VOICE RECOGNITION SYSTEM IMPLEMENTATION AND LABORATORY EXERCISE

BY

Richard Calvin Sanders

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

MASTERS OF ENGINEERING
in
Industrial and Systems Engineering

APPROVED:

Dr. R.J. Reasor, Chairman

Dr. C.P. Koelling

Dr. J.C. Woldstad

April, 1995
Blacksburg, Virginia

c.2

LD
5655
V851
1995
C663
c.2

VOICE RECOGNITION SYSTEM IMPLEMENTATION AND LABORATORY EXERCISE

by

Richard Calvin Sanders

Project Committee Chairman:

Roderick J. Reasor, Ph.D., P.E.

(ABSTRACT)

Efficient and accurate data collection is vital to the successful management of modern businesses, providing crucial information for decision making. Automatic data collection systems reduce the amount of paperwork, time, and labor needed and the complexity of generating and using this information. Voice recognition systems are one of the newest systems being utilized for automatic data collection. This method uses voice interfacing, offering the possibility of interacting with the humans most natural and best-developed communication skill- speech.

It will be important for industrial engineers to understand how this newer technology can assist in making an operation more efficient. This project was conceived to meet this goal. Steps to completing this project included doing a literature survey on voice recognition, becoming familiar with the voice recognition equipment in the Automatic Data Collection Systems Laboratory at Virginia Tech, setting up the network, and then designing the lab activities for using the equipment.

The main emphasis of this project was the design of the lab activities. These labs provide students with practical hands on experience in using voice recognition systems and gives them a basic understanding of its theory and applications. The set of four laboratory activities includes exercises on voice training, text to speech synthesis, improving the accuracy of voice systems, and using a voice system in conjunction with another automatic data collection device (a bar-code wand).

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

Chapter 1: INTRODUCTION

1.1 Overview

Efficient and accurate data collection is vital to the successful management of modern businesses, providing crucial information for decision making. However, many companies still use paper-based methods to collect data: workers write data on pre-printed forms, then this information is key entered into a computer. When a worker has to look up to obtain data and then shift their gaze down at a piece of paper to record that data, their eyes become more easily fatigued. This often increases the number of mistakes made in recording data and reduces the legibility of their handwriting. The data entry person who then transcribes the data can introduce more errors into the data, resulting from typos or attempts to interpret sloppy handwriting. The hands-free, eyes-free nature of voice data collection equipment allows the operator to concentrate on the task at hand with both their eyes and hands. An automatic data collection method can increase the accuracy of the information recorded and reduce the amount of time it takes to produce crucial reports from collected data.

Voice recognition systems (VRS) are one of the newer technologies being utilized for data collection and can be the key to improving data collection productivity. Many voice systems have the capability to verify data as it is entered to prevent invalid information from being recorded. This also eliminates transcription errors since data is entered directly into the computer. Hence, voice based systems are supposed to gather, verify, and compile reports with greater speed, ease, and accuracy than writing and typing.

While the aim of all data capture technologies is to accurately get information into a computer so it can be acted upon or used in a variety of ways, VRS is one of the only applications that can be used when human judgment is needed and / or when both hands and eyes must be free for gathering information. Voice technology can also be applied where other data collection methods cannot be used. For example, in environments where dirt would ruin keyboards or where safety requires that workers keep constant eye contact with manufacturing equipment such as low-light environments, or where you have big gloves and small buttons on a hand-held terminal. Voice interfaces offer the possibility of interacting with computers using the humans most natural and best-developed communication skill- speech.

Voice recognition has a number of synonyms including Voice I/O, Voice Data Collection, Interactive Voice Systems, Voice Data Input Systems, and a host of other proprietary terms such as Voice Net, Voice Navigator, Voice Data and others.

Of the voice recognition systems generally available to industry today, only a few have reached the point where any worker can strap on the headset, speak whatever he / she wants into the microphone and have the system recognize, decode, and interact with accuracy. However, even these systems have some limitations (vocabulary, speed, etc.)

Whether we realize it or not voice recognition systems are everywhere and will be increasing in the number of applications that it will be used. Currently, voice recognition is being used in car phones so that all that needs to be said is "call home" and the car phone will dial the persons home number automatically. Another simple form of voice recognition is used when a person punches his / her account number over the phone and has an automated bank line give the person their account balance or the last five transactions to their account. Also, Kurzwill, DragonDictate, IBM, and Philips all have speech processors that allow a person to "speak" instead of typing in a word processing environment (although these systems are by no means perfected- see Labriola in reference section)

With this technology being deemed as the wave of the future, "Speech recognition is one of the business consumer's hot buttons," says Bob McBreen, Product Manager for Microsoft's Windows Sound System (Meisel, p.113), it will become important for students to be knowledgeable of its theory and applications. Thus, the purpose of this project is to give students hands-on experience with this new and innovative technology.

1.2 Project Objective

The objectives of this project are to:

1.) study the theory and applications behind voice recognition systems.

2.) define the technology and system components.

3.) use the voice recognition equipment in the Automatic Data Collections Systems Lab (ADCSL) to build and integrate voice recognition systems in conjunction with the bar-coding and radio frequency systems already being put into place to form an integrated manufacturing data collection laboratory.

4.) design a series of laboratory activities to help students gain hand on experience with the use of voice recognition for data collection.

1.3 Scope of the Project

The scope of this project is limited to the integration of the voice recognition equipment in the Automatic Data Collection Systems Lab at Virginia Tech. Several lab exercises will be designed for students to learn the basic theory and applications behind voice recognition systems. In conjunction with existing radio frequency and bar-coding equipment being designed by two other graduate students, the ADCSL will showcase an integrated network of the latest technologies in data collection.

Chapter 2: LITERATURE REVIEW

2.0 Automatic Data Collection Technologies

Automatic identification and data collection systems (AIDCS) have become an integral part of the manufacturing environment. In today's competitive marketplace the bottom line liquidity of a company (eg. quick ratio and current ratio) has become even more important in measuring a company's profitability and sometimes whether it will survive or not. Maintaining control of raw materials and finished good inventories has become the key to effective operations management.

Using AIDCS helps by tightening control from warehouse receipt to customer shipment leading to faster order completion, improved resource utilization, and lower inventory investment. Maintaining this control requires more disciplined material flow, accurate and timely transaction recording, and prompt exception detection at receiving, at production, upon storage and retrieval, and at shipping. This further proves that there is a need for an AIDCS.

Automatic identification and data collection systems reduce the amount of paperwork, time, and labor needed and the complexity of generating and using information. They read source data or code and convert it to digital form for use by computers and processors for controlling equipment, generating reports, or making activity transactions. Accurate identification lowers operating costs and improves material throughput, handling efficiency, real-time accuracy, security and audit-trail accuracy.

A primary benefit of automatic identification systems is timely and accurate data collection. "Traditional collection methods produce error rates ranging from one in 30 for handwritten documents to one in 300 for keyboard input" (Soltis, p.55). Automatic identification systems typically have an accuracy range of one error in three million entries (Soltis, p.55).

Another benefit is speed. Automatic identification devices process hundreds of characters per second. "This is substantially faster than manual data entry at five to seven characters or keyboard entry at 10 to 15 characters per second" (Soltis, p.55).

Automatic identification systems use the following types of technology:

1.) bar code scanning,

2.) optical character readers (OCR),

3.) magnetic stripes,

4.) machine vision systems,
5.) radio frequency,
6.) smart cards, and
7.) voice recognition (described in sections 2.2-2.9).

Each of these technologies will be briefly described in order to provide a holistic view of the automatic data collection options available in today's marketplace.

BAR CODE SYSTEM

Bar-coding is the predominant automatic ID technology used to collect data. It is used for a variety of different applications, including item tracking, inventory control, time and attendance recording, monitoring work-in-process, quality control, check-in / check-out, sortation, order entry, document tracking, controlling access to secured areas, and many others.

A bar-code consists of an array of differing width parallel bars and spaces within which information is encoded (Figure 2.1). The way the bars and spaces are arranged is called its symbology. The characteristics of the different symbologies include the character set, type (discrete or continuous), number of element widths, length (fixed or variable), and the X dimension (width of narrow elements). A few common symbologies include UPC, EAN, INTERLEAVED 2 of 5, CODE 128, CODE 39, and CODABAR.

The bar-code is usually the size of a label and is attached to an object, for instance, an item in a grocery store. This code serves as an identifier for the object that it is attached to. This "identifier" is also part of a database where the other details associated with this particular object are stored. Once the identifier has been decoded, information regarding this particular product can be retrieved from the database. The decoding and access of the database is usually controlled by the system software.

A typical bar coding system consists of three elements:

1.) The bar-code label, which contains the machine readable symbol. The symbol contains information encoded in the wide and narrow bars and spaces.
2.) The bar code scanner, which reads the bar code and convert the optical information into electrical signals. To detect the bars and spaces of a bar code symbol, the scanner's internal light source is directed by its optics onto that symbol. Light reflected from the symbol's white spaces is directed to a photodiode detector inside the

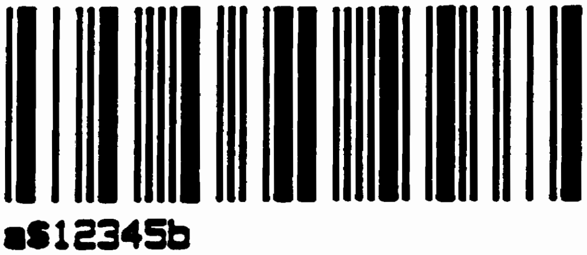


FIGURE 2.1: Bar-Code Symbol

scanner, which in turn generates a small current proportional to the amount of light returned. This analog current is amplified and then converted into a digital waveform by a circuit known as the waveshaper. The digitized signal is then sent to decoding circuitry by the signal processing circuitry.

3.) The decoder is an electronic package that receives signals from the scanner and interprets them into meaningful data. This data is then sent either directly to a host computer or is temporarily stored in the memory of the data collection device. (Librescu, p.6)

OPTICAL CHARACTER RECOGNITION

Optical character recognition (OCR) is a technology which has been used in commercial applications since the 1950's. Optical character recognition (OCR) is used to read "human readable" text on packages and cases (Figure 2.2). OCR requires a near contact reader, which eliminates scanning from a distance. An OCR scanner examines the printed characters using two-dimensional technology. It looks at both the vertical and horizontal axes as part of the decoding process.

Aside from page scanning, OCR technology has not substantially impacted industrial automatic identification systems. OCR equipment usage has declined in recent years due to:

1.) low first read rate with semi-skilled operators,

2.) lack of an automatic omni-directional OCR scanner for check-out counters,

3.) high substitution error rate compared to bar codes, and

4.) difficulty to scan at a distance
(Ackley, p.1.8).

MAGNETIC STRIPES

Magnetic stripe technology uses the magnetic field of an encoding head to record magnetic flux reversals. This information is placed onto a layer of magnetic material similar to that on an audio or video tape. The layer, called a magnetic stripe, is generally attached to the front or back of a paper or plastic card (ex. credit card). A decoder reads the flux reversals and translates them into letters and numbers for processing by a computer.

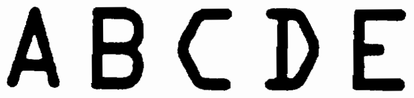


FIGURE 2.2: Optical Character Code

It is possible to encode a great deal of information onto a magnetic stripe and change the information at a later date. Magnetic stripe data is stored as a series of regions with differing magnetization, just like computer tapes or floppy disks. Typically, magnetic stripes are employed in a read-only mode.

Although commonly used for financial transactions, magnetic stripe technology has not been widely adopted for general tracking applications because of:

1.) the unavailability of non-contact scanning equipment,
2.) environmental considerations,
3.) the inability of conventional printing methods to encode magnetic information, and
4.) higher labeling costs when compared to printing technologies
(Ackley, p.1.4).

MACHINE VISION SYSTEMS

A machine vision system consists of a high resolution television camera interfaced to a computer via signal processing circuitry. This set-up uses video technology to present a picture to a processor which digitizes the image to generate a numeric representation of the picture. This is then placed in the computer's memory. The picture may be numbers, letters, bar-codes, or any image presented to the camera. Software programs are then used to process this scene representation to obtain the desired information. Many of today's machine vision systems perform this processing with specialized electronic circuitry (hardware) as opposed to software because of the tremendous speed improvement that hardware-based processing provides.

Such systems are appropriate for a variety of applications such as automatic identification, measurement and inspection, robot guidance and control, materials handling and sorting, and a variety of natural and medical sciences (eg. x-ray interpretation and cartography).

RADIO FREQUENCY DATA COMMUNICATIONS (RF / DC)

Radio frequency data communications is not an automatic identification alternative, but a complementary technology that can be used with automatic identification technologies to communicate from distant locations to host computers in real-time (eg. in conjunction with a voice recognition system). The terminals communicate data over a wireless RF/DC link to an RF base station which relays information to a host computer. RF data communications offer the capability of an on-line, real-time communications link without wires.

RF/DC are typically used in materials handling and in retail applications. In the material handling industry, RF/DC allows shipping, receiving, storage, retrieval, order picking, pick-slot replenishment and other instructions to be transmitted directly to / from terminal operators and the host computer. On the retail side, RF terminals are used for price verification, order entry, and direct store delivery (DSD).

SMART CARDS

The smart card is a method of automatic identification that uses a credit card-sized plastic card with one or more microchips embedded in them. Typically, they are programmable, containing a microprocessor chip and a large database. The microprocessor manages the security entry to one or more application databases. The data in a smart card is generally accessible using a card reading device. Applications of a programmable smart card range from storing input / output information to a numerical control machine to controlling flexible manufacturing cells and for machine assembly control.

The term "smart card" is also applied to plastic cards that only contain memory and are used for applications such as coin replacement (eg. the copy card used to operate the copy machines at Virginia Tech's library) or units of inventory. These integrated circuit read-only memory (IC ROM) cards or IC memory cards have no programmability but can contain a large amount of data. They are similar in concept to magnetic stripe cards, but the data is hidden and the card can hold far more data.

Smart cards can also carry magnetic stripes or embossed characters that contain some of the data stored in the smart card's memory. This allows the card to be used like an ordinary plastic card with pre-existing terminals. Data is added and deleted through interaction with the operating system. The memory is reusable which permits data to be read, written, and rewritten.

2.1 Overview of Voice Recognition Systems (VRS)

2.1.1 General Systems

Hardware

The physical hardware comprising a voice system can range from a single PC card to a network of many workstations linked to a large computer for more intensive processing. The various hardware options also differ in their range of capabilities. Some common hardware configurations are described below:

• ***PC card***- one of the simplest voice system is a PC card that plugs into your existing PC. A PC card is usually capable of recognition and synthesis. A headset or handset with speakers and a microphone is often used to interact with the PC card. (Figure 2.3)

• ***Input station with optional output***- This hardware option consists of a recognizer with optional small readout display or keypad, a microphone for entering input, and a communication link (such as a RS-232 or a radio transmitter) to another computer where data will be stored and processed. Some voice units can send collected data as ASCII text files for import into many popular software packages. An input station may also be capable of output, in which case it will also include a voice synthesizer and a speaker (Figure 2.4).

• ***Workstation***- A so-called "workstation" earns its name because it is a complete voice system capable of prompting the user for input; accepting, recognizing, and verifying input; generating (via synthesis or record and playback) and broadcasting voice output; and storing and manipulating collected data, including the generation of reports and communication with other computer systems. A workstation often takes the form of a PC plus an internal card (hardware) for recognition and synthesis; some form of microphone and speaker combination; and software that both guides the user through the data collection process and generates reports from that data (Figure 2.5). Also available are portable voice data collection devices that contain both the voice recognition and synthesis in the portable unit. These beltworn voice data collection units have the ability to perform calculations, make branching decisions, include database information from a PC and format collected data. Data can be collected in portable batch mode or in real-time mode where data are returned over a radio network (Figure 2.6).

• ***Multi-user network***- This voice system configuration consists of a host computer connected to a network (LAN, WAN, etc.) of input nodes. Each node in the network is usually an input station or a work station where users actually speak into the voice system to record information (Figure 2.7). The data collected at each node are transferred to the host computer where they are stored along with the voice system application and report-generation software.

Software

Voice system software falls into two categories: operating systems and applications.

Operating System Software

Operating system software coordinates all the tasks running on a computer:

• A single-tasking disk operating system (DOS) is the simplest operating system for a voice system which can handle only one operation on at a time. If you use a PC, the operating system that runs your computer is most likely single-tasking.

• Multi-tasking (OS/2 and UNIX) operating systems are used in multi-user voice system networks to allow many users to input data simultaneously. They coordinate communication between the host computer and each of the nodes on the network.

• Custom operating system software can also be created to handle specialized needs.

Application Software

Application software is used to help solve a particular data collection problem. A voice application is a computer program that causes the voice system to interact with a user. Application and application building software is available in many forms:

• Vocabulary / grammar definition software which allows the user to specify the words that the voice system application will understand and to set up simple applications.

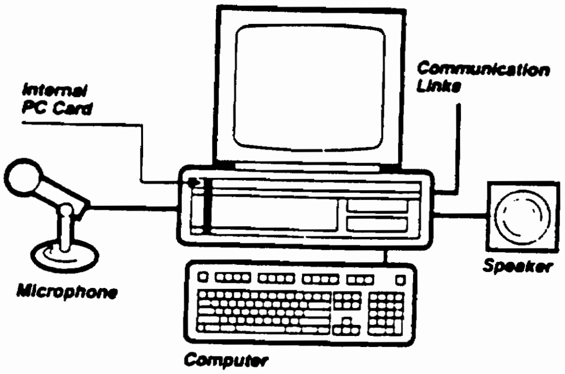


FIGURE 2.3: A PC card

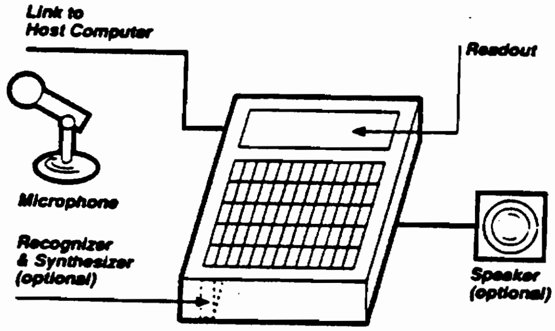


FIGURE 2.4: An input station with optional output

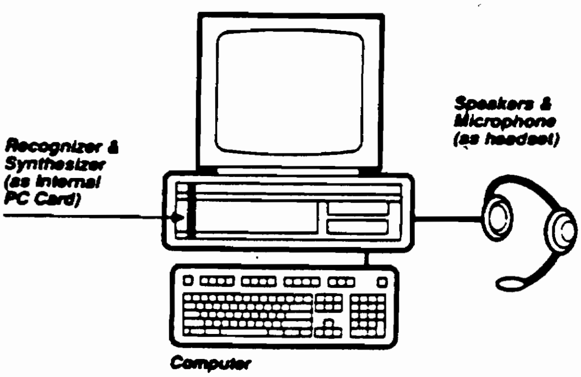


FIGURE 2.5: A typical workstation

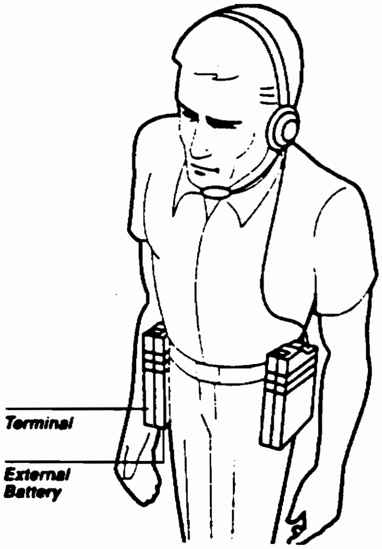


FIGURE 2.6: Voice data collection device

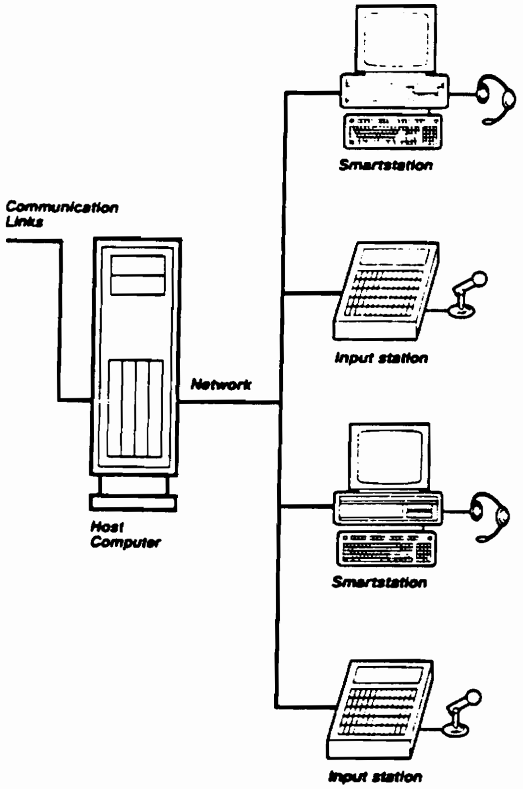


FIGURE 2.7: A multi-user voice system network

• Packages, or off-the-shelf, application software are geared toward solving a certain kind of problem, such as package handling, inspection, inventory control, and receiving / shipping. Off-the-shelf applications eliminate the need to spend time developing an application, but because they are designed to handle a general problem may not be specialized enough to fully solve the users data collection problems.

Application development toolkits are designed for flexibility and ease of use in creating sophisticated application solutions. Typically, this software will run on a variety of computer platforms and has a graphical user interface (eg. menus, graphics mouse commands). These user-friendly application development tools are an intermediate-level solution to the application development problem. They enable a company to build its own application rather simply and inexpensively, and also allows modification of the application whenever the need arises. With this type of application software, often the user can create the data collection process much like creating a flow chart. As long as a person can visualize and organize how the data collection process works, a person can build a task. This software can work for any process: inspection, material handling, or order picking. If more customized software is needed or it is expected that the voice application may have to be modified on short notice, this type of software is probably the best solution.

At the other end of the spectrum are custom applications prepared by system integrators. Such voice system software is developed exactly to a company's needs and preferences. Like any custom work, custom voice system applications are much more expensive than off-the-shelf applications. A custom application may also require reliance upon a systems integrator to perform any future modifications to the system.

2.1.2 System Components

The purpose of the previous section was to give a holistic impression of a voice recognition system. This section gives a breakdown of its general components. These include:

Headset with microphone- allows the operator to speak to the voice unit, and the voice unit to speak to the operator. This headset with microphone connects to the voice unit through an interconnect cable.

The voice unit (approximately 1 lb.)- worn on the operator's hip (using a belt), the voice unit is completely portable. The operator can collect data anywhere in the plant without restriction. The voice unit contains a CPU, internal battery (to prevent data loss in case the battery pack is accidentally disconnected), and an internal clock for date/time stamps. The voice unit also contains both the voice recognition and speech synthesis (not contained at the PC). The voice unit also has the ability to perform calculations.

Battery pack (approximately 1 lb.)- powers the voice unit. The battery is also worn on the operator's hip using a belt.

Optical Communications Pod (OCP)- is a method that allows data transfer between the PC and the voice units.

Radio option (optional)- includes a base radio unit that connects to the PC through a serial port and portable radio units that connect to the voice units through an interconnect cable. With this option, data can be sent in real-time from the operators throughout the factory to the PC (through the base radio) to provide real-time data collection.

Support software- as described above this includes the voice unit operating software, the allocations software (used to control the voice dialogue), and the data communications software that controls the transfer of data between the voice units and the host computer.

2.2.3 Other Voice Unit Features

The list below gives additional features of a voice system that can possibly be added if it is not a standard feature (depends on the system). When purchasing a voice recognition system if one or more of these features is deemed to be a necessity, then this must be kept in mind when choosing a system.

Talk-ahead capabilities- if the operator knows the answers to the next few questions, the person can speak these answers without waiting for the questions. The unit recognizes these responses and will not delay the process by continuing to ask all of these questions after the operator has already answered them.

Echo- the unit by default echoes all the operator's responses so the operator can verify that the unit correctly heard him.

Help features- by saying the word "help," the unit will speak to the operator all the valid responses to the current data collection question, or provide other assistance as you specify in the application building software.

Audio menus- by pressing and holding the control button on the voice unit, you can do the following:

1.) switch the current operator using the voice unit (multiple operators can be stored in one unit- used with multiple shift operations).
2.) change the volume of the voice unit.
3.) retrain selected words in the task (if the operator is having recognition problems for a given word).
4.) can be used in conjunction with other data collection technologies like bar-coding and radio frequency.

2.2 Advantages and Disadvantages of Voice Technology

Advantages

In collecting data, VRS competes with methods such as paper and pencil, bar code scanning, and keyboard / keypad data entry. However, VRS offers many advantages over these other systems.

Bar-coding is an excellent technology to read a tag and record that a particular item is in a particular location at a particular time. The limitation of bar-coding, however, is that it is unable to record human judgment and cannot be used in applications where hands and eyes must be free to collect data. Hand held terminals, bar-code with key pads, or bar-code with pencil pads, are excellent portable terminals but also have limited use when the operators hands and eyes must focus on a process.

Voice data collection is the best technology when there is a need to record human judgment and it is particularly useful when the operator's eyes and hands are busy, which is the case in many applications. Voice provides greater productivity because more data can be collected quickly. For example, most people easily speak at rates of 200 words per minute, yet few can type better than 60 words per minute. Once spoken, the data can be transferred directly to a computer for processing. Management now has immediate availability of information (real-time system). Voice is easy to use and requires limited operator training. Many operators are trained in less than a day and no keyboard skills are required.

Other instances where it is advantageous to use voice recognition systems include:

1.) where the application requires confirmation of information in a step-by-step manner in order to continue a process or to revise instructions,

2.) where data consists of more than just numbers,

3.) where information being gathered is about an item rather than printed on it, and

4.) workers are on the move on foot or in vehicles.

Disadvantages

When voice recognition and synthesis technology first became available to industry, many companies found it severely lacking in its capabilities. This has changed to the point that now many companies using voice-based data collection systems will not allow this fact to be publicized, considering it as major competitive advantage. Voice technology still has its limitations, however:

1.) Input must be consistently spoken in order to reduce recognition errors. User training is often required to operate these systems successfully,

2.) Large vocabularies can be cumbersome to use due to the processing time needed for recognition, and

3.) Vocabulary size must often be sacrificed to achieve recognition in speaker-independent systems or continuous speech systems.

2.3 Difference in VRS Systems

2.3.1 Types of Voice Systems

Voice systems differ due to several different parameters. These will be described in the sections that follow. In general there are four types of voice systems.

* ***Voice recognition-*** This is a system where a user speaks into a microphone and the processor recognizes the words and turns them into output data that causes an action or is sent to a host computer.

* ***Voice recognition and voice response-*** This system will also give a voice response to the information if appropriate or prompt the user for the next entry.

* ***Voice response-*** The user may use a keypad to enter information or a touch-tone phone pad. A dial-up could give the user a menu of responses or data types available. Additional keying narrows the data to that required. An example of *voice response* is when you call a telephone information service. When you call for information, the operator asks the city and name, and then calls up the appropriate directory page on a screen. Finding the name requested, a code is activated and a synthesized voice then gives you the phone number. The operator goes on to the next call while the computer is giving you the number.

* ***Voice inquiry response***. This type of system is also activated by using a keyboard or keypad on a telephone. The user has a coded menu to select information. There is no other interaction between the user and the system. An example of *voice inquiry response* would be when a person queries his / her bank balance from a 24-hour automated service line or dial an informational 900 number offering a touch-tone menu selection.

2.3.2 Speaker Dependent vs. Speaker Independent Systems

Some systems are differentiated by whether they are speaker dependent or a speaker independent system. General definitions and the differences between the two types of systems are given below.

Speaker dependent- means that every worker must speak the appropriate task words and phrases into the system so the system can record how the speaker's words sound in order to recognize them later. These words are stored in a look-up table.

Speaker independent systems- can understand the majority of users without pre-training the system. Some systems even recognize foreign accents. Speaker independent systems are more versatile, but for both accuracy and security, the speaker dependent systems make up the majority of installed systems.

2.3.3 Discrete, Connected, and Continuous Voice Input

Systems can be differentiated by how the voice data is inputted. They include discrete, connected, and continuous voice input.

Discrete voice input- require the user to say one word at a time and wait for the system to recognize it. Recognition can be verified by the computer using synthesized speech or pre-recorded / programmed responses such as a simple beep. Also, many systems have a redundant color coding feature, for instance, a green light would signify "recognition" where a red light would signify "non-recognition." Discrete word recognition systems recognizes a series of spoken words when more than 250 milliseconds of silence separates each word.

Connected voice input- requires users to speak in pre-defined word strings or phrases. Words out of sequence will not be recognized. This extends the capability and versatility of the VRS beyond the discrete word systems but sets up exact speaking rules. Connected voice input systems recognizes a series of spoken words when at least 50 but not more than 250 milliseconds of silence separates each word.

Continuous voice input systems- recognize multiple words in any order without requiring that they be in strings or phrases. This can be done without training the worker or retraining the system in most cases. This feature is desirable for a VRS used in a factory environment. Continuous voice input systems recognizes a series of spoken words when less than 50 milliseconds of silence separates each word.

Problems can occur in interpreting input data using either connected or continuous voice input systems due to co-articulation. This is explained in the next section.

2.3.4 Co-articulation

What makes *connected and continuous voice input systems* so difficult to interpret is co-articulation. This is when speaking a pair of words such as "seven nine" or "test tube" people will generally omit the consonant which starts the second word, and will say something like "seven'ine" or "test'ube." This phenomenon is called co-articulation and while it was developed to make sentences easier to say it makes it very difficult for developers of voice recognition systems. (Edgar, p.576)

2.4 General Steps to VRS Data Recognition

The general steps (see Figure 2.8) a VRS goes through to record data includes:

***Step 1*:** The computer asks questions using speech synthesis.

***Step 2*:** Spoken data is inputted through a microphone.

***Step 3*:** A signal processor converts spoken data into digital form (A/D converted)

***Step 4*:** Spoken data is stored using speech-to-text synthesis.

***Step 5*:** A voice synthesizer, which is a combination of hardware and software, converts text (digital data) to speech (analog data) using speech synthesis.

***Step 6*:** A speaker or headset broadcasts the output. This is done for one of three reasons:

1.) To insure that the inputted data has been correctly interpreted.

2.) To give error messages.

3.) To give prompts for questions (back to Step 1).

2.4.1 Asking Questions

The computer uses speech synthesis (described in a later section) to produce human sounds in the form of a question. There are two basic techniques to have the computer ask questions:

1.) A human could pre-record each question and have the system playback the question in the form of a prompt (known as record and playback).

2.) The second way is to type each question and use speech synthesis to convert the text to speech.

Speech synthesis is the most common technique being used today because it offers maximum flexibility and is easy to implement.

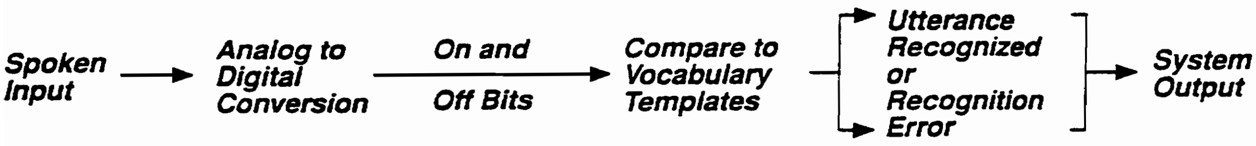


FIGURE 2.8: Steps in the voice recognition process

2.4.2 Speech Input and A/D Conversion

All voice recognition systems operate similarly. The person speaks into a microphone; by the Nyquist Limit (a theorem that states that to accurately capture an analog signal you need a sampling rate of twice the frequency), a sample rate of 8 kHz to 10 kHz is used to capture a number of different frequencies that comprise speech. Once the VRS has digitized the spoken words, its software applies a reverse Fourier transform to the digitized data, mapping the frequencies in the digitized speech to discrete ranges, or vectors. These vectors represent phonemes, the basic sounds such as "uh" or "ee"- that make up words (described in section 2.4.4) (Labriola, p.144).

2.4.3 Speech-to-Text Synthesis

Speech recognition software operates on the digitized speech signal to perform one of two tasks: convert spoken words to text (speech-to-text) or perform a spoken command (text-to-speech). After A/D conversion, the spectral representation of each basic speech sound (each phoneme) is shown as a characteristic form that shows up in the spectrum. Then, the spectrum of the incoming speech is compared to a library of word models, which are acoustic models of how the word sounds. This is done in order to see which is the best match.

Acoustic models usually show how the spectrum of the speech varies over the duration of the word. The simplest acoustic model describes a single typical way in which a word is pronounced. This template represents a frequency time picture of the word. An unknown spoken word can then be identified by comparing it to templates of all the words in the vocabulary and finding the best match.

Dynamic Time Warping (DTW)

A word varies slightly each time you pronounce it (eg. speaking a little quicker or slower). In an attempt to handle this variation, the first generation of speech-recognition software mathematically distorted the time axis of each template until it best fit the unknown speech, a technique called Dynamic Time Warping, or DTW.

DTW proved too limited an approach for larger vocabularies, continuous speech, and speaker independence. A single template could not describe the full variability of pronunciations.

Language Model

Another method that has helped improve recognition accuracy is a language model. The simplest form of a language model is a list of words or phrases that you can speak legitimately in the current context. If the system asks how many copies to print, it is expecting a number. If the system asks for verification of a command to delete a file, it is expecting a "yes" or a "no."

A more complex language model assigns a likelihood to word sequences. A language model can indicate that "they read" is more likely than "they red," allowing the speech-recognition algorithm to use word context to improve accuracy

2.4.4 Text-to-Speech Synthesis

All text-to-speech systems initially transform text input into a sequence of sound symbols, usually phonemes, diphones, or demi-syllables. The complexity of English requires between 500 and 1,000 mapping rules to derive most pronunciations. Some (limited) success has been achieved in using neural networks to model English pronunciations based on large numbers of examples, but commercial systems generally use a rule-based approach. Even then, high-quality systems include an exception dictionary to cover "anomalous" pronunciations. Still other systems include special rules to pronounce loan words from other languages. To further increase the quality of text-to-speech synthesis, limited syntactic analysis is used to determine sentence structure and augment the string of sound symbols with pitch and duration information.

There are two main uses of speech synthesis. It is used to reply to the speaker to confirm information or to provide help. Generally, most systems have a synthesized speech capability that is used to lead a user through a sequence or program (asking questions), or reply to spoken input by saying words such as "say again," "yes," "no," or "next." It may ask the worker to repeat the last spoken input as a confirmation. Text-to-speech capability also allows the host processor to generate messages that can be spoken to the operators or call on pre-programmed messages as required. The next section describes two synthesis techniques:
1.) mathematical modeling (formant synthesis) and 2.) synthesis by concatenation.

Synthesis techniques

Two synthesis techniques are in common use: mathematical modeling of the waveform generated by the human speech production apparatus (formant synthesis) and splicing pre-recorded segments of speech (synthesis by concatenation).

Formant synthesizers use the sound symbols to define a sequence of acoustic targets, then interpolate the acoustic signal between these targets to mimic the dynamics of the human voice.

Concatenation-based systems achieve the same effect by selecting a sequence of pre-recorded elements corresponding to the sound symbols in context and then smoothing the junctures between these elements. Since they are based on the actual human voice, concatenation systems tend to produce a richer and more natural-sounding signal, but requires more storage space than formant synthesizers (Rudnicky, p.55).

Concatenation-based systems can be further categorized into two groups: time domain and frequency domain systems. Time domain systems store and concatenate the individual elements using the amplitude of each sample directly, while frequency domain systems transform the signal into a spectral representation which allows easier manipulation of pitch and duration, but reduces the quality of the outputted voice.

Error rates in speech synthesis

While intelligibility is quite high with error rates as low as 3.25% for individual word perception (compared to 0.53% on natural speech) on a standardized test, existing synthesizers still do not sound natural. Text-to-speech systems are noticeably lacking in the quality of intonational characteristics or "prosody." Prosody encompasses the timing, intensity and pitch, as well as some of the co-articulation effects (described above) that change the way words sound when spoken together instead of individually.

This fairly low error rate can be misleading. The error rate for the perception of complete sentences is also over five times as large as that of natural human speech (4.7% synthesized vs. 0.8% for natural speech). The experiment from which these results were taken was conducted under controlled laboratory settings. Therefore, it is safe to assume that use in a real-world setting would provide higher error rates. Progress in synthesizer technology appears to have reached an asymptote with "successively small improvements seemingly difficult to achieve" (Rudinsky p.55).

Types of recognition errors

Some typical errors that may be detected by the voice recognition system include:

1.) person begins speaking too soon or too late,
2.) person speaks too loudly or too softly,
3.) person does not speak at all (the time period to speak has run out),
4.) person speaks fewer than the required number of responses (eg. speaks four instead of five digits for a US zip code),
5.) person speaks "fillers," such as "Yes, five" or "I want five" instead of "five" (does not have word spotting capability, which is described in the next section),
6.) spoken utterance is close to more than one word, such as "S" and "F," "Austin" and "Boston," "five" and "nine" etc., and
7.) person mumbles as he / she speaks

(examples taken from Edgar, p.289).

Phonomes and Diphones

All text-to-speech systems initially transform text input into a sequence of sound symbols, usually either phonemes (see Appendix I for a listing of phonemes) or diphones.

Phonemes- the smallest units of identifiably different sounds. Webster's Collegiate Dictionary provides a phonetic spelling of every word: "voice," for example, is phonetically spelled \'vóis\. A phonetic alphabet is used to describe in detail how the word is spoken. For instance, hyphens are used to separate syllables. Since "voice" is a single syllable, no hyphens are used. The high set mark ' indicates that the stress is at the beginning of the word (you say "VOiss," not "voISS," to use an informal phonetic spelling). The various vowel and consonant sounds are assigned specific marks. The "s" in \'vóis\, for example, represents the hard "s" as in "kiss" rather than the soft "s" as in "noise" (Edgar, p.312).

Disyllables or diophones- breaking down sounds even further, for example, pairs of sounds such as \vo\ and \is\, which might themselves be concatenated to make \vóis\.

The phonemes, diphones, or demi-syllables are put together in various ways by the software to create words and sentences to query, prompt, or reply to verbal or tonal input.

In addition, the complexity of English requires between 500 and 1,000 mapping rules to derive most word pronunciations.

2.5 Vocabulary

The vocabulary of a VRS is the set of utterances (words) which can be recognized. The following are some points that need to taken into consideration when choosing a voice recognition system including vocabulary features and vocabulary size:

2.5.1 Vocabulary Features

These are some features of a voice recognition systems vocabulary.

The words in the vocabulary

This is an obvious point in that if you want your system to recognize digits (numbers) along with letters, the selected vocabulary must have this capability.

Sub vocabularies

There are two different types of sub vocabularies: 1) trained (recognizes correct words only) sub vocabularies and 2) trained sub vocabularies (recognizes correct words and has the capability to reject other utterances). If you have a digit vocabulary, but only want to recognize "one," "two," and "three," you can ask the recognizer to restrict its matches to these options. However, a sub vocabulary trained on samples "one," "two," "three" and specifically trained to reject other digits and other random utterances will achieve better results. A particularly important example is "Yes" / "No" vocabulary, which should be as reliable and effective as possible, since this capability plays a key role in most systems. Specific "Yes" / "No" sub vocabularies may be provided as options as well as larger sub vocabularies which may include "Yes" and "No" in addition to other words ("Help," "Cancel," "Stop" etc.).

Exception Dictionary Look-up

Dictionaries are used to isolate exceptions to common pronunciation rules. The exact stage where the dictionary is used depends on the detailed strategy used by the generator. One example is how the generator interprets the English word "off," which is pronounced with a "v" rather with an "f" sound. This exception in English has a relatively large number of exceptions to spelling pronunciation rules, especially when proper names are pronounced.

Range of accents and dialects

There is a trade-off between accuracy and specific targeting of particular accents or dialects. For example, a US English vocabulary might produce acceptable recognition percentages and accuracy for any native born American speaker. However, a vocabulary targeted on New York (speech created from samples collected only from native New Yorkers), for example, will be more accurate if the target group of users of the application is based only in New York. This is an important consideration that should be brought up with a vendor about their strategy used to train the vocabulary when purchasing voice recognition equipment. The tolerance of accents and dialects not used in the building of the vocabulary will also vary depending on the voice recognition technology (for example-speaker dependent or independent system).

Ambient Noise

Different systems display different tolerances to echo, static, background noise, poor microphones, etc. that can greatly effect the accuracy of the system. This is an important consideration that needs to be taken into account especially, for instance, if the system needs to be used in the plant floor environment.

Voice Cut-Through

This refers to a caller's ability to interrupt a menu prompt by speaking before the prompt plays to completion. This may come in two variants, a simple "barge" where the user says anything, the prompt stops playing, a beep is played and the user then speaks a selection, or a "true" cut-through where the spoken selection itself can interrupt the prompt. True cut-through requires more signal processing memory than barge cut-through.

Word Spotting

This is the process of picking out certain key words from responses. For example, the word "No" may be picked out from "No, thank you," or the number "five" can be picked out from "send me five tickets please." This feature allows the speaker both the ease of using a continuous speaking format and the speed of a discrete system.

Ability to change vocabularies in an application

An application may require several different vocabularies. Depending on the memory capacity of the voice recognition board and other factors, there may be limits on the number of active vocabularies in the system at any one time.

Adaptive and learning vocabularies

It may be possible for an application to collect voice samples from the live usage of the system. These samples can later be added to the vocabulary, allowing the system to adapt to its users in a continual process of improvement.

2.5.2 Vocabulary Size

Vocabulary size could also be a consideration in selecting a system. In the article "Straight Talk" (Labriola, p.148) the author states that systems can now recognize from 10,000 to 35,000 words. The larger the vocabulary, usually the slower the response time. Most industrial applications have a 100 to 500 word vocabulary. The word list is seldom fully accessed. A vocabulary of 50 to 100 words is large enough for most applications. Some suppliers say system accuracy significantly degrades with larger vocabularies. Vocabulary size will vary by VRS manufacturer and influence which system is appropriate for the application.

2.6 Data Collection Modes

In order to install any data collection system, one must consider the three most common data collection modes. They are (1) portable batch mode, (2) portable real-time mode, and (3) fixed real-time mode. The portable batch mode consists of a data collection terminal in which the collected data is stored in the terminal for later uploading to another computer for processing. The portable real time mode takes the collected data and transmits it to the base computer by means of a FM radio transmission. The real-time fixed mode consists of a data collection terminal hard-wired to the data processing computer. Each mode has its advantages and disadvantages based on the data collection environment.

2.7 Guidelines for Choosing a Voice System

Before jumping into a project using voice technology to collect data, one should establish the specifications of the application. Consideration should be given to the following issues.

Vocabulary size- the size of the vocabulary will influence which vendor's technology is appropriate for the application. Ranges tend to be grouped into categories of <200, <1,000, <10,000, and >10,000 words

Ambient noise- the background noise will determine which microphone/headset combination is best for the application. Headsets range from lightweight, inexpensive models to very heavy models with a better ability to filter out loud noise. There is also the issue of operator comfort with the headset.

Discrete vs. continuous recognition- discrete recognition consists of a one word at a time pace of data entry which is not natural to most people. Continuous recognition allows the operator to input data in a natural speaking rate which most people are more comfortable with.

Collection mode- will the operator be seated or walking around? Is the data required in real time or could it be uploaded at a later time? The answer to these questions will help determine the collection mode: either portable batch, portable real time or fixed workstations in real time.

Ease of set-up and use- who will be using the system and how computer literate will the operators be? The voice system should be considered a tool for the operator and management. It should help a process and not slow down the operators or managers.

Software development- management should be determined who will write and maintain the software. What format will it take? Will the voice vendor or a third party reseller develop the application and the output files.

Mobility / Portability- does the application require that voice system operators use portable and / or radio units to collect data? If so, will the wireless communication technology used in the voice system (usually RF / DC) function properly in the data collection environment and for what distance?

Interfaces to other systems- with what other system does the voice system need to interface? Does the voice system come with interface(s) that meet the necessary application needs? How easy is it to use the interface(s)?

Maintenance- are service and support for the voice system available from the vendor or will these services be provides in-house.

Cost justification and return-on-investment (ROI)- How much will the voice system cost? What additional expenses will be incurred due to development of the application and the training of users? What quantitative benefits can you expect as a result of using voice-based data collection? How long will it take to see the ROI?

Once these issues are determined, the voice vendors should be contacted to review the hardware and software available. The issue of standard vs. custom software should be addressed and who will maintain the system over time.

2.8 Applications

This section describes five examples of areas in which voice technology are being applied: quality control inspection, materials handling, laboratory settings, manufacturing, and in multilingual banking:

Quality control in the automotive industry was one of the first to successfully apply voice data collection. The reasons are simple. They have a large object to inspect, a lot of data to be collected, the item being inspected is of high value, and it is moving on a production line. People have to use their hands to touch the cars (open doors, windows, insert keys, unscrew various items, etc.), place stickers on the cars, and at the same time record what they find. Voice technology is a natural for this application (Hemphill, p.5).

Textile manufacturers have been using voice data collection for years. Their application consists of inspecting the bolts of material as they come off the loom. A trained inspector has to look for defects such as pulled strings, discoloration, uneven weave, sags, etc. Once a defect is noted and the exact coordinates are determined, the operator can continue to look for more defects. Since the material is moving, the inspector must be fast in spotting and recording the defects. An inspected bolt of material will then move to a cutting operation in which cutting is done around the defects. The more precise the inspection, the greater the yield, and the less scrap and rework (Soltis, p.55).

Laboratory data collection is another area where data is being collected by voice in a quality control application. Here is a situation where the inspector is looking into a microscope, has total concentration of his eyes on the object being inspected, needs his hands to position the object and adjust the scope due to various depth of field focusing, and at the same time be recording the findings. Without voice technology, the operator must take his eyes and hands away from the microscope to make a notation of a condition found. Once the notation is complete, the operator must go back to the microscope, reposition his eyes, hands, the microscope, and the object in order to continue the process. With a voice data collection system, the operator simply speaks the data as the observation continues. Voice data collection provides a great improvement in operator efficiency. Less errors are made, the data is immediately available for analysis, and the operator experiences less eye strain and fatigue (Stovicek, p.27).

Manufacturing Application. During the manufacture of printed circuit boards and semiconductor wafers, each chip or circuit is inspected for defects through a high power microscope. The inspector notes any defects along with brief comments as to their nature and likely source. Using voice recognition saves the labor costs of transcription of notes, and allows the operator to maintain peak visual concentration, thus improving throughput (Stovicek, p.26).

Multilingual banking. A real-time spoken language translator, the Voice English/Spanish Translator (VEST), for example uses a text-to-speech synthesis system recognizes about 450 words, determines the language, and "speaks" the translated sentence in less than one second. This system was developed by scientists at AT&T Bell Labs, Yorktown Heights, NY, and Telefonica Investigacion y Desarollo, Madrid, Spain, the system simulates a banking transaction or currency exchange. The limited vocabulary is modeled on a language guide for tourists. Translation is done by an AT&T BT-100 machine that uses 127 parallel digital signal processors (DSPs) capable of one billion floating point operations per second (Haskin, p.54).

2.9 Case Studies

While the previous section described general areas where voice technology is being applied, this section was included to give specific examples of how this technology is being used in industry today.

2.9.1 Saturn Paint Quality Inspection by Voice

Paint inspection team members constantly use their hands and eyes to collect data. They use their hands to feel the quality of the paint or primer and to make minor repairs when possible (for example, sanding, polishing, or buffing). They use their eyes to visually inspect the cars panels. For these reasons Saturn wanted to find a data collection method that matched their needs.

The data collection technology options were: traditional pencil and clipboard methods, bar-code readers and hand-held terminals, and voice data collection.

After ruling out the traditional pencil and paper data since it was too time consuming and an error prone process, they looked at bar-coding and voice systems. It was decided that the bar-code was a good system for automatic identification, however, it is not good for recording human judgment, specifically, defect types and locations. The only way bar-code readers would work to record defects would be to have a book containing bar-code labels (many clothing stores utilize this methodology) for all the different codes and locations. This solution was much too time consuming. For applications needed by the Saturn team it was deemed that voice recognition technology would work the best. With voice units, the operator can enter the collected data by voice while using their hands and eyes to inspect the cars and perform minor repairs (Hemphill, p.5).

Saturn Paint Quality Inspection by Voice: Sample Voice Dialogue

This is an example of the dialogue between the operator and a voice recognition system. The unit asks the operator questions and the operator provides responses. The terminal has the capability to ask different questions (branching) on what operator response is given.

A sample dialogue is as follows:

Question	**Answer**
Operator Number?	29293933
Side of line you're on?	East
Buck Number?	3930399
Color of Panel?	Saturn Blue
Any defects for this panel?	Yes?
How many?	1
What is the defect?	Scratch?
Where is the defect?	Top left?
Disposition?	Good (means the operator was able to fix the problem)
Buck Number?	.....(the process is repeated)

(Hemphill, p.6)

2.9.2 Military Applications (USAF)

A technique that can automatically identify the voices of different speakers or determine what language each is using is under development by United States Air Force in its Rome (N.Y.) Laboratory. This work shows promise for a variety of military and civil applications (Klass, p.57).

During military operations when the communications are jammed with friendly as well as enemy voice communications, the new techniques could be used to automatically select and switch enemy radio operators to appropriate communications intelligence (Comint) monitors.

In a test conducted by USAF using nine different languages, it was possible to automatically identify each speaker's language with an accuracy of about 85%. The next step in this application is to improve its accuracy under noisy conditions. Accuracy with more typical noisy radio-type samples, with a signal-noise ratio of 10 dB, dropped the accuracy to about 45% (Klass, p.57)

2.9.3 Auto Auctions

With 28 U.S. sites and 25 in the United Kingdom, Nashville-based ADT Automotive sells more than a million cars yearly. Previously, the company checked in merchandise as it arrived via hard-wired headsets that allowed workers on the lot to verbally relay a car's vehicle identification number (VIN), odometer reading, and color to a data-entry person inside the office. However, this made labor costs very expensive necessitating a two-person operation. Then the company switched to letting employees write check-in data on clipboards, which was to later be key entered by someone else into a computer. This worked a little better, however, there were a lot of errors especially with the 17-character alphanumeric strings that make up VIN numbers, which auto makers have been bar coding (in Code 39) since 1992.

For these reasons, ADT Automotive moved to a voice recognition system. Their system includes a headset with an earpiece and a noise-canceling microphone, plus a belt that holds a CSL4000 VoiceLink radio and PSC 5317 bar code scanner capable of reading labels through windshields.

Now when a car is delivered, a worker typically scans in its VIN number or voice enters the information if made before 1992, then tells the system its color, mileage and options, along with the name of its seller, and if applicable, a "floor price" beneath which the auctioneer is not to sell the vehicle. In addition, key words recognized by the system allow ADT personnel to enter remarks about vehicle damage, odometers that have flipped past their mechanical limits, etc. All information is spoken back to the worker for verification.

While ADT bought its system more to improve its overall efficiency than to save money, voice recognition technology has provided ADT with an estimated annual savings of one million dollars. ADT's biggest savings occurred in the area of labor reduction, needing only one-half the number of people required to check in cars than before (Jesitus, p.18).

Chapter 3: REVIEW OF ADCSL VOICE RECOGNITION SYSTEMS AND THE PROJECT LIFE CYCLE

3.1 The ADCSL- an Overview

The Automatic Data Collection Systems Lab consists of three automatic identification and data collection technologies:

1.) Bar Code Systems,
2.) Radio Frequency Data Collection Systems, and
3.) Voice Recognition Systems.

The portion of the laboratory that is designated for this project consists of the Voice Recognition Data Collection Systems. The purpose of this chapter is to discuss lessons learned about voice recognition systems and to provides an analysis of how they must be improved before wide spread usage becomes the norm. This is followed by a discussion of the hardware and software available in the ADCSL.

The topics discussed in this chapter include:

1.) Current Applications of Voice Recognition Systems,
2.) Future Applications of Voice Recognition Systems,
3.) Current Problems with Voice Recognition Systems, and
4.) A review of the hardware and software available in the lab for developing a voice recognition system.

3.2 Current applications of voice recognition systems

Voice recognition systems are currently being used for:

1.) word processing,
2.) inspection / quality control,
3.) automatic data collection,
4.) language translation, and
5.) basic commands for manipulating computer operating system shells.

By far the largest application for voice recognition systems is quality control inspection (see a case example on Saturn, Chapter 2.9.1). Users of voice recognition systems for quality control include the automotive industry, textile manufacturers, and the military. These groups use voice recognition systems mainly because of the need for the operators hands and eyes to be free during data collection.

3.3 Future applications of voice recognition systems

Future applications for voice recognition systems may include:

1.) ***security***- a person will be able to open doors or enter a restricted area by recognition of his/her specific voice pattern (shown in popular science fiction programs, such as Star Trek and Tech War). The system will be able to collect many different types of data, including who entered the system, at what time, for what length of time, etc.

2) ***commands for manipulating computer operating system shells***- someday there will be a voice shell where a computer can be turned on and off, commands issued, and memos / data entered, all by a person's voice. As shown by the popular movie, "Disclosure," this may be used in conjunction with virtual reality to create the ultimate computer system.

3) ***word processing***- currently there is not a voice recognition system on the market where a person can install a voice recognition system and immediately type a memo by voice. Typing memos by voice could become the norm in ten to fifteen years.

4) ***tailored word processing packages***- for special groups like doctors and lawyers this system would have all the features described above and have the ability to process terminology specific to their interest groups (eg. medicine, law, etc.).

3.4 Current Problems with Voice Recognition Systems

It will be interesting to see if any or all of these applications develop in the near future. However, before these voice recognition systems can become "reality," current problems with voice recognition systems need to be addressed and solutions found to these problems. Current problems with voice recognition systems includes:

1.) ***range of accents and dialects.*** How will a system react to a New Yorker's accent in comparison with someone from Mississippi? Will there be separate software packages depending on what part of the country you are from?

2.) ***continuous voice input.*** There is currently no system that allows a person to accurately speak in a continuous manner.

3.) ***co-articulation.*** What makes continuous voice input systems so difficult to interpret is co-articulation. This deals with speaking a pair of words and having the words blend together (eg. "test'tub" instead of test tube, see Chapter 2.3.4 for more details). Somehow this problem will need to be overcome.

4.) ***word spotting.*** This is the process of picking out certain key words from responses when a person speaks "fillers," such as "Yes, five" or "I want five" instead of "five." To be successful voice systems must allow the user to speak fillers and allow words to be spoken in any order.

5.) ***cost justification and return-on-investment.*** As voice recognition systems become more widely used, the costs will fall to a more reasonable level. Currently, systems cost between $1,000 for voice dictation software up to $25,000 for an entire voice recognition system.

3.5 Technical Resources

All of the equipment for the voice recognition system has been obtained from VOICE CONNEXION based in Irvine, California. The following sections give a brief description of the available software and hardware:

3.5.1 Hardware

IntroVoice VI Voice I/O with headset microphone (IBM and Compatible Voice Subsystem)

IntroVoice VI is a complete voice input / output system. The minimum requirement is that the system is run on at least an IBM PC/XT/AT, PS / 2 Models 25 and 30 or compatible. The system is supplied with an IntroVoice VI circuit board, microphone, speaker, and utility software.

Any program under MS / PC DOS may be operated by voice or keyboard through the supplied memory resident Voice Executive software (VEXEC). The IntroVoice VI allows PC users to input and retrieve information as well as command the operation of application programs and system functions by voice input and output. IntroVoice VI is supposed to increase productivity and reduce errors by allowing on-line data entry, especially where the hands and/or eyes are busy.

Micro IntroVoice modular voice I/O system with built-in microprocessor and serial interface and headset microphone

Micro IntroVoice (MIV) control program facilitates operation of the Micro IntroVoice through the PC's RS-232 asynchronous communications port. This program was designed to make it easy to take full advantage of the voice recognition and robotic text-to-speech synthesis features of the Micro IntroVoice. The MIV control program may be operated under Windows as a DOS application.

PTVC-756 Portable Transaction Voice Computer with 1 Mg (includes MS DOS)

The PTVC-756 is a portable transaction voice computer that is designed to handle most data collection, analysis, and communication applications. This hand held system has the features of an IBM desktop, but also has the added versatility of voice recognition and synthesis and bar coding.

The data acquisition device features voice recognition with 500 words per user and unlimited text-to-speech voice output for prompting and verification. It also has a high contrast 16 line by 21 character display, a built-in serial port, and up to one megabyte of RAM memory.

3.5.2 Support Software

WinVoice

WinVoice is a program which allows Windows, Windows Accessories and Games, application programs written for Windows (such as Word and Excel) and standard DOS application programs to utilize the voice recognition and synthesis capabilities of the Micro IntroVoice. WinVoice is a program provided by Voice Connexion which when recognition is activated, emulates the keyboard and mouse via voice control.

3.6 Project Life Cycle and Methodology

Phase 1: Recognition Phase

This first step was needed to become familiar with the voice recognition equipment in the lab. These tasks include reading the manuals, understanding how to use the equipment, and testing the capabilities of the equipment.

Phase 2: Network Configuration and Installation Phase:

The second phase involved defining the network architecture requirements needed to support the voice recognition system. The network was set-up. All the voice recognition equipment was interfaced to function as an automatic data collection system.

Phase 3: Laboratory experiments design

Everything learned from the entire project was incorporated into this phase. A set of laboratory experiments was designed to give students hands-on experience in using the voice recognition tools. The labs included the following tasks:

* voice training
* improving the accuracy of the system
* text to speech synthesis
* bar code scanning

3.7 Summary

To date, speech recognition has been largely confined to niche applications. However, in the future voice recognition will play a major role in many applications including banking, security, dictation, and data collection. It will be important that industrial engineers understand how this new technology can assist in making an operation more efficient.

This project was intended to delve into the voice recognition process and to provide students with hands-on exercises and a basic understanding of the methodology behind voice recognition technology.

We are still a long way from the capabilities of voice recognition shown in the movie *2001: A Space Odyssey's* (the movie featured a computer that was really good at voice recognition): "Open the pod bay doors, HAL!"; but we are getting much closer!

Chapter 4: MICRO INTROVOICE LAB

4.1 Introduction to the Micro IntroVoice Lab Activity

The purpose of this lab is to give students hands on experience in using voice recognition equipment, specifically the Micro IntroVoice. The lab includes training a vocabulary with the subjects voice patterns, using voice recognition in the DOS environment, testing the text-to-speech synthesis function, and improving the accuracy of the Micro IntroVoice by changing the microphone gain, retraining a single/pair of words, and looking at what would happen if the baud rate was changed.

After doing the lab activity students should understand the basics of voice recognition and have gained practical experience in using a voice recognition system. Students should recognize voice recognition's potential for future use in applications involving automatic data collection.

The next section describes how a vocabulary is created in the Micro IntroVoice followed by a section explaining what was done for this particular lab activity.

4.2 How to Create a Vocabulary

This section describes the process of vocabulary development and the elements of the vocabulary editor provided by the Micro IntroVoice. These steps were undertaken to create the lab exercise that is to be trained by the students in the Voice Recognition Lab Activity using the Micro IntroVoice.

4.2.1 Elements of a Vocabulary

Vocabularies define a set of words that are to be spoken into the microphone and recognized by the Micro IntroVoice. Vocabularies also define what the Micro IntroVoice does when it recognizes a specific word. The MIV Control Program provides all the tools needed to create, edit, load, save, and print a vocabulary. Each word in a vocabulary consists of three fields or elements, including the spoken word field, the key replacement field, and the next vocabulary field. Each of these fields are described below.

Spoken Word Field

The spoken word field contains the words or phrases to be spoken into the microphone to cause commands or data to be input to a program. These are the words that will be recognized by the Micro IntroVoice when it is in voice recognition mode. These are also the words or phrases that the user will be prompted to speak when the vocabulary is being trained with his/her particular voice patterns.

When the Micro IntroVoice recognizes a word or phrase, the characters and/or commands of the key replacement field associated with the recognized spoken word are sent to the computer and/or processed by the Micro IntroVoice. Also, the subvocabulary(s) specified in the next vocabulary field (if any) become the new active subvocabulary(s).

The first word in the spoken word list (normally "Begin"), is a dummy word that does not get trained, but the Micro IntroVoice behaves as if this word was spoken when recognition mode is first activated. That is, the data in the corresponding key replacement field is processed and the subvocabulary(s) specified in the corresponding next vocabulary field becomes the active subvocabulary(s). This allows for the inclusion of a start-up message or command when the Micro IntroVoice enters recognition mode and to specify the initial active subvocabulary (the next vocabulary field corresponding to "Begin" must specify at least one subvocabulary).

The spoken word list also includes subvocabulary labels. A spoken word becomes a subvocabulary label when it is placed within parenthesis (see Chapter 4.4.2 for an example). All words or phrases following a label until the next label (or the end of the list) are part of the subvocabulary named by that label. Large vocabularies are generally divided into subvocabularies and the spoken word list must contain at least one subvocabulary label.

Key Replacement Field

The key replacement field defines the response for the corresponding spoken word when that word is recognized by the Micro IntroVoice. The response can include data that is sent from the Micro IntroVoice to the users computer and/or commands that are processed by the Micro IntroVoice. Data includes regular text and non-printable keystrokes (such as F1 or BACKSPACE). Commands are special non-printable characters that provide access to things such as the Micro IntroVoice's text to speech synthesizer. The following table lists the available Micro IntroVoice commands, which keystrokes enter those commands, and a description of what each command does.

TABLE 4.1: Micro IntroVoice Commands

Command	Keystroke	Description
Again	F2, A	Causes the key replacement string of the most-recently recognized word to be repeated until another utterance is detected by the program.
Toggle Speech	F2, G	This command toggles the voice output enable state. If voice output is enabled, the Quit, Speak, Echo, and Voice commands work together to send text to the Micro IntroVoice's text to speech synthesizer. If voice output is disabled, these commands have no effect.
Rubout	F2, J	Causes the Micro IntroVoice to transmit n BACKSPACE characters (ASCII 8). Where n is the length of the command sent for the most recently recognized word. This has the effect of erasing the text of the last command in most programs. If this command is repeated with no other intervening command, a single BACKSPACE is transmitted.
Help Menu	F2, M	Causes the Micro IntroVoice to send the first 36 words in the active subvocabulary across the serial port. If you are using the VKEY program, this command will display a window with the list of active words, thereby aiding the user by displaying available commands.
Quit	F2, Q	This command works in combination with the Echo, Voice, and Speak commands. It indicates the end of the text that is to be placed into the speech synthesis buffer (which was started with the Echo or Voice command). Note that the text is not spoken until the Speak command is encountered.
Speak	F2, S	This command works in combination with the Echo, Voice, and Quit commands. It causes the text currently in the speech synthesis buffer to be spoken by the Micro IntroVoice's text to speech synthesizer
Echo	F2, T	Causes the Micro IntroVoice to place the text that follows this command into the speech synthesis buffer in addition to being transmitted across the serial port to the PC. This continues until either the Speak or Quit command is encountered.
Voice	F2, V	Causes the Micro IntroVoice to place the text that follows this command into the speech synthesis buffer only. This allows the spoken text to be completely independent of the data transmitted across the serial port to the PC. This continues until either the Speak or Quit command is encountered.

The key replacement field can contain up to 255 characters. The Micro IntroVoice reserves 40,000 bytes (a byte is equal to one character) for all the key replacement fields of the entire vocabulary.

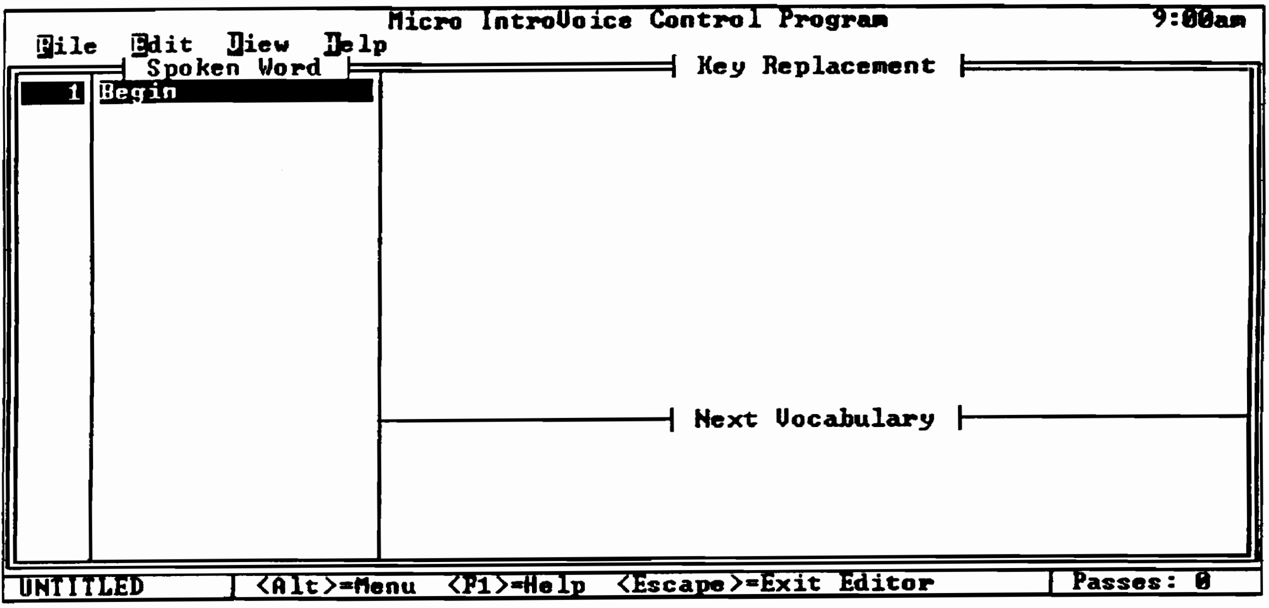
Next Vocabulary Field

The next vocabulary field contains the subvocabulary(s) that become active after the corresponding spoken word has been recognized. This feature allows limitation of the words that are "active" at any one time and changes the active subvocabulary based on which words are spoken. If the next vocabulary field is blank, then the subvocabulary(s) active when the corresponding word is recognized remains active.

When recognition is first activated, the active subvocabulary(s) are determined by the first word in the spoken word list (the first word must specify at least one defined subvocabulary). Vocabularies can be specified as a subvocabulary label that has been defined in the spoken word list, a word range such as 5-21 or 7-7, or the previous subvocabulary. The previous subvocabulary specifier indicates that the subvocabulary(s) active before the current one will again become active once the corresponding spoken word is recognized. If the next vocabulary field contains the previous subvocabulary specifier, it can contain no other data.

4.2.2 The Vocabulary Editor

The MIV Control Program features a vocabulary editor that allows the user to create and edit vocabulary files. The vocabulary editor is accessed when a new vocabulary is created with the File/New command (ALT, F, N) and when an existing vocabulary is edited with the File/Edit command (ALT, F, E). When creating a new vocabulary, the display appears as shown in the following illustration.



```
                              Micro IntroVoice Control Program                    9:00am
File  Edit  View  Help
        Spoken Word                          Key Replacement
    1  Begin

                                             Next Vocabulary

UNTITLED          <Alt>=Menu  <F1>=Help  <Escape>=Exit Editor          Passes: 0
```

FIGURE 4.1: Creating a New Vocabulary

The menu bar that runs along the second row of the display has now changed to contain only the commands related to the vocabulary editor. Commands are selected from the vocabulary editor menu in the same way they were selected from the main menu. There is detailed on line help available for all of the items in the vocabulary editor menu.

Basically, the vocabulary editor display is divided into three windows. Each of these windows corresponds to the three fields of the vocabulary as previously described. Note that the spoken word window lists as many spoken words as will fit in the window. In contrast, the key replacement and next vocabulary windows display only the information corresponding to the currently selected spoken word. The selected spoken word can be changed using the arrow keys to move the highlight bar up and down in the spoken word list. The cursor can be moved to different windows using the TAB and SHIFT+TAB keys.

When creating a new vocabulary, the vocabulary editor automatically inserts the word "Begin" at the start of the spoken word list. As previously explained, this word is not trained but does serve a special purpose. The information in the key replacement window that corresponds to the first word in the spoken word list is processed when the Micro IntroVoice first enters voice recognition mode. In addition, the vocabulary(s) specified in begin's next vocabulary window determines the initial active subvocabularies.

4.3 Vocabulary used in the Lab Activity

This section describes the vocabulary developed for the Voice Recognition Lab Activity using the Micro IntroVoice. The specific file is entitled EXAMPLE.RCS. Section 4.4.2 shows a listing of the file, which contains a sixteen word list to be trained by the students during the lab exercise. EXAMPLE.RCS also shows the key replacement and next vocabulary columns as described above.

Command numbers two through eleven were chosen because they are all basic commands used in DOS. This set includes the Date, Time, Path, Check Disk, Wide Directory, Directory, Clear Screen, Return, Path, and Set commands. Commands twelve through sixteen use special commands specific to the Micro IntroVoice program. The following table describes the special commands used in EXAMPLE.RCS.

TABLE 4.2: Special Commands used in EXAMPLE.RCS

Number	Spoken Word	Description
12	Welcome	Speaks the phrase, "Welcome to the Micro IntroVoice." Does nothing else.
13	Toggle Speech	Toggles the speech on and off. In the off position commands are still sent to the screen, but nothing is outputted through the speaker.
14	Help Menu	Displays the Help Menu
15	Stop Listening	Deactivates voice recognition. The only way to reactivate is to state the "Attention Computer" command.
16	Attention Computer	Reactivates voice recognition.

4.3.1 Why these specific commands were selected

These particular vocabulary words were selected for several reasons. One reason is that there are a limited set of DOS commands to select from. Thus, most of the standard commands are included. Another reason the set of fifteen words was chosen is that the recognition accuracy gets worse as the number of words that the program must differentiate among increases. This set of words was also selected because it contains few similar sounding words, which helps to improve accuracy. One exception that was left intentionally is the wide directory and directory commands. Lastly, time was another factor that was taken into account. It takes between eight to twelve minutes to train fifteen words and approximately one hour and fifteen minutes to do the entire lab activity. A larger word set would make the lab activity extremely long.

4.4 Micro IntroVoice Lab Activity

The following section contains the elements of the lab developed for the Micro IntroVoice. This includes:

4.4.1 Basics in using the Micro IntroVoice (this is to be read before attending the lab session)

4.4.2 Listing of file EXAMPLE.RCS

4.4.3 Voice Recognition Lab Activity for the Micro IntroVoice

I. Training a vocabulary and implementing voice recognition

II. Testing the text to speech synthesis function of the Micro IntroVoice

4.4.4 Micro IntroVoice Lab Write-Up

4.4.5 Keyboard Reference

General Operations

Help

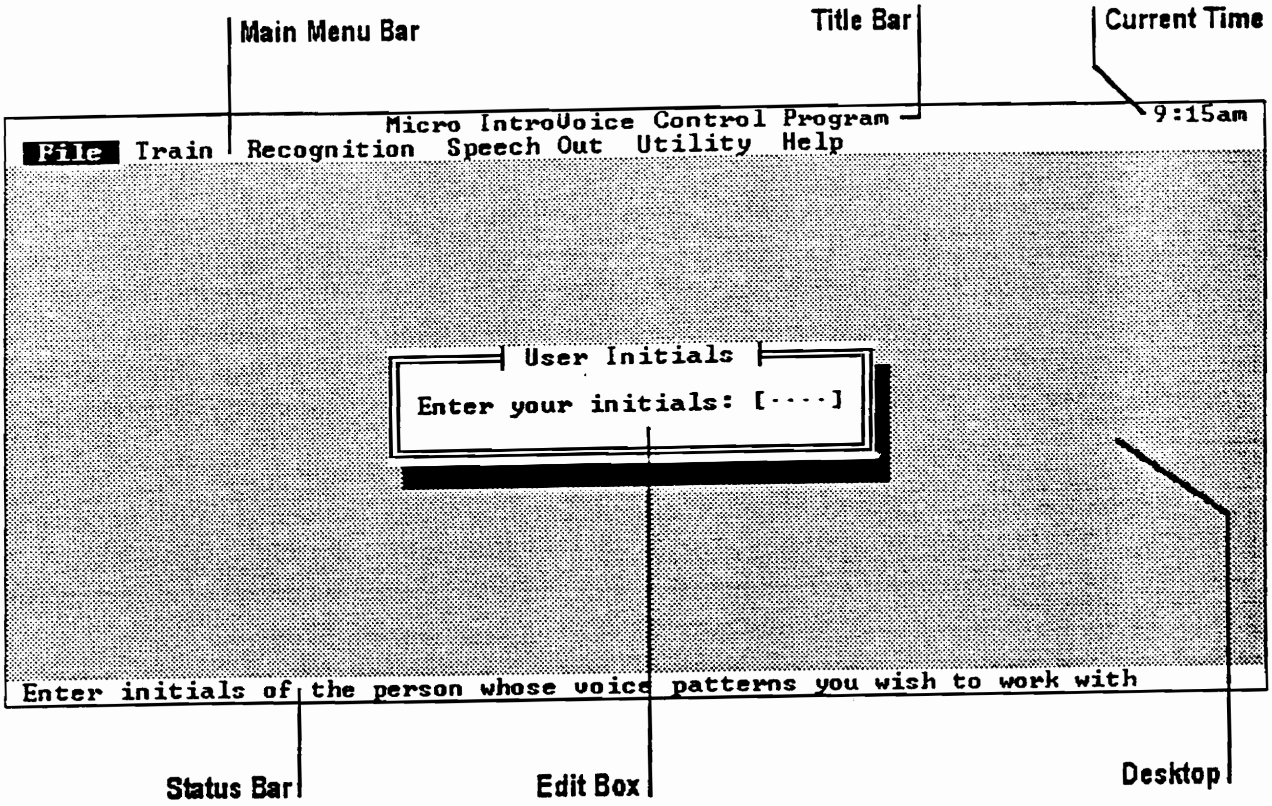
4.4.1 Basics in using the Micro IntroVoice

Basics in using the Micro IntroVoice

(these sections are to be read before coming to lab)

Elements of the MIV Control Program Screen

Some of the elements of the MIV Control Program Screen are indicated below.


Main Menu Bar
Title Bar
Current Time
Micro IntroVoice Control Program
9:15am
File Train Recognition Speech Out Utility Help
Status Bar
Edit Box
Desktop

User Initials

Enter your initials: [····]

Enter initials of the person whose voice patterns you wish to work with

FIGURE 4.2: MIV Control Program Screen

The main menu bar is where commands are selected. At the bottom of the screen is the status bar. The status bar displays information about the current operation and should always be the first place you look for assistance. The largest portion of the screen is used by the area called the desktop. This area is like the top of a desk in the sense that things you work with (windows, edit boxes, list boxes, message boxes, etc.) appear over this area. The figure shows the User Initials edit box (described in detail later) displayed on the desktop.

Selecting Commands

Commands direct the MIV Control Program to perform specific operations and are selected from the pull-down menus. Many commands can also be invoked using keyboard shortcuts.

Using the Menus

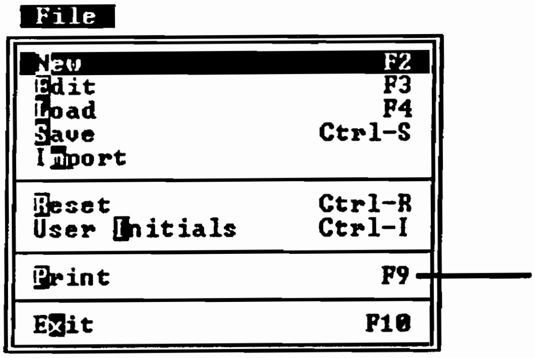
Commands can be selected from the pull-down menus that run along the second line of the display. When the program starts up, the file menu drops down automatically. To select a command from the menu when no menus are currently open, press the **ALT key** to activate the main menu. You can then select a submenu by using the **arrow keys** to move the highlight bar to the item you want and pressing **ENTER**, or by pressing the **bold letter** that appears within the desired selection. Once a pull-down menu drops down, you can use the same method to make pull-down menu selections. Press **ESCAPE** to close the menu without making a selection.

A good way to become familiar with the features of the MIV Control Program is by browsing the menus. Once a pull-down menu is open, you can open the adjoining submenu by pressing the **left or right arrow key**. As you move the menu selection bar with the **up and down arrow keys**, the status bar displays a brief description of the highlighted command. If you want more detailed information, move the highlight bar to the command in question and press **FI**. This will open the **Help** window with a description of that command.

Keyboard Shortcuts

In addition to using the menus, the most frequently used commands can be executed with a single keystroke or hotkey without ever activating the menu. The MIV Control Program is designed so that when you are unfamiliar with a certain area of operation, you can browse the menus looking for the command you want. Then, if there is a command that you use often, you can save time by learning the hotkey associated with that command.

The hotkey for a given command is displayed in the menu to the right of the associated command. For example, the hot key to print a vocabulary file is F9.



F9 is the hotkey for Print

FIGURE 4.3: Hotkey F9

Note that hotkeys are not available when the menu is active

Getting Help

The MIV Control Program has extensive on line help for virtually every aspect of the program. Help is context sensitive which means that it displays information about what you were doing when help was activated.

Using the Help System

Help is accessed by pressing **F1**, or by selecting help from the Help menu. To get help about the current operation, press **F1**. To view a list box that contains all of the available help topics, press **CTRL+F1**, or select the Help/Help Index command (**ALT, H, I**). Some help topics are too large to fit inside the help window at one time. Use the **arrow keys, PGUP, PGDN, HOME**, and **END** to scroll the viewing area. Press **ESCAPE** to close help.

Although the user's manual provides substantial information about the Micro IntroVoice and the MIV Control Program, the on line help actually contains more extensive information about specific commands. To get help on a specific command, open the menu and move the highlight bar to the command and press **F1**.

Hypertext Links

The MIV Control Program's help system implements a hypertext viewing system in which many help topics contain links to other help topics. This makes it easy to branch to related help topics and then back again. Links are words that are displayed in bold letters. While the help window is open, you can press **TAB** and **SHIFT+TAB** to move the highlight bar among the currently visible links. Press **ENTER** to view the help topic referenced by the selected link. To return to the previous help topic, press **BACKSPACE**.

4.4.2 Listing of file EXAMPLE.RCS

```
File: EXAMPLE.RCS      16 word(s)       Wed Mar 15 1995 08:49 PM
=======================================================================
Num  Spoken Word      Key Replacement                        Next Vocab
-----------------------------------------------------------------------
  1  Begin                                                   main
     (main)
  2  Date             DATE[Enter][Enter]
  3  Time             TIME[Enter][Enter]
  4  Path             PATH[Enter]
  5  Check Disk       CHKDSK[Enter]
  6  Wide Directory   DIR /W[Enter]
  7  Directory        DIR /P[Enter]
  8  Clear Screen     CLS[Enter]
  9  Return           [Enter]
 10  Path             Path[Enter]
 11  Set              Set[Enter]
 12  Welcome          [Echo]Welcome to Micro IntroVoice[Spea
                      k][Esc][Enter]
 13  Toggle Speech    [Toggle Speech]
 14  Help Menu        [Help Menu][Voice]Help Menu Displays t
                      he active Recognition Vocabulary[Speak
                      ]
 15  Stop Listening   [Help Menu][Voice]Say Attention Comput on_off
                      er to Re Activate Recognition of the m
                      ain vocabulary[Speak]
     (on_off)
 16  Attention Comp   [Help Menu][Voice]The DOS vocabulary i [Previous]
     uter             s active[Speak]
```

4.4.3 Voice Recognition Lab Activity- Micro IntroVoice

VOICE RECOGNITION LAB ACTIVITY
Micro IntroVoice

I. Training a Vocabulary and Implementing Voice Recognition

Step 1: Starting the Micro IntroVoice Program
The MIV Control Program is started as you would start any other DOS program or command.

- To start the MIV Control Program

1. Turn on the MIV control box by switching the slide lever. A GREEN light signifies activation. The speaker volume, which is located next to the headset input may need to be adjusted later. So note its location for future reference.

2. Change the directory to C:\MIV by using the DOS command **CD\MIV**.

3. Type the following command at the DOS prompt:
   MIV, then press the **ENTER key**

Step 2: Entering Your User Initials
When the MIV Control Program first starts up, it displays an edit box that prompts you to type in your initials. When the edit box appears, **type in up to 3 initials**, then press **ENTER**.

Changing the User Initials
If a new user wants to begin working with the MIV Control Program, they can change the user initials without restarting the program by selecting the File/User Initials command **(ALT, F, I)** and **entering their initials**.

Note: Make sure each group member uses different initials. If two people use the same initials the older data will be lost!

Step 3: Importing a Vocabulary
To work with the example vocabulary file, you will need to import **EXAMPLE.RCS**. The Import File command will make a copy of a vocabulary file with someone else's initials and give the new copy your initials (the original file is left unchanged). In effect, this command will make your own personal copy of a vocabulary file.

- <u>To import the example vocabulary</u>

1. Select the File/Import File command **(ALT, F, M)**.

2. Type in the initials of the file extension you want to import (in this case, **RCS**) and press **ENTER**. A list box displays all the vocabulary files in the current directory with the RCS filename extension.

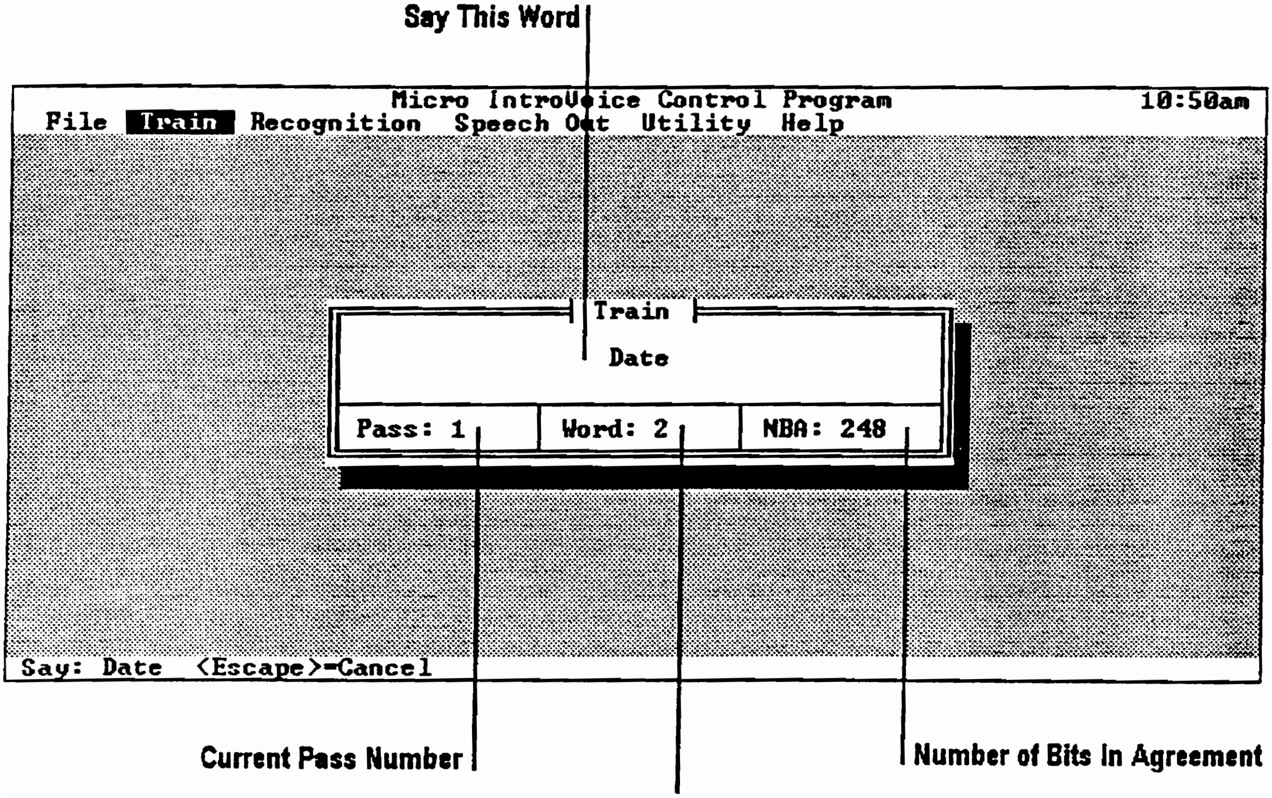
3. Use the **arrow keys** to move the highlight bar to **EXAMPLE.RCS** and press **ENTER**.

Step 4: Voice Training

Before you can use the voice recognition of the Micro IntroVoice, you must first train the Micro IntroVoice to recognize your voice as you say the words in the **EXAMPLE** vocabulary.

Training a Vocabulary

Now you can train the vocabulary. The number of training passes required can vary, but the best results occur when 7 to 10 passes are made. For this exercise you will make 7 passes. The following illustration shows the train window during training.



```
                   Micro IntroVoice Control Program                 10:50am
 File  Train  Recognition  Speech Out  Utility  Help

                      Train
                      Date
         Pass: 1   |   Word: 2   |   NBA: 248

Say: Date   <Escape>=Cancel
```

FIGURE 4.4: Train Window

- To train the vocabulary loaded in the Micro IntroVoice

1. Select the Train/Train command **(ALT, T, T)**.
   An edit box prompts you to type in the number of training passes.

2. Type 7 and press **ENTER**.

3. Keep the microphone close to the corner of your mouth and speak each word as you are prompted in the train window. Be consistent even if the system prompts you to repeat the same word. Say phrases such as "Toggle Speech" in a continuous flowing manner without any pause. The microphone is very sensitive to noise, so be sure to pronounce all words clearly.

   After 7 passes have been performed, a message box asks if you want to save the trained voice patterns to disk.

4. Press **Y**.
   An edit box appears and asks you to type the name of the vocabulary file that the voice patterns are to be saved to.

5. Since the word EXAMPLE already appears in the edit box, just press **ENTER**. If you pressed a key that caused EXAMPLE to no longer appear in the edit box, type EXAMPLE and press **ENTER**.

   The status bar (see FIGURE 4.2) displays the progress while the voice patterns are saved from the Micro IntroVoice to your personal copy of the EXAMPLE vocabulary file (EXAMPLE.xxx).

Step 5: Activating Voice Recognition
The fun begins. Now that the vocabulary is trained, you can put the power of voice to work!

Since the trained EXAMPLE vocabulary is already loaded in the Micro IntroVoice, you don't need to load it. You can simply activate recognition

- To activate voice recognition

1. Select the Recognition/Recognition Mode command **(ALT, R, R)**.

2. Use the print-out of the EXAMPLE vocabulary (given in your handout) to view the words in the vocabulary. Speak those words into the microphone. The Recognition

window displays the data sent from the Micro IntroVoice based on the key replacement field of the recognized words.

Step 6: Exiting DOS with Recognition Active

Often, you will close the recognition window by pressing ESCAPE. When you do, the MIV Control Program will take the Micro IntroVoice out of recognition mode. However, sometimes you will want to quit the MIV Control Program and return to DOS with recognition still active. For this exercise, exit to DOS with recognition still active.

- <u>To exit to DOS with recognition active</u>

These steps assume that the Recognition window is currently open and recognition is active. If not, follow the steps on activating voice recognition

1. Press **CTRL+F10**
   The MIV Control Program terminates and voice recognition remains active.

Step 7: Putting Voice to Work

When you exit to DOS with recognition active, the Micro IntroVoice is ready to send data to your computer as soon as a word is recognized; however, DOS by itself will not respond to the Micro IntroVoice. The C:\MIV directory contains a program called **VKEY** that allows you to run DOS and most DOS applications by voice.

The VKEY Program

To test the EXAMPLE vocabulary that you've trained, use the VKEY utility included on the C:\MIV directory. **VKEY** will receive data sent by the Micro IntroVoice and pass it to DOS as though it had been typed at the keyboard.

- <u>To enter DOS commands by voice</u>

1. Type **VKEY** and press **ENTER**.

2. Speak the words from the EXAMPLE vocabulary into the microphone. VKEY sends the key replacement data to the DOS command line.

Step 8: Improving the Accuracy of the Micro IntroVoice

As you probably noticed, there are a lot of accuracy problems with the Micro IntroVoice. Words that sound similar are sometimes confused with one another. There are several parameters that can be changed to try and improve the accuracy rate of the system.

Read/Set Microphone Gain

An important parameter to the recognition of speech is the GAIN. In general, in a noisy environment, human users tend to increase their vocal effort to hear their own speech.

Hence, in a noisy environment the GAIN may be reduced to maintain a good signal-to-noise ratio. The GAIN may be set over a range of 0 to 255 counts. The default value for the Micro IntroVoice is 128 counts. A reduction of the counter by a factor of 2 corresponds to 6 dB, the range from 255 down to 0 is equivalent to providing 48 dB of audio attenuation relative to the maximum gain of 255.

- To change microphone gain

1. Select the Utility/Gain command **(ALT, U, G)** to change the gain level. Make sure you record the gain level that you selected for the lab write-up.

2. Activate voice recognition **(ALT, R, R)**.

3. Speak the words from the EXAMPLE vocabulary into the microphone. Note any differences in word recognition.

4. Repeat the process until you have several data points.

Training a Single Word

From time to time you may find you are getting poor recognition of a particular word, or find that two words are being confused with each other by the Micro IntroVoice. The single Word Train command **(ALT, T, W)** allows you to train a single word from a vocabulary.

- To Retrain Word(s)

1. Select the single Word Train command **(ALT, T, W)**.

2. Type the word to be trained at the prompt as it appears in your file EXAMPLE.xxx.

3. Speak the word into the microphone and repeat for 7 passes.

4. Micro IntroVoice will then ask you if you want to update your file EXAMPLE.xxx. Type **Y**.

Baud Rate

The default setting for the baud rate is 9600 baud. This could be lowered to slow down the rate of data input and may increase the accuracy. However, lowering the baud rate would entail disconnecting the RTS (pin 4) of the dB-25 female connector attached to the Micro IntroVoice. Just note that the baud rate can be changed. **(DON'T ACTUALLY DO IT!!!)**

II. Testing the text to speech synthesis function of the Micro IntroVoice

This command allows experimentation and confirmation of the 1200 text-to-speech rules contained by the Micro IntroVoice. For example, the company name "Hughes" is mispronounced. However, if the name is spelled as "Hewz," it is pronounced correctly. The Micro IntroVoice was accompanied with software that provides an exception dictionary to cover "anomalous" pronunciations. However, these words must be entered separately so that whenever Hughes is typed the synthesizer will pronounce "Hewz" in its place. This is a long process and time does not allow for the training of exceptions in this lab activity. Note that this process can be done. To get a feel for how the text-to-speech synthesis works follow the commands listed below.

Step 9: Text to speech synthesizer

- <u>To enter text-to-speech synthesis mode</u>

1. Select Speech out, text to speech (**ALT, S, T**).

2. At the prompt type in the word "Hughes" then press the **ENTER key**. Note the pronunciation.

3. Type in the word "Hewz." Note the difference in pronunciation.

4. Next, type in another word, for instance your name and see how the text-to-speech . synthesizer pronounces it. Alter the spelling of your name to see if it affects the pronunciation (eg. Doctor Pat Koelling --> Doctor Pat Kehling).

Step 10: Exiting the Micro IntroVoice Program

- <u>To end the MIV Control Program</u>

1. Turn off the MIV control box by switching the slide lever. The GREEN light turning off signifies deactivation.

2. Change the directory to C:\MIV by using the DOS command **CD\MIV.**

3. Type the following command at the DOS prompt:
   MIV, then press the **ENTER key.**

4. The combination of turning the Micro IntroVoice off and re-entering the MIV Control Program should de-activate VKEY.

4.4.4 Micro IntroVoice Lab Write-up

Micro IntroVoice
Lab Write-up

1. Discuss some advantages and limitations of using voice recognition for data collection.

2. What effect did changing the parameters (example: retraining a single/pair of words, increasing/decreasing the gain) have on the accuracy rate of the Micro IntroVoice?

3. What is baud rate and how can it affect voice recognition?

4. Name some potential applications for voice recognition technology. Why is voice recognition better suited for the applications that you chose than other data collection technologies?

5. Which "other" words (example: your name) did you input into the text to speech synthesizer? What were the results?

6. Did the output of text to speech synthesis give an accurate reflection of the words you inputted? Why or why not?

7. Do you feel that the Micro IntroVoice is a tool that can be used in a real-world environment such as a manufacturing plant floor? Why or why not?

4.4.5 Keyboard Reference for the Micro IntroVoice

Keyboard Reference
Micro IntroVoice
General Operations

These tables list the keystrokes that are active from the MIV Control Program's main screen.

Press:	To:
F1	Activate context-sensitive help.
CTRL+F1	Display the help index with a list of all the available help topics.
ALT	Activate main menu.
CTRL+I	Change the user initials.
CTRL+M	Run Micro IntroVoice memory test.
CTRL+R	Send a reset command to the Micro IntroVoice.
F10	Exit the program and return to DOS.

File operations

Press:	To:
F2	Create a new vocabulary.
F3	Edit an existing vocabulary.
F4	Load a vocabulary from disk to the Micro IntroVoice.
F9	Print a vocabulary file.
CTRL+S	Save a vocabulary from the Micro IntroVoice to disk.

Training a vocabulary

Press:	To:
F5	Train the vocabulary currently loaded in the Micro IntroVoice.
F6	Perform update passes to vocabulary currently loaded in the Micro IntroVoice.
CTRL+W	Train a single word of the vocabulary currently loaded in the Micro IntroVoice.

Voice Recognition

Press:	To:
F7	Enter voice recognition mode.

Text to Speech

Press:	To:
F8	Speak a string using text to speech synthesizer.

Micro IntroVoice
Help

The following table lists the keystrokes that are active when the help window is open.

Press:	**To:**
F1	Open the help index with a list of all the available help topics.
ESCAPE	Close the help window.
DOWN ARROW	Scroll down one line of help text.
UP ARROW	Scroll up one line of help text.
PGDN	Scroll down one page of help text.
PGUP	Scroll up one page of help text.
HOME	Move to the beginning of the current help topic.
END	Move to the end of the current help topic.
TAB	Move the selection bar to the next visible hypertext link.
SHIFT+TAB	Move the selection bar to the previous visible hypertext link.
ENTER	Jump to the help topic referenced by the currently highlighted hypertext link.
BACKSPACE	Jump to the previous help topic.

Chapter 5: WINVOICE LAB

5.1 Introduction to WinVoice

The purpose of this lab is to give students hands on experience in using voice recognition equipment, specifically the software application WinVoice. WinVoice is a program which allows Microsoft Windows, Microsoft Windows accessories and games, application programs written for Microsoft Windows (such as Microsoft Word and Excel), and standard DOS application programs to utilize the voice recognition and synthesis capabilities of the Micro IntroVoice. The difference is that the Micro IntroVoice is activated in DOS, where WinVoice in conjunction with the Micro IntroVoice is activated by clicking on icons in Microsoft Windows.

The objectives of this lab include voice training, improving the accuracy of the system, and application in the Windows environment. Not only does this lab give students the opportunity to see another voice recognition system, but this lab exposes the limitations of voice recognition more readily than the previous lab activity. In attempting to "type" a message by voice the students can see first hand that there is still a lot of work that must be done before typing memos by voice becomes an everyday occurrence. Students will also see that there are many opportunities for research in the voice recognition field.

5.2 How to Create a Vocabulary

Since WinVoice uses the MIV Control Program of the Micro IntroVoice, the process of vocabulary development and the elements of the vocabulary editor are the same as that described in Chapter 4: Micro IntroVoice.

5.3 Vocabulary used in the Lab Activity

The vocabulary to be used in this lab exercise, WINVOICE.RCS, consists of 63 words. To train this vocabulary with the users voice pattern, making 7 passes of the 63 word vocabulary, takes approximately 20 minutes. Chapter 5.4.1 contains a copy of the WINVOICE.RCS vocabulary list.

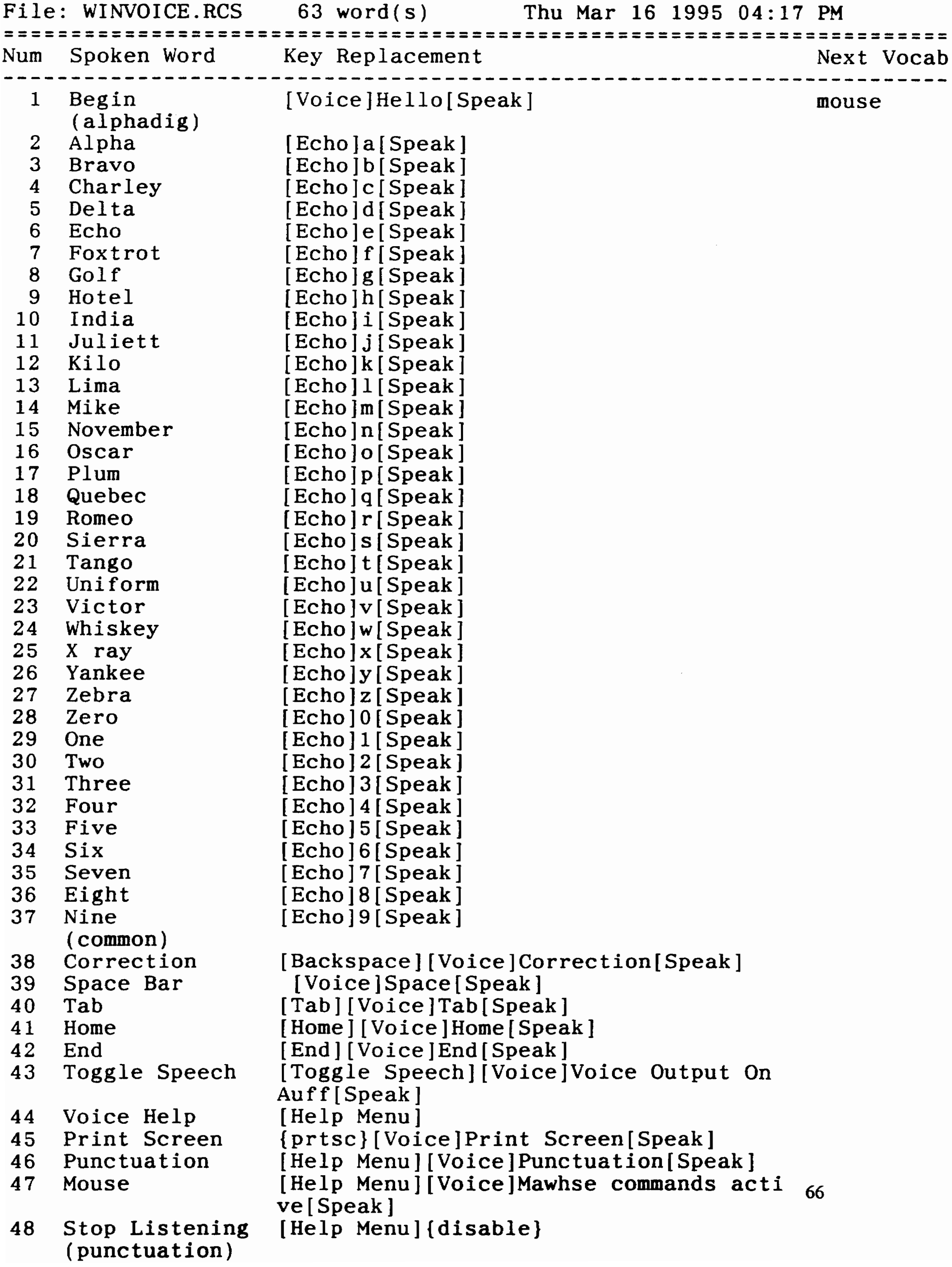
The vocabulary consists of only mouse commands and words representing the letters of the alphabet. In order to properly use WinVoice in a word processing package, approximately 130 words would need to be trained. However, this list of words would take over one hour to train. Because of the length of training time it was deemed inadequate for use as a lab activity, thus the shorter vocabulary was formulated. Even with the shorter vocabulary, the accuracy rate of this vocabulary should be noticeably worse than the EXAMPLE.RCS vocabulary trained in the previous lab activity. This is mainly due to the increase in the number of words trained. Also, the recognition speed is substantially slower as the system searches through more words in order to make a match.

5.4 WinVoice Lab Activity

The following section contains the elements of the lab developed for the application software WinVoice. This includes:

5.4.1 Listing of file WINVOICE.RCS
5.4.2 Voice Recognition Lab Activity for WinVoice
5.4.3 WinVoice Lab Write-Up
5.4.4 Keyboard Reference for the Micro IntroVoice
General Operations
Help

5.4.1 Listing of file WINVOICE.RCS



```
File: WINVOICE.RCS    63 word(s)       Thu Mar 16 1995 04:17 PM
========================================================================
Num  Spoken Word      Key Replacement                             Next Vocab
------------------------------------------------------------------------
  1  Begin            [Voice]Hello[Speak]                         mouse
     (alphadig)
  2  Alpha            [Echo]a[Speak]
  3  Bravo            [Echo]b[Speak]
  4  Charley          [Echo]c[Speak]
  5  Delta            [Echo]d[Speak]
  6  Echo             [Echo]e[Speak]
  7  Foxtrot          [Echo]f[Speak]
  8  Golf             [Echo]g[Speak]
  9  Hotel            [Echo]h[Speak]
 10  India            [Echo]i[Speak]
 11  Juliett          [Echo]j[Speak]
 12  Kilo             [Echo]k[Speak]
 13  Lima             [Echo]l[Speak]
 14  Mike             [Echo]m[Speak]
 15  November         [Echo]n[Speak]
 16  Oscar            [Echo]o[Speak]
 17  Plum             [Echo]p[Speak]
 18  Quebec           [Echo]q[Speak]
 19  Romeo            [Echo]r[Speak]
 20  Sierra           [Echo]s[Speak]
 21  Tango            [Echo]t[Speak]
 22  Uniform          [Echo]u[Speak]
 23  Victor           [Echo]v[Speak]
 24  Whiskey          [Echo]w[Speak]
 25  X ray            [Echo]x[Speak]
 26  Yankee           [Echo]y[Speak]
 27  Zebra            [Echo]z[Speak]
 28  Zero             [Echo]0[Speak]
 29  One              [Echo]1[Speak]
 30  Two              [Echo]2[Speak]
 31  Three            [Echo]3[Speak]
 32  Four             [Echo]4[Speak]
 33  Five             [Echo]5[Speak]
 34  Six              [Echo]6[Speak]
 35  Seven            [Echo]7[Speak]
 36  Eight            [Echo]8[Speak]
 37  Nine             [Echo]9[Speak]
     (common)
 38  Correction       [Backspace][Voice]Correction[Speak]
 39  Space Bar         [Voice]Space[Speak]
 40  Tab              [Tab][Voice]Tab[Speak]
 41  Home             [Home][Voice]Home[Speak]
 42  End              [End][Voice]End[Speak]
 43  Toggle Speech    [Toggle Speech][Voice]Voice Output On
                     Auff[Speak]
 44  Voice Help       [Help Menu]
 45  Print Screen     {prtsc}[Voice]Print Screen[Speak]
 46  Punctuation      [Help Menu][Voice]Punctuation[Speak]
 47  Mouse            [Help Menu][Voice]Mawhse commands acti
                     ve[Speak]
 48  Stop Listening   [Help Menu]{disable}
     (punctuation)
```

```
File: WINVOICE.RCS    63 word(s)       Thu Mar 16 1995 04:18 PM
========================================================================
Num  Spoken Word     Key Replacement                         Next Vocab
------------------------------------------------------------------------
 49  Hyphen          -                                       [Previous]
     (mouse)
 50  Up              {mmove up}[Voice]Mawhse Up[Speak]
 51  Down            {mmove down}[Voice]Mawhse Down[Speak]
 52  Left            {mmove left}[Voice]Mawhse left[Speak]
 53  Right           {mmove right}[Voice]Mawhse Right[Speak
                     ]
 54  Faster          {mfaster}[Voice]Faster[Speak]
 55  Slower          {mslower}[Voice]Slower[Speak]
 56  Nudge Right     {mnudge right}[Voice]Nudge right[Speak  28-37 mouse
                     ]
 57  Nudge Left      {mnudge left}[Voice]Nudge left[Speak]   28-37 mouse
 58  Nudge Up        {mnudge up}[Voice]Nudge up[Speak]       28-37 mouse
 59  Nudge Down      {mnudge down}[Voice]Nudge down[Speak]   28-37 mouse
 60  Stop            {mstop}[Voice]Stop[Speak]
 61  Click           {lbuttondblclk}[Voice]Double click[Spe
                     ak]
 62  Previous        [Help Menu][Voice]Returning to main vo  alphadig co
                     cabulary[Speak]                         nctuation
     (voice_on_off)
 63  Attention Comp  {disable}[Help Menu]                    [Previous]
     uter
```

5.4.2 Voice Recognition Lab Activity- WinVoice

VOICE RECOGNITION LAB ACTIVITY
WinVoice

I. Training a Vocabulary and Implementing Voice Recognition

Step 1: Starting the Micro IntroVoice Program
The MIV Control Program is started just like you would start any other DOS application in windows.

• To start the MIV Control Program

1. Turn on the MIV control box by switching the slide lever. A GREEN light signifies activation. The speaker volume, which is located next to the headset input may need to be adjusted later. So note its location for future reference

2. If you are in DOS enter windows. Then **double click** on the Applications icon.

3. Find the MIV Control Program and **double click** on the icon.

Step 2: Entering Your User Initials
When the MIV Control Program first starts up, it displays an edit box that prompts you to type in your initials. When the edit box appears, **type in up to 3 initials**, then press **ENTER**.

Changing the User Initials
If a new user wants to begin working with the MIV Control Program, they can change the user initials without restarting the program by selecting the File/User Initials command **(ALT, F, I)** and **entering their initials**.

Note: Make sure each group member uses different initials. If two people use the same initials the older data will be lost!

Step 3: Importing a Vocabulary
To work with the example vocabulary file, you will need to import **WINVOICE.RCS**. The Import File command will make a copy of a vocabulary file with someone else's initials and give the new copy your initials (the original file is left unchanged). In effect, this command will make your own personal copy of a vocabulary file.

- To import the example vocabulary

1. Select the File/Import File command **(ALT, F, M)**.

2. Type in the initials of the file you want to import (in this case, **RCS**) and press **ENTER**. A list box displays all the vocabulary files in the current directory with the RCS filename extension.

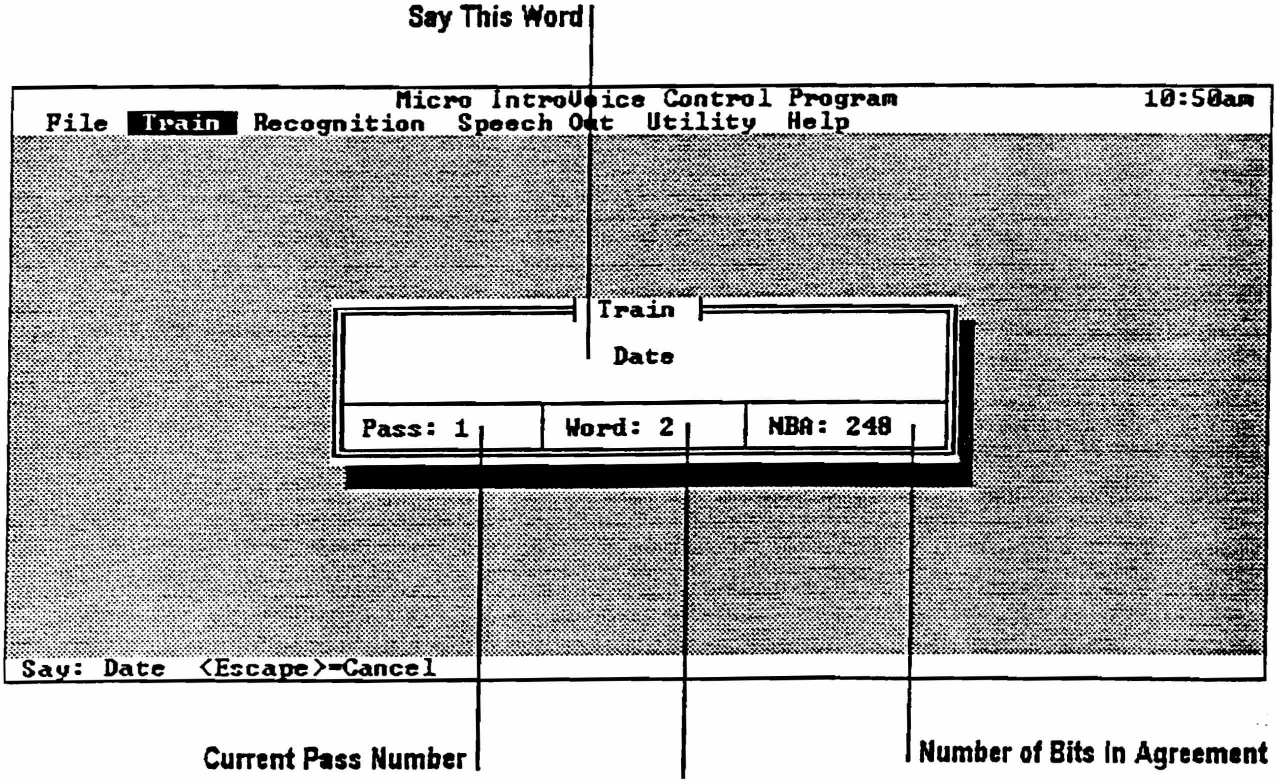
3. Use the **arrow keys** to move the highlight bar to **WINVOICE.RCS** and press **ENTER**.

Step 4: Voice Training

Before you can use the voice recognition of the Micro IntroVoice, you must first train the Micro IntroVoice to recognize your voice as you say the words in the **WINVOICE** vocabulary.

Training a Vocabulary

Now you can train the vocabulary. The number of training passes required can vary, but the best results occur when 7 to 10 passes are made. For this exercise you will make 7 passes. The following illustration shows the train window during training.



Say This Word

Micro IntroVoice Control Program 10:50am

Train

Date

Pass: 1

Word: 2

NBA: 248

Say: Date <Escape>=Cancel

Current Pass Number

Number of Bits In Agreement

FIGURE 5.1: Train Window

- To train the vocabulary loaded in the Micro IntroVoice

1. Select the Train/Train command **(ALT, T, T)**.
   An edit box prompts you to type in the number of training passes.

2. Type 7 and press **ENTER**.

3. Keep the microphone close to the corner of your mouth and speak each word as you are prompted in the train window. Be consistent even if the system prompts you to repeat the same word. Say phrases such as "Toggle Speech" in a continuous flowing manner without any pause. The microphone is very sensitive to noise, so be sure to pronounce all words clearly.

   After 7 passes have been performed, a message box asks if you want to save the trained voice patterns to disk.

4. Press **Y**.
   An edit box appears and asks you to type the name of the vocabulary file that the voice patterns are to be saved to.

5. Since the word WINVOICE already appears in the edit box, just press **ENTER**. If you pressed a key that caused WINVOICE to no longer appear in the edit box, type WINVOICE and press ENTER.

   The status bar displays the progress while the voice patterns are saved from the Micro IntroVoice to your personal copy of the WINVOICE vocabulary file (WINVOICE.xxx).

Step 5: Activating Voice Recognition

The fun begins. Now that the vocabulary is trained, you can put the power of voice to work!

Since the trained WINVOICE vocabulary is already loaded in the Micro IntroVoice, you do not need to load it. You can simply activate recognition

- To activate voice recognition

1. Select the Recognition/Recognition Mode command **(ALT, R, R)**.

2. Use the print-out of the WINVOICE vocabulary (given in your handout) to view the words in the vocabulary. Speak those words into the microphone. The recognition window displays the data sent from the Micro IntroVoice based on the key replacement field of the recognized words.

Step 6: Exiting DOS with Recognition Active
Often, you will close the recognition window by pressing ESCAPE. When you do, the MIV Control Program will take the Micro IntroVoice out of Recognition mode. However, sometimes you will want to quit the MIV Control Program and return to windows with recognition still active. For this tutorial, exit to windows with recognition still active.

- To exit to windows with recognition active

These steps assume that the Recognition window is currently open and recognition is active. If not, follow the steps on activating voice recognition

1. Press **CTRL+F10**. The MIV Control Program terminates and you will return to the applications folder of windows

2. **Double click** on the WinVoice Icon (see below). This activates voice recognition.

FIGURE 5.2: WinVoice Icon

After double clicking on the icon the following screen should pop up

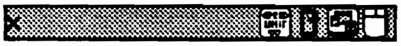


FIGURE 5.3: WinVoice Pop-up Screen

Step 7: Putting Voice to Work
1. Use the **(mouse)** subvocabulary to maneuver the mouse via voice as you would do it manually.

2. Enter the word processing package by using only voice commands.

3. After entering the word processor, say the word "Previous." (see file WINVOICE.RCS). This command changes the active subvocabulary from (mouse) to the (alphadig), (common), and (punctuation) subvocabularies.

4. By voice, type the group members names and social security numbers using the format below:

name (TAB) social security number (PERIOD)
<u>example:</u>
richard sanders   216-02-5461.

Note: the "correction" command can be used if the program misinterprets any of your alphanumeric commands (and it will!)

5. **Print** the file. This is to be turned in along with your WinVoice Lab Write-up.

6. To return to the (mouse) subvocabulary, say **"mouse."** This will make the (mouse) subvocabulary the only active subvocabulary.

Step 8: Exiting the Micro IntroVoice Program

- <u>To end the MIV Control Program</u>

1. Manually exit the word processing package.

2. Go back to the applications window and click once on the active WinVoice icon (see below).

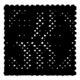


FIGURE 5.4: Active WinVoice Icon

3. Turn off the MIV control box by switching the slide lever. The GREEN light turning off signifies deactivation.

5.4.3 WinVoice Lab Write-up

WinVoice
Lab Write-up

1. How was the accuracy rate of WinVoice in comparison to the Micro IntroVoice? Why do you think there are differences in accuracy rates?

2. What other commands would you include in the mouse subvocabulary? Why?

3. Do you feel that typing a letter using WinVoice is a feasible alternative? Why or why not?

4. What improvements would you make to the WinVoice program to make it more user friendly?

5. What are some potential research opportunities in utilizing the voice for word processing? What problems need to be overcome before typing a memo via voice becomes an everyday occurrence?

(make sure to attach the WINVOICE.xxx file to your lab write-up)

5.4.4 Keyboard Reference for WinVoice

Keyboard Reference
WinVoice
General Operations

These tables list the keystrokes that are active from the MIV Control Program's main screen.

Press:	**To:**
F1	Activate context-sensitive help.
CTRL+F1	Display the help index with a list of all the available help topics.
ALT	Activate main menu.
CTRL+I	Change the user initials.
CTRL+M	Run Micro IntroVoice memory test.
CTRL+R	Send a reset command to the Micro IntroVoice.
F10	Exit the program and return to DOS.

File operations

Press:	**To:**
F2	Create a new vocabulary.
F3	Edit an existing vocabulary.
F4	Load a vocabulary from disk to the Micro IntroVoice.
F9	Print a vocabulary file.
CTRL+S	Save a vocabulary from the Micro IntroVoice to disk.

Training a vocabulary

Press:	**To:**
F5	Train the vocabulary currently loaded in the Micro IntroVoice.
F6	Perform update passes to vocabulary currently loaded in the Micro IntroVoice.
CTRL+W	Train a single word of the vocabulary currently loaded in the Micro IntroVoice.

Voice Recognition

Press:	**To:**
F7	Enter voice recognition mode.

Text to Speech

Press:	**To:**
F8	Speak a string using text to speech synthesizer.

WinVoice
Help

The following table lists the keystrokes that are active when the help window is open.

Press:	**To:**
F1	Open the help index with a list of all the available help topics.
ESCAPE	Close the help window.
DOWN ARROW	Scroll down one line of help text.
UP ARROW	Scroll up one line of help text.
PGDN	Scroll down one page of help text.
PGUP	Scroll up one page of help text.
HOME	Move to the beginning of the current help topic.
END	Move to the end of the current help topic.
TAB	Move the selection bar to the next visible hypertext link.
SHIFT+TAB	Move the selection bar to the previous visible hypertext link.
ENTER	Jump to the help topic referenced by the currently highlighted hypertext link.
BACKSPACE	Jump to the previous help topic.

CHAPTER 6: INTROVOICE VI

6.1 Introduction to the IntroVoice VI Lab Activity

The purpose of this lab is to give students hands on experience in using voice recognition equipment, specifically the IntroVoice VI. The system is supplied with an IntroVoice VI circuit board which plugs into an expansion slot of a Model 25 IBM / compatible computer or higher. Any program under MS/PC DOS may be operated by voice or keyboard through the supplied memory resident Voice Executive software (VEXEC). The IntroVoice VI allows PC users to input and retrieve information as well as command the operation of application programs and system functions by voice input and output.

The IntroVoice VI was first introduced in 1986. The system currently in the ADCSL is revision 5.44 made in 1990. The program should be noticeably less aesthetically pleasing and not as user friendly as the newer Micro IntroVoice (Chapter 4), WinVoice (Chapter 5), or PTVC-756 (described in Chapter 7) systems. However, even after using this older system, students should recognize the voice recognition's potential for future use in applications involving automatic data collection.

The next section describes how a vocabulary is created in the IntroVoice VI followed by a section explaining what was done for this lab activity.

6.2 How to Create a Vocabulary

IntroVoice VI contains a complete set of utilities for creating application vocabularies. From the DOS command line, the editor may be accessed in two ways: (1) by running the batch file VCREATE.BAT or (2) from a menu option after loading the Voice Utility Program (VUP).

Creating a vocabulary begins with determining the set of spoken words for controlling and/or entering data with one or several application programs. Each vocabulary file may contain from 5 to 500 spoken words with up to 250 spoken words active at any time in as many as 15 subvocabularies.

The following figure shows an example of the editing menu for IntroVoice VI. This is followed by a summary listing of cursor positioning control keys within the editor, editing functions and keys, a method for entering non-printable characters and special key replacement characters used to cause execution of built-in voice commands.

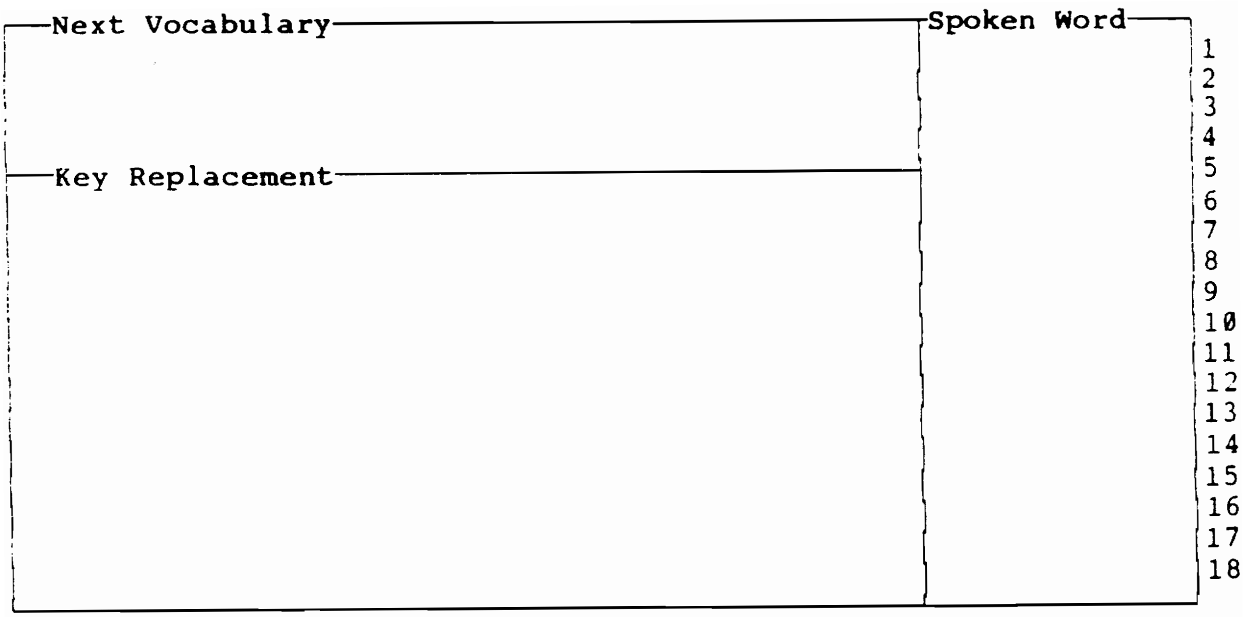


FIGURE 6.1: Vocabulary Editing/Creating Menu

6.2.1 Cursor Positioning Control Keys within the Editor

- within a window area: use **UP**, **DOWN**, **LEFT**, and **RIGHT** arrows.
- in the spoken word area: use **PAGE UP** and **PAGE DOWN** when the number of items in the spoken word area exceeds 18.
- to move between window areas: use the **TAB** key.

6.2.2 Editing Controls

- to delete a character: use the **DELETE** key.
- to insert a character: type the new character(s)
- to delete a spoken word (line): use the **CTRL-BACKSPACE** or **F10**.
- to insert a new spoken word (line): use the **INSERT** key.

6.2.3 To Return from a Subvocabulary- PREVIOUS vocabulary

- in the next vocabulary area press the F1 key followed by the **Return** key.

6.2.4 Non-Printable Characters

In the key replacement area press the **F1 key** followed by the **non-printable keystroke(s)**, for example the "ALT" or "Tab" keys.

6.2.5 Built-In Voice Commands and Associated Key Replacements

Table 6.1 lists the built in commands of the IntroVoice VI. The spoken word chosen to cause a built-in voice command key replacement may be altered at the users discretion. For example "Upper Case" may be selected to be the word "Upper." The table also provides a spoken word description and the associated key replacement to create the built-in commands.

TABLE 6.1: Built-in Commands

Spoken Word	Keystroke	Description
Rubout	F1-J	Clears (deletes) the last key replacement string
Help Menu	F1-M	Displays up to 30 items in the menu, 3 columns of up to 15 characters by 10 rows of text. Removed by pressing ANY KEY or saying a word in the vocabulary
Menu Left	F1-L	Left shifting of the Help Menu via voice or manually by pressing the left shift key
Menu Right	F1-R	Right shifting of the Help Menu via voice or manually by pressing the right shift key
Again	F1-A	Repeats the previous voice command
Caps Lock	F1-U	Provides capability to change case by spoken command
Upper Case	F1-F	Changes all lower case key replacements into upper case characters
Lower Case	F1-P	Changes all upper case key replacement into lower case characters
Control Keys	F1-O	Enables the user to activate "CTRL" key function by voice
Function	F1-W	Activates the Function keys by voice
Shift-Function	F1-X	Activates the Shift-Function Keys, Shift-F1 through Shift-F10 by voice.
Control-Function	F1-Y	Activates the Control-Function Keys, Ctrl-F1 through Ctrl-F10, by voice
Alt-Function	F1-Z	Activates the Alt-Function Keys, Alt-F1 through Alt-F10, by voice
Echo	F1-T	Echoes the key replacement sequence to the monitor & concurrently translates & buffers the corresponding phonemic sequence for synthesized voice output
Voice	F1-V	Produces speech synthesizer voice output only. The "Voice" control code translates the key replacement sequence (krs) & buffers the corresponding phonemic sequence for synthesized voice output. If the krs is followed by the control code, "Speak," the buffer is spoken immediately. If the krs is followed by "Quit" the key replacement is displayed on the monitor but held in buffer until the "Speak" code occurs
Speak	F1-S	Transfers phonemic data buffer to the speech synthesizer output. This command works together with "Echo," "Voice," and "Quit" commands

Toggle Speech	F1-G	Activates and deactivates synthesis output by voice command w/o changing your vocabulary
Quit	F1-Q	Used to terminate the command Function-Key after saying exactly one function number (digit)
Active Multiple Choice Window	F1-C & #	IntroVoice VI software supports changing of programs, by voice, using Multiple Choice Version 2.31. This software allows multiple programs to reside in memory, enabling rapid program switching by voice.
Pause	F1-K	Causes termination of additionally defined voice activate key replacements until an input is received from the keyboard
Delay	F1-I	Causes a delay of approximately one second before the rest of the key replacement string is activated

6.2.6 Creating Non-printable Character(s) for Key Replacement

All of the following key(s) can be accomplished by using the F1 key preceding the non-printable character(s) while creating or editing your vocabulary.

ALT-0 through ALT-9
ALT-A through ALT-Z
CTRL-A through CTRL-Z
F1 through F10
SHIFT-F1 through SHIFT-F10
ALT-F1 through ALT-F10
CTRL-F1 through CTRL-F10
ESCAPE
TAB
BACK TAB
INSERT
DELETE
END
HOME
PAGE DOWN
PAGE UP
CURSOR UP
CURSOR DOWN
CURSOR LEFT
CURSOR RIGHT
RETURN
CONTROL PRINT-SCREEN
CONTROL BREAK
BACK SPACE

6.2.7 Steps for Creating Vocabularies

1. ***Specifying Spoken Words***

A list of vocabulary items and subvocabulary labels are specified in the right hand column beneath the heading SPOKEN WORD (see Figure 6.1). The first word of the vocabulary is used to open the first subvocabulary in the list and is generally used to display the first Help Menu. Each subvocabulary list is preceded by a label identifying the associated word list. For each subvocabulary designated, a starting (label) and an (end) statement must appear. The (end) statement tells the system that the preceding spoken word was the last element of the defined list. Subvocabulary designators (labels) and (end) statements must be enclosed in parenthesis to distinguish them from spoken vocabulary items. A spoken word and a label must not have identical spelling.

2. ***Specifying Next Vocabulary***

The subvocabulary chosen to become active upon recognition of a spoken word is specified in the upper left hand corner of the editor display. Only selected words need to specify a Next Vocabulary. In the next vocabulary entering the key sequence F1 followed by Return (window return), causes a return to the previous subvocabulary.

3. ***Specifying Key Replacement***

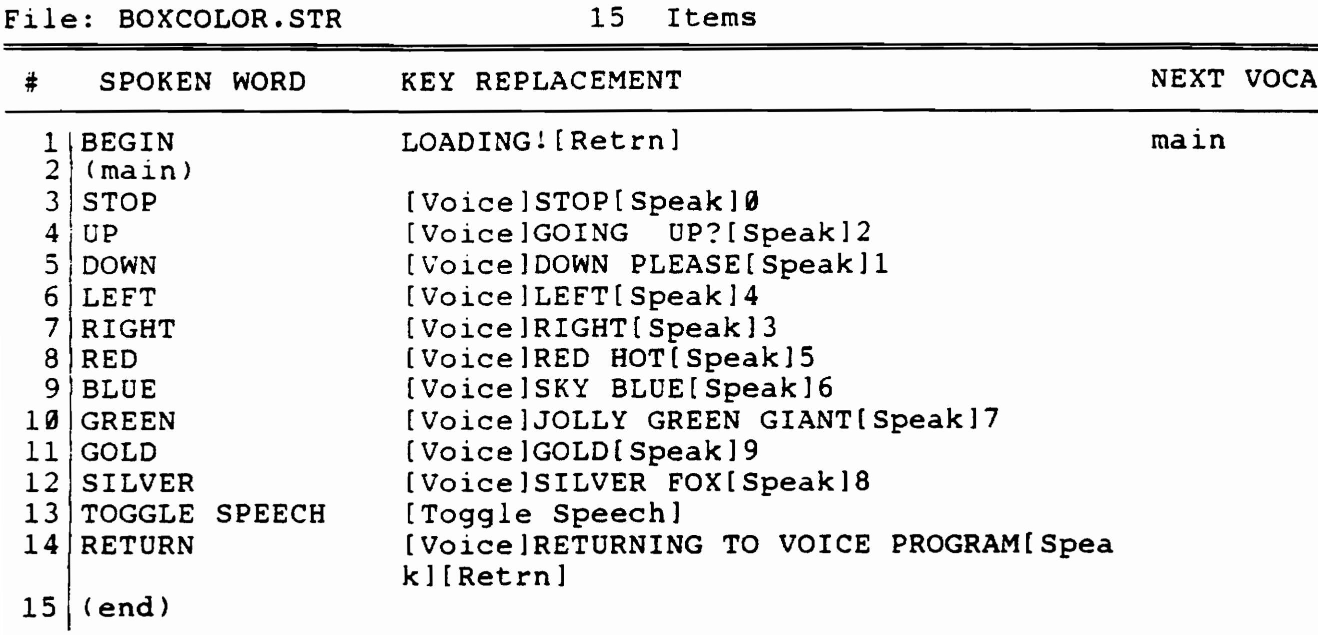
In the largest box on the screen, the user enters the key replacements or list of corresponding keystrokes and voice prompting / verification that will be generated when each spoken command is recognized. Help menus for sub-vocabularies are also entered in this editing window.

6.3 IntroVoice VI Lab Activity

The following section contains the elements of the lab developed for the IntroVoice VI. This includes:

6.3.1 Listing of file DEMONSTRATION
6.3.2 Voice Recognition Lab Activity for the IntroVoice VI
6.3.3 IntroVoice VI Lab Write-Up

6.3.1 Listing of file DEMONSTRATION



```
File:  BOXCOLOR.STR                   15  Items
═══════════════════════════════════════════════════════════════════════
 #   SPOKEN WORD      KEY REPLACEMENT                            NEXT VOCA
───────────────────────────────────────────────────────────────────────
  1|BEGIN             LOADING![Retrn]                            main
  2|(main)
  3|STOP              [Voice]STOP[Speak]Ø
  4|UP                [Voice]GOING  UP?[Speak]2
  5|DOWN              [Voice]DOWN PLEASE[Speak]1
  6|LEFT              [Voice]LEFT[Speak]4
  7|RIGHT             [Voice]RIGHT[Speak]3
  8|RED               [Voice]RED HOT[Speak]5
  9|BLUE              [Voice]SKY BLUE[Speak]6
 1Ø|GREEN             [Voice]JOLLY GREEN GIANT[Speak]7
 11|GOLD              [Voice]GOLD[Speak]9
 12|SILVER            [Voice]SILVER FOX[Speak]8
 13|TOGGLE SPEECH     [Toggle Speech]
 14|RETURN            [Voice]RETURNING TO VOICE PROGRAM[Spea
   |                  k][Retrn]
 15|(end)
```

6.3.2 Voice Recognition Lab Activity- IntroVoice VI

VOICE RECOGNITION LAB ACTIVITY IntroVoice VI

I. Training a Vocabulary and Implementing Voice Recognition

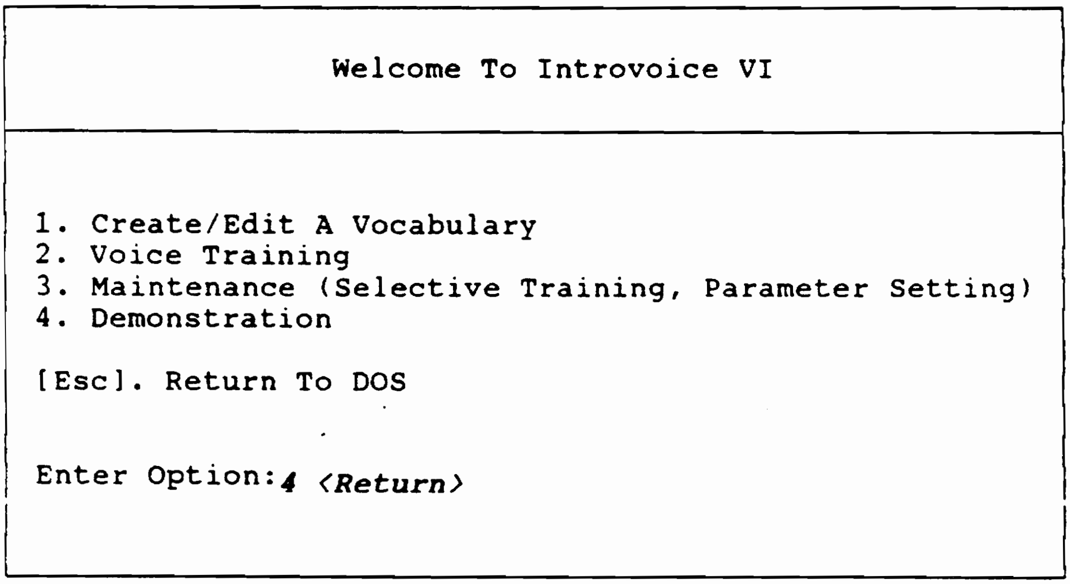
Step 1: Starting the IntroVoice VI
The IntroVoice VI is started as you would start any other DOS program or command.

- <u>To start the IntroVoice VI</u>

1. Change the directory to C:\IV6 by using the DOS command **CD\IV6**.

2. Load the Voice Executive software by typing the following command at the DOS prompt: **VEXEC**, then press the **ENTER key.**

3. Load the voice utility program by typing the following command at the DOS prompt: **VUP**, then press the **ENTER KEY**. The main menu will appear as shown below.



```
                Welcome To Introvoice VI

 1. Create/Edit A Vocabulary
 2. Voice Training
 3. Maintenance (Selective Training, Parameter Setting)
 4. Demonstration

 [Esc]. Return To DOS


 Enter Option:4 <Return>
```

FIGURE 6.2: Main Menu

Step 2: Loading the Demonstration Vocabulary

- To import the example vocabulary

1. From the main menu, select item 4- Demonstration

Step 3: Training a vocabulary

Now you can train the vocabulary. The number of training passes required can vary, but the best results occur when 7 to 10 passes are made. This exercise requires only **5 passes**.

- To train the vocabulary loaded in the IntroVoice VI

1. Next you are prompted by the screen display to speak into the microphone to train the 12 word demonstration vocabulary. The word training sequence is "STOP", "UP", "DOWN", "LEFT", "RIGHT", "RED", "BLUE", "GREEN", "GOLD", "SILVER", "TOGGLE SPEECH", and "RETURN."

   The first word in the vocabulary is prompted on the display. Hold the microphone approximately two fingers from your mouth. Say the prompted word clearly into the microphone. Successive words are prompted. Continue saying each word when prompted. If the same word appears after it is spoken, repeat the word. You may have to repeat some words a number of times. Do not change your vocal effort or question the machines request for repetition. Try as much as possible to be consistent and speak in a natural manner. The system requires such repetition in order to ensure that the current spoken word sufficiently agrees with prior occurrences of speaking the prompted word. Multiple words should be spoken quickly, without a pause, as if saying a phrase in connected speech.

   The number of times you have trained the vocabulary appears at the top of the screen. The training score also appears in the lower right-hand corner. When you train a vocabulary for the first time, the score for each word will be the maximum, that is, 128. The second training pass will initially lower the score. A word is properly trained if it scores 120 or more with variation less than + or - five points.

2. When the last word in the file has been spoken you have completed **one pass**. The system will then prompt you to update the vocabulary, type **Y**.

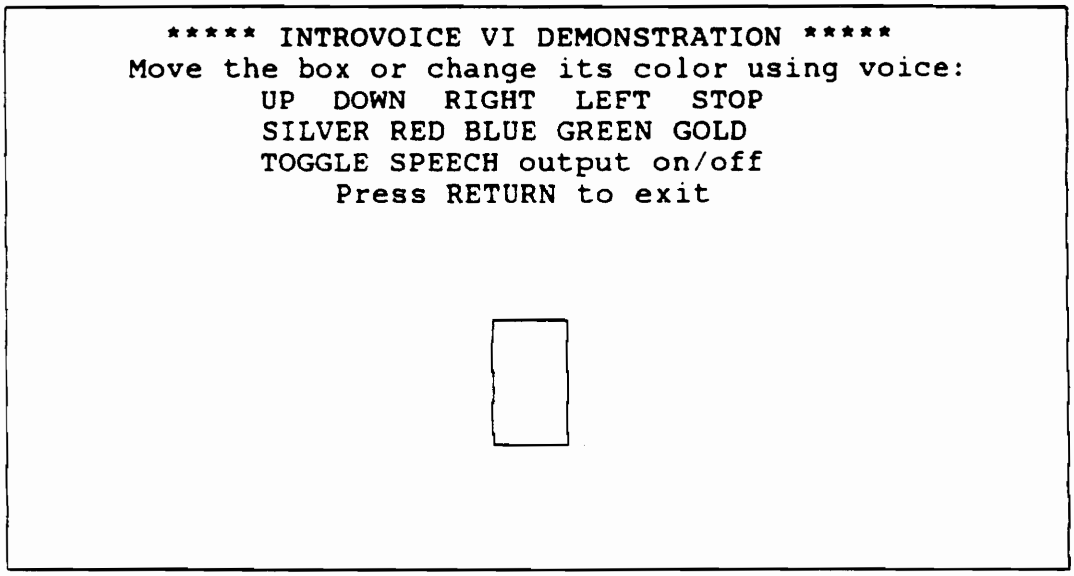
   Note: when you finish training the vocabulary, it is saved by the voice utility program after each training pass.

3. Repeat the process until you have trained the vocabulary **five times**. On the final training pass, when the update prompt appears for the last time, enter **N**.

Step 4: Putting the Voice to Work

1. When the screen appears as shown below press the voice "ON/OFF" switch to activate recognition. Have fun! When you have finished with the box demo, return to the main menu by **pressing RETURN** or by **saying "RETURN."**

   The NUM LOCK key is the default switch used to turn ON and OFF voice recognition. Note a single "beep" is given indicating that recognition is active. Two "beeps" indicate that recognition has been switched off.



```
          ***** INTROVOICE VI DEMONSTRATION *****
       Move the box or change its color using voice:
              UP  DOWN  RIGHT  LEFT  STOP
              SILVER RED BLUE GREEN GOLD
              TOGGLE SPEECH output on/off
                  Press RETURN to exit
```

FIGURE 6.3: IntroVoice VI Demonstration

6.3.3 IntroVoice VI Lab Write-up

IntroVoice VI
Lab Write-up

1. Discuss some advantages and limitations of using voice recognition for data collection.

2. Is the VUP program user friendly? If no, how would you improve the programs to make it more user friendly?

3. Name some potential applications for voice recognition technology. Why is voice recognition better suited for the applications that you chose than other data collection technologies?

4. Do you feel that the IntroVoice VI is a tool that can be used in a real-world environment such as a manufacturing plant floor? Why or why not?

CHAPTER 7: Portable Transactions Voice Computer (PTVC-756)

7.1 Introduction to the PTVC-756 Lab Activity

The purpose of this lab is to give students hands on experience in using voice recognition equipment, specifically the hand held Portable Transactions Voice Computer, PTVC-756. The lab includes training a vocabulary with the subjects voice patterns, improving the accuracy of the system, message prompting and verification, and bar-code scanning.

After doing the lab activity students should understand the basics of voice recognition and have gained practical experience in using a voice recognition system. Students should recognize voice recognition's potential future use in applications involving automatic data collection (either as a stand alone system or in conjunction with another system).

7.2 How to Create a Vocabulary

The Portable Transactions Voice Computer and the IntroVoice VI (Chapter 6) create vocabularies in the same manner, using the batch file, VCREATE.BAT. A vocabulary can also be made in the Voice Utility Program (VUP) of the IntroVoice VI and transferred to the PTVC-756 via its file send and receive utility (see next section). See Chapter 6.2 on "How to Create a Vocabulary" for an in depth discussion on the process of creating and editing a vocabulary.

7.2.1 File Send and Receive Utility

This utility allows data and program file transfers between a PC and the Portable Transactions Voice Computer. A serial cable, furnished with the PTVC-756, allows direct file transfers between the PC, via COM1, and the PTVC-756 serial port. The PTVC-756 System Master diskette contains in the directory A:\756 the file TXRXPC. Built into the EPROM of the PTVC-756 is the file TXRX750. To transfer files run the program TXRXPC on the PC or compatible and then on the PTVC-756 run the program TXRX750.

7.3 System Configuration

A block diagram of the PTVC-756 is shown below. The block diagram is subdivided into the three sections associated with the systems three circuit boards. Central to the PTVC-756, in the top housing of the system, is the 80C88 CMOS processor board with the proprietary Telxon controller IC and 128 Kb of EPROM. The EPROM's provide a few useful utility files and allows MS-DOS operation. The top housing also contains the RAM expansion board. A lithium battery provides complete RAM backup, even with the 6 AA cells completely removed. The bottom housing contains the third circuit board with the Voice Input/Output electronics, bar-code interface, and serial communication electronics. The system, configured for MS-DOS, requires a minimum of 512 KB of static RAM and directly addresses and supports up to 1 Mb.

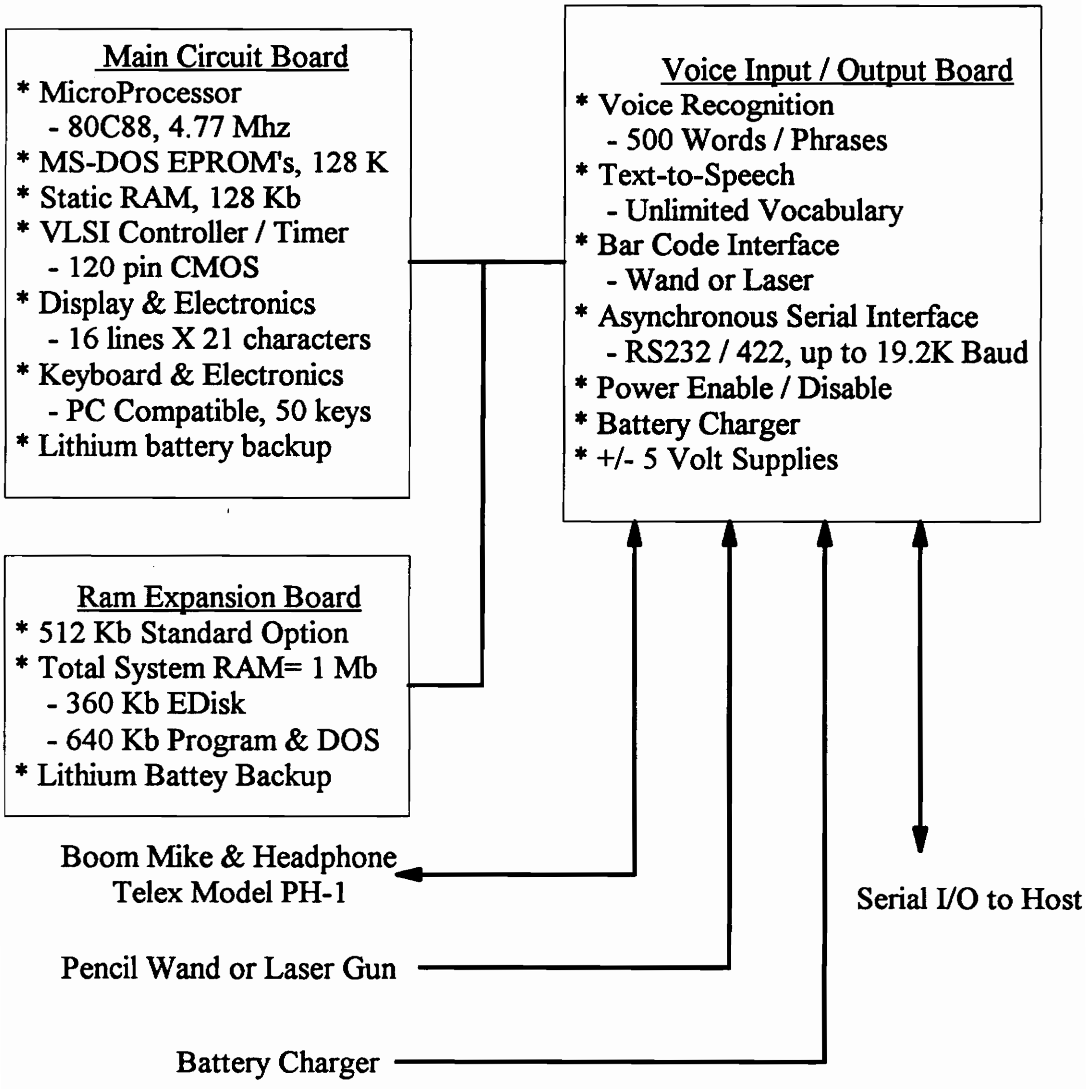


FIGURE 7.1: Block Diagram of PTVC-756

7.4 Portable Transactions Voice Computer (PTVC-756) Lab Activity

The following section contains the elements of the lab developed for the Portable Transactions Voice Computer. This includes:

7.4.1 Basics in using the Portable Transactions Voice Computer (this is to be read before attending the lab session)
7.4.2 Listing of file BARCODE.COM
7.4.3 Voice Recognition Lab Activity for the Portable Transactions Voice Computer
7.4.4 Portable Transactions Voice Computer Lab Write-Up

7.4.1 Basics in using the PTVC-756

Basics in using the PTVC-756

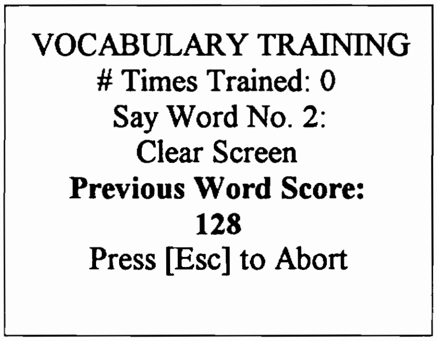
(these sections are to be read before coming to lab)

Training a Vocabulary

Say each prompted word in a firm natural consistent voice. If the display screen flashes and the same word appears, repeat the word until a new word appears, but be consistent in the manner you repeat each word. Do not shout at or question the system while training. When a short phrase, such as "Clear Screen" is prompted say the phrase in a connected fashion, do not pause between words. During training, relax and try to keep a fairly constant volume and speaking rate. Try to use a posture similar to that you will use when operating the system, for example, standing or sitting.

A minimum of 3 training passes will allow fair recognition if used over a short time span. For good recognition and stability over an extended time, 5 to 7 training passes are recommended.

Some increase in recognition accuracy typically occurs up to about 10 training passes. Training should proceed until the average score over the set of words being trained is in the neighborhood of 120 (see FIGURE 7.2). In the first training pass the score of all words is the maximum of 128 by default. Thereafter, the score falls and then increases with the number of training passes. For this lab activity the number of training passes will be set to 7.



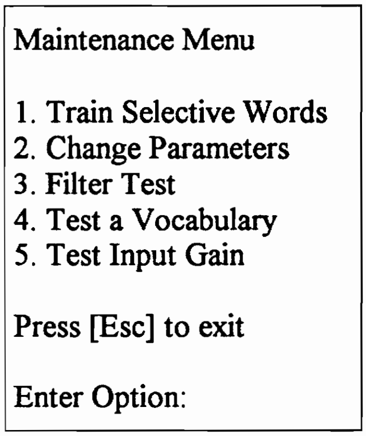
VOCABULARY TRAINING
# Times Trained: 0
Say Word No. 2:
Clear Screen
Previous Word Score:
128
Press [Esc] to Abort

FIGURE 7.2: Previous Word Score

In voice recognition each spoken word/phrase is compared with the active recognition training patterns and the training pattern having the highest score with respect to the spoken input is chosen. If the score falls below the recognition Reject Threshold (RTHL) the utterance is rejected (beep will be heard). The default Reject Threshold score is 106. This adjustable parameter is a function of the number of training passes. The higher RTHL the better the rejection of out-of-class utterances and spurious sounds.

Improving Voice Recognition Accuracy

Improvements in voice recognition accuracy can be made by using one or more of the following utilities of the PTVC-756. The general functions of the five Voice Maintenance options are briefly described below. Some of these functions will be put to use in the Portable Transactions Voice Computer Lab Activity.



Maintenance Menu

1. Train Selective Words
2. Change Parameters
3. Filter Test
4. Test a Vocabulary
5. Test Input Gain

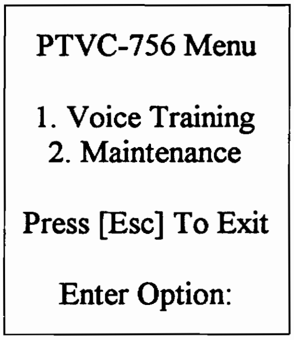
Press [Esc] to exit

Enter Option:

FIGURE 7.3: Maintenance Menu

1. Train Selective Words

If you add a word to the end of the vocabulary or if you are having recognition problems with individual words, you can train the INDIVIDUAL words rather than the entire vocabulary using menu option 1 (see below). Selective word training should not be exercised if the corresponding vocabulary has been trained less than five training passes unless a new word is being appended to an existing training vocabulary.



PTVC-756 Menu

1. Voice Training
2. Maintenance

Press [Esc] To Exit

Enter Option:

FIGURE 7.4: Main Menu

Why do recognition substitution errors occur?

A typical cause of recognition substitution errors is a "poor" voice reference pattern (template). During each training pass each word in the vocabulary is sequentially prompted, and, except the first pass, each input word is classified and scored relative to its reference pattern. The maximum score for any word is 128. The higher the score the better the prompted spoken word matches the reference pattern for that word. If this score does not exceed a reject threshold (RTHL), which is increased with training passes, the word is reprompted. This, in general, prevents a voice reference pattern from being modified by unwanted sounds, such as background noises, coughing, breathing noises, saying a prompted word in an inconsistent manner between training passes, or even accidentally saying a word different than the prompted word.

The following table lists the reject threshold score versus training passes:

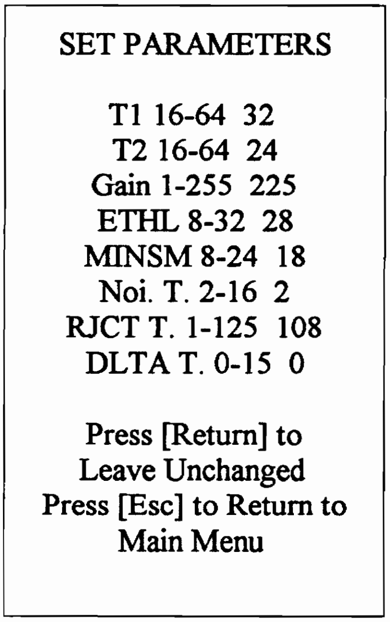
TABLE 7.1: RTHL

Training Pass	Reject Threshold (RTHL)
1	0
2	97
3 & 4	101
5 & up	105

The value of the reject threshold (RTHL) versus training passes must be low enough to accept the considerable variability of a single speaker repeating the same word and yet high enough to reject significant inconsistencies for the prompted training word. Because of this compromise, even with the RTHL training table, it is possible for one or more "poor" templates to occur.

2. Change Parameters

If you select item 2 and then enter the "password" keys, !@#, you may alter the various system parameters. It is recommended that only Reject Threshold (RJCT T.), Delta Reject Threshold (DLTA T.), and the Audio Input Gain (Gain) be changed



```
     SET PARAMETERS

      T1 16-64  32
      T2 16-64  24
    Gain 1-255  225
     ETHL 8-32  28
    MINSM 8-24  18
    Noi. T. 2-16  2
   RJCT T. 1-125  108
    DLTA T. 0-15  0

    Press [Return] to
    Leave Unchanged
  Press [Esc] to Return to
       Main Menu
```

FIGURE 7.5: Set Parameters

Reject Threshold and Delta Reject Threshold

During recognition, an incoming word is compared and scored with all reference word patterns of the active subvocabulary(s). A maximum score of 128 can ideally be attained.

If: Score >= RTHL and exceeds δt

The word with the highest score is recognized and the corresponding key replacement output.

Where: RTHL= Reject Threshold
δt= Delta Reject Threshold

If either: Score < RTHL
or δt<selected constant
The word is rejected, typically indicated by a "BEEP."

The value of RTHL is a function of training passes. To eliminate unwanted recognition response of spoken words not in the vocabulary, coughing, background sounds, etc. the value of RTHL should be set as large as possible. The factory default value of 108 works well for five training passes. To optimize performance increase RTHL two counts for each additional training passes (max. value= 120).

The value δt= winning score (highest) - runner up score. The runner up score is the nearest neighbor of the non-winner. A large delta, δt, means the recognized word is well separated from the winning word and is reliably classified. The default value of δt=0. This insures that if a word exceeds RTHL it is recognized and not rejected. The value delta, δt, may be increased until excessive rejects are occurring when saying active vocabulary words. In general it is not recommended to set delta to a value over 3.

Audio Input Gain

For noisy environments, the microphone gain should be decreased. The Input Gain is the only parameter that is changed during the lab activity.

3. Filter Test

You can check to verify that the 16 bandpass filters are operational with menu option 3. These filters have a small output offset with the mike OFF and should graph the output of your speech for continued long duration speech sounds, such as the vowels "AH" or "EE." For valid filter data to be displayed a trained vocabulary file must be loaded prior to selecting this menu option. Use the batch file TESTFILT.BAT prior to selecting this menu option to insure a trained vocabulary is loaded.

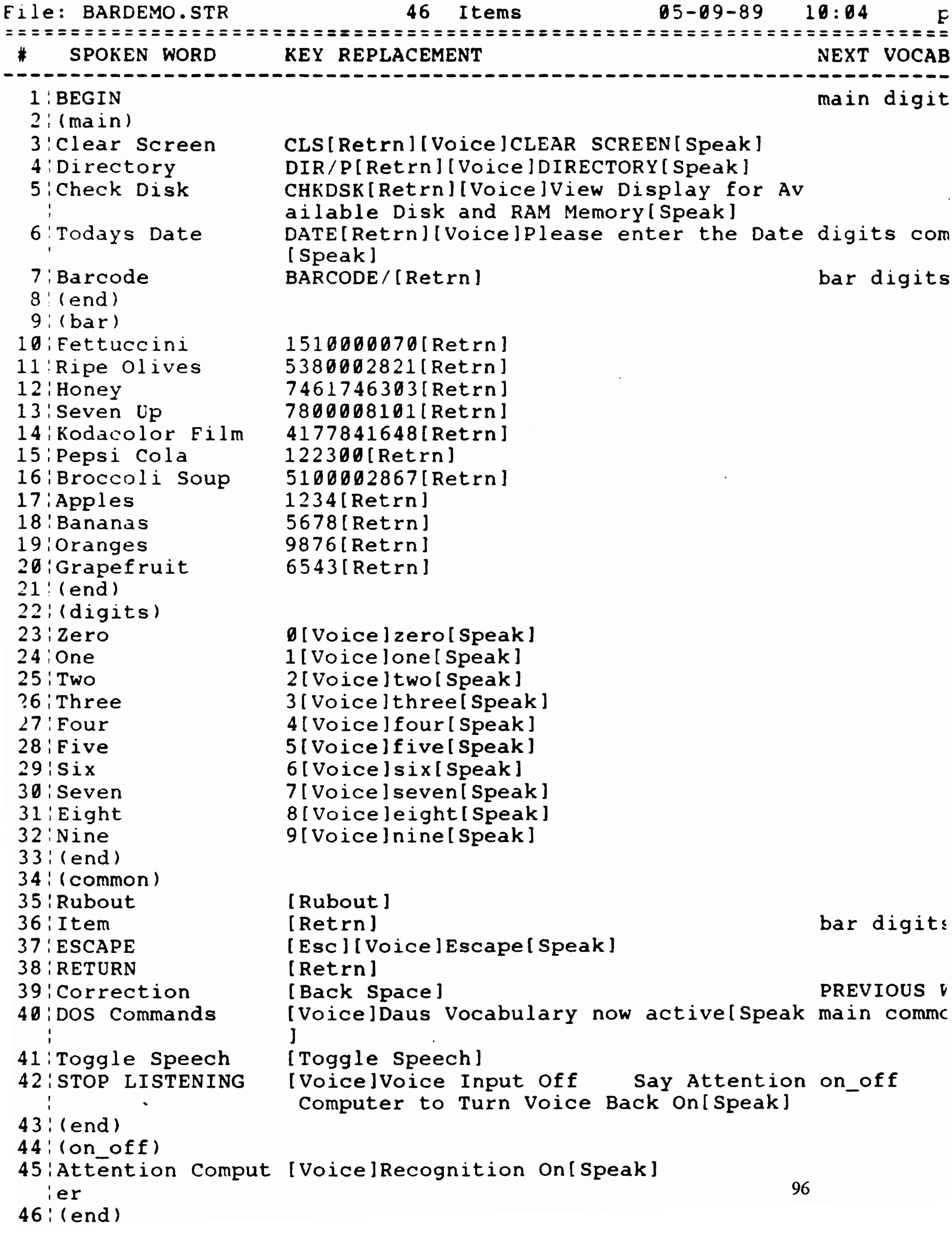
4. Test a Vocabulary

You can test the system to determine if it is confusing two similar sounding words with menu option 4. An average difference score of **4** or greater for **7** test repetitions of the same word yields reliable and accurate recognition.

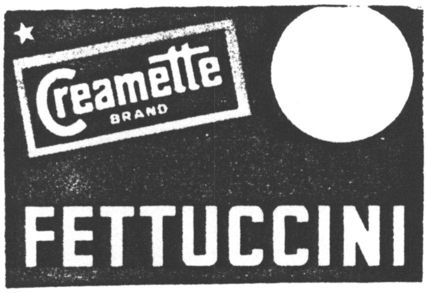
5. Test Input Gain

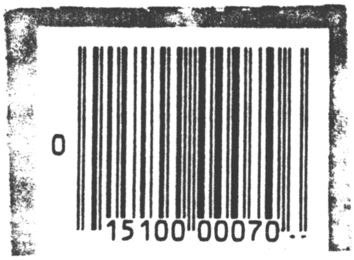
This options requests that you say the ten digits, 0 through 9. The peak energy in each of the 16 filters for all of the spoken digits is computed and displayed, followed by the mean and peak value over the ten digit vocabulary. This option is extremely useful in establishing the gain for a given microphone and background noise level.

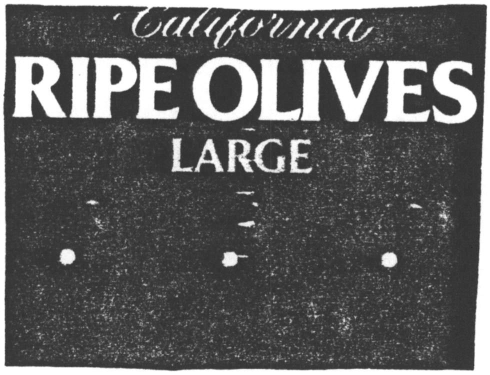
7.4.2 Listing of file BARDEMO.COM

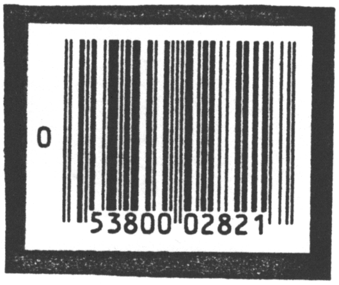


```
File: BARDEMO.STR                 46  Items          05-09-89   10:04      P
=============================================================================
 #   SPOKEN WORD      KEY REPLACEMENT                           NEXT VOCAB
-----------------------------------------------------------------------------
   1¦BEGIN                                                      main digit
   2¦(main)
   3¦Clear Screen     CLS[Retrn][Voice]CLEAR SCREEN[Speak]
   4¦Directory        DIR/P[Retrn][Voice]DIRECTORY[Speak]
   5¦Check Disk       CHKDSK[Retrn][Voice]View Display for Av
    ¦                 ailable Disk and RAM Memory[Speak]
   6¦Todays Date      DATE[Retrn][Voice]Please enter the Date digits com
    ¦                 [Speak]
   7¦Barcode          BARCODE/[Retrn]                           bar digits
   8¦(end)
   9¦(bar)
  10¦Fettuccini       1510000070[Retrn]
  11¦Ripe Olives      5380002821[Retrn]
  12¦Honey            7461746303[Retrn]
  13¦Seven Up         7800008101[Retrn]
  14¦Kodacolor Film   4177841648[Retrn]
  15¦Pepsi Cola       122300[Retrn]
  16¦Broccoli Soup    5100002867[Retrn]
  17¦Apples           1234[Retrn]
  18¦Bananas          5678[Retrn]
  19¦Oranges          9876[Retrn]
  20¦Grapefruit       6543[Retrn]
  21¦(end)
  22¦(digits)
  23¦Zero             0[Voice]zero[Speak]
  24¦One              1[Voice]one[Speak]
  25¦Two              2[Voice]two[Speak]
  26¦Three            3[Voice]three[Speak]
  27¦Four             4[Voice]four[Speak]
  28¦Five             5[Voice]five[Speak]
  29¦Six              6[Voice]six[Speak]
  30¦Seven            7[Voice]seven[Speak]
  31¦Eight            8[Voice]eight[Speak]
  32¦Nine             9[Voice]nine[Speak]
  33¦(end)
  34¦(common)
  35¦Rubout           [Rubout]
  36¦Item             [Retrn]                                   bar digits
  37¦ESCAPE           [Esc][Voice]Escape[Speak]
  38¦RETURN           [Retrn]
  39¦Correction       [Back Space]                              PREVIOUS V
  40¦DOS Commands     [Voice]Daus Vocabulary now active[Speak main commc
    ¦                 ]
  41¦Toggle Speech    [Toggle Speech]
  42¦STOP LISTENING   [Voice]Voice Input Off    Say Attention on_off
    ¦                  Computer to Turn Voice Back On[Speak]
  43¦(end)
  44¦(on_off)
  45¦Attention Comput [Voice]Recognition On[Speak]
    ¦er
  46¦(end)
```

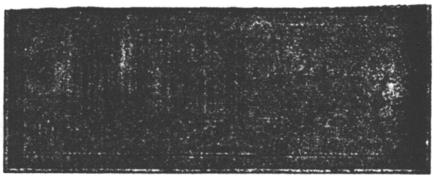

Creamette
BRAND
FETTUCCINI


0
15100 00070

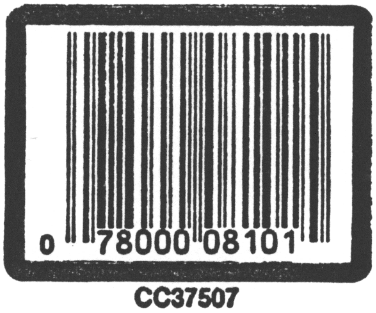

California
RIPE OLIVES
LARGE

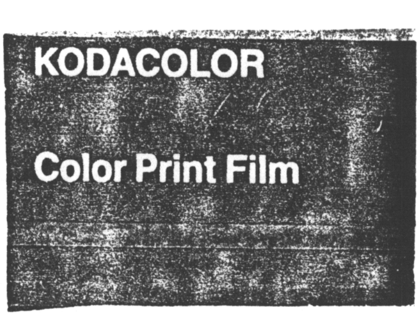

0
53800 02821

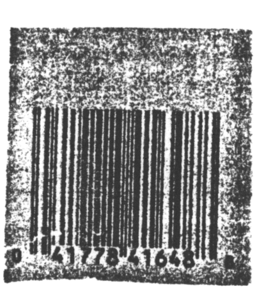
cucamonga
HONEY

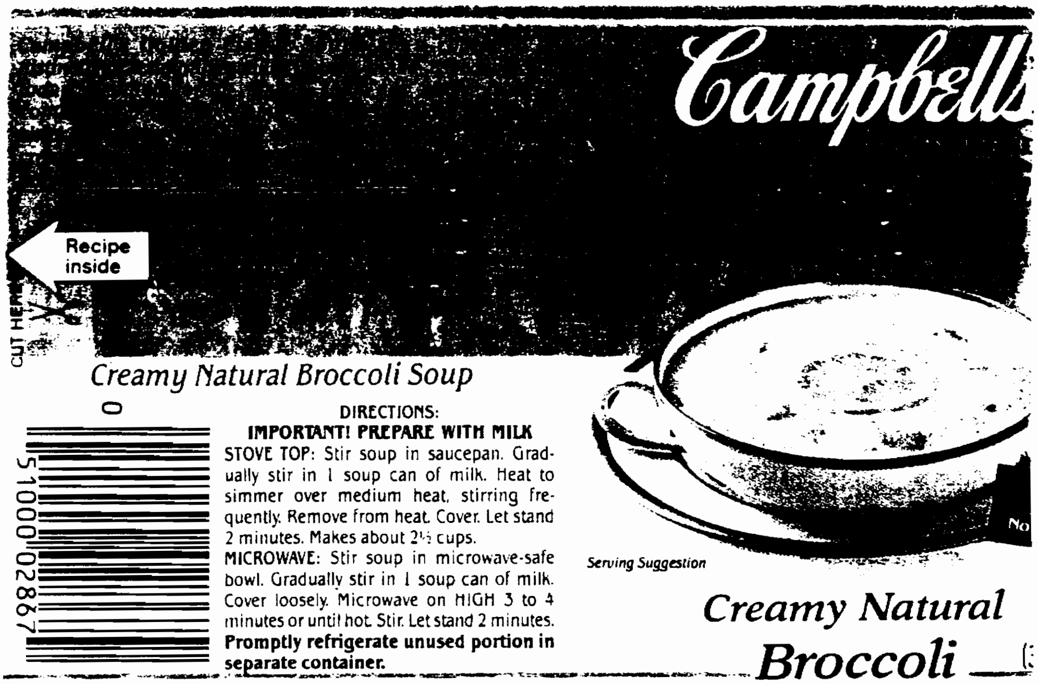


7UP


0
78000 08101
CC37507


KODACOLOR
Color Print Film


0
41778 41648



Campbell's

Recipe inside

CUT HERE

Creamy Natural Broccoli Soup

0 51000 02867

DIRECTIONS:
IMPORTANT! PREPARE WITH MILK

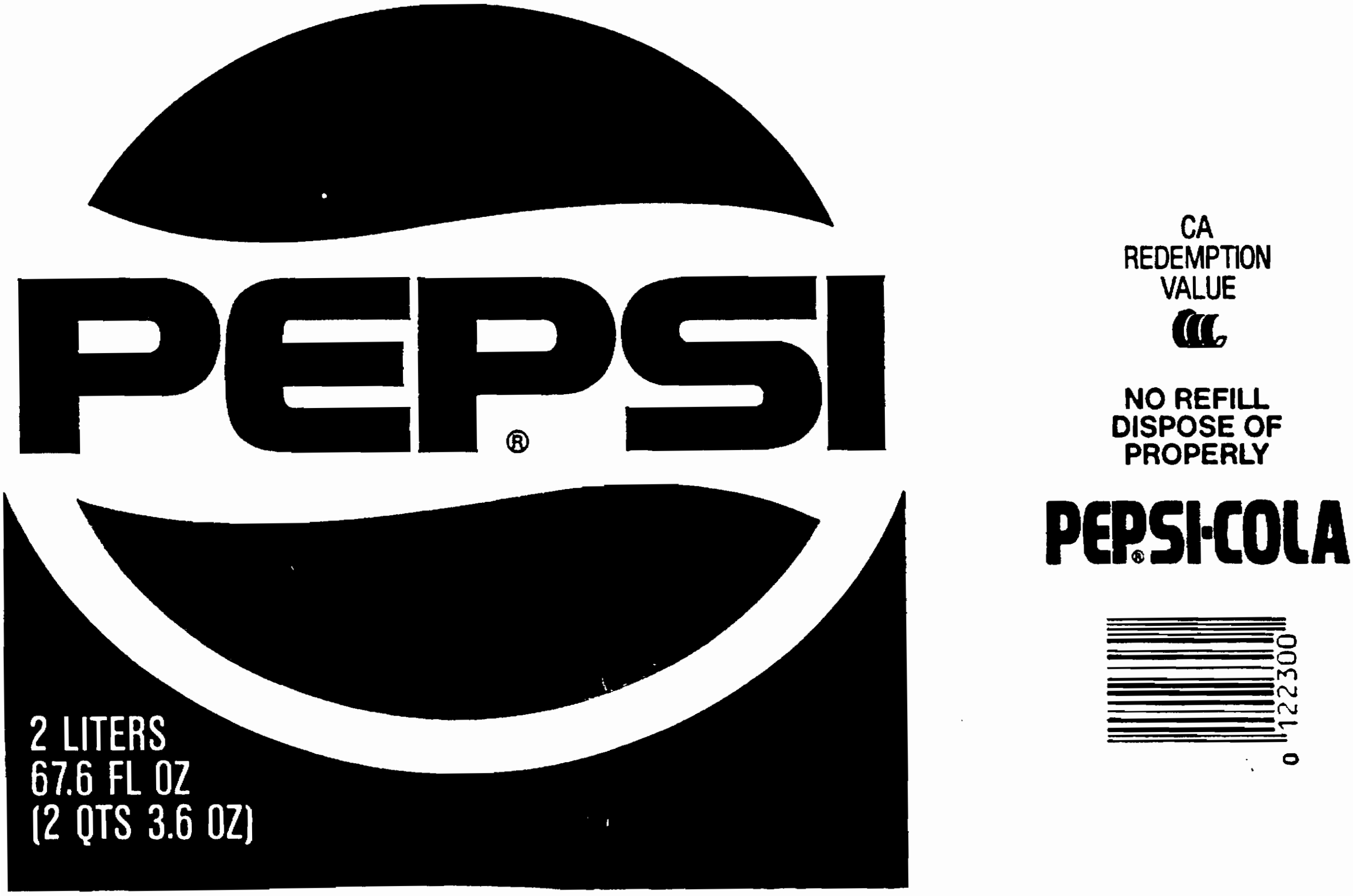
STOVE TOP: Stir soup in saucepan. Gradually stir in 1 soup can of milk. Heat to simmer over medium heat, stirring frequently. Remove from heat. Cover. Let stand 2 minutes. Makes about 2½ cups.

MICROWAVE: Stir soup in microwave-safe bowl. Gradually stir in 1 soup can of milk. Cover loosely. Microwave on HIGH 3 to 4 minutes or until hot. Stir. Let stand 2 minutes.

Promptly refrigerate unused portion in separate container.

Serving Suggestion

Creamy Natural
Broccoli



PEPSI®

2 LITERS
67.6 FL OZ
(2 QTS 3.6 OZ)

CA
REDEMPTION
VALUE

NO REFILL
DISPOSE OF
PROPERLY

PEPSI®-COLA

0 122300

7.4.3 Voice Recognition Lab Activity- Portable Transactions Voice Computer

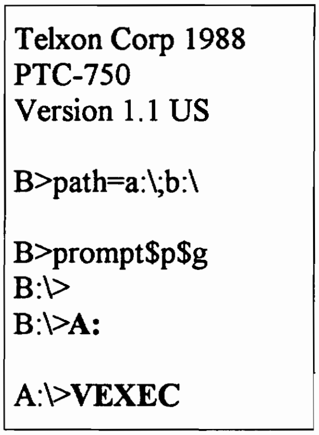
VOICE RECOGNITION LAB ACTIVITY
Portable Transactions Voice Computer- PTVC 756

I. Training a Vocabulary and Implementing Voice Recognition

Step 1: Starting the PTVC-756

- To start the PTVC-756

1. To begin first press the **"ON/OFF"** key.
2. Next sequentially press the keys **"CTRL"**, **"ALT"**, and **"DELETE."**
3. Then press **"FUNC"** followed by **"CAP LK"**.
   Capital letters make reading of the display easier. When the 16 by 21 character display of the 756 appears as shown below do steps 4. and 5.



```
Telxon Corp 1988
PTC-750
Version 1.1 US

B>path=a:\;b:\

B>prompt$p$g
B:\>
B:\>A:

A:\>VEXEC
```

FIGURE 7.6: PTVC Display

4. Type **VEXEC** at the A:\ prompt and then press **RETURN**.

   Note: VEXEC is a batch file which loads SW_POWER, PWR_ON, TTS, RECOG, RECOVER, and SAY/" ~2 HELLO. SW_POWER is a program that allows PWR_ON or PWR_OFF to be executed at any time. PWR_ON turns power on to the voice circuitry. TTS is the text-to-speech program. RECOG is the voice recognition program. The program RECOVER is used to reinstate the recognition Gain and Offset of the digital to analog converter when the 756 is shut OFF and later turned back ON. The SAY/" ~2 HELLO is a command which sets the audio output level to the value ~2 (range ~0 to ~9) and informs the user the system is ready for use by saying the word

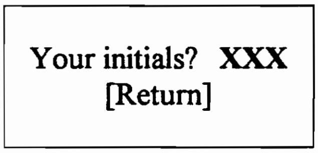
"HELLO." The audio output level will remain at this level unless changed to a new level by the SAY command.

5. Next, type **BAR.BAT** to load the pencil wand.
6. Next type **VUP** at the prompt. This loads the Voice Utility Program (VUP).

Step 2: Importing a Vocabulary

1. Select the Voice Training option, **1**.
2. Next enter the vocabulary file name, **BARDEMO**. Then you will be asked if you have ever trained this vocabulary before. Type **N,** since this is the first time you have trained the vocabulary.

3. When the display prompts as shown in FIGURE 7.7 enter **your three initials,** symbolized in the figure as **XXX**. Following training, a voice training pattern file for the BARDEMO vocabulary will be automatically saved to the Edisk. This file name will be called **BARDEMO.XXX** where **XXX** are the initials of the person who trained the vocabulary.
   Note: Make sure each group member uses different initials. If two people use the same initials the older data will be lost!



Your initials? **XXX**
[Return]

FIGURE 7.7: Entering your Initials

Step 3: Training a Vocabulary

Say each prompted word in a firm natural consistent voice. If the display screen flashes and the same word appears repeat the word until a new word appears, but be consistent in the manner you repeat each word. Do not shout at or question the system while training. When a short phrase, such as "Clear Screen" is prompted say the phrase in a connected fashion, do not pause between words. During training, relax and try to keep a fairly constant volume and speaking rate. Try to use a posture similar to that you will use when operating the system, for example, standing or sitting.

A minimum of 3 training passes will allow fair recognition if used over a short time span. For good recognition and stability over an extended time, 5 to 7 training passes are recommended. For this exercise you will make 7 passes.

- To train the vocabulary

1. Having pressed the **"Enter" key** the 756 screen displays:

```
VOCABULARY TRAINING
  # Times Trained: 0
   Say Word No. 2:
     Clear Screen
 Previous Word Score:
         128
 Press [Esc] to Abort
```

FIGURE 7.8: Vocabulary Training

2. **Say each prompted word**. Each time the screen request, shown below, appears your voice pattern file is automatically updated and saved to the Edisk.

3. **After each training pass**, the screen will display as shown below and requests updating the existing training patterns. Enter as shown below:

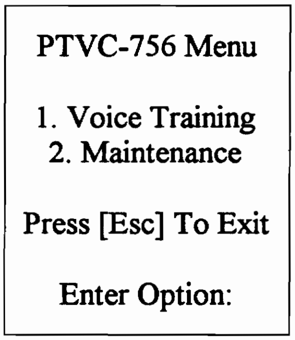
```
VOCABULARY TRAINING
  # Times Trained: 0
  Say Word No. 45:
 Attention Computer
Previous Word Score:
        128
Press [Esc] to Abort

 Wish to Update? Y
```

FIGURE 7.9: Updating Training Patterns

4. After 7 training passes, when the screen displays FIGURE 7.9, press the **"ESCAPE"** key to abort vocabulary training

5. When the main menu of the Voice Utility Program reappears, as shown below, press the **"ESCAPE" key** again. This should give you the A:\ prompt.



PTVC-756 Menu

1. Voice Training
2. Maintenance

Press [Esc] To Exit

Enter Option:

FIGURE 7.10: Main Menu Reappears

Step 4: Activating Voice Recognition
The fun begins. Now that the vocabulary is trained, you can put the power of voice to work!

- To activate voice recognition

1. Press the **"FUNC"** and then **"SCR LK" keys**. **One beep** will be heard to indicate recognition is active. Press the "FUNC" and "SCR LK" keys again and **two "Beeps"** will be heard indicating recognition is off.

2. Use the print-out of the BARDEMO vocabulary (included in your handout) to view word list. Speak those words into the microphone. The output should be displayed on the PTVC-756.

3. State the word **Barcode** to activate the subvocabularies barcode, digits, and common. The PTVC will now prompt you for the item to be entered. This data can be inputted by keyboard, voice, or by using the pencil wand (Note: the bar code labels to be scanned are included in your handout).

   eg. ***using the keyboard***
   Type in **"1510000070"** and press **"Enter."** The system will repeat the numbers entered, state what the item is, and then tell you how much the item costs. Next the system will ask for the quantity that you want. Since the "digits" subvocabulary is already active you can enter the number via keyboard. Press **"1"** and then press **"Enter."**

   using the pencil wand
   Scan the barcode using the pencil wand. Since the "digits" subvocabulary is active you can enter the quantity by keyboard or by voice. For example, state **"1"** or press **"1"** and then state or press **"Enter."**

using voice
Say **"Fetuccini"** and the PTVC-756 automatically repeats the UPC number code. Then enter the quantity via voice. For example, state **"54"** (five then four).

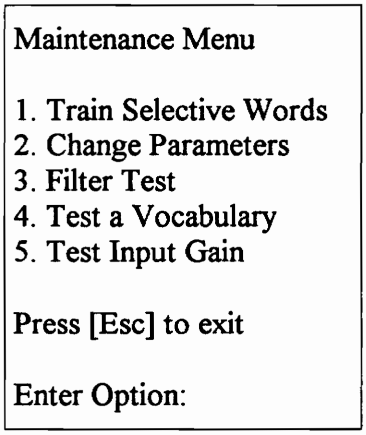
4. Try other examples using the voice, pencil wand, keyboard, or a combination of these commands.

   Note: at this time it may be a good idea to read the lab write-up, since there are specific questions regarding this portion of the lab activity.

Step 8: Improving the Accuracy of the Portable Transactions Voice Computer

There may be accuracy problems with the Portable Transactions Voice Computer. For example, words that sound similar are sometimes confused with one another. There are several parameters that can be changed to try and improve the accuracy rate of the system.

When the Voice Utility Program is loaded, select **option 2** to obtain the Maintenance Menu, as shown below. This will display the following menu:



```
Maintenance Menu

1. Train Selective Words
2. Change Parameters
3. Filter Test
4. Test a Vocabulary
5. Test Input Gain

Press [Esc] to exit

Enter Option:
```

FIGURE 7.11: Maintenance Menu

Select Menu Option #1: Training Selective Words

Selective (single) word training clears (erases) the selected reference word and then reprompts you to train the same number of training passes as you have trained the other words in the vocabulary.

Choose a single word or pair of words that the system is confusing and retrain the system. A discussion question is included in the Lab Write-up.

- To Train Selective Words

1. From the maintenance menu, select **1**
2. The program requests the application **<filename>**. Type the name of the file containing the vocabulary that you want to modify.
3. When the program requests your initials, **type your initials** exactly as entered when you trained the vocabulary.
4. Use **Space Bar** to scroll through the vocabulary until you find the word you want to retrain.
5. Press **Return**
6. **Say the word** clearly into the microphone
7. When a word is sufficiently trained, the reference pattern is saved and the initial training screen returns. Next train another word or press ESCAPE and return to the maintenance menu.

Select Menu Option #5: Test Input Gain
Select the Test Input Gain, menu option 5. The display prompts:

```
Say: '1234567890'
Then Press [ESC]
  To Continue_
```

FIGURE 7.12: Test Input Gain

After saying 1 through 0 press **Escape** to view the peak values and a graph of these peaks for the prior spoken data sequence. All 16 peaks, one for each filter, are listed and graphed.

On pressing **Escape** a second time the mean energy detected by the 16 filters and the maximum peak energy value over the entire spoken sequence is computed. The display with typical values is shown below.

```
Mean Filter Value:
96
Peak Filter Value:
158

[ESC] to Exit:

Press [ENTER] to
Cont. Gain Test_
```

FIGURE 7.13: Filter Values

Write these numbers down. A discussion question is included in the Lab Write-up. The Test Input Gain should yield a Mean Filter Value typically between 80 and 120.

Select Menu Option #2: Change Parameters

1. From the maintenance menu select **2**.
2. Enter the password **!@# [ENTER]**
3. The following parameters will then be displayed:

```
SET PARAMETERS

T1 16-64  32
T2 16-64  24
Gain 1-255  225
ETHL 8-32  28
MINSM 8-24  18
Noi. T. 2-16  2
RJCT T. 1-125  108
DLTA T. 0-15  0

Press [Return] to
Leave Unchanged
Press [Esc] to Return to
Main Menu
```

FIGURE 7.14: Change Parameters

The only parameter to be changed in this lab activity is the **Audio Input Gain**. The Audio Input Gain settings may be modified to best suit to your voice intensity, microphone type and background noise level. For example, in a noisy environment, you normally speak louder thus you should use a lower Audio Input Gain to minimize background noise input.

Experiment with increasing and decreasing the Gain. Note any difference in speech recognition. A discussion question is included in the Lab Write-up.

Note: It is recommended that **only** Reject Threshold, Delta Reject Threshold and Audio Input Gain be changed. In this exercise, only the Audio Input Gain is changed.

7.4.4 PTVC Lab Write-up

Portable Voice Transactions Computer
PTVC-756
Lab Write-up

1. Why does the word score start low and then increase with the number of training passes?

2. What is the price of broccoli soup? What is the price of two pounds of oranges?

3. Was entering data via voice, pencil wand, or by keypad most effective/efficient? Why? In what specific situations would you select data entry by voice? keypad? bar-code scanning?

4. What word(s) did you choose to retrain? Why? What were the results?

5. What was your mean filter value and peak filter value? Are they within the acceptable range?

6. What were your results from increasing and decreasing the gain? What gain gave the best overall accuracy? Why? What outside factors (if any) were involved?

7. How accurate was voice recognition in conjunction with the bar-code scanner?

8. Discuss some advantages and limitations of using voice recognition for data collection.

Chapter 8: RESULTS AND CONCLUSIONS

8.1 Summary

The purpose of this project was to become familiar with the voice recognition systems in the Automatic Data Collection Systems Laboratory at Virginia Tech. The objectives of the project included doing a literature survey on voice recognition systems, setting up the equipment in the ADCSL, and designing laboratory activities for undergraduates in industrial engineering.

After completing this project it can be concluded that voice recognition technology has not progressed to the point necessary for wide spread implementation of voice recognition systems. However, voice systems have been successful for niche applications. These systems typically use limited vocabularies for specific tasks.

For an in depth discussion of the problems that need to be overcome in order for wide spread implementation of voice recognition systems to become reality, see Chapter 4, Section 4: Current Problems with Voice Recognition Systems.

8.2 Latest Voice Recognition Systems!?

This section describes two of the newest voice recognition systems on the market, including IBM's VoiceType Dictation System and the Macintosh's Quadra 840AV Speech Recognition Software.

The latest issue (March, 1995) of the magazine Windows has an article entitled, “Sluggish Speech Recognition.” The article describes the VoiceType Dictation System, IBM's newest voice recognition / dictation package. IBM's newest system recognizes only isolated words, rather than word groups, making the user pause unnaturally between each word. The author of the article concludes that "Anyone who can type well should be able to key in text faster than VoiceType Dictation can recognize it" (Nicolaisen, p.40). This system costs $999. This is just one example that proves that voice recognition systems have not developed to the point necessary for wide spread implementation to occur.

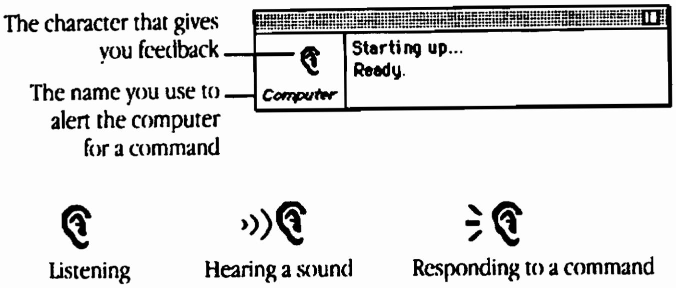
Macintosh has introduced the Quadra 840AV, which comes standard with speech recognition software that enables it to recognize and execute certain spoken commands.

Data can only be inputted discretely by key words or phrases. Phrases are spoken in connected fashion.

The status window of the speech recognition software has three parts:

1. An animated drawing that represents what the speech-recognition software is doing.
2. The name you should use to address the computer (the default name is "Computer").
3. Text that displays what the computer hears and how it is responding.

When the status window indicates that the computer is ready to receive , each command is started with the name "Computer." (see Figure 8.1) A complete listing of the voice command capabilities of the Quadra 840AV is shown in Appendix II.



Starting up...
Ready.

The character that gives you feedback
The name you use to alert the computer for a command
Computer
Listening
Hearing a sound
Responding to a command

FIGURE 8.1: Status Commands for the Quadra 840 AV

The voice recognition systems in the Automatic Data Storage Collection Laboratory, including the Micro IntroVoice, IntroVoice VI, Portable Transactions Voice Computer, and the WinVoice software program, are not the latest and greatest systems out on the market; but are newer systems that much better? The newest systems would be nice to have, but they would not provide the significant increase in accuracy rate needed to justify their purchase, as proven by the discussion in this section. Also, new equipment would not provide students with any more instruction in using voice recognition systems, which would be the main purpose for their purchase. However, if research needed to be done on voice recognition, it is recommended that newer equipment be requisitioned.

8.3 Strengths and Weaknesses of Labs

This section discusses the strengths and weaknesses of each lab. As described in this section, Labs #1 and #4 are the best overall labs.

Lab #1 (Micro IntroVoice)

Strengths- best overall lab (along with Lab #4), uses one of the newest voice systems in the ADCSL, contains a section on text-to-speech synthesis, user friendly (with pull-down windows), and produces a high accuracy rate.

Weaknesses- can only be used in a DOS environment.

Lab #2 (WinVoice)

Strengths- uses the same control program as the Micro IntroVoice (same comments on being user friendly and having a high accuracy rate), applied in Windows based programs.

Weaknesses- long training period (the lab has 62 vocabulary words).

Lab #3 (IntroVoice VI)

Strengths- lab includes a graphical demonstration (eg. a cursor moves up, down, left, and right via voice command).
(this lab would be best utilized for lab tours of the Automatic Data Collection Systems Laboratory to demonstrate the use of voice recognition systems)

Weaknesses- poor accuracy, difficult to train, not user friendly.

Lab #4 (PTVC-756)

Strengths- best overall lab (along with Lab #1), uses one of the newest voice systems in the ADCSL, showcases a portable system, used in conjunction with another automatic data collection device (bar-code wand), user friendly, and produces a fairly high accuracy rate.

Weaknesses- control program is the same as the IntroVoice VI control program.

8.4 Comparison of Voice Recognition Systems

The four systems utilize either the Micro IntroVoice Program (Micro IntroVoice) or the Voice Utility Program (IntroVoice VI) to create, edit, and train vocabularies. This is shown in Table 8.1. Also, shown in the table is the first development date of each system along with the date of the most current version displayed in the Automatic Data Collection Systems Laboratory at Virginia Tech.

Table 8.1: Dates of Development

System	Control Program	First Version	Current Version
Micro IntroVoice	MIV	1992-1993	1992-1993
WinVoice	MIV	1992-1993	1992-1993
IntroVoice VI	VUP	1986	1990 revision 2.0
PTVC-756	VUP	1988	1989 revision 5.42

As discussed in the previous section, there are distinct strengths and weaknesses of each system. The age of the system is one important factor to consider. In general, the older systems had a lower accuracy rate and were not as user friendly as the newer systems. Thus, the age of the system, which most likely corresponds to the technology utilized, is a factor that can not be overlooked. The next section gives a more in depth discussion of the differences between the Micro IntroVoice and IntroVoice VI systems.

8.5 Micro IntroVoice vs. IntroVoice VI

The Micro IntroVoice has its own 8 MHz. V-25 microcomputer with 128 kilobytes of battery backed RAM (expandable to 512 kb) and 32 kilobytes of EPROM (expandable to 128 kb) and in addition, all of the audio spectrum analysis and synthesis electronics necessary for voice recognition and text-to-speech synthesis.

These hardware features with new and advanced firmware allow Micro IntroVoice to extend beyond the capabilities of its forerunner, the PC plug-in IntroVoice VI board. A brief comparison is shown below.

TABLE 8.2: Comparison of the IntroVoice VI and the Micro IntroVoice

Parameters	IntroVoice VI	Micro IntroVoice
Vocabulary Size	500 words	1,000 words
Active Vocabulary Size	250 words	1,000 words
Key Replacement String Size	28 kb max.	40 kb
Operates with Any Computer	No	Yes
Operable with most Terminals	No	Yes
Standalone Operable	No	Yes

8.6 Accuracy of the Systems

The two control programs were tested for their accuracy rates. As shown in the table below these systems have reasonably good accuracy rates, but what is good enough? In order for voice recognition systems to become more widely used, they must have at least a 95% accuracy rate. A higher accuracy rate along with an increase in speed are two functions that must be improved before people will begin using voice recognition systems for more than niche applications.

TABLE 8.3: Accuracy of the Control Programs

Control Program	Sample Size	# of Correct Responses	Accuracy Rate
Micro IntroVoice	2000	1715	85.75%
IntroVoice VI	2000	1507	75.35%

8.7 Conclusions

The future of voice recognition systems looks very promising. Unfortunately the future is not now. Before wide spread implementation of voice systems can become a reality current problems, including dealing with different accents and dialects, continuous voice input, co-articulation, and word spotting (see Chapter 3.4), need to be addressed. System accuracy and reliability must also be improved. These points were well proven by problems that arose while attempting to demonstrate the voice systems during the defense of this project. For these reasons, it is recommended that voice recognition systems not be sought after as an "end all" solution to ones data collection needs. A thorough evaluation needs to be made before deciding to purchase any voice recognition system. Reading this project report to gain a better understanding of the strengths and weaknesses of voice systems is a step in the right direction.

Even though voice recognition technology has not been perfected, it is still important for undergraduate industrial engineers to understand the basics of voice recognition technology. The purpose of this project was to explore the voice recognition process and to provide students with some hands on exercises to gain practical experience in using voice recognition systems. It should be noted that all four labs do not need to be completed to gain this understanding. A lab was designed for each voice system to showcase the capabilities of each system. As discussed in Chapter 8.3, if only one lab is to be completed, either Labs #1 or Lab #4 is recommended. After completing a lab exercise, not only should students come away with the basics of voice recognition, but they should also have enough insight to see their potential for use in future applications involving automatic data collection. For now voice recognition systems will probably still be used mainly for niche applications, but who knows what the future may bring!

"Computer, I command you to turn off."
[Computer off]

REFERENCES

Ackley, H.S. (1991). Bar Code Symbol Quality Control. SCAN-TECH 91, AIM USA, Everett, WA.

AIM USA Publication (1991). Voice-Based Data Collection. AIM USA. Pittsburgh, PA..

Dambrot, S.M. (1992). Verbex speaks to Japan- through Mitsubishi. Electronics, p.5-6.

Edgar, B. (1994). PC-Based Voice Processing. New York: Flatiron Publishing.

Etter, B.D. (1989). Voice Controls for Manufacturing Environments. Manufacturing Review, p.242-249.

Foster, P. (1992). Speech Recognition. Chelsea, MI: Bookcrafters.

Hemphill, D. (1991). Saturn Paint Quality Inspection by Voice. SCAN-TECH 91 Proceedings, AIM USA, Pittsburgh PA.

Jesitus, J. (1994). Voice Recognition Eases Auto Auctions. Automatic I.D. News, p.18.

Klass, P.J. (1992). Military, Civil Applications Seen in USAF Voice Identification System. Information Technology, p.67.

Librescu, J. (1994). Bar Code Data Collection System Implementation and Laboratory Exercise, Virginia Tech Project.

Labriola, D. (1995). Straight Talk. Windows Sources, p.144-160.

Lee, K. (1990). The Spoken Word. Byte, p.225-232.

Meer, A.T. (1994). Car manufacturer rolls out quality information management system. Automatic I.D. News, p.22.

Meisel, W.S. (1993). Talking to your Computer. Byte, p.113-120.

Miller, M.J. (1994). Conversations with My PC. PC Magazine, p.79-80.

Nicolaisen, N. (1995). Sluggish Speech Recognition. PC Magazine, p.40.

Plecko, E. (1991).Radiation Dosimetry Badge Inspection by Voice. SCAN-TECH 91 Proceedings, AIM USA, Pittsburgh PA.

Robinson, G.M. (1992). Harnessing the Power of Speech. Design News, p.19-20.

Rudnicky, A.I. (1994). Survey of Current Speech Technology. Communications of the ACM, p. 52-57.

Salicce, R.L. (1991). Voice Data Collection- Today!. SCAN-TECH 91 Proceedings, AIM USA, Pittsburgh PA.

Schwind, G.F. (1994). Voice Recogniton Deserves a Second Listen. Material Handling Engineering, p.63-67.

Schwind, G.F. (1988). Voice data entry: identification by conversation. Material Handling Engineering, p.69-73.

Stovicek, D. (1991). Voice I/O in Manufacturing. Automation, p.26-27.

Subrata, D. (1002). The Power of Speech. Byte, p.151-160.

APPENDIX I: List of Phonemes

LIST OF PHONETIC SYMBOLS

The following is a list of the phonetic symbols used in this book, with short particulars of their meanings. It is to be understood that most symbols have at times values deviating from what may be considered to be their most usual values.

(1) *Consonant Letters of the Roman Alphabet.*

p, b, t, d, k, g (hard), m, n, l, f, h have their customary values.

The others are as follows:

c breathed palatal plosive; also used when convenient to denote the affricate tʃ (Eng. *ch*).

j voiced palatal fricative; also the corresponding semi-vowel (Eng. *y* in *yet*).

q breathed uvular plosive.

r stands for the various r-sounds of different languages (Eng., Fr., Ger., etc.); replaced by other symbols, ɹ, ɾ, ʀ, when necessary.

s as in Eng. *see*.

v as in Eng. *ever*.

w as in Eng. *well*.

z as in Eng. *lazy*.

(2) *Other Consonant Letters.*

ʈ, ɖ retroflex plosives.

ƫ, ᶊ, etc., palatalized t, s, etc.; represented in digraphic transcription by tj, sj, etc.

ᵵ velarized t; also used for pharyngalized t.

ɟ voiced palatal plosive; also used when convenient to denote the affricate dʒ (Eng. *j*).

ʔ glottal stop.

ʔ Cairene Arabic glottal stop, when corresponding to classical q.

ɱ labio-dental nasal.

ɳ retroflex nasal.

ɲ palatal nasal (Fr. *gn*).

ŋ velar nasal (Eng. *ng*).

ƞ Japanese syllabic nasal.

ɬ breathed l (Welsh *ll*).

ɫ velarized (dark) l.

ɺ a sound intermediate between d and l.

ʎ palatal lateral (Ital. *gl*).

ɾ single flap tongue-tip r.

ɽ retroflex flap.

ʀ rolled or flapped uvular r (one variety of Parisian *r*).

ɸ, β bi-labial fricatives.

θ, ð dental fricatives (Eng. *th*-sounds in *thing, then*).

ɹ fricative tongue-tip r; also the corresponding frictionless continuant, and a retroflexed variety of this; also r-coloured ə.

ʃ, ʒ palato-alveolar fricatives (Eng. *sh*, Fr. *j*).

ᵴ velarized s; also used for pharyngalized s.

ç breathed palatal fricative (one variety of the Ger. *ich*-sound).

ɕ, ʑ alveolo-palatal fricatives (Polish *ś*, *ź*).

ɣ voiced velar fricative.

ħ breathed pharyngal fricative.

ʕ voiced creaky sound made with contracted larynx and pharynx (Arabic ‘ain).

ɦ voiced h.

ɥ consonantal y (= y̆).

ʇ dental click.

ʖ lateral click.

tʃ, dʒ, ts, etc. affricates (one mode of representation).

ph, th, etc. aspirated plosives.

(3) *Vowel Letters.*

i as in Fr. *si*, and sounds near to this; used also when convenient for the Eng. short *i* as in *sit*.

e as Fr. *é*, and shades of sound near to this; also used in place of ɛ when possible, e.g. in transcribing the Eng. vowel in *set*.

ɛ as Fr. *è*, and sounds near to this; sometimes used for the Eng. vowel in *set*.

a as in Parisian Fr. *là*, and sounds near to this; also used in the Simplified Transcription of the Eng. short vowel in *hat* and the long vowel in *half*.

ɑ as in Parisian Fr. *las*, and sounds near to this; used for the Eng. long vowel in *half* in narrowed transcription.

ɔ as *o* in Fr. *porte*, and sounds near to this; used in narrow transcriptions of Eng. *saw* (long), *hot* (short), Ger. *Sonne*, etc.

o as in Fr. *beau*, and sounds near to this; replaces ɔ in broad transcriptions of Eng. *saw*, *hot*, Ger. *Sonne*, etc.

u as in Ger. *Schuh*, and sounds near to this, e.g. the vowels in Fr. *coup*, Eng. *too*; used also when convenient for the Eng. short *u* in *put*, *book*.

y close lip-rounded i, as Fr. *u*, Ger. *ü*.

ø close lip-rounded e, and sounds near to it, e.g. the Fr. vowel in *peu*.

œ open lip-rounded ɛ, and sounds near to it, e.g. the Fr. vowel in *œuf*.

ɒ open lip-rounded ɑ; used for Eng. vowel in *hot* in narrow transcription.

ʌ unrounded ɔ; also used for Eng. vowel in *cup*.

ɤ unrounded close o.

ɯ unrounded u.

ɨ vowel intermediate between i and ɯ.

ʉ vowel intermediate between y and u (= a lip-rounded ɨ).

ɩ a lowered and retracted variety of i; used in Tswana, and in transcribing Scottish and American pronunciation of *sit*, etc.; also in narrow transcription of Southern British pronunciation of such words.

ᵻ a lowered variety of ɨ.

ɷ a lowered variety of u, or a very close variety of o; used in Tswana, and for the Igbo ɵ, also used in transcribing American pronunciation of *book*, etc., and in narrow transcription of Southern British pronunciation of such words.

ʏ a lowered and retracted variety of y; used for Ger. short y in narrow transcription.

ə unrounded central vowel (schwa), as Eng. *a* in *along*.

ɐ a lower variety of central vowel, as in Eng. *sofa* or in Lisbon Portuguese *para*; also sometimes used to denote the quality of long əː in narrow transcriptions of Southern English.

æ a raised a or a very open ɛ; used for Eng. short a in narrow transcription.

a˞, ɔ˞, r-coloured a, ɔ.

ə˞ r-coloured ə; also represented by ɹ or ɚ.

An index letter means that the sound is that of the main letter modified in the direction of that indicated by the index. Thus h^{ç} means a ç-like variety of h, and ʒ^{z} denotes a sound intermediate between ʒ and z.

(4) *Diacritic Marks.*

˜ nasalization; ɛ̃ = nasalized ɛ.

̥ devoicing; n̥, l̥, ʒ̥ = unvoiced n, l, ʒ.

̬ voicing; s̬ = z, t̬ = American "voiced t."

̣ close variety; ẹ = a very close e, ạ = æ.

̨ open variety; ę = a sound between French *é* and *è*.

˒ lips more rounded.

˓ lips more spread.

+ advanced variety; u+ or u̟ = a sound between u and ʉ.

- or ̠ retracted variety; a- or a̠ = a sound between a and ɑ.

˔ raised variety; a˔ or a̝ = ạ = æ.

˕ lowered variety; e˕ or e̞ = ę.

¨ central vowel; ü = ʉ, ë = a high variety of ə.

ʻ slight aspiration after p, t, etc.

ʼ glottal stop accompanying p, t, etc.; glottal contraction accompanying continuant sounds, Danish *stød*.

˘ indication that a vowel is consonantal: y̆ = ɥ; also used to mean that a continuant sound is very short, as m̆ in Sinhalese m̆b, ĕ in Tswana ĕò.

̩ under a letter (or over it if the letter has a tail below) means that the sound is syllabic; n̩ = syllabic n.

⁀ simultaneous pronunciation of two sounds, e.g. Provençal m͡ŋ.

ː length mark.

ˑ half length.

ˈ at the beginning of a syllable denotes strong stress.

ˈˈ at the beginning of a syllable denotes extra strong stress.

ˌ at the beginning of a syllable denotes medium (secondary) stress.

ˉ (thus ā or ˉa) high level tone.

¬ in Burmese, a level tone ending with a slight fall and pronounced with creaky voice; in Tswana a high level tone requiring that the next succeeding high tone shall be slightly lower.

ˍ (thus a̱ or ˍa) low level tone; in Vietnamese, low tone combined with a creaky voice.

ˊ (thus á or ˊa) high rising tone, or rising tone without implication of height.

ˏ (thus a̗ or ˏa) low rising tone.

ˋ (thus à or ˋa) high falling tone.

ˎ (thus a̖ or ˎa) low falling tone; in Vietnamese, low tone combined with breathy voice.

ˆ (thus â or ˆa) rising-falling tone.

‹ Panjabi low rising tone.

▾ (preceding the syllable) Vietnamese rising tone combined with creaky voice.

ᵕ (preceding the syllable) Vietnamese rising tone combined with breathy voice.

APPENDIX II: Voice Command Functions of the Macintosh Quadra 840AV

Voice Command Functions of the Macintosh Quadra 840AV

What you can say
* Hello.
When you can say it: **anytime.**
What it does: **the computer replies, "Hello, welcome to Macintosh."**

* What time is it?
When you can say it: **anytime.**
What it does: **the computer replies, the computer tells you the time.**

* Zoom window
When you can say it: **anytime.**
What it does: **the computer replies, this command has the same effect as clicking the zoom box in the active window.**

* What day is it?
When you can say it: **anytime.**
What it does: **the computer replies, the computer tells you the date.**

* Close all windows
When you can say it: **anytime.**
What it does: **this command has the same effect as clocking the close box in the active window with the option key held down.**

* Print...copies
* Print from...to...
* Print page...
* Print pages...to...
When you can say it: **anytime.**
What it does: **the computer prints the specified number of copies or pages of the active document (when the Finder is active, the computer prints the selected document).**

* Is file sharing on?
When you can say it: **when the Finder is active.**
What it does: **the computer opens the Sharing Setup control panel and depending on the current state of file sharing, the computer says, "File sharing is on," or "File sharing is off," or "File sharing is starting up."**

* Start file sharing

* Stop file sharing
When you can say it: **when the Finder is active.**
What it does: **the computer opens the sharing setup control panel, clicks the appropriate button, and closes the control panel.**

* Open...[any item in the Apple menu]
When you can say it: **anytime.**
What it does: **the computer opens the item.**

* Switch to...[any open program]
When you can say it: **anytime.**
What it does: **the computer makes the open program active.**

* You can say any menu command
When you can say it: **anytime.**
What it does: **the computer performs the menu command.**

* You can say the name of any button in a dialog box
When you can say it: **anytime.**
What it does: **the computer clicks the button.**

* You can say the name of any item in the speakable items folder
When you can say it: **anytime.**
What it does: **the computer opens the item.**

* Restart
* Shut Down
When you can say it: **when the computer Finder is active.**
What it does: **the computer asks, "Are you sure you want to restart?" or "Are you sure you want to shut down?" and waits for a reply.**