted a few times on words that it did not recognize. Repeating the word usually got the wizard going again, but sometimes it repeatedly beeped regardless of my efforts to pronounce the word it was stuck on.

To work around this problem, click the Back button, then click on the story again. You will be returned to a spot near where the wizard halted and then you can continue.

Since continuous reading can be tiring, it is good that the wizard allows you to pause at any point - mid-sentence or not. If the phone rings, you can click on a

pause button so that the wizard won't be confused by what you say. The "Click to Resume" button gets you going again.

The hardest practice to master is dictating everything in your normal voice. You will be tempted to alter your voice, subtly or not, in a misguided effort to prevent the ViaVoice recognizer from making mistakes. Ironically, mistakes are good! Each time you say something normally, ViaVoice makes a mistake, and you correct the mistake, ViaVoice gets better. That is an important part of the process of training ViaVoice.

User Training

With most software programs, the user needs to learn one new approach to working with information. To learn the Web, the user masters hyperlinks, Web addresses, forward and back buttons, and maybe some search engine rules. To use a spreadsheet program, the user learns about columns, rows, formulas, sorting and perhaps a bit about graphs. Often skills learned in one program carry over to another.

To become quite comfortable with voice recognition software, a new user must move up a number of learning curves simultaneously. Skills carry over from the world of word processing, but many of them have very unfamiliar new twists.



IBM ViaVoice Voice Center

Speaking Naturally

Speaking naturally does not come naturally to new users of the voice recognition software. If left to our own devices, we tend to over-enunciate, slow down, speak more distinctly, and change our pronunciations of words when voice recognition software does not perfectly recognize what we say. Each of these behaviors is self-defeating. Speaking naturally is the goal.

To learn the effective use of voice recognition software, new users need the help of experienced instructors who will listen to them, helping them to avoid problematic adjustments. Besides teaching users to speak truly naturally, instructors have a number of other important skills to develop in their students.

Navigation

We're used to the mouse and cursor arrows, but to dictate rapidly with voice recognition software, we're better off using voice commands to navigate around our documents. To do so, we need to memorize a separate vocabulary and also get used to pausing briefly before and after dictating a navigation command.

IBM ViaVoice comes with a Quick Reference Guide divided into 11 sections. Relating to navigation are:

- Cursor Movement
- Desktop Navigation
- Internet Navigation

These sections are not intended to be comprehensive, but rather list the most frequently used commands. For example, 10 commands are listed under Cursor Movement, covering movement by word, line, page and beginning and end of document.

In addition to the navigation functions that duplicate the use of the cursor arrows, Home, End, Page Up and Page Down keys, voice recognition programs need to duplicate the mouse user's ability to directly select a word. ViaVoice offers this capability with the "Select <text>" command and the "Select this" command.

"Select <text>" should move the cursor to a previous occurrence of the one or more words spoke by the user after saying, "Select." In practice, this function may prove problematic. The user must pause briefly after saying, "Select." Speaking too quickly caused ViaVoice to type the word "Select" and the <text>. Another problem is that the ViaVoice often jumps to a wrong word, and then seems to balk at moving from that word in response to further "Select" commands. Finally, there is the unavoidable, occasional problem that the word you want to select and ones that sounds like it occur in multiple places in the preceding text. ViaVoice lets you move from a false hit to the next occurrence by saying, "Try Again." That moves the selection highlighting to the next occurrence of the word, but does not help if the word you really want is a homonym rather than another literal occurrence of the currently selected word.



Punctuation

Attorneys experienced with dictating may assume that lawyers with more than a few years of experience all know how to dictate. Currently, it's safer to assume that they are experienced typists rather than experienced with a Dictaphone. That means the process of inserting punctuation verbally is a new skill that most users need to learn in order to effectively use the voice recognition software.

Commands for punctuation marks like "period," "comma," and "exclamation point" are easy to remember, but hard to remember to insert consistently when speaking in sentences.

Using Help

ViaVoice comes with excellent, online help tools. Help screens are accessible using voice commands. “What can I say?” is the command you use to find out your options. You can further narrow the Help that will be displayed by saying, for example: “What can I say for Cursor Movement?”

Because of the broad set of commands, it is important for users to master the use of Help so that they won’t waste time finding new commands they need.

The complete ViaVoice documentation is available as an Adobe Acrobat PDF file that can be copied to the hard disk as a part of installing ViaVoice.

Formatting

Though the demands of formatting a legal brief are heavy, even a relatively short letter may contain italics, special capitalization, indented paragraphs and a list or table. ViaVoice automates part of the process of formatting, capitalizing the first letter of each sentence, putting a space after each period, and capitalizing a surprising variety of proper nouns that it recognizes. But much of the formatting is left up to the author and microphone.

Simply to capitalize a name such as “Drew” requires that the user say: “Capitalize on drew capitalize off.” Inserting paragraphs and line breaks is easier. Just say, “New line” or “New paragraph.”

A new feature available in ViaVoice is Natural Commands, which are available in MS Word, MS Excel, and MS Outlook. You need not say a command in precisely the right order for it to work. Natural Commands are mentioned below in Special Features.

Correction

“Scratch that” becomes an important part of a voice recognition user’s vocabulary. It erases whatever ViaVoice just typed and lets the user try again. But overuse of “Scratch that” will slow down the process of ViaVoice adapting to an individual’s voice and pronunciations. “Correct that” is the command ViaVoice needs to hear in order to improve its accuracy. This command is discussed further, below.



ViaVoice Correction Window

ViaVoice offers a good method for entering words with unusual spellings. Say “Spell mode,” then say each letter of the word continuously. When done, say “Return” or “Cancel” to return to normal dictation mode.

Whether the user can quickly and accurately make corrections may well determine whether the user will persist in using the product and will realize a true increase in efficiency. This aspect of using voice recognition software does not necessarily come quickly or easily. So here is another area where skilled instruction can make the difference between success and failure.

Speech File Building

Though ViaVoice is surprising accurate out of the box, individual speech files are used to improve its accuracy. Speech files are individual to each user. They are started during the initial “break-in” process during which the user reads passages from a story. They are refined as corrections are made to misrecognized words using the Correction window. That is how ViaVoice learns to recognize words accurately.

Correction By User

As a user dictates new documents at a computer, the correction window pops up in response to the command, “Correct this.” ViaVoice presents a quick list of the most probable alternatives to the incorrect word wants to correct. Choosing one is as simple as saying, “Pick 3,” to pick the third word on the list.

If the word you want does not appear in the correction list, you can try re-dictating. Otherwise, you must resort to the keyboard to type in the word. Surprising, ViaVoice does not have a “Spell Mode” that would allow you to spell out the word in the Correction window. It has this mode outside the window. Dragon NaturallySpeaking seems to do a better job allowing re-dictation within Correction window and does allow the user to spell out a word in that window.

Correction By Assistant

A big advantage that ViaVoice has over Dragon NaturallySpeaking is its option for storing voice recordings right with the text version of a dictated document. This feature permits an assistant more power and flexibility in working with a dictated document.

Rather than having a busy attorney lose billable hours on the rather tedious process of correcting misrecognized words, an assistant can perform the same function.

First, the attorney dictates a document, saving the document and associated voice recording file on the network. Next, the assistant, who also has ViaVoice, proofreads the document, stopping at misrecognized words and correcting them with the Correction window. In order to determine accurately which word the attorney wanted, the assistant can replay the recording of the attorney's voice saying the word in question. The assistant must then make sure that the improved speech files for that attorney are copied back to the computer that the attorney uses for voice recognition. This process results in increasing ViaVoice accuracy while sparing the attorney from the time-consuming extended break-in period.

Important Note: For all dictated documents that go outside the law office, weigh very carefully the importance of having a second set of human eyes proofread them. Try as you will, if you dictated a document, you will not find all of the words that ViaVoice or any voice recognition software has misrecognized. These errors are insidious. They don't show up in a Spellchecker, since voice recognition software cannot make classic typos; it can only use words that are in or have been added to its dictionary. The erroneous words will be especially hard to spot, since they "sound like" the correct word. Misrecognized words that you don't catch can be *very* embarrassing. Rely on a second person to proofread anything important.

Special Word Lists

To accelerate the process of improving the accuracy of ViaVoice, the user can draw on any of several resources to build the ViaVoice dictionary of words it recognizes. A special legal version is available at extra cost. A built-in feature allows the user to add up to 64,000 words to ViaVoice's personal vocabulary.

- **Analyze Documents**

From Tools / Analyze My Documents, the user can point ViaVoice at a collection of his or her documents. ViaVoice will scan through the documents, identifying words that do not yet exist in its vocabulary. The user then has the opportunity to train ViaVoice to recognize these words.

- **Find New Words**

Instead of using the Correction Window to correct text as you dictate, the user can make corrections with the keyboard during or after dictation. After the text is

corrected, the user can have ViaVoice analyze the document to find new words or phrases that need to be trained.

Special Features of ViaVoice

Natural Commands

For MS Word, MS Excel and MS Outlook (from Office 97 and Office 2000), ViaVoice supports natural commands for formatting documents. This feature is a big step forward. The commands needed to format documents are so numerous that it is difficult even for very quick learners to master the whole set of literal commands. Natural commands give the user flexibility in the words used and the order in which they are used to communicate commands.

Natural commands should be preceded with the word “Computer” so that they are not interpreted as ordinary dictation. (The word “Computer” will be recognized as a command if the user pauses briefly before and after saying the word.)

For example: “Computer, move to top of page.”

Without Natural Commands, the user would have to dictate, “Move to beginning of page.” The difference may not seem large, but across a wide set of commands, it measurably reduces frustration.

Voice Recordings

ViaVoice has the option to save the audio portion of a dictated document in a Session file. These files take up a good deal of hard disk space, but disk space has become inexpensive. Session files allow attorneys to become even more efficient during the extended break-in period and afterwards by relying on an assistant to handle the correction of misrecognized words. An assistant can refer to both the text and the sound files when proofing a document dictated by an attorney using ViaVoice.

Macros

Users can create special phrases, such as “inside-address,” that, when dictated, will result in the insertion of fully formatted boilerplate text. These are not difficult to create and can become quite involved. The challenge is to organize and document them well so that it will be easy to use them on a day-to-day basis. The macro command itself must be recorded by each user, but the contents of a macro do not have to be recreated. They can be exported from one user to another.

Dragon NaturallySpeaking

Installation and Setup

Dragon Software

Installing Dragon was simple. Insert the CD, click Next a number of times, fill in your name and registration information. It installed only 225 files taking up 97 MB of space, far less than IBM ViaVoice. Surprisingly, it made no entries in the Windows registry and does not require a reboot after installation.

Hardware Setup

Dragon has you read out loud twice for about 15 seconds each time to check microphone volume and signal-to-noise ratio.

Training

As noted above, mastering speech recognition requires an unexpected amount of training. The software needs to be trained to recognize the user's voice accurately. The user must develop dictation skills and master essentially two new languages: screen navigation and document formatting.

The process of training Dragon to recognize your particular speech begins with a choice of five selections:

- Alice's Adventures in Wonderland
- Charlie and the Chocolate Factory
- Dave Barry in Cyberspace
- Dogbert's Top Secret Management Handbook
- Talking to your Computer

You read through the selection so that Dragon can get used to the sound of your voice and match what you say with the words in its dictionary.

Next Steps

Showing its maturity, Dragon NaturallySpeaking offers three choices tailored to users with varying levels of experience.

- Start the Tutorial
- See What's New
- Begin Dictating

The tutorial provides a great blend of a helpful audio track, animated demos, hand and live interactive practice. Starting with the tutorial is strongly recommended.

An invaluable aid in learning the program is the Quick Start pamphlet. Because it is short and full of helpful color diagrams, users are drawn to it. In addition, it has a flip-out Command Quick Reference listing its many commands in intelligent groupings.

Results of Training NaturallySpeaking

The success of the General Training process, reading the story to the computer, determines the accuracy of speech recognition. Surprisingly, we experienced bad results using a supposedly good microphone. The results with a Telex USB microphone specially designed to improve voice recognition performance were unacceptable. In fairness to Dragon, then microphone has been around in a laptop case for the last two years. Switching to the much less-expensive included microphone yielded excellent results.

To give you an idea of the differences between these results, here are the two paragraphs with recognition errors highlighted in yellow:

Telex USB Microphone (well-traveled)

Dragon NaturallySpeaking lessened off your computer instead of-date you can stop the Microsoft Word, corral Word for it, female, chat, and virtually any other programs and normally tight

Included Microphone

Dragon NaturallySpeaking lets you talked your computer instead of typing. You can talk to Microsoft Word, Corel WordPerfect, e-mail, chat, and virtually any other programs in which you normally type.

Special Features of Dragon NaturallySpeaking

Fast Correction Process

The process for correcting recognition mistakes flows more smoothly than the ViaVoice process. Dragon also makes it easy to improve the speech files so that future recognition mistakes will be avoided. The entire process is easily controlled by voice.

Easy Formatting Commands

Dragon shines at formatting text the way we expect it to be formatted. There is even an option for putting either one or two spaces after a period.

Navigation Commands

Our experience with Dragon's navigation commands was very positive. For example, both the phrases "move to end of line" and "go to end of line" had the desired effect.

Dictation Shortcuts

The \$99 Standard version of Dragon allows the creation of one-liner shortcuts such as your e-mail address or firm name. The \$179 Preferred version supports the insertion of boilerplate paragraphs, full addresses and phrases.

Dragon NaturallySpeaking Version 6 Coming Soon

If you are about to buy Dragon Naturally Speaking, be sure that your vendor will upgrade you from version 5 to version 6 at no charge. One such vendor is Freedom of Speech, <http://www.freedomofspeech.com>, a nationally recognized speech recognition vendor and training company.

Hardware

Sound Card

Either a high quality sound card that is approved for your software or a USB microphone is important for a successful implementation of voice recognition. As some users have found out, a good sound card may not be sufficient, especially in a noisy notebook computer. A USB microphone bypasses the sound card entirely and removes a potentially troublesome variable from the equation.

It is unfortunate how many users have become completely disillusioned about voice recognition because of an acoustic problem. Investing \$70 in a USB headset microphone not only can solve the problem, but also can preserve a potentially huge time investment in developing personal speech files through the Correction process.

Speech files recorded on one PC with a USB microphone stand a very good chance of being perfectly usable for dictation on another PC using the same type of USB microphone. The same cannot be said for normal microphones that require sound cards. Speech files are unlikely to be usable for dictation on other computers when recorded through a sound card and normal microphone.

Processor

Consider a 500 MHz Pentium III as a minimum processor for effective voice recognition at this writing. The faster, the better, because the recommended minimum increases each year.

RAM

Consider 256 MB of RAM as a bare minimum. The whole point is to become more efficient. Why save a few dollars, and then find yourself slowed way down.

By investing in 512 MB of RAM, you allow yourself to have multiple programs open and still be able to use speech recognition efficiently.

Microphone

I believe the most important factor in a microphone now is whether it is USB (Universal Serial Bus) or not, our bad experience with an older USB mic notwithstanding. Your speech files will likely be usable for dictation on PCs other than the one where they were created. See Sound cards, above.

Microphone placement is also important. It is best to use a headset so that you can position the microphone consistently close to your mouth and out of your air stream. It is convenient to be able to rotate the microphone away from your mouth and back so that you can drink something without removing the headset.

Hand-held Dictation Unit

The selection of products in the hand-held dictation unit niche is rapidly changing. For anyone considering the purchase of such a unit, I suggest you try out the hand controls in a store. You will be using them a lot, so you'll want them to work well for you.

Voice Recognition Pricing

PC Connection prices for speech recognition software

As of November 2001 <http://www.pconnection.com>

Dragon NaturallySpeaking

Dragon NaturallySpeaking Professional Edition 5.0	\$479.00
Dragon NaturallySpeaking Preferred Edition 5.0	\$164.95
Dragon NaturallySpeaking Standard (at Micro Center retail store)	\$ 99.95
Dragon NaturallySpeaking Essentials (at Micro Center retail store)	\$59.95
Corel WordPerfect Office 2002 Pro (with Dragon NaturallySpeaking)	\$399.95
Corel WordPerfect Office 2002 Pro Upgrade (with Dragon)	\$219.95

IBM

ViaVoice Pro USB 9.0	\$199.95
ViaVoice Standard 9.0	\$99.95
ViaVoice Personal 9.0	\$27.95
ViaVoice Web 9.0	\$29.95

More Resources

Open Directory Project, voice recognition topic:

http://dmoz.org/Computers/Speech_Technology/

Freedom of Speech, national specialist in voice recognition technology

<http://www.freedomofspeech.com>



Wells H. Anderson is president of Wells Anderson Legal Tech Services in Minneapolis. His firm works with attorneys and staff who want to implement better computer systems. Mr. Anderson focuses on tailoring Time Matters practice management software and training to needs of each law office.

Winner of the Legal Technology Consultant of the Year Award for 2000 from TechnoLawyer, Mr. Anderson presents on legal software topics at international conferences and participates actively in legal technology listservers.

Wells Anderson

Active Practice LLC
Edina, Minnesota USA

phone 800.575.0007

web <http://www.activepractice.com>

Helping lawyers leverage technology

© Copyright 2000 Wells Anderson