International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 02 Issue: 04 | July-2015 www.irjet.net p-ISSN: 2395-0072

EMOTION DETECTION AND GENDER IDENTIFICATION USING AUDIO SIGNALS

Dr. A. Narendra Babu[1], A. Jyothi[2]

[1] *Professor, Department of Electronics & Communication Engg., Lakireddy Balireddy College of Engg., Andhra Pradesh, India*

[2] *P.G. Student, Department of Electronics & Communication Engg., Lakireddy Balireddy College of Engg., Andhra Pradesh, India*

---------- *** ----------

***Abstract*-Emotion is often defined as a complex state of feeling that results in physical and psychological changes that influence thought and behaviour. Emotion modelling and recognition has drawn extensive attention from disciplines such as psychology, cognitive science and engineering. The main goal is to identify the emotional or physical state of a human being from his or her voice. A speaker has dissimilar stages throughout speech that are recognized as emotional aspects of speech and are integrated in the so named paralinguistic aspects. The database considered for emotion recognition is based on audio signals. Fundamental frequency plays a vital role in Gender identification. Based on empirical mode decomposition method detect the dynamically evolving emotion patterns. Classification features are based on the instantaneous frequency and the local oscillation within every mode. The proposed system uses the pre-processing filters to remove the noise and therefore the emotional states were identified efficiently.**

Key words: *Emotion recognition, empirical mode decomposition, fundamental frequency, gender identification.*

I. INTRODUCTION

HUMAN emotions are psycho physiological experiences that affect all aspects of our daily lives. Emotions are complex processes comprised of numerous components, including feelings, bodily changes, cognitive reactions, behavior, and thoughts. Various models have been proposed by considering the ways in which these components interact to give rise to emotions, but at the moment there is no single formulation that is universally acceptable. Modeling emotions is a very challenging problem that has drawn a great deal of interest from the emerging field of human-computer interaction.

The objective is to design systems that can automatically identify emotional states, which would revolutionize applications in medicine, entertainment, education, safety, etc. The main difficulty in formulating these models lies in the fact that we must rely on visible manifestations of emotions to produce and verify them since the latent factors that generate emotions are unobservable.

The first step in modeling any phenomenon is data collection. Every human being has a distinct voice which can be used to identify the person. This identification is similar with the concept of fingerprints. Voice mainly refers to the sound made by human being for talking, screaming, laughing etc. These sounds are produced using the vocal folds (Vocal Cords) which are the primary sound source. Various components at different levels of multitude are mixed together for composition of human voice. Due to this voice of each human being is different and unique. These components are named as pitch, tone, and rate. The rate at which the vocal cords vibrate is referred as pitch. Higher the number of vibration per second, higher will be the pitch which results in high sound of voice band.

The fundamental frequency of voiced speech of adult male lies between 85 to 180 Hz and that of women lies between 165 to 255 Hz. Therefore most of the speeches have their fundamental frequencies below the voice frequency band. The ranges of signals which are

heard by human beings have the frequency range from 20 Hz to 20 kHz.

II. EMOTION DETECTION SYSTEM

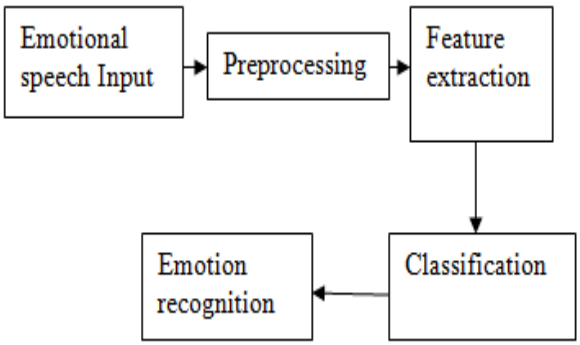


Fig 1: Block Diagram of Emotion Detection System

Combination of various stages forms an emotion recognition system as shown in above fig1. The first basic step of this system contains the task of collecting the speech samples which contain emotions. These samples play a vital role in creating the database of speech .This database is further utilize as an input to the emotion recognition system. After the first step, the second phase is pre-processing of the input signal. The pre-processing of speech signal is done so that the corresponding emotions in it can be identified. This phase mainly focuses on improving the quality of input signal. This signal is then further used to collect the various features. After this system needs to use the classifier so that the speech signals can be classify according to the emotions. These extracted features are further used to provide recognizing the emotions.

III. METHODOLOGY

The proposed methodology for audio signal feature extraction is based o n the Empirical Mode Decomposition. EMD is adaptive and the basis of the decomposition is self-defined, which makes it suitable for the analysis of complex underlying phenomena. In addition, since EMD operates in the time domain it favours the exploration of silently evolving trends such as low arousal emotions. However, it turns out that adaptivity is a mixed blessing. There are two important problems with regard to uniqueness and mode mixing that arises as a consequence of the adaptive nature of EMD. Evaluating the EMD algorithm on two instances of the same signal may result in incomparable decompositions. Moreover, ea c h mode is not captured in a single decomposition—similar modes are present at various decomposition levels of the same signal. In this work, address these issues by making use of a bivariate extension of EMD (BEMD). BEMD acts on two signals x_I (t) and x_S (t) (as opposed to EMD which acts on a single signal). In this work, the second signal (x_S (t)) is a synthetic waveform, standardized over all emotional states and subjects, and designed to act as a decomposition guide. T h e proposed frame work is comprised of two independent steps:

1. Signal Decomposition:

Huang et al. [3] proposed the Empirical Mode Decomposition as a way to empirically decompose a nonstationary, nonlinear signal into a number of IMFs, each of which represents a distinct oscillatory activity. An IMF is a function satisfying certain explicit properties [3]:

1. The number of extrema and the number of zero crossings must be equal or differ at most by one.

2. The mean of the envelopes defined by the maxima and the minima is zero for every sample.

These rules naturally force one mode of oscillatory activity in the IMF, since between two successive extrema no riding waves are allowed. The algorithm for the detection and extraction of IMFs is adaptive and iterative. Once an IMF is found, it is removed from the signal and the algorithm iterates on the residual in order to find more oscillatory modes. Fast oscillations are detected first. Given a signal x (t), EMD operates as follows:

a. Detect local maxima x max (i) and minima x min (j) of x (t)

b. Interpolate among x max (i) to get an upper envelope x up (t) and x low (t) for minima, respectively.

c. Compute the average o f envelopes $m(t)=\frac{x\,up(t)+x\,low\,(t)}{2}$

d. Subtract from signal u(t)=x(t)-m(t).

This is a sifting process, which is terminated when u (t) meets the IMF criteria. If it does, u (t) will be describing a underlying oscillation of x (t), referred to herein as d (t). EMD continues with sifting on the residual r (t) =x (t)-d (t). The original signal can then be expressed as

$$x(t) = \sum_{i=1}^{N-1} di(t) + r(t)$$

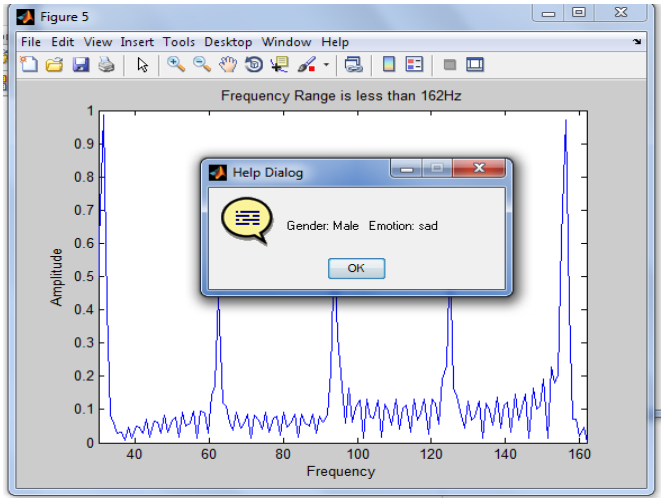
Where d (t) denotes the i^{th} IMF extracted from the signal x (t) and r (t) is the final residual. Note that by definition, r (t) is not an IMF. It has been observed [4], [5], [6], [7] that the first few IMFs carry the quasi periodicity property of signal.

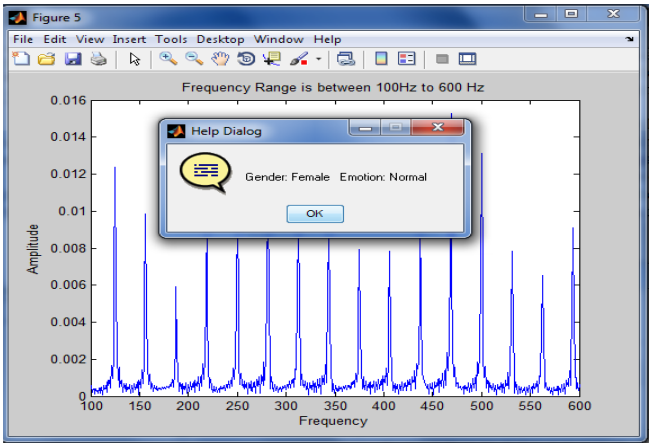
2. Feature Extraction:

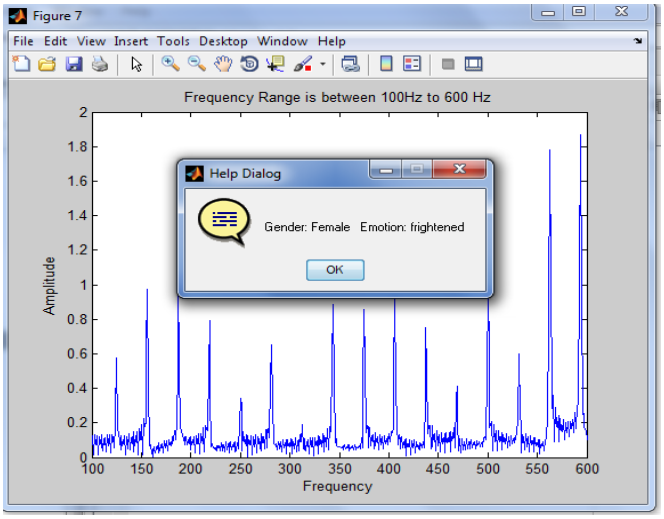
The IMFs are time domain signals carrying information about oscillation activity. Comparisons in the time domain are not straightforward since the IMFs $d_i(t)$ have to be aligned with similar modes from other recordings. Therefore, it is important to design features that summarize the oscillatory activity within every IMF. In this work, use two types of features—the Hilbert instantaneous frequency and a measure of local oscillation

IV. RESULTS AND DISCUSSION

Gender	Emotion	Amplitude
Males	Neutral	0.059-0.166
Males	Sad	0.542-2.33
Males	Happy	0.42-2.85
Males	Anger	0.33-1.72
Males	Fear	0.51-2.45
Females	Neutral	0.057-0.154
Females	Sad	0.51-3.04
Females	Happy	0.41-2.64
Females	Anger	0.28-1.98
Females	Fear	0.46-3.1







V. CONCLUSION

In emotion research, it is very important to collect meaningful data. It is very difficult to design an experimental setup that can induce the same emotion in every subject, especially if the same stimuli are used across all subjects. Different characters, varying moods, and the inability to accurately self-report an emotional experience may significantly affect the outcome of a study.

Despite the difficulties, establishing emotions from internal manifestations of the body is worthwhile for human computer interaction systems. For the naive user, hiding emotions with regard to cardiac reactions is difficult, while, for behaviourally suppressive individuals, physiological patterns can provide hints of emotion. The obtrusive nature of the acquisition, however, poses restrictions on the number of sensors that can be worn and subsequently to the signals that can be collected. The majority of the approaches in the literature rely on fusion of various physiological sources for emotion detection.

To determine multiple F0 in mixed signals and gender identification of those F0'S. The basic idea of our approach is to identify the predominant pitch and then subtract it from the signal to get the residual signal for next iteration. Our approach results the number of persons in the signal along with their gender. The experiments results show that our approach performs well for human voiced signal and is able to identify their gender. The proposed system identifies emotions with gender.

REFERENCES

1. K. Kim, S. Bang, and S. Kim, "Emotion Recognition System Using Short-Term Monitoring of Physiological Signals," Medical and Biological Eng. and Computing, vol. 42, no. 3, pp. 419-427, May 2004.
2. R.L. Mandryk and M.S. Atkins, "A Fuzzy Physiological Approach for Continuously Modeling Emotion during Interaction with Play Technologies," Int'l J. Human Computer Studies, vol. 65, no. 4, pp. 329-347, 2007.
3. N.E. Huang, Z. Shen, R.R. Long, M.L. Wu, Q . Zheng, N.C. Yen, and C.C. Tung, "The Empirical Mode Decomposition and Hilbert Spectrum for Nonlinear and Nonstationary Time Series Analysis," Proc. Royal Soc. London, vol. 454, pp. 903-995, 1998.
4. C. Zong and M. Chetouani, "Hilbert-Huang Transform Based Physiological Signals Analysis for Emotion Recognition," Proc. IEEE Int'l Symp. Signal Processing and Information Technology, pp. 334-339, Dec. 2009.
5. A. Arafat and K. Hasan, "Automatic Detection of ECG Wave Boundaries Using Empirical Mode Decomposition," Proc. IEEE Int'l Conf. Acoustics, Speech, and Signal Processing, pp. 461-464, Apr. 2009.
6. Human Voice. Available at: http://en.wikipedia.org/wiki/Human_voice.
7. [6] J. X. Zhang, M. G. Christensen, S. H. Jensen, and M. Moonen, "A Robust and Computationally Efficient Subspace-Based Fundamental Frequency Estimator," *Transactions on Audio, Speech, and Language Processing, IEEE*, vol. 18, no. 3, pp. 487 –497, March 2010.
8. W. W. Zhao and T. Ogunfunmi, "Formant and Pitch Detection Using Time Frequency Distribution," *International journal of speech technology*, pp. 35-49, 1999.

BIOGRAPHIES

Dr. A. Narendra Babu is a professor in Department of Electronics & Communication Engineering since 2009 at Lakireddy Balireddy College of Engineering, Mylavaram, Krishna District, Andhra Pradesh. He is young Scientist in DST Project. His research area is **Atmospheric Sciences using Radar & Satellites and Cognitive studies.** He is a life member IETE, Indian Science Congress, and ISTE.

A.Jyothi is a M.Tech (Systems & Signal Processing) Student in Department of Electronics & Communication Engineering at Lakireddy Balireddy College of Engineering, Mylavaram, Krishna District, Andhra Pradesh.

