Voice Recognition Technology:

Adaptive Studies for Success

By Michael Siebenaler

Research Proposal for Research in Technical Writing graduate class

Bowling Green State University

April 2004

Instructor: Dr. Judith R. Edminster, Department of English

Bowling Green State University

Introduction

Since I purchased the voice recognition programs, Dragon Naturally Speaking® and Lifeline®, a Playstation2™ (PS2) game, I've enjoyed great benefits from the technology, but also encountered many challenges.

I've heard various detracting attitudes and statements from other people about the technology. Some people didn't like technology at all, while some people didn't think it was worth their time. I didn't find much research about voice recognition technology (VRT) that explained why a program would work for one user, but not another. I feel social research would assess identifiable traits and reasoning why a person would choose not to utilize voice recognition technology. I believe this research can be best achieved by having the subject experience the training process of VRT programs and analyze this research data.

Research actions will attempt to address the following areas:

1. Assess social perception (surveys/discussions/pre-test to training)
2. Training process/articulation
3. Users describe the training process
4. Assess perceptions after training (surveys/discussions/post-test to training).

Though the training processes I encountered were tedious at times and required a lot of patience, I felt it was worth the time and effort. The results I've achieved from VRT saved me countless hours of work and stress on my hands. The Dragon Naturally Speaking program also improved my oral skills and allowed me to multitask during the process of my work (allowing use both hands to read resources and references, etc.), while the Lifeline program provided

entertainment through a unique experience that combined important communication methods with VRT.

My research project will investigate the social aspects and attitudes behind voice recognition technology (VRT) software available for personal use, namely Dragon Naturally Speaking and Lifeline (plus the PS2 system and necessary peripherals including a special USB headset). These specific programs were chosen to narrow the scope of voice recognition technology specifically for this research, so subjects wouldn't discuss other voice recognition processes (e.g. telecommunications experiences such as saying number options, like credit card number, or verbally answering "yes" or "no"; programming speed dial options by voice, etc.). These other voice recognition processes may not be recognized by the subjects, but its importance includes other areas such as telecommunications because subjects experience these processes at a high frequency. Subject approval is dependant on the local research advisory board.

Research questions

Many references I reviewed discussed social benefits, but mainly concentrated on negative social aspects such as frustration during program training, perceived usability of the software and/or skepticism of technology. The research questions involve negative and positive social aspects of voice/speech recognition technology. These questions address the traits and reasoning people would or would not use VRT.

1. In which ways can VRT be socially viewed as a discourse process, even as an alternative or in combination with traditional interfaces?
2. What methods of VRT training produce the most successful results and practices?

Both research questions will require the subjects to complete training process for both programs. The level of assistance provided by the facilitators only originates from help documentation contained in each respective program. The research data results will also analyze social views of VRT including:

- Measurements of the social comfort of the subjects using the application
- Communication factors such as orality/editing/writing

Literature Review

"Speech is one of the most natural means of exchanging information for humans…enlisting the possible applications of such as system capable of understanding natural human speech is a task limited only by human imagination (Ganapathtrafu, 1999)."

VRT has reached a crucial point in society where it will become widely used or go the way of the BETA tape and have limited usability. Regardless of social views, VRT is a growing technology that many people are using.

VRT incorporates four core technologies (Dreyer, 2002)

1. Speech to text (speech recognition)
2. Text to speech (speech synthesis)
3. speaker identification and verification
4. natural language understanding

The technology must account for the variables of simultaneous voice to text process as the user speaks into a microphone. Then the sound is analyzed through a sound card, language model and speech engine on a computer. The data is converted into digital signal that can be processed by the computer (a.k.a. sampling) through programming models and user

identification data achieved through voice training. Next, VRT attempts to achieve synchronicity between text and dialogue (including speech rate, technology processes and lag times). The user sees recognized speech written as words on computer and the resulting data involves various editing processes.

"Telling a computer where to make an addition, for instance, still remains cumbersome. It is frequently easier to make the required edits manually (Clabby, 2002)." I have frequently experienced this first-hand keeping one finger on the delete key of my keyboard while using VRT. Methods used by the subjects will probably be different, but will be observed and noted.

The following information will address VRT history, barriers, available programs, progress and success:

Brief History

Late 1940's: U.S. Department of Defense wanted to expedite the reception and translation of intercepted Russian messages. The project failed but led to government-sponsored research in voice recognition development.

1952: Bell Laboratories identified digits spoken over the telephone.

1959: Massachusetts Institute of Technology (MIT) developed a system that identified vowel sounds with 93 percent accuracy.

1966: The next phase of the MIT project produced a system that recognized 50 vocabulary words.

1970's: Carnegie Mellon University researchers created software that could recognize complete sentences. However, it required the power of 50 computers to process the information.

1980's: Voice recognition software began to show up commercially. It required that users use discreet speech by enunciating each word and pausing between. The cost ran in the $1,000 range.

Late 1980's: Prices had fallen to about $800 but software still required discreet speech.

1996: Charles Schwab implemented a speech recognition system for its customer interface. On seeing Schwab's success, other companies quickly followed suit. Software prices fell, some as low as $150. People could dictate to their computers without pausing at a rate of 120 words per minute.

1997: Naturally Speaking became available. It used continuous speech, thus eliminating the need for a pause between each word.

2000: NetByTel launched recognition technology that allowed users to fill out a Web-based data form over the phone. (Bruce, 2002)

Barriers

"Even though human communication through speech appears to be extremely easy, mathematically modeling the underlying processes has proven to be one of the grand challenges of modern computing (Deshmukh, 1999)."

Pools of knowledgeable computer professionals are limited, but these professional must also understand the capabilities and applications of voice recognition technology to fully expand this field. The potential is there, but the gap is great as described in the following quote:

> "A shortage of computer literate and capable white collar information workers has plagued business worldwide for over 25 years…Billions of people on this planet cannot

type-but almost everyone on this planet can speak. For business, speech recognition technologies hold the promise of helping to overcome one of the biggest impediments to computer use-keyboard input-thereby reducing costs related to training and education (Clabby, 2002)."

VRT has numerous social challenges to overcome involving accuracy rates, speech variations/stresses, computer limitations and various stigmas.

Accuracy rates

"Accuracy has improved dramatically over the past two years – Recently, speech recognition has improved for certain application environments. With accuracy rates that many vendors claim exceed 99 percent, speech recognition is now a viable technology for command-and-control applications as well as for speech dictation. Further, issues such as distortion (noise from diverse surrounding environments) can be overcome, depending on the equipment used and the location of the user (Clabby, 2002)."

The association of VRT technology with an accuracy rate also creates a social stigma. If it's not 100%, it's not any good. "If this fledgling technology is to realize its potential, its developers will probably need to achieve at least 99% accuracy. The last 5% is proving extremely hard to achieve. (Pringle, 2001)." What program is 100%? 90% or above seems acceptable, but 98% or better is ideal by current standards. There's no method in the either program (Dragon Naturally Speaking or Lifeline) that could measure accuracy any way – facilitators would spell check at the end of each session and count the words then compare error

words with the total words to get a percentage. A challenging task, but it would provide valuable research data especially when compared with conventional writing.

"Experience with the system produce individual rates, as words are added to the non vocabulary, as this system learns individual pronunciations as practice dictating increases. However, even with frequent use, an error rate of about 2-5% will remain. (Gardner-Bonneau, 1999)." The research will have to account for this estimate when compiling the data.

Speech variations/stresses

The research will address the relation between human exertion and the ability of speech recognition software to correctly recognize human speech. Typically the following results occurs – "the higher the level of exertion, the lower the accuracy rate (Entwistle, 2003)."

The research will also consider how the subject's speech variations due to physical stresses and/or conditions would affect VRT applications. Many companies using biometrics applications for special privilege/access to facilities and information see this variation as a possible security risk. Most companies recommend combining VRT with another biometric technology, such as with a retinal scan (Gardner-Bonneau, 1999).

Noise variations and distortions in the subject's background also play an important role. Most social situations frequently include people talking in the background (no silent places anymore unless you're in a library). Dynamic environments cause a lot of distortion especially for a wireless system and subjects can not always find private area.

Also, physical changes over time, such as aging, should be considered.

Computer limitations

One of the biggest problems that speech recognition systems experienced in the past was related to PC hardware-specifically CPU processing speed and data storage. Storing large speech files can burn up disk space like there's no tomorrow… computer memory is now inexpensive making it possible to process speech data in main memory. More main memory means that the data to be processed is moved closer to the processor rather than having to be accessed from disk or some other input output device, hence commands and dictation can be more quickly interpreted and displayed or acted upon (Clabby, 2002). In VRT, such pauses do not necessarily appear at each word end, which significantly increases the number of possible processing paths and the memory required to keep track of these paths.

Large processing power and the speech synthesis process can also produce large limitations:

- Incorrect determination of word and/or sentence ending
- Educated guesses at received speech that doesn't conform to system parameters
- Amount of memory
- Variables - speed, accent, mumbling, omitting syllables, pitch and fluctuation

Frustrated users may even revert back to traditional interfaces like the mouse, keyboard or touch screen when VRT doesn't meet their expectations. "What we dictate into a microphone appears immediately on our computer screen, in the same way as when we type at a keyboard (Hartley, 2003)." VRT has replaced some traditional interfaces, but has been considered only as an alternative in other cases. "Consumers believe that speech is a technology that will create

more usable applications, even when faced with the relative failure of current technology to live up to their expectations or fulfill their practical needs (Johnson, 2001)."

The potential of VRT and its potential cost saving exists, but just hasn't been fully realized yet. "University-level research studies have shown that businesses can expect to realize new efficiencies by enabling computer users to tell their system or application what they are trying to do, rather than trying to figure out layers in the drop-down menu, a command would take the desired action-and efficiency frequently translates into additional cost savings (Clabby, 2002)."

Various stigmas

VRT seems to have many stigmas – it's only for the disabled; it takes too much time, etc. "Most people are not willing to spend 1, 3, or 10 hours training their computer to understand their speech patterns. Programs will have to be developed to adapt to speakers more quickly (Clabby, 2002)."

An ideal vocabulary database that achieves high accuracy rates requires long training sessions when using speaker dependant systems. Subjects need patience and significant blocks of time allotted to achieve this goal. Time issues still continue after the process is completed, mostly regarding extensive editing time to correct errors such as incorrect grammar and misspellings either by the VRT program's misrecognition or human error. From a user's point of view, ideal use of VRT might include a universal system, which recognizes and synthesizes voice input from any subject that required no training, but the majority of VRT applications involve speaker dependant systems.

More efficient typing is a major benefit. “The average typist can type between 40 and 80 words per minute. But individuals using speech recognition technology can type (using speech) at 160 words a minute-2 to 4 times faster then using a keyboard (Clabby, 2002).” VRT research results have also showed significant improvement in a subject’s written work (Bruce, 2002)

Talking and interaction with technology also represents a major stigma, for example, by “telephone all voices are disembodied, and…requires no interaction with the caller. However many people find telephone interactions with machines uncomfortable (Pitt & Edwards, 2003).” Research subjects often have to describe their VRT experience because no one is on the other end giving them feedback, commonly encountered in human to human contact. Here are some responses from VRT research subjects:

> “I experience dictating as very differently from writing. It feels far more contrived, in fact stilted, and I am often surprised to see my senses come out reasonably correct grammatically speaking. At the same time, I find dictating physically far easier than tapping at a keyboard with two fingers...
>
> There are times when I just cannot get a sentence right via dictating, even if I do lots of corrections. I sense that there is here a substantial difference, but don’t have the linguistic tools to analyze it…
>
> I don’t think that the move from typing to dictating-when using a word-processor for both-is anywhere near as making as the leap from using a typewriter to using a word processor. It is the latter which is the huge leap, and, as I see it, the main difference lies in what is involved in correcting (Hartley, 2003).”

Use of three dimensional faces and/or figures that interact with users during the VRT process have been proposed. This type of embodiment would need to address many difficult social and communication aspects like altering the way a user communicates to others because of the chosen media device of voice recognition and combinations with other devices.

Overall, technology is driving the market but nobody is considering the social aspects. You either learn the technology or get left behind. Here's another viewpoint that addressed the social effects of speech recognition technology:

> "You'll want a natural human language command interaction with your computer system, and you'll want a more natural way to navigate and manipulate objects in 3D worlds. The latest advances in speech recognition combined with other technologies such as artificial intelligence and expert systems – and the fact that businesses are starting to jump on the speech recognition trend- lead me to concluding that speech recognition technology has finally come of age (Clabby, 2002)."

Available programs

"Multifunctional/upgradeable [VRT] programs can be used…to run spreadsheets, to word processing, played games, edit digital content and do thousands of other things (Clabby, 2002)." VRT programs are often termed such names as command-and-control or continuous dictation, but this research will concentrate on two popular programs that I have direct access to: Dragon Naturally Speaking and the PS2 VRT game, Lifeline.

"Command and control systems can be shown to work for certain specialized applications, the question then becomes when will command and control systems become

effective tools for general human to computer interfaces (Clabby, 2002)?" Command and control systems eventually could be phased into voice only commands where a user could make complex navigation commands, access menu options and other tasks, which involves many factors mentioned in the following quote:

> "The aim of a continuous speech recognition system is…to provide an efficient and accurate mechanism to transcribe human speech into text. To make such a system ubiquitous, it is important that the system be able to handle a large vocabulary, and be independent of speaker and language characteristics such as accents, speaking styles, deficiencies (particularly important in spontaneous speech), syntax, and grammar (Matouések & Mautner, 2003)."

Dragon Naturally Speaking®

The training time is much longer than Lifeline and has more challenges. For example, when the microphone is activated by accessing the mouse, the VRT is open and active the microphone can pick up any sound. I often find myself holding my breath between writing because breaths usually show up as words like "in", "and", "a", "no" and "or". Menu shortcuts are also activated when I say words like "users" or "menu" when spoken as a single word with a pause between each word. Faster, continuous speech with words like "users" or "menu" without pauses did not activate menu shortcuts. The following study overview is a reminder that progress directly relates to the training process and social, personal goals and perceptions of voice recognition technology:

> "In fact, speaking in very short phrases or individual words can lessen the accuracy of this system as it uses the context of a phrase to help decipher what has been said. They did not correct all the recognition errors made in the training sessions, which is necessary if Dragon Naturally Speaking is to become a more accurate and recognizing what has been said. However, despite these problems, all but one of the participants reached an accuracy level of 80 percent or above when reading short phrases. It seems reasonable to assume that people who can train the program in the standard way are more likely to be successful. Thus, Dragon Naturally Speaking would appear to have more potential for someone who can read aloud long sentences of complex words and has fluent speech with only mild distortions and articulatory errors (Bruce, 2002)."

Lifeline

The technology of the program Lifeline is comparable to Dragon Naturally Speaking, even surpassing it at times, but doesn't have near as many practical uses as Dragon does. Lifeline allows a user to manipulate players on a space station using VRT through a special USB headset and activated through the game controller, which allows the subject can speak into the headset.

Lifeline has a significantly shorter training time and seems to work best when the subject speaks at a faster rate than Dragon. Unlike Dragon, Lifeline doesn't show the speech output in text form so errors are measured in the training section as (highest to lowest) "Brilliant", "Excellent", "Good" and "Bad" (prompts the subject to try again until achieving at least a "Good" rating), but not in gameplay (errors just occur in the results – e.g. the program navigates

you to the wrong area, completes the wrong action, etc.). Subjects must exercise analytical skills so they can use VRT to efficiently navigate the environments (e.g. “Go to the white object on the table, etc.). The ideal media vehicle for VRT constantly shifts, but currently the following quote compares PC computer with game consoles as a VRT vehicle:

> “Game consoles have strong advantages over PC architectures. Sony's PlayStation 2 qualifies from a technology standpoint because it acts as a supercomputer. These 'super' powerful game consoles use their power to display depth, light and shadow, and other special effects that make computer screens and television screens, live with vivid 3-D imagery. The sheer power of these dedicated controllers gives gamers audio and visual experiences that mimic real-life 3-D environments… a game consoles graphics are generally better than a PC's due to the consoles superior graphics processing power… another advantage that game consoles have over PC's is the ability to process 128 bit data streams… most personal computers can process only 32 bits of data at a time (Clabby, 2002).”

Navigation and manipulation of the Lifeline program will likely create more challenges for subjects compared to using a PC for Dragon. “No matter how enjoyable many gamers find today's 3-D gameplay, there are thousands upon thousands of people who have difficulty…but, if a game is fun, gamers will eventually master the navigation and controls (Clabby, 2002).”

Lifeline does have the unique advantage of personalizing VRT with actual, manipulatable human characters. “Users are able to reduce the complexity of movement in and about virtual, three-dimensional worlds (Clabby, 2002).” Many future applications of VRT have proposed the

incorporation of 3-D graphics faces that talk to subjects. The visuals create a different experience that can involve the subject on a deeper level than text-only VRT programs like Dragon Naturally Speaking.

Progress

"Applications such as automated call attendance, voice styling, dictation, customer service systems, stock quotes systems, airline and travel systems and the like that use speech recognition and artificial intelligence technologies are growing like wildfire (Clabby, 2002)."

The SALT forum has made advances to term VRT as a speech interface markup language which consist of a small set of XML elements, with associated attributes, object properties, events and methods, which apply a speech interface to web pages. SALT can be used with HTML, XHTML and other standards to write speech interfaces for both voice-only (e.g. telephony) and multimodal applications. The design principles of the SALT forum (forum, 2004) call for:

- Clean integration of speech with web pages
- Separation of the speech interface from business logic and data
- Power and flexibility of programming model
- Reuse existing standards for grammar, speech output and semantic results
- Wide range of available VRT devices
- Minimal cost of authoring across modes and devices

Much of the current VRT progress concentrates on dividing each word divided into sections, ideally by very small measures such as the millisecond, to improve the speech synthesis or "mapping" process. The simplified example figure below displays the computer mapping process and how different words might be distinguished:

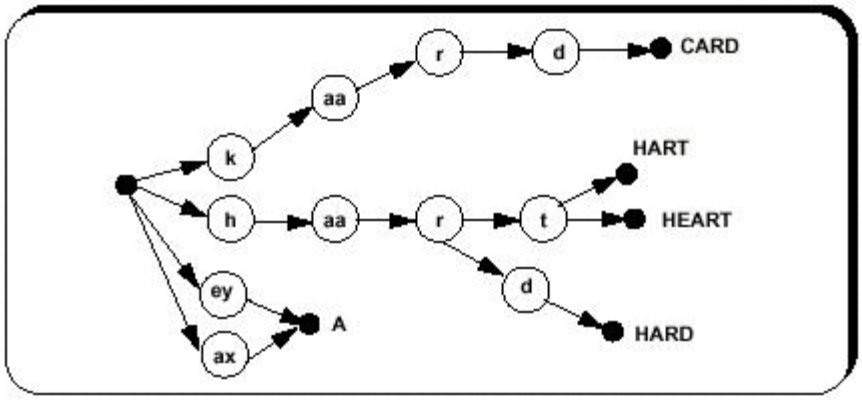

Figure taken from Deshmukh, Ganapathiraju and Picone (IEEE Signal Processing Magazine, vol. 16, no. 5, September 1999**).**

"Speech recognition systems today accomplish this task through an excruciatingly detailed analysis of the speech signal. Since we don't really know where words or phrases begin or end in the signal, or whether pauses occurred between the words, it is extremely productive to let the recognition system decide for us (in an optimal manner) where these units exist in the signal (Deshmukh, 1999)."

VRT seeks to be a more visual discourse by combining the technology with facial graphics on computer screen and other interfaces. "People will be more comfortable interacting with a machine if it behaves more like a person. It is possible to use a human façade in synthetic speech in the form of a visual animation of a person (Pitt & Edwards, 2003)."

The current wireless technology has also made great progress. "Instead of character driven keyboard like interfaces…we will see RPCs, palms, and mobile computing devices become capable of taking and acting upon speech commands (Clabby, 2002)."

VRT has made advances in transcribing methods for televisions broadcasts, equipment for full video-editing applications and audio indexing tools. “Users will be able to search for video segments that talk about specific subjects such as land mines, plane crashes or the President. In this case, the audio indexing can be used as a part of a video cataloger or editor to help in automating the editing of videos (Pringle, 2001).”

VRT has also helped disabled and handicapped individuals make great contributions and make advancements in society not previously possible. Here’s a feedback example from a handicapped research subject named MS (Bruce, 2002):

> “Using Dragon speaking naturally has helped me a great deal, I think the main thing it has enabled me to do, is slow my speech down to an acceptable slower pace. Because I am now dyslexic, my spelling is very bad, and my grammar likewise. Without Dragon I would be unable to write any letters at all. So for example, I can use it to communicate if I want to cancel a standing order or any other bills I want to amend. I’ve learnt how to use the e-mail facility, consequently I can write to my babies if I’m lucky, and they’ll respond by writing back.”

“One big impediment to the use of speech technology such as continuous dictation had been that it took a long time to teach your computer how to say things. That time frame has dropped from four hours to less than 2 (Clabby, 2002).” Once the computer limitations have been overcome, then complicated applications such as readjusting columns or creating tables could be much easier to accomplish. VRT is not a common application, but associating VRT

with a widely used application such as Windows, would also create useful data in a research setting and will be considered.

"Overall the bountiful benefits greatly outweigh the negatives and as hardware and software continue to improve the recognition's accuracy rates continue to increase and the acceptance of early adopters becomes visionary (Dreyer, 2002)."

Success

VRT has made necessary and predictable improvements such as gradually reducing the amount of program training and improving the accuracy rating, but many elements of VRT success have direct social implications in several areas and disciplines.

VRT has assisted the disabled and people with other communication impairments, so they can achieve more in society. The elimination of interfaces such as the keyboard, not only saves stress on hands and helps prevent carpal tunnel syndrome, improving the health of society overall. VRT has materialized in household devices, hands free car devices (e.g. CD player) and educational tools, especially for subjects who have stronger oral skills.

VRT has also reduced various physical strains including carpal tunnel syndrome and arthritis, which has reduced the amount of health care costs. VRT also promotes more physical activity that can enhance the overall process. "A voice recognition system frees one's hands and eyes. This allows additional activities to take place while composing…for example; writers can read and shuffle papers, while dictating…a dictionary can be kept open for reference when dictating a foreign language. Authors can walk about while writing or sit with their eyes closed. Possibly, these additional freedoms enhance thinking while writing, and reduce the time needed for planning (Hartley, 2003)."

Methodology

The research will compare two VRT programs, Lifeline and Dragon Naturally Speaking. The research will also measure data including human error during training and while subjects read training instructions. All activities of subjects will be recorded, even if they're not doing what they're instructed. The differences between these two programs may create challenges during the data collection and require extra time and resources to analyze differences like the following:

- Dragon Naturally Speaking – users will see what they're saying – bubbles appear at bottom or top left of computer screen.
- Lifeline – no visual of actual text, only waves during the initial training session and color graphics showing if the VRT is active or not.

Independent and dependant variables will be used to acquire data and analysis cause and effect relationships of social views related to VRT.

The following information outlines the main research aspects of each research question:

First question

1. In which ways can VRT be socially viewed as a discourse process, even as an alternative or in combination with traditional interfaces?

Traditional interfaces

- Lifeline combines VRT with a controller
- Dragon Naturally Speaking combines VRT with keyboard/mouse

Variables

- Independent – VRT software
- Dependant – subjects' reactions to VRT

Second question

2. What methods of VRT training produce the most successful results and practices?

VRT training improves social view and maximizes the program's usability.

Variables

- Independent – VRT software training
- Dependant – program usability

Facilitators should not assist the subjects during the training and refer them to each program's respective help documents when questioned by the subjects.

The proposed research will investigate if and how people to use VRT, then explore how their perceptions match the ideal goals of VRT. Individual and group traits and emotions will be compared to the actual methods of VRT software and systems (e.g. a subject alters his/her view about VRT taking a long time to train, when they complete the training in an acceptable time frame). Pre-test and post-test questions will be the main source for the comparison. The user will then go through the actual training and related VRT processes which will provide data to eventually assess what makes each person and/or group reach a satisfactory view of VRT.

The following information taken from Sergay Rybas' thesis work will serve as the model for measuring and acquiring usability data during the research process:

Measuring usability, according to an extended definition offered by Nielsen in Usability Engineering, includes an inspection of five usability components: learnability, efficiency, memorability, errors, and user's satisfaction.

Learnability/Ease of Learning. "The system must be easy to learn so that the user can rapidly start getting some work done with the system" (26). To measure learnability, Nielsen suggests "picking some users who have not used the system before and measure the time it takes them to reach a specific level of proficiency in using it" (29).

Efficiency of Use. "The system should be efficient to use, so that once the user has learned the system a high level of productivity is possible" (26). The measurement of efficiency, in Nielsen's opinion, is possible through involving experienced users and measuring time it takes these users to perform some typical test tasks.

Memorability. "The system should be easy to remember, so that the casual user is able to return to the system after some period of not having used it, without having to learn everything all over again" (26). "One way to measure memorability is to perform a standard user test with casual users who have been away from the system for a specific amount of time, and measure the time they need to perform some test tasks" (32).

Errors. The system should have a low error rate, so that users make few errors during the use of the system, and so that if they do make errors they can easily recover from them (26). An error in Nielsen's definition is "any action that does not accomplish the desired

goal" (27); "the system's error rate is measured by counting the number of such errors while performing a specific task" (27).

Subjective Satisfaction. "The system should be pleasant to use, so that users are subjectively satisfied when using it; they like it" (26). Nielsen suggests measuring satisfaction by "asking the users for their subjective opinion" (34) (Rybas, 2003). Specific questions will measure these opinions on a Likert scale ranging from strongly disagree, "1", to strongly agree, "5".

Ordinal measurements of performance time, error rate and feedback (as mentioned in detail above) will constitute the scales of measurement. These scales of measurement will be applied to the five components of usability will be put into these categories:

- Processing
- Understanding/benefits
- Challenges/errors
- Articulation/satisfaction

If the subjects are trained in VRT, it may be assumed that their usability of VRT increases. The usability testing will incorporate a control group and possibly a treatment group consisting of randomly assigned subjects, including disabled subjects. The subject group will also have varied bias/values that could affect the VRT object under research, namely the VRT programs.

The treatment group will then receive the treatment (one value of the independent variable) and the control group will receive no treatment or the standard treatment (the second value of the independent variable).

I have the programs/resources available for the research and, as a writer; I plan to produce documentation according to the research methodology while consulting an appointed research partner who will also help with quantitative aspects of the research, an area that I have limited experience in. Facilities would be located in a central position on campus for convenient subject access including separate rooms for individual training and research data acquisition.

A computer technology expert with experience in VRT would also be incorporated to assess/analyze the research equipment such as the microphone and sound card(s).

I feel it's important to understand social aspects of VRT to further educate the subjects, which consequently, educates the public. This research will also expose subjects to actual programs so they can discover the numerous practical applications/advantages that I've personally enjoyed. The research will also help accentuate the writing/editing process, recognize specific errors (program and human) and help make VRT a more common interface.

Uses for VRT have been based on society and their feelings about the technology. Society has a lot of power in deciding the success or failure of this technology, so people must understand the negative social views as well as the positive social views.

Questions during VRT training will focus on why the subjects do or do not accept VRT taking technology, user ability and personal feelings into account.

Many research questions have listed options to more easily collect data within parameters of the test and will have additional space for subject feedback. Research questions will also

incorporate the Likert scale to assess the quality of subject's personal experience using a 5-point range from strongly disagree, "1" to strongly agree, "5".

Research will include a general pre-test and post-test what the subjects like/dislike about VRT programs. This data will help analyze whether the research questions have been proven wrong, or has survived enough tests to be considered. Pre-test information will include demographical information (i.e. level of education, age, occupation, etc.), typing speed, general impression of VRT (to be compared with post-test) and an assigned number to keep track of the subjects. Post-test information will include general impressions of VRT, explanations and/or justifications of actions and thought processes, VRT usability and future plans regarding VRT. Questions in the pre-test and post-test will include general criticism, comfort level, analysis of VRT, descriptions of VRT, descriptions of VRT processes/training and program elements (relative to both programs and only one or the other).

Qualitative research would consist of data collection such as surveys and focus groups. Quantitative research would include statistics mainly produced by the research partner with extensive research experience. I've experienced limited amounts of statistics - many of the data models I saw in the VRT books that I researched were currently beyond my comprehension.

The research will also come from the social perspective which will look at situation from social construction or certain practices, such as computer interfaces like a controller or keyboard/mouse and analysis the subject's expectations and/or dependence on these interfaces.

Human errors and program errors will be measured. Human error will consider an average percentage from a sampling at the beginning, middle and end of training (it would too much time and resources to measure each instance).

Program errors will be organized into in three categories:

1. Substitution error – program misrecognizes spoken word
2. Absence error – word spoken not recognized
3. Addition error - program

Dragon Naturally Speaking will use general statistics and a computed accuracy rating (using total number of words and measured errors) calculated into a percentage. Lifeline measurements will involve general statistics and the initial training ratings of “Brilliant”, “Excellent”, “Good” and “Bad” (prompts the subject to try again).

Conclusion

VRT practices and general knowledge will advance through:

- practical use
- new research possibilities
- a better understanding of facts that will allow a more appropriate course of action for parties involved in VRT

Social attitudes and VRT advantages/disadvantages will be treated in an objective manner. Overall, this research will be simple and inexpensive to design, but could be time-consuming. Various methods will provide predictable and measurable data, while stressing a close relationship between task and the subject’s action. Combined activity of the subject and technology and the transfer of learning from one interface to another will also provide helpful data.

This research will find identifiable traits and reasons why a given person would or would not choose to use VRT. This project is worth researching because it advances knowledge about VRT and has many practical applications that anyone can use.

References

Bajaj, Sanjay. 2004. Expanding the Reach of Speech Recognition. *Voice Recognition*, 16.

Bruce, C. E., Anne; Coleman, Michael. (2002). Writing with voice: an investigation of the use of a voice recognition system as a writing aid for a man with aphasia. *International Journal of Language & Communication Disorders, 38*(2), 131-148.

Clabby, J. (2002). *Visualize this: collaboration, communication, and commerce in the 21st century*. Upper Saddle River, NJ: Prentice Hall PTR.

Deshmukh, N. G., Aravind; Picone, Joseph. (1999, June 15). Hierarchical Search for Large Vocabulary Conversational Speech Recognition. *IEEE SP Magazine*, 58.

Dreyer, K. M., Amit; Thrall, James. (2002). *PACS - a Guide to the Digital Revolution*: Verlag New York, Inc.

Entwistle, M. (2003). The Performance of Automated Speech Recognition Systems under Adverse Conditions of Human Exertion. *International Journal of Human-Computer Interaction, 16*(2), 127-140.

Forum, S. (2004). *Speech Application Language Tags (SALT) Technical White Paper*. Retrieved March 15, 2004, from www.saltforum.org/Saltforum/downloads/SALTTechnicalWhitePaper.pdf

Ganapathtrafu, A. P., Joseph. (1999, June 15). Electronic & Computer Engineering. *IEEE SP Magazine*.

Gardner-Bonneau, D. (1999). *Human factors and voice interactive systems*. Boston: Kluwer Academic Publishers.

Hartley, J., Sotto, Eric; Pennebaker, James. (2003). Speaking Versus Typing: a Case Study of the Effects of Using Voice-recognition Software on Academic Correspondence. *British Journal of Education Technology, 34*(1), 5-16.

Johnson, G. C., Lynne. (2001). "You Talking to Me?" Exploring Voice in Self-Service User Interfaces. *International Journal of Human-Computer Interaction, 13*(2), 161-186.

Matouések, V., & Mautner, P. (2003). *Text, speech and dialogue: 6th International Conference, TSD 2003, éCeskâe Budéejovice, Czech Republic, September 8-12, 2003 : proceedings*. Berlin; New York: Springer.

Pitt, I., & Edwards, A. D. N. (2003). *Design of speech-based devices : a practical guide*. London ; New York: Springer.

Pringle, D. (2001). *Video-editing: IBM's Speech Technology to the Rescue*. Retrieved February 19, 2004, from LeJournal/article.asp-articleIndex=308.htm

Rybas, S. (2003). *Online Education and User's Diversity: Usability as a Defining Feature of Effectiveness.* Bowling Green State University.