Chapter 10
NORTHERN ILLINOIS UNIVERSITY

Department of Electrical Engineering
DeKalb, IL 60115

Principal Investigators:

Mansour Tahernezhadi (815)-753-8568
Xuan Kong (815)-753-9942

A PORTABLE DEVICE FOR MONITORING VOCAL INTENSITY

Designers: Von Monmany Richard Kwarciany, Jeff Schonhoff
Client Coordinator: Kelly Hall, Department of Communicative Disorder
Supervising Professors: Drs. M. Tahernezhadi, X. Kong
Department of Electrical Engineering
Northern Illinois University (NIU)
DeKalb, IL 60115

INTRODUCTION

A portable device for monitoring vocal intensity is designed for patients with speech impairment. It is a portable device to assist speech-language pathologists and their patients in acquiring information on episodes of overly high volume in everyday conversations. The device incorporates the TMS320C50 DSK board, a low-cost digital signal processor (DSP) board equipped with 14-bit input/output analog-to-digital (ADC) and digital-to-analog (DAC) converters. It is battery operated and wearable around the waist. Signals from either a throat or lapel microphone are sampled and amplified with an adjustable gain through the onboard A/D converter. Digital signal processing algorithms coded in TMS320C50 assembly language perform voice intensity level calculations on the sampled speech signal. The DSK board will generate a beep sound if the voice intensity exceeds a threshold preset by the clinician. The pattern of the audible sound is programmable based on the preference of the patient.

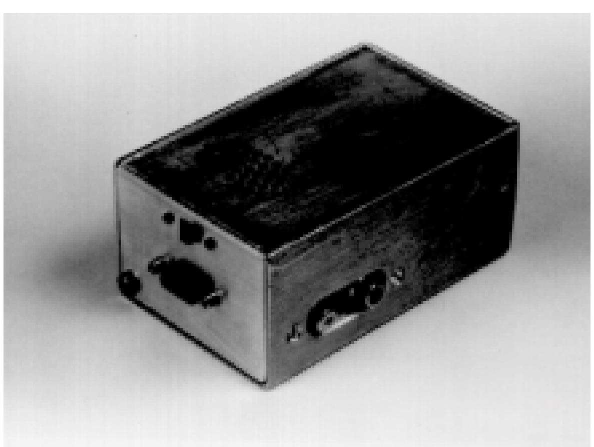

Figure 10.1. Portable Speech Analyzer

SUMMARY OF IMPACT

Speaking loudly may result in vocal cord pathologies such as vocal nodules. Chronic hoarseness, breathiness, or complete loss of voice are the most common symptoms of vocal nodules. Abnormal vocal quality and/or loss of voice not only can be devastating to one's job performance, but can also have profound psychological effects. The primary method of treating patients with voice disorders related to vocal abuse involves monitoring of vocal habits. Typically, patients are asked to self-monitor loudness throughout the day and make adjustments in their phonatory behaviors as indicated. It is not surprising that this treatment strategy is often unsuccessful. A continuous monitoring device would be much more effective in objectively determining an individual's natural voice pattern.

The monitoring device can also serve as a reminder (through beep sounds) when excessively high intensity of voice is detected. Whenever the normal intensity range is exceeded, the portable unit produces an audible tone alerting the patient. In addition, a total daily count is registered to further assist the clinician in monitoring a patient’s progress toward recovery.

The portable speech analyzer meets the needs of many clients. It gives them a sense of control to bring their voice level down when alerted, thus preventing irreversible damage to the vocal cords. The availability of the portable unit also leads to fewer office visits and, in turn, to a reduction in the healthcare cost for the patient.

TECHNICAL DESCRIPTION

The main design requirements for this project were that it be: 1) portable and as small as possible for the patient to wear around the waist without discomfort; and 2) microprocessor-based, allowing easy entering of patient-dependent parameters and maintenance of a log of episodes of interest.

The design incorporated the Texas Instruments TMS320C50 digital signal processor (DSP) DSK board. The fixed point TMS320C50 DSP provides 10 K of on chip RAM with an instruction cycle of 50 nsec. The DSK board comes with the Texas Instruments TLC32040C Analog Interface Circuit (AIC). The AIC is a highly integrated component that combines the function of a 14-bit A/D, a 14-bit D/A, input anti-aliasing filter, output reconstruction filter, and a serial CPU interface. The AIC can be programmed for various sampling rates, anti-aliasing frequencies, and input gains. For the portable speech analyzer, the input speech signal from the microphone pre-amplifier is sampled at 8 KHz by the AIC.

The algorithm for the portable speech analyzer utilizes two buffers each of 240 points (one frame). While one buffer is being filled, the other one would be processed. Four buffer status flags are used to control processing. The RBUFSEL flag is set to control one of the two input buffers are currently being filled. PBUFSEL controls which buffer is being processed. BUF1RDY and BUF2RDY indicate that the corresponding buffer is full and ready to be processed. The main processing loop continuously checks for a full buffer. Upon finding a full buffer, the subroutine for the power calculation is called. During the processing, the other buffer is filled by the interrupt service routine (ISR). In the power subroutine, the power associated with the 240 samples in the buffer is calculated. The calculated power value is then added to a running sum that calculates the averaged power of 10 consecutive frames. The averaged power is then compared with a set threshold. If the averaged power exceeds the threshold, then a warning tone is generated. The output warning could also be realized in the form of a mechanical vibration.

The size of the portable unit is 4 5/16 by 2 11/16 by 2 1/16 inches. The housing contains the DSK board, the microphone preamplifier, the beeper unit, and the batteries. The portable unit consumes three 9-volt alkaline batteries. The DSK board itself runs from two 9-volt batteries (+9 and –9 volts) and the microphone preamplifier also requires a separate 9-volt supply.

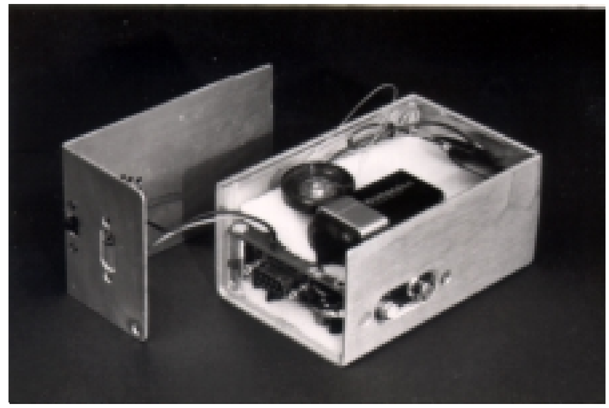

Figure 10.2. Internal View of Portable Speech Analyzer.

The current consumption during the loading of code from PC to the DSK is 240 mA with a minimum required voltage of 6 volts. During processing, the current consumption drops to 100 mA with a minimum required voltage of 4 volts. In order to save battery life, in the absence of any input voice signal, the processor is put into the idle mode (via software), where the current consumption is further dropped to 50 mA. The overall continuous running time for the portable unit is approximately 3 hours.

A three-digit display can also be incorporated to the design to display the number of times excessive volume occurred. A programmer BCD counter with asynchronous RESET (74HC160) is used. The BCD output is directly fed to the BCD-to-Seven Segment latch/decoder/display driver (74HC4511). A common cathode 7 segment LED display is directly connected to 74HC4511. This arrangement is identical for all three digits. However, the “ripple-carry out” output of the counter driving the least significant digit must be given as the clock input to the next counter. The RESET and LOAD pins of the counters are tied to the supply.

The device was tested for various levels of high volume voice signals, and in each case the device was successful.

The final cost of the project is approximately $200.

A DSP-Based Device for Isolated Word Recognition

Designers: Y. K. Chao, J. Skicewicz, M. Schell
Client Coordinator: Kelly Hall, Department of Communicative Disorder
Supervising Professors: Drs. M. Tahernezhadi, X. Kong
Department of Electrical Engineering
Northern Illinois University (NIU)
DeKalb, IL 60115

INTRODUCTION

An isolated word recognition system is developed for training children with phonetic disorders. The device serves as an effective tool in assisting the speech-language pathologist to help children overcome problems with the pronunciation of words. The device performs the task of isolated word recognition using real-time digital signal processing algorithms. The device utilizes the TMS320C31DSK board, a low-cost floating-point digital signal processor (DSP) board equipped with 14-bit input/output analog-to-digital (ADC) and digital-to-analog (DAC) converters. An uttered word received by the microphone is sampled and amplified with an adjustable gain through the onboard A/D converter. Upon acquiring the sampled speech signal, digital signal processing algorithms coded in TMS320C31 assembly language perform isolated word recognition algorithms. The DSK board generates a pre-recorded sound if the word is pronounced correctly.

SUMMARY OF IMPACT

Children with phonetic disorders often substitute one speech sound for another. Based on the substitution pattern, a specialist prescribes certain exercises to correct the problem. Close monitoring of the child's practice is crucial in determining the effectiveness of therapy. Parental involvement is desirable but may not always be possible. A computer-based monitoring system can be very effective in providing feedback to the child as to whether a desired word has been produced correctly or not. With more and more families having personal computers, children with articulation problems will benefit from customized software and a dedicated digital signal processing (DSP) board to correct phonetic disorders.

Typically, the child with an articulation problem will have a pattern of substitution (e.g., “wabbit” for “rabbit”). With this design, the clinician can also monitor the rate of substitution. The DSP-based device gives the child a sense of independence and control.

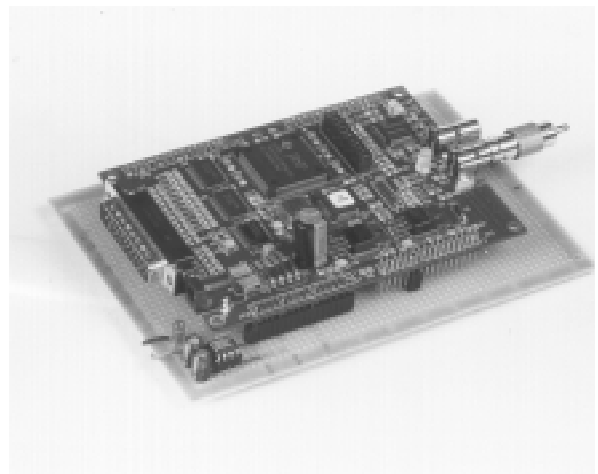

Figure 10.3. Word Recognition Device.

The availability of this system may also lead to fewer office visits, which ideally translates into a reduction of the healthcare cost for the child.

TECHNICAL DESCRIPTION

The word identification/verification system had the following requirements: limited vocabulary, isolated word, and high signal-to-noise ratio (SNR). A PC-based system with a dedicated DSP board is developed to accomplish this task. The design is carried out on the Texas Instruments TMS320C31 digital signal processor (DSP) DSK board. The floating point TMS320C31 DSP provides 2 K of on chip RAM with an instruction cycle of 40 nsec. The DSK board comes with the Texas Instruments TLC32040C Analog Interface Circuit (AIC). The AIC is a highly integrated component that combines the function of a 14-bit A/D, a 14-bit D/A, input anti-aliasing filter, output reconstruction filter, and a serial CPU interface. The AIC can be programmed for various sampling rates, anti-aliasing frequencies, and input gains. The input speech signal from the microphone pre-amplifier is sampled at 8K Hz by the AIC. Since, the on chip

RAM of 2K is not sufficient for the application of this project, an additional 32 K bytes of external SRAM is designed and appended to TMS320C31 DSK board.

For the 32K external SRAM all necessary signals are tapped through dual row headers. The male end is soldered to the DSK board, while the female is soldered to the memory board. The address line A0-A14 are buffered using two 74HCT244 (octal unidirectional buffer). The buffered addressed lines are connected to the SRAM address lines. The data lines are not buffered. The chip select signals for the SRAMs are derived from the DSK board. This is the SRAM/signal decoded by the PAL 22v10 on the DSK. The memory is mapped beginning 80A000h. An onboard regulator is necessary since the DSK board is capable of supplying only 50 mA for expansion purposes.

Software design process was divided into component problems, the solutions for which result in the overall speech recognition system. The first problem is to determine if a word has begun. This is by calculating the signal energy, and how many zero crossings take place within a 40-sample (5 ms) buffer. Once the word has definitely begun, a 160- (20 ms) sample buffer of speech is taken in order to start processing. The 20 ms buffer is compatible with speech applications, because this is a period that speech spectra stay relatively stationary. Next, the input speech samples are windowed by a 30 ms Hamming window in order to obtain a more distinct main lobe as compared to the side lobes in frequency, and to reduce leakage. The 160-point buffers are filled by use of an interrupt service routine (ISR), which is enabled while processing of the previous buffer is initiated.

Processing of the present buffer is done first by calculating the auto-correlation of the 20 ms speech buffer. The buffer is modeled by a 10-coefficient all-pole filter when the system is in the training module. These auto-correlation coefficients are used to obtain the linear predictive coding (LPC) coefficients to describe an all-pole discrete time filter. The LPCs are obtained using Levinson-Durbin algorithm. The number of LPC coefficients determines the accuracy of the filter response as compared to the true signal spectrum. For this design, the number of coefficients is assumed to be 10, sufficient enough to describe the signal spectrum yet small enough for the amount of available memory, and meeting real-time processing requirements. Storage of the word reduces to the number of buffers filled in a given word, multiplied by 10 LPC coefficients. These templates are stored while the processor is still in the training mode.

In the recognition mode, the auto-correlation functions of the spoken word are passed through an inverse all-pole FIR filter described by the LPC coefficients stored in permanent memory. The output of the filter represents the error energy between the present input signal and the response of the template filter. If the word is a match, the output of the filter should yield very low error energy. If the word is not a match, the output of the filter should yield very high error energy. Time warping, to adjust for the speed of pronunciation, is also implemented in the recognition algorithm. This is done by running a segment of autocorrelation through three different templates, each of which describes the word in different consecutive time frames. The template that produces the lowest error energy is considered the correct time frame match between the auto-correlation and the corresponding template.

A second problem is to determine, while in training or in recognition mode, when the word has ended. This is done through the use of high and low thresholds. These thresholds are compared numerous times to the present energy of the signal. When the thresholds have been logically exceeded, a bit in the artificial flags register is set to indicate that the word is over.

Upon the recognition of an uttered word, a prerecorded sound is played to provide feedback about the correct pronunciation to the child. If need be, other types of feedback can also be provided. Although at this point a PC is required to load the program, the device can be made standalone, portable, and less costly by employing flash memory. The design was tested for different words and recognized all the correctly pronounced words without error.

The final cost of the project without a PC is approximately $200.