International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395 -0056
Volume: 02 Issue: 03 | June-2015 www.irjet.net p-ISSN: 2395-0072

Application of Speaker Recognition on Biometric

Sumanta Karmakar[1], Soumitra Mukhopadhyay[2]

[1]*Assistant Professor, ECE Department, Asansol Engineering College, West Bengal, India*

[2]*M.Tech Student, ECE Department, Asansol Engineering College, West Bengal, India*

Abstract: ***Speaker recognition is the process of determining which registered speaker provides a given utterance followed by the process of accepting or rejecting the identity claim of a speaker. This paper reports on an experimental study involving signal processing in both time and frequency domain, and to receive a small bit of insight into the principles of speech analysis. This was accomplished by recording four speech segments from each person in our classroom, all of them varying slightly. Comparisons and analysis were then made on each signal, depending upon the instructions given by Dr. Qi.***

Keywords: *degradation, pitch, formant, enhancement*

1. Introduction:

This project entails the design of a speaker recognition code using MATLAB. Signal processing in the time and frequency domain yields a powerful method for analysis. MATLAB's built in functions for frequency domain analysis as well as its straightforward programming interface makes it an ideal tool for speech analysis projects. Speech editing was performed as well as degradation of signals by the application of Gaussian noise. Background noise was successfully removed from a signal by the application of a 3rd order Butterworth filter. A code was then constructed to compare the pitch and formant of a known speech file to 83 unknown speech files and choose the top twelve matches. Development of speaker identification systems began as early as the 1960s with exploration into voiceprint analysis, where characteristics of an individual's voice were thought to be able to characterize the uniqueness of an individual much like a fingerprint. The early systems had many flaws and research ensued to derive a more reliable method of predicting the correlation between two sets of speech utterances. Speaker identification research continues today under the realm of the field of digital signal processing where many advances have taken place in recent years. In the current design project a basic speaker identification algorithm has been written to sort through a list of files and choose the 12 most likely matches based on the average pitch of the speech utterance as well as the location of the formants in the frequency domain representation. In addition, the basic filtering of high frequency noise signals with the use of a Butterworth filter as well as speech editing techniques has been performed.

2. Design Approach:

This multi faceted design project can be categorized into six different sections:

1. speech editing
2. speech degradation
3. speech enhancement
4. pitch analysis
5. formant analysis
6. waveform comparison

Speech analysis was a simple cut-and-paste type procedure. Speech degradation and speech enhancement were related sections, in which a signal was taken, noise was added, and then a lowpass filter was used to help diminish that noise. Pitch analysis was a useful way to roughly tell if a speaker was male or female based on the average pitch derived from the pitch contour. Formant analysis was a slightly more useful approach that could actually be used to help distinguish between members of the same sex. And, finally, waveform comparison made use of both the pitch and formant analyses to find the closest three files to a pre-defined reference file.

3. Experimental Results and Analysis: (using MATLAB)

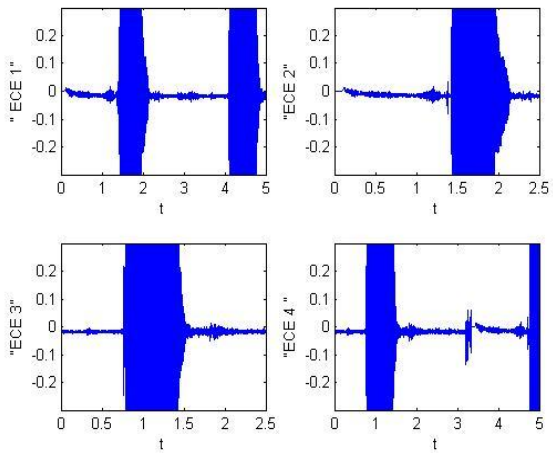

Fig.1 Speech editing

The first direct problem to solve was **speech editing**. The "slow signal" was read into MATLAB using the function wavread(). This function takes a wav file and turns it into an

array of numbers that can be graphed to simulate a spectrograph recording of that same speech file. After that file is graphed, each syllable of "ECE 1" is clearly visible. Since there are five syllables in "ECE 2", the section containing the first five bursts of the plot were read into a temporary variable, and the rest of the original variable was read into a second temporary variable. Then these two temporary arrays were put back into a third new variable in reverse order and the wavwrite() function was used to create a new wav file. This new wav file was then plotted and listened to in order to confirm that it said "ECE 4 Signals and Systems". In the **Fig.1**, Plot 1 is the original signal. Plot 2 is the first half. Plot 3 is the second half. Plot 4 is the recombined signal. It is quite obvious even from these simple plots that the order of the words has been reversed. Also, running the MATLAB code for this section creates a new wav file called a01backwards.wav.

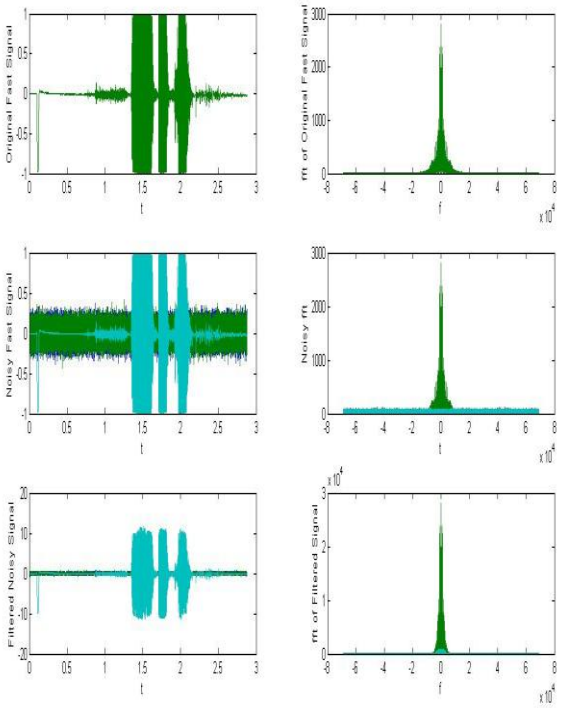

Fig.2 Speech degradation and enhancement

Speech degradation is an application of data compression of digital audio signals containing speech. **Speech enhancement** aims to improve speech quality by using various algorithms. First, the "fast signal" was read into MATLAB in the same way as was done in speech editing. Then a random noisy signal was generated at a variance specified by the user and was added on top of the fast signal. Then a lowpass filter was created at a cutoff frequency specified by the user by using butter() and was applied to the noisy signal by using filter(). All signals and their fft's were plotted in these sections to show that the noise had been added and then diminished. Also, these signals were turned back into wav files for listening purposes using wavwrite(). **Fig.2** shows the plot of the original, noisy, and de-noised files (as well as the fft's of those files) using a variance of 0.08 and a cutoff of 0.04.

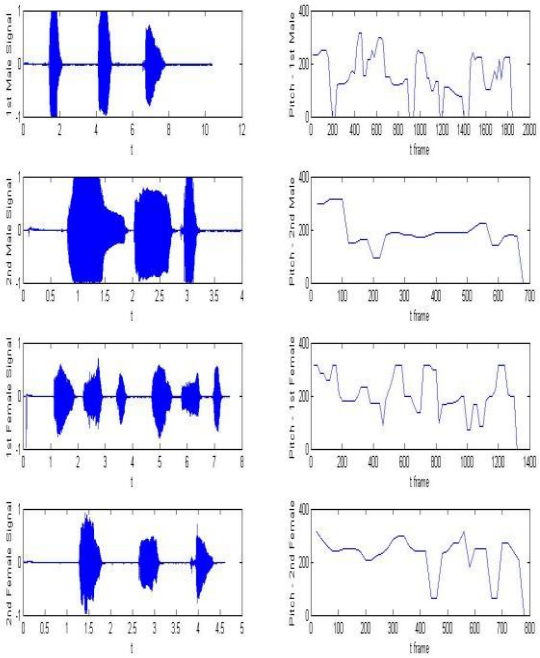

Fig.3 Pitch analysis

The fourth section was **pitch analysis**, and while it was only required to analyze a single signal, several signals were analyzed instead to show how pitch can easily be used to tell male from female when the speakers are all saying roughly the same thing for the same length of time. Signals were read in using wavread(). Dr. Qi's function pitch() was then called, which returns a time frame for the plot, a frequency pitch contour, and an average pitch. All signals were plotted, as were their pitch contours, but it was found that average pitch had the most bearing on male versus female determination. Two males and two females were analyzed in order to show trends between the two groups.

While the pitch contours in **Fig.3** do not do much to convince the user that a particular signal is male or female, the average pitches do. That fprintf() output to the screen was:

The 1st average male pitch frequency is 175.1076 Hz.

The 2nd average male pitch frequency is 193.7725 Hz.

The 1st average female pitch frequency is 213.5543 Hz.

The 2nd average female pitch frequency is 232.1559 Hz.

This clearly shows that the males are well below two hundred, while the females are well above. So, pitch analysis

is a useful tool in speech recognition, as least as far as gender.

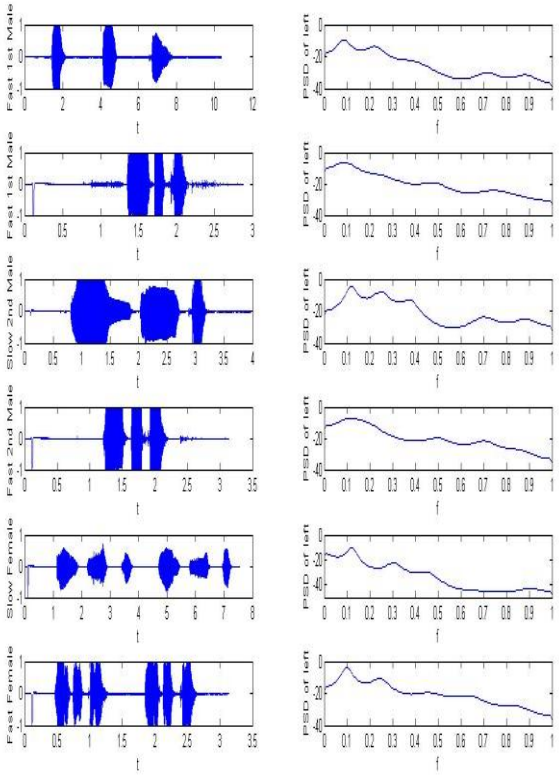

Fig.4 Formant analysis

Formants are defined by Gunnar Fant as "the spectral peaks of the sound spectrum of the voice". In speech science and phonetics, formant is also used to mean an acoustic resonance of the human vocal tract. It is often measured as an amplitude peak in the frequency spectrum of the sound, using a spectrogram or a spectrum analyzer, though in vowels spoken with a high fundamental frequency, as in a female or child voice, the frequency of the resonance may lie between the widely-spread harmonics and hence no peak is visible. Formants are the distinguishing or meaningful frequency components of human speech and of singing.

In acoustics, it refers to a peak in the sound envelope and/or to a resonance in sound sources, notably musical instruments, as well as that of sound chambers. Any room can be said to have a formant unique to that particular room, due to the way sound may bounce differently across its walls and objects. Room formants of this nature reinforce themselves by emphasizing specific frequencies and absorbing others, as exploited, for example, by Alvin Lucier in his piece *I Am Sitting in a Room*.

The fifth section of the project was **formant analysis.** Both slow and fast signals were read in for three different people so that formant peak comparisons could be done both between the slow and fast signals of each individual person, but also to show differences between two different people. Wavread() was used for input, and Dr. Qi's formant() was used to analyze them. As a rather important side note, the formant() function provided on the web was slightly modified so that it would also return the indices of each of the formant peaks, which it did not originally do even though they were calculated inside the function.

Psd contour plots from the formant analysis section are more helpful than pitch when determining a particular speaker, however, as can be seen from the plots in **Fig.4.**

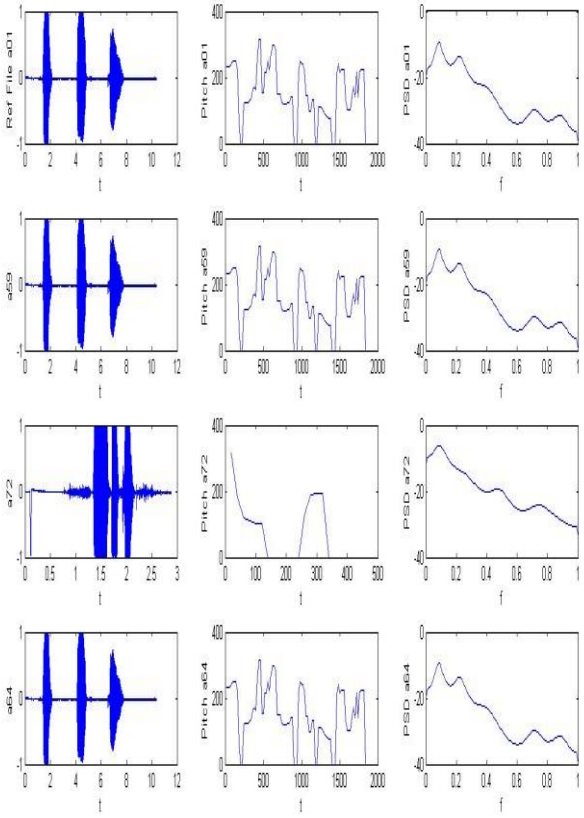

Fig.5 Waveform comparison

Finally, **waveform comparison** was done between the slow signal and all eighty-two other signals. There were several choices as to the best way to compare these signals, including direct psd (formant) comparison, comparison of the psd peaks, comparison of the first few samples in the pitch contour, comparison of a histogram of the pitch contour, and comparison of the average pitches. In the end, it was decided that using both the indices of the psd peaks and the average pitches would be a good comparison. Files were read in and compared in a "for" loop. The differences in the average pitches and psd peaks were then put into an array

and sorted. Ignoring the first element of each sorted array the three closest files were found, and their pitch contours and formants were plotted, shown in **Fig.5.** The point of this was to hopefully show that the computer would pick the other three files recorded by the same person as the three that were closest to the reference.

The computer picked a59.wav, a72.wav, and a64.wav as the closest to the reference file of a03.wav. Unfortunately, only a64.wav is also a file created here, with noise in the background. However, this cannot be seen as a total loss, because the computer did at least get one of the matches correct.

4. Conclusion

Speech editing is nothing more than moving about some arrays of numbers. Enhancement filters can be used to remove both natural and intentional noise, to a reasonable extent. And pitch and formant analysis can be used to give a general idea of whether two speakers are the same person or not. The defect, however, is obvious in the waveform comparison. While these approaches can be used to give a rough estimate or to aid in human decisions about whether two voices are the same, computer programs like these are simply not advanced enough to be completely automated and foolproof. In other words, this is not a "black box" where you do not have to know anything about how the program works and just expect an accurate answer based on a certain set of inputs. Other things that we would like to explore in the subject include Delta-Cepstrum coefficients and perceptual linear predictive coefficients in order to see how much they could help with or replace pitch and formant analysis. Maybe a combination of all four would give a much higher confirmation percentage.

5. References

[1] http://cslu.cse.ogi.edu/HLTsurvey/ch1node9.html
[2] Signal Processing of Speech – *Frank J. Owens 1993*
[3]*Mathworks*® matlab – http://www.mathworks.com/products/matlab/
[4] Dr. Jacek M. Kowalski – Department of Physics, UNT - http://myuntcourses.com/default.aspx
[5] MFCC code - http://www.speech-recognition.de/matlab-examples2.html
[6]Springer Handbook of Speech Processing – *Benesty, Sondhi, Huang – 2008* [7] blueface technology - http://www.blueface.ie
[8]Speaker Recognition Using MFCC and Vector Quantization Model"

[9] Topic on "Extraction of Pitch and Formants and its Analysis to identify 3 different emotional states of a person" ijcsi.org/papers/IJCSI-9-4-1-296-299.pdf

MATLAB" Chesapeake Information Based Aeronautics Consortium August 2005.

[10]"Biometrics Comparison Chart." Court Technology Laboratory. Retrieved 07 Nov. 2003 <http://ct.ncsc.dni.us/biomet%20web/BMCompare>.

[11]Markowitz, Judith A. "Voice Biometrics". Communications of the ACM. Vol. 43, No. 9. September 2000. p66-73

BIOGRAPHIES :

Sumanta karmakar :

Currently working as Assistant Professor, ECE Department, Asansol Engineering College. B.E. from B.U., M.Tech from NITTTR, Kolkata and pursuing PhD from ISM, Dhanbad. A Silent worker all throughout; never demands anything in return.

Soumitra Mukhopadhyay :

Currently studying M.Tech from Asansol Engineering College, ECE Dept. (specialization: Communication).
A good student throughout. Completed B.Tech from Dr. B.C.Roy Engineering College, Durgapur,West Bengal. Residing at Burdwan, West Bengal.

