

United States  
Department of  
Agriculture

Forest Service

National Technology &  
Development Program

1019 1808—SDTDC

Inventory and Monitoring  
November 2010

# Voice-Activated Personal Data Recorder/Field Data Recorder (PDA/FDR) or Laptop Application



|  |  |  |
|--|--|--|
|  |  |  |
|  |  |  |
|  |  |  |





## Voice-Activated Personal Data Recorder/Field Data Recorder (PDA/FDR) or Laptop Application

### Rey Farve, Project Leader

U.S. Forest Service

San Dimas Technology & Development Center

San Dimas, California

November 2010

---

The information contained in this publication has been developed for the guidance of employees of the Forest Service, U.S. Department of Agriculture, its contractors, and cooperating Federal and State agencies. The Forest Service assumes no responsibility for the interpretation or use of this information by other than its own employees. The use of trade, firm, or corporation names is for the information and convenience of the reader. Such use does not constitute an official evaluation, conclusion, recommendation, endorsement, or approval of any product or service to the exclusion of others that may be suitable.

The U.S. Department of Agriculture (USDA) prohibits discrimination in all its programs and activities on the basis of race, color, national origin, age, disability, and where applicable, sex, marital status, familial status, parental status, religion, sexual orientation, genetic information, political beliefs, reprisal, or because all or part of an individual's income is derived from any public assistance program. (Not all prohibited bases apply to all programs.) Persons with disabilities who require alternative means for communication of program information (Braille, large print, audiotape, etc.) should contact USDA's TARGET Center at (202) 720-2600 (voice and TDD). To file a complaint of discrimination, write USDA, Director, Office of Civil Rights, 1400 Independence Avenue, S.W., Washington, D.C. 20250-9410, or call (800) 795-3272 (voice) or (202) 720-6382 (TDD). USDA is an equal opportunity provider and employer.



---

## Table of Contents

|  |    |
|--|----|
| 1. Background.....   | 1  |
| 2. Investigation of State of Automatic Speech Recognition (ASR)<br>Technology for PDRs/FDRs..... | 1  |
| 2.1 Automatic Speech Recognition (ASR) Technology.....   | 2  |
| 2.2 ASR Technology and Mobile Computing Devices (PDRs/FDRs) .....                                | 2  |
| 2.2.1 Categories of Mobile Computing Devices.....  | 2  |
| 2.2.2 Current Applications of ARS Technology .....   | 3  |
| 2.3 State of ASR Technology for Mobile Devices .....   | 5  |
| 2.4 Implementing ASR Technology on Mobile Devices .....  | 6  |
| 2.5 Speech Recognition Applications and Software Development Kits.....                           | 6  |
| 3. Evaluation of ASR Technology for Laptop and Desktop Computers .....                           | 8  |
| 3.1 How the Evaluation Was Conducted.....  | 9  |
| 3.2 Dragon Naturally Speaking Software .....   | 9  |
| 3.2.1 Features .....   | 10 |
| 3.2.2 Enrollment Process .....   | 10 |
| 3.2.3 Advantages and Disadvantages .....   | 10 |
| 3.2.4 Percent Error Rate.....  | 11 |
| 3.3 IBM's ViaVoice Software .....  | 11 |
| 3.3.1 Features .....   | 12 |
| 3.3.2 Enrollment Process .....   | 12 |
| 3.3.3 Advantages and Disadvantages .....   | 12 |
| 3.3.4 Percent Error Rate.....  | 12 |
| 3.4 Windows XP, Vista, and Windows 7 Speech Recognition Software.....                            | 13 |
| 3.5 Summary of the Evaluation of DNS and ViaVoice .....  | 13 |
| 4. Literature Cited .....  | 15 |
| 5. Appendix.....   | 17 |



---

## 1. Background

Wendy Stein of the Chequamegon-Nicolet National Forest submitted this proposal that requested the National Technology and Development Program evaluate the state-of-the-technology of voice recognition software for personal data recorders/field data recorders (PDRs/FDRs).

The proposer envisioned the end product as an enhancement of the current Forest Service field data recorder applications that might use voice recognition software to eliminate much of the manual manipulation needed for applications. Using voice commands, one might navigate to a specific screen and data entry field, complete the field data entry, as well as access a list of values or data dictionary. The proposer was especially interested in eliminating the incontinence of using gloved hands (in cold weather) while trying to operate and enter data in the field into PDRs.

The Inventory and Monitoring Steering Committee broadened the scope of the proposer's request to include an investigation of the state of voice activation technology for laptop personal computers (PCs), especially technology that decreases (or eliminates) the need for a keyboard.

San Dimas Technology and Development Center (SDTDC) was aware that voice activation technology had been available for desktop and laptop computers for several years. The status of voice activation use for PDRs/FDRs, however, was less certain. (The advanced features of speech recognitions require significant computing power that PDRs/FDRs do not possess.) As such, SDTDC considered the objective of the investigation was to:

1. Investigate the state-of-technology of voice recognition for field data recorders (section 2).
2. Evaluate popular, existing voice recognition software packages that are used with laptops/tablet PCs (section 3).



*Figure 1—Personal data recorders and field data recorders (PDR/FDR) differ from personal computers in the computing power necessary for speech recognition.*

## 2. Investigation of State of Automatic Speech Recognition (ASR) Technology for PDRs/ FDRs

Marc Todd, SDTDC computer programmer (retired), performed this state-of-the-technology investigation. It provides a general discussion of automatic speech recognition (ASR) technology, what currently restricts PDRs/FDRs from fully accommodating some features of ASR technology, and what would be needed for PDRs/FDRs to provide the most ASR features.

---

## 2.1 Automatic Speech Recognition Technology

The human brain's ability to process the sound of the human language and speech while distinguishing it from extraneous noises is an extremely complicated process. The task of programming any computer to mimic the brain's ability to process human speech represents a profound challenge. Today's ASR technology is meeting that challenge, achieving a 99-percent accuracy rate in some cases (Alwang 2002, Consumer Search 2009).

As modern ASR technology advances, it necessarily demands more computing power and hard-drive storage capacity. This technology exists primarily as software, which is designed and compiled to run under specific operating systems to the degree that the operating system will support it. Currently, the more powerful ASR technologies require at least the processing power of an up-to-date desktop PC or equivalent (laptop, tablet, or ultramini PC). Limited features of ASR technology, however, will run on less powerful devices, such as PDA/FDRs and cell phones.

As newer versions of mobile operating systems are developed, they increasingly accommodate more features of ASR technology. Nevertheless, it remains the limited computing power and hard-drive storage capacity of the smaller hand-held mobile devices that restricts the sophistication of the ASR technology that can be installed and run. It is not likely that hand-held mobile devices will acquire the power necessary to run advanced ASR in the foreseeable future.

The features of ASR technology: (1) command-and-control applications, (2) simple speech-to-text functionality, and (3) advanced speech-to-text functionality will be discussed in detail below, especially as they relate to PDRs/FDRs.

## 2.2 ASR Technology and Mobile Computing Devices (PDRs/FDRs)

### 2.2.1 Categories of Mobile Computing Devices

Mobile computing devices generally fall into three categories. The more complicated mobile devices (tablet PCs) are operated by desktop operating systems, just as laptops and desktop PCs are. At the other extreme are PDAs/FDRs, which have much less computing power and storage capacity. As such, they operate by using a less advanced mobile operating system. Since cell phones (especially smart phones) are able to wirelessly communicate with a server to augment their computing power, they fit in between these two extremes.

| PDA's and FDR's   | Cell Phones   | Tablet and Ultramini PCs   |
|---|---|--|
|  |  |  |
| Mobile operating system; identical functionality (except keypad)                    | Mobile operating system; simple and advanced (more or less like a PDA)              | Desktop operating system; equivalent to laptops                                      |

Figure 2—Typical mobile computing devices and their operating systems.

**Mobile Operating Systems.** These systems are designed to run on devices having relatively limited computing power and restricted hard-drive storage capacity. These have been either specifically designed for mobile devices (Symbian® OS), or they may represent a minimal version of a desktop operating system (such as Windows® Mobile). In either case, the original equipment manufacturer usually installs only those features of the operating system that it deems necessary to accomplish the purpose of the mobile device. Therefore, it may be that some devices will not support speech recognition even though that functionality might be accommodated by the provider of the operating system. (This is becoming less and less the case as mobile hardware technology advances.) <http://www.symbian.org/>

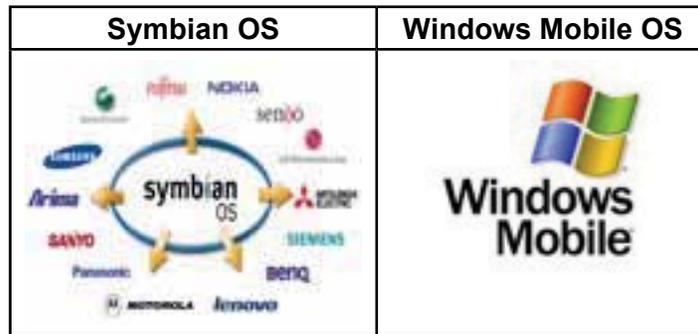


Figure 3—Two common operating systems for mobile devices.

### 2.2.2 Current Applications of ARS Technology

**Command-and-Control Applications.** These applications represent a feature of ASR technology installed to a device in the form of a stand-alone program, which normally runs in the background. This program serves to voice-enable control of the device for the purposes of turning the device on and off, starting programs, and so forth. The user might be seen as speaking to the device itself. The program simply compares the words spoken by the user to only those contained in its limited database of recognized words or simple verb-object phrases, and then issues the corresponding command to the device. The user may be able to add new words to the database and to associate any given word with the desired functionality.

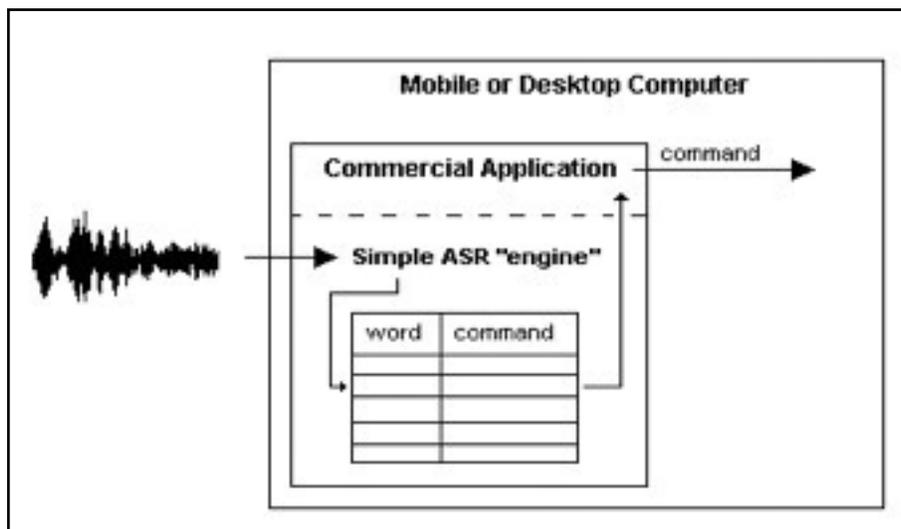


Figure 4—Diagram of simple command-and-control of mobile and desktop computers.

**Speech-to-Text Functionality (Simple).** This is a particular feature of ASR technology, which is made available to application developers for inclusion into their programs by means of a software development kit (SDK). With speech-to-text functionality a user is effectively speaking to a particular program. The ASR components of the program compare the discrete words spoken by the user to those contained in a database of recognized words (much larger than that employed by command-and-control applications). If a match is found, then the word is automatically inserted into the currently active text field of the program. (This type of speech recognition is sometimes referred to as discrete word recognition.) The user can usually add new words to the database, and these might be integrated with the predictive-typing capabilities of certain mobile devices. Command-and-control capabilities also may be available.

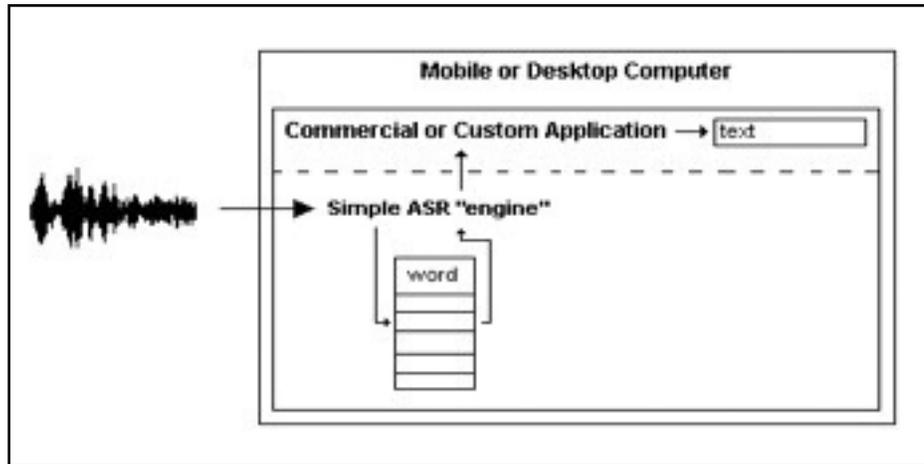


Figure 5—Diagram of simple speech-to-text functionality of mobile and desktop computers.

**Speech-to-Text Functionality (Advanced).** Again, ASR technology is made available to application developers for inclusion into their programs. Multiple languages may be supported and vocabulary is virtually unlimited. Sophisticated rules of grammar and other enhancements are incorporated into the software, which allows a user to speak continuously and more naturally. (This is often referred to as continuous speech recognition.) **The ASR components supplying this advanced functionality require significant processing power and hard drive storage that is usually limited to desktop PCs or equivalent.**

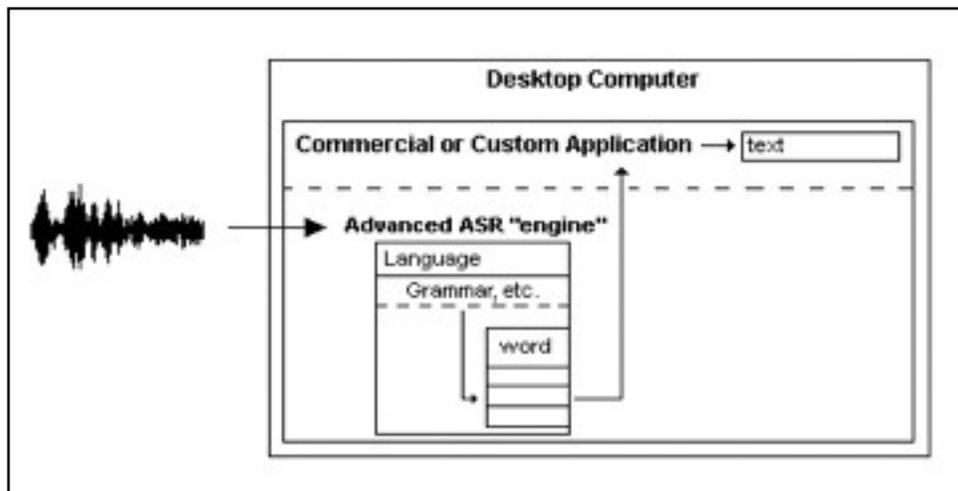


Figure 6—Diagram of advanced speech-to-text functionality that is only available for desktop computers.

**Distributed Speech Recognition (DSR).** DSR makes advanced speech-to-text functionality available to mobile devices (PDA/FDRs, and cell phones) that do not possess the required computing power or hard-drive storage. A wired or wireless connection to a computer (server) powerful enough to run advanced ASR software is required. A limited subset of ASR technology is installed to the mobile device (integrated into an application by the developer), which serves to preprocess the user's speech and to wirelessly transmit that information to the server. The server then fully processes the information and sends the appropriate text back to the mobile device, which subsequently inserts it into the currently active text field. (Obviously, the requirement of having to wirelessly transmit data to a server is a significant limiting factor for practical field use.)

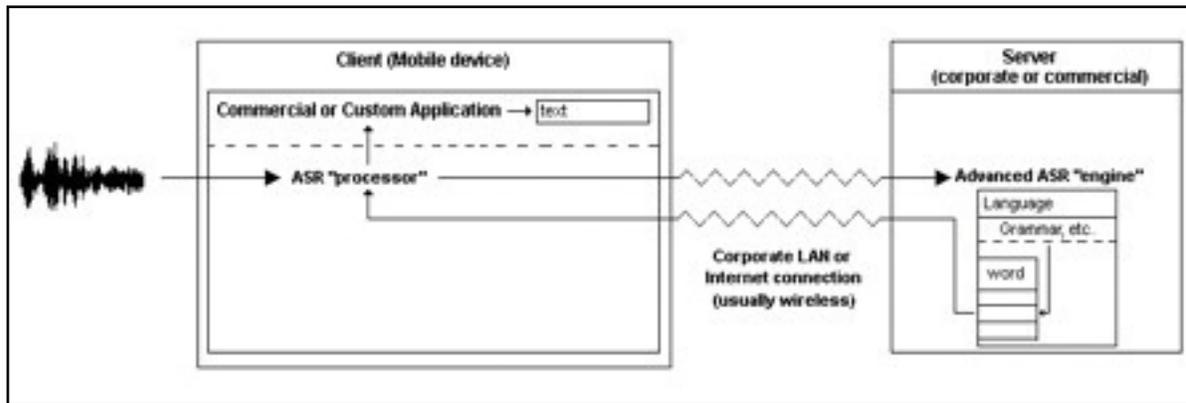


Figure 7—Diagram of how distributed speech recognition allows a mobile device to achieve advanced speech-to-text functionality.

### 2.3 State of ASR Technology for Mobile Devices

In general, PDRs/FDRs and cell phones can only run programs implementing simple ASR. Programs that implement simple ASR often require some minimal training of the speaker (user) and some configuration of the software or its associated databases. Most ASR technology applicable to PDRs/FDRs now targets only Microsoft's Mobile OS (operating system) version 5.0 and later (though at least one supplier still provides a product for Pocket PC 2002). This is because Mobile OS 5.0 provides a system interface for ASR engines to original equipment manufacturers and to application developers. Beginning with Windows Mobile OS 6.5, the operating system also provides inherent wireless DSR services via Microsoft's subsidiary Tellme.

Most of the smaller development companies working with ASR, along with their relevant technologies, are being purchased by a few large companies (notably Microsoft®, Nuance®, and Loquendo), which target their products toward the largest possible user groups. In the realm of mobile devices, ASR technology is being targeted toward advanced cell phones with functionalities, such as automated dialing and Web browsing, and toward in-vehicle navigation systems to support hands-free operation.

A few specialized companies are incorporating ASR technology into a few of their vertical (or niche) market solutions. At present, companies that develop ASR for mobile devices seem to be focused exclusively on warehouse inventory management systems, hospital information management systems, and, to a lesser extent, to voice-enabled applications for law enforcement vehicles<sup>1</sup>. The vertical market for advanced speech-to-text functionality is primarily the medical and legal profession (especially for transcription of reports/documents).

<sup>1</sup> Demand for Mobile ASR is predicted to triple in the next 4 years. Most companies will likely continue to address voice-dialing and voice-web-browsing; but as the technology improves it may become even better suited to other applications.

---

Command-and-control applications are commercially available as off-the-shelf software for mobile devices, including most of the PDA/FDRs commonly in use by the Forest Service (dependent upon the version of the operating system). Software development kits providing both simple and advanced speech-to-text capabilities also are commercially available for integration into custom software applications by developers. Sources for a number of these products are listed in section 6, along with links to the Web sites.

## 2.4 Implementing ARS technology on mobile devices

Consider the following items prior to utilizing ASR technology on mobile devices.

- The mobile device must either have a built-in microphone or be capable of accepting microphone input (wired or wireless). Some implementations of ASR technology have specific microphone requirements. In all cases: the higher the quality of the microphone, the better the results<sup>2</sup>.
- Speech-to-text functionality would need to be incorporated into each individual application by the developer(s). The use of even simple speech-to-text functionality with Forest Service applications would require the rewriting of at least some parts of existing applications and might further require significant rewrites to accommodate some of the more advanced features of Windows OS 5.0.
- The expansion of cell coverage or of agency remote communication capabilities may allow the use of distributed speech recognition technology which, in turn supports advanced speech to text on PDRs/FDRs in remote locations in the not too distant future.
- As an alternative to reconfiguring software for PDRs/FDRs, one might consider using more powerful mobile devices, such as tabletPCs or ultramini PCs, as these are capable of running more powerful and accurate implementations of ASR technology. Chest packs might be designed to make carrying such devices in the field more convenient. Employing different devices, however, would require further rewrites of agency software.

## 2.5 Speech Recognition Applications and Software Development Kits for Mobile Devices



Figure 8—Typical voice enabled mobile devices.

---

<sup>2</sup> Lumsden et al. (2007) and (2008) has reported the efficacy of different microphones and under different levels of background noise to support data entry by speech onto tablet PCs.

---

**Command-and-Control Applications.** Below are a few commercial software programs that allow a user to dictate very simple voice commands to PDRs and smart phones. The software allows the user to exercise limited hands-free control over the mobile device by the execution of simple commands, allowing the user to launch programs, look up contacts or phone numbers, or make phone calls. Most have SDKs that allow a software engineer to create applications or interfaces with other applications.

- [Microsoft Voice Command](#). Works with mobile devices that operate by using Windows Mobile 5.0 and better. Supports Bluetooth headsets. Not supported on some devices (such as HP Ipaq because of the Bluetooth stack implementation on that device).
- [Fonix VoiceCentral](#). Windows Mobile 5.0 and better (older versions support back to Pocket PC 2002). Supports Bluetooth headsets. SDK available ([Voiceln](#)).
- [Vito Voice2Go](#). Windows Mobile 5.0 and better.
- [Speereo Voice Launcher](#). Windows Mobile 2003SE and better. SDK available.
- [NowSpeak Mobile](#). SDK only.
- [Nuance VoCon 3200](#). SDK only.

**Speech-to-Text Software Development Kits.** Below are some commercial products that provide for simple speech-to-text features via device-resident ASR and advanced speech-to-text features through a server.

#### **Device-resident ASR**

- [AccuCode® AO:Voice](#). Works better on newer mobile devices that operate by using Windows Mobile 5.0 and better. Supports DSR. Multiplatform (works for desktops and tabletPCs as well). Employs AccuSpeechMobile (below). Voice enables even commercial off-the-shelf software.
- Vanguard Voice Systems, [AccuSpeechMobile](#) works best with newer PDRs/FDRs that operate on Windows Mobile 5.0 and better. Supports Bluetooth headsets, but works best with plug-in headsets.

See Vanguard Voice Systems Web site for a [video demonstration of how voice commands could be used to power an application on a PDR](#).

(Note: Vanguard developers assist clients develop the software so that it can drive their intended application.)

- [SVOX Mobile](#). This software has been used primarily with smart phones to provide text-to-speech functionality. It also is being used with Google Maps and in high-end automobiles (voice-enabled in-car systems) to provide simple hands-free voice commands. See demonstration of SVOX controlling smart phones on the SVOX Web page: <http://www.svox.com/Demos.aspx>



Figure 9—AccuSpeechMobile, Mobile Voice Platform from Vanguard Voice Systems (see [AccuSpeechMobile Brochure](#)).

### Client/Server Distributed Speech Recognition (DSR)

Nuance is one of the leaders in developing ASR technology and software, especially for providing DSR for mobile devices. Software developed by Nuance has focused largely on providing server-based speech recognition and speech-to-text functionality for mobile phones and smart phones. (See Nuance brochure at: [Mobile speech brochure](#).)

Two products by Nuance that provide advanced ASR for mobile devices are:

- [Nuance Mobile Gateway](#)
- [Nuance Mobile Speech Platform](#)

See a demonstration using Nuance on a smart phone: [Nuance demonstration](#)

### 3. Evaluation of ASR Software for Laptop and Desktop Computers

ASR software for desktops and laptops has been on the market for over a decade. As stated previously (section 2.2.2), ASR technology has advanced such that speech-to-text functionality now utilizes sophisticated rules of grammar and other enhancements to allow the user to speak more naturally (continuously) as opposed to the discrete (word-by-word) calling that was necessary for earlier speech recognition software (discrete voice recognition).<sup>3</sup>

The most notable (and popular) ASR software is Dragon Naturally Speaking distributed by Nuance Communications. Dragon Naturally Speaking was first introduced in 1997 (version 1.0). The most recent version (version 10) was released in 2008. IBM's ViaVoice (also distributed by Nuance) is considered a less expensive alternative to Dragon Naturally Speaking (about \$230 for ViaVoice versus \$700 for Dragon NS). As might be expected ViaVoice also has fewer features and capabilities than Dragon Naturally Speaking. In this evaluation SDTDC (we) compared the ASR features and capabilities of Dragon Naturally Speaking (v.10) and IBM's ViaVoice (Professional USB Edition, v.10) on a standard Forest Service Hewlett Packard laptop PC.

<sup>3</sup>Continuous voice recognition systems do not hear individual words or phrases but rather continuous streams of sounds that are identified as sound patterns (phonemes). The software uses large collections or libraries to compare the patterns heard. The software transcribes text based on the sound pattern - not the individual words. Over time, a user builds a personal library of sound patterns that eventually improves the software's recognition accuracy (Di Petta and Woloshyn 2001).

---

### 3.1 How evaluation was conducted.

Several investigators previously have performed evaluations of ASR software for personal computers. Several of these reviews performed a head-to-head comparison of Dragon Naturally Speaking versus ViaVoice. Devine et al. (2000) compared and evaluated ASR software packages and their out-of-the-box ability to accurately record dictations for medical records, and Weinzetl (2005) evaluated commercial ASR software's effectiveness for dictating reports by law enforcement officers. Tamez (2003) reported on ASR software's effectiveness in command-and-control of a Navy vessel's engine and rudder during maneuvers. Drevensted et al. (2007) discussed the advantages of using ASR for dental practices.

The method we employed in this evaluation was based on the procedures used in some of these previous investigations.

We evaluated the two ASR softwares (Dragon Naturally Speaking and IBM ViaVoice) using the voice of a single staff person with the same microphone and on the same laptop PC at the SDTDC office. We followed the directions provided in the user manuals relative to training the software (enrollment) to recognize an individual's voice and speech pattern<sup>4</sup>. After the enrollment period, the tester practiced with each ASR for about 2 hours to become more familiar with the software's features.

Following this period, we tested the software for its accuracy in performing dictation by determining the word error rate.

The percent error rate was calculated as:

$$\text{Percent Error Rate} = \frac{[(\text{insertion} + \text{substitutions} + \text{deletions})/\text{total words}] \times 100}{}$$

For each ASR software program, the tester read three separate (different) passages using his/her normal speaking voice, in a normal manner, and in a quiet office setting. The tester recorded the reading of the passage on a digital recorder to ensure that the passage was read correctly.

A second reading of the passages was conducted on a separate occasion to determine if the software learned to better recognize the user's voice and thereby reduce the error rate. (The three passages that were read are provided in the appendix.)

### 3.2 Dragon Naturally Speaking

As previously mentioned, Dragon Naturally Speaking (DNS) was one of the earliest software programs marketed for speech recognition and it is considered a leader in the voice recognition market. DNS software is available in five different editions. We tested the Professional Edition (cost: \$665).



Figure 10-Dragon Naturally Speaking (version 10) - Professional Edition.

<sup>4</sup>The software license allows the purchaser to install the software on up to five computers. Although the license is single-user, the license is extended to five immediate family members who can also create their own user profiles in the program.

---

DNS Professional allows for complete hands-free use of a personal computer. A user can convert speech-to-text in any Windows program, and command-and-control any application on a PC's desktop.

DNS specifically markets to the disabled community and claims that the Professional Edition is certified under Section 508 of the Amendments of 1998 to the Rehabilitation Act of 1973 (29 U.S.C. 794d) (see Nuance 2008).

### 3.2.1 Features

**Command-and-Control.** DNS Professional allows the user to control any Windows application with voice commands. The user has hands-free voice control of Microsoft Internet Explorer, MS Word, MS Outlook,

Lotus Notes, and MS Excel. The software has limited control of Mozilla Firefox.

The software allows the user to launch Internet Explorer and perform Google or Wikipedia searches by using voice commands only.

**Speech-to-Text.** DNS converts speech-to-text in DragonPad (similar to MS WordPad), MS Word, and email messages (Outlook or Lotus Notes).

The software allows the user to create custom voice commands (like macros) to automate repetitive tasks and boilerplate text. A user also can download his/her dictation from a Nuance-certified digital recorder to a PC to achieve automatic transcription.

DNS claims its speech-to-text accuracy rate is 99 percent.

A full description of the features available for the various DNS editions is found at their Web site: <http://www.nuance.com/naturallyspeaking/products/product-matrix.asp>

The New York Times, Technology Section posted (August 8, 2008) a video demonstration of DNS, which is provided at the link below:

<http://video.nytimes.com/video/2008/08/07/technology/personaltech/1194817477305/when-your-computer-listens-to-you.html>

### 3.2.2 Enrollment Process

The tester, following the instructions in the user manual, completed the enrollment process in about 30 minutes. Enrollment is a two-step process that involves: (1) setting up the audio level of the microphone, and (2) creating a user profile by training the software's algorithms to match the individual's speech. (This is done by reading large portions of suggested text so the software recognizes the user's speech pattern.) The instructions and effort to complete the process was minimal. After enrollment the user is given an option to adapt DNS to the user's writing style. This can increase the voice recognition accuracy of the software. This adaptation process can add 5 to 30 minutes to the user profile training.

### 3.2.3 Advantages and Disadvantages

DNS software is straightforward and easy to use. There are many available resources for technical support and training, such as a printable online user's guide, quick start card, command reference card, and a cheat sheet of tips and commands. There are also tutorials and on-screen help. (Also see section 3.5 for a discussion of training and support also offered by [Microref.com](http://Microref.com).)

---

DNS will open many applications by voice command on the desktop but others it does not recognize. It will allow voice command-and-control of Internet Explorer, which allows the user to access Web sites. However, once you are directed to the selected Web page, typically the user needs to use the mouse to navigate further within it.

With practice, command-and-control becomes easier and more fluid, but initially a user is tempted to use the keyboard and mouse instead of waiting for the software to correctly recognize voice commands.

### 3.2.4 Percent Error Rate

The word error rate of DNS is reported below. DNS had a 1-percent error rate with the first passage on the first reading test. (Note: DNS claims an accuracy rate of 99 percent.) The other passages appear to be somewhat more complex and, therefore, the software generated a higher error rate (4-7 percent) with the first test.

In the second test, the software appeared to learn and decreased the error rate somewhat.

*Table 1—Percent error rate of Dragon Naturally Speaking in performing speech-to-text.*

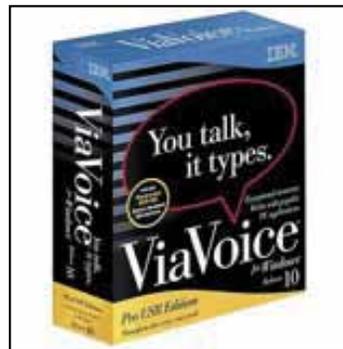
|  | Passage 1 (%) | Passage 2 (%) | Passage 3 (%) |
|--|---------------|---------------|---------------|
| 1st reading test   | 1             | 7             | 4             |
| 2nd reading test   | 1             | 2             | 4             |
| Percent Error Rate = [(insertion + substitutions + deletions)/total words] x 100 |               |               |               |

### 3.3. IBM's ViaVoice

ViaVoice is probably the second most popular voice recognition software available. It was released to the market in 1997. In 2003, IBM gave global distribution rights to SoftScan. In 2005, SoftScan merged with

Nuance Communications (the distributor of Dragon Naturally Speaking). ViaVoice software has not been updated since 2003.

ViaVoice comes in [four Editions](#). We tested the Pro USB Edition (v.10) (cost: \$180).



*Figure 11-IBM's ViaVoice (version 10) - Pro USB edition.*

Most reviewers of ViaVoice (see summary discussion in section 3.5) consider this software to be a reasonable, less expensive speech recognition alternative to DNS. ViaVoice has many of the speech-to-text features of DNS, but, in general, lacks the enterprise solutions that DNS offers, for example DNS can store user profile information on a server to allow multiple users of one computer.

However, since ViaVoice, has not been updated since 2003, it will likely continue to fall further behind DNS in features.

---

### 3.3.1 Features

#### Command-and-Control

Compared to DNS, ViaVoice has more limited command-and-control of Windows applications and Internet Explorer. ViaVoice, however, can command-and-control: MS Word, MS Excel, MS PowerPoint, and MS Outlook. It does not allow for control of Lotus Notes.

The software allows limited use within Internet Explorer. It does not perform Google or Wikipedia searches. The only voice command it responds to is selecting tabs and toolbars.

#### Speech-to-text

ViaVoice converts speech-to-text in SpeakPad (which is similar to MS WordPad) and in MS Word. ViaVoice also allows for speech-to-text input into any application where text can be entered by the keyboard, such as into cells of spreadsheets like Excel).

Like DNS, ViaVoice allows the user to create custom voice commands (macros) to automate repetitive tasks and boilerplate text.

ViaVoice does not download dictation from a digital recorder to a PC to achieve automatic transcription. The distributor makes no accuracy rate claims.

See the [ViaVoice, Pro USB Edition User Manual](#) for all the software's features.

### 3.3.2 Enrollment process

The tester completed the enrollment process in about 3 hours by following the instruction in the user manual. The process is very similar to the enrollment process for DNS.

The instructions and effort to complete the process were difficult and time consuming, primarily because it took quite a bit of troubleshooting to eventually discover that the headset that came with the software was not working<sup>5</sup>.

### 3.3.3 Advantages and Disadvantages

The software is reasonably responsive to voice commands. There is a helpful user's guide and command reference card that comes with the software. There is also a voice center that offers access to most ViaVoice functions. The only desktop application available with a voice command is Internet Explorer. (This may apply only to Microsoft 2007 because the user's guide claims you can open applications on the desktop.)

As with DNS, the novice user typically becomes impatient when the software fails to recognize voice commands, and the experienced keyboard user often is tempted to use the keyboard and mouse instead.

### 3.3.4 Percent Error rate

Compared to DNS, ViaVoice had a higher error rate on the initial test for reading all passages. However, the software showed significant improvement in the second test as the error rate decreased to 3-4 percent for all passages. (Note: ViaVoice makes no claims of the software's accuracy rate.)

---

<sup>5</sup>The tester received multiple error messages during the software installation process, and technical support could not be obtained from Nuance via phone or email for ViaVoice issues. The tester eventually used the headset from the DNS software and was finally able to complete the enrollment process.

Table 2—Percent error rate of ViaVoice in performing speech-to-text.

|  | Passage 1 (%) | Passage 2 (%) | Passage 3 (%) |
|--|---------------|---------------|---------------|
| 1st reading test   | 4             | 10            | 11            |
| 2nd reading test   | 4             | 3             | 4             |
| Percent Error Rate = [(insertion + substitutions + deletions)/total words] x 100 |               |               |               |

### 3.4. Windows XP, Vista, and Windows 7 Speech Recognition

Most users of PCs that are operated by Windows XP operating system are probably unaware that XP supports speech recognition. Computers with Windows XP typically come with a speech recognition engine (SRE) already installed. Once activated and trained to your voice, the SRE allows the user to perform simple command-and-control of MS applications and adequate speech-to-text functions. Microsoft admits, however, that the SRE is not designed to provide for complete hands-free operation.

The SRE, however, does not work with Microsoft Office 2007 applications. Since Forest Service computers have MS Office 2007 and the SRE is not available, we made no attempt to evaluate XP's speech recognition functions.

The reader should be aware, however, that speech recognition functionality is available for Microsoft Vista and MS 7 Operating Systems. <http://www.microsoft.com/windows/windows-vista/features/accessibility.aspx?tabid=2&catid=1> Some independent reviews, such as Muchmore 2008 and ConsumerSearch 2009, consider the speech recognition software of these systems significantly more advanced than the XP Operating System and go as far as to report it as good as DNS in performing routine speech recognition tasks. If Forest Service computers are upgraded to these new operating systems, readers might be interested in testing these speech recognition capabilities.

We provide the following links for the interested reader:

[Microsoft's Web site with information on Vista's speech recognition capability](#)  
[Video demonstration of Vista Speech Recognition.](#)

### 3.5. Summary of Evaluation of DNS and ViaVoice

Either DNS or ViaVoice require a significant investment of time to become proficient in the use of the software's features. However if the potential user spends a significant amount of time typing, the time investment might be worthwhile. This is probably truer for DNS software, as it will likely improve in subsequent versions.

As previously stated (section 3.1), several reviewers performed head-to-head tests of DNS versus ViaVoice. The results of those tests were, in general, similar to what we report. DNS typically outperformed ViaVoice in word error rate and in ability to command-and-control applications (Alwang 2002; Ferrill 2005; Wood 2007; ConsumerSearch 2009). Note, however, that Devine et al. 2000 reported that ViaVoice 98 was more accurate than DNS Medical Suite (v.3) for transcribing dictation.

The user that is seriously interested in ASR probably should consider the DNS software as it is the recognized leader and it has survived challenges from other software competitors for over a decade. Furthermore, companies like [Microref.com](#) are dedicated to the promotion, implementation, and utilization of speech recognition technology for business purposes, especially DNS. They have developed products, accessories, and training services to assist DNS users.

ViaVoice is generally considered a less expensive alternative to DNS. ViaVoice offers many of the features of DNS at a fraction of the cost. We found customer support for ViaVoice to be very limited.



---

#### 4. Literature Cited

- Alwang, G. 2002. Voice recognition: getting better. P.C. Magazine. 21(4): 29.
- ConsumerSearch. 2009. Voice recognition software: full report. <http://www.consumersearch.com/voice-recognition-software/review>. Updated October 2009. Accessed November 4, 2009.
- Devine, E. C.; Gaehde, S. A.; Curtis, A. C. 2000. Comparative evaluation of three continuous speech recognition software packages in the generation of medical reports. Journal of American Medical Informatics Association. 7(5): 462-468.
- Di Petta, T.; Woloshyn, V. E. 2001. Voice recognition for on-line literacy: continuous voice recognition technology in adult literacy training. Education and Information Technology. 6(4): 225-240.
- Drevenstedt, G. L.; McDonald, J. C.; Drevenstedt, L. W. 2005. The role of voice-activated technology in today's dental practice. Journal of American Dental Association. 136: 157-161.
- Ferrill, P. 2005. Can you hear me now?: Improvements prep voice-recognition technology for a wide range of users. Federal Computer Week. 19(12): 30-34.
- Lumsden, J.; Kondratova, I.; Durling, S. 2007. Investigating microphone efficacy for facilitation of mobile speech-based data entry. Proceedings of the British HCI Conference. Lancaster, UK. National Research Council of Canada. September 2007.
- Lumsden, J.; Durling, S.; Kondratova, I. 2008. A comparison of microphone and speech recognition engine efficacy for mobile data entry. Proceedings of the International Workshop on Mobile and Networking technologies for social applications (MONET 2008). Monterey, Mexico. Nov. 9-14, 2008. 519-527.
- Muchmore, M. 2008. Dragon naturally speaking 10 - review. PC Magazine <http://www.pcmag.com/article2/0,2817,2327354,00.asp> (accessed March 2, 2009).
- Nuance. 2008. Dragon Naturally speaking professional for government: A white paper from Nuance Communications. August 2008.
- Tamez, D. J. 2003. Using commercial off-the-shelf speech recognition software for conning U.S. warships. Monterey, CA: Naval Postgraduate School. 92 p. Thesis.
- Weinzetl, M. P. 2005. Analysis of voice recognition software as an aid to the efficiency of the report writing process in law enforcement. St. Paul, MN: Concordia University. 84 p. Thesis.
- Wood, L. 2007. Is speech recognition finally good enough? <http://www.computerworld.com>, dated May 18, 2007 (Accessed February 22, 2010.)



---

## 5. Appendix

Three passages that were read to test the speech-to-text Percent Error Rate of Dragon Naturally Speaking and ViaVoice.

### 5.1 1st Reading Test - President Abraham Lincoln's Gettysburg Address

Gettysburg, Pennsylvania

November 19, 1863

(271 words)

"Four score and seven years ago our fathers brought forth on this continent, a new nation, conceived in Liberty, and dedicated to the proposition that all men are created equal.

"Now we are engaged in a great civil war, testing whether that nation or any nation so conceived and so dedicated, can long endure. We are met on a great battlefield of that war. We have come to dedicate a portion of that field, as a final resting place for those who here gave their lives that that nation might live. It is altogether fitting and proper that we should do this.

"But, in a larger sense - we can not dedicate; we can not consecrate; we can not hallow - this ground. The brave men, living and dead, who struggled here, have consecrated it, far above our poor power to add or detract. The world will little note, nor long remember what we say here, but it can never forget what they did here. It is for us the living, rather, to be dedicated here to the unfinished work which they who fought here have thus far so nobly advanced. It is rather for us to be here dedicated to the great task remaining before us; that from these honored dead we take increased devotion to that cause for which they gave the last full measure of devotion; that we here highly resolve that these dead shall not have died in vain; that this nation, under God, shall have a new birth of freedom; and that government of the people, by the people, for the people, shall not perish from the earth."

### 5.2 2nd Reading Test - from San County Almanac by Aldo Leopold

(257 words)

"We were eating lunch on a high rimrock, at the foot of which a turbulent river elbowed its way. We saw what we thought was a doe fording the torrent, her breast awash in white water. When she climbed the bank toward us and shook out her tail, we realized our error: it was a wolf. A half-dozen others, evidently grown pups, sprang from the willows and all joined in a welcoming melee of wagging tails and playful maulings. What was literally a pile of wolves writhed and tumbled in the center of an open flat at the foot of our rimrock.

"In those days we had never heard of passing up a chance to kill a wolf. In a second we were pumping lead into the pack, but with more excitement than accuracy; how to aim a steep downhill shot is always confusing. When our rifles were empty, the old wolf was down, and a pup was dragging a leg into impassable side-rocks.

"We reached the old wolf in time to watch a fierce green fire dying in her eyes. I realized then, and have known ever since, that there was something new to me in those eyes-something known only to her and to the mountain. I was young then, and full of trigger-itch; I thought that because fewer wolves meant more deer, that no wolves would mean hunters' paradise. But after seeing the green fire die, I sensed that neither the wolf nor the mountain agreed with such a view."

---

### 5.3 3rd Reading Test - from Round River by Aldo Leopold

(234 words)

“Conservation is a state of harmony between men and land. By land is meant all of the things on, over, or in the earth. Harmony with land is like harmony with a friend; you cannot cherish his right hand and chop off his left. That is to say, you cannot love game and hate predators; you cannot conserve the waters and waste the ranges; you cannot build the forest and mine the farm. The land is one organism. Its parts, like our own parts, compete with each other and cooperate with each other. The competitions are as much a part of the inner workings as the cooperations. You can regulate them-cautiously-but not abolish them.

“The outstanding scientific discovery of the twentieth century is not television, or radio, but rather the complexity of the land organism. Only those who know the most about it can appreciate how little we know about it. The last word in ignorance is the man who says of an animal or plant: “What good is it?” If the land mechanism as a whole is good, then every part is good, whether we understand it or not. If the biota, in the course of eons, has built something we like but do not understand, then who but a fool would discard seemingly useless parts? To keep every cog and wheel is the first precaution of intelligent tinkering.”



