GENDER RECOGNITION USING SPEECH PROCESSING TECHNIQUES IN LABVIEW

Kumar Rakesh[1],Subhangi Dutta[2] and Kumara Shama[3]

[1]IMT – Ghaziabad, India
kumarrakesh@ieee.org
[2]Wipro VLSI, Bangalore, India
graymalkin7@gmail.com
[3]HOD, ECE Department, MIT, Manipal, India
shama.kumar@manipal.edu

ABSTRACT

Traditionally the interest in voice-gender conversion was of a more theoretical nature rather than founded in real–life applications. However, with the increase in biometric security applications, mobile and automated telephonic communication and the resulting limitation in transmission bandwidth, practical applications of gender recognition have increased many folds. In this paper, using various speech processing techniques and algorithms, two models were made, one for generating Formant values of the voice sample and the other for generating pitch value of the voice sample. These two models were to be used for extracting gender biased features, i.e. Formant 1 and Pitch Value of a speaker. A preprocessing model was prepared in LabView for filtering out the noise components and also to enhance the high frequency formants in the voice sample. To calculate the mean of formants and pitch of all the samples of a speaker, a model containing loop and counters were implemented which generated a mean of Formant 1 and Pitch value of the speaker. Using nearest neighbor method, calculating Euclidean distance from the Mean value of Males and Females of the generated mean values of Formant 1 and Pitch, the speaker was classified between Male and Female. The algorithm was implemented in real time using NI LabVIEW.

KEYWORDS

Speech analysis, Speech recognition, Speech processing, Gender detection, Detection algorithms

1. INTRODUCTION

1.1. Problem Definition

The aim of this paper is to identify the gender of a speaker based on the voice of the speaker using certain speech processing techniques in real time using LabVIEW. Gender-based differences in human speech are partly due to physiological differences such as vocal fold thickness or vocal tract length and partly due to differences in speaking style. Since these changes are reflected in the speech signal, we hope to exploit these properties to automatically classify a speaker as male or female.

1.2. Proposed Solution

In finding the gender of a speaker we have used acoustic measures from both the voice source and the vocal tract, the fundamental frequency (F_0) or pitch and the first formant frequency (F_1) respectively. It is well-known that F_0 values for male speakers are lower due to longer and thicker vocal folds. F_0 for adult males is typically around 120 Hz, while F_0 for adult females is around 200 Hz. Further adult males exhibit lower formant frequencies than adult females due to vocal tract length differences.

Linear predictive analysis is used to find both the fundamental frequency and the formant frequency of each speech frame. The mean of all the frames is calculated to obtain the values for each speaker. The Euclidean distance of this mean point is found from the preset means of the male class and the female class. The least of the two distances determines whether the speaker is male or female. The preset mean points for each class is found by training the system with 20 male and 20 female speakers.

1.3. Benefits

Automatically detecting the gender of a speaker has several potential applications or benefits:

- Facilitating automatic speaker recognition by cutting the search space in half, therefore reducing computations and enhancing the speed of the system.
- Enhance speaker adaptation as a part of an automatic speech recognition system.
- Sorting telephone calls by gender for gender sensitive surveys.
- Identifying the gender and removing the gender specific components, higher compression rates can be achieved of a speech signal and thus enhancing the information content to be transmitted and also saving the bandwidth.

2. LITERATURE REVIEW

2.1. Speech Processing Overview

Speech processing techniques and various sound extractions has been discussed extensively in theory over a long period of time. We utilized some of the concepts developed over the time to implement the real time gender recognition module in LabVIEW

2.1.1. Framing: Framing is implemented after initial noise elimination of the speech signal. The recorded discrete signal *s(n)* has always a finite length N_{total}, but is usually not processed whole due to its quasi-stationary nature. The signal is framed into pieces of length $N<<N_{total}$ samples. The vocal tract is not able to change its shape faster than fifty times per second, which gives us a period of 20 milliseconds during which the signal can be assumed to be stationary. The length *N* of the frames is based on a compromise between time and frequency resolution. Usually, an overlapping of the individual frames is used so as to increase precision of the recognition process.

2.1.2. Windowing: Before further processing, the individual frames are windowed. Windowed signal is defined as

$$s_w(n) = s(n).w(n)$$

where $s_w(n)$ is the windowed signal, *s(n)* is the original signal *N* samples long, and *w(n)* is the window itself.

2.1.3. Pre-Emphasis: Pre-emphasis is processing of the input signal by a low order digital FIR filter so as to flatten spectrally the input signal in favor of vocal tract parameters. It makes the signal less susceptible to later finite precision effects. This filter is usually the first order FIR filter defined as

$$s_p(n)=s(n)-a.s(n-1)$$

Where a is a pre-emphasis coefficient lying usually in an interval of (0.9 to1), s(n) is the original signal, and $s_p(n)$ is a pre-emphasized signal.

2.1.4. Features Extraction Using Linear Predictive Analysis (LPC) Feature extraction is a crucial phase of the speaker verification process. A well-chosen feature set can result in quality recognition just as a wrongly chosen feature set can result in a poor recognition. The basic discrete-time model for speech production consists of a filter that is excited by either a quasi-periodic train of impulses or a random noise source. The parameters of the filter determine the identity (spectral characteristics) of the particular sound for each of the two types of excitation.

The composite spectrum effects of radiation, vocal tract and glottal excitation are represented by a time-varying digital filter whose steady-state system function is of the form

$$H(z) = \frac{G}{1 + \sum_{k=1}^{p} a_k z^{-k}}.$$

where a_k are the filter coefficients and G is the gain factor.

The basic idea behind linear predictive analysis is that a speech sample can be approximated as a linear combination of past speech samples. By minimizing the sum of the squared differences over a

finite interval between the actual speech sample and the linearly predicted ones, a unique set of predictor coefficients can be determined. In the all-pole model, therefore, we assume that the signal s(n), is given as a linear combination of its past values and the excitation input u(n),

$$s_n = -\sum_{k=1}^{P} a_k s_{n-k} + G u_n$$

3. RELATED WORK

3.1 Work in Pitch Detection

There is a substantial amount of work on the frequency of the voice fundamental (F_0) in the speech of speakers who differ in age and sex. The data reported nearly always include an average measure of F_0, usually expressed in Hz. Typical values obtained for F_0 are 120 Hz for men and 200 Hz for women. The mean values change slightly with age. Many methods and algorithms are in use for pitch detection divided into two main camps, time domain analysis and frequency domain analysis.

3.1.1 Time Autocorrelation Function (TA):

A commonly used method to estimate pitch (fundamental frequency) is based on detecting the highest value of the Autocorrelation function (ACF) in the region of interest. For a given discrete time signal *x(n),* defined for all *n*, the autocorrelation function is generally defined as

$$R_x(\tau) = \lim_{N\to\infty} \frac{1}{2N+1} \sum_{n=-N}^{N} x(n) \cdot x(n+\tau),$$

- (1)

If *x(n)* is assumed to be exactly periodic with period *P*, i.e., *x(n)=x(n+P)* for all *n,* then it is easy to show that the autocorrelation *Rx (τ)= Rx(τ+P)* is also periodic with the same period. Conversely, periodicity in the autocorrelation function indicates periodicity in the signal. For non-stationary signals, such as speech, the concept of a long-time autocorrelation measurement as given by (1) is not really suitable. In practice, short speech segments, consisting of only N samples, are operated with. That is why a short-time autocorrelation function, given by equation (2), is used instead.

$$R_x(\tau) = \frac{1}{N} \sum_{n=0}^{N-1-m} x(n) \cdot x(n+\tau), \quad 0 \le \tau \le T$$

- (2)

where *N* is the length of analyzed frame, *T* is the number of autocorrelation points to be computed. The variable *τ* is called lag, or delay, and the pitch is equal to the value of lag *τ*, which results in the maximum *R(τ)*.

3.1.2 Average Magnitude Difference Function (AMDF)

The AMDF is a variation of ACF analysis where, instead of correlating the input speech at various delays (where multiplications and summations are formed at each value), a difference signal is formed between the delayed speech and original, and at each delay value the absolute magnitude is taken. For a frame of *N* samples, the short-time difference function AMDF is defined by the relation (3):

$$D_x(\tau) = \frac{1}{N} \sum_{n=0}^{N-1-m} \left| x(n) - x(n+\tau) \right|, \quad 0 \le \tau \le T$$

- (3)

where *x(n)* are the samples of input speech and x(n+τ) are the samples time shifted on τ samples. The difference function is expected to have a strong local minimum if the lag τ is equal to or very close to the fundamental period.

Unlike the autocorrelation function, the AMDF calculations require no multiplications, which is a desirable property for real-time applications. PDA based on average magnitude difference function has relatively low computational cost and is easy to implement.

3.1.3 Cepstrum Pitch Determination (CPD)

Cepstral analysis also provides a way for the fundamental frequency estimation. The cepstrum of voiced speech intervals contains strong peak corresponding to the pitch period. Generally, the cepstrum is defined as an inverse Fourier transformation of the logarithmic spectrum of signal. For pitch determination, the real part of cepstrum is sufficient. The real cepstrum *C(k)* of the discrete signal *s(n)* can be calculated by (4):

$$C(k) = \frac{1}{N}\left\| \sum_{p=0}^{N-1} S(p) \cdot e^{-j2\pi \cdot pk/N} \right\|, \quad (4)$$

where *S(p)* is logarithmic magnitude spectrum of *s(n)*

$$S(p) = \log\left\| \sum_{n=0}^{N-1} s(n) \cdot e^{-j2\pi \cdot np/N} \right\|. \quad (5)$$

The cepstrum is so-called because it turns the spectrum inside out. The x-axis of the cepstrum has units of quefrency (1/frequency). The cepstrum consists of peak occurring at a high quefrency equal to the pitch period in seconds and low quefrency information corresponding to the formant structure in the log spectrum. The cepstral peak corresponding to pitch period of voiced segments is clearly resolved and quite sharp. Hence, to obtain an estimate of the fundamental frequency from the cepstrum a peak is searched for in the quefrency region 0.002-0.02s corresponding to typical speech fundamental frequencies (50-500Hz).

3.1.4 Using LPC Parameters

This algorithm proposed by Markel is called the Simple Inverse Filtering Tracking method. The input signal after low-pass filtering and decimation is inverse filtered to give a signal with an approximately flat spectrum, which corresponds to the error signal. The digital inverse filter is given by

$$A(z) = 1 + \sum_{i=1}^{M} a_i z^{-i},$$

where M is specified. It is required to find the coefficients a_i, i = 1, 2 ,.... M such that the energy measured at the filter output { y_n} is minimized. The purpose of the linear predictive analysis is to spectrally flatten the input signal. If it were essentially flat except for random perturbations about a constant value (the case for unvoiced sounds) the transformed results would have a major peak at the time origin with low-amplitude values for all other terms. If the spectrum were essentially flat except for a definite periodic component whose peaks are separated by *Fo* (corresponding to a voiced sound) the transformed sequence would have a main peak at the origin with a secondary peak at *P* = l/Fo. The short-time auto-correlation of the inverse filtered signal is computed and the largest peak in the appropriate range is chosen.

3.2 Work in Formant Tracking

The speech waveform can be modeled as the response of a resonator (the vocal tract) to a series of pulses (quasi-periodic glottal pulses during voiced sounds, or noise generated at a constriction during unvoiced sounds). The resonances of the vocal tract are called formants, and they are manifested in the spectral domain by energy maxima at the resonant frequencies. The frequencies at which the formants occur are primarily dependent upon the shape of the vocal tract, which is determined by the positions of the articulators (tongue, lips, jaw, etc.). In continuous speech, the formant frequencies vary in time as the articulators change position.

The formant frequencies are an important cue in the characterization of speech sounds, and therefore, a reliable algorithm for computing these frequencies would be useful for many aspects of speech research, such as speech synthesis, formant vocoders, and speech recognition.

3.2.1 Linear Prediction Coding Method:

This frequently used technique for formant location involves the determination of resonance peaks from the filter coefficients obtained through LPC analysis of segments of the speech waveform . Once the prediction polynomial *A(z)* has been calculated, the formant parameters are determined either by "peak-picking'' on the filter response curve or by solving for the roots of the equation A(z) = 0. Each pair of complex roots is used to calculate the corresponding formant frequency and bandwidth. The computations involved in "peak-picking" consist of either the use of the fast Fourier transform with a sufficiently large number of points to provide the prescribed accuracy in formant locations or the evaluation of the complex function $A(e^{j\theta})$ at an equivalently large number of points .

3.2.2 Cepstral Analysis Method:

An improvement on the LPC analysis algorithm adopted the cepstral spectrum coefficient of LPC to acquire the parameters of formant. The log spectra display the resonant structure of the particular segment; i.e., the peaks in the spectrum correspond to the formant frequencies. The robustness of the improved algorithm was better when acquiring the formant of the fragment of vowel.

3.2.3 Mel Scale LPC Algorithm:

This algorithm combines a linear predictive analysis together with the Me1 psycho-acoustical perceptual scale for F1 and F2 estimation. In some speech processing applications, it is useful to employ a non linear frequency scale instead of the linear scale in Hz. In the analysis of speech signals for speech recognition, for example, it is common to use psychoacoustic perceptual scales, specially the Me1 scale**.** These scales result from acoustic perception experiments and establish a nonlinear spectral characterization for the speech signal. The relation between **the** linear scale (f in Hz) and the nonlinear Me1 scale (M in Mel) is given by

$$M = 2595 \log_{10}\left[\left(\frac{f}{700}\right) + 1\right]$$

The Discrete Fourier Transform in Mel scale (DFT-Mel) for each speech segment is first computed by sampling the continuous Fourier Transform at frequencies uniformly spaced in the Mel scale. The autocorrelation of the DFT-Mel is next calculated, followed by computation of the LPC filter in the Mel scale by the Levinson-Durbin algorithm. The angular position of its poles furnishes the formant frequencies in the Mel scale. The frequencies in Me1 are then converted in Hz by using the inverse of the above equation.

4. IMPLEMENTATION IN LAB VIEW

4.1. Approach To Implementation LabVIEW

In order to determine the gender of a speaker we have used two features, pitch or fundamental frequency and the first formant to implement a nearest neighbor classifier. The flowchart for our system is as shown in Figure 1:

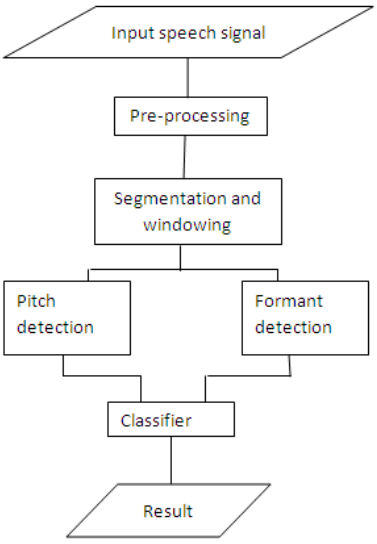

Figure 1. Flowchart of Algorithm used for Implementation in LabView.

4.2. Using LabVIEW to Detect Formants and Pitch

Several methods can be used to detect formant tracks and pitch contour. The most popular method however is the Linear Prediction Coding (LPC) method. This method applies an all-pole model to simulate the vocal tract.

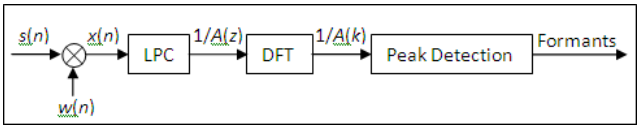

Figure 2. Flow chart of formant detection with the LPC method

Applying the window w(n) breaks the source signal s(n) into signal blocks x(n). Each signal block x(n) estimates the coefficients of an all-pole vocal tract model by using the LPC method. After calculating the discrete Fourier transform (DFT) on the coefficients A(z), the peak detection of 1/A(k) produces the formants.

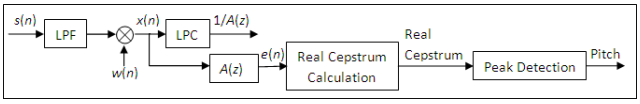

Figure 3.Flow chart of pitch detection with the LPC method

This method uses inverse filtering to separate the excitation signal from the vocal tract and uses the real cepstrum signal to detect the pitch. The source signal s(n) first goes through a low pass filter (LPF), and then breaks into signal blocks x(n) by applying a window w(n). Each signal block x(n) estimates the coefficients of an all-pole vocal tract model by using the LPC method. These coefficients inversely filter x(n). The resulting residual signal e(n) passes through a system which calculates the real cepstrum. Finally, the peaks of the real cepstrum calculate the pitch.

4.3. Implementing in LabVIEW

To input the parameters for pre processing during the generation of formants and pitch, using control palette, various parameters were fed into the system. Figure 4 shows the front panel containing various input control functions

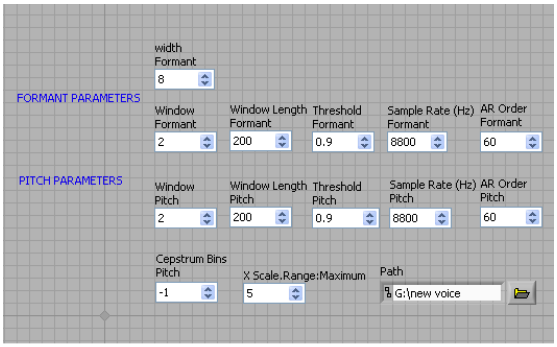

Figure 4. Data input interface in front panel

The values of Formant and Pitch generated after processing of the input speech signal based on input parameters were displayed on the front panel of every sample and were also exported in a Microsoft Excel sheet

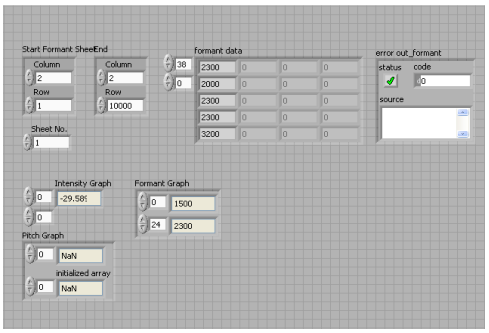

Figure 5. Reading Values of Formant generated by the program

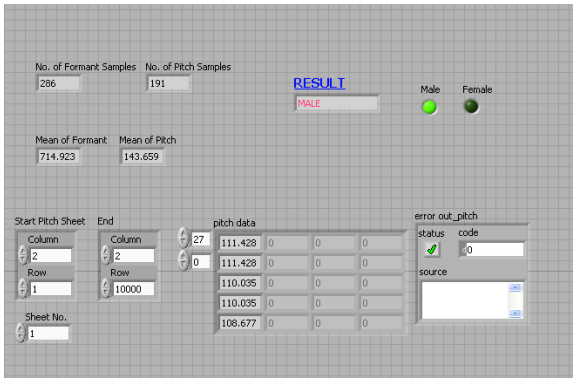

Figure 6. Output of Result and Pitch values of all samples

The program may read from a .wav file as specified by the path into an array of waveform or may also take speech input in real time using Microphone as the requirement may be. For taking real time input, the VI was configured with sampling rate=22050 Hz and maximum length of speech signal=4 s.

The signal is first band limited by a **3.5** kHz bandwidth low-pass filter to eliminate the high frequency noise. After that it is re-sampled with a **0.4** decimation factor to obtain a sampling frequency of 8.8kHz. The digitized speech signal is then put through a first-order FIR filter called a pre-emphasis filter. It results in an amplification of the higher frequency components and also serves to spectrally flatten the signal. The output of the pre-emphasis filter x(n) is related to the input s(n) by the difference equation:

$$s(n) = x(n) - ax(n-1) , \quad 0.9 <= a <= 1$$

We used a = 0.98

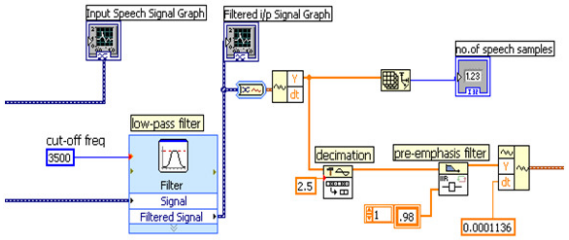

Figure 7. Pre-Processing of input speech signal

The speech signal is then blocked into frames of N samples each and then processed by windowing each individual frame by a Hamming window so as to minimize the signal discontinuities at the beginning and end of each frame. If we define the window as w(n), 0<=n<=N-1 , then the result of windowing the signal is the signal, x1(n)=x(n)w(n). The hamming window has the form:

$$w(n)=0.54-0.46\cos((2*pi*n)/(N-1))$$

The Scaled Time Domain Window VI included in the LabVIEW Signal Processing Toolkit has been used for this purpose.

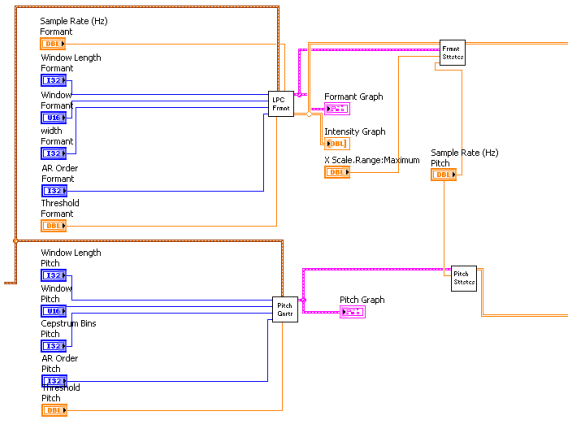

Figure 8. Input of Parameters for Pitch and Formant Generation

The formant location detection technique used by us involves the determination of resonance peaks from the filter coefficients obtained through LPC analysis of segments of the speech waveform. Once the prediction polynomial *A(z)* has been calculated using FFT method, the formant parameters were determined by "peak-picking'' technique. The Advanced Signal Processing Toolkit includes the Modeling and Prediction Vis that has been used to obtain the LPC coefficients or AR model coefficients.

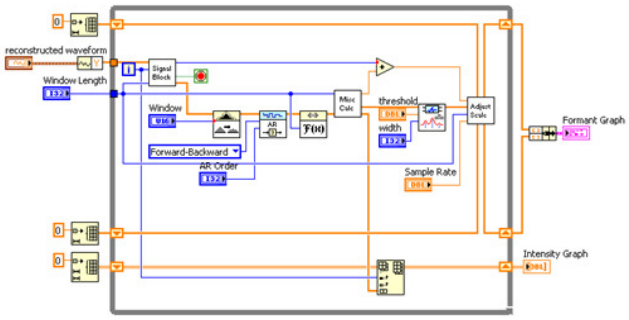

Figure 9. Formant Generation Sub VI

The pitch detection method uses inverse filtering to separate the excitation signal from the vocal tract response by using the linear prediction coefficients in a FIR filter. Cepstral analysis is used to determine the pitch period by calculating the real cepstrum of the filter output which gives a sharp peak in the vicinity of the pitch period. A peak detector is used with a preset threshold to determine the location of the peak. Inherent in the fundamental frequency extraction is the voiced-unvoiced decision. If the maximum value of the frame exceeds threshold, the frame is classified as voiced and the location of peak corresponds to the pitch period, T0. Otherwise, the frame is classified as unvoiced. The pitch period for voiced frames is converted to fundamental frequency by F0[Hz]=1/T0[s].

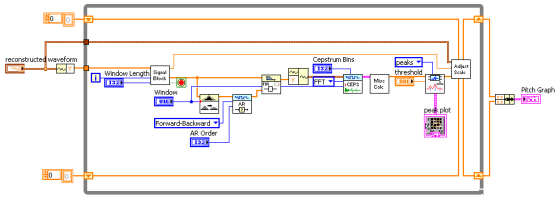

Figure 10. Pitch Generation VI

A For Loop and shift registers were used to extract data elements for mean calculation.

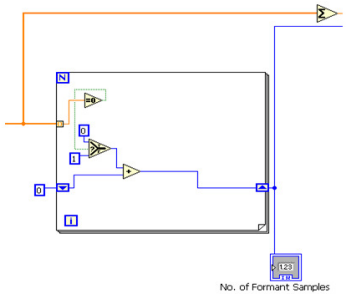

Figure 11. Calculation of number of Formant samples generated sum of all sample values to find Mean value of formant

The distances of the speaker's values from the mean of each class (male and female) were found using the Boolean, Numeric and Comparison palletes. The output is displayed on the front panel both as LED indicators as Well as text.

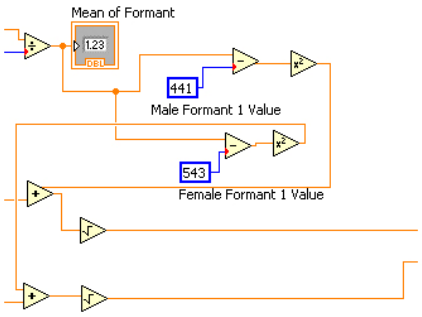

Figure 12. Calculation of Euclidean Distance from Male and Female Mean values for Classification

5. CONCLUSIONS

5.1. Results

As per Hartmut Traunmüller and Anders Eriksson, typical value of F_0 for Male is 120 Hz and 210 Hz for Females. Using the Table 6.1 and Table 6.2, mean value of F_1 of Males and Females for the vowels is calculated. For male F_1 is calculated to be 387 Hz and for Females F_1 mean value 432 Hz.

TABLE 1. VALUE OF FORMANT 1 & 2 OF MALES FOR DIFFERENT VOWELS

MALE Vowels	F1 (Hz.)	Band (Db)	F2 (Hz.)	Band (Db)
A	609	78	1000	88
E	400	64	1700	81
I	238	73	1741	108
O	325	73	700	80
U	360	51	750	61

TABLE 2. VALUE OF FORMANTS 1 & 2 OF FEMALES FOR DIFFERENT VOWELS

MALE Vowels	F_1 (Hz.)	Band (Db)	F_2 (Hz.)	Band (Db)
A	650	69	1100	95
E	500	75	1750	104
I	330	89	2000	114
O	400	86	840	109
U	280	70	650	132

These values of Mean of Pitch and Formant of Males and Females were fed into the system for discrimination between Male or Female speakers by finding the Euclidean distance of a speaker's mean pitch and formant from these two mean values on a 2 dimensional plot. Random samples of 20 males and 20 females were fed into the system to check the efficiency and functioning of the same. Table 6.3 lists the data obtained after running the system.

TABLE 3. RESULT OF GENDER DETECTION LABVIEW PROGRAM

Sample No.	Formant 1	Pitch	Detection	Result
1	505.501	160.754	MALE	CORRECT
2	695.574	176.046	FEMALE	CORRECT
3	1460.23	173.303	FEMALE	CORRECT
4	885.016	155.239	MALE	CORRECT
5	761.23	173.209	FEMALE	CORRECT
6	714.923	143.659	MALE	CORRECT
7	522.094	173.754	MALE	CORRECT
8	679.827	147.77	MALE	CORRECT
9	784.476	189.635	FEMALE	CORRECT
10	476.131	143.516	MALE	CORRECT
11	748.839	190.553	FEMALE	CORRECT
12	1160.29	181.695	FEMALE	CORRECT
13	1080.04	147.02	MALE	CORRECT
14	859.214	179.296	FEMALE	CORRECT
15	971.238	186.185	FEMALE	CORRECT
16	1095.92	158.924	MALE	CORRECT
17	574.057	149.024	MALE	CORRECT
18	698.892	171.991	FEMALE	CORRECT
19	709.21	189.829	FEMALE	CORRECT
20	1310.904	177.549	FEMALE	CORRECT
21	469.689	139.753	MALE	CORRECT
22	671.406	179.597	FEMALE	CORRECT
23	939.004	134.418	MALE	CORRECT
24	809.216	151.235	MALE	CORRECT
25	798.291	183.719	FEMALE	CORRECT
26	749.231	157.821	MALE	CORRECT
27	914.619	191.605	FEMALE	CORRECT
28	861.681	178.216	FEMALE	CORRECT
29	711.191	174.997	FEMALE	CORRECT
30	910.353	183.512	FEMALE	CORRECT
31	2150.81	169.182	MALE	INCORRECT
32	627.236	152.43	MALE	CORRECT
33	731.902	190.041	FEMALE	CORRECT
34	829.621	148.051	MALE	CORRECT

5.2. Conclusion

Considering the efficiency of the results obtained, it is concluded that the algorithm implemented in LabView is working successfully. Since the algorithm does not extract the vowels from the speech, the value obtained for Formant 1 were not completely correct as they were obtained by processing all the samples of the speech. It was also observed that by increasing the unvoiced part in the speech, like the sound of 's', the value of pitch increases hampering the gender detection in case of Male samples. Likewise by increasing the voiced, like the sound of 'a', decreases the value of pitch but the system takes care of such dip in value and results were not affected by the same. Different speech by the same speaker spoken in the near to identical conditions generated the same pitch value establishing the system can be used for speaker identification after further work.

5.3. Further Work

By identifying the gender and removing the gender specific components, higher compression rates can be achieved of a speech signal, thus enhancing the information content to be transmitted and also saving the bandwidth. Our work related to gender detection showed that the model can successfully be implemented in Speaker Identification, separating the male and female speaker to reduce the computation involved at later stage. Further work is also needed with regard to formant calculation by extracting the vowels from the speech. While working on formants we concluded that including formant for gender detection would make the system text dependent.

REFERENCES

[1] Eric Keller, "*Fundamentals Of Speech Synthesis And Speech Recognition*".

[2] Lawrence Rabiner, "*Fundamentals of Speech Recognition*".

[3] Milan Sigmund. "*Gender Distinction Using Short Segments Of Speech Signal*".

[4] John Arnold. "*Accent, Gender, And Speaker Recognition With Products Of Experts*".

[5] Florian Metze, Jitendra Ajmera, Roman Englert, Udo Bub; Felix Burkhardt, Joachim Stegmann; Christian M¨Uller; Richard Huber; Bernt Andrassy, Josef G. Bauer, Bernhard Little. "*Comparison of Four Approaches to Age And Gender Recognition For Telephone Applications*".

[6] Hui Lin, Huchuan Lu, Lihe Zhang. "*A New Automatic Recognition System Of Gender, Age And Ethnicity*".

[7] E. Jung, A. Scwarbacher, R Lawlor. "*Implementation of Real Time Pitch Detection For Voice Gender Normalization."*

[8] Fan Yingle Yi Li And Tong Qinye. "*Speaker Gender Identification Based On Combining Linear And Nonlinear Features*".

[9] Eluned S Parris And Michael J Carey. "*Language Independent Gender Identification*".

[10] W. H. Abdulla & N. K. Kasabov. "*Improving Speech Recognition Performance Through Gender Separation*".

[11] Huang Ting Yang Yingchun Wu Zhaohui. "*Combining Mfcc And Pitch To Enhance The Performance Of The Gender Recognition.*"

[12] Tobias Bocklet1, Andreas Maier, Josef G. Bauer, Felix Burkhardt, Elmar N¨Oth. "*Age And Gender Recognition For Telephone Applications Based On Gmm Supervectors And Support Vector Machines*".

[13] D.G.Childers, Ke Wu, K.S.Bae & D.M.Hicks. "*Automatic Recognition Of Gender By Voice*".

[14] Yen-Liang Shue and Markus Iseli. "*The Role Of Voice Source Measures On Automatic Gender Classification*".

[15] Deepawale D.S., Bachu R., Barkana B.D. "*Energy Estimation Between Adjacent Formant Frequencies To Identify Speakers' Gender*".

[16] Yu-Min Zeng, Zhen-Yang Wu, Tiago Falk, Wai-Yip Chan. Robust "*Gmm Based Gender Classification Using Pitch And Rasta-Plp Parameters Of Speech*".

[17] John D. Markel, "*Application of a Digital Inverse Filter for Automatic Formant and F_o Analysis*".

[18] Roy C. Snell and Fausto Milinazzo, "*Formant Location From LPC Analysis Data*".

[19] Stephanie S. Mccandless, "*An Algorithm For Automatic Formant Extraction Using Linear Prediction Spectra*" .

[20] John Makhoul,"*Linear Prediction: A Tutorial Review*".

[21] Antonio Marcos de Lima Araujo Fiibio Violaro, "*Formant Frequency Estimation Using A Mel Scale Lpc Algorithm*".

[22] Ekaterina Verteletskaya, Kirill Sakhnov, Boris Šimák, "*Pitch detection algorithms and voiced/unvoiced classification for noisy speech"*

Authors

Mr. Kumar Rakesh received the B.E. degree in 2010 in Electronics & Communication Engineering from Manipal Institute of Technology. He is presently pursuing MBA in Finance from Institute of Management Technology, Ghaziabad and would graduate in the year 2012. He has participated and won several national level events. His research interests include Signal Processing, Digital Electronics & Business Valuation.

Ms. Subhangi Dutta is working as a Verification Engineer in the Analog Mixed Signal group of Wipro Technologies. She completed her B.E. in Electronics and Communication from Manipal Institute of Technology, Manipal in 2010.

Dr.Kumara Shama was born in 1965 in Mangalore, India. He received the B.E. degree in 1987 in Electronics and Communication Engineering and M.Tech. degree in 1992 in Digital Electronics and Advanced Communication, both from Mangalore University, India. He obtained his Ph. D Degree from Manipal University, Manipal in the year 2007 in the area of Speech Processing. Since 1987 he has been with Manipal Institute of Technology, Manipal University, Manipal, India, where he is currently a Professor and Head in the Department of Electronics and Communication Engineering. His research interests include Speech Processing, Digital Communication and Digital Signal Processing. He has published many Research papers in various journals and conferences.