

Gender Recognition in Labview using a Pitch Analysis VI

Aronee Dasgupta¹, Astha Nachrani², Sahil Chakraborty³, Yogitha S R⁴

¹Department of Electronics and Telecommunication Engineering

^{3,4}Department of Electronics and Communication Engineering

²Department of Computer Science and Engineering

Abstract: The veracity and acceptance of speech processing has come a long way in the last few years. Gender recognition which is one of the important applications of speech processing is found useful in biometric security and automated telephonic communications. In this paper we pre-process the input signal, segment it into frames, segregate into components that allow pitch and format detection, and classify the gender based on average pitch and format value of the system.

Keywords: Gender Recognition, Feature Extraction, Linear Predictive Analysis, Mel Frequency Cepstral Coefficients, Vocal Tract Filter, LabVIEW

1. Introduction

Speech signals originally being analogue in nature are transformed into digital signals for being processed by the computers. Representation of speech signals occurs in two ways: time domain and frequency domain, the latter being more superior. In frequency domain, x-axis defines frequency and y-axis defines magnitude in dB. In time domain, speech signals are assorted into three families namely voiced, unvoiced and silent speech. Development of a gender recognition system depends on the formation and structure of speech [1].

The space of speech generated by humans is enclosed between 50 Hz to 3400 Hz and the frequencies from 20 Hz to 20 KHz of the audible range. Adult males have a fundamental frequency between 85 Hz to 155 Hz while adult females have it between 165 Hz to 255 Hz. The range for Infants being much higher is enclosed between 250 Hz to 650 Hz.

Gender recognition for speech applications has not been researched in the academia widely. This is an open research area. Although, there are some attempts at implementing this system, it fails to perform in noisy environments. Since there is noise associated with almost every process in the real world, it demands for a highly robust recognition system. The motivation of this thesis is to design a robust gender recognition system which works well even in noisy environments[2]. It is also a good idea to combine various biometrics when getting gender information about the people. But due to the system complexity, cost of implementation and scarcity of data, it is more reasonable to use speech signals in order to recognize gender of a speaker since it is one of the easiest form of data that can be accessed.

2. Literature Survey

The aim of speech analysis techniques is to analyze the speech signal and estimate the parameters necessary for

speech processing applications [3]. To derive these parameters, frequency domain representation of speech signals is needed. Various speech processing techniques are then employed. The challenge here is to find the most efficient method of speech processing. There are three most commonly used processing methods namely Fourier analysis, cepstral analysis and linear prediction analysis.

A. Fourier Analysis

Fourier analysis is the traditional method of computing the phase and amplitude spectra of speech signal. It employs fast Fourier transform (FFT) algorithm and window functions to compute short-time Fourier values[5]. We know that the speech signal is comprised of excitation-source and linear signal spectra. Fourier analysis cannot estimate the two spectral components separately. In order to overcome this limitation, speech analysis technique should perform the short time analysis of the speech signal in such a way that the two components of speech are separately available for further processing. Two such techniques are cepstral analysis and linear prediction analysis

B. Cepstral Analysis

Many applications demand the separation of excitation source and signal spectra. The cepstral analysis technique provides an effective way of separating the two components by transforming the product of the two in the frequency domain into their sum and also making use of the fact that the two signals have different spectral characteristics[6]. This technique, thus, provides a procedure for computing spectral envelope and pitch period.

C. Linear Prediction Analysis

Like the cepstral analysis technique, linear prediction analysis technique also provides a method to separate the excitation-source and the signal spectra of the speech model[7]. To achieve this, it assumes an all-pole model for the linear system. Excitation to this model is either a single impulse or a white random noise sequence.

3. Design and Simulation

The required design for the Pitch analysis engine and parameters selected for simulation and their specific values will be elucidated in this section. The parameters which were selected for the pitch analysis are window types, window size, array order and windowing technique.

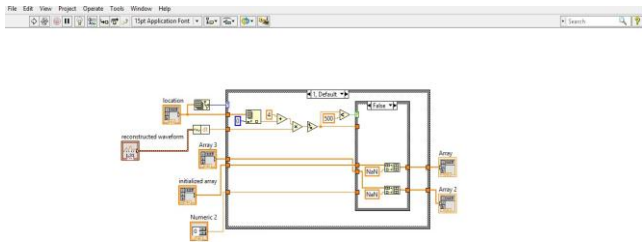


Figure 3.1: Pitch Adjust Scale VI

Fig 3.1 shows the pitch adjust scale VI. The pitch adjust scale VI is used to scale the pitch after the cepstrum coefficients are analyzed and calculated.

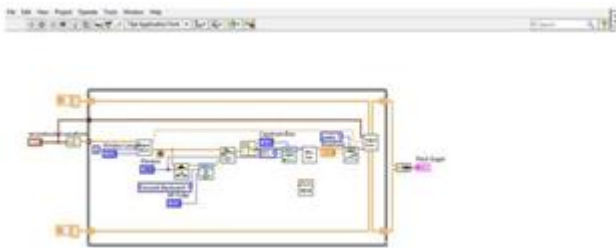


Figure 3.2: Pitch Generate VI

Fig 3.2 shows the generate pitch VI. The window length parameter and the pre-processed signal is fed into the signal block VI. Fast Fourier Transform is performed on output from the signal block to yield the Cepstrum coefficients after passing through a FIR filter.

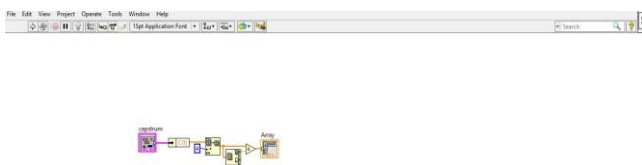


Figure 3.3: Cepstrum Calculation VI

Figure 3.3 shows the Cepstrum calculation VI. The Cepstrum consists of two elements. An element from the excitation sequence (a pulse train for voiced speech) is present in the higher frequencies. The other element originates from the vocal tract impulse response and is present in the lower frequencies.

Table 1: Parameters for Pitch Analysis

Window Length	20 ms to 50 ms
Method	Linear Predictive Coding and MFCC
Window Type	Hamming, Hanning, Flat Top, Blackman
Array Order	20 – 100
No of Samples	1400 – 1600
Sampling Rate	22 kHz
Threshold	260 Hz
Peak Frequency for Men	256 Hz
Least Frequency for Women	272 Hz

4. Observations and Results

After running the speech processing engine, observations have been made from the results obtained from the speech processing unit. Relevant diagrams have been included for a better understanding of the simulation. The window length was chosen as 20ms and the auto correlation order (AR) was also chosen as 20. The number of samples taken were 10, both for male and female voices. The pitch analysis was carried out and from twenty voice samples, the mean pitch of the male voice was always below 142 Hz and the mean pitch of the female voice was 198 Hz.

A decision circuit was used to gauge the gender of the speaker using these thresholds. The mean pitch for each of the voice sample was exported in to a spreadsheet. The pitch analysis was carried out using different window function viz, Hanning, Hamming, Blackman and Flat Top. The cut off frequency for the low pass filter was maintained at 1000 Hz in each of the analysis.

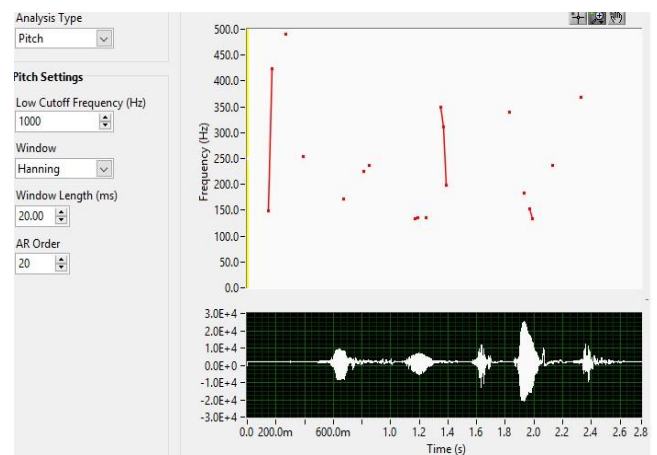


Figure 4.1: Pitch Analysis for Hanning Window

Figure 4.1 shows the pitch analysis with Hanning window. The cut off frequency is chosen as 1000 Hz and the window length as 20ms. The auto correlation order is 20. The pitch contours are plotted along with an intensity graph of the voice signal.

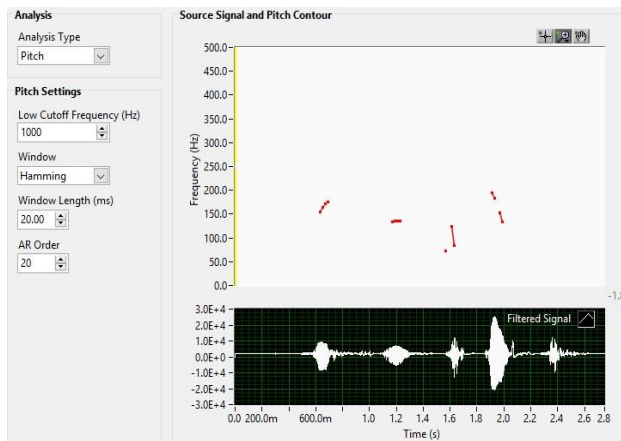


Figure 4.2: Pitch Analysis for Hamming Window

Figure 4.2 shows the pitch analysis with Hamming window. The cut off frequency is chosen as 1000 Hz and the window length as 20ms. The array order AR is taken as 20. The pitch contours are plotted along with an intensity graph of the voice signal.

5. Conclusion

A system is designed using LabVIEW to detect, isolate and analyze the pitch voice samples. The pitch and values of the voice samples are fed into the system. The system was trained by using twenty different voice samples and the mean pitch was established. A new voice sample was fed to this trained system which detected the gender of the input sample voice accurately by comparing the pitch of the given voice sample with the mean value.

6. Future Scope

The pitch based model gave recognition values up to 95% which can be considered as a good value. But in order to increase the performance of the system some additional steps can be implemented. First of all, the training data was small in size. So with the training data size much bigger, the system performance can get better. The system can be trained on faster computers so that the training time can be reduced. Pitch was used as a feature in this technique. Various other speech features can be added into the system which will eventually increase the performance of the system.

References

- [1] H. Erokyar, "Age and Gender Recognition for Speech Applications based on Support Vector Machines," Scholar Commons, Tampa, 2014.
- [2] S. A. Hadei, "A Family of Adaptive Filter Algorithms in Noise Cancellation for Speech Enhancement," International Journal of Computer and Electrical Engineering, vol. 2, no. 2, pp. 67-78, 2010.
- [3] K. S. a. M. G. Kunjithapatham Meena, "Gender Classification in Speech Recognition using Fuzzy Logic and Neural Network," The International Arab Journal of Information Technology, vol. 10, no. 5, pp. 477-485, 2013.

- [4] N. A. Meseguer, "Speech Analysis for Automatic Speech Recognition," Institutt for elektronikk og telekommunikasjon, Norwegian University of Science and Technology, 2009.
- [5] H. A. Murthy, "Algorithms for Processing Fourier Transform Phase of Signals," Speech and Vision Laboratory, Indian Institute of Technology, Madras, 1991.
- [6] M. S. S. A. M. M. a. A. A. A. Eslam Mansour Mohammed, "LPC and MFCC Performance Evaluation with Artificial Neural Network for Spoken Language Identification," International Journal of Signal Processing, Image Processing and Pattern Recognition, vol. 6, no. 3, 2013.
- [7] a. R. P. Bishnu Prasad Das, "Recognition of Isolated Words using Features based on LPC, MFCC, ZCR and STE, with Neural Network Classifiers," International Journal of Modern Engineering Research (IJMER), vol. 2, no. 3, 2012.