Voice Recognition Eye Test

Madhavi Priya

[1]Student, Dept. of Electronics & Communication Engineering, Dronacharya group of institution, India

---***---

Abstract - *In this paper, an attempt has been made to review on design of voice recognition eye test that is capable of performing the entire eye test on its own. This project attempted to design and implement a voice recognition system that would identify different users based on previously stored voice samples. Each user inputs audio samples with a keyword of his or her choice. This input was gathered but successful processing to extract meaningful spectral coefficients was not achieved. These coefficients were to be stored in a database for later comparison with future audio inputs. Afterwards, the system had to capture an input from any user and match its spectral coefficients to all previously stored coefficients on the database, in order to identify the unknown speaker.*

***Key Words*: Recognition, Eye, Design, Database, Processing**

1. INTRODUCTION

It starts by displaying large letters and waits on the user to guess which letter has been displayed. The user speaks his/her guess into the microphone. The speech recognition portion uses energy threshold to make sure background noise does not interfere with the user's guess. The system will then determine the next step based on whether the user guessed the displayed letter correctly or not. If the user guesses enough correctly, the text size will continue getting smaller until the user either reaches the minimum text size (corresponding to 20/20) or starts guessing enough incorrectly. If the user guesses too many letters incorrectly, the system will display the result corresponding to the current text size (e.g. 20/30). The thresholds for amount guessed and amount guessed correctly are set in the code. Our eye test uses a dictionary containing the letters "A", "E", "I", "R", and "L" with the same 6 possible text sizes for each of these letters.

1.1 High Level Design

Rationale

The idea is that we want to bring the common Snellen eye test to the household. The test is very simple in that it requires the user to stand a distance away from a chart and read letters to estimate one's visual acuity. So we thought, why not bring the test from the eye doctor's office to the household? In this way, people can take the initiative in taking care of their health in a way that does not require having to drive over and wait until it is their turn to complete the test which, in the end, takes a fraction of the time spent driving and waiting. Also, this can reduce the time it takes to receive a proper diagnosis for one's eyesight when one visits the doctor after having taking the household exam. Also, our eye test could be used in the doctor's office to streamline the process of determining a person's visual acuity since so many of these tests are performed each day

1.2 BPF Design

The range of spoken frequencies is 300 Hz to 3400 Hz. This means we only need to worry about frequencies in that range, so we can filter out the rest. We set up our band-pass filter and determined the resistance and capacitance values according to the cutoff frequency equation below. This equation applies to both the cutoff frequency for the low-pass filter and the high-pass filter.

$$fc = \frac{1}{2\pi RC}\ Hz$$

Equation for Cutoff Frequency Calculation

In the above equation, we chose R1 = 5.1kΩ and C1 = 96.2nF for the low-pass filter, while R2 = 326Ω and C2 = 169.6nF for the high-pass filter. Under this setup, the actual cutoff frequencies are 324Hz and 3058Hz respectively.

We also set up our band-pass filter amplifier to have a proper gain and DC bias. This is because the output of our microphone circuit is inputted into the ADC of the PIC32, which expects an input in the range of 0 to 3.3V so that the ADC can gain more accuracy. We spoke into the microphone and determined its output alone to be in the range of ±13mV. The equation for the gain calculation is simply: Gain = Rg/R2. As a result, we set Rg = 437kΩ to obtain a gain of 1340 so that the circuit's output could be in the proper range for the ADC. Some graphs of our band-pass filters output can be found in the appendix. We have a voltage divider applied to the microphone signal to provide a Vdd/2 DC bias before serving as input to the band-pass filter.

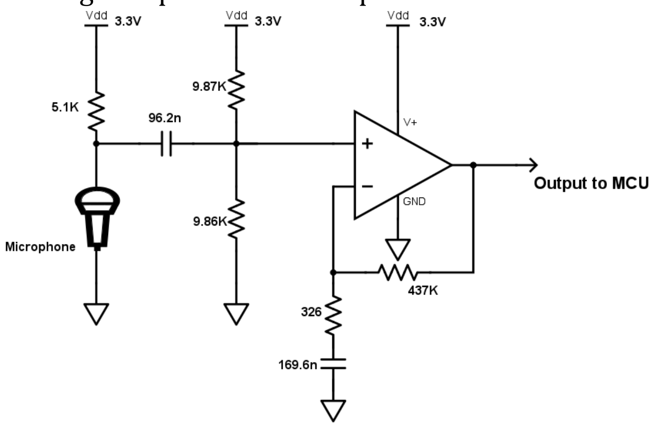

Fig.1. Band-Pass Filter Circuit

2. LOGICAL STRUCTURE

The overall structure of how our speech recognition eye test is implemented is shown in the below figure. Our lab starts by displaying a large letter. Then, the user speaks into the microphone system (microphone, filters and amplifiers). The output of the microphone system is inputted into the PIC32 and sampled at a rate of 8 kHz. The PIC32 stores the first and last 256 samples in two separate arrays. These arrays are then passed into an FFT to switch from the time-domain into the frequency domain. Then, a Mel transform is applied to the FFT results to average the frequencies into 10 bins. These bins are then sorted, along with their corresponding indices. The indices are then compared to the dictionary to see which letter in the dictionary most closely matches to the spoken letter. Once the spoken letter has been determined, it is inputted into the FSM used for our vision test. The FSM will then determine if it is ready to print the results to the TFT for the user to see or if it needs to display a new letter and repeat the process in order to get more information from the user.

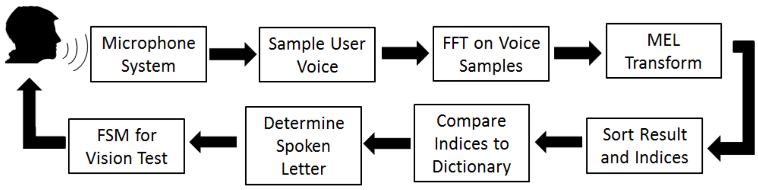

Fig -2: Logical Structure Of Voice Recognition Eye Test

Our eye test works using the standard approach of determining how accurate someone's vision is 20 ft away from the display compared to what the average person can see at 20 ft away. We refer to the Snellen eye chart to help determine the formula for a person's eyesight. We also had to choose text sizes on the PIC32 that closely line up to the proper proportions for the Eye chart.

Using our project, a person could perform the Snellen eye test without the need to leave their house. The results could then be stored and even printed to a computer. If a doctor is familiar with the project, he could then use the data to conduct further eyesight testing instead of spending time needing to measure the person's eyesight. Of course, the Snellen eye test should only serve as a guiding tool for one's eye care and one should always consult with the eye doctor for further concerns.

2.1. Integration Of Hardware And Software

A majority of our code was software based, which meant integrating the hardware and software together did not require much since there was very little hardware. We simply needed to connect the TFT to the PIC32 and the output of the microphone circuit to the ADC of the PIC32.

3. CONCLUSIONS

The results show reasonably good success in recognizing continuous speech from various speakers, for a large vocabulary. The different modules were analyzed in their respective domains and were successfully verified for different speech input files. The obtained results can be improved by fine tuning the system with larger training databases. The next step would be to recognize live speech, which would require more resources including larger speech databases, acoustic models and exhaustive vocabularies to produce good recognition results

REFERENCES

[1] Geoffrey Z, Picheny M (2004) Advances in large vocabulary continuous speech recognition. Adv Comput 60:249–291

[2] Carlson R (2002) Dialogue system. Slide presentation, speech technology, GSLT, Göteborg, 23 Oct 2002. http://www.speech.kth.se/~rolf/gslt/GSLT021023_dialogue.pdf

[3] Rolf C, Granström B (1997) Speech synthesis. In: Hardcastle WJ, Laver J (eds) The handbook of phonetic sciences. Blackwell Publishers Ltd, Oxford.

[4] He X, Deng L, Wu C (2008) Discriminative learning in sequential pattern recognition. IEEE Signal Process Mag 25(5):14–36