Comprehensive Survey on Recognition of Emotions from speech

[1]Khushboo Mittal, [2]Parvinder Kaur
[1]Student, [2]Asst.Proffesor
[1]Computer Science and Engineering
[1]Shaheed Udham Singh College of Engineering and Technology, Tangori, Punjab

---

***Abstract* - Emotional state of human is very important in the medical field. Emotions refer to the complex psychological process of human being. This could help in human interaction with environment. There are various commonly used methods in emotional recognition from speech. The use of ineffective technique is one of the main drawback which is faced in emotion recognition in speech database. So, this paper has described various types of methods that have been used for the recognition of emotions from the speech. In addition to this various speech features and model has been discussed in the paper.**

***Keywords* - Emotion Recognition, Happy, Sad, Anger, Feature Extraction, Classification.**

---

1. INTRODUCTION

Emotion speech is a crucial area of interest in Human Communication Interface systems [1]. The system firstly recognizes the emotions then work accordingly [2]. There are various modules that made a HCI system like speech to text transformation, feature extraction, database, classification etc [3]. The main step of emotional recognition is the use of effective database. There are some rules that must be followed in the emotion recognition process like:

- Real-world emotions or acted ones
- Who utters the emotions How to simulate the utterances
- Balanced utterances or unbalanced utterances
- Utterances are uniformly distributed over emotions [4]

Emotions play an extremely important role in human mental life. It is a medium of expression of one's perspective or his mental state to others. It is a channel of human psychological description of one's feelings [5]. Emotions are a key part of speech. Automatically detecting emotion in a recording can enhance human computer interaction. It also enables various other kinds of analyses, such as search for paralinguistic phenomena, the honesty of the speaker, etc.

In human interactions there are many ways in which information is exchanged (speech, body language, facial expressions, etc.). A speech message in which people express ideas or communicate has a lot of information that is interpreted implicitly [6]. This information may be expressed or perceived in the intonation, volume and speed of the voice and in the emotional state of people, among others. The speaker's emotional state is closely related to this information. In evolutionary theory, it is widely accepted the "basic" term to define some emotions. The most popular set of basic emotions: happiness (joy), anger, fear, boredom, sadness, disgust and neutral [7].

Earlier researchers has utilized many techniques in this context like HMM (Hidden Markov Model), SVM (Support Vector Machines), GFCC) gammatone frequency cepstral coefficients), MFCC (Mel frequency cepstral coefficients), GMM (Gaussian Mixture Model). Each technique has its own advantages and drawbacks according to the database used.

The remaining paper is organized as Section 2 will gives the overview of emotional recognition system, Section 3 describes the emotional speech properties, Section 4 includes commonly used algorithms with their steps and finally in the end conclusion is described.

2. FRAMEWORK OF EMOTIONAL RECOGNITION SYSTEM

Below figure describes that firstly input speech signal is fed to the features extraction module [8]. In feature extraction module various types of features that has been described in section 3 will be extracted from speech signal. After that feature selection will be done. In the end classification will be done that uses various types of classifiers like HMM, SVM, NN etc.

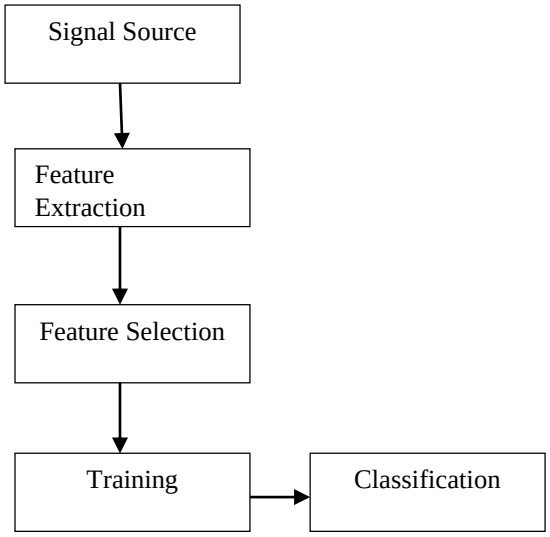

Figure1. Emotion Recognition Model

3. CHARACTERISTICS OF EMOTIONS IN SPEECH

Table1. Features of Speech

Features	Research Group	HAPPY	AGRESSIVE	SAD
Bandwidth	Kwon et.al [8]	Low	Low	Very low
Pitch	Petrushin [9]	High	Very high	Very low
Energy	Altun et.al [10]	High	Very high	Very low
Intensity	Petrushin [9]	High	Very high	Low
Speech rate	Shami et.al [10]	High	High	Low
Spectral features	Rong et.al [11]	High	high	Very low

4. EMOTIONAL SPEECH RECOGNITION METHODS

4.1 GTCC (Gammatone Cepstral Coefficient)

Gammatone Cepstral Coefficient is another FFT-based feature extraction technique in speech recognition systems [12]. The technique is based on the Gammatone filter bank, which attempts to model the human auditory system as a series of overlapping band pass filters. Figure 2 shows the block diagram of GTCC

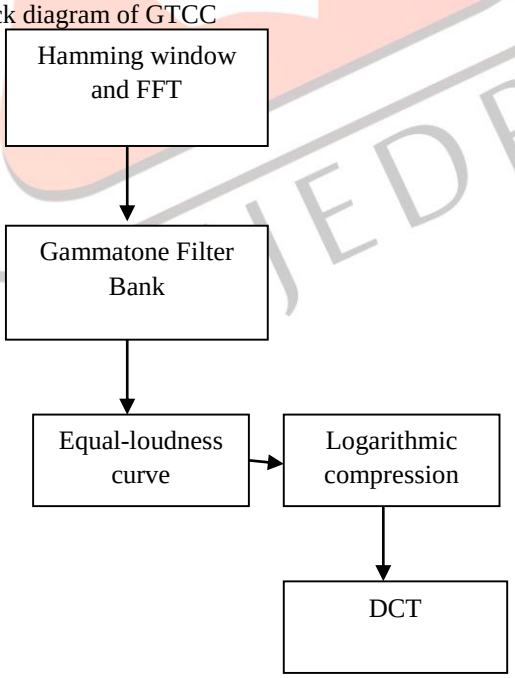

Figure2 GTCC Process

Algorithmic Steps:

- Divide speech into frames
- Convert frames from time domain to frequency domain
- Use of Gammatone filters
- Application of logarithmic to get features of loudness
- Find DCT

Drawbacks

1 The reduction of dimensionality is not given.
2 The main disadvantage lies in choosing an appropriate window size.

4.2 GA (Genetic Algorithm)

GAs were first described by John Holland in the 1960s and further developed by Holland and his students and colleagues at the University of Michigan in the 1960s and 1970s. Genetic algorithms (GAs) are computer programs that take off the processes of biological growth in order to explain problems and to make evolutionary systems [13]. Specify the problem to be solved and a bit-string illustration for candidate solutions, the simple GA works as follows:

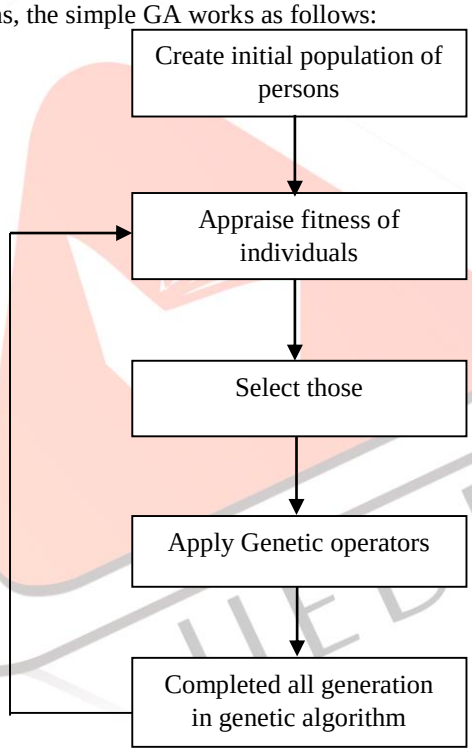

Figure 3 Genetic Algorithm Process

Algorithmic Steps:

[Start] Generate irregular populace of n chromosomes

[Fitness] Evaluate the wellness f(x) of every chromosome x in the populace

[New population] Create another populace

[Selection] Select two guardian chromosomes from a populace

[Crossover] with hybrid likelihood traverse the folks to shape posterity (kids).

[Mutation] with a transformation likely to change new posterity at every locus (position in chromosome).

[Replace] Use new produced populace for a further run of calculation

[Loop] Go to step 2.

Drawbacks

1 Some specific optimization issues could not be resolved using genetic algorithms.
2 There is no unconditional guarantee such that a genetic algorithm will discover a universal optimum.
3 Corresponding several artificial intelligence methods, the genetic algorithm could not possibly promise persistent optimization retort times.
4 Genetic algorithm applications in handles that are accomplished in real time are restricted due to some arbitrary solutions as well as convergence, in some other words this also means that the over-all population is cultivating, nevertheless this could possibly not be supposed for an individual inside this population. For that reason, it is irrational to utilize genetic algorithms intended for on-line controls in real frameworks deprived of any testing them first on a simulation system.

4.3 NN (Neural Network)

Neural network mainly consists of the layers. Layers are made up of nodes. There are mainly three types of layers in the neural architecture: Input layer, hidden layer and output layer. On the basis of the output layer weights are assigned to get output accordingly to input layer. Most of the ANNs contains the learning rule that helps in the maintenance of the weights according output layer. There are many types of the learning rules in neural network but delta rule is common rule that has been used these days [14]. Some useful properties of ANN are:

- Adaptability
- Non-linearity
- Uniformity of analysis and project
- Mapping between Input-Output
- Fault-tolerance

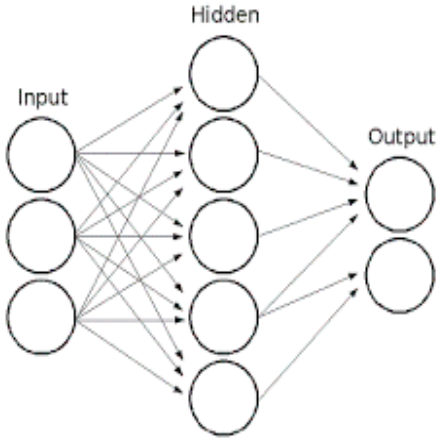

Figure 4 Neural Network Model

Algorithm Steps:

- Assign weights to inputs
- Adjust weights according to output
- Till optimum output has been achieved.
- Stop

Drawbacks

1 The Vapnik–Chervonenkis dimension of neural networks is uncertain.
2 Neural networks could not be reinstructed.
3 Neural networks frequently display patterns related towards those presented through humans.
4 Processing of time series information in neural networks is a very intricate topic.
5 Neural networks are not an auxiliary for understanding any kind of issue intensely.

4.4 PSO (Particle Swarm Optimization)

Particle Swarm Optimization is evolutionary algorithm based on swarms. PSO share many features with other evolutionary algorithms [15]. The system is initialized with number of populations. Then searching for optima is done. Unlike GA, PSO has no operators like mutation, fitness etc. In PSO there are potential solutions called PSO. Each particle in PSO moves after another particle in its space for searching of new solution.

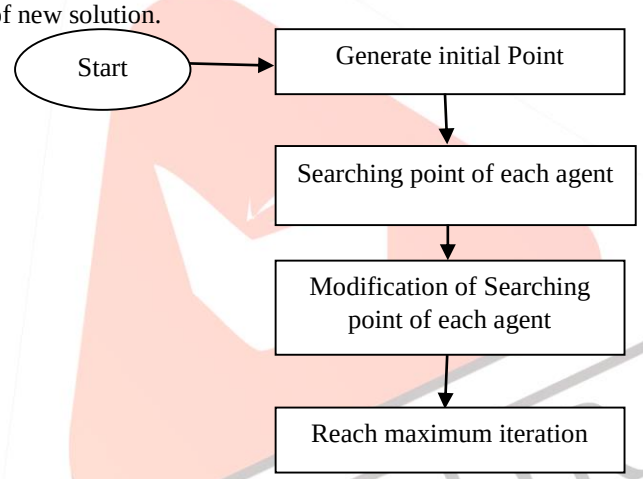

Figure 5 PSO Model

Algorithmic Steps:

- Create initial particles.
- Evaluate objective function of each particle.
- Choose new velocities
- Update each particle location.
- Iterate until solution is reached.

Drawbacks

1 The method without any problems undergoes from the fractional optimism that causes the less exact on the guideline of the aforementioned rapidity as well as the direction.
2 The system could not work out the issues of scattering as well as optimization.
3 The technique could not able to work out the issues of non-coordinate framework, for instance the explanation to the energy area as well as the moving guidelines of the particles in the energy area.

5. CONCLUSION AND FUTURE SCOPE

Recognizing the emotions in speech is a complex process because it was influenced by gender, female and male, culture, environment and experiences. The main and crucial step of Speech Recognition Systems is to recognize the speech effectively. Here word "effectively" means to recognize the speech accurately on the basis of features extracted whether it is SAD, HAPPY etc. Recognizing the emotions in speech is a complex process. Complexity of this process is about that the environment or source that generates speech is difficult to identify due to some environmental factors which includes noise or unwanted signals. This paper describes the process of emotions recognition in speech and also various methods has been discussed that can be used as in combination for future work.

REFERENCES

[1] I. Luengo, E. Navas, I. Hernáez, and J. Sánchez, "Automatic emotion recognition using prosodic parameters", in Proc. INTERSPEECH, 2005, pp.493-496.
[2] D. Neiberg, K. Elenius, and K. Laskowski, "Emotion recognition in spontaneous speech using GMMs", in Proc. INTERSPEECH, 2006.
[3] B. Schuller, G. Rigoll, and M. Lang, "Hidden Markov model based speech emotion recognition," in Acoustics, Speech, and Signal Processing, 2003.Proceedings. (ICASSP '03). 2003 IEEE International Conference on, 2003, pp. II-1-4 vol.2.
[5] N. Amir, O. Kerret, and D. Karlinski, "Classifying emotions in speech: a comparison of methods", in Proc. INTERSPEECH, 2001, pp.127-130.
[6] C. Lili, J. Chunhui, W. Zhiping, Z. Li, and Z. Cairong, "A method combining the global and time series structure features for emotion recognition in speech," in Neural Networks andSignal Processing, 2003. Proceedings of the 2003 International Conference on, 2003, pp.904-907 Vol.2.
[7] C.M. Lee and S. Narayanan, "Emotion Recognition Using a Data-Driven Fuzzy Interface System". In the Proceeding of the European Conference on Speech Communication and Technology, pp: 157-160, 2003.
[9] Petrushin, V. A, "Emotion Recognition in Speech Signal: Experimental Study, Development and Application". In the Proceedings of the International Conference on Spoken Language Processing, pp: 493-496, 2000.
[8] O.W. Kwon, K. Chan, J. Hao, and T. Lee, "Emotion recognition by speech signals," in EUROSPEECH 2003, GENEVA, pp. 125-128, 2003.
[10] H. Altun and G. Polat, "Boosting selection of speech related features to improve performance of multi-class SVMs in emotion detection", presented at Expert Syst. Appl., 2009, pp.8197-8203.
[11] J. Rong, G. Li, and Y.P. Chen, "Acoustic feature selection for automatic emotion recognition from speech", presented at Inf. Process. Manage. 2009, pp.315-328.
[12] X. Valero and F. Alias, "Gammatone Cepstral Coefficients: biologically Inspired Features for Non-Speech Audio Classification, " IEEE Transactions on Multimedia, vol. 14, no. 6, pp. 1684-1689, Dec. 2012.
[13] Shi Hong Chu Combination of GA and SDM to Improve ANN Training Efficiency, 2003: Shu-Te University
[14] V. Petrushin, "Emotion in speech: Recognition and applicationto call centers," in Artificial Neural Networks in Engineering(ANNIE), St. Louis, Missouri, 1999, pp. 7–10.
[15] http://mnemstudio.org/particle-swarm-introduction.htm