ACHIEVING A HANDS-FREE COMPUTER INTERFACE USING VOICE RECOGNITION AND SPEECH SYNTHESIS

James R. Evans
Software Engineering Division (TISAD)
Ogden Air Logistics Center
Hill Air Force Base, Utah 84056
Phone 801-777-4527
Evansj@software.hill.af.mil or
jevans2@sisna.com

Wayne A. Tjoland and Lloyd G. Allred
Software Engineering Division (TISAD)
Ogden Air Logistics Center
Hill Air Force Base, Utah 84056

Abstract - The software designers at Hill Air Force Base have developed a voice recognition and speech synthesis system (Voice Control) for use with the F-16 Analog Test Station Sustainment (FATSS) project. The Voice Control system is reliable, speaker independent, and has a total added hardware price tag of under $50.00 per station. In contrast to traditional voice recognition systems, operator training is not required. OO-ALC has developed a general-purpose internal interface (Voice Control) to the Speech Recognition and Text-To-Speech engines provided by Microsoft. Voice Control can be accessed by any 32-bit Windows software, which has windows messaging capability. This is available to standard programming languages such as LabWindows CVI, Borland or Microsoft C/C++, Visual Basic, or even commercial packages such as Lotus Notes, and Microsoft Word. Through Voice Control, the computer uses both video and voice prompts to request input from the operator. The operator is allowed to enter data and to control the software flow by voice command or from the keyboard or mouse. The Voice Control system allows for dynamic specification of a grammar set, or legal set of commands. The use of a reduced grammar set greatly increases recognition accuracy. The computer voice enables the operator to focus his attention away from the computer screen, which is required for activities such as probing a circuit card and taking readings. When the operator takes readings, the computer, to insure reliable entry, echoes his voice entries. With electronic tuning, speech synthesis allows the operator to hear the resulting reading, enabling him to focus on the circuit card instead of the constantly turning his head to see the computer screen. This paper describes the capability and functionality of the Voice Control system.

INTRODUCTION

As part of the FATSS upgrade to the Honeywell 2600, OO-ALC/TISAD has devised a method for integrating voice recognition into the system. When both of an operators hands are occupied in testing and probing an electronic circuit card, it can be extremely annoying to put down the probe and disconnect the ground strap in order to enter data or to operate the computer and/or review the computer response. The process becomes more tedious when a circuit needs to be tuned in order to achieve a required specification. Using a combination of voice recognition and speech synthesis, OO-ALC has achieved a hands-free computer interface which allows the technician to probe and tune electronic circuit cards without having to remove his hands from the circuit card or to focus his eyes on the computer screen. The **Voice Control** application was devised for use with the Microsoft Speech API version 4.0, which can be obtained for free over the Internet from Microsoft. The system is very reliable, speaker independent, and has a total price tag of under $50. In contrast to traditional voice recognition systems, no training is required. The system uses a limited set of grammar, or legitimate commands, providing a restricted context, thus enabling very high reliability. To prevent inadvertent errors, verification from the user is required on crucial commands.

OPERATION

The operator chooses to activate the Voice Control application from a switch on the Test Executive main screen. Once activated, **Voice Control** starts off in "Sleep" mode. In this mode the only commands that the computer will recognize are

COMPUTER WAKE UP
COMPUTER SLEEP
SHOW GRAMMAR

This is the default grammar that will always be recognized by **Voice Control**. When the operator issues the command for the computer to wake up, additional commands are added to the active grammar set. The operator can then choose from the set of legal commands based on the current situation. The standard commands used on the FATSS station are:

    BEGIN TEST
    ADVANCE
    GO
    NOGO
    LOCAL PRINT
    TEST MODE 1
    TEST MODE 2
    TEST MODE 3
    TEST MODE 4
    STATION CLEAR.
    COMPUTER WAKE UP
    COMPUTER SLEEP
    SHOW GRAMMAR

When the ATLAS is running, and a different response is needed from the operator, the grammar is altered to allow only valid commands for the current situation. An example of this would be if the ATLAS was expecting a numerical answer. It this situation, the valid commands would be:

    INTEGERS
    COMPUTER WAKE UP
    COMPUTER SLEEP
    SHOW GRAMMAR

Many visitors are surprised to see this system regularly demonstrate perfect numerical recognition even for a first time user. Once the valid integer is entered, either by voice command, or from the keyboard, the grammar is be changed back to the standard commands. Crucial commands such as **STATION CLEAR** may require verification to prevent an abnormal abortion of a test. Under such circumstances, **YES** and **NO** would become the only legal commands available, preventing inadvertent errors. Because of the limited list of commands, and the fact that the grammar set is context sensitive, high reliability of voice recognition can be achieved.

INTERFACE DESCRIPTION

The OO-ALC voice interface (**Voice Control**) consists of a System Tray application, which is launched by a switch on the Test Executive main panel. The operator can enable or disable the application by clicking on the **Voice Control** icon in the system tray with the mouse. The operator can also restore the Voice Control application window to normal size by right clicking the icon in the system tray. From the Voice Control Window the operator has access to application settings and to the Microphone Setup Wizard. This wizard should be run when a new operator begins using the system, or when the amount of background noise in the area changes. The Voice Control Window may be seen in Figure 1.
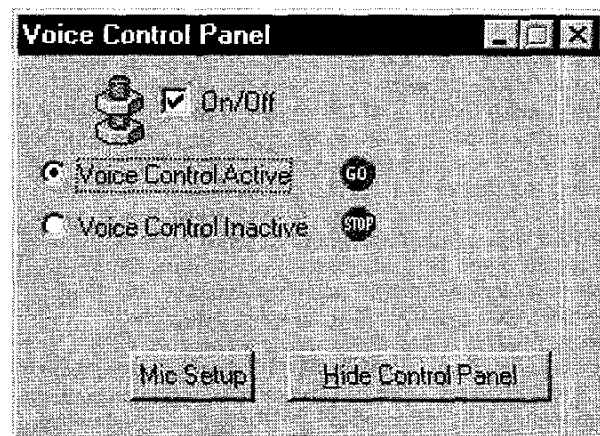

Figure 1.

When the operator selects HIDE CONTROL PANEL the window is minimized once again to the system tray.

The voice processing is done in the background. When a legitimate command is recognized, a windows message is sent to the Test Executive. Because Window's software is event driven, this enables the user to enter data through traditional means (mouse or keystroke) when convenient. The choice of grammar is controlled from the Test Executive to allow only meaningful commands based on the situation. The command, **ADVANCE**, would not be accepted when a **GO** or a **NOGO** response is required. Modular programming allows for a single windows message being sent from the Test Executive to the Voice Control to set grammar to commonly expected responses.

SYSTEM FEATURES

During a test, when the operator is tuning a circuit card, the computer voice can be used to echo the readings until an appropriate value is obtained. The computer can also provide voice commands to tell the technician where to probe. When a numerical response is required, the technician can enter values by speaking them or by entering them on the keyboard. The computer can then echoes the values for verification. This allows the operator to operate the test station hands free, while running a TPS that was never intended for operation in a hands free environment.

PORTABILITY

The **Voice Control** application is a 32-bit windows application which opens a session with the Microsoft's Text-To-Speech and Voice Command Recognition engines. The complexity of establishing and maintaining this interface is hidden from the Test Executive. The **Voice Control** system maintains the speech recognition grammar. When the Test Executive needs to modify the grammar, it sends a windows message to the **Voice Control** application. When a phrase from the current grammar is recognized, the **Voice Control** application sends a windows message back to the Test Executive. When the Test Executive needs to issue a verbal command to the user, it sends a "Speak This" windows message to the **Voice Control** application, which then issues the actual request to the Text-To-Speech engine. This scheme is very portable to other windows based applications, and should ease the way to incorporating Voice Recognition and Speech Synthesis into standing ATE applications which run under Windows. The main reason for implementing **Voice Control** as a separate application was the difficulty involved in integrating this capability into the Nation Instruments Test Executive application. The National Instruments CVI C compiler could not link correctly with the Microsoft SAPI libraries, therefore the easiest solution was to create the **Voice Control** as a separate executable and build it with the Microsoft Visual C++ compiler.

Hardware Requirements

The hardware required for this system is readily available and very affordable. A good quality "Noise Canceling" headset microphone is the most important thing. For the FATSS station upgrade, OO-ALC/TIS chose the VXI Parrott 10-3 headset. It has a single earpiece, which is an import safety concern in a shop environment, and an excellent "Noise Canceling" boom microphone. The 10-foot cord is ample for use with the FATSS test station. This headset microphone is available for less than $50.00 making this an excellent choice. The PC soundcard is also an important concern when considering Voice Recognition. A good quality PCI audio card such as the Ensoniq PCI Audio, or the Creative Labs PCI Audio should be sufficient. Both of these cards were tested with good results. Older ISA sound cards, such as the Creative Labs SoundBlaster 16 can be used, but will be less reliable. Also, older sound cards may not work in "Full Duplex" mode, and when this situation is detected, the software automatically switches to a "Half Duplex" mode. Most systems today come with a sound card as standard equipment, so this may not represent an added cost to implementing voice recognition.

CONCLUSION

The **Voice Control** interface developed by OO-ALC/TISAD for use on the FATSS test station is a simple, reliable and low cost solution to integrating Voice Synthesis and Text To Speech capability into Windows based ATE applications.