INTERNATIONAL JOURNAL OF ADVANCE RESEARCH, IDEAS AND INNOVATIONS IN TECHNOLOGY

ISSN: 2454-132X
Impact factor: 6.078
(Volume 6, Issue 4)
Available online at: https://www.ijariit.com

Parameter extraction for audio signal

Neha Sohani Mohanty
nehasohani97@gmail.com
College of Engineering and Technology, Bhubaneswar, Odisha

ABSTRACT

SPEECH is a signal which consisting of various complex information, like the information contain in a signal which to be communicate, language, speaker, region, emotions etc. Speech Processing is one of the important parts of signal processing and finds applications in Audio mining, Human computer interfaces, Assistive technologies, Telecommunication, Security etc. For the speaker identification and emotion analysis Gender Identification is a pre requites. By analyzing the features of different voice signals of male & female it can be determined that which parameters are creating a difference in between the male voice & female voice. For this a comparison between the different parameters present in a voice signal like energy, mean, median and standard deviation is done. It observed that in the graph of mean and standard deviation one could differentiate between a male voice and female as the statistical results are shown. In this paper an input voice signal taken as .wav file and added some noise or unwanted signal; the unwanted noise signals are removed from the original signal by reducing the intensities of high frequencies by averaging and Gaussian window for the smoothing and conv function to removing the pseudo numbers that added as a noise to the original signal.

Keywords: *Energy, Mean, Median, Standard deviation*

1. INTRODUCTION

Communication between human beings is always done by exchanging gestures of emotions and feelings which are recognised by some experiences and knowledge. These expressions are conveyed in speech form. Emotions are part and parcel of human life and these exchanges of emotions are highly influence in decision making [1]. To understand the emotions of a human Gender Identification is primarily required [2]. It this paper various features that affect a speech signal taken into consideration and some signal processing techniques are used. The approach is to calculate which features carry more information and to combine these features to get a better recognition rate [1].

2. ANALYSIS SPEECH SIGNAL

The proposed block diagram of voice recognition is shown in fig. 1. Main stages of automatic sound recognition are feature extraction and classification. For the extraction of features of a voice signal MFCC is widely use as it gives better accuracy in a noisy environment and for classification a Support Vector Machine (SVM) is used by which the different data which are extracted from a voice signal can be classified as in my class male or female [3-5]. Before analysing the feature extraction technique, we have done a signal analysis to know the silent parameters of a voice signal like energy, mean, median and the standard deviation. Before analysing the voice signal there is a pre-analysis is present in which we have taken an input voice signal as .m4a file; any the unwanted noise signals are removed from the original signal by reducing the intensities of high frequencies by averaging and Gaussian window for the smoothing and *conv* function to removing the pseudo numbers that added as a noise to the original signal [6]. In Figure 1 we see the different block that may be used for parameter extraction and classification of voice signals. In this paper we are presenting only the parameter extraction from different audio samples using MATLAB. The different parameters like the energy, mean, median and the standard deviation of the audio signal will be extracted using MATLAB.

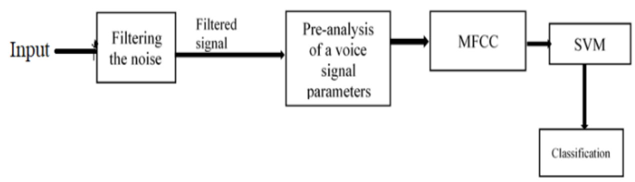

Fig. 1: Proposed block diagram

3. COMPUTATION OF DIFFERENT PARAMETERS

Before analysing the feature extraction technique, signal analysis is done to know the different silent parameters of a voice signal like energy, mean, median and the standard deviation [7-8]. By computing these parameters of any signal, it can be used to recognize the source of that voice signal. Moreover, by analysing these features; the recognition of speaker can be done or the identification of speaker can be done.

3.1 Energy of a signal

It is the area under the squared magnitude of the considered signal. By computing the energy of the male voice and female voice it can be observed the difference in between them.

$$E = \int_{-\infty}^{\infty} |x(t)|^2 \, dt$$

Where E = energy; $x(t)$ = continuous time signal

3.2 Mean of a signal

The mean is the average amplitude of the signal over time. In this method add all the samples together which are present in a signal and divide by N.

$$\mu = \frac{1}{N}\sum_{i=0}^{N-1} x_i$$

Where μ= energy; x_i = continuous time signal; N= no. of samples

3.3 Median of a signal

The median is the value at which half of the observations in the sample have values smaller than the median and half have values larger than the median. The median is often used as the measure of the "Centre" of a signal because it is less sensitive to outliers.

3.4 Standard deviation

The standard deviation is a measure of how far the signal fluctuates from the mean.

$$S = \sqrt{\frac{\sum_{i=1}^{N}(x_i - \mu)^2}{N-1}}$$

Where S =standard deviation; N= no. of observation; x_i= observed value of a sample item

4. RESULTS AND DISCUSSION

The different audio samples from persons of different age groups were taken and their voice was recorded for the word “good morning”. The .m4a file was read into MATLAB and any background noise was removed by filtering. The different parameters were extracted for further analysis.

Sample 1 was of a young girl in her twenties, sample 2 was of a teenage boy, sample 3 was collected from a female of around 40 years and sample 4 was of a mid-forties male. Sample 5 and 6 were from senior citizens female and male respectively. The sample 7 was taken from the same person as sample 6 to analyze the difference in data. The extracted parameters have been tabulated as in Table 1 below.

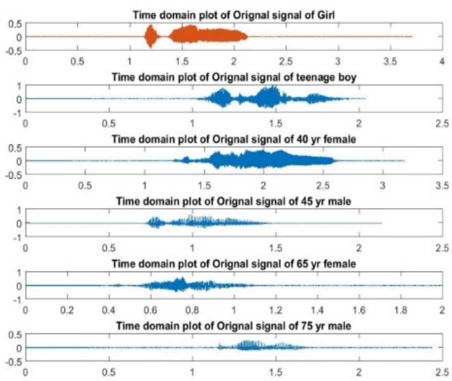

Fig. 2: Time Domain representations of different audio samples

Table 1: Statistical data of different parameters present in voice signal

Parameters	Energy	Mean	Median	Standard deviation
Sample 1	1.1404e+04	3.0745	0.0003	26.2073
Sample 2	1.2081e+08	5.5398	0.0014	36.190
Sample 3	9.1314e+07	3.6055	0.0007	25.257
Sample 4	4.4504e+08	2.9387	0.0003	21.5364
Sample 5	3.5621e+07	3.4683	0.0015	19.0187
Sample 6	6.3527e+08	1.2415	0.0006	7.589
Sample 7	1.6530e+08	6.2179	0.0006	33.2596

From the data found in Table 1 above, comparison between the different parameters like energy, mean, median and standard deviation were observed. On comparison it was found that the energy of the samples collected from the male members were more than the female members. But still there is a drawback like when the male voice and the female voice have equal energy, mean or standard deviation then it can’t be differentiated. To avoid this kind of problem there are some feature extraction techniques like Mel-frequency Cepstral Coefficients (MFCC) can use as this technique is very efficient in noisy environment. After the extraction of these features they must be classification according to the requirements; so, there is a use of support vector machine (SVM) as this is widely used for classification of datasets.

5. CONCLUSION

From the data found in Table 1 above, comparison between the different parameters like energy, mean, median and standard deviation were observed. On comparison it was found that the energy of the samples collected from the male members were more than the female members. But still there is a drawback like when the male voice and the female voice have equal energy, mean or standard deviation then it can’t be differentiated. To avoid this kind of problem there are some feature extraction techniques like Mel-frequency Cepstral Coefficients (MFCC) can use as this technique is very efficient in noisy environment. After the extraction of these features they must be classification according to the requirements; so, there is a use of support vector machine (SVM) as this is widely used for classification of datasets.

6. REFERENCES

[1] Kunxia Wang, Ning An, Bing Nan Li, Yanyong Zhang and Lian Li “Speech Emotion Recognition Using Fourier Parameters” IEEE Transactions on Affective Computing, Vol. 6, No.1, 2015.

[2] K. Sri Rama Murty and B. Yegnanarayana “Combining Evidence From Residual Phase and MFCC Features for Speaker Recognition ” IEEE Signal Processing Letters, Vol. 13, No 1, 2006.

[3] Wei HAN, Cheong-Fat CHAN, Chiu-Sing CHOY and Kong-Pang PUN “An Efficient MFCC Extraction Method in Speech Recognition ” 0-7803-9390-2/06, ©2006 IEEE.

[4] Ahmed Kamil Hasan Al-Ali, David Dean, Bouchra Senadji, Vinod Chandran and Ganesh R. Naik “Enhanced Forensic Speaker Verification Using a Combination of DWT and MFCC Feature Warping in the Presence of Noise and Reverberation Conditions ” 2017, DOI 10.1109/ACCESS.2017.2728801.

[5] Md. Sahidullah, Goutam Saha “Design, analysis and experimental evaluation of block based transformation in MFCC computation for speaker recognition ” 0167-6393/$, Elsevier B.V.

[6] AboElenein, N. M., Amin, K. M., Ibrahim, M., & Hadhoud, M. M. (2016, May). "Improved text-independent speaker identification system for real time applications". In Electronics, Communications and Computers (JEC-ECC), 2016 Fourth International Japan-Egypt Conference on (pp. 58-62). IEEE.
[7] Rathor, S., & Jadon, R. S. (2017, July). "Text independent speaker recognition using wavelet cepstral coefficient and butter worth filter". In 2017 8th International Conference on Computing, Communication and Networking Technologies (ICCCNT) (pp. 1-5). IEEE.
[8] Martinez, J., Perez, H., Escamilla, E., & Suzuki, M. M. (2012,February). "Speaker recognition using Mel frequency Cepstral Coefficients (MFCC) and Vector quantization (VQ) techniques". In Electrical Communications and Computers (CONIELECOMP), 2012 22nd International Conference on (pp. 248-251). IEEE.