Scientific Journal of Impact Factor (SJIF): 4.72

e-ISSN (O): 2348-4470
p-ISSN (P): 2348-6406

International Journal of Advance Engineering and Research Development

Emerging Trends and Innovations in Electronics and Communication Engineering - ETIECE-2017

Volume 5, Special Issue 01, Jan.-2018 (UGC Approved)

Speech Based Gender Classification

Mohamad Idris Wani[1], Bareen Farooqi[2], Nida Wani[3], Haider Mehraj[4]

[1,2,3,4] *Department of Electronics and Communication Engineering,*
Baba Ghulam Shah Badshah University, Rajouri-185234, J&K

Abstract--*It has been seen that a wide variety of systems or spaces require efficient gender recognition schemes so as to ensure their usage by a designated gender and hence to safe guard such sensitive systems and spaces from the wiles of an impostor. The aim of this this paper is to introduce such system of gender classification and that the rendered services of these sensitive systems are only used by a legitimate gender. One of the most unique characteristic of a person is his\her speech. Speech based gender classification aims at extracting various features of human speech and dictates, how these features can be used for gender classification. This paper gives an in-depth analysis as to how human speech can be used for gender classification.*

Keywords— *Gender, Energy, Features, Frequency, Power, Recognition, Speech.*

I. INTRODUCTION

Gender of a person is one of the most important factors influencing their social communication and interaction. It contains a wide range of characteristic information regarding the difference between male and female. Successful gender classification is essential and it is critical for many applications such as human-computer interaction and computer-aided physiological analysis. Some have also proposed for automatic gender classification using the features that are derived from human bodies or from their behaviors [1]. Humans are very advanced in identifying the gender of a person, one person can identify another person's gender just by looking at their face or just by listening to their voice. This is possible by virtue of the large distinctiveness present is various human features. Humans mostly use face or voice for gender recognition however, other prominent feature's such as gait can be used for gender classification. Such features like voice, gait and face can be used in computers or biometric systems for classifying human male and female genders.

In the past few decades machine learning has grown exponentially and it is very reliable to use the above-mentioned features for gender classification in computers and biometric systems. A machine can be trained to extract different features of a person and based on these features gender of a person can be recognized.

One of the most important biometric feature of a human is face [2]. For solving the problem of gender classification using facial features many techniques have been developed, and may use support vector machine, geometric features based method, graph matching method and neural network based method [2]. While using face for gender recognition 'Local features' or 'Global features' can be considered. In local features or Geometric based feature extraction certain facial points of the face such as nose and eye are considered [3]. However, the problem with these features is that some of the useful information is lost [4]. In Global features or Appearance-based feature extraction, whole face is considered for feature extraction instead of using certain facial points [5].

In Webster's New Colligate Dictionary gait is defined as "as a manner of walking", however speaking in technical terms gait is much more than a walking pattern. Gait is an idiosyncratic feature of an individual which may be determined from weight of an individual, limb length and posture combined with characteristic motion [6]. When features like walking patterns, jogging, climbing of a person are considered for gender classification it is known as Gait based gender classification. In this approach features such as waist-hip ratio, shoulder-hip ratio is used for gender classification [7]. Although standing silhouettes can be used for gender classification but it is a very challenging task. In the earlier gait based gender classification experiments trackers were joined to the main joints of the body however appearance based gait features can be easily acquired at lower computational cost as compared to the model based gait features [8].

Speech or voice is of the most interesting areas in gender classification. Speech is an important means of communication. Speech is the most significant characteristic of speech. Pitch is most commonly used feature for gender classification as it differs male and female voice [9] Either time domain analysis or frequency domain analysis can be used in speech processing [10]. When measurements are performed directly on the speech signal to extract the information about the speaker it is time

domain analysis while as in frequency domain analysis information about the speaker is computed from the frequency content which is then used to form a spectrum [10].

Speech not only contains the information about the communication but also contains the information about the speaker itself. Features such as Frequency, Power Spectral Density and Energy of a speech signal can be quantified for recognizing the gender of the speaker.

Speech processing is the study of speech signals and the processing methods of these signals. It basically involves a representation in digital domain with the conditions of limiting the bandwidth, sampling at a certain rate and storing each sample with an adequate resolution. These speech signals are used to communicate among people and it not only consists of the information but also carries information regarding the speaker of a particular speech that is to say the gender can be classified. Gender classification is a method of classifying whether the person is male or female.

Aspects of speech processing include acquisition, manipulation, storage, transfer, and output of speech signal. There are many applications of speech processing such as speaker recognition, speaker verification, language translation, and many others. One of the most important applications of speaker verification is gender classification by which it is possible to classify the two genders that is male andfemale using speech signal. Both the genders have different values for parameters such as frequency, power etc. and thereby using these parameters it is possible to differentiate the genders.

In our country India, a huge amount of money is being spent in save guarding our women from harassment in public area. One such example can be seen in subways, trains, local buses wherein a few sections have been dedicated for female passengers only. However, it is always seen that local goons are making use of such services which are meant for the opposite gender. By implementing a gender classification system, it would be possible to restrict the entry of such persons to these areas and to a large extent harassment of woman can be eliminated. But it is not possible to implement such a scheme without the use of proper technology. With a current concern of security worldwide speaker identification has received great deal of attention among speech researchers. By proposing this novel system, we intend to save guard the integrity of different genders.

Based on power, energy and frequency we have implemented a complete gender classification classifier to identify a particular gender in addition to description theoretical and experimental analysis have also been provided along with implementation details.

II. IMPLEMENTATION

The standard database completed by James M. Hillenbrand of University of Western Michigan is used. This database has voice samples of 540 males and 576 females. Out of these 44 samples of both male and female are collected and recorded in .wav format. The sound waves being encoded using PCM and sampled at 16,000 Hz. The implementation is divided into three parts:

1. *Training part*
2. *Threshold (Averaging)*
3. *Testing part*

Features like power spectral density, frequency at maximum power carry speaker information. These features can be tracked well by varying the frequency characteristics of the vocal tract and variation in excitation.

Three parameters have been used for evaluating the results that is during training theses three parameters are calculated and threshold for each is calculated separately and the testing specimen will be declared as male or female depending of the value of flag. These three parameters are:

1) Energy of speech signal.
2) Power of speech signal.
3) Fundamental Frequency.

The energy of a signal can be defined as one which has finite energy and zero average power. Therefore, it can be said [11]:

$$0 < E < \infty \; and \; P = 0$$

Where E is energy and P is the power.

On the other hand, power signal is one which has finite average power and infinite energy given as:

$$0 < P < \infty \; and \; E = \infty$$

The power of an energy signal is zero but the energy of power signal is infinite.

Power and energy content of a signal is often calculated for communication applications in speech processing. For any signal in discrete form, the power and energy can be calculated using equations [12]:

$$E = T \sum_{n-0}^{N-1} x^2[n]$$

$$P = \frac{1}{N} \sum_{n-0}^{N-1} x^2[n]$$

Here, 'T' is duration of signal, x[n] is discrete samples of signals at regular interval.

The power spectral density denoted by PSD is the frequency response of a random periodic signal and tells us where the average power is distributed as a function of frequency [13].

$$P = \lim_{T \to \infty} 1/T \int_{-T/2}^{T/2} x^2(t)\, dt$$

For average long-term speech spectrum, the maximum energy is in the 250 Hz to 500 Hz band. The lower frequency band corresponds to vowel sounds, the higher frequency bands in the 2 KHZ to 4 KHZ correspond to consonant sound.

The fundamental frequency of an adult male ranges from 85 Hz to 180 Hz and that of an adult female from 165 Hz to 225 Hz. Therefore, it can be said that the pitch of man's voice falls under low frequency whereas women's voice is of high pitch type.

1)TRAINING PART

During the training part of the gender classifier, we first train our classifier with some speech signals and calculate the energy, fundamental frequency and power characteristics of both male and female speech signals. Training is done to train the classifier to correctly classify whether a particular speech signal belongs to male or a female. Training is essential for any classifier to achieve desired degree of accuracy.

2)THRESHOLD (Averaging)

Once the training is done mean of the parameter is taken to acquire the thresholds. When the testing is being done these parameters are compared with the threshold and the decision is taken in favor of a particular gender which is favored by at least two parameters.

3) TESTING PART

On completion of the averaging process the testing segment starts. The calculated parameters for the testing speech signal will be compared to their respective thresholds which have been calculated beforehand during the training phase. If two or more than two parameters favor one gender than the final decision will be in that particular genders favor.

Noise cancellation is a method for reducing any unwanted sound by the addition of second noise specifically designed to cancel out the first. Active noise control (ANC) is also sometimes known as noise cancellation or active noise reduction (ANR).

Active noise cancellation is done by Fourier transform. Noise cancelling uses this formula to break the external noise into simpler waves and then emit identical waves that are shifted out of phase so that the peaks of each new one corresponds to the trough of its counterpart and as such unwanted sound is removed.

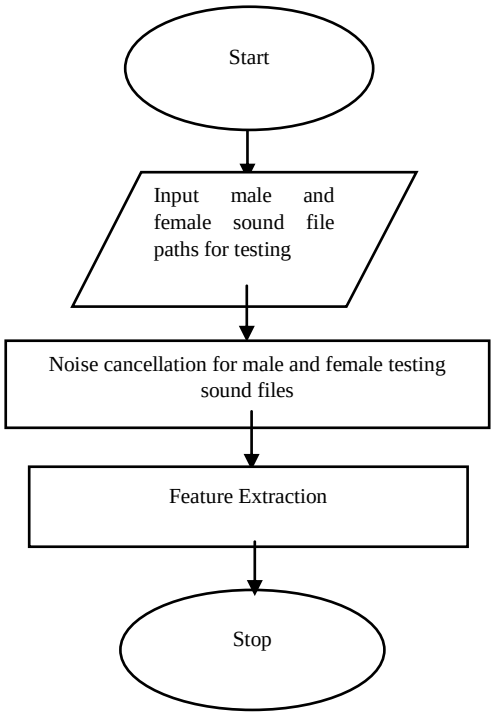


Figure 1. Training Phase

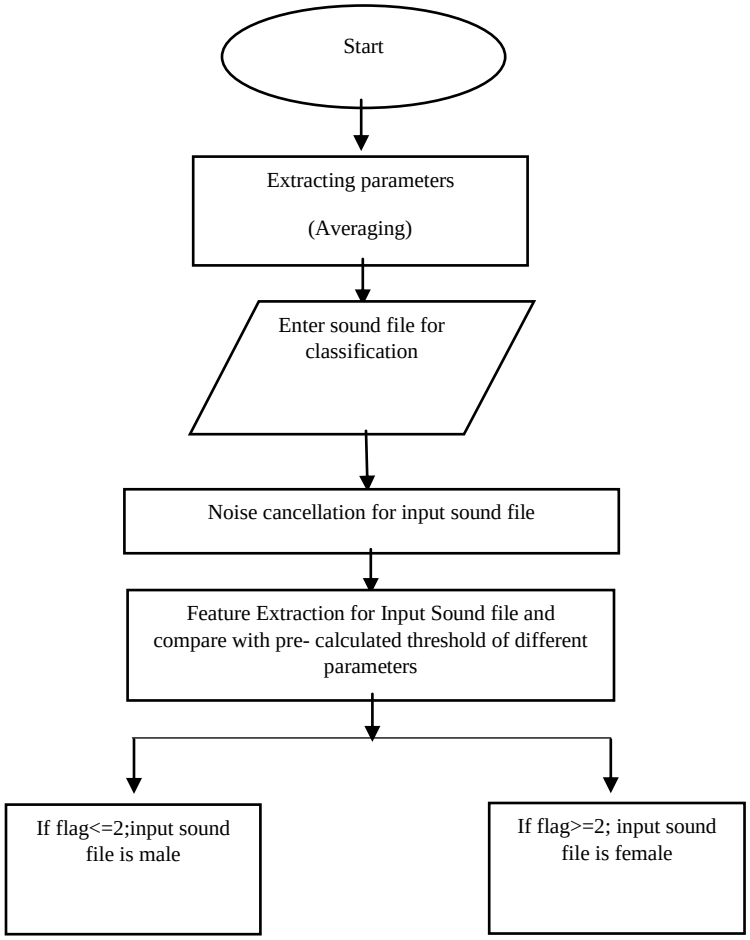


Figure 2 Testing Phase

Figure 1 & 2 is a pictorial representation of the processes taking place in the technique. Firstly, for training the classifier both male and female speech signals are given to the classifier and different features of human speech under consideration are extracted to form a threshold. In the testing phase a random speech signal is taken and its features are extracted. The extracted features of the testing speech signal are compared to the pre-calculated threshold values and the decision is made in favour of a particular gender.

III. RESULTS

The training phase consists of training the classifier about the parameters of energy, average power and frequency. Most of the test samples belonging to the two different genders were correctly classified using the averaging technique.However, there are a few number of miss identified signals in the database that hasbeen used. The miss identified sound waves are those in which the gender recognition is wrongly done that is to say male is classified as female and vice versa. We trained our classifier with the following data presented in the below given table (Table 2):

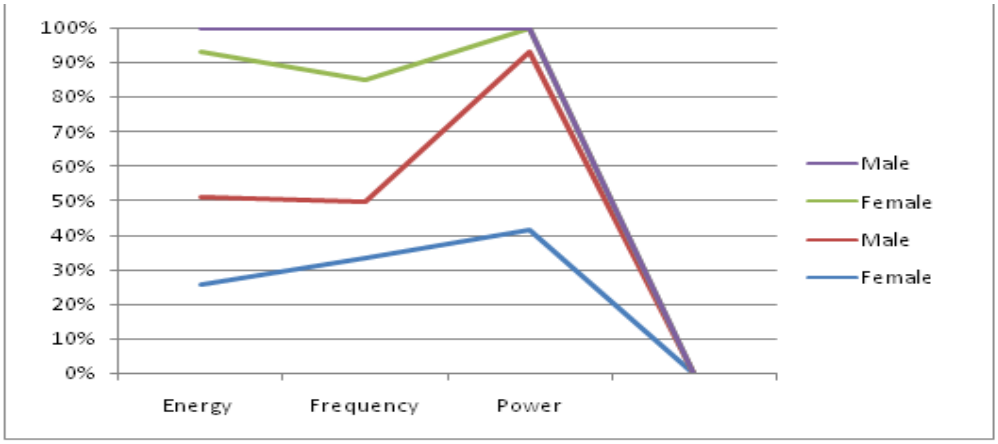
Table 1: *Results for correct gender*

Speaker	Result	Gender
m22.wav	Male	Male
w24.wav	Female	Female

Table 2:*Results for training set*

GENDER/AUDIO	ENERGY	AVERAGE POWER	FREQUENCY
MALE/m01.wav	584.692	6.9853	6194
FEMALE/w01.wav	707.047	1.388e+04	3125

For recognition or more approximately the testing phase, the results for gender classifier yielded the following results for gender classification.



Graph 1 *Showing Energy, Frequency and Power variations of samples taken from training set.*

There may be two types of errors a system can make one mistaking one biometric measurement from two different persons to be from the same person called as false match. Mistaking two biometric measurements from the same person to be from two different persons called as false non- match. These two types of errors are often termed as false accept or false match rate (FMR) and false reject or false non- match rate (FNMR). The equation for calculating the values of FMR and FNMR are as follows:

FMR = P (D1|H0)

FNMR = P (D0|H1)

Where

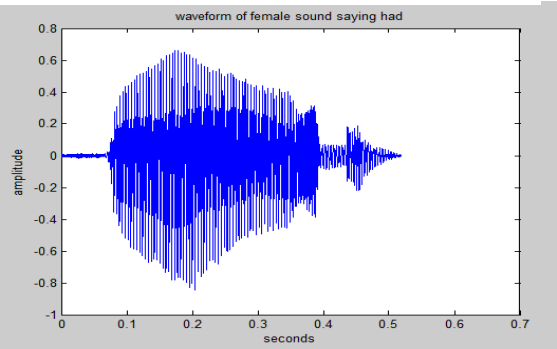
D0 is not the person who she claims to be, D1 is the person who she claims to be and H0 is the input which does not come from the same person, H1 input comes from the same person.

There are many voice signals whose output result is wrongly classified i.e. the miss classified signals.

Table 1*Miss-classified audio signals*

Speaker	Result	Actual Gender
w05.wav	Male	Female
m02.wav	Female	Male

The percentage of accuracy rate of classification performance of the system with different speakers was calculated in this report and the total system accuracy yields up to around 79%.



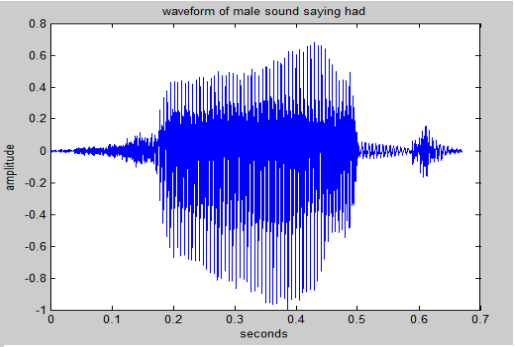
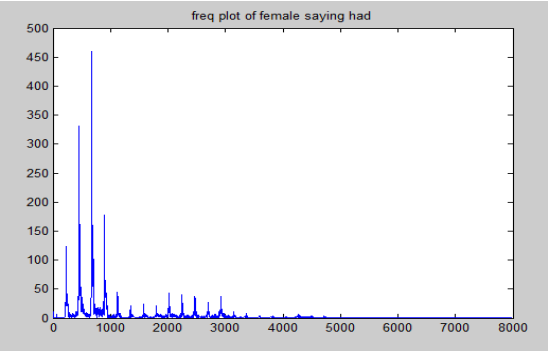


*Figure 3 Waveform of a female saying had.*Figure 4 *Waveform of a male saying had.*



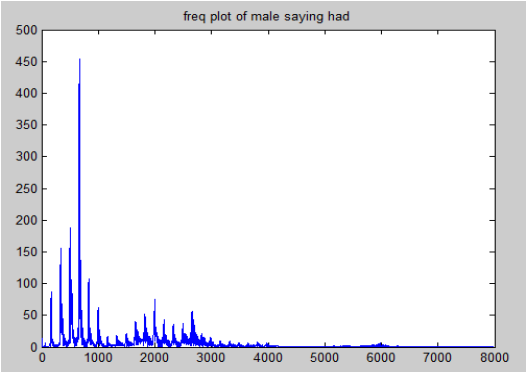


Figure 5 Frequency Plot of a female saying had.

Figure 6. Frequency Plot a male saying had.

IV. CONCLUSION AND FUTURE WORK

During our study it was seen that male and female speech features do not vary appreciably. How so ever, the two genders have unique speech features and these features can be used to recognize a gender as male or female. Also it was seen that words spoken at lower frequencies where the vowels whereas words spoken at higher frequencies where the consonants. Pitch of male voice was quite lower than that of female voice and using this variation a threshold was set. Three features namely Energy, Fundamental Frequency and Power of the human speech were extracted for training the classifier, same features of the test speech were extracted and compared with the threshold and decision was made in favour of a gender on the basis of a counter. If two or more than two features were found belonging to a particular gender, then the voice belonged to that particular gender. It is thus obvious that any system assuring reliable gender classification must necessarily involve a biometric component. Gender classification using speech is an interesting topic. In this correspondence, a gender classification method which combines human knowledge with audio features is proposed. The proposed method gives good results, and correct classification rate is even higher than that achieved by human observers. All the above given results show that despite different tones of speech, speech could be used to identify the gender of a person.

One of the most emerging technologies in machine learning is the imitation of working of a biological neuron using machines, this technology is known as artificial neural network. In future, the genders can be classified genders using face and gait features and combine all the three features hence to create a multimodal system for gender classification.

A personal verification method using both face and speech can be proposed, so that the rate of a biometric verification system can be improved. Such systems are meant to provide higher level of security because more than one biometric feature of the user is requested. This makes it hard for an intruder to fool the system because more than one feature is required at the same time.

REFERENCES

[1] Available online at :https://arxiv.org/abs/1507.05122

[2} AnushriJaswante Dr. Asif Ullah Khan Dr. BhupeshGour "Gender Classification Technique Based on Facial Features using Neural Network" Appeared in International Journal of Computer Science and Information Technologies, Vol. 4 (6), 2013, 839-843

[3] Sajid Ali Khan, Maqsood Ahmad, Muhammad Nazir and Naveed Riaz "A Comparative Analysis of Gender Classification Techniques" Appeared in International Journal of Bio-Science and Bio-Technology Vol. 5, No. 4, August, 2013

[4] A. M. Burton, V. Bruce and Dench, "What are the difference between men and women? Evidence from facial measurements perception", (1993).

[5]B. Golomb, D. Lawrence and T. Sejnowski, "Sexnet: A Neural Network identifies Sex fromhuman faces", Advance in neural information processing systems, (1990), pp. 572-577.

[6] L. Lee, W.E.L. Grimson "Gait Analysis for Recognition and Classification" Appeared in Proceedings of the Fifth International Conference on Automatic Face and Gesture Recognition (FGR02)0-7695-1602-5/02.

[7]T. Anitha, M. Ramya "Gait Analysis for Gender Classification Using CASIA Gait Database" Appeared in International Journal of Innovative Research in Science, Engineering and Technology Vol. 2, Issue 2, February 2013.

[8] Shiqi Yu, Kaiqi Huang "A Study on Gait-Based Gender Classification" Appeared in IEEE TRANSACTIONS ON IMAGE PROCESSING, VOL. 18, NO. 8, AUGUST 2009

[9] Available online at :http://ieeexplore.ieee.org/document/7342709/

[10] BhagyalLaxmi Jena, AbhishekMajhi, Beda Prakash Panigrahi, "Gender Classisification by Speech Analysis" Sillicon Institute of Technology.

[11] Sanjay Sharma "Signals: Energy and Power Signals" in "Signals and Systems (with MATLAB programs)",7th Edition, Published by S.K. Kataria and Sons, India.

[12]Available online at :http://www.gaussianwaves.com/2010/01/calculating-power-and-energy-content-of-a-signal-in-matlab-2/

[13]faculty.etsu.edu/blanton/lab_3_psd.doc&sa=U&ved=0ahUKEwil4PSBoabSAhUS02MKHf5tDCWQFggfMAU&USg=AFQjCNFR-4TVsWq3e01as0LcMXNWF_fsCg

