Volume 6, Issue 5, May 2016 ISSN: 2277 128X
International Journal of Advanced Research in Computer Science and Software Engineering
Research Paper
Available online at: www.ijarcsse.com

Feature Extraction Approach in Emotional Speech Recognition System

Jay Patadia, Alpa Reshamwala
Department of Computer Engineering, NMIMS, Mumbai,
Maharashtra, India

Abstract— Technology is playing a good role in human's day to day life. Use your devices by your own voice are one of the favorite tasks which people wants to go with. On the other hand speech has biometric characteristics so; it contains normal features and behavioral biological features. This type of features cane be extracted based on their application. Normal features are used for some simple application like speech to text. But behavioral type of features has emotion which reflects mankind nature. Emotion can be detected by human easily but computer cannot detect that easily. So, this type of high-tech features is used for understanding person's mood. This paper is written for explaining speech recognition system and extraction of features from it. This document also infers that emotion identification is detected by any devices and also useful for making some intelligent system for helping humans for their life. In the end conclusion come with suggestion that speech recognition is better for understanding human behavioral according to their mood and it can also understandable by any devices.

Keywords— speech recognition techniques, feature extraction in speech recognition, Intelligent application

I. INTRODUCTION

Voice is a part of biometric and it contains behavioural information. Voice processing can be done by two types [5]

- Speech recognition
- Speaker Identification

In speech recognition it recognizes the speech what user is speaking an in speaker identification it identify the user who is speaking. For any devices speech recognition or identification cane be done by two type of techniques,

- Text dependent

In this type of recognition text are decided already for making this system and it will work on that predefined words or command only. So, Users are restricted to use that command only.

- Text Independent

In this type of recognition there is no predefined text and it can work for any word or sentence. So, users are free to use it.
For any speech recognition system works under same flow as discussed below
From the figure-1 we can show some basic steps

- Acoustic analysis of single- to analyse the signal
- Digitization – convert analogue signal to digital signal
- Remove noise – remove noise for better performance or use noise cancelation mice
- Normalization – normalize all the data for ease of extraction
- Feature Extraction – recognize speech as per application concern by different algorithm and techniques.
- Database generation – generate training and testing phase database
- Matching – match query input with generated data base for generating output
- Output – check generated output is appropriate or not.

As I have discussed in upper part that for feature extraction in speech recognition can be done as per the application concern. It should be dependent on what type of application we want to implement. Here this paper is discussed about emotional features and how to get it extracted from speech.
In Affected Science emotion is classify in 24 different scales and extracting this all feature by using algorithms like Mel Frequency Cepstral Co-efficient (MFCC), Support Vector Machine (SVM) , AdaBoost, Deep Neural Network (DNN), Aerosol and valence method and etc. are suggested by many author In referred paper.[4, 5, 6, 18, 19, 20]
Following paper contain information about Mel Frequency Cepstral Co-efficient (MFCC),which is useful methods for feature extraction.

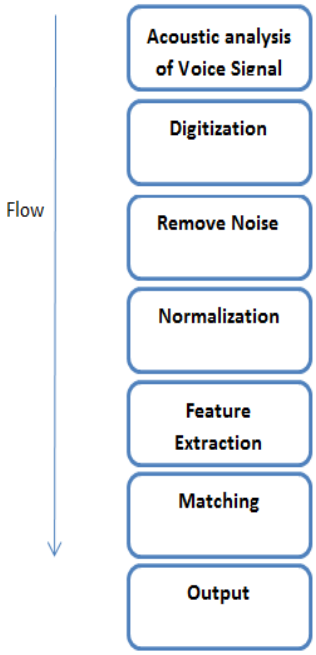

Figure-1 simple Flow diagram

II. RELATED ALGORITHM

A Mel Frequency Cepstral Co-efficient (MFCC)

Feature extraction by Mel Frequency Cepstral Coefficient (MFCC) is based on human hearing perception [6, 7]. MFCC is used for feature extraction frame by frame in speech recognition. It has two types of filter which are spaced linearly at low frequency below 1000 Hz and logarithmic spacing above 1000Hz. A subjective pitch is present on Mel Frequency Scale to capture important characteristic of phonetic in speech. It has six computational steps as follow in figure-2.

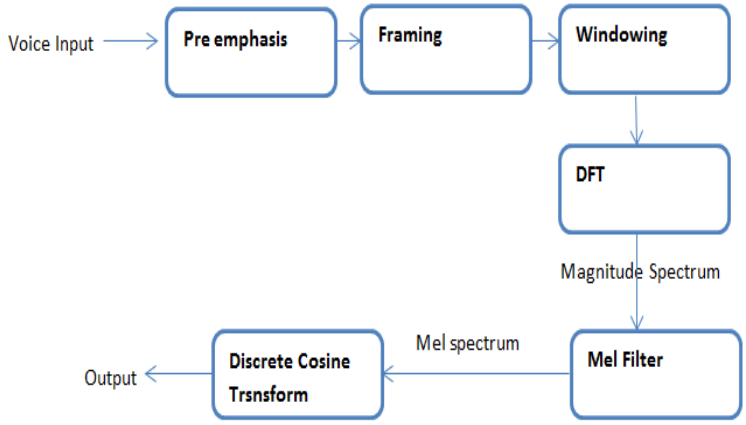

Figure-2 MFCC steps

- Pre-emphasis
  It refers to a filter within some frequency band and function of this step is to emphasize the higher frequencies and to increase energy of the signal.
- Framing
  In this step divides large signal into small-small frames with considering some overlapping. Since signal is not stable so have to make it stable for some time period for extracting feature, this time period is call as frame rate too.
- Windowing
  Humming windowing is for windowing and in this step signal is multiply by window function. It joints signal at the edges of the previous frame and avoid discontinuity in the beginning as well as in the end.
- DFT(Discrete Fourier Transform)
  In this step we are using fast Fourier Transform because it is computationally efficient algorithm of DFT and it was used because each and every number of N sample frames has to be converting from time domain to frequency domain.
- Mel Filter
  Mel filter bank consists of triangular filters which are used to calculate the weighted sum of the filter spectral components [7]. Mel scale is approximated by output and given frequency is computed as a Mel.

- Discrete cosine Transform(DCT)
  DCT is use in this step to convert Mel log spectrum into time domain. After that getting result is call as MFCC. The set of coefficient is call acoustic vector. DCT is applied to the log spectral-energy vector.

III. EXPERIMENT AND RESULTS

Feeling recognition from discourse is not that much simple errand. Here, I have utilized database of Torronto which is accessible on reference [24]. In this database there are 8 user and every users 200 speech files.

Table -1 dataset

User	**Young and Old**			
Emotion	***Happy***	***Sad***	***Natural***	***Angry***
Total wav file	200	200	200	200

After collecting database our next step is to extract behavioural features using MFCC feature extraction. For generating feature database we have use MFCC feature extraction, for that input is emotional voice file and output is 24 emotional feature of that file. It means each file give 24 features as feature vector.
After generating the feature database next step is to find which emotion is present. For find the emotion we classify the generated feature database based on emotional database. It is like each emotion has its own feature database and our input signal will be compared with this entire feature database. Next step is to check the emotion detection rate, so for that have select 30 files per emotion and get truly emotion detection table which is shown in table -2. Also figure 3 shows bar chart of that table so it looks perfect for analysis.

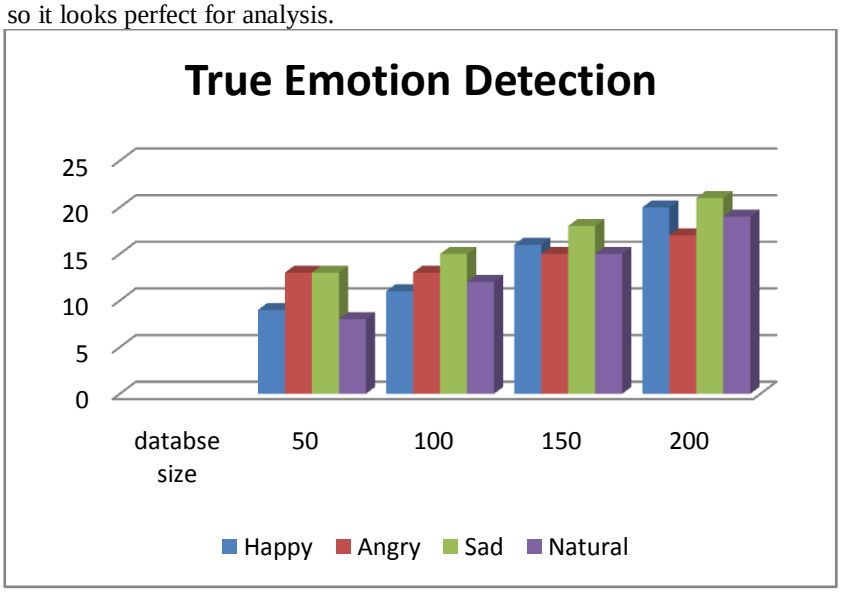

Figure-3 Bar Chart of true emotion detection

Figure-3 shows continuous increment in emotion detection so if we increase the size of database for feature extraction it will definitely increase the rate of true emotion detection and getting the percentage table of true emotion detection which shows data is in table -3.

Table -2 True emotion Detection

	True emotion detection out of 30 input			
	Happy	**Angry**	**Sad**	**Natural**
Database size				
50	9	13	13	8
100	11	13	15	12
150	16	15	18	15
200	20	17	21	19

Table -3 Percentage Table of True Emotion Detection

	Happy (%)	**Angry (%)**	**Sad (%)**	**Natural (%)**
Database Size				
50 file	30	43.33333	43.33333	26.66667
100 file	36.66667	43.33333	50	40
150 file	53.33333	50	60	50
200 file	66.66667	56.66667	70	63.33333

IV. CONCLUSIONS

Speech recognition system is widely use in today's technology for developing hands-free devices. use of our voice and get features from it and according to that feature we can identify user and recognition of speech. Beyond this, speech is also containing behavioural biometric information and from that emotion related features can be extracted by different techniques. This paper also give experiment and result which shows that if we increase the size of database it will definitely increases the rate of True emotion detection. Apart from this after getting emotion we can identify persons mood and accordingly we can make an application or intelligent system which will help human for their routine and give batter lifestyle.

ACKNOWLEDGMENT

This literature review was supported by my mentor Prof. AlpaReshamwala. I am grateful to her for sharing her pearls of wisdom with me during this literature review.

REFERENCES

[1] SadaokiFurui, KiyohiroShikano, ShoichiMatsunaga, Tatsuo Matsuoka, Satoshi Takahashi, and Tomokazu Yamada, RECENT TOPICS IN SPEECH RECOGNITION RESEARCH AT NTT LABORATORIES. 1994.

[2] Raja Sukumar A and SarinSukuma A, and Firoz Shah A and Babu A into P Key-Word Based Query Recognition In a Speech Corpus By Using Artificial Neural Networks , Second International Conference on Computational Intelligence, Communication Systems and Networks in 2010.

[3] NiladriSekharDey, RamakantaMohanty and K. L. chugh Speech and Speaker Recognition System using Artificial Neural Networks and Hidden Markov Model, International conference on Communication Systems and Network Technologies(IEEE) 2012

[4] Jasdeep Singh Bhalla and AnmolAggarwal, Using Adaboost Algorithm Along with Artificial Neural Networks for Efficient Human Emotion Recognition From Speech, IEEE, 2013

[5] Suma Swamy and K.V Ramakrishnan,, AN EFFICIENT SPEECH RECOGNITION SYSTEM ,Computer Science and Engineering: An International Journal (CSEIJ), Vol. 3, No. 4, August 2013

[6] LindasalwaMuda, MumtajBegam and I. Elamvazuthi, "Voice Recognition Algorithms using Mel Frequency Cepstral Coefficient (MFCC) and Dynamic Time Warping (DTW) Techniques,JOURNAL OF COMPUTING, VOLUME 2, ISSUE 3, MARCH 2010, ISSN 2151-9617.

[7] Dalmiya C.P, Dr.Dharun V.S and Rajesh K.P," An Efficient Method for Tamil Speech Recognition using MFCC and DTW for Mobile Applications", Proceedings of 2013 IEEE Conference on Information and Communication Technologies (ICT 2013).

[8] Martin Borchert and Antje Diisterhoft, "Emotions in Speech - Experiments with Prosody and Quality Features in Speech for Use in Categorical and Dimensional Emotion Recognition Environments", Proceeding ofNLP-KE in 2005.

[9] Sabine Deligne, SatyaDharanipragada, Ramesh Gopinath, BenoîtMaison, Peder Olsen and Harry Printz, "A Robust High Accuracy Speech Recognition System for Mobile Applications", IEEE TRANSACTIONS ON SPEECH AND AUDIO PROCESSING, VOL. 10, NO. 8, NOVEMBER 2002.

[10] Kil-Ram Ha, Jung-Hyun Kim, Jeh-SeonYoun and Kwang-Seok Hong, "Speech Recognition-Based Mobile Geo-Mashup Application Technology", Third International Symposium on Intelligent Information Technology Application in 2009.

[11] George E. Dahl, Dong Yu, Li Deng and Alex Acero, "Context-Dependent Pre-Trained Deep Neural Networks for Large-Vocabulary Speech Recognition", IEEE TRANSACTIONS ON AUDIO, SPEECH, AND LANGUAGE PROCESSING, VOL. 20, NO. 1, JANUARY 2012.

[12] SafAsghar and Lin Cong, "ROBUST SPEECH RECOGNITION FOR MOBILE APPLICATIONS", IEEE in 1999.

[13] Dr. H. B. Kekre and VaishaliKulkarni, "Comparative Analysis of Automatic Speaker Recognition using Kekre's Fast Codebook Generation Algorithm in Time and Transform Domain", International Journal of Computer Applications (0975 – 8887) inVolume 7– No.1, September 2010.

[14] DR. H. B. Kekre and VaishaliKulkarni, "Performance Comparison of Speaker Recognition using Vector Quantization by LBG and KFCG", International Journal of Computer Applications (0975 – 8887) in Volume 3 – No.10, July 2010.

[15] Ms.Rupali S Chavan and Dr. Ganesh. S Sable, "An Overview of Speech Recognition Using HMM", International Journal of Computer Science and Mobile Computing, IJCSMC, Vol. 2, Issue. 6,, pg.233 – 238 in June 2013

[16] Ashok Shigli, Ibrahim Patel and Dr. K. SrinivasRao, "A SPECTRAL FEATURE PROCESS FOR SPEECH RECOGNITION USING HMM WITH MFCC APPROACH", National Conference on Computing and Communication Systems (NCCCS) in 2012

[17] Hemakumar G and Dr.Punitha P., "Speaker Dependent Continuous Kannada Speech Recognition Using HMM", 2014 International Conference on Intelligent Computing Applications in 2014.

[18] Sanghoon Jun, Seungmin Rho, Byeong-jun Han and Eenjun Hwang, "A Fuzzy Inference-based Music Emotion Recognition System", Institution of Engineering and Technology, page no: 673-677, 2008

[19] Igor Bisio, Alessandro Delfino, Fabio Lavagetto, Mario Marchese and Andrea Sciarrone, "Gender-Driven Emotion Recognition Through Speech Signals for Ambient Intelligence Applications", IEEE Transactions On Emerging Topics In Computing, Volume 1, No. 2, December 2013

[20] NorhaslindaKamaruddin, Abdul Wahab Abdul Rahman and NorSakinah Abdullah. "Speech Emotion Identification Analysis based on different Spectral Feature Extraction Methods", Information and Communication Technology for The Muslim World (ICT4M), 2014 The 5th International Conference, 17-18 Nov. 2014 PP. 1-5

[21] Naoyuki Kubota, Fumio Kojima and Toshio Fukuda, "Self-Consciousness and Emotion for A Pet Robot with Structured Intelligence", in IEEE ,2001 PP. 2786-2791

[22] Yoav Freund and Robert Schapire, "Adaboost" from site: "http://en.wikipedia.org/wiki/AdaBoost"

[23] Jameslyons,"MFCC feature Extrection" site: http://practicalcryptography.com/miscellaneous/machine-learning/guide-mel-frequency-cepstral-coefficients-mfccs/

[24] Emotional Speech Database site : https://tspace.library.utoronto.ca/handle/1807/24490

