580.471 Principle of Design of Biomedical Instrumentation
Challenge Project Report
Report Prepared By: Alejandro Stark & Matthew Trachtenberg

Voice Recognition

Project Objective

The main goal of our project was to create a device that could recognize one's voice as a unique biometric signal and compare it against a database to choose the person's identity or deny an unregistered person while being as standalone as possible.

Introduction

We were able to create a standalone program in Labview (does not require the user to have Labview) that uniquely identifies voices by analyzing the harmonics present in the voice. The user must produce a sound within some defined frequency range until a set number of samples are gathered. Additionally, our hardware consists of a microphone with a bandpass filter for irrelevant noise and an amplifier.

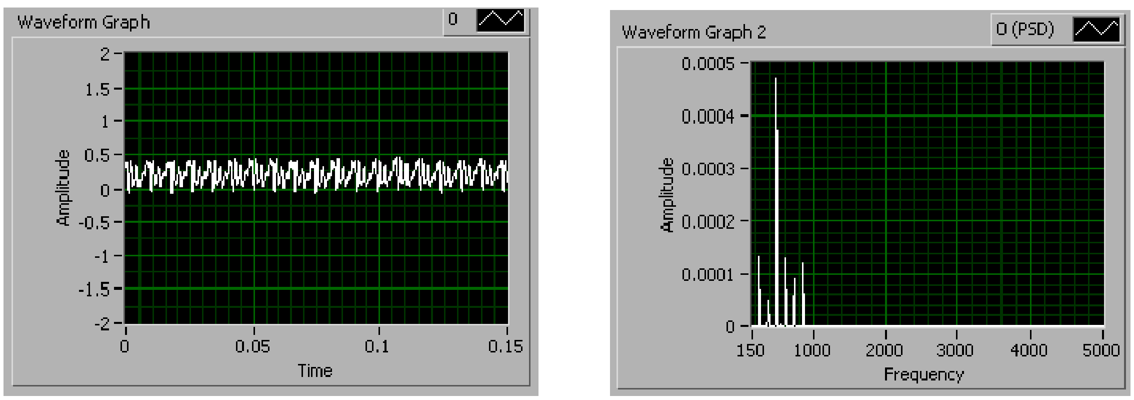

Figure 1: The first Waveform Graph contains the signal from the microphone. Waveform Graph 2 is the spectrum of the signal.

A harmonic of a wave is an integer multiple of the fundamental frequency (pitch). Due to this property, each harmonic is periodic at the fundamental frequency. The human ear cannot differentiate individual harmonics but instead recognizes their combination as the timbre of a tone. The three major auditory cues recognized are the pitch, timbre and loudness. Due to the uniqueness of each person's vocal chords, and thus their voice, individuals will have a specific harmonic pattern (timbre) if enough harmonics are analyzed at the same pitch. The harmonic pattern we are investigating is the presence or absence of certain harmonics. For example two people may each have a first, second, and third harmonic of the same note while only one may have the fourth. Additionally, when the same harmonic is present, the relative amplitude of the signal can vary greatly between two people. Relativity is important here due to the influence of loudness in the signal. The louder the signal, the greater the measured amplitude will be. In order to negate this effect, all signals are normalized relative to the frequency with the greatest amplitude. Figure 1 contains graphs of the signal from the microphone in addition to the signal spectrum. The spectrogram provides amplitude as a function of frequency. This graph makes the relative magnitude of each harmonic very clear. In figure 2 is the waveform of two different users producing the same pitch. It is clear that the wavelength is similar between the two but that the harmonics are considerably different.

Figure 2: Different waveforms for the same note, as produced by different people

Design Process

Part 1: Sensor Selection and Powering

In order to analyze a vocal biometric signal we needed a microphone. The microphones that were available were very affordable and gave a reliable signal. We selected an active microphone, which needs to be connected to a power supply and constantly has current flowing through it; as opposed to a passive microphone, which produces its own voltage from the sound waves. The main advantage of this type of sensor is that the voltage variation is higher, which results in a higher signal to noise ratio. In order to supply the microphone with the appropriate power, we took +9 Volts from a battery and fed it to a 5 Volt regulator. Then we used a voltage divider with two equal resistors (1000 Ω) to get 2.5 Volts. The 5 Volt regulator also provided power for the operational amplifiers mentioned in the following section.

Part 2: Filtering and Amplification of the Signal

The next step in the project is analog signal processing. We aimed to eliminate noise and to amplify the signal so that it would be more clearly defined. We used a low-pass filter to eliminate frequencies above 17000 Hz because these frequencies are above the range that the human voice can produce and would be background noise. Additionally we implemented a high-pass filter to ignore frequencies below 150 Hz because we were getting interference from the ubiquitous 60 Hz signal, and the human voice very rarely produces any sound below this range. Finally, we amplified the filtered signal with a gain of 10 in an attempt to make the signal less susceptible to small amplitude noise between the circuit board and the computer. In order to power the operational amplifiers necessary for the two filters and the amplifier, we needed -5 Volts in addition the +5 Volts. We used two batteries to give us +9 and -9 Volts. Then, a -5 Volt regulator was employed to give the negative voltage necessary. A ten Volt differential was used rather than an 18 Volt differential because it gave satisfactory data while using almost half the power from the batteries.

Step 3: Output to the Computer

After amplifying the signal, we sent it to the analog input on a Labview compatible A/D converter connected to the computer. This only consisted of the signal wire after filtering and amplification, and the ground cable (since our project uses a virtual ground, this cable is very important). The computer receives this information via serial port.

Step 4: Labview Data Storage and Analysis

The central part of our system is a Labview program. This program is attached and consists of three main parts. The first part is the code used to record the user's data. It consists of a command to read the values from the A/D converter, send them into a harmonics processing function and from there, the 11 harmonic values are read and averaged to get a footprint of the user's voice. The frequency has a tolerance of 100 Hz to make it easier for the user to hit the right frequency. There is also a cue switch that will generate a sample of the right frequency through the computer's sound card (and from there to speakers or headphones connected to the computer). This array of 11 harmonic frequencies (footprint) and the name typed in the input area are sent to the next big block, which is the operational program. This program has two cases: working mode and adding users mode (or management mode). The management mode allows the system manager to add new users and delete the entire database (a password is required to verify the manager). Once in this mode, the database file will be read, and every time the save button is pressed, a new user will be entered in the system. In working mode, the program reads the database file, and cycles through all the users after the sample has been recorded, and finds the best match. It does so by comparing the difference between each value in the array (absolute difference) and then averaging these differences to find one value that reflects how close the match was. After finding this, it is compared against a threshold to make sure it is not only the best match, but an actual user. If this threshold is not reached, the system does not allow access and the "Access Denied!" LED lights up. The user name also changes to "not registered". If the user is correctly identified, pressing the enter button will clear the data, since there was no output designed into the program. But if this were an actual product, the enter button would also send the confirmed identity into the security device being controlled.

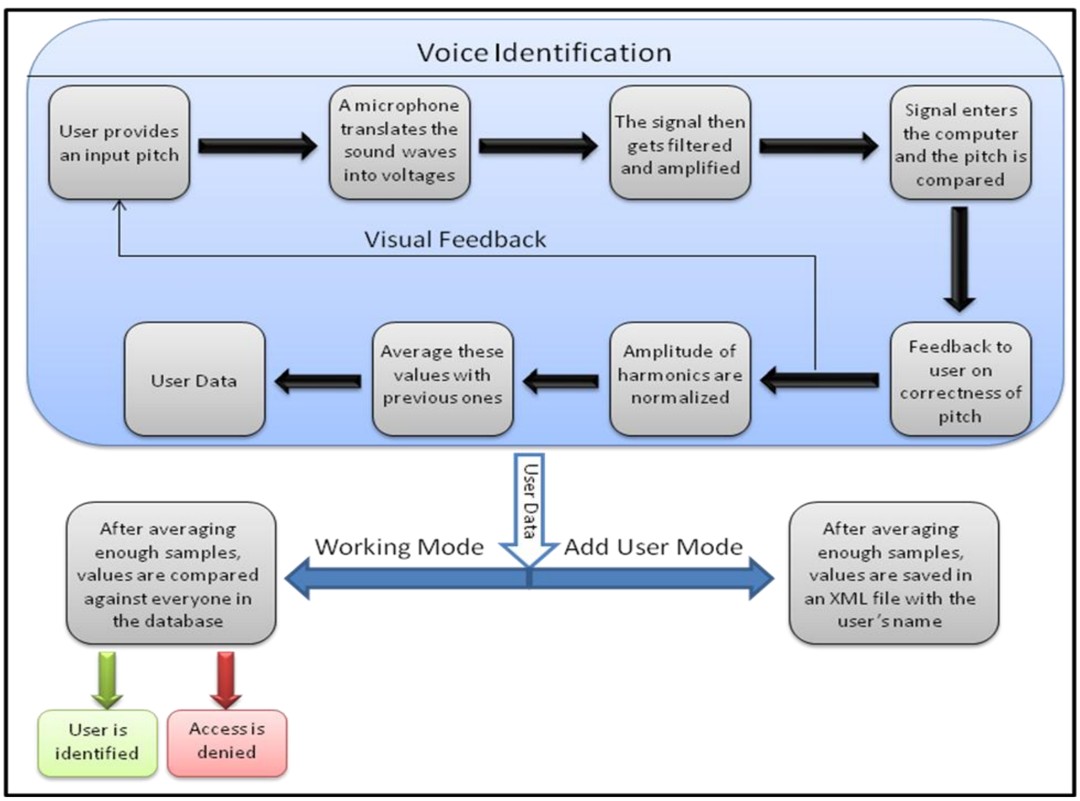

Figure 3: Block Diagram of the Design Process

Problems Encountered

1. Accuracy of identification
2. Finding a user-friendly fundamental frequency and acceptable frequency range
3. Catching errors from trying to load non-existent files that would freeze Labview
4. Choosing an appropriate threshold to indicate an unregistered user
5. Soldering

Possible Design/Device Improvements

1. Using a rail-to-rail op-amp might have given better results, and maybe we could have used a 3.3V regulator to lower current even more, therefore achieving better battery life
2. A new comparison algorithm would help with the accuracy of detecting people. It turned out to be too difficult to compare each and every value for each possible harmonic in the time that we had.
3. Further research into the science behind voice harmonics would help clarify the number of harmonics we have to analyze. Also, it would help to determine how wide a frequency range is allowed to determine a useful sample to analyze.
4. Some people had trouble producing a sound at the desired frequency. We began work on a version that would analyze multiple fundamental frequencies but it was more work than we had time to complete.
5. If the A/D converter were placed inside the box, the signal to be processed could be sent to the computer using multiple digital data transmission standards, serial port, USB, or Bluetooth. This would enable the computer to be further away from the recognition box and the A/D converter box connected to the computer would no longer be needed (lowering the initial setup cost).
6. For the product to be standalone the operation buttons and LED indicators from the GUI should be placed on the box. This is fairly simple as it only requires a few LEDs, an LCD screen, a button, and a digital output that can be connected to the device our voice recognition should activate (door, light, etc.). A PIC would be implemented so these can be controlled.). Additionally, a PIC's processing power should be enough for the whole software to be programmed into it to eliminating the need for Labview altogether
7. A printed circuit board could save some space inside the box.
8. The GUI was merely functional, it can be further improved to make it more aesthetically pleasing and simple to use.
9. The program would be more effective if it prevented multiple entries of the same user.
10. A double microphone array (such as used in noise cancelling microphones) would make the system much more reliable. A second microphone used to record ambient noise and subtract it from the main one could also work.

Potential Biomedical Applications

The potential biomedical applications of such a device could include security protection for a computer, a bank vault, or for a phone system that requires identification before releasing sensitive information. Some conditions that would require voice as a biometric signal over any other technique would be one for amputees and quadriplegics. They would be unable to use fingerprint identification, which is the most common method implemented, or any other handbased tool such as signature analysis or mouse-interface analysis. Another use is for remote identification. Whenever one communicates with a large company over the phone, a series of personal questions are used in an attempt to confirm one's identity. However, your mother's maiden name and where you were born is not always the most confidential information in this age of the internet where every snippet of information is instantly at your fingertips. Voice identification would be a much more secure way to do this and would be easy to complete over the phone.

Attachments:

1. Circuit Diagram for the entire device
2. Labview Screenshots
3. The VI file used