

## A Review: Speech Emotion Recognition

G.N. Peerzade<sup>1\*</sup>, R.R. Deshmukh<sup>2</sup>, S.D. Waghmare<sup>3</sup>

<sup>1\*</sup>Department of CS and IT, Dr. B. A. M. U, Aurangabad (MS), India

<sup>2</sup>Department of CS and IT, Dr. B. A. M. U, Aurangabad (MS), India

<sup>3</sup>Department of CS and IT, Dr. B. A. M. U, Aurangabad (MS), India

\*Corresponding Author: [peerzadegulnaz@gmail.com](mailto:peerzadegulnaz@gmail.com)

Available online at: [www.ijcseonline.org](http://www.ijcseonline.org)

Received: 23/Feb//2018, Revised: 28/Feb/2018, Accepted: 21/Mar/2018, Published: 30/Mar/2018

**Abstract**— In Human Computer Interaction (HCI) area speech emotion recognition is one of the popular topic in the world. Many researchers are engaged in developing systems to recognize different emotions from human speech. This is done to make HCI and human interface more effective and develop systems like humans. In this paper we have stated the basics of speech emotion recognition system and reviewed different feature extraction and classification technique for the system. Features are classified as Elicited features, Prosodic features and Spectral features. Different classifying techniques are used to classify different emotions from human speech like Hidden Markov Model (HMM), Gaussian Mixtures Model (GMM), Support Vector Machine (SVM), Artificial Neural Network (ANN), K-nearest neighbor (KNN). Performance of classifiers are also discussed shortly. Different applications where speech emotion recognition systems are used are also discussed in last section of the paper.

**Keywords**— Emotion, Speech, Emotional Speech database, Elicited features, HMM, GMM, SVM, ANN, KNN, Application.

### I. INTRODUCTION

Speech processing is one of the most widely used research area. Many researchers around the world is working on different technologies in speech processing. Speech processing came into existence in 1920 when a celluloid toy was made called 'Radio Rex'. This toy was the first speech recognition machine which works on acoustic energy released by the vowel 'Rex' i.e. 500 Hz. Speech recognition system was firstly developed in 1952 by Davis in Bell Laboratory, US which recognizes digits from 0-9 in male voice. Speech processing contributed too many challenges to researchers like continuous speech recognition, emotion recognition etc. Emotions are not only understood by facial expressions but also with speech. Every speech of human being is associated with an emotion. Emotions are very important as they allows an individual to understand the feelings. Speech shows the emotions of the persons that whether they are happy or sad etc. Thus human computer interaction (HCI) came to a new challenge of identification of emotions by speech. HCI should be more specific to understand the human natural emotions. There are different types of human emotion like yelling, crying, dancing, laughing, stamping, teasing etc. [1].

In speech emotion recognition system different feature extraction techniques are used and to recognize emotions by using different classifiers. Three main type of features are classified as Elicited features, Prosodic features and Spectral

features [2]. For spectral features different technologies are used like MFCC, LPCC and MEDC. For prosodic features technologies used are pitch, intensity, fundamental frequency, loudness, glottal parameters etc. Classification techniques for classifying emotions are used such as Hidden Markov Model (HMM), Gaussian Mixtures Model (GMM), Support Vector Machine (SVM), Artificial Neural Network (ANN) etc. [3].

This paper contains sections like complete description of speech emotion recognition system and then follows the feature extraction techniques, classification techniques and different applications. Last section comprise conclusion.

### II. SPEECH EMOTION RECOGNITION SYSTEM

Emotion recognition system targets at finding different emotions from human speech. There are modules in speech emotion recognition system given in fig. 1 such as speech corpus (text), emotional speech input, pre-processing, feature extraction and selection, classification and emotional speech output. Six basic type of emotions can be classified by using speech emotion recognition system [4]. These are fear, anger, neutral, sadness, happy and surprise [5, 6, 7]. In speech corpus (text) a text corpus is made for effective future classification so that same text is repeated by different individuals.

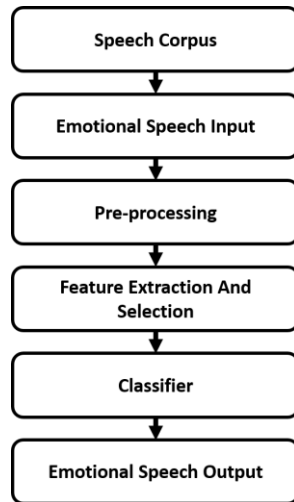


Figure 1. Speech Emotion Recognition System

Emotional Speech Input, this step consists of speech samples which are collected from different individuals. These individuals speak these words with different emotions according to categorizations of words for different emotions [8]. In pre-processing noise is removed from samples. These noise are parts which are not needed even spaces are considered as noise. This is done by using software like PRAAT and Audicity.

### III. FEATURE EXTRACTION

Human speech consists of many parameters which shows the emotions comprise in it. As there is change in emotions these parameters also gets changed. Hence it's necessary to select proper feature vector to identify the emotions. Features are categorized as excitation source features, spectral features and prosodic features. Excitation source features are achieved by suppressing characteristics of vocal tract (VT). Spectral features used for emotion recognition are Linear prediction coefficients (LPC), Perceptual linear prediction coefficients (PLPCs), Mel-frequency spectrum coefficients (MFCC), Linear prediction cepstrum coefficients (LPCC), perceptual linear prediction (PLP). The accuracy of differentiating different emotions can be achieved by using MFCC, LFPC, LPC, PLP, and RASTA-PLP [9, 10]. Prosodic features used for emotion recognition are pitch, energy, intensity. Statistical measurements are also used to distinguish emotions like minimum, maximum, standard deviation, range, mean, median, variance, skewness, kurtosis etc. of features.

**Linear prediction cepstrum coefficients (LPCC):** LPCC helps us to get characteristics of channel of speech. There are different channel characteristics of a person through different emotional states. LPCC is used to define vowels more effectively.

**Mel-frequency spectrum coefficients (MFCC):** MFCC has a simple calculation, anti-noise, good ability of distinction and many other advantages. It is commonly used feature of speech. There are different steps (fig 2) in MFCC feature extraction such as pre-processing, framing, windowing, FFT, Mel Filterbank and Frequency wrapping, logarithm, discrete cosine transform. [11]

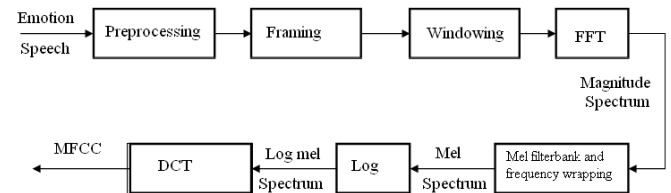


Figure 2. MFCC feature extraction

### IV. CLASSIFIERS

After extracting features of speech, it is essential to select a proper classifier in speech emotion recognition. Classifiers are used to classify emotions. Different classifiers used are Hidden Markov Model (HMM), Gaussian Mixtures Model (GMM), Support Vector Machine (SVM), Artificial Neural Network (ANN) etc. Classifiers can also be combined and used in making a hybrid model.

**SVM:** It is an algorithm used in classification and pattern recognition problems. This classifier is used to separate features from other features. The data is being categorize by constructing N-dimensional hyper-plane which gives larger functional margin i.e. large distance between nearest data points [12].

**HMM:** Hidden Markov Model is also called as First order markov chain. In this the internal working is hidden from the observer. HMM is used to catch temporal structure of data. HMM differentiates between different emotions by getting trained with extracted features. The input given to the SVM is the output of this model [13].

### V. APPLICATION

There are many applications of speech emotion recognition system. Few of them are listed here intelligent toys, psychiatric diagnosis, conversation with robots, mobile based speech recognition system, emotion recognition in call centre where emotions of customer can be identified and can help to get better quality of service, intelligent tutoring system, in car board system, prosody in dialog system, sorting of voice mail, computer games, lie detection [14, 15].

### VI. CONCLUSION

In this paper, we have reviewed the basics of speech recognition system. We have also given a short description of

feature extraction techniques and classifiers. MFCC is most commonly used feature extraction technique.

#### ACKNOWLEDGMENT

This work is supported by University Grants Commission., Aurangabad, Maharashtra, India. The authors would like to thank Dept. of Computer Science and Information Technology, Dr. Babasaheb Ambedkar Marathwada University Authorities for providing the infrastructure and necessary support for carrying out the research.

#### REFERENCES

- [1] V. B. Waghmare, R. R. Deshmukh, P. P. Shrishrimal, G. B. Janvale, "Emotion Recognition System from Artificial Marathi Speech using MFCC and LDA Techniques," Proc. of Int. Conf. on Advances in Communication, Network, and Computing, CNC, Elsevier, 2014.
- [2] R. A. S. Nilofer, R. P. Gadhe, R. Deshmukh, V. B. Wasghmare, P. P. Shrishrimal, "Automatic emotion recognition from speech signals: A review", International Journal of Scientific and Engineering Research, vol. 6, no. 4, April 2015.
- [3] S. Emerich, E. Lupu, A. Apatean, "Emotions Recognitions by Speech and Facial Expressions Analysis", 17th European Signal Processing Conference, 2009.
- [4] A. Mohanta and U. Sharma, "Bengali speech emotion recognition," 2016 3rd International Conference on Computing for Sustainable Global Development (INDIACom), New Delhi, 2016, pp. 2812-2814 (IEEE).
- [5] D. D. Joshi, M. B. Zalte, "Speech Emotion Recognition: A Review", IOSR Journal of Electronics and Communication Engineering (IOSR-JECE), pp. 34-37, January-February 2013.
- [6] A Mohanta, U Sharma, "Human Emotion Recognition through Speech", Advances in Computer Science and Information Technology (ACSIT), vol. 2, no. 10, pp. 29-32, April-June 2015.
- [7] A. S. Utane, S. L. Nalbalwar, "Emotion Recognition through Speech", International Journal of Applied Information Systems (IJ AIS), pp. 5-8, 2013.
- [8] Ayadi, M. E., Kamel, M. S., & Karray, F. "Survey on speech emotion recognition: features, classification schemes, and databases". Pattern Recognition, 44, 572-587 2011..
- [9] Idris I., Salam M.S. "Improved Speech Emotion Classification from Spectral Coefficient Optimization". Lecture Notes in Electrical Engineering, vol 387. Springer, 2016.
- [10] Pao TL., Chen YT., Yeh JH., Cheng YM., Chien C.S. "Feature Combination for Better Differentiating Anger from Neutral in Mandarin Emotional Speech", LNCS: Vol. 4738 Berlin:Springer 2007.
- [11] Yashpalsing Chavhan, M. L. Dhore, Pallavi Yesaware, "Speech Emotion Recognition Using Support Vector Machine", International Journal of Computer Applications, vol. 1, pp.6-9, February 2010.
- [12] A.P. Wanare, S.N. Dandare, "Human Emotion Recognition From Speech", Int. Journal of Engineering Research and Applications, vol. 4, no. 7, pp. 74-78, July 2014.
- [13] A. Joshi, "Speech Emotion Recognition Using Combined Features of HMM & SVM Algorithm," International Journal of Advanced Research in Computer Science and Software Engineering, pp. 387-392, 2013.
- [14] A. B Ingale and D. S. Chaudhari, "Speech Emotion Recognition", International Journal of Soft Computing and Engineering (IJSCE), March 2012, pp. 235-238.
- [15] A. S. Utane and S.L. Nalbalwar, "Emotion Recognition through Speech", International Journal of Applied Information Systems (IJ AIS), 2013, pp. 5-8.

#### Authors Profile

*G. N. Peerzade*, Pursued Bachelor of Engineering in Computer science and Engineering in Savitribai Phule Womens Engineering College, Aurangabad (MS) 431001 India in 2016, Pursuing Master of Technology under the Faculty of Engineering and Technology in Computer Science and Engineering in Department of Computer Science and IT, Dr. Babasaheb Ambedkar Marathwada University, Aurangabad (MS) India.



*Dr. R. R. Deshmukh*, Professor and Former Head of Department of Computer Science and IT, Dr. Babasaheb Ambedkar Marathwada University, Aurangabad (MS) 431001 India. His research interests are Human computer Interaction, Digital Speech Signal processing, Data Mining, Image Processing, Pattern Recognition, artificial Intelligence, Computational Auditory Scene Analysis (CASA), Neural Networks, GIs and Remote



Sensing, Sentiment Analysis etc. He is Fellow member and Chairman of IETE Aurangabad Centre, The Institutions of Electronics and Telecommunication Engineers (FIETE). Senior Member of Association of Computer Electronics and Electrical Engineers (ACEEE). Life member of CSI (Computer Society of India), The Indian Society for Technical Education (ISTE), Indian Science Congress Association (ISCA), Institute of Doctors Engineers and Scientists (IDES). Member of International Association of Engineers (IAEng), computer science Teachers Association. (CSTA), IEEE. DST-FIST Program Coordinator, Coordinator of Global Initiative on Academic Networks (GIAN) Coordinator of Maharashtra. State Marathi World Dictionary Development Committee, Govt. of Maharashtra in the subject of Information Technology and Computer Science. Edited ten books and published more than 165 research papers.

*S.D. Waghmare*, Research Fellow, Department of Computer Science and IT, Dr. B. A. M. University, Aurangabad (MS).

