Evaluating the Effects of Automatic Speech Recognition Word Accuracy

Hope L. Doe

Thesis submitted to the Faculty of the Virginia Polytechnic Institute and State University in partial fulfillment of the requirements for the degree of

Master of Science
In
Industrial and Systems Engineering

Dr. Brian M. Kleiner, Chair
Dr. Andrew W. Gellatly
Dr. Robert C. Williges

July 10, 1998
Blacksburg, Virginia

Keywords: Automatic speech recognition, word accuracy, user satisfaction

Evaluating the Effects of Automatic Speech Recognition Word Accuracy

Hope L. Doe

ABSTRACT

Automatic Speech Recognition (ASR) research has been primarily focused towards large-scale systems and industry, while other areas that require attention are often over-looked by researchers. For this reason, this research looked at automatic speech recognition at the consumer level. Many individual consumers will purchase and use automatic software recognition for a different purpose than that of the military or commercial industries, such as telecommunications. Consumers who purchase the software for personal use will mainly use ASR for dictation of correspondences and documents. Two ASR dictation software packages were used to conduct the study. The research examined the relationships between (1) speech recognition software training and word accuracy, (2) error-correction time by the user and word accuracy, and (3) correspondence type and word accuracy. The correspondences evaluated were those that resemble Personal, Business, and Technical Correspondences. Word accuracy was assessed after initial system training, five minutes of error-correction time, and ten minutes of error-correction time.

Results indicated that word recognition accuracy achieved does affect user satisfaction. It was also found that with increased error-correction time, word accuracy results improved. Additionally, the results found that Personal Correspondence achieved the highest mean word accuracy rate for both systems and that Dragon Systems achieved the highest mean word accuracy recognition for the Correspondences explored in this research. Results were discussed in terms of subjective and objective measures, advantages and disadvantages of speech input, and design recommendations were provided.

Acknowledgements

I would like to thank Dr. Brian M. Kleiner, Dr. Robert C. Williges, and Dr. Andrew W. Gellatly for their time and support in advising me throughout this research process. I appreciate the guidance and encouragement you provided me in conducting this research.

I would like to dedicate this to Hepsie L. Nickelson, my grandmother, who will never be forgotten.

I would like to thank my family, you all provided me with so much love and support. Mom, thank you for your prayers, encouragement, and your sincere belief in me. I would also like to thank my sister and brother for their friendship and support. I cannot thank you enough for all you have done.

I would like to thank my friends for their support and encouragement

Table of Contents

List of Tables

List of Figures

CHAPTER 1 INTRODUCTION

Background

For many years, since the earliest days of computing, enabling machines to understand human speech has been a goal of researchers (Randall, 1998). This is in part due to the belief that speech is the ultimate human/machine interface, primarily because speech comes very natural to most (Randall, 1998). Automatic speech recognition (ASR) researchers continue to make significant technological advances in the area. In the past, speech has been available but very costly. Today speech recognition software for computers is not only commercially available but also reasonably priced. The significant strides being made by numerous manufacturers to provide consumers with reasonably priced software is leading to increased reliability and popularity by consumers.

Automatic speech recognition technology is used for several applications and by numerous individuals from doctors and lawyers to students and teachers. Automatic speech recognition technology permits human speech signals to be used to carry out preset activities. Once the system detects and recognizes a sound or string of sounds, the recognizer can be programmed to execute a predetermined action (Barber and Noyes, 1996). However, speech input presents advantages and disadvantages over other input methods.

Many groups of individuals have and are benefiting from ASR in human machine-interaction, human-to-human communications, and as a means of control in their immediate environment in which they live or work. Researchers are concentrating efforts in this area particularly because they realize that voice recognition may become the next primary user interface. Thus the subjective opinions of the user in using such systems is important and designing or redesigning in order to meet the expectations of the user (Preece, 1993)). Issues such as how users must train the systems and what is involved during this training, is important in examining users expectations and preferences.

A keen interest in automatic speech recognition lies within human-machine interaction, specifically interaction with computers. Today in most schools, businesses, and increasingly in homes, computers are being used to augment daily life. Individuals use computers to manage everything from business transactions to completing homework assignments. Despite the fact that automatic speech recognition can be used in an increasing number of applications, certain physical and psychological environments are still deemed inappropriate for this technology (Barber and Noyes, 1996). They are domains in which there is high ambient noise levels, elevated levels of stress, and extremes of vibration, pressure, and acceleration (as found in the aircraft cockpit) (Barber and Noyes, 1996).

The creation of speech recognition technology software is revolutionizing the way people receive and process information. Users can now enter text and data into a personal computer verbally. This new technology allows users to voice commands in order to perform tasks that would typically require a mouse to open menus or move the cursor. Speech recognition software can be used in conjunction with a PC or Mac and the aid of a microphone headset. Determining if and how the human, the system, or both should perform or carry out a task associated with using such systems becomes important. This process is known as function allocation (Wilson and Corlett, 1990). Therefore, function allocation should be addressed with respect to automatic speech recognition.

Speech recognition has made considerable progress in the past years. Systems have emerged and continue to emerge with impressive accuracy (Lee, Hon, Reedy, 1990). Constraints such as 1) speaker dependence, 2) isolated words (discrete speech), and 3) small vocabulary are what most systems seek to overcome. The most difficult constraint for systems to overcome has been found to be speaker independence (Lee, Hon, Reedy, 1990). Manufacturers have been successful in producing speaker dependent systems. Speaker dependent systems require a speaker to “train” the system before reasonable performance can be expected. Systems that comprehend isolated word recognition have been in existence for many years. However, error rates increase drastically from isolated-

word to continuous speech recognition. A 280 percent error rate increase from isolated -word to continuous speech recognition was reported in a study done by Bahl et al (1981). However, recent advances allowed for continuous speech recognition systems to be introduced and research concentration is being placed in this area. Continuous speech research thrives because only through continuous speech can desired speed and naturalness of man -machines communications be achieved (Lee, Hon, Reedy, 1990).

Large vocabulary produces some problems and constraints. As a system's vocabulary increases, the number of confusable words (i.e., words that the system may mistake for another because they are closely related in pronunciation) increases. Despite the fact that ASR systems are error-prone, users of the systems do expect satisfactory results (Wilpon, 1995). However, large vocabulary systems are still needed for many applications, such as dictation (Lee, Hon, Reedy, 1990).). Therefore, word accuracy results obtained and error-correction procedures of ASR systems become an issue

Problem Statement

This research seeks to examine how speech recognition software system training (i.e., training the system to recognize a user's speech) and varied levels of error-correction time affect word accuracy. This study examines the relationships between (1) speech recognition software system training and the system's overall performance (i.e., word accuracy), (2) error-correction time by the user and improved word accuracy, and (3) correspondence type and word/command accuracy. The system's overall performance is also examined in regards to user satisfaction.

Research Objectives

The practitioner literature provides indirect information on what proportion of system training is necessary for a system to achieve an acceptable level of performance. However, there is no research to support if system-required training

produces satisfactory results for the user and satisfactory performance for the system.

The objectives of this research are to:

(1) Determine what level (i.e. percentage) of word accuracy is produced by speech recognition software system-required training.
(2) Determine whether and to what extent word accuracy increases with varied levels of error-correction time.
(3) Determine if the level of word accuracy achieved by the system affects users' satisfaction.

Research Questions and Hypotheses

Questions that the research address and the corresponding hypotheses are presented below.

Research Question 1- What is the relationship between the type of correspondence dictated and word/command accuracy rate?

Hypothesis- Business correspondences will achieve the greatest word accuracy rate.

Research Question 2- What is the relationship between varied levels of error-correction time by the user and word accuracy?

Hypothesis- Increased error-correction time by the user will provide an increased word accuracy rate.

Research Question 3- What is the relationship between varied levels of error correction time and user satisfaction?

Hypothesis- User satisfaction will be influenced negatively by lower word accuracy recognition for the shorter periods of error-correction versus the increased error-correction condition.

Research Variables

This section describes the independent and dependent variables to be used in the research (See Figure 1). The independent variables that will be manipulated in the study are error-correction time, system type, and correspondence type. The

two dependent variables that will be used in the study are word accuracy (percentage of words/commands recognized correctly), and user satisfaction levels.

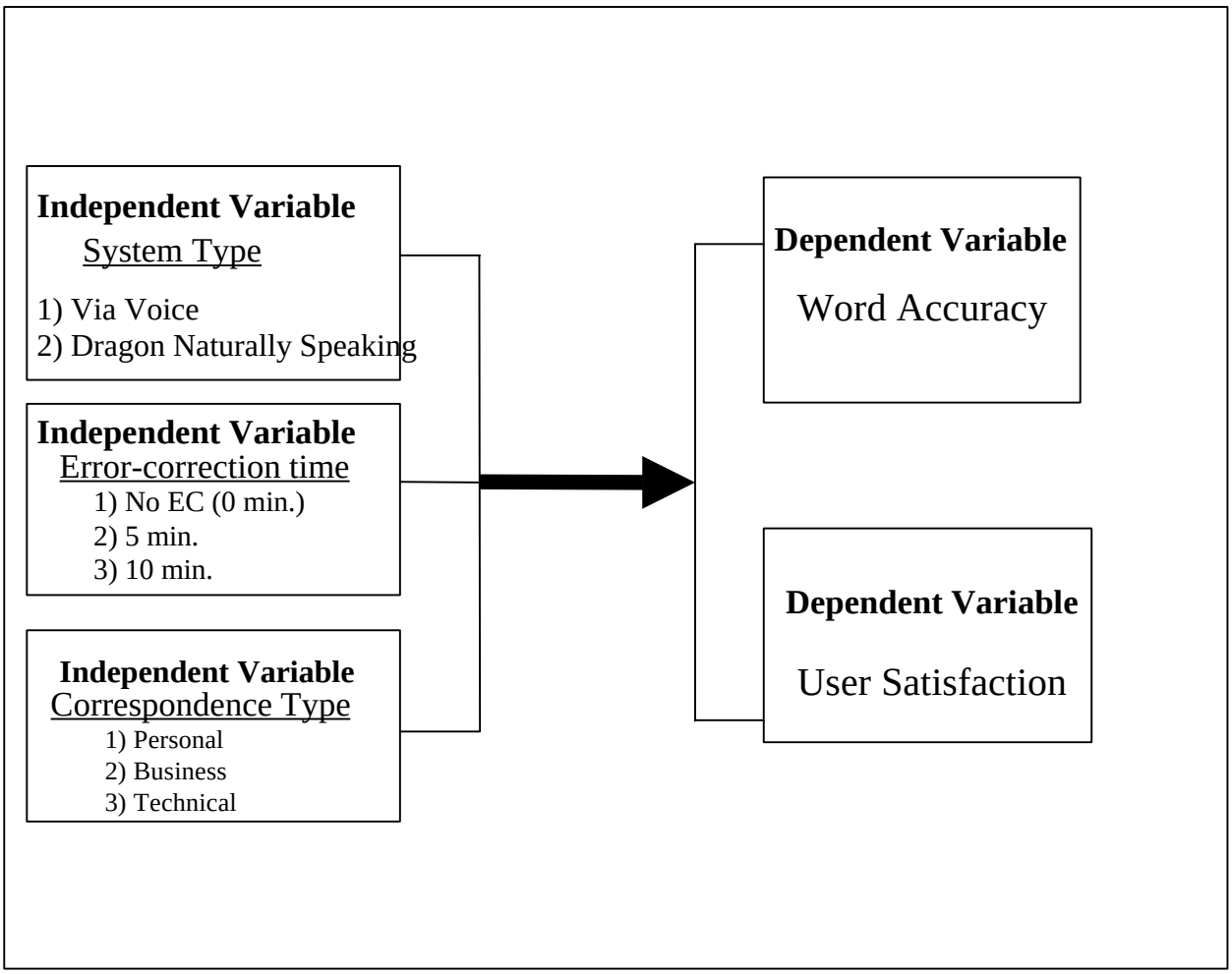


Figure 1. Research Model

Chapter 2 Literature Review

Humans have the ability to communicate with other humans in various ways, which includes but is not limited to, body gestures, the printed text, pictures, drawings, and voice (Shaefer, 1995). However, voice communication is used widely in our daily activities. Since speech has been demonstrated to be an effective and efficient way for humans to express ideas and requests, it does not come as a surprise that a desire exists to communicate with machines by voice. This is in part due to very obvious advantages: 1) the natural mode of communications is speech, 2) when a human's hands and/or eyes are occupied, voice control is especially appealing, and 3) handicapped individuals could benefit from voice communication (Schaefer, 1995).

Despite the continuous technological advances being made in relation to computers and their use, problems with the human-computer interface still exist. Norman (1988) stated that users were not well-served by existing practices and that the problem requires dedicated efforts, with new techniques of software engineering, new evaluation procedures, and specialized groups of interface designers.

Human-Machine Communication through Voice Input

The voice-processing field encompasses five broad technology areas: 1) voice coding, 2) voice synthesis, 3) speaker recognition, 4) speech recognition and 5) spoken language translation. Voice coding is the process of compressing the information in a voice signal so as to either transmit it or store it over a channel whose bandwidth is significantly smaller than that of the uncompressed signal (Rabiner, 1995). Voice coding technology has been widely used in network transmissions and has been utilized in cellular systems and used as a driving force for security applications in the U.S. government. The storage of voice messages in voice mailboxes is considered one of the most important applications of voice coding used for the purpose of storage. The digital telephone answering machine

also relies heavily on voice coding in which both voice prompts and voice messages are compressed and stored in the machine's local memory.

Voice synthesis is the process of creating a synthetic replica of a voice signal to transmit a message from a machine to a person, with the purpose of conveying the information in the message (Rabiner, 1995). Several key applications have emerged and continue to emerge: a voice server for assessing electronic mail messages remotely over a dialed-up telephone line, automated order inquiry, remote student registration, and proofing of text documents, and providing names, addresses, and telephone numbers in response to directory assistance given.

Speaker recognition can be defined as the process of either identifying or verifying a speaker by selecting individual voice characteristics (with the main purpose of restricting access to information, networks, or physical demands). Speaker recognition technology is one of the many applications where the computer can outperform a human (Rabiner, 1995). The computer is able to identify a speaker from a given population or can verify an identify claim from a named speaker with greater accuracy than that of a human.

Speech recognition can be stated as the process of extracting the message information in a voice signal so as to control the actions of a machine in response to spoken commands (Rabiner, 1995).

Spoken language translation is the process of recognizing the speech of a person talking in one language, translating the message content to a second language, and synthesizing an appropriate message in the second language for the purpose of providing two-way communication between people who do not speak the same language. Spoken language translation relies heavily on speech recognition, speech synthesis, and natural language processing and is the long-term goal of voice processing technology (Rabiner, 1995).

Speech Recognition

Speech, a stream of utterances, produce time varying sound pressure waves of different frequencies and amplitudes. Speech recognition occurs when a corresponding sequence of discrete units (i.e., phonemes, words, or sentences) are derived from sound waves or acoustical waveforms (Moore, 1994). The goal of most, if not all computer-based speech recognition systems is to model human speech recognition. However, computer-based systems do not yet have the capability and flexibility of understanding speech as humans.

Two types of speech recognition have emerged in the PC market place. The first type enables one to speak commands, such as “bold” or “new window”, to the software. Such capability requires a sound board, a microphone, and software that will add speech capabilities to the application. Dictation software is the second type. The principal goal of the second is to emulate the familiar business in which a manager dictates some type of correspondence to a secretary (Randall, 1998). Dictation software has been and is being designed to save time on typing (Ross, 1997).

Technologies such as automatic speech recognition and text-to-speech have been under development since the early days of computer technology. Automatic speech recognition had made significant progress by the 1980s and was able to make practical speech-driven data entry systems (Oberteuffer, 1995). Automatic speech recognition’s development has been carried out by companies and universities. The early 1990s provided us with voice command systems for personal computers and telephone-based systems (Oberteuffer, 1995). Today, users have access to very powerful, large-vocabulary systems for the creation of text entirely by voice.

However, computer-based systems do not yet have the capability and flexibility of understanding speech as humans.

> “Most computer-based systems use a similar process for speech recognition. In the first stage, the computer receives speech input and the signal is converted from an analog signal to a digital signal in a digital

signal processor (DSP). The DSP conversion produces a digitized representation of the acoustic signal. Most systems use a vector quantization (VQ); the VQ representation is used as algorithms have been produced that reduce the amount of data storage and computation time. In the second stage, the digital signal is compared to digitized speech patterns stored in databases" (Moore, 1994, p. 8).

ASR devices can usually accommodate three types of speech: 1) isolated word recognition, 2) connected word recognition, and 3) connected speech recognition (i.e., continuous speech recognition) (Barber, 1991). Isolated or discrete word recognition is the simplest speech type because it requires the user to pause between each word. Connected word recognition is capable of analyzing a string of words spoken together, but not at normal speech rate. While connected speech recognition or continuous speech allows for normal conversational speech. Such devices may require a user to train the system referred to as speaker-dependent or talker-dependent. Devices that do not require a user to train the system is referred to as speaker-independent or talker-independent.

Understanding continuous speech, natural or conversational speech, is the goal of ASR systems today. However, when words are spoken in a natural flow (i.e., continuous speech), they become more difficult to recognize since there are no pauses between words and phrases. A speech recognizer is then faced with the task of "guessing" where one word ends and another begins. The "guessing" is where the statistical analysis takes place to produce the most likely word or words to produce a correct sentence. Search algorithms and grammar modeling can improve the recognition in continuous speech (Moore, 1994). Figure 2 depicts a general system for training and recognition (Makhoul and Schwartz, 1995).

The first step in the training and recognition process is feature extraction. Feature extraction is performed to reduce the variability of the speech signal (Makhoul and Schwartz, 1995). During training, the process of estimating speech model parameters from actual speech data occurs. Once the system receives the training speech, the text of the speech, and the phonetic spellings of all the words,

the phonetic Hidden Markov Model (HMM) is estimated automatically using a forward-backward algorithm (Makhoul and Schwartz, 1995). It is important that the lexicon (i.e., vocabulary) contain words that would be expected to occur in future data. Grammar is another aspect of training that is needed to aid in the recognition. Grammar places constraints on the sequences of the words that are allowed. Without grammar, all words would be considered equally likely at each point in an utterance (Makhoul and Schwartz, 1995). The recognition process also starts with feature extraction. Once given the sequence of feature vectors, the word HMM models and the grammar, the recognition is a large search among all possible word sequences for that word with the highest probability to have generated the computed sequence of feature vectors (Makhoul and Schwartz, 1995).

The Hidden Markov Model (HMM) has been identified as the most widely used statistical model for continuos speech (Acero, 1993). The HMM is a statistical model that uses two transitions between states to quickly search through a database. Two sets of probabilities are provided for each transition: 1) going to the next stage and 2) defining the conditional probability that a word is correct (Moore, 1994).

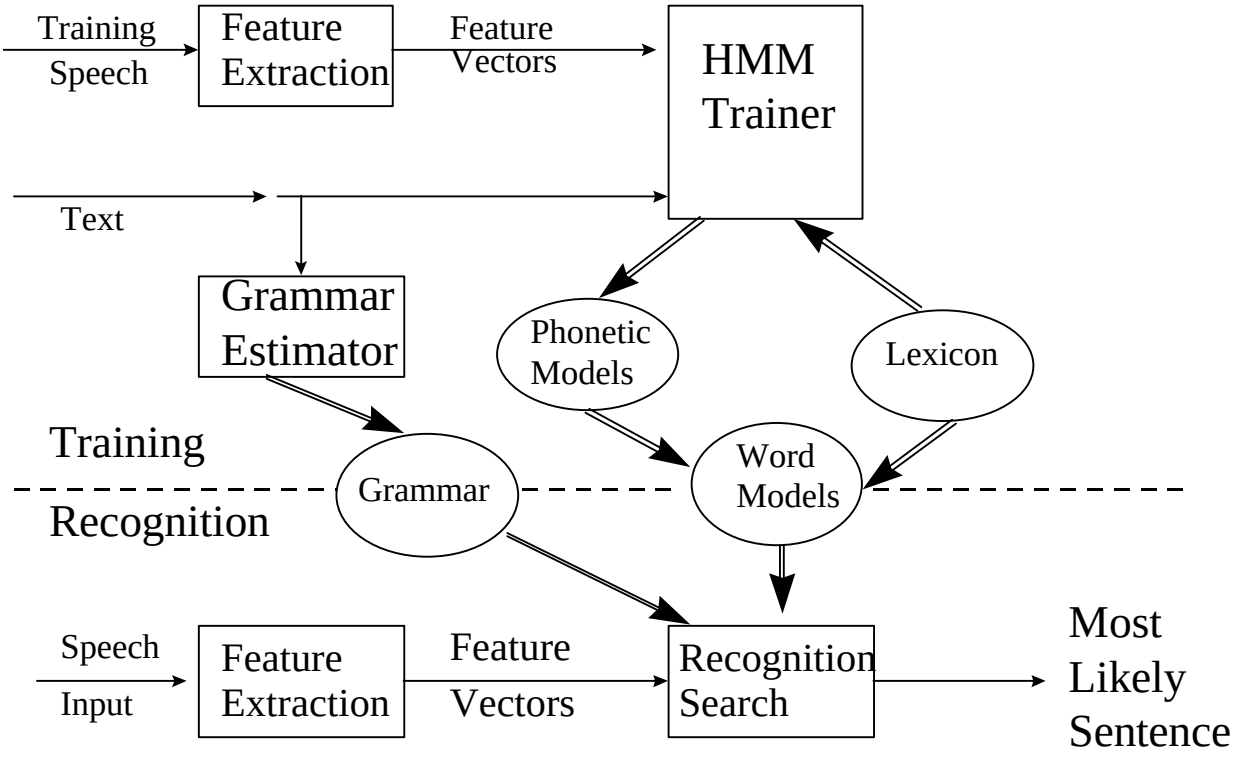


Figure 2. General system for training and recognition (Makhoul and Schwartz, 1995)

User Interfaces for voice applications

A successful human-machine interaction or human-human interaction is one that accomplishes the task at hand efficiently and easily from the humans perspective (Kamm, 1995). In designing an effective user interface for voice application, three major considerations must be taken into account: 1) the information requirements of the task, 2) the limitations and capabilities of the voice technology, and 3) the expectations, expertise, and preferences of the user (Kamm, 1993). From a human factors perspective, the users' expectations and preferences are important factors. Many new users will expect a human-computer voice interface to allow the same conversational speech style that is used between humans. For this reason, three common behaviors of humans are very difficult to overcome: 1) speaking in a continuous manner, 2) anticipating responses and

speaking at the same time as the other talker, and 3) interrupting pauses by the other talker as implicit exchange of turn and permission to speak.

Novice and expert users will have different expectations and needs. Novice or infrequent users will likely require instructions and/or guidance through a system as they try to build a cognitive model of how the system works and how he or she should interact with the system. While experienced users may want to bypass instructions and move through the interaction more efficiently, a successful user interface for automated system will accommodate for the needs of novice users and the preferences of expert users (Kamm, 1995).

A major goal of speech recognition systems is to limit erroneous actions. Providing the user with feedback about the application's state and to request verification that the system's interpretation is what the user intended is one way to limit mistaken actions. However, providing feedback and eliciting confirmation for each fragment (i.e., piece) of information exchanged between the user would most likely result in inefficient interaction. Therefore in some instances, when the user is provided sufficient information to establish that the system's response was correct it may be reasonable to forgo some of the exchanges (Kamm, 1995).

Error recovery procedures are an inevitable requirement in a user interface. The aim of error recovery procedures "is to prevent the complete breakdown of the system into an unstable or repetitive state that precludes making progress toward task completion (Kamm, 1993, p. 10039)." Error recovery requires the cooperation of the user; both the system and the user must be able to initiate error recovery sequences. The first step in detecting errors is the feedback and confirmation dialogues.

Human-Computer/Communication Interfaces

Designers and researchers alike realize that different users have different needs and that different stages of interaction may exist for a single user. Norman (1988) identified four possible distinct stages of a person interacting with a computer: intention, selection, execution, and evaluation. Each stage of

interaction has different methods, goals, and even needs. Therefore it becomes important to realize that an interface for one stage may not be appropriate for another.

Therefore, in developing effective human-computer interfaces, allocation of functions to be performed by the user becomes one of the most important categories of design decisions (Brown, 1988). Despite the fact that allocating functions to be designed by the user or the computer should be based on the capabilities of both, decisions regarding allocation are often either based on hardware, software, and cost concerns, or made without any explicit analysis of the allocation of functions. Allocation includes making decisions like the following (Brown, 1988):

1) Will the user be required to commit the commands needed to perform a particular task to memory, or will a list of available options be presented?
2) Will the user be required to perform mental arithmetic on displayed data, or will the computer system calculate and display the data in the form required to perform the user's task?
3) Will the software keep track of previous user entries in a multiple step procedure, permitting the user to correct an error in a later step without starting the whole procedure over?
4) Will the display highlight suspect parameters to draw to the user's attention? Or will the software monitor all parameters automatically and recommend actions to the users?

Allocation of functions to be performed by the user is an important area in regards to ASR. Users are required to use commands in order to perform error-correction tasks. ASR users may also benefit from highlighted information to inform them of problems or if they are speaking to quickly or not loud enough for the system to interpret what they are saying.

Visual, Auditory, and Tactile Modalities

For many years, human factors engineers have been concerned with how information is displayed. In some instances, the selection or design of displays used for transmitting information and the selection of the sensory modality is a predetermined conclusion, such as using vision for road signs (Sanders and McCormick, 1993). However, when there is an option, certain advantages of one over another can depend on many considerations. Due to its ability to obtain the user's attention, audition tends to have an advantage over vision in observation (vigilance) types of tasks. Sanders and McCormick (1993) provided an extensive comparison of audition and vision which indicates the kinds of circumstances in which each of the two modalities tend to be more useful. The comparisons are based on considerations of substantial amounts of research and experience relating to the two sensory modalities. The tactile sense has relevance in specific situations; such as with blind persons and other special circumstances when the visual and auditory sensory modalities are overloaded (Sanders and McCormick, 1993). However, the tactile sense is not used very extensively as a means of transmission of information. Tactile displays have been mainly used as substitutes for hearing, especially as aids to the deaf and hearing-impaired and as substitutes for seeing, aiding the blind.

In determining the kinds of displays that would be preferable for a specific type of information, one must look at the nature of the information in question. In selecting a display modality, a major decision is whether to use an auditory or a visual form of presentation. Figure 3 depicts when auditory or visual presentation should be used.

Use Auditory Presentation if:	Use Visual Presentation if:
1. The message is simple.	1. The message is complex.
2. The message is short.	2. The message is long.
3. The message will not be referred to later.	3. The message will be referred to later.
4. The message deals with events in time.	4. The message deals with location in space.
5. The message calls for immediate action.	5. The message does not call for immediate action.
6. The visual system of the person is over-burdened.	6. The auditory system of the person is over-burdened.
7. The receiving location is too bright or dark-adaptation integrity is necessary.	7. The receiving location is too noisy.
8. The person's job requires moving about continually.	8. The person's job allows him or her to remain in one position.

Figure 3 When to Use the Auditory or Visual Form of Presentation

(Sanders and McCormick, 1993)

Applications

To date, there is no theory of tasks and environments that predict when voice would be a preferred modality of human computer communication (Cohen and Oviatt, 1995). However, a number of situations have been identified in which spoken communications with machines would be advantageous: when the user's hands or eyes are busy, a limited keyboard and/or screen is available, disabled users, and when pronunciation is the subject matter of computer use, and when natural language is preferred (Cohen and Oviatt, 1995).

Spoken interaction with machines is a situation in which a user's hands' and/or eyes are busy performing another task. When users are able to use speech

to communicate with a machine, they are free to pay attention to their task, as opposed to them breaking away to use a keyboard or other input device (i.e., beneficial for automobile drivers and in many cockpit control situations). Many field studies of high accurate speech recognition systems with hands/eyes-busy task have found that spoken input leads to higher task productivity and accuracy (Cohen and Oviatt, 1995).

Telephone based applications

Telephone based applications that replace or augment operator services are the most prevalent current use of speech recognition (Cohen and Oviatt, 1995). Hundreds of millions of callers each year are assisted, resulting in tremendous savings. Speech recognizers used for telecommunications applications accept limited vocabulary. However, certain key words are the input, and the system is expected to function with high reliability. One of the most challenging potential application of telephone-based spoken language technology is the interpretation of language where two callers speaking different languages can engage in a conversation with the aid of a spoken language translation system (Cohen and Oviatt, 1995). The largest ongoing commercial application is the automation of operator services. Initially, by simply using the words “yes” and “no”, many telephone companies saved hundreds of millions of dollars a year (Seelbach, 1995). Services have now been expanded to include selection of payment such as “collect,” “person-to person,” and “third party,” as well as help commands (e.g., “operator”). Applications used in the early 1990s have been and are currently being expanded to handle larger vocabularies, “out-of-vocabulary words,” and the ability to speak over prompts or “barge in”.

The telecommunications industry is constantly striving to provide the products and services that people will desire. The industry realizes that automatic speech recognition is one of the technologies that will become common and that it will provide users with more freedom on when, where, and how they access information (Wilpon, 1995).

Applications for users with disabilities

Voice technology can also be used to assist users with disabilities. The motorically impaired users could use speech recognition as a means to control certain household appliances and wheelchairs. The possibility of having spoken input through the use of speech recognition systems may even become a prescribed therapy for carpal tunnel syndrome (Cohen and Oviatt, 1995). Individuals with carpal tunnel may be prescribed to use automated speech recognition systems in place of using a typewriter or computer. Even limited speech recognition increases control for individuals with disabilities (Seelbach, 1995).

Military and Government

The Army foresees many applications of human-machine communication by voice (See Table 1). Three major uses include: 1) Command and Control on the Move (C2OTM), 2) the Soldier's Computer, and 3) voice control of radios and other auxiliary systems in Army helicopters (Weinstein, 1995). C2OTM is an Army program whose focus is to ensure the mobility of command and control for potential future needs. Since typing is often a poor input medium for mobile users, whose eyes and hands may be busy, a voice or speech-based input medium may be beneficial. Foot soldiers could use speech recognition to enter reports that could be transmitted to command and control headquarters. Repair and maintenance in the field can be simplified through voice access providing repair information. The soldier's computer, an Army Communications and Electronics Command program, responds to the information needs of the modern soldier. Speech recognition can be essential for control of radios and other devices in Army helicopters (Weinstein, 1995). Navy applications include: aircraft carrier flight deck control and information management, SONAR supervisor command and control, and combat team tactical training. The objective of the aircraft carrier flight deck control and information management application is to provide speech recognition for updates to aircraft launch, recovery weapon status, and maintenance information. The Air Force has had a vested interest in speech input/output for the cockpit and proposes

to include human -machine communication by voice (Weinstein, 1995). Cockpit applications range from voice control of radio frequency settings to an intelligent Pilot's system. The Federal Bureau of Investigation (FBI) also has numerous potential applications for speech and language technology in criminal investigations and law enforcement. Functions of interests to FBI agents include 1) voice check-in, 2) data or report entry, 3) rapid access to license plate or description-based data, 4) covert communication, 5) rapid access to map and direction information, and 6) simple translation of words or phrases (Weinstein, 1995)

Table 1. Matrix of human-machine communication applications by voice interest of military and government users

Users	Data Entry	Data Access	Command & Control	Training	Translation
Soldier	**	*	*	*	*
Naval Officer	**	**	**	**	
Pilot	**	*	**		
Agent	**	**		*	*
Commander		**	**		**

** = primary application

* = additional application (Adapted from Weinstein, 1995)

Problems with Speech Recognition

Automatic speech recognition is often viewed as a mapping from the speech signal to a sequence of discrete entities such as phonemes (i.e., speech sounds), words, and sentences (Makhoul and Schwartz, 1995). A major obstacle in obtaining high-accuracy recognition is the large variability in the speech signal characteristic. The three components of variability are: linguistic variability, speaker variability, and channel variability. Linguistic variability includes the effects of phonetics, phonology, syntax, semantics, and discourse on the speech signal. Speaker variability includes intra- and interspeaker variability and the effects of coarticulation. Channel variability includes the effects of background noise and the transmission channels (e.g., microphone, telephone, and reverberation). The above-mentioned variabilities sometimes interfere with the intended message and the problem must be unraveled by the recognition process.

Robustness against speech variation is one of the most important issues in speech and speaker recognition. There are many causes of speech variation. The main causes of speech variation can be classified based on whether they originate in the speaking and recording environment, the speakers themselves, or the input equipment, indicated in Table 2. Additive noises can be classified as stationary or

nonstationary, with the most typical nonstationary noise being other voices. In addition, noise can be classified according to whether they are correlated or uncorrelated to speech.

Table 2. Main causes of speech variation

Environment	Speaker	Input Equipment
Speech-correlated noise- reverberation, reflection	Attributes of speakers- dialect, gender, age	Microphone (transmitter) Distance to the microphone
Uncorrelated noise- additive noise (stationary, nonstationary)	Manner of speaking- breath and lip noise, stress, rate, level, pitch, cooperativeness	Filter Transmission system- distortion, noise, echo Recording equipment

(Adapted from Furui, 1995)

Current research issues and new speech recognition challenges

The major focus of speech research is now on producing systems that are accurate and robust but that do not impose unnecessary constraints on the user (Atal, 1995). Speech technology has advanced to the point where it is now useful in various applications. However, the prospect of a machine understanding speech as humans do is still far away. Using human performance as a benchmark shows us how far researchers are from the goal. Major roadblocks faced by the current technology must be removed for speech technology to be widely used. Current research issues include (Atal, 1995):

- Ease of use -if speech technology is not easy to use, it will have limited applications
- Robust performance- the capability of a recognizer working well with different speakers and in the presence of noise

- Automatic learning of new words and sounds- can the systems learn to recognize new sounds or words automatically
- Grammar of spoken language- since the grammar for spoken language is different from that used in carefully written text
- Control of synthesized voice quality- can more flexible intonation rules be used
- Integrated learning for speech recognition and synthesis- can methods be developed for the training of both the recognizer and synthesizer in an integrated manner.

Another factor behind the progress that has been achieved in ASR is the application of hidden Markov models (HMMs). In applying speech recognition or synthesis technology to real services, algorithms become very important (Nakatsu and Suzuki, 1995). However, the algorithms suffer from fundamental shortcomings that must be overcome, such as robustness of algorithms (Nakatsu and Suzuki, 1995).

The major issues in training and recognition are: 1) training and generalization (i.e., whether the trained patterns characterize the speech of only the training set or whether they also generalize to speech that will be present in actual use), 2) discriminative training (i.e., what are the most appropriate discriminant functions of speech patterns), 3) adaptive learning (i.e., can the learning of discriminant functions be adaptive), and 4) artificial neural networks (i.e., what is the potential of neural networks in providing improved training and recognition for speech patterns) (Atal, 1995). Other speech recognition research challenges include: 1) better handling of the varied channel and microphone conditions, 2) better noise immunity, 3) better decision criteria, better out-of-vocabulary rejection, better understanding and incorporation of task syntax and semantics and human interface design into speech recognition system, more human-sounding speech, and easy generation of new voices, dialects, and languages (Wilpon, 1995).

There are many dimensions of difficulty for speech recognition applications (Roe, 1995): (1) speaker independence, (2) expertise of the speaker, (3)

vocabulary confusability, (4) grammar perplexity, (5) speaking mode and (6) user tolerance of errors. Speaker independence becomes a problem because it is difficult to recognize all voice types and all dialects. Regarding the expertise of the speaker, Roe (1995) stated that people typically learn how to get good recognition results with practice. A larger vocabulary is more likely to contain confusable words or phrases that can lead to recognition errors and some applications may only permit certain words to be used given that the appropriate preceding word is used in the sentence. The speaking mode encompasses issues regarding rate and coarticulation. User tolerance of errors is a major issue since most systems remain error-prone.

Advances in Speech Recognition and Future Predictions

Speech recognition has provided numerous advances in the area. These advances include word spotting; barge in; rejection; subword units; adaptation; noise immunity and channel equalization; proper name pronunciation; and address, date, and number processing (Wilpon, 1995). For more specifics on the above see Wilpon article.

Speech technologies still remain error-prone despite advances in reliability. For this reason, Wilpon (1995) believes that successful products and services will be those with the following characteristics:

Simplicity- Successful speech recognition systems will be natural to use.

Evolutionary Growth - Applications will be extensions of existing systems.

Tolerance of Errors - Since it is likely that a speech recognizer will make some errors, inconvenience to the user should be minimized.

As do many researchers, Levinson and Fallside (1995) recognize the difficulty of technological forecasting and do not link their predictions of automatic speech recognition to any specific date. However, speech synthesis and recognition systems are expected to play important roles in advanced user-friendly human-machine interfaces by the year 2001 (Furui, 1995). Speech recognition systems services will include databases access and management, numerous order-

made services, dictation and editing, electronic secretarial assistance, robots, automatic interpreting telephony, security control, and aids for the handicapped (Furui, 1995). Furui (1995) also stated that future speech recognition technology should have the following features:

- Few restrictions on tasks, vocabulary, speakers, speaking styles, environmental noise, microphone, and telephones,
- Robustness against speech variations,
- Adaptation and normalization to variations due to environmental conditions and speakers,
- Automatic knowledge acquisition for phonemes, syllables, words, syntax, semantics, and concepts,
- The ability to process discourse in conversational speech (e.g., to analyze context and accept ungrammatical sentences),
- Naturalness and ease of human-machine interaction, and
- Recognition of emotion.

Table 3 depicts broad projections for speech recognition that are and will become available in commercial systems in the next decade. An ultimate system should be capable of robust speaker-independent or speaker-adaptive, continuous speech recognition and no restrictions on vocabulary, syntax, semantics, or task would exist (Furui, 1995).

In the near future, speech recognition will become a component of computer-based aids for foreign language reading. However, use for such an application will require a degree of robustness that may not be considered in other speech recognition applications (Cohen and Oviatt, 1995). From the viewpoint of applications, other features become important (Furui, 1995): (1) Incentive for customers to use the systems, (2) Low cost, (3) Creation of new revenues for suppliers, (4) Cooperation on standards and regulation and (5) Quick prototyping and development.

Table 3. History of and Projections for speech recognition

Year	Recognition Capability	Vocabulary Size	Applications
1990	Isolated/connected words, Whole-word models- word spotting, finite-state grammars, constrained tasks	10-30	Voice dialing, credit card entry, catalog ordering, inventory inquiry, transaction inquiry
1995	Continuous speech Subword recognition-elements, stochastic language models	100-1000	Transaction processing, robot control, resource management
1998	Continuous speech Subword recognition-elements, language models representative of natural language, task-specific semantics	5000-20,000	Dictation machines, computer-based secretarial assistants, database access
2000+	Continuous speech Spontaneous speech-grammar, syntax, semantics; adaptation, learning	Unrestricted	Spontaneous speech- interaction, translating telephony

(Adapted from Rabiner and Juang, 1995)

Speech recognition systems have been used to a limited extent in performing in-vehicle tasks in automobiles. However in the future, ASR may be used to a greater extent in performing in-vehicle tasks, such as adjusting the volume of the radio. Gellatly (1997) stated that speech recognition systems being considered for use in automobiles should have certain parameters: (1) The system should be speaker adaptive or at least speaker independent, (2) The system should allow for continuous speech, and (3) The command vocabulary should be large enough to allow users to say common words related to the task being performed.

In the future, consumer products, voice input/output-capable hardware for PCs, telephone applications, and large-vocabulary text generation systems will

dominate developments in speech interface technology (Oberteuffer, 1995). By the end of century, it is very likely that speech recognition and text-to-speech systems will be applied to hand-held computers, especially with the speech interface being ideally suited to such devices due to its small space requirements and low cost.

Speech recognition and synthesis technologies are affected more than other recent technologies by specific application factors and user interfaces issues. Successful commercialization of theses technologies will not happen unless system integrators and human factors professionals are involved at an early stage (Seelbach, 1995).

Research Motivation

Oberteuffer (1995) differentiates the automatic speech recognition market into six major segments as shown in Table 4. The 1990's sparked significant growth in three of the segments due to new applications: speech to text, computer control, and telephone. The computer control segment grew due to the number of small and large companies that introduced speech input/output products for a few hundred dollars (Oberteuffer, 1995).

Table 4. Automatic speech recognition market segments (adapted from Oberteuffer, 1995)

Segments	Applications
Computer Control	Disabled, CAD
Consumer	Appliances, Toys
Data Entry	QA Inspection, Sorting
Speech-to-text	Text Generation
Telephone	Operator Services, IVR
Voice Verification	Physical Entry, Network Access

Establishing methods for measuring the quality of speech recognition system is important. Objective evaluations are essential to technological development in the speech processing field. Such evaluation methods can be classified into two categories: 1) Task evaluation (creating a measure capable of evaluating the complexity and difficulty of tasks) and 2) technique evaluation (formulating both subjective and objective methods for evaluating task) (Furui, 1995). Therefore, research that can aid in establishing such methods would be beneficial.

Summary

With the significant number of computer users and the inexpensive availability of software that support automatic speech recognition, continuous research in the area regarding its usability and effectiveness is needed. Some consideration has been given to various commercial applications regarding automatic speech recognition. Since in the past most automatic speech recognition systems used isolated word speech, some research in the area exists. However, due to technological advances, advanced research in almost any area related to automatic speech recognition systems is warranted.

As in many areas, research issues regarding large-scale systems and industries as it relates to ASR receives the most attention. However, other areas

that require significant attention are often over-looked. For this reason, this research intends to look at automatic speech recognition at the consumer level. Many individual consumers who will purchase and use automatic software recognition from a different aspect than that of commercial industries, such as telecommunications. Consumers that purchase the software for personal use will mainly use the ASR for dictation of correspondences and documents. This research intends to examine ASR software packages used in conjunction with personal computers for the purpose of dictation and to assess effectiveness and user satisfaction of such systems.

Chapter 3 Methodology

Subjects

Subjects for this experiment were undergraduate and graduate students in the Industrial and Systems Engineering department at Virginia Tech who responded to a general request for participants in an Automatic Speech Recognition experiment. Five male and eight female subjects were used for the actual experiment and one subject was used for pre-testing.

A questionnaire was used to ensure that the subjects' first language is English and he/or she had not used automatic speech recognition for the purpose of dictation in the past (See Appendix A). There were no age or gender restrictions.

Experimental Design

A 2 x 3 x 3 within-subjects design with two dependent measures was used. The subjects received each treatment condition. The within-subjects variable, System Type variable had two levels: 1) IBM Via Voice and 2) Dragon Systems NaturallySpeaking. The second factor, Correspondence Type, had three levels: (1) Personal Correspondence, (2) Business Correspondence, and (3) Technical Correspondence. The within-subjects variable, Error-Correction Time, had three levels: 1) no error-correction time (initial results), 2) five minutes of error-correction time, and 3) ten minutes of error-correction time. During level one, no error-correction time, subjects did not receive error-correction time and word recognition accuracy was based solely on initial system training. During level two, five minutes of error-correction time, subjects received five minutes to correct errors made by the system and word accuracy was then assessed. During level three, ten minutes of error-correction time, subjects received ten minutes to correct errors made by the system during dictation and word accuracy was then assessed. Table 5 represents the experimental design with subject assignment to

treatment conditions. The dependent measures assessed word/command accuracy and user satisfaction.

Table 5. Experimental Design with subject assignments

		Error Correction		
System Type	Correspondence Type	No Error-Correction	5 minutes	10 minutes
IBM Via Voice Gold	1) Personal 2) Business 3) Technical	S1-13	S1-13	S1-13
Dragon Systems Naturally Speaking Preferred	1) Personal 2) Business 3) Technical	S1-13	S1-13	S1-S13

Facilities

The experiment was performed in the Macroerogonomics and Group Decision Systems Laboratory in the Human Factors Engineering Center at Virginia Tech.

Software and Equipment

Two commercially available speech recognition software packages were used in the experiment: 1) IBM- Via Voice Gold and 2) Dragon NaturallySpeaking Preferred (See Table 6). Both of the systems provided features such as continuous speech, voice commands, and multiple users on a single PC and can be purchased for under $200.

IBM Via Voice Gold allows users to dictate text and control the computer by voice. Via Voice Gold is a high -performance speech recognition product that can be used with Microsoft Windows 95 or Windows NT Version 4.0. With suggested initial system training, Via Voice Gold can understand words commonly used in business documents and correspondence. It has a base vocabulary of 20,000 words and allows for users to add up to a total of 64,000 words and commands.

Dragon NaturallySpeaking Preferred is a basic word processor that users can speak to and control by voice commands. Dragon NaturallySpeaking Preferred can be used to compose e-mail messages, create reports, draft letters, and edit proposals just by speaking. While a user dictates at a normal pace, what he or she says will appears as text in the document window.

Table 6. A Comparison of System Requirements

System

Requirements	IBM Via Voice Gold	Dragon NaturallySpeaking Preferred
Processor Speed	Pentium 150 MHz or faster	Pentium 133 MHz or faster
Operating System	Windows 95 or Windows NT 4.0	Windows 95 or Windows NT 4.0
Hard Disk Space	125 MB available hard disk space	65 MB free hard disk space
RAM	32 MB Ram for Window 95, 48 MB Ram for Windows NT 4.0	32 MB Ram for Windows 95, 48 MB Ram for Window NT
Sound Card	16-bit sound card or built-in audio system	Creative Labs Sound Blaster 16 or 100% compatible or Mwave sound card

Procedure

Once Institutional Review Board (IRB) approval was received, data collection was performed in two phases: (1) Pre-testing and (2) Data Collection. IRB Review and Approval, is a requirement of the university for research involving human subjects. A copy of the IRB proposal package has been attached to the document (See Appendix B). Phase 1, pre-testing was done to pilot test the research method and provide the experimenter with an opportunity to carry out the experimental protocol. Phase 2, was data collection, each data collection session was organized in the following manner:

1. Subjects completed the informed consent form found in Appendix B that provided a written explanation of the experiment and its purpose.
2. Then subjects completed a short screening questionnaire to ensure they met minimum criteria requirements found in Appendix A.
3. Next the subjects trained the system based on specified system requirements by reading aloud a number of paragraphs and went through a

Power Point Presentation in error-correction (see Appendix E). Subjects then proceed to complete the three levels of the independent variable, correspondence type by dictating 3 paragraphs to assess the system's word accuracy rate and were given a 5-mintue interval for error-correction and a 10-minute interval for error-correction (see Appendix C).

4. Finally, the subjects were administered a user-satisfaction survey found in Appendix D.

Data Analysis

This section describes the data analyses methods that were used in response to the research questions posed by the research. A three-way Analysis of Variance (ANOVA) using system type, correspondence type, and error correction time as the factors was used to analyze the data. In addition to the method stated above, a Wilcoxon two-tail test and ANOVA were performed to determine if there was any statistically significant difference in user acceptability between the two systems.

Chapter 4 Results

The two dependent variables (word accuracy and user satisfaction) were analyzed using separate analysis of variance (ANOVA) procedures. Additionally, a Wilcoxon two-tail test was performed to determined if there was any statistically significant difference in word accuracy objective and subjective results assessing user satisfaction between the two systems (IBM Via Voice Gold and Dragon Systems Naturally Speaking) used. The Statistical Analysis System (SAS) Version 6.11 and MINITAB Version 10.2 for Windows computer software were used to perform the statistical analyses.

Sample Demographics

A pre-experimental questionnaire was used to collect some general information about the subjects and ensure the subjects met the minimum requirements to participate in the experiment (See Appendix A). Thirteen students (5 males and 8 females) from the Industrial and Systems Engineering Department at Virginia Tech were used in the study (one subject was used during pretesting). Five juniors, seven seniors, and one graduate student participated in the study. None of the participants had used Automatic Speech Recognition for the purpose of dictation.

Word Accuracy

ANOVA results for word accuracy are shown in Table 7. The alpha level was set at 0.05 for all tests of significance. Word accuracy recognition or word accuracy percentage rates for each condition were found using the formula:

$$\text{Word Accuracy} = \frac{\text{\# of words correctly recognized} * 100}{(100 - \text{\# of words/commands skipped} - \text{\# of words mispronounced})}.$$

The main effects of Error Correction and Correspondence Type were significant at p= 0.0001 and p= 0.0004 respectively, as were the interactions of

System x Correspondence Type and Error Correction Time x Correspondence Type at p=0.0368 and p=0.0463. A Newman-Keuls post hoc analysis was performed to determine which Error Correction levels were significantly different and the results are shown in Table 8.

Table 7. Analysis of Variance Word Accuracy

Source	df	SS	MS	F	p
Between					
Subject	12	15606.760	1300.563		
Within					
System	1	5246.427	5246.427	4.22	0.0624
System* Subject	12	14924.128	1243.677		
Error Correction	2	8576.102	4288.051	30.87	0.0001
Error Correction* Subject	24	3333.341	138.889		
System * Error Correction	2	702.803	351.401	2.66	0.0908
System* Error Correction* Subject	24	3175.974	132.332		
Correspondence	2	2156.384	1078.192	10.92	0.0004
Correspondence* Subject	24	2370.059	98.752		
System *Correspondence	2	547.188	273.594	3.8	0.0368
System* Correspondence* Subject	24	1727.256	71.9690		
Error Correction * Correspondence	4	588.512	147.1282	2.62	0.0463
Error Correction *Correspondence *Subject	48	2694.376	56.132		
System * Error Correction * Correspondence	4	181.042	45.260	1.16	0.3403
System *Error Correction * Correspondence* Subject	48	1873.179	39.024		
Total	**233**	**63703.538**			

Table 8. Newman-Keuls Results for the Main Effect of Error Correction Time on Word Accuracy.

Error Correction Time	Mean	N	SNK Grouping
no error-correction	72.897	78	C
5 min.	82.256	78	B
10 min.	87.538	78	A

(Note means w/ different letters are significantly different.)

A Newman-Keuls post hoc analysis was performed to determine which Correspondence levels were significantly different and the results are shown in Table 9. The results indicated that word accuracy achieved by the systems for the Personal Correspondence were significantly better than that of Business and Technical Correspondences. The differences in word accuracy results for the business and technical correspondence were also significant.

Table 9. Newman-Keuls Results for the Main Effect of Correspondence Type on Word Accuracy

Correspondence Type	Mean	N	SNK Grouping
Personal	85.03	78	A
Business	77.85	78	B
Technical	79.81	78	B

(Note means w/different letters are significantly different.)

Interactions

Two two-way interactions were significant: System Type x Correspondence Type ($p=0.0368$) and Error-Correction Time x Correspondence Type ($p=0.0463$).

System Type x Correspondence Type

An interaction occurs when the relationship between one independent variable and the subjects' behavior depends on the level of a second dependent variable. According to a Newman-Keuls post hoc analysis of the unconfounded comparisons of the interaction between System Type and Correspondence Type, there was a statistically significant difference between word accuracy results of the Via Voice System for Business Correspondence and the word accuracy of Via Voice for the Personal Correspondence (see Table 10). The word accuracy results of the Via Voice system's Business, Technical, and Personal Correspondences were significantly lower than the word accuracy of the Dragon NaturallySpeaking System for the each of the correspondence types (Business, Technical, and Personal Correspondences respectively) evaluated. No statistically significant difference existed between Dragon system's Personal, Business, and Technical Correspondence. Figure 4 shows the two-way interaction between system type and correspondence type.

Table 10. Newman-Keuls analysis of the effect of System Type and Correspondence Type on word accuracy

Increasing Rank Order									
		1	2	3	4	5	6		
Treatment Means		Sys_VC_B 72.92	Sys_VC_T 76.17	Sys_VC_P 81.35	Sys_DC_T 83.53	Sys_DC_B 84.69	Sys_DC_P 88.77	r	$CD_{0.05}$
	1	-------	3.25	8.43*	10.61*	11.77*	15.85*	6	7.25
	2		------	5.18	7.36*	8.52*	12.6*	5	6.92
	3			------	2.18	3.34	7.42*	4	6.47
	4				------	1.16	5.24	3	5.85
	5					-------	4.08	2	4.84

*Statistically significant at $\alpha = 0.05$

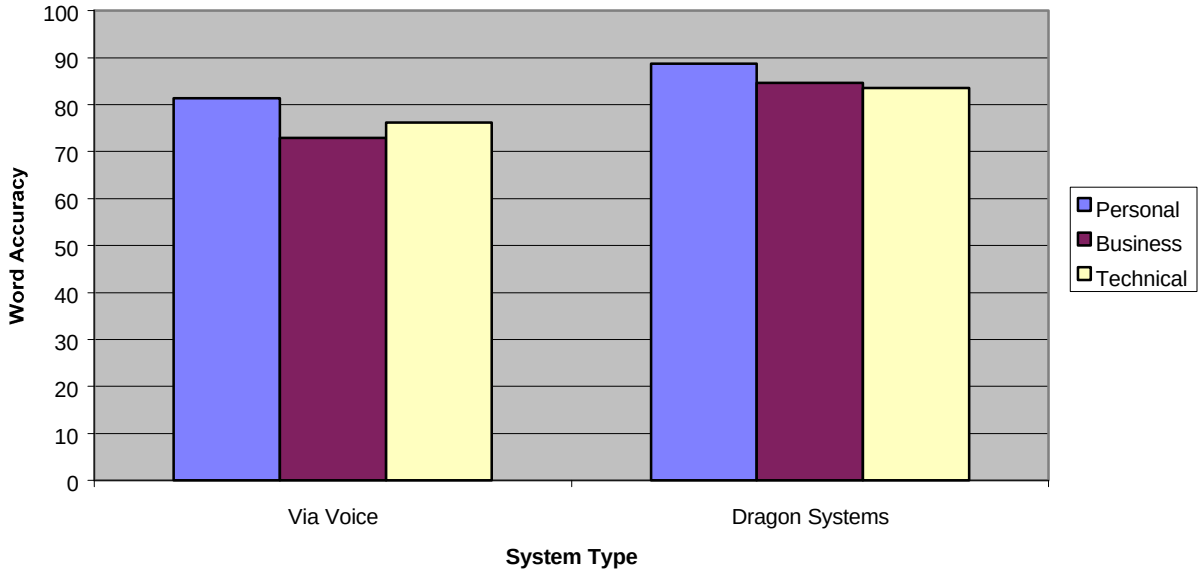


Figure 4. Mean plot of the effects of System Type and Correspondence Type interaction on Word Accuracy

Error-Correction Time x Correspondence Type

According to a Newman-Keuls post hoc analysis of the unconfounded comparisons of the interaction between Error-Correction Time and Correspondence Type, there was a statistically significant difference between word accuracy results obtained without error-correction time for the Business Correspondence and no error-correction time for the Personal Correspondence (see Table 11). No significant difference existed between the Business and Technical Correspondence for the no error-correction condition. A significant difference also existed between the Business and Personal Correspondence for the 5 minutes of error-correction condition, but no significant difference existed between the Business and Technical Correspondences for this condition. The Newman-Keuls analysis showed no statistically significant difference between word accuracy results obtained for the 10 minutes of error-correction condition for the three correspondence types.

There was a significant difference between no error-correction and five minutes of error-correction for the Business and Technical Correspondences. A significant difference did not exist between the five minutes of error-correction and ten minutes of error-correction for the Business and Technical Correspondences. For the Personal Correspondence the opposite was true; there was a significant difference between five minutes of error-correction time and ten minutes of error-correction time, but no difference between no error-correction and five minutes of error correction. However, for all three correspondence types, there was a significant difference between no error-correction and ten minutes of error-correction. When the subjects were allotted ten minutes of error-correction time to dictate the three correspondence types, the word accuracy results were higher than when no error-correction time and five minutes of error-correction time were given. Figure 5 shows the two-way interaction between Error-Correction Time and Correspondence Type.

Table 11. Newman-Keuls analysis of the effect Error-Correction and Correspondence Type on word accuracy

		1	2	3	4	5	6	7	8	9		
Treatment Means		EC_0C_B 68.35	EC_0C_T 70.15	EC_5C_B 80.00	EC_0C_P 80.19	EC_5C_T 81.73	EC_5C_P 85.04	$EC_{10}C_B$ 85.19	$EC_{10}C_T$ 87.54	$EC_{10}C_P$ 89.88	r	$CD_{0.05}$
	1	------	1.18	11.65*	11.84*	13.38*	16.69*	16.84*	19.19*	21.53*	9	6.80
	2		--------	9.85*	10.04*	11.58*	14.89*	15.04*	17.39*	19.73*	8	6.639
	3			---------	0.19	1.73	5.04	5.19	7.54*	9.88*	7	6.448
	4				--------	1.54	4.85	5.00	7.35*	9.69*	6	6.213
	5					-------	3.31	3.46	5.81	8.15*	5	5.934
	6						---------	0.15	2.50	4.84	4	5.567
	7							---------	2.35	4.69	3	5.050
	8								--------	2.34	2	4.200

*Statistically significant at $\alpha = 0.05$

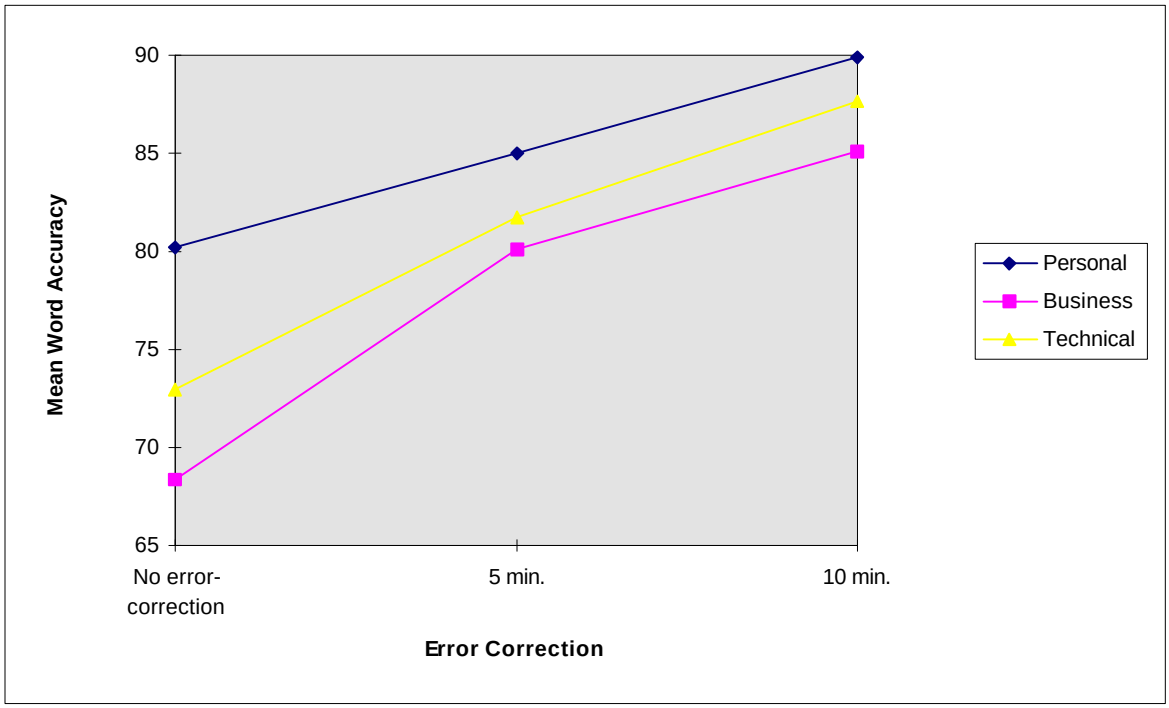


Figure 5. Mean plot of the effects Error-Correction Time and Correspondence Type interaction on word accuracy.

Subjective Measures

User Satisfaction

ANOVA results for user satisfaction are shown in Table 12. User satisfaction/acceptability results were obtained from the subjects after they completed the experiment. The subjects rated user satisfaction from zero to 100 (See Appendix D). Subjects were instructed to rate the five and ten minutes of error-correction levels on error-correction procedure ease of use and the results obtained after each correction condition. They rated the correspondence types on how they felt the system did in recognizing (in terms of word accuracy) the various types of correspondences. The overall/final opinion was a rating based on how the subjects felt with the systems performance. The main effect Opinion was found to be significant. No interactions were found to be significant.

A Newman-Keuls post hoc analysis was performed to determine which Opinion levels were significantly different and the results are shown in Table 13.

Table 12. Analysis of Variance User Satisfaction

Source	df	SS	MS	F	p
Between					
Subject	12	4681.192	390.099		
Within					
System	1	1212.980	1212.980	1.9	0.1930
System* Subject	12	7650.935	637.577		
Opinion	5	5752.903	1150.580	12.41	0.0001
Opinion* Subject	60	5562.346	92.7057		
System * Opinion	5	474.7500	94.7057	0.87	0.5044
System* Opinion* Subject	60	6521.833	108.697		
Total	**155**	**31856.939**			

Table 13. Newman-Keuls Results for the Main Effect of Opinion on User Satisfaction

Opinion	Mean	N	SNK Grouping
5 min. of Error-Correction	61.923	26	B
10 min. of Error Correction	80.269	26	A
Personal Correspondence	78.462	26	A
Business Correspondence	76.615	26	A
Technical Correspondence	72.769	26	A
Overall/Final Opinion	77.077	26	A

(Note means w/ different letters are significantly different.)

Subjective data was also gathered to gain an understanding as to how the subjects felt about using the two systems. The survey addressed system-required training, the subjects' feelings toward dictation, remembering commands, error-correction procedures, and overall performance (See User Satisfaction Survey in Appendix D). The subjects received the subjective survey after using each system. A five point Likert-type scale with the following categories: strongly disagree (S/D), disagree (D), undecided, agree (A), and strongly agree (S/A) was used.

The two systems were compared for each statement using a Mann-Whitney Confidence interval and test (also referred to as a 2-sample Wilcoxon rank sum test). The Wilcoxon test was performed to determined if there was any statistically significant difference in subjective results assessing user satisfaction between the two systems used (Via Voice Gold and Dragon Systems Naturally Speaking). The Wilcoxon test for statement five indicated a main effect of system ($p=0.012$). No other significant effects were found. Question five addressed if the subjects' dictated the paragraphs as they would in normal conversation. The figures below provide the frequency results for each question (See Figures 6-13).

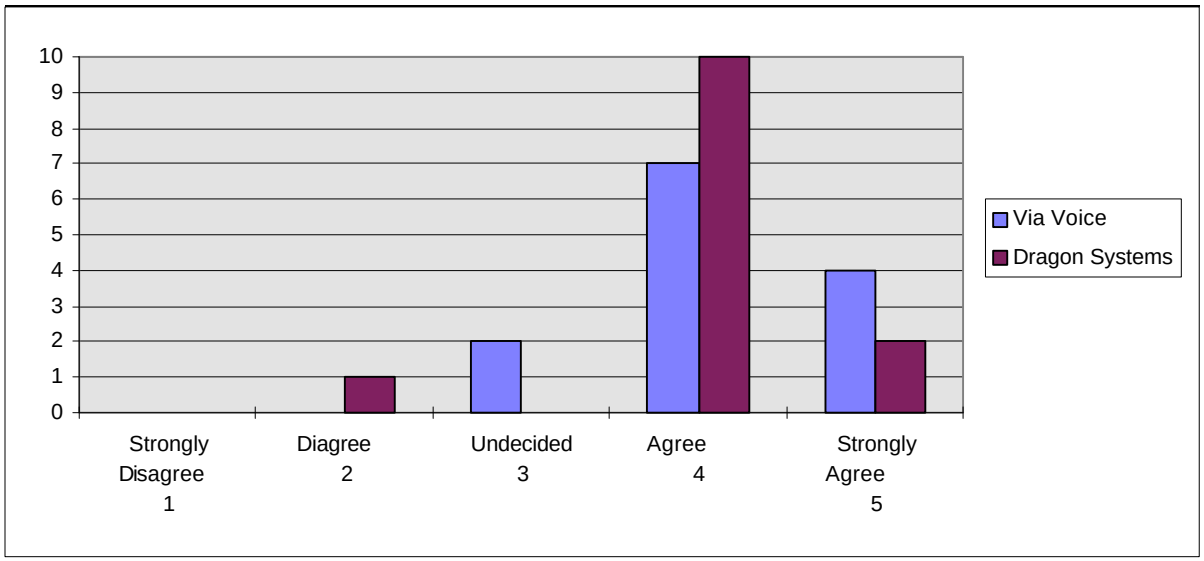


Figure 6: Frequency counts for Statement 1: The system-required training aided in word accuracy.

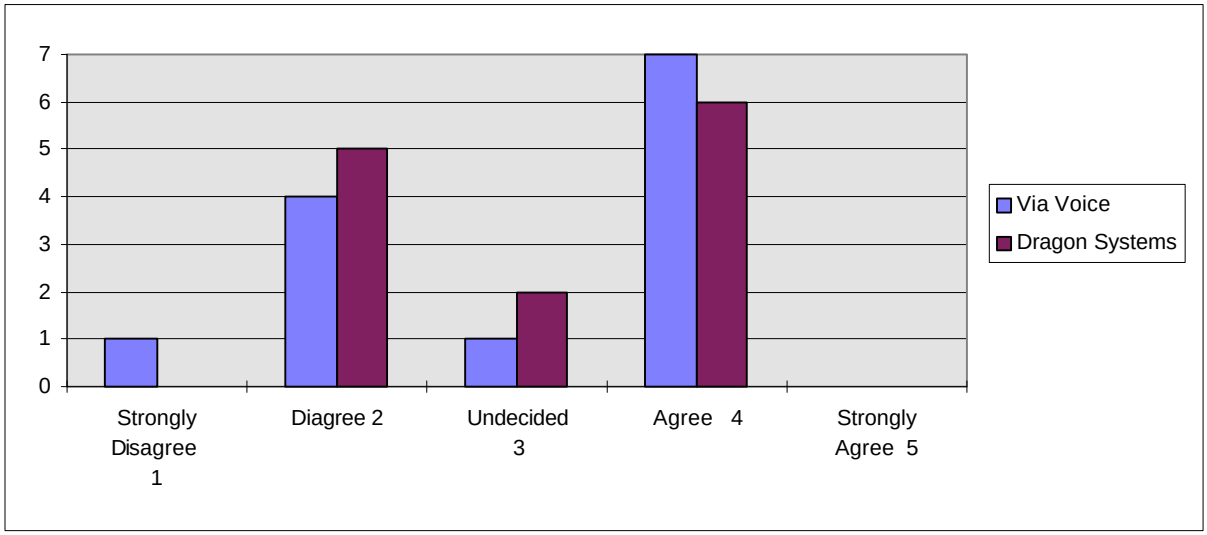


Figure 7. Frequency count for Statement 2: During the system-required, I experienced fatigue.

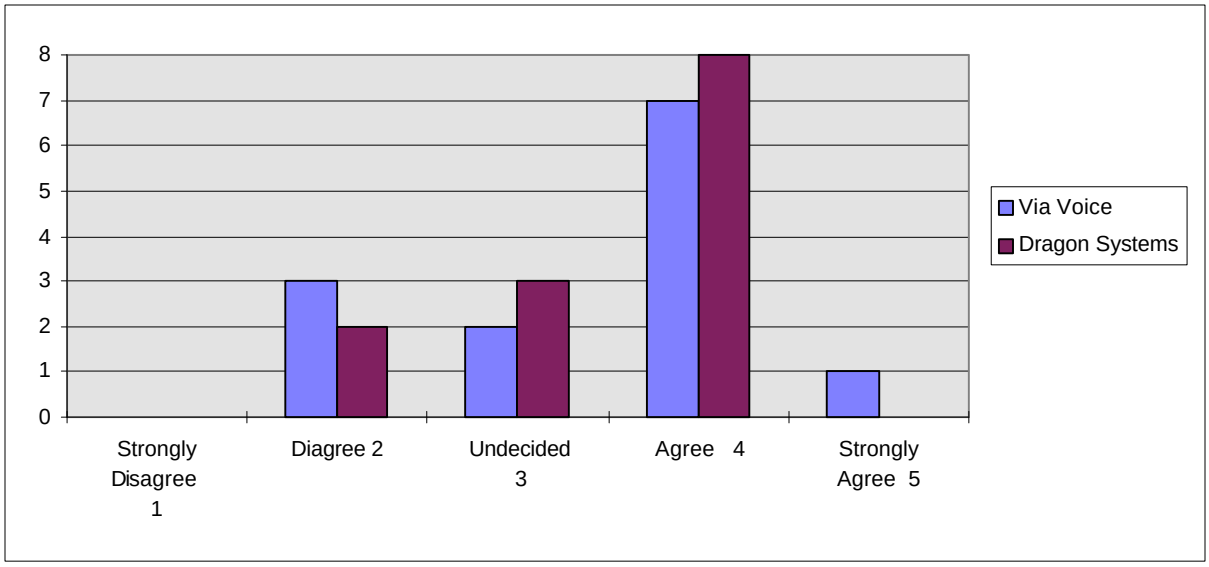


Figure 8. Frequency count for Statement 3: I felt the system-required training was adequate.

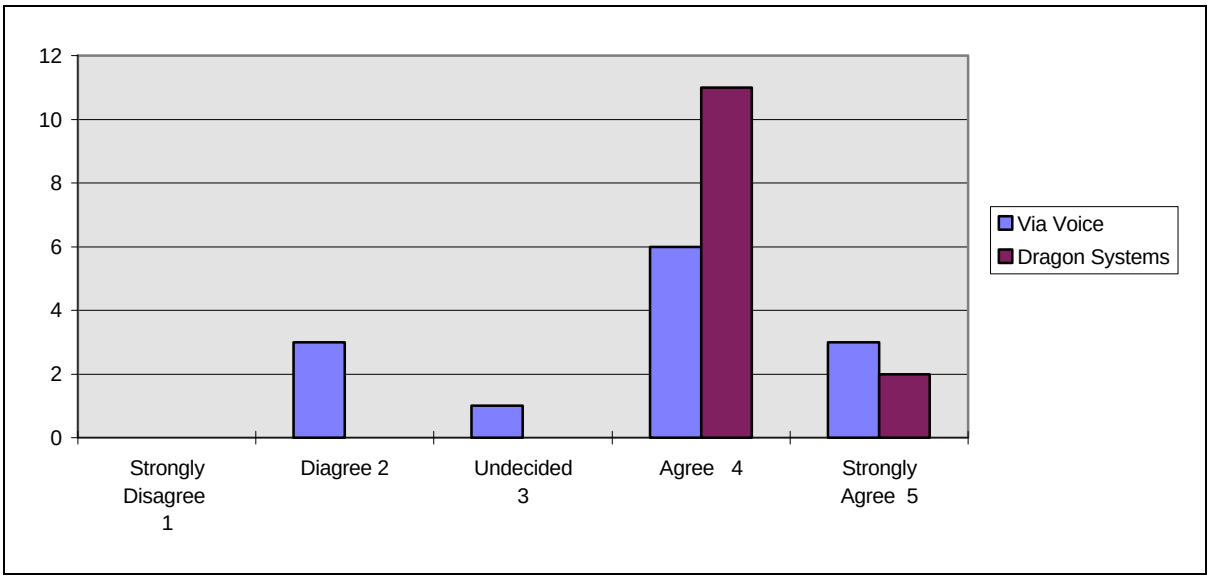


Figure 9. Frequency count for Statement 4: I felt comfortable while dictating using the system.

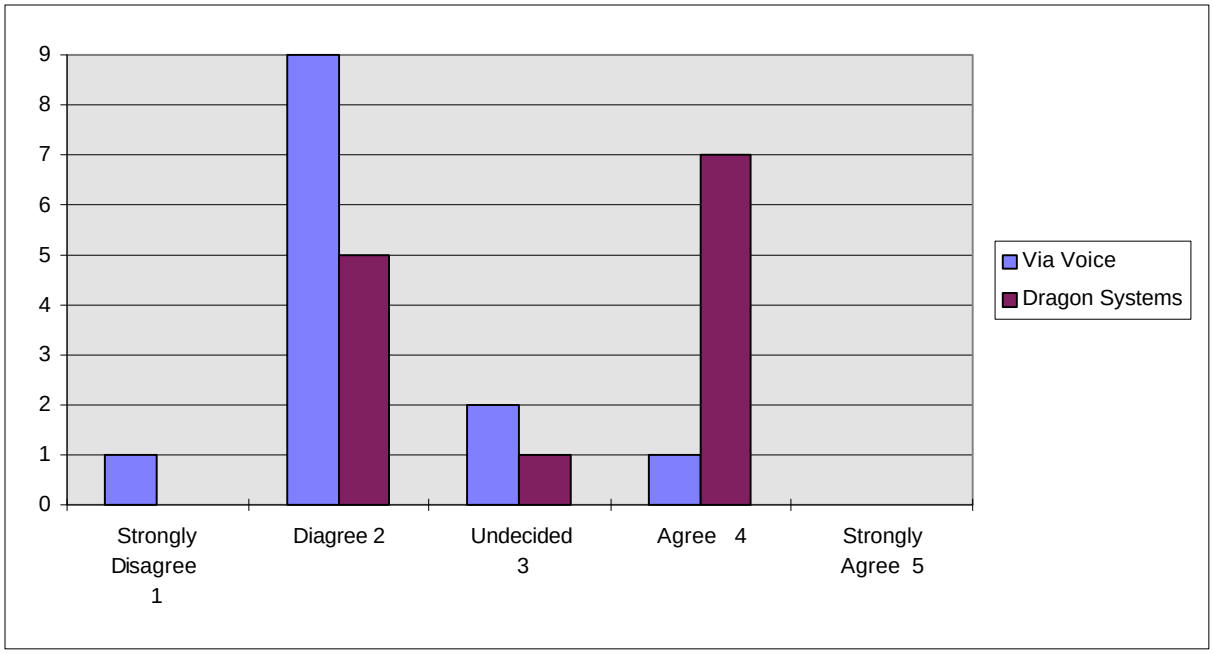


Figure 10. Frequency count for Statement 5: I felt that I dictated the paragraphs as I would in normal conversation. (Significant effect found)

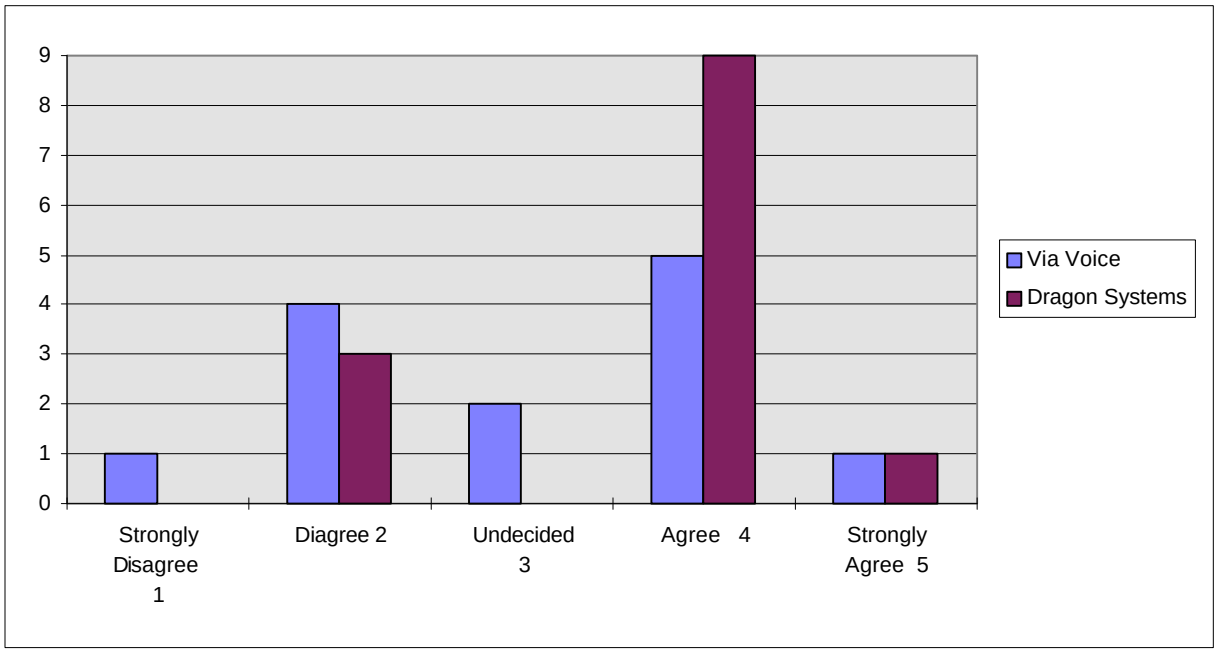


Figure 11. Frequency count for Statement 6: I had no problem remembering commands.

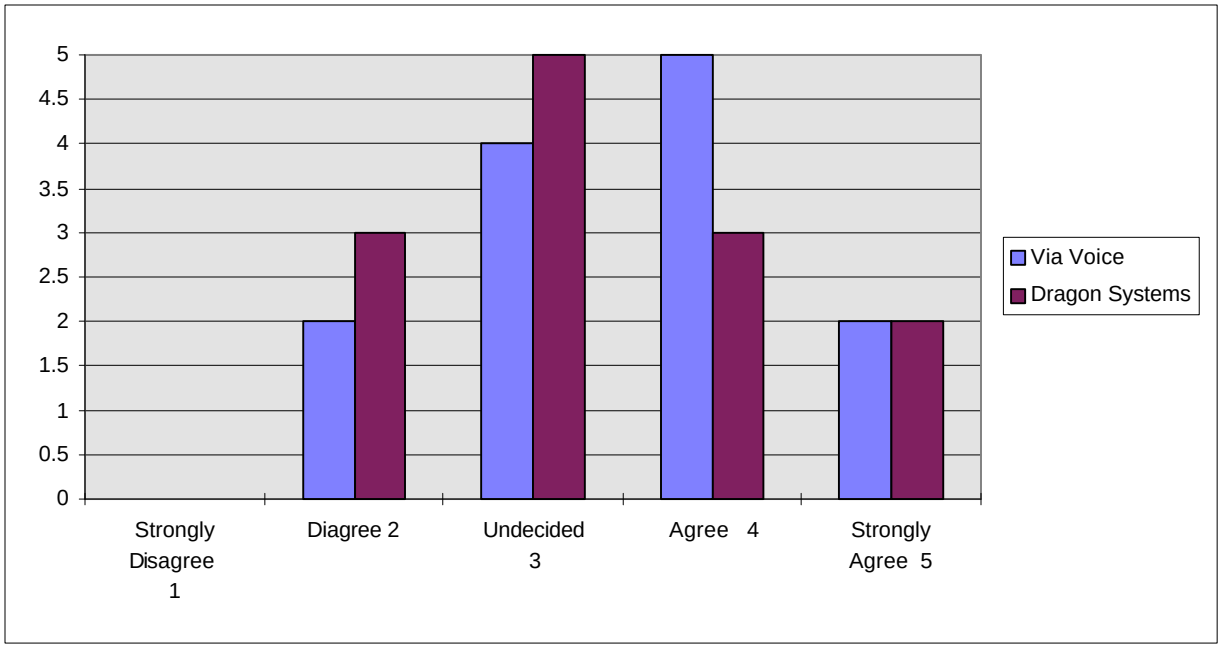


Figure 12. Frequency count for Statement 7: I felt the error-correction procedure was tedious.

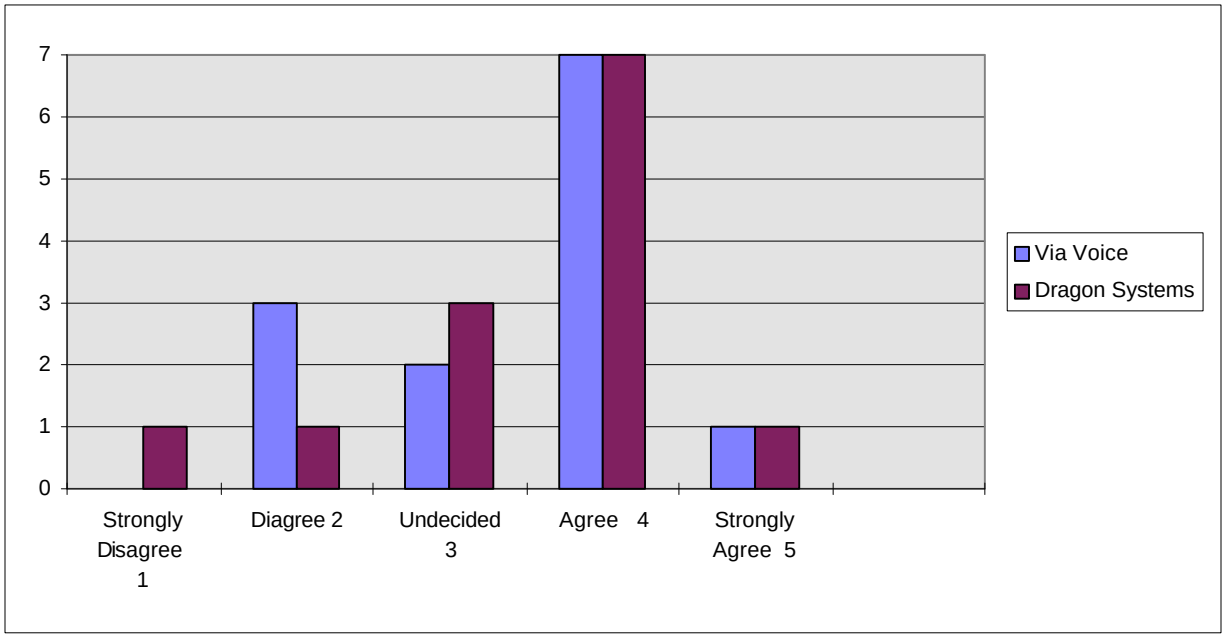


Figure 13. Frequency count for Statement 8: Overall, I was pleased with the speech recognition software's performance

Additional Post-Hoc Analyses

A Wilcoxon test was used determine if a statistically significant difference existed from the objective results and subjective ratings of the two systems (Via Voice Gold and Dragon Systems Naturally Speaking Preferred). Personal, Business, and Technical Correspondence word accuracy results and Personal, Business, and Technical Correspondence subjective ratings results were analyzed. A statistically significant difference was found to exist between the two systems for the Business Correspondence Word Accuracy results ($p=0.0083$) and the Technical Correspondence User Satisfaction rating results ($p=0.0317$). No other significant differences were found.

Via Voice

A Pearson-r correlation coefficient was calculated to test whether Via Voice's word accuracy results were related to the user satisfaction ratings obtained for each type of correspondence (i.e., personal, business, and technical). A correlation coefficient was also obtained to test whether Via Voice's final word accuracy results (after the 10-min. error-correction level) and the subjects' overall user satisfaction ratings were related. A t-test of significance was used to determine if the correlation coefficients were significantly different from zero. The results showed that the correlation between the Personal Correspondence word accuracy and user satisfaction ratings ($r=0.701$) and the Technical Correspondence word accuracy and user satisfaction ratings ($r=0.698$) were significant.

Table 14. Pearson-r correlation coefficients for Via Voice

Relationship	Pearson-r correlation coefficient
Personal Correspondence Word Accuracy and Satisfaction Ratings	r = 0.701
Business Correspondence Word Accuracy and Satisfaction Ratings	r = 0.407
Technical Correspondence Word Accuracy and Satisfaction Ratings	r = 0.698
Overall Word Accuracy and Satisfaction Ratings	r = 0.256

Dragon Systems

A Pearson-r correlation coefficient was also calculated to test whether Dragon Systems NaturallySpeaking's user satisfaction ratings were related to the word accuracy results obtained for each type of correspondence (i.e., personal, business, and technical). A correlation coefficient was also obtained to test whether Dragon System's final word accuracy results (i.e., results obtained after the 10 min. error-correction level) and the subjects' overall user satisfaction ratings were related. A t-test of significance was used to determine if the correlation coefficients were significantly different than zero. The results showed that the Personal Correspondence word accuracy results and user satisfaction ratings (r=0.570) was found to be significant.

Table 15. Pearson-r correlation coefficients for Dragon Systems Naturally Speaking

Relationship	Pearson-r correlation coefficient
Personal Correspondence Word Accuracy and Satisfaction Ratings	r = 0.570
Business Correspondence Word Accuracy and Satisfaction Ratings	r = 0.518
Technical Correspondence Word Accuracy and Satisfaction Ratings	r = 0.431
Overall Word Accuracy and Satisfaction Ratings	r = 0.449

Chapter 5 Discussion and Conclusions

The results obtained from this experiment partially support the assertion that commercially available automatic speech recognition software systems can provide users with acceptable word accuracy results and user satisfaction. Novice users participated in the experiment. However, the above claim extends to individuals who will use the systems frequently and the systems overall performance over a period of time. Therefore, observing frequent users over a period of time should be considered in order to obtain a better understanding of automatic speech recognition systems capabilities with frequent use. This claim will be evident as the three hypotheses that motivated this research are evaluated.

Hypothesis One

Hypothesis one stated that Business Correspondence would achieve the greatest word accuracy results. Word accuracy results for the Personal Correspondence achieved the greatest word accuracy results for both systems. The System Type x Correspondence Type interaction was significant with respect to Word Accuracy for the Via Voice System, but not for the Dragon System. These results failed to support this hypothesis. This finding was not expected because the producers of Via Voice Gold claim the software works best when text that resembles general business correspondence is dictated and therefore the assumption was made that other commercially available systems would also dictate business types of correspondences better. However, due to the fact that personal correspondence can resemble (i.e., contain similar words and/or use names of people and places-proper nouns) business correspondence, the results obtained are conceivable.

However, there are issues associated with the correspondence types resembled that could have affected the results. The correspondence types evaluated in the study were devised with the expectation of depicting three different types of correspondences. One could choose to argue that they were

really not that different from each other. However, the issue of how a correspondence type is classified could vary from person to person. The Business Correspondence resembled in the study was from an apology written from a company to a customer concerning a product that did not meet the customer's satisfaction. The Personal Correspondence was a segment of a reminder note to a group of individuals planning an event. While, the Technical Correspondence was an excerpt taken from a technical paper written by a graduate student.

Dragon Systems NaturallySpeaking Preferred obtained mean word accuracy results of greater than 80% for the Personal, Business, and Technical Correspondence (87.7%, 84.69%, 83.55% respectively). Via Voice only obtained a mean word accuracy result of greater than 80% for the Personal Correspondence. The greatest results were achieved during dictation of Personal Correspondence which corresponds with the Correspondence main effect finding. This finding could be a result of the Personal Correspondence not including any technical terms or proper nouns that could have given the systems trouble in interpreting the words dictated. The Business Correspondence included proper nouns such as "Merchant Mart" and words that could be easily interpreted as other words (i.e., confusability issue). While the Technical Correspondence included technical terms such as ergonomically and non-ergonomically, that were not easily recognizable for the systems.

The type of correspondence being dictated is important when using Via Voice, correspondences resembling the Personal correspondence evaluated in the experiment would achieve greater results over correspondences that resemble the Business and Technical correspondences evaluated. Dragon System's is more likely to achieve comparable word accuracy results for correspondences resembling the three types of correspondences evaluated.

Hypothesis Two

Hypothesis two stated that increased error-correction time by the user would provide increased word accuracy results. The results support this

hypothesis. The subjects were allowed to correct errors beginning with the correspondence type of their choice or by switching between correspondences. During the five minutes of error-correction, it was observed that a majority of the subjects began correcting the correspondence types with the most errors and that included terms they felt the systems were having difficulty understanding (Business and Technical correspondences) and went to the Personal Correspondence last which had the least errors if time permitted. However, during the ten minutes of error-correction, a majority of the subjects began with the Personal Correspondence which had the least number of errors, and then went on to the other two Correspondences because of the significant time and effort that was placed on the Business and Technical Correspondence during the five minutes of error-correction condition.

The 10 minutes Error-Correction time condition out performed the five minutes Error-Correction time and No Error-Correction time condition. The greatest results were achieved after the 10 minutes error-correction condition for Personal correspondence which supports the Correspondence and Error-Correction main effect findings. The 10 minutes of error-correction condition was also rated highest by the subjects in terms of user satisfaction and acceptability (80.27 out of 100). However, from this research, it is not known if 10 minutes of error-correction time is the optimal time. Supplementary research could be conducted to determine user satisfaction and word accuracy results obtained after longer periods of error-correction time and at what point user satisfaction would begin to decrease. Subjects could have rated the 10 minutes of error-correction condition the highest because of the increase in word accuracy results for the three types of correspondences and them becoming more comfortable with the error-correction procedures after the initial five minutes of error-correction condition. This finding is consistent with those reported by Schurick et al. (1985) in which they concluded that user error-correction provided an increase in recognition accuracy. Despite the fact that word accuracy results may increase with increased error-correction time, users of the ASR systems could become discouraged from

its use if a significant amount of time is required in correcting errors. It is possible that error-correction time may diminish with use.

Occasional errors by speech recognizers are expected and because this is so, inconvenience to the user should be minimized (Wilpon, 1995). Despite relatively high error rates that may result from using such systems, many systems are acceptable if the error tendency is natural and matches the principles of human hearing and perception (such as the system making a mistake with similar-sounding and closely-related words) (Furui, 1995). Therefore, it becomes very important that recognition errors are easy to correct and the system does not repeat the same errors.

Hypothesis Three

Hypothesis three stated that user satisfaction would be influenced negatively due to lower word accuracy recognition for the shorter period of error correction versus the increased error-correction time of 10 minutes. Subjects were instructed to rate the five and ten minutes of error-correction levels on error-correction procedure ease of use and the results obtained after each correction condition. The subjects' mean user satisfaction rating after five minutes of error-correction was 61.9 out of a possible rating of 100. While the subjects' mean user satisfaction rating after 10 minutes was 80.2 out of a possible rating of 100. These results support this hypothesis. Via Voice received a mean rating of 59 out of 100 for five minutes of error-correction results and 80.15 out of 100 for ten minutes of error-correction time. While Dragon Systems NaturallySpeaking improved going from 63.46 to 80.38. Subjects could have rated this condition the highest because of the increase in word accuracy results for the three types of correspondences and them becoming more comfortable with the error-correction procedures after the initial five minutes of error-correction condition.

The higher user satisfaction ratings for the increased error-correction condition (10 minutes of error-correction) could be due to the following

observations: 1) When subjects were provided increased error-correction time, word accuracy results increased for all of the correspondence types studied and 2) The subjects also appeared to be more comfortable during the 10 minutes of error-correction time condition because of the practice and experience gained during the first error-correction condition. Levision and Roe (1990) reported that users of commercially available speech recognition systems do not expect error-free results when beginning to use such systems. However, with continued use, improved results are expected and user satisfaction is affected by the word accuracy recognition obtained.

Subjective and Objective Measures

In assessing the performance of the systems, both subjective and objective measures provided vital information. Subjective and objective measures provided information about a system in order to improve features within an interface or assess a completed system (Preece, 1993). In this experiment, objective measures provided results regarding word accuracy recognition of the two systems evaluated. While the subjective measures provided opinions of the user concerning the accuracy of the systems, acceptability, and ease of use.

Rating scale items were used in the user satisfaction survey. Only one statement of the eight subjective questions asked after the experiment indicated a main effect of system (see Appendix D). Question 5 addressed if the subjects dictated the paragraphs as they would in normal conversation. A majority of the subjects felt that while using the Via Voice System they did not dictate as they would in normal conversation. Despite the fact that subjects were encouraged to dictate as they would in normal conversation, it seemed that many subjects felt if they were to dictate slower, Via Voice would be able to recognize what they were saying better. Roe and Wilpon (1993) noted that people sometimes modify their speech habits to use speech recognition systems (just as they would when leaving messages on an answering machine). Overall, the subjects agreed that they were pleased with the speech recognition software's performance. This could in part be

a result of the subjects being first-time users and fascinated by the capabilities of the systems.

Subjects were also allowed to rate user satisfaction based on their perceptions of how accurate the systems were in recognizing a particular type of correspondence and the observed results after error-correction. No statistically significant difference was found between user satisfaction ratings for the three types of correspondences evaluated in the study and the overall satisfaction of the two systems. One likely reason why subjects did not perceive a significant difference in the correspondence types as was evident in the objective results is that most subjects quickly reviewed the results on the screen and were not given the exact word accuracy results for each correspondence type, but did see that the word accuracy was improving after the error-correction conditions. Additionally, the subjects may not have perceived a difference among the correspondence types. The overall satisfaction rating was based on overall word accuracy recognition and the subjects' feelings toward the error-correction procedures for the two systems. The subjects gave the two systems a mean overall rating of 77 out of 100. Neither system received outstanding mean overall ratings (Via Voice received a 75.62 out of a possible 100 and Dragon Systems received a 80.08 out of a possible 100). These subjective results partially support the objective measures of word accuracy. Dragon Systems outperformed Via Voice in word accuracy recognition and the subjective ratings of the systems support the results in a similar fashion.

Speech Input

Speech input, the process by which human speech is received and processed by a computer, offers both advantages and disadvantages over other input methods such as keyboards, special keys and other facilities (i.e., cursor control keys, screen keys), automatic scanners such as bar code readers and document scanners, and other input devices such as a dataglove, mouse, or joystick (Preece, 1993).

A major advantage of speech input is the training of new users. Because speech is a natural form of communication, training new users is much easier than with other input devices (Preece, 1993). For this experiment, users did not have to be trained to speak a certain way; users were only required to train the systems. System training required the subjects to read numerous sentences, so that the system could adapt to the speech files of the user and users were encouraged to speak clearly and loud enough for the system to understand him or her. Dictation differences were observed; specifically, some subjects dictated clearer than others, some subjects dictated monotonically, and some dictated with enthusiasm and expression. It was observed that during system training, monotone speakers would have to repeat words or sentences more frequently to further aid the system in understanding his or her speech.

However, speech input does suffer some disadvantages. Automatic speech recognizers have limitations, specifically in being able to distinguish between similar-sounding words or phrases (i.e., to and two). This limitation was evident in the experiment. Both systems had difficulty in interpreting similar-sounding words. This limitation was especially obvious in the Business Correspondence word accuracy results.

Unlike other input devices, speech input recognition is subject to background noise interference. The subjects performed the experiment in a room with only one other individual present (the experimenter). Background noise was a factor, the air conditioner and other computers were on during the experiment. However, background noise was minimized to control for substantial interference. Still, depending on where such systems are used background noise interference could significantly impede the systems performance.

Design Recommendations

Realizing that automatic speech recognition is an emerging technology, some guidelines as to how to evaluate speech recognition systems can be made from performing this research. The first is to use numerous kinds of

correspondences and evaluate the word accuracy results achieved by the systems, taking into account the various documents or correspondences an individual may choose to dictate. The second guideline would be to compare the system training time with the word accuracy results achieved for the systems. The third guideline would be to evaluate the systems error-corrections procedures to ensure that users are comfortable with the particular methods of the systems.

This research examined automatic speech recognition with the individual consumer in mind, since many consumers will choose to purchase such software because of its low cost. Realizing that consumers who purchase the software for personal use will mainly use the ASR for dictation of correspondences and documents, the research examined ASR software packages used in conjunction with personal computers for the purpose of dictation and assessed word accuracy recognition and user satisfaction of such systems.

Designing for the user is a key principle of human factors professionals. In designing for the user, the objective becomes to satisfy the user's needs. Once the needs of the user are understood, speech based user interfaces can be designed or better designed. In understanding the needs and expectations of the user, the goal becomes to make the system easy to use and have it obtain acceptable results.

Function allocation is important in regards to automatic speech recognition. Function allocation is the process of determining which will carry out the function and how the function will be carried out (Wilson and Corlett, 1990). Addressing or readdressing the issue of function allocation should be evaluated from an error-correction procedure standpoint in ASR systems. Functions associated with using ASR systems can be implemented by the human involved in the process, the machine or system, or a combination of both. In many cases, function allocation involves human and machine interaction. With the error-correction procedures of both systems, the subjects could highlight the incorrect word or words and an error-correction dialogue box would appear. This dialog box allowed the subjects to choose the correct word from a list of similar-sounding words and if the correct word was not listed the user could type or spell the

correct word. The two systems could incorporate a feature that would scan the sentences or paragraphs and when a sentence or word in a sentence does not make sense to the system (in terms of context) have a dialogue box automatically appear. Then the user could choose from a list the correct word or phrase and if the correct phrase is not given type it in.

The major implication derived from these research findings is that word accuracy recognition will be influenced not only by system training and how clearly a user dictates, but the time and effort that is spent in correcting errors so that the system can adapt the users speech files and the subjective feelings of the user in using the systems (for example, how they feel about system training and error-correction). Systems should be designed with the goal of meeting user expectations. Therefore, systems should provide the user with error-correction procedures that are easy to remember and perform.

Based on comments from and observations of the subjects in the experiment and empirical results, the following design recommendations are made with respect to system and user interfaces. A significant main effect of System was found for System Training Time. Training Via Voice's system to recognize the speech patterns of the user took significantly longer than Dragon System NaturallySpeaking. Via Voice had a mean system training time of 58.9 minutes. The material that the subjects had to read was about speech recognition technology and how it can be used to benefit society. Subjects commented that they became restless during training because of the nature of the material and complained about the amount of time required to train the system. One recommendation is for Via Voice (like Dragon Systems NaturallySpeaking Preferred) to provide more entertaining material for the users to read. Dragon Systems had a mean training time of 31 minutes. The subjects were allowed to choose from two fun excerpts (Dave Barry in Cyberspace and Dogbert's Top Secret Management Handbook) and one science fiction excerpt (3001: The Final Odyssey). The subjects enjoyed reading the material and seemed to speak more clearly because they did not think of it so much as a task; it was entertaining and

time was not a negative factor. Despite the fact that Via Voice's system training was significantly longer than Dragon Systems, Dragon System produced higher word accuracy results.

It is recommended that both systems incorporate an interactive error-correction training session. In the experiment, a Power Point Presentation (see Appendix E) was developed for each system to provide users with error-correction procedures. The subjects were also provided with paper and a pen to develop a "help sheet" (if desired). Only two of the subjects actually used their help sheet while correcting errors. Based on general observations and comments from subjects, an error-correction tutorial that includes actual interaction or practice with the specific system would have been beneficial. The practice session would differ from system training, in that subjects will not only be provided with instructions on error- correction procedures, but an avenue to apply what they learn in order to gain a full understanding of the procedure. This could possibly improve the user opinions of the error-correction procedures being tedious. The tedious feelings toward error-correction procedures could have been a result of the subjects being uncomfortable about their ability to remember and perform the procedure correctly.

Dragon Systems is also recommended to include an added error-correction feature. When correcting errors in Via Voice, the subject hears how they dictated the highlighted or chosen word or phrase. This particular feature provided users with immediate feedback in order to determine if they did not pronounce the word correctly or if they needed to pronounce the word more clearly. The subjects seemed to appreciate this feature because it served as a constant reminder for them to speak clearly.

In summary, both systems are provided with suggested recommendations that can aid in improving user satisfaction and user acceptability. Via Voice Gold is recommended to provide users with more entertaining material to dictate during system training. Dragon System NaturallySpeaking Prefferred is recommended to incorporate an error-correction feature that will allow the user to hear how he or

she dictated a word or phrase that is incorrectly interpreted by the system. While both systems are encouraged to provide an interactive error-correction training session for users.

In conclusion, both ASR systems provided users with some degree of word accuracy using a medium (speech) that is comfortable to most. User satisfaction was not overwhelmingly high for either system. Based on this research, one system would not be recommended over the other. However, it can be concluded that Dragon system training was preferred by subjects over Via Voice. The subjects had to make significant corrections using both systems. Both Dragon System NaturallySpeaking Preferred and Via Voice Gold have areas in which improvement is warranted.

Future Research

While the implications of the study are somewhat limited by looking at only two systems, the correspondences, and the error-correction times used for this research, a reference point for future research is provided. This research does provide worthwhile information with respect to the performance of commercially available ASR systems and novice users' acceptability and satisfaction with the systems.

The results of the study showed that both systems could improve in the area of user satisfaction and word accuracy recognition. Thus, reevaluating the needs and expectations of the user would be beneficial. Research addressing which error-correction procedure methods are preferred by users and why would be helpful. System-required training should also be re-examined to determine if more training time should be required in order for users to obtain satisfactory results. Users expect that the system-required training will provide satisfactory results and their expectations exceed that of the systems capabilities for first-time users. Independent variables such as gender, age, novice and experienced users should also be evaluated with respect to word accuracy and user satisfaction. Another important study that will prove beneficial is one that examines background noise

levels and word accuracy and user satisfaction results. Evaluating users and word accuracy results, and ratings of the systems over an extended period of time to test if satisfactory results are achieved with increased use would be advantageous. The following areas or applications also require future research in ASR: users with disabilities, telecommunications (operator service), automobile in-vehicle tasks, and government and military.

Summary

This research investigated the effects of systems, error-correction time, and correspondence types on word accuracy and user satisfaction. The experimental results provide information regarding how users' subjective ratings relate to word accuracy recognition achieved by the system. It was found that word recognition accuracy achieved does effect user satisfaction (or their subjective ratings). It was also found that with increased error-correction time, word accuracy results improved. Additionally, the results found that Personal Correspondence achieved the highest mean word accuracy rate for both systems and that Dragon Systems achieved the highest mean word accuracy recognition for the Correspondences explored in this research. The results of the study will expand the current research area with respect to Automatic Speech Recognition systems

References

Acero, A. (1993). Acoustical and environmental robustness in automatic speech recognition. Boston, MA: Kluwer Academic.

Atal, B.S. (1995). Speech technology in 2001: New research directions. In Proceedings of the National Academy of Science. Vol. 92, pp. 10046-10051.

Baber, C. (1991). Speech Technology in Control Room Systems: A Human Factors Perspective. New York: Ellis Horwood.

Barber, C. , and Noyes, J. (1996). Automatic speech recognition in adverse environments. Human Factors, 38, 142-156.

Bahl, L.R., Bakis, R., Cohen, P.S., Cole, A.G., Jelinek, F., Lewis, B.L., and Mercer, R.L. (1981). "Speech recognition of a natural text read as isolated words." Presented at the IEEE International Conference on Acoustics, Speech and Signal Processing.

Cohen, P.R., and Oviatt, S.L. (1995). The role of voice input for human-machine communication. In Proceedings of the National Academy of Science. Vol. 92, pp. 9921-9927.

Flanagan, J. (1995). Research in speech communication. In Proceedings of the National Academy of Science. Vol. 92, pp. 9938-9945.

Furui, S. (1995). Toward the ultimate synthesis/recognition systems. In Proceedings of the National Academy of Science. Vol. 92, pp. 10040-10045.

Gellatly, A.W. (1997). The use of speech recognition technology in automotive applications. Doctoral Dissertation, Virginia Polytechnic Institute and State University, Blacksburg, VA.

Kamm, C. (1995). User interfaces for voice applications. In Proceedings of the National Academy of Science. Vol. 92, pp. 10031-10037.

Kato, Y. (1995). The future of voice-processing technology in the world of computers and communications. In Proceedings of the National Academy of Science. Vol. 92, pp. 10060-10063.

Lee, K., Hon, H., and Reddy, R. (1990). An overview of the sphinx speech recognition system. IEEE Transactions of Acoustics, Speech, and Signal Processing, Vol. 38 (1), 35-45.

Levision, S.E., and Fallside, F. (1995). Speech technology in the year 2001. In Proceedings of the National Academy of Science. Vol. 92, pp. 10038-10039.

Levision, S. and Roe D. (1990). A perspective on speech recognition. In IEEE Communications Magazine, pp. 28-34.

Levitt, H. (1995). Processing of speech signals for physical and sensory disabilities. In Proceedings of the National Academy of Science. Vol. 92, pp. 9999-10006.

Lieberman, M. (1995). Computer speech synthesis: Its status and prospects. In Proceedings of the National Academy of Science. Vol. 92, pp. 9928-9931.

Makhoul, J., and Schwartz, R. (1995). State of the art in continuous speech recognition. In Proceedings of the National Academy of Science. Vol. 92, pp. 9956-9963.

Moore, D.W. (1994). Automatic speech recognition for electronic warfare verbal reports. Unpublished master's thesis, Virginia Polytechnic Institute and State University, Blacksburg, VA.

Nakatsu, R., and Suzuki, Y. (1995). What does voice-processing technology support today? In Proceedings of the National Academy of Science. Vol. 92, pp. 10023-10030.

Norman, D.A. (1988). The psychology of everyday things. New York: Basics Books.

Oberteuffer, J.A. (1995). Commercial applications of speech interface technology: An industry at the threshold. In Proceedings of the National Academy of Science. Vol. 92, pp. 10007-10010.

Preece, J. ed. (1993). A guide to usability: human factors in computing. Addisson Wesley, The Open University.

Randall, N. (1998). Computers take a memo. In PC Magazine. Vol. 17 (1), pp. 235-237.

Roe, D.B. (1995). Deployment of human-machine dialogue systems. In Proceedings of the National Academy of Science. Vol. 92, pp. 10017-10022.

Roe, D. and Wilpon, J. (1993). Whither speech recognition: The next 25 Years. In IEEE Communications Magazine. pp. 54-61.

Sanders, M.S., and McCormick, E.J. (1993). Human Factors in Engineering and Design (7th ed.). New York: McGraw-Hill.

Schafer, R.W. (1995). Scientific bases of human-machine communication by voice. In Proceedings of the National Academy of Science. Vol. 92, pp. 9914-9920.

Seelbach, C. (1995). A perspective on early commercial applications of voice-processing technology for telecommunications and the aids for the handicapped. In Proceedings of the National Academy of Science. Vol. 92, pp. 9989-9990.

Weinstein, C.J. (1995). Military and government applications of human-machine communication by voice. In Proceedings of the National Academy of Science. Vol. 92, pp. 10011-10016.

Wilpon, J.G. (1995). Voice-processing technologies-Their application in telecommunications. In Proceedings of the National Academy of Science. Vol. 92, pp. 9991-9998.

Wilson, J. and Corlett, E. (1990). Evaluation of human work. London: Taylor & Francis.

Appendix A: Questionnaire

Questionnaire

Instructions: Please answer the following questions honestly and place the appropriate response in the space provided.

1. Are you an engineering undergraduate or graduate student at Virginia Tech.

   Yes No

2. Are you a Freshman, Sophomore, Junior, Senior, or Graduate Student?

   Freshman Sophomore Junior Senior Graduate Student

3. Is English your "first language?"

   Yes No

4. Where is your hometown (i.e., where are you from)?

---

5. Are you a male or female?

   Male Female

6. Have you ever used Automatic Speech Recognition?

   Yes No

7. Have you ever used Automatic Speech Recognition for Dictation?

   Yes No

_____ _____

Signature Date

Appendix B: IRB Package

Experimental Protocol

for:

Evaluating the Effects of Automatic Speech Recognition

1. Justification of Research Project

The purpose for the research is to examine how speech recognition software system-required training and varied levels of error-correction time affect word accuracy. This study will examine the relationships between (1) speech recognition software system-required training and the system's overall performance (i.e., word accuracy), and (2) error-correction time by the user and improved word accuracy in performing dictation tasks.

With the significant number of computer users and the inexpensive availability of software that support automatic speech recognition, continuous research in the area regarding its usability and effectiveness is needed. Some consideration has been given to various commercial applications regarding automatic speech recognition. Since in the past most automatic speech recognition systems used isolated word speech, some research in the area exist. However, due to technological advances, advanced research in almost any area related to automatic speech recognition systems is warranted.

As in many areas, research issues regarding large-scale systems and industries as it relates to ASR receives the most attention. However, other areas that requires significant attention are often over-looked. For this reason, this research intends to look at automatic speech recognition at a smaller level. Many individual consumers who will purchase and use automatic software recognition from a different aspect than that of commercial industries, such as telecommunications.

Consumers that purchase the software for personal use will mainly use the ASR for dictation of correspondences and documents. This research intends to examine ASR software packages used in conjunction with personal computers for the purpose of dictation and to assess user satisfaction of such systems. The

creation of speech recognition technology software is revolutionizing the way people receive and process information. Users can now enter text and data into a personal computer verbally. This new technology software allows users to voice commands in order to perform tasks that would typically require a mouse to open menus or move the cursor. Speech recognition software can be used in conjunction with a PC or Mac and the aid of a microphone headset.

2. **Procedures**

Subject Population

Subjects for this experiment will be recruited from undergraduate and graduate classes in the engineering department. 13 subjects will be used, one will be used during pre-testing and development and 12 will be used for the actual data collection. Subjects will be paid $10.00 per session for their participation in the experiment.

A questionnaire will be used to ensure that the subjects' primary language is English and he/or she has not used automatic speech recognition for the purpose of dictation in the past. There will be no age or gender restrictions.

Experiment Tasks

The experiment has subjects performing five tasks. The five tasks are:

Task 1- Training the system which will take approximately 45-60 minutes for System 1 (Dragon NaturallySpeaking Preferred) and approximately 45-60 minutes for System 2 (IBM Via Voice Gold),

Task 2- Dictating 3 paragraphs to assess word accuracy which will take approximately 5-10 minutes,

Task 3- Correcting errors made by the system during dictation for a period of 5 minutes to assess word accuracy rates which will take approximately 5-10 minutes.

Task 4- Correcting errors made by the system during dictation for a period of 10 minutes to assess word accuracy rates which will take approximately 10-15 minutes.

Task 5- Completing a user satisfaction survey.

Subjects will be required to attend 2 sessions, a session using System 1 and a session using System 2. Both sessions will last approximately 2 hours.

3. **Risks and Benefits**

This experiment presents minimal risks to the subjects volunteering for the research (no more than using a computer). Subjects may experience some fatigue due to the length of the experiment and the amount of dictation time required. However, subjects will be given a short break after training the system. There are no direct benefits to the subject from this research (besides payment). Students will not be encouraged to participate in the experiment by means of promises nor guarantees of benefits. However, subject participation should provide possible improvements to user-developed and system-required training and the overall functioning of automatic speech recognition software systems that support dictation.

4. **Confidentiality/Anonymity**

The data gathered will be treated with confidentiality. Soon after subjects have participated their name will be separated from their data. A coding system will be used to identify data by subject number only (e.g., Subject No. 3).

5. **Informed Consent**

Please see attached sheets labeled- Informed Consent for Participants.

6. **Biographical Sketch**

Dr. Brian Kleiner, Chairperson

Dr. Kleiner is an associate professor, in the Department of Industrial and Systems Engineering at Virginia Polytechnic Institute and State University. He is also director of the Macroergonomics and Group Decision Systems Laboratory. Dr. Kleiner's research interests focus on macroergonomics, sociotechnical systems research and design of cross-cultural/minority, team-based, and virtual/agile systems; function allocation in automation and job design; Group Decision Support Systems/Computer-supported cooperative work; Performance measurement. He also has a background and experience in management systems

engineering, human factors engineering, safety, quality management, strategic management, benchmarking, environmental management, organizational assessment, and process improvement. Dr. Kleiner's publications in professional journals include articles on socio-technical systems, automation and job design, process control, human performance, benchmarking, productivity, TQM and performance measurement.

Hope Doe

Ms. Doe received a B.S. degree in Industrial Engineering from South Carolina State University in Orangeburg, SC in 1996. She interned for two summers with the Department of Energy in Aiken, South Carolina and Las Vegas Nevada. Hope's internships responsibilities utilized both industrial engineering and human factors engineering concepts and principles. She is currently pursuing a Master's Degree in Industrial and Systems Engineering (human factors option) from Virginia Tech. She served as a Graduate Teaching Assistant (GTA) for Work Methods and Measurement Engineering, Introduction to Human Factors, and Senior Design. Hope is currently an active member of the Human Factors and Ergonomics Society. Her research interest in human factors include macroergonomics, automatic speech recognition, and training.

Virginia Polytechnic Institute and State University

Informed Consent for Participants

Title of Project: Evaluating the Effects of Automatic Speech Recognition Word Accuracy

Investigators: Hope L. Doe, Industrial and Systems Engineering graduate student

Dr. Brian Kleiner, Industrial and Systems Engineering Professor

I. The Purpose of the Research

The purpose for the research is to examine how speech recognition software system-required training and varied levels of error-correction time affect word accuracy. This study will examine the relationships between (1) speech recognition software system-required training and the system's overall performance (i.e., word accuracy), and (2) error-correction time by the user and improved word accuracy in performing dictation tasks.

II. Procedures

In the study you will be asked to thoroughly read the informed consent form and complete a screening questionnaire. You will be asked to perform five tasks in the experiment. The five tasks are:

Task 1- Train the system which will take approximately 45-60 minutes for System 1 (Dragon NaturallySpeaking Preferred) and approximately 45-60 minutes for System 2 (IBM Via Voice Gold),

Task 2- Dictate 3 paragraphs to assess word accuracy which will take approximately 5-10 minutes,

Task 3- Correct errors made by the system during dictation for a period of 5 minutes to assess word accuracy rates which will take approximately 5-10 minutes.

Task 4- Correct errors made by the system during dictation for a period of 10 minutes to assess word accuracy rates which will take approximately 10-15 minutes.

Task 5- Complete a user satisfaction survey.

The research experiment will take place in Whittemore 568, the Macroergonomics Lab in the Human Factors Engineering Center. You will be asked to attend 2 sessions, a session using System 1 and a session using System 2. Each session will last approximately 2 hours.

III. Risks

This experiment presents minimal risks to the subjects volunteering for the research (no more than normal computer use). You may experience some fatigue due to the length of the experiment and the amount of dictation time required. However, you will be given a five minute break during the experiment.

IV. **Benefits of this Project**

There are no direct benefits to the you from this research. I cannot encourage you to participate in the experiment by means of promises nor guarantees of benefits. Your participation should however provide possible improvements to user-developed and system-required training and the overall functioning of automatic speech recognition software systems that support dictation.

V. **Extent of Anonymity and Confidentiality**

The data gathered will be treated with confidentiality. Soon after you have participated your name will be separated from the data. A coding system will be used to identify your data by Subject number only (e.g., Subject No. 3).

VI. **Compensation**

You will be paid $10 per session for the time you actually spend in the experiment. Payment will be made immediately after you have finished your participation.

VII. **Freedom to Withdraw**

You should know that at any time you are free to withdraw from participation in this research program without penalty and will be paid for sessions completed.

VIII. **Approval of Research**

This research project has been approved, as required, by the Institutional Review Board for Research Involving Human Subjects at Virginia Polytechnic Institute and State University and the Department of Industrial and Systems Engineering.

IX. **Subject's Responsibilities**

I voluntarily agree to participate in this study. I have the following responsibilities:

1. I should not volunteer for participation in this research if "English" is not my primary.
2. I should not volunteer for participation in this research if I have used Automatic Speech Recognition software for the purpose of dictation.

X. **Subject's Permission**

I have read and understand the Informed Consent and conditions of this project. I have had all my questions answered. I hereby acknowledge the above and give my voluntary consent for participation in this project.

If I participate, I may withdraw at any time without penalty. I agree to abide by the rules of this project

_____ _____

Signature Date

Should I have any questions about this research or its conduct. I may contact:

Hope Doe, Principal Investigator	552-6476
Dr. Brian Kleiner, Faculty Advisor	231-4926
H. T. Hurd, Chair IRB Research Division	231-9359

Appendix C: Paragraphs used for Dictation

Personal Correspondence

I would like to meet with you on Friday, May, 7, to discuss plans for the up-coming conference. Our group has to complete the registration brochures and plan the banquet that will be held on the last day of the conference. I believe that we can have the brochures out to members by the end of the month. However, everyone will need to come to the meeting with their assigned tasks completed in order to meet this goal. Let's work to make this the best conference ever.

Business Correspondence

I am sorry that the item you purchased from our store did not meet your satisfaction. Thank you for returning it. Because you are covered by our 30-day guarantee policy, I am enclosing a credit voucher that covers the cost of the item. I sincerely hope that you will return to our store again to purchase other items that you may need. Our sales staff will do their best in serving you. Thank you for shopping at Merchant Mart. We look forward to serving you again.

Technical Correspondence

The proposed research seeks to determine if ergonomically-designed or non-ergonomically designed hand tools provide the least muscular strain in various wrist postures. After determining which hand tool provides the least muscular strain, the results of the study can be distributed or assessable to manufacturers and consumers to assist in decreasing the number of cumulative trauma disorders and problems associated with human strain. An ergonomically- and non-ergonomically- designed hammer will be used in the study. The subjects will hammer nails into a plywood board using both hammers in three wrist positions.

Appendix D: User Satisfaction Survey

User Satisfaction Survey

Subject # _____

System 1- Dragon Systems System 2- IBM Via Voice

1. The system-required training by the system aided in word accuracy.

Strongly Agree	Agree	Undecided	Disagree	Strongly Disagree

2. During the system-required training, I experienced fatigue.

Strongly Agree	Agree	Undecided	Disagree	Strongly Disagree

3. I felt the system-required training was adequate:

Strongly Agree	Agree	Undecided	Disagree	Strongly Disagree

4. I felt comfortable while dictating using the system.

Strongly Agree	Agree	Undecided	Disagree	Strongly Disagree

5. I felt that I dictated the paragraphs as I would in normal conversation.

Strongly Agree	Agree	Undecided	Disagree	Strongly Disagree

6. I had no problem remembering commands (such as “New Paragraph”).

Strongly Agree	Agree	Undecided	Disagree	Strongly Disagree

7. I felt the error-correction procedures were tedious.

Strongly Agree	Agree	Undecided	Disagree	Strongly Disagree

8. Overall, I was pleased with the speech recognition software’s performance.

Strongly Agree	Agree	Undecided	Disagree	Strongly Disagree

9. Additional Comments:

10. Rate the System from 1-100 based on user acceptability.

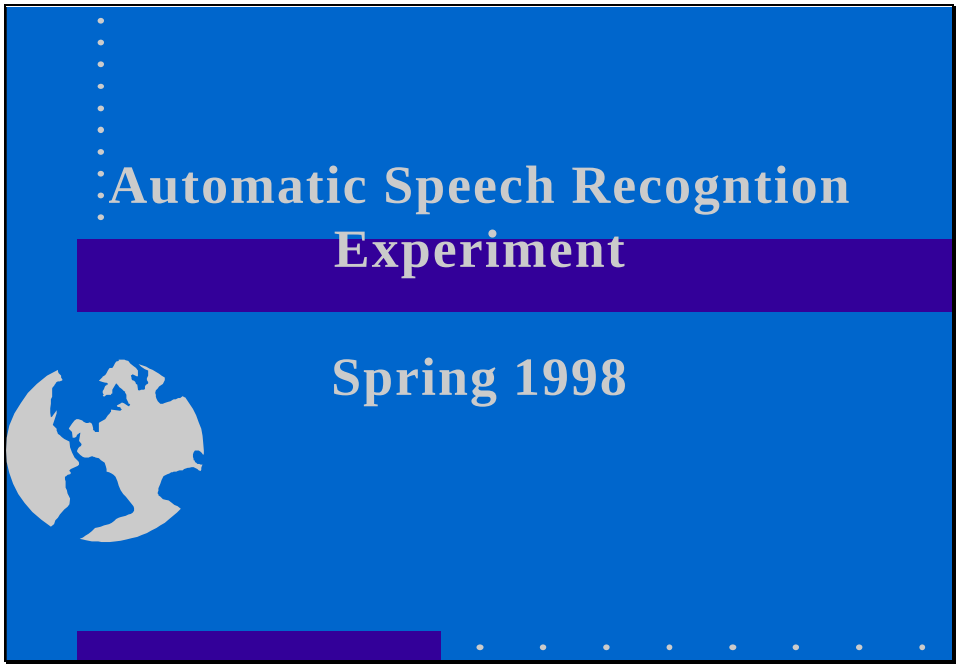
_____	_____	
5- min EC- time	10-min EC- time	
_____	_____	_____
Personal Cor.	Business Cor.	Technical Cor.
_____		
Overall		

EC= error correction Cor. = Correspondence Type

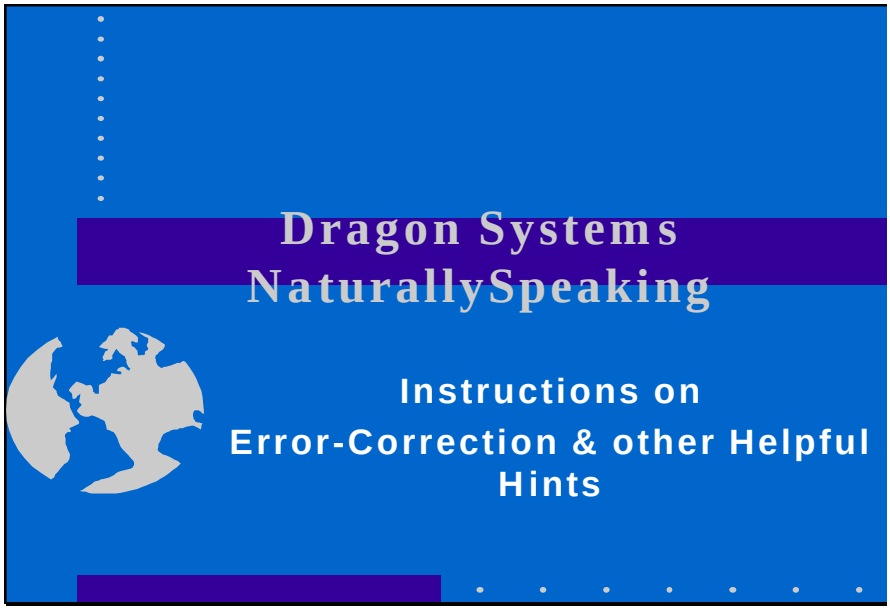
APPENDIX E: POWERPOINT PRESENTATIONS

PowerPoint for Dragon System Naturally Speaking

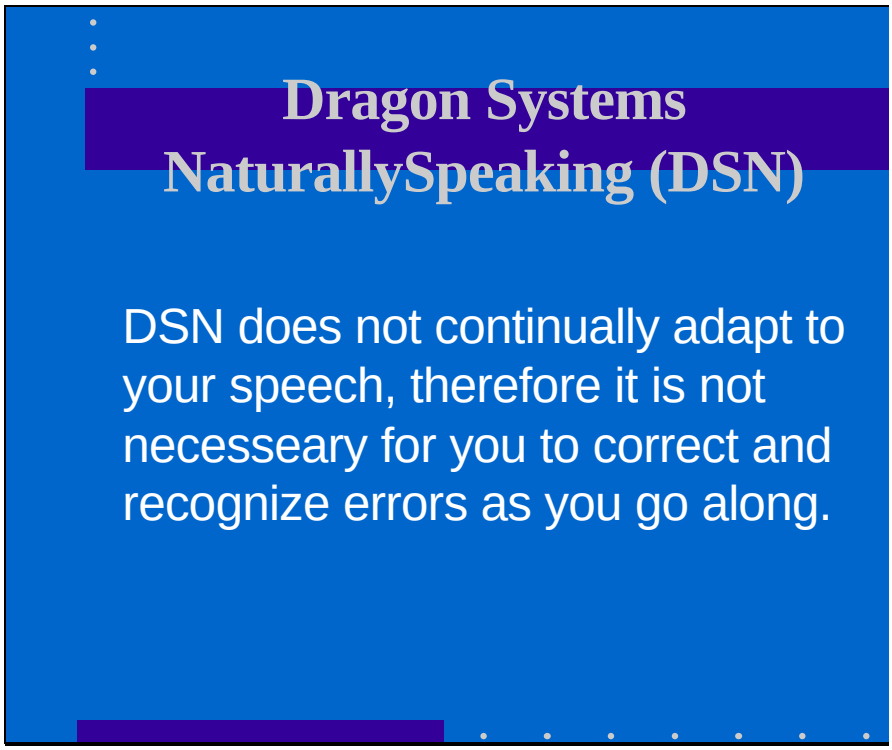
Slide 1



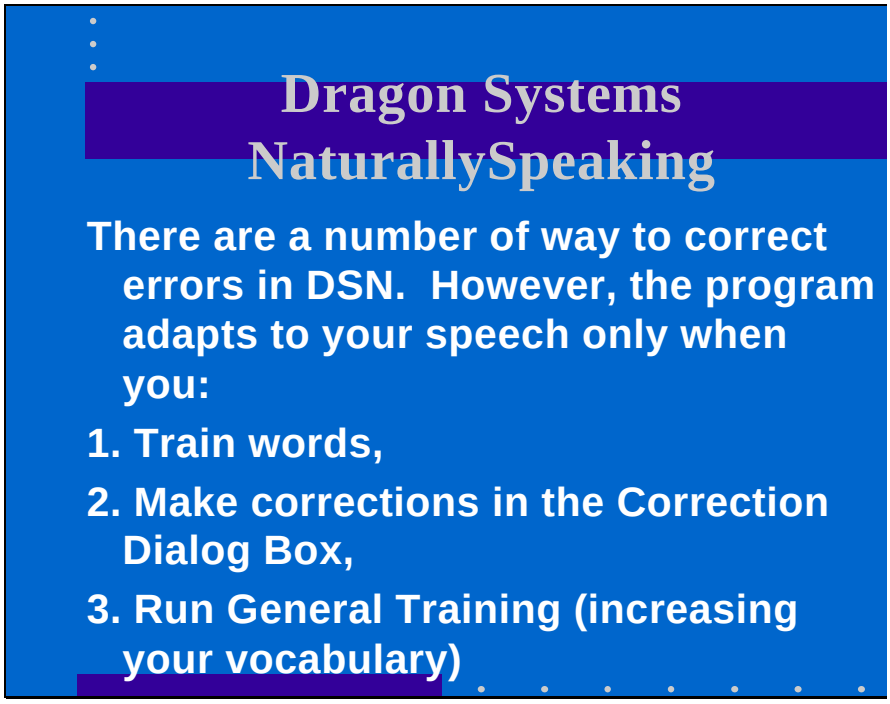
Slide 2



Slide 3



Slide 4



Slide 5

1. Training Words

To train words that Dragon NaturallySpeaking has misrecognized, you should:

1. Select Tools from the pull-down menu,
2. Select “Train Words”
3. Type the word you want to train
4. Press “Record” and say the word and press done when complete

Slide 6

2. Using the Correction-Dialog Box

1. Say “Correct (what ever word it is you want to correct)
2. The correction-dialog box will appear allowing you to choose from a given list the correct word or type in the correct word and then have you train the system.

Slide 7

3. General Training

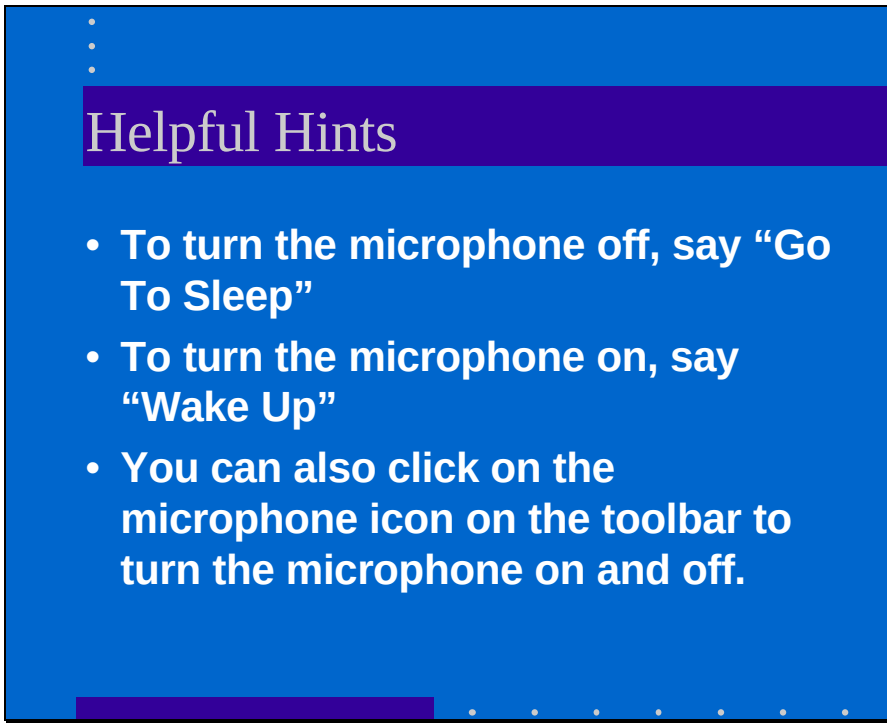
You can also go back to general training and increase your speech files so that system can more accurately determine what you are saying.

Slide 8

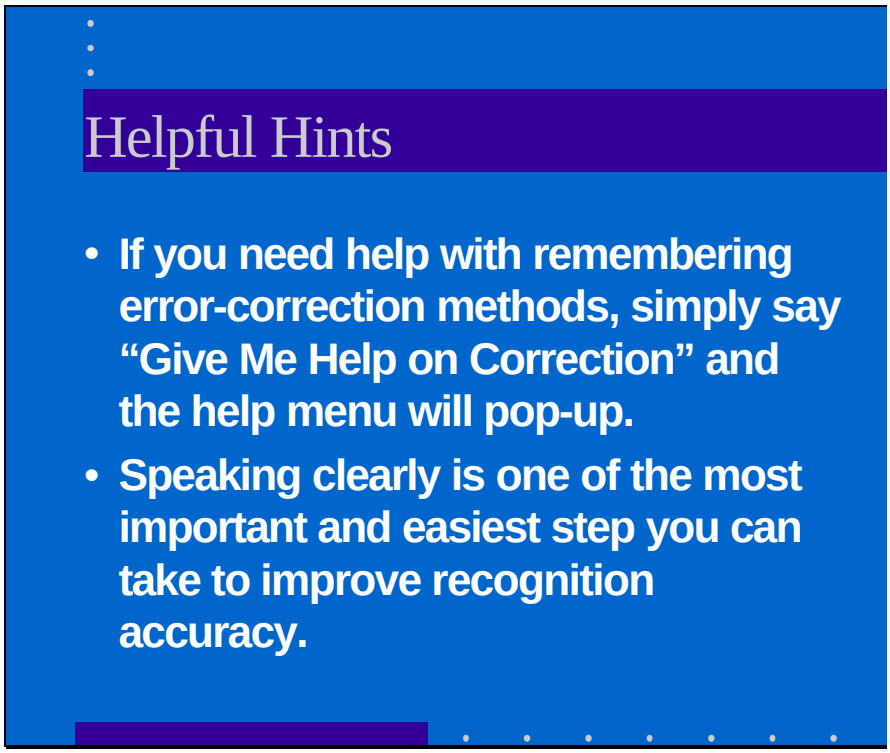
IS THAT IT?

No, there are other ways you can correct errors. However, it does not help the system improve recognition word accuracy. Therefore, you will use those methods that should increase recognition accuracy.

Slide 9



Slide 10



Slide 11

Helpful Hints

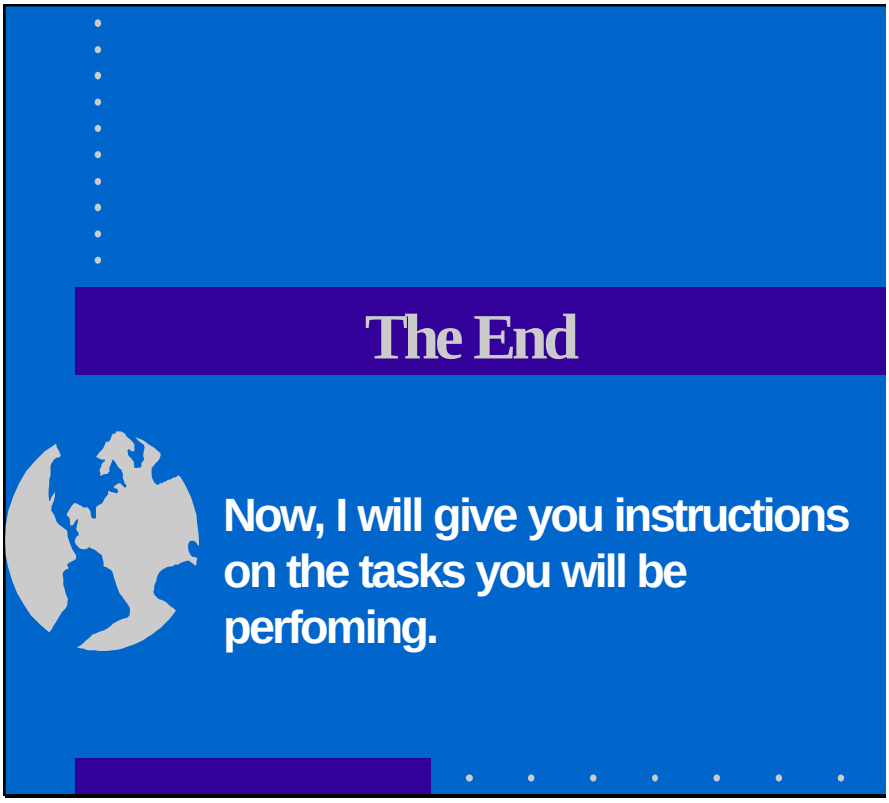
- **If extra words, for example “and” or “the” often appear in your document when you pause the program may be interpretting your breath as speech. Try moving the microphone slightly farther from your mouth and to the side.**

Slide 12

Other Features

There are many things that Dragon Naturally Speaking can do that will not be demonstrated in the experiment. However, if time allots (and you would like), you may explore some of the other features at your leisure.

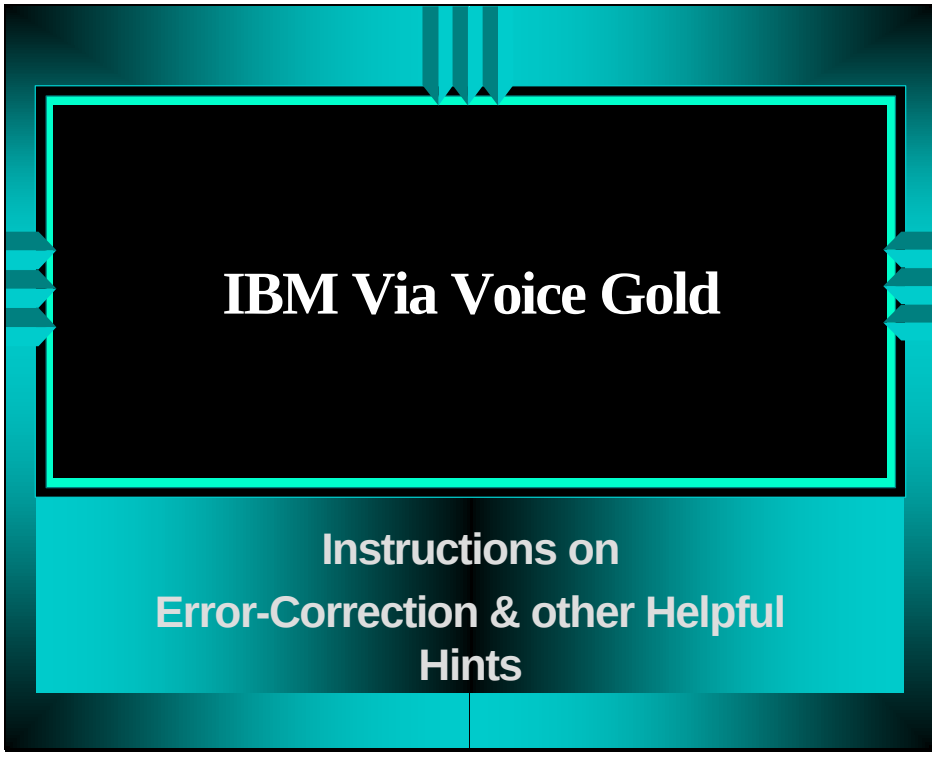
Slide 13


The End
Now, I will give you instructions on the tasks you will be perfoming.

PowerPoint Presentation for Via Voice

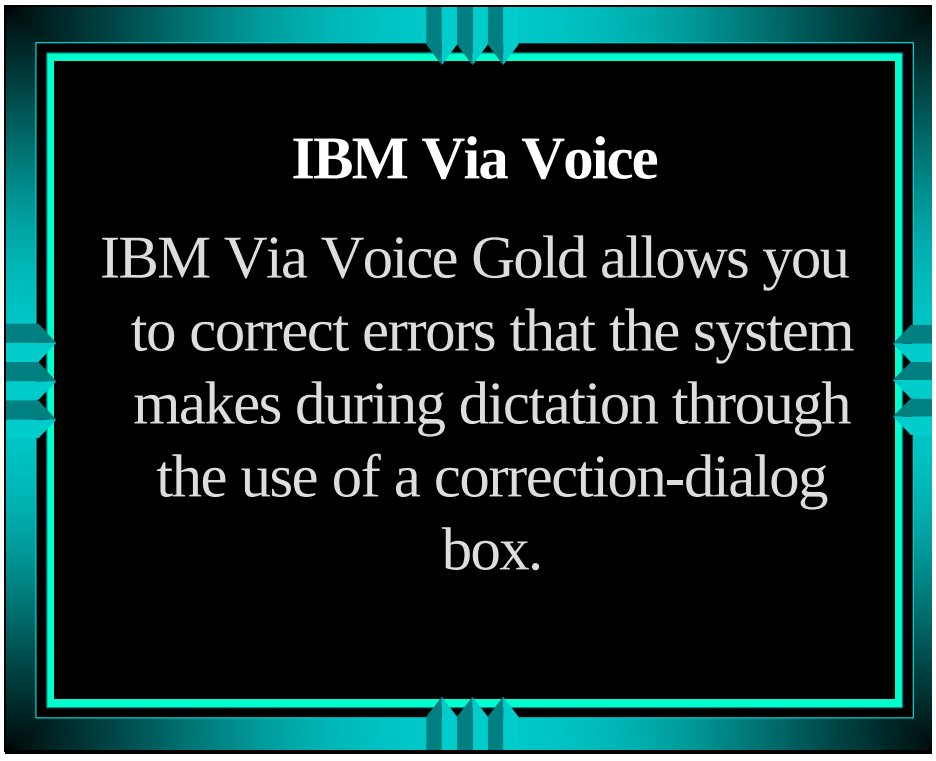
Slide 1

Slide 2

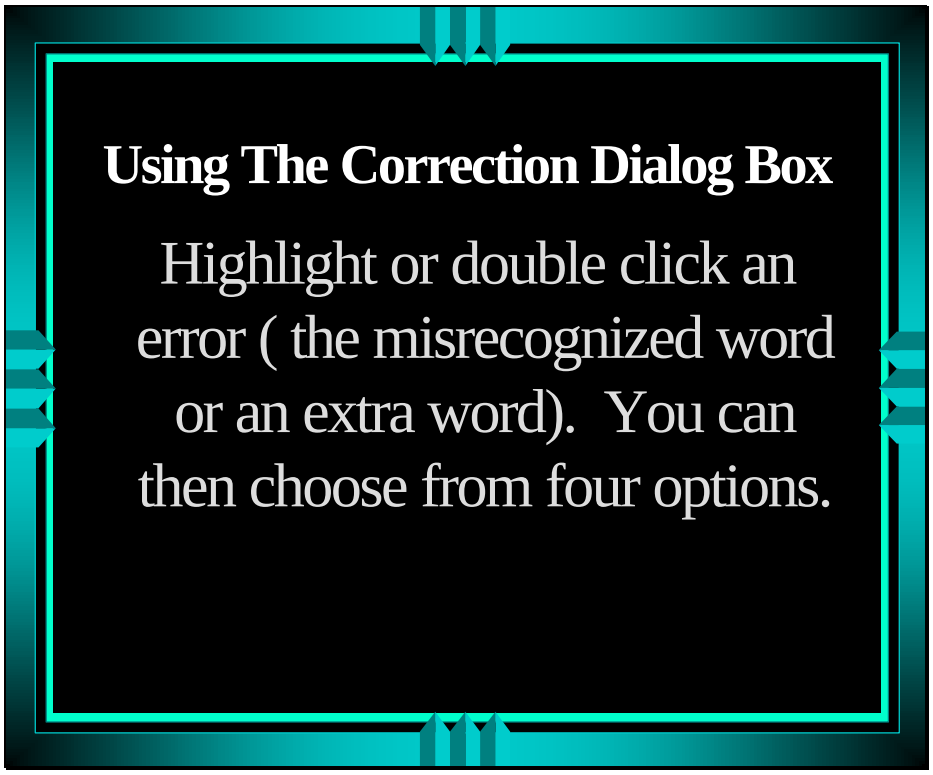


Slide 3

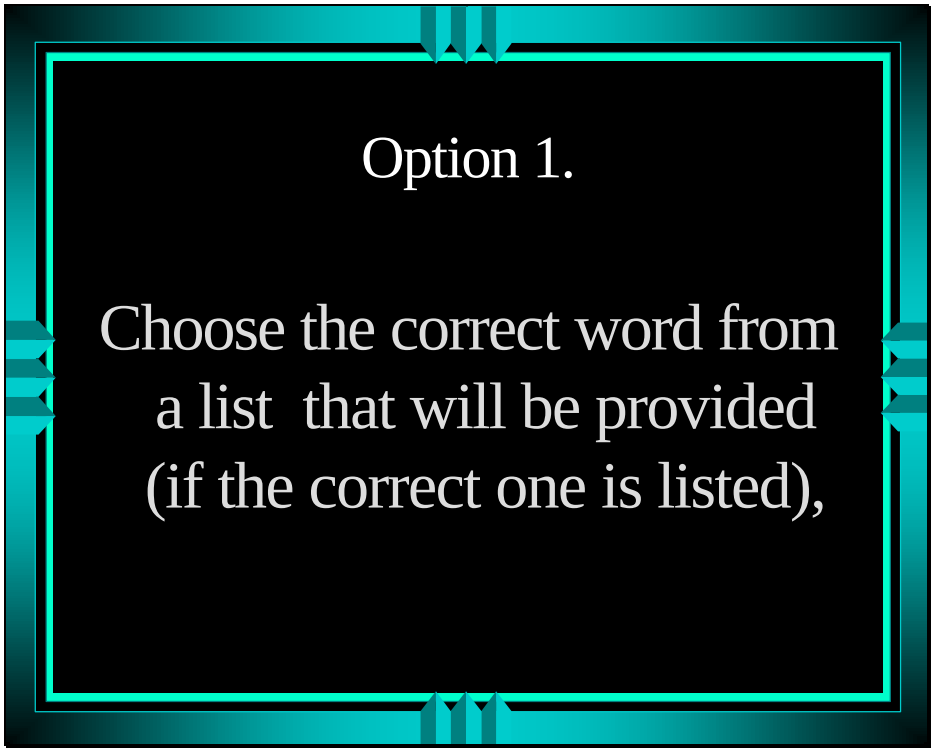
Slide 4



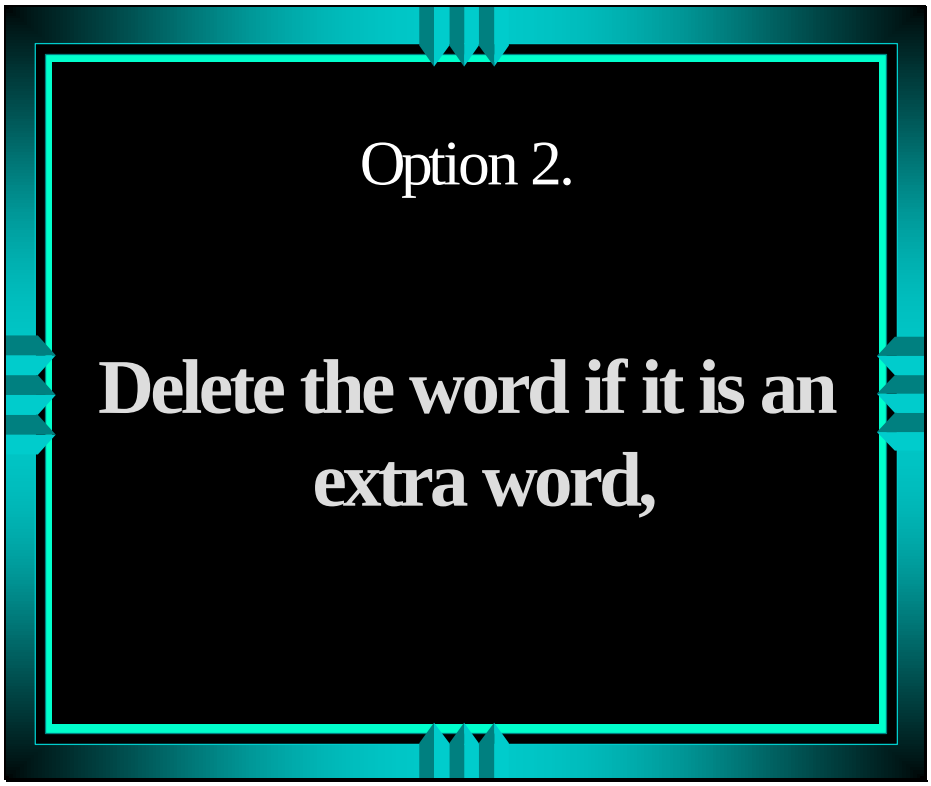
Slide 5



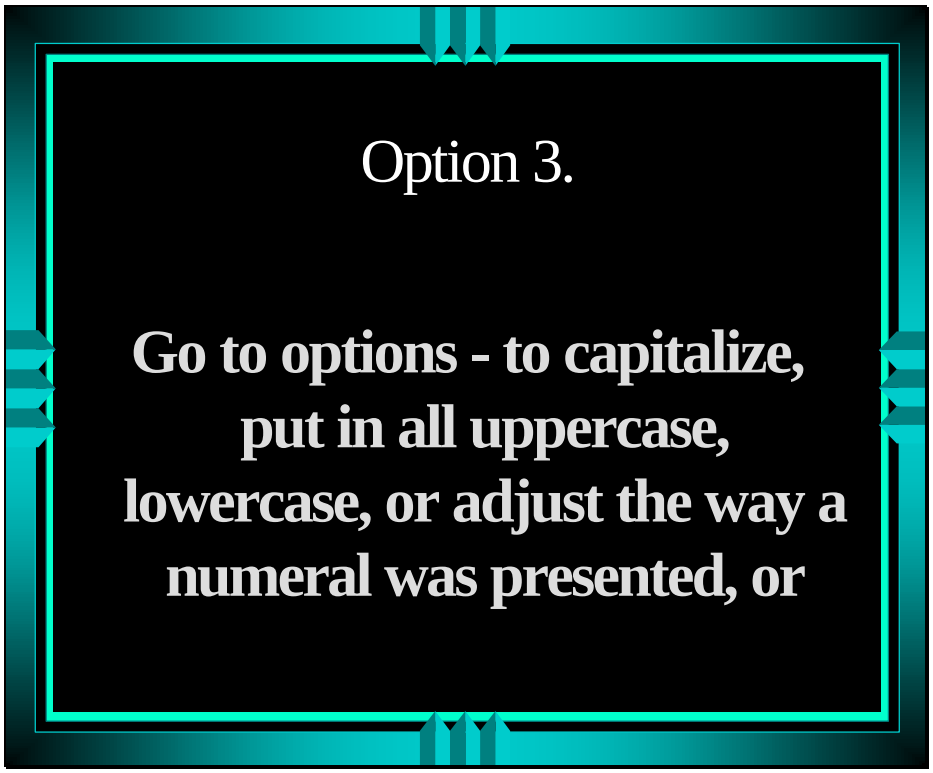
Slide 6



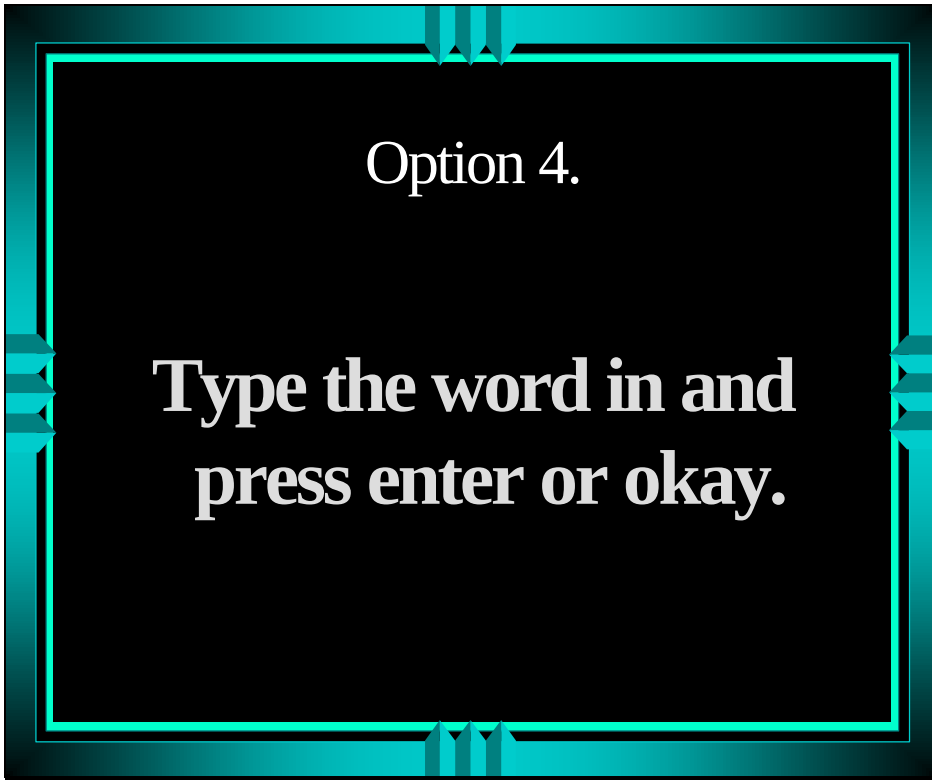
Slide 7



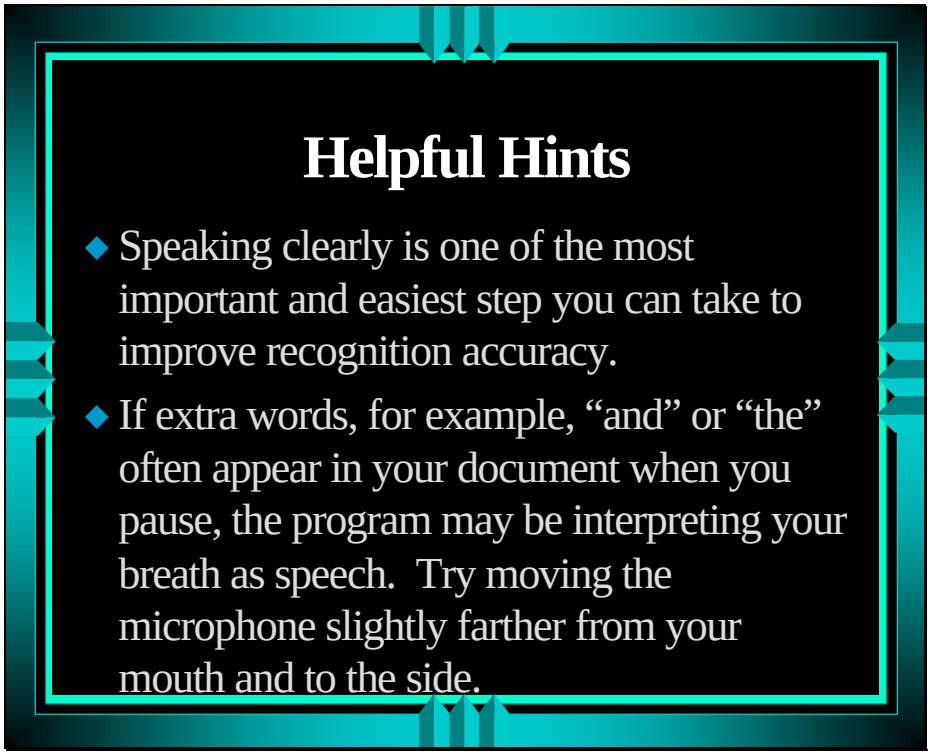
Slide 8



Slide 9



Slide 10



Slide 11


Helpful Hints
To Start Dictation, you can to go to Dictation in the pull down menu and click “Begin Dictation” or say “Begin Dictation.”
To End Dictation, say “Stop Dictation.”

APPENDIX F: RAW DATA

Raw Data For Word Accuracy Results

Subject	System Type	Error-Correction Time	Correspondence Type	Word Accuracy
1	1	1	1	97
1	1	1	2	72
1	1	1	3	86
1	1	2	1	99
1	1	2	2	80
1	1	2	3	89
1	1	3	1	100
1	1	3	2	96
1	1	3	3	93
1	2	1	1	96
1	2	1	2	79
1	2	1	3	96
1	2	2	1	96
1	2	2	2	81
1	2	2	3	92
1	2	3	1	98
1	2	3	2	97
1	2	3	3	96
2	1	1	1	95
2	1	1	2	77
2	1	1	3	80
2	1	2	1	89
2	1	2	2	81
2	1	2	3	85
2	1	3	1	93
2	1	3	2	94
2	1	3	3	73

2	2	1	1	72
2	2	1	2	75
2	2	1	3	74
2	2	2	1	87
2	2	2	2	89
2	2	2	3	94
2	2	3	1	90
2	2	3	2	96
2	2	3	3	96
3	1	1	1	85
3	1	1	2	70
3	1	1	3	83
3	1	2	1	80
3	1	2	2	91
3	1	2	3	93
3	1	3	1	87
3	1	3	2	93
3	1	3	3	96
3	2	1	1	58
3	2	1	2	57
3	2	1	3	74
3	2	2	1	60
3	2	2	2	57
3	2	2	3	76
3	2	3	1	83
3	2	3	2	68
3	2	3	3	79
4	1	1	1	35
4	1	1	2	20
4	1	1	3	20

4	1	2	1	39
4	1	2	2	31
4	1	2	3	25
4	1	3	1	46
4	1	3	2	31
4	1	3	3	50
4	2	1	1	89
4	2	1	2	90
4	2	1	3	69
4	2	2	1	89
4	2	2	2	91
4	2	2	3	82
4	2	3	1	93
4	2	3	2	90
4	2	3	3	88
5	1	1	1	54
5	1	1	2	34
5	1	1	3	42
5	1	2	1	78
5	1	2	2	91
5	1	2	3	83
5	1	3	1	92
5	1	3	2	68
5	1	3	3	94
5	2	1	1	93
5	2	1	2	73
5	2	1	3	80
5	2	2	1	95
5	2	2	2	89
5	2	2	3	86

5	2	3	1	93
5	2	3	2	90
5	2	3	3	88
6	1	1	1	79
6	1	1	2	63
6	1	1	3	76
6	1	2	1	73
6	1	2	2	79
6	1	2	3	68
6	1	3	1	86
6	1	3	2	75
6	1	3	3	93
6	2	1	1	91
6	2	1	2	87
6	2	1	3	79
6	2	2	1	93
6	2	2	2	95
6	2	2	3	90
6	2	3	1	92
6	2	3	2	93
6	2	3	3	93
7	1	1	1	80
7	1	1	2	71
7	1	1	3	68
7	1	2	1	96
7	1	2	2	82
7	1	2	3	85
7	1	3	1	91
7	1	3	2	86
7	1	3	3	99

7	2	1	1	95
7	2	1	2	95
7	2	1	3	93
7	2	2	1	97
7	2	2	2	95
7	2	2	3	96
7	2	3	1	98
7	2	3	2	97
7	2	3	3	99
8	1	1	1	87
8	1	1	2	60
8	1	1	3	78
8	1	2	1	87
8	1	2	2	62
8	1	2	3	95
8	1	3	1	90
8	1	3	2	98
8	1	3	3	85
8	2	1	1	79
8	2	1	2	73
8	2	1	3	73
8	2	2	1	87
8	2	2	2	77
8	2	2	3	80
8	2	3	1	97
8	2	3	2	84
8	2	3	3	86
9	1	1	1	74
9	1	1	2	63
9	1	1	3	59

9	1	2	1	86
9	1	2	2	57
9	1	2	3	89
9	1	3	1	85
9	1	3	2	81
9	1	3	3	93
9	2	1	1	89
9	2	1	2	74
9	2	1	3	63
9	2	2	1	91
9	2	2	2	76
9	2	2	3	69
9	2	3	1	94
9	2	3	2	86
9	2	3	3	89
10	1	1	1	81
10	1	1	2	66
10	1	1	3	56
10	1	2	1	79
10	1	2	2	75
10	1	2	3	96
10	1	3	1	96
10	1	3	2	83
10	1	3	3	99
10	2	1	1	70
10	2	1	2	76
10	2	1	3	52
10	2	2	1	85
10	2	2	2	83
10	2	2	3	80

10	2	3	1	92
10	2	3	2	90
10	2	3	3	84
11	1	1	1	54
11	1	1	2	12
11	1	1	3	62
11	1	2	1	97
11	1	2	2	84
11	1	2	3	99
11	1	3	1	90
11	1	3	2	87
11	1	3	3	95
11	2	1	1	91
11	2	1	2	90
11	2	1	3	86
11	2	2	1	94
11	2	2	2	95
11	2	2	3	93
11	2	3	1	97
11	2	3	2	99
11	2	3	3	97
12	1	1	1	68
12	1	1	2	58
12	1	1	3	52
12	1	2	1	72
12	1	2	2	67
12	1	2	3	46
12	1	3	1	74
12	1	3	2	70
12	1	3	3	60

12	2	1	1	85
12	2	1	2	65
12	2	1	3	57
12	2	2	1	79
12	2	2	2	89
12	2	2	3	67
12	2	3	1	88
12	2	3	2	93
12	2	3	3	81
13	1	1	1	90
13	1	1	2	81
13	1	1	3	75
13	1	2	1	93
13	1	2	2	87
13	1	2	3	77
13	1	3	1	96
13	1	3	2	91
13	1	3	3	74
13	2	1	1	98
13	2	1	2	96
13	2	1	3	91
13	2	2	1	90
13	2	2	2	96
13	2	2	3	90
13	2	3	1	96
13	2	3	2	79
13	2	3	3	96

Raw Data For User Satisfaction

Subjects	System	Opinion	User Satisfaction
1	1	1	75
1	1	2	95
1	1	3	90
1	1	4	80
1	1	5	85
1	1	6	90
1	2	1	50
1	2	2	90
1	2	3	85
1	2	4	80
1	2	5	80
1	2	6	80
2	1	1	60
2	1	2	90
2	1	3	75
2	1	4	50
2	1	5	60
2	1	6	70
2	2	1	50
2	2	2	95
2	2	3	75
2	2	4	80
2	2	5	85
2	2	6	80
3	1	1	70
3	1	2	70
3	1	3	70

3	1	4	70
3	1	5	75
3	1	6	80
3	2	1	20
3	2	2	50
3	2	3	70
3	2	4	70
3	2	5	60
3	2	6	50
4	1	1	75
4	1	2	75
4	1	3	60
4	1	4	50
4	1	5	50
4	1	6	80
4	2	1	85
4	2	2	85
4	2	3	90
4	2	4	90
4	2	5	90
4	2	6	90
5	1	1	50
5	1	2	80
5	1	3	85
5	1	4	70
5	1	5	60
5	1	6	78
5	2	1	85
5	2	2	90
5	2	3	95

5	2	4	85
5	2	5	90
5	2	6	90
6	1	1	40
6	1	2	70
6	1	3	65
6	1	4	75
6	1	5	50
6	1	6	60
6	2	1	80
6	2	2	95
6	2	3	85
6	2	4	82
6	2	5	82
6	2	6	88
7	1	1	70
7	1	2	73
7	1	3	90
7	1	4	75
7	1	5	60
7	1	6	72
7	2	1	70
7	2	2	95
7	2	3	95
7	2	4	90
7	2	5	90
7	2	6	93
8	1	1	50
8	1	2	70
8	1	3	85

8	1	4	75
8	1	5	75
8	1	6	85
8	2	1	70
8	2	2	90
8	2	3	70
8	2	4	70
8	2	5	70
8	2	6	85
9	1	1	50
9	1	2	80
9	1	3	75
9	1	4	80
9	1	5	75
9	1	6	75
9	2	1	60
9	2	2	75
9	2	3	80
9	2	4	90
9	2	5	70
9	2	6	80
10	1	1	50
10	1	2	100
10	1	3	90
10	1	4	85
10	1	5	85
10	1	6	88
10	2	1	75
10	2	2	80
10	2	3	80

10	2	4	70
10	2	5	70
10	2	6	75
11	1	1	20
11	1	2	80
11	1	3	70
11	1	4	80
11	1	5	80
11	1	6	70
11	2	1	70
11	2	2	80
11	2	3	90
11	2	4	90
11	2	5	80
11	2	6	90
12	1	1	65
12	1	2	69
12	1	3	60
12	1	4	75
12	1	5	40
12	1	6	55
12	2	1	60
12	2	2	70
12	2	3	60
12	2	4	70
12	2	5	80
12	2	6	70
13	1	1	80
13	1	2	90
13	1	3	90

13	1	4	90
13	1	5	80
13	1	6	60
13	2	1	80
13	2	2	50
13	2	3	60
13	2	4	70
13	2	5	70
13	2	6	70

Raw Data for System Training Time

Subjects	System	Training
1	1	55
2	1	60
3	1	63
4	1	59
5	1	50
6	1	75
7	1	58
8	1	60
9	1	61
10	1	52
11	1	50
12	1	60
13	1	63
1	2	35
2	2	30
3	2	32
4	2	33
5	2	27
6	2	29
7	2	31
8	2	35
9	2	30
10	2	32
11	2	28
12	2	28
13	2	33

VITA

Hope L. Doe

OBJECTIVE Industrial Engineer; interests in manufacturing and human factors.

EDUCATION

VIRGINIA POLYTECHNIC INSTITUTE AND STATE UNIVERSITY Blacksburg, VA
Master of Science Industrial and Systems Engineering July 1998
Overall GPA: 3.6

SOUTH CAROLINA STATE UNIVERSITY Orangeburg, SC
Bachelor of Science Industrial Engineering Technology May 1996
Overall GPA: 3.8, ABET Accredited School of Engineering

WORK EXPERIENCE

VIRGINIA POLYTECHNIC INSTITUTE AND STATE UNIVERSITY, INDUSTRIAL AND SYSTEMS ENGINEERING DEPARTMENT Blacksburg, VA
Graduate Teaching Assistant 8/96 - 5/98

- Conducted Work Measurement and Methods Engineering laboratory classes for undergraduate students.
- Graded laboratory reports of students weekly.
- Assisted in teaching-related responsibilities for the Introduction to Human Factors and Senior Design courses.

DEPARTMENT OF ENERGY, NEVADA OPERATIONS Las Vegas, NV
Intern 5/96 - 8/96

- Assisted in preparing the Human Factors Engineering Concept of Operations for the Mined Geological Disposal System Yucca Mountain Project.

WESTINGHOUSE, SAVANNAH RIVER SITE Aiken, SC
Intern 5/95 - 8/95

- Developed an alternative system for the Analytical Lab Cost Charging for the Industrial Engineering Department.
- Reconstructed the Monthly Status Reports for the department's customers.

THE WHITE HOUSE Washington, DC
Intern 5/94 - 8/94

- Acted as a liaison between local and state-elected Governmental officials and Native Americans and the Office of Intergovernmental Affairs.
- Assisted in the organization and implementation of White House events for Governmental officials and Native Americans.

HONORS/ ACTIVITIES

Institute of Industrial Engineeers
Outstanding Graduate Teaching Assistant
Human Factors and Ergonomics Society
Agnes Jones Jackson Scholar
Southern Region Educational Board Scholar