An Audio-Visual Teaching Aid (AVTA) with Scrolling Display and Speech to Text over the Internet

Davood Khalili, Wan-Young Chung

Division of Internet Engineering, Dongseo University

San 69-1 Churye-2-Dong, Sasang-Ku, Busan, 617-716, Korea

dkhalili@yahoo.com

Abstract

In this Paper, an Audio-Visual Teaching aid (AVTA) for use in a classroom and with Internet is presented. A system, which was designed and tested, consists of a wireless Microphone system, Text to Speech conversion Software, Noise filtering circuit and a Computer. An IBM compatible PC with sound card and Network Interface card and a Web browser and a voice and text messenger service were used to provide slightly delayed text and also voice over the internet for remote learning, while providing scrolling text from a real time lecture in a classroom. The motivation for design of this system, was to aid Korean students who may have difficulty in listening comprehension while have, fairly good reading ability of text

This application of this system is twofold. On one hand it will help the students in a class to view and listen to a lecture, and on the other hand, it will serve as a vehicle for remote access (audio and text) for a classroom lecture. The project provides a simple and low cost solution to remote learning and also allows a student to have access to classroom in emergency situations when the student, can not attend a class. In addition, such system allows the student in capturing a teacher's lecture in audio and text form, without the need to be present in class or having to take many notes. This system will therefore help students in many ways.

I. Introduction

With the advent of major improvements in voice recognition software, it is now possible to design and build systems that can convert speech to text with high degree of accuracy. Some of the available software can reach recognition accuracies near perfection with a learning system and creation and continuous enhancement of a person's voice model. One of the applications and adaptation of voice to text is an Audi-Visual Teaching Aid. The primary goal for implementing this system is to provide visual information in real time so that Korean students (in our case) with poor ability of listening comprehension, may improve their understanding a lecture given by foreign professors with various accents.

The secondary objective is to use speech to text conversion and relay the result along with the text over the internet, for the students who may not be able to attend a class

or take the class under distant learning plan. The transmission of Audio and text is easily possible using the readily available chatting messengers. Since the high quality video transmission in real time would require a system extended bandwidth without any bottlenecks, it is not possible to use video as the primary means of lecture transmission. So, the this project will alleviate the problem and will provide to enhancement of remote learning without additional cost or expensive equipment or the need for high speed internet access.

II. Implementation

For the implementation and design of an AVTA, two phases for the project were considered. The first phase was the study of various voice recognition software, noise cancellation circuits and RF links between wireless microphone and the PC-sound card. The block diagram of the system is shown in Figure 1. This system consists of readily available modules, but further work was needed to implement transfer of the text to parallel port and transmission of data to LCD (eventually to be replaced by LED) display and design of a reliable Wireless microphone.

The wireless microphone was designed and tested and proved to be suitable for the application. The problem with existing modules is the size, and the weight, which are source of inconvenience for a lecturer.

Since the source code of the IBM speech engine was not accessible, the direct transfer of text to a parallel port was not possible, so the interface S/W to send display data to printer port had to be generated.

The speech recognition Engine used in this project is from IBM(VIAVOICE) which is now available for less than $200. The S/W is capable of receiving sound over the USB port, so, as an extension of this project an RF transmitter/receiver system with USB output may be designed to improve system response by reducing latency.

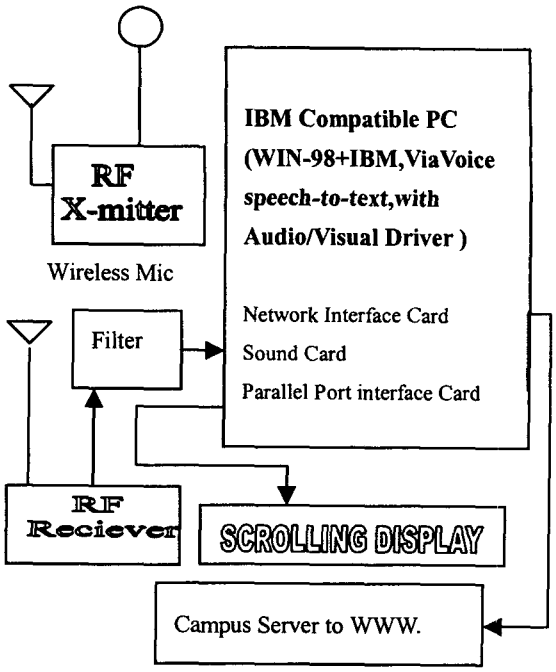

Figure 1. AVTA system

The system considered here, is dependent of individual voice model. This is due to IBM's ViaVoice S/W and its inherent characteristics.

III. Wireless Microphone system

For flexibility of movement of an instructor, a wireless Microphone system was considered. Several wireless systems are available off-the-shelf. The operation of these systems can be 49MHZ, 88-108MHZ, and 900 MHZ. For our experiments we used a standard FM radio receiver with good selectivity operating at a quiet frequency and feeding the output of the receiver to a sound card. A a low noise FM transmitter which was designed is shown in Figure 2. In addition, we tested the system using a 49MHZ system from Tandy Corporation. With slight modification of our FM transmitter we were able to use it at 49 MHZ with the 49 MHZ-receiver. The motivation for our transmitter design was to reduce the size and weight of the wireless microphone. In future by using an 12 volt N-Size

battery, we will reduce the weight significantly

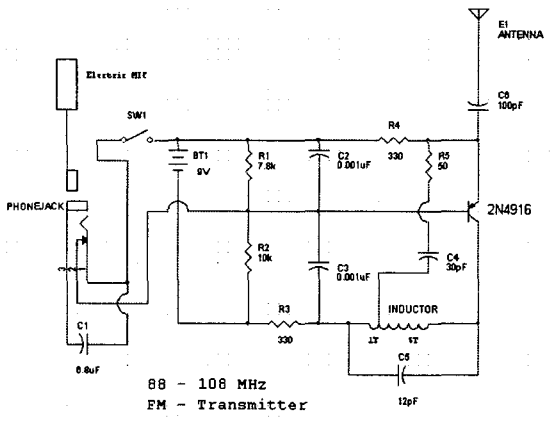

Figure 2. FM transmitter

IV. Accuracy of Voice Conversion

With proper system design and low noise modules, the accuracy of recognition and conversion of the latest IBM S/W can be great but it requires training and creation of an individual's speech model file, which is portable and can be transferred from machine to machine and over time, it can be improved significantly.

V. Lecture(Text+Audio)over Internet

Upon conversion of speech to text the text along with an instructor's voice are relayed over the internet for the students who may not be able to attend a class or take the class under distant learning plan. The transmission of Audio is easily possible using the readily available messengers (e.g. Yahoo, MSN), but since video transmission is real time would require a system extended bandwidth without any bottlenecks, it may not always be possible to use that as the primary means of lecture transmission, so, the **AVTA** will alleviate this problem.Since the existing messengers allow conferencing, transmission of text of speech along with the speech is possible without the need for any additional hardware.

VI. System Requirements for AVTA

The AVTA can be used with any of the following Operating systems Windows 95, 98, ME, XP, NT.

For Windows 95, 98 & ME user, to run this application, these files are needed:

1. SMAPI Engine (IBM) – ViaVoice
2. SpeechRecognition.exe

For Windows NT/2000 and Windows XP user requires installing I/O port device driver – PortTalk. (Download at http://beyondlogic.org/porttalk/porttalk.htm) To run the application, you need following files:

1. SMAPI Engine (IBM) – ViaVoice
2. Speech.bat
3. PortTalk.sys
4. AllowIo.exe

The Hard Disk space requirement for AVTA is 500 Meg Bytes.

VII. Running the Program

The Operation of this AVTA tool is very simple.

For Windows 95, 98 & ME user can double click the **SpeechRecognition.exe** file in the directory **C:\program files\ViaVoice\bin** . For Windows NT/2000 and Windows XP user, can double click the **speech.bat** file from **C:\program files\ViaVoice\bin**

Upon running the program, speech to text conversion can get initiated by voice command. For example, "Dictate to Word" will open a Microsoft Word file and the text of the speech will be displayed in that window.

We can easily use a Internet messenger (e.g., YAHOO-Messenger instead of Word to dictate into.

The latest version of ViaVoice provides USB interface for Microphone. For our application, using a wireless system, we need to do some modification in order to use the USB port. A noise cancellation chip (programmed DSP from Micronics) comes with the IBM supplied S/W and microphone. But the performance of this chip needs to be evaluated further in order to optimize the system, since we use a wireless microphone system as opposed to a headset. The plan is to identify the noise type and level in a classroom setting in view of Echo, and environmental factors and design self tuning DSP if necessary, to enhance the AVTA system performance for error free operation and optimal speed.

The following is a Run-time example:

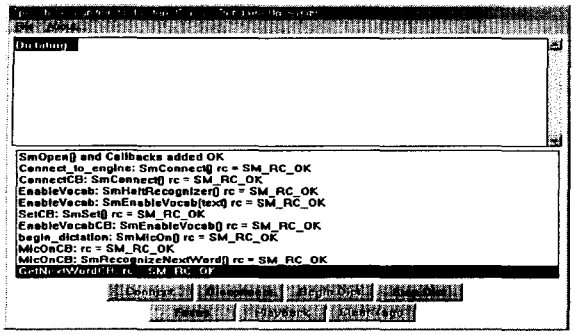

Figure 3

Speech to text conversion on Display (LED) will begin by manual or voice commands (dictate to ..,stop dictation click "Stop Dict". It will return as below diagram.

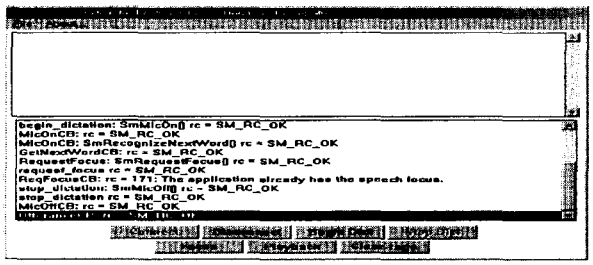

Figure 4. Run time example

VIII. Scrolling Display Unit (LCD Design & Interfacing)

For initial testing of the system, We considered an LCD Module rather than a large scale LED system. IBM Viavoice S/W does not provide the source-code so, we had to send the text from IBM to Printer port by writing additional code. The LCD to Parallel Port printer is straightforward and is described here.

This is the interfacing for the Parallel Port. It is a 20 Character x 2 Line LCD Module to Parallel Port. These LCD Modules very common and simple to work.

The schematic for Parallel port to LCD is shown in Figure 5.

The schematic is quiet simple. The LCD panel's *Enable* and *Register Select* is connected to the Control Port. The Control Port is an open collector / open drain output. While most Parallel Ports have internal pull-up resistors, there is a few which don't. Therefore by incorporating the two 10K external pull up resistors, the circuit is more portable for a wider range of computers, some of which may have no internal pull up resistors.

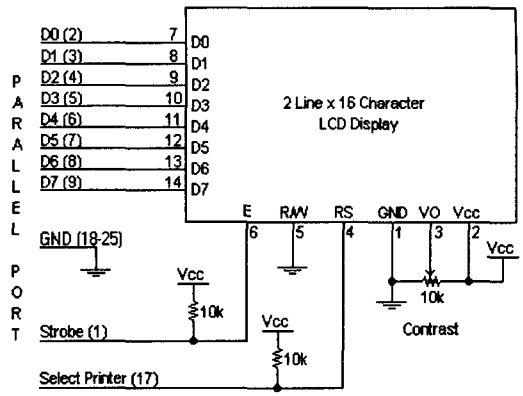

Figure 5. LCD to Parallel Port Interface

We make no effort to place the Data bus into reverse direction. Therefore we hard wire the *R/W* line of the LCD panel, into write mode. This will cause no bus conflicts on the data lines. As a result we cannot read back the LCD's internal Busy Flag which tells us if the LCD has accepted and finished processing the last instruction. This problem is overcome by inserting known delays into our program.

The 10k Potentiometer controls the contrast of the LCD panel. Nothing fancy here. As with all the examples, I've left the power supply out. You can use a bench power supply set to 5v or use a onboard +5 regulator. Remember a few de-coupling capacitors, especially if you have trouble with the circuit working properly. The 2 line x 20 character LCD modules are available from a wide range of manufacturers and should all be compatible with the HD44780.

IX. Future Work

The test results will be published in future after the system has undergone self-learning and improvement of individual voice model. In the future, it is the intent, to beta-test this system over the internet for providing the text of a speaker's speech in almost real time. This application will be useful for academic centers and well as serving the individuals with hearing impairment.