Matlab based Gender Recognition using Voice Processing

Shaheda Eqbal[1], Ms. Taru Tevatia[2]
[1]M.Tech Student, ECE Department, Al-Falah University, Dhauj, Faridabad, Haryana, India
[2]Assistant Professor, ECE Department, Al-Falah University, Dhauj, Faridabad, Haryana, India

Abstract- The speech signal processing has a wide range of applications in most of the technical fields. In speech processing, gender identification plays an important role. This paper expresses a comparative investigation of speech signals to produce automatic gender classification. Gender classification by speech signal is used to recognize the gender of the speaker by analyzing various features of the voice sample. In this paper, a system, developed for speech encoding, analysis, synthesis and gender identification is presented. A typical gender recognition system can be divided into front-end system and back-end system. The task of the front-end system is to extract the gender related information from a speech signal and represents it by a set of vectors called feature. Features like power spectrum density, frequency at maximum power carry speaker information. The feature is extracted using First Fourier Transform (FFT) algorithm. The task of the back-end system (also called classifier) is to create a gender model to recognize the gender from his/her speech signal in recognition phase. This paper also presents the digital processing of a speech signals (pronounced "A" and "B") which are taken from 10 persons, 5 of them are Male and the rest of them are Female. Power Spectrum Estimation of the signal is examined .The frequency at maximum power of the English Phonemes is extracted from the estimated power spectrum. The system uses threshold technique as identification tool. The recognition accuracy of this system is 80% on average.

Keywords- Gender Recognition, Feature Extraction, First Fourier Transform (FFT), Font-end, Back-end.

I.   INTRODUCTION

Signal is a physical quantity that varies with respect to the independent variable like time, space, etc. Signal values can be represented in zero's and one's. Processing of digital signal by using digital computer is called as Digital Signal Processing. According to Webster's dictionary, speech is the expression or communication throughout in speakers. Speech is the most important thing to express our thoughts. Speech signal is used to communicate among people. It not only consists of the information but also carries the information regarding the particular speaker. From which the speaker is male or female can be recognized [8]. The meaning of Gender Recognition (GR) is recognizing the gender of the person whether the speaker is male or female[1]. The Information about gender, age, ethnicity, and emotional state are the important ingredients that give rich behavioural information. Such information can be obtained from the speech signal [10]. In this paper, an unknown speaker is compared to a database of some known speakers. The best matching system is taken as the recognition decision. From the Recognition decision we conclude whether the given voice sample is generated by a male or female [8].

II.   LITERATURE REVIEW

 Gender Recognition of a speaker has various potential applications. In Sparse Estimation Technique [5], the extracted fundamental frequency is used for finding auto correlation function of speech signal for identifying the gender. The reported gender recognition accuracy is 90 to 95%. But the fundamental frequency (F0) estimation may be difficult in case of noisy environment. In this situation more complex methods can be proposed to improve F0 estimation. In Mel Frequency Cepstral Coefficients (MFCC)[2], features are extracted for voiced vowels where the distinction of male/female is most significant & accuracy is better. This technique does not propose the method for extraction of segments corresponding to selected vowels from the continuous speech audio stream.

In Two Stage Classification Method [10], frequency is estimated for gender recognition. The accuracy is increased. The achieved accuracy is low. If the estimated frequency is far from the zone then male and female frequencies are overlapped. Therefore it is difficult to identification. In Lab VIEW technique [6], identification of the gender and removing gender specific components, higher compression rate can be achieved. Here the information is enhanced to save the Bandwidth. This method does not extract the vowels from speech.

The value is obtained for formant1 were not completely correct as they were obtained by processing all the samples of speech. Hence it is difficult to identify the gender. Linear Predictive Coding (LPC) method [7] used for Feature extraction. A well chosen feature can result in quality recognition. PDA based on average magnitude difference function has relatively low computational cost and it is easy to implement. In which wrongly chosen feature can result in poor recognition. By the use of Machine Language algorithm (MLP) [3], 92.5% accuracy will be achieved. But it is difficult to detection of females as compared to males.

Seven different methods are combined [9] for Automatic gender and speaker recognition. Three baseline subsystems

Gaussian Mixture Model (GMM), Mel Frequency Cepstral Coefficients (MFCC), Support Vector Machine (SVM) and four subsystems SVM based on UBM weight posterior probability super vector, sparse representation based on UBM weight posterior probability super vector, SVM based on GMM Maximum Likely hood Linear Regression(MLLR), SVM based on the polynomial expansion of the coefficients. It is suitable for large scale online adaptive learning due to its property of no new training effort required. But the performance is low. To increase the performance, weighted summation based fusion of these seven subsystems at the score level is demonstrated. The result obtained is 3.1% and 3.8% which is the absolute improvement.

Gaussian Mixture model [1] helps to extract higher level information from the speech signal. This method is a simple and effective. The disadvantage of this technique is the higher levels of information will not provide good performance and may need to be fused with more traditional acoustic-based systems.

III. PROPOSED METHODOLOGY

In this paper, the energy based thresholding technique is used as a classifier. First, the voice samples of both male and female are recorded in a file and it is stored in a recorded file. Then the feature (energy) is extracted from the voice sample and it is referred as a known value. Then the unknown voice sample is taken for analysis and the feature is extracted. The extracted feature is referred as unknown value. The unknown value is compared with the known value. If it is matched, then we conclude that whether the speaker is a male or female. Feature Extraction Energy is used as a feature and it is extracted by estimating the power spectrum. The voices samples of both male and females are recorded in a file. Then voice samples are given as the input to the recognition system. Then FFT is applied to the voice samples. Then the power spectrum is also estimated from the FFT applied signal. Then the energy is extracted from the power spectrum. From that the threshold energy is calculated. For the unknown voice sample, energy is extracted by the same method and it is compared with the estimated threshold energy. The procedure for extracting the energy is shown in figure 1.
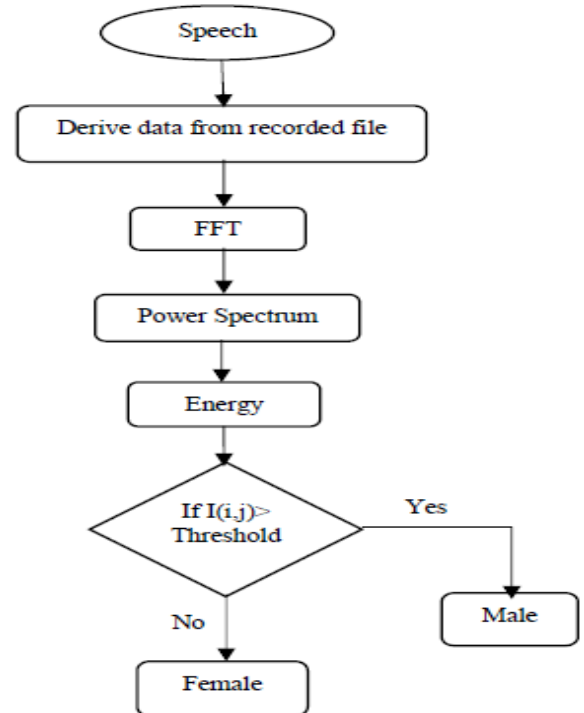

Fig.1: Flow chart of Proposed method

IV. RESULTS AND DISCUSSION

Gender recognition is the process of translating the speech in systems. It is developed for a person to authenticate or verify the identity of a speaker as a part of security purpose. In this paper, the speaker is identified by the estimation of energy as a threshold value. This estimated energy is compared with the threshold energy. If the energy is greater than the threshold then the voice sample is produced by the male. If it is less than the threshold then the voice sample is produced by the female.
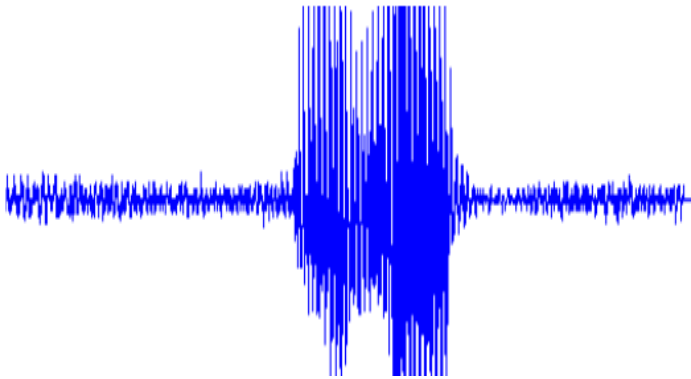

Fig.2: Speech signal of female.

Fig.3: Speech signal of male.

Figure 2 and figure 3 represents the speech signal of female and male respectively.
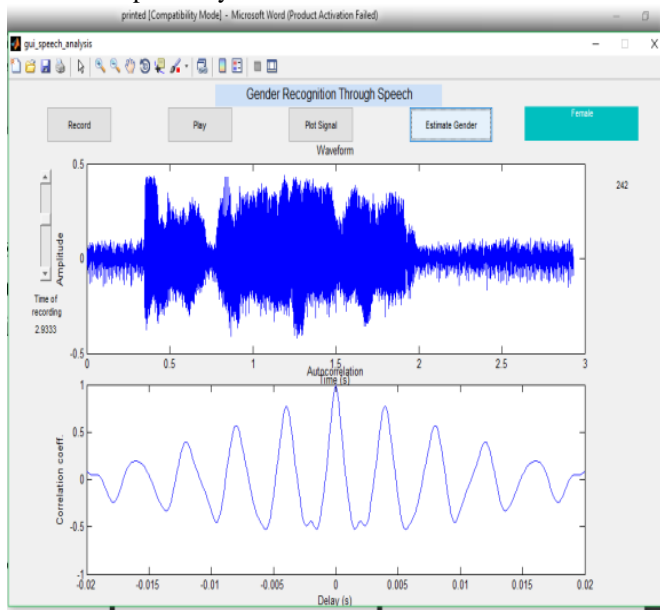

Fig.4: output female

V. CONCLUSION

In this paper the main goal was to develop a gender recognition system using speech signal.[10] The feature selection is one of the most important factors in designing a gender recognition system. From the study of different previous research works it was observed that among the different features the power spectrum results in best classification rate. Thus the power spectrum has been selected as the feature for classification. Among the different technique, the statistical analysis and threshold technique is simple in computation and produces very good results. This is why this method was selected for pattern comparison in recognition process to obtain improved performance. The average recognition accuracy is 80 %. From experimental result, it can be seen that recognition rate decreases as the number of speaker increases. Therefore, the system efficiency is decreases, if the number of reference speakers in the speaker database increases. For recognition constant thresholds have been used. If we could use dynamic threshold for recognition it might produce more accurate and better recognition results.

VI. REFERENCES

[1]. Prabhakar, S., Pankanti, S., and Jain, A. "Biometric recognition: security and privacy concerns" IEEE Security and Privacy Magazine 1(2003), 33-42.
[2]. Huang X., Acero, A., and Hon, H.-W. "Spoken Language Processing: a Guideto Theory, Algorithm and System Development" prentice-Hall, New Jersey, 2001.
[3]. Martin, A., and Przybocki, M. Speaker recognition in a multi-speaker environment.In Proc. 7th European Conference on Speech Communication and Technology (Eurospeech 2001) (Aalborg, Denmark, 2001), pp. 787–90.
[4]. Tomi Kinnunen " Spectral Feature for Automatic Voice-independent Speaker Recognition" Depertment of Computer Science, Joensuu University,Finland. December 21, 2003.
[5]. John R. Deller, John G Proakis and John H. L. Hansen, "Discrete- Time Processing of Speech Signals" Macmillan Publishing company, 866 Third avenue, New York 10022.
[6]. Rabiner Lawrence, Juang Bing-Hwang, "Fundamentals of Speech Recognitions", Prentice Hall New Jersey, 1993, ISBN 0-13-015157-2.
[7]. Md. Saidur Rahman, "Small Vocabulary Speech Recognition in Bangla Language", M.Sc. Thesis, Dept. of Computer Science & Engineering, Islamic University, Kushtia-7003, July-2004.
[8]. Md. Sadek Ali, Md. Shariful Islam and Md. Alamgir Hossain, Gender Rrecognition Using Speech Signal , (IJCSCIT), February 2012.
[9]. Ming li, Kyu J.Han, Shrikanth Narayan, Automatic speaker age and gender Recognition using Acostic and prosodic Level Information Fusion, Computer speech and Language, vol. 27, pp. 151–167
[10]. Yakun Hu, Dappeng wu and Antonio Nucci, pitch Based Gender Identification with Two Stage Classification, International Journal of Security and Communication Networks., vol. 5, Issue 2, pp. 211– 225, 2012.