Emotion Recognition from Speech: A Survey

Rani P. Gadhe,Shaikh Nilofer R. A., V. B. Waghmare, P. P. Shrishrimal, R. R. Deshmukh

Abstract— Emotion recognition from speech is an important area in research that represents human-computer interaction. The main purpose of this paper is to present literature review of different features and techniques used for speech emotion recognition. The survey represents the importance of choosing different classification model and features for speech emotion recognition. Speech emotion recognition databases are also reviewed in this paper for the purpose of identifying the number of speakers, language used and emotion classification till date.

Index Terms— Classification Model, Emotion Recognition, , Feature Selection, Feature Extraction, Human-Computer Interface, Speech Database, Speech Processing.

1 Introduction

SPEECH is one of the most natural forms of communication between human and computer. Speech signal is one of the fastest methods of communications between humans. Therefore the speech can be more efficient and fast method of interaction between machine and human [1]. Speech is a complex signal which contains information about the message, speaker, language and emotions. An emotion makes speech more expressive and effective. Different ways like laughing, yelling, teasing, crying, etc, are used by humans to express their emotions [2].

Emotion recognition through speech is an area which is increasingly attracting the attention the field of pattern recognition and speech signal processing in recent years. Automatic emotion recognition pays close attention to identify emotional state of speaker from voice signal. It is important medium of expressing humans perspective or fillings and his or hers mental state to others. Humans have natural ability to recognize emotions through speech information but the task of emotion recognition for machine using speech signal is very difficult since machine does not have sufficient intelligence to analyze emotions from speech [3].

Machine can detect who said and what is said by using speaker identification and speech recognition techniques but if we implied emotion recognition system through speech then machine can also detect how it is said [4]. As emotions plays an important role in rational actions of human being there is a desirable requirement for intelligent machine human interfaces for better human machine communication and decision making. Emotion recognition through speech means detection of the emotional state of human through feature extracted from his or her voice signal. Emotion recognition through Speech is particularly useful for applications in the field of human machine interaction to make better human machine interface.

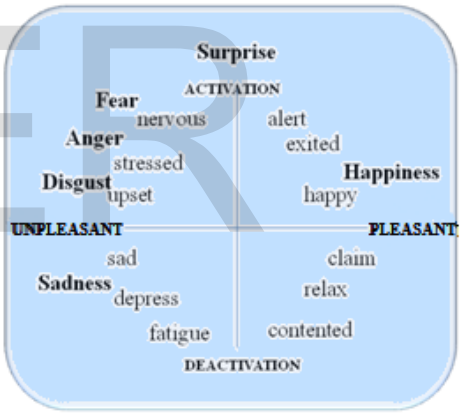

Fig 1: Schematic map of affect [5].

As shown in above Fig. 1, emotion states can be placed within a dimensional model of two or three affective dimensions (Fig. 1). The dimensions are usually valence (from positive to negative) and arousal (from high to low), sometimes a third dimension like stance (from open to close) is added. A dimensional model allows for a continuous description which is very suitable for spontaneous emotions.

Speech emotion recognition has many more applications in our daily life. Some of these applications include:

1. Most important application is for enhancing the interaction between human and machine.
2. In psychiatric diagnosis, lies detection.
3. For analysing the behavioural study in call centre conversation between customer and employee.

- *Author name is currently pursuing masters degree program in electric power engineering in University, Country, PH-01123456789. E-mail: author_name@mail.com*
- *Co-Author name is currently pursuing masters degree program in electric power engineering in University, Country, PH-01123456789. E-mail: author_name@mail.com*
(*This information is optional; change it according to your need.*)

4. In aircraft cockpits, for the better performance speech emotion recognition system trained for stress speech.
5. For understanding the criminals behaviour that would help to criminal department for analysis.
6. Emotion recognition in robotic, it would be more realistic and enjoyable, if they understand and express the emotions like human [6]

The rest of the paper is organized as follows: existing speech corpora i.e. database review, review of different classification techniques used in emotion recognition system and conclusion.

2 Database Review

Emotion speech database collected for the variety in research. Speech corpora used for developing the recognition of emotion from speech system, it can be divided in three ways as follows:

1. Simulated based emotional speech database: Where the database is collected from actors, experienced and trained artists. In this artist express their natural words or sentence in different emotions. This one is the easier way to collect the database. More that 60% of databases are collected in this way.
2. Induced emotional speech database: This type of database is collected by simulating artificial emotional situation, without knowledge of the speaker. This database may be more natural as compare to simulated database, but there may be a problem when speaker know that they are being recorded, then they are not that much of expressive.
3. Natural emotional speech database: These types of emotions are sometimes difficult to recognize. Natural emotions also called underlying emotions. Natural database can be recorded form call centre conversation and emotional conversation between public places and so on.

2.1 Comparision of the Databases

Databases with emotional speech are not only essential for psychological studies, but also for automatic emotion recognition, as standard methods are statistical and need to learn by examples. Generally, research deals with databases of acted, induced or completely spontaneous emotions. Of course, the complexity of the task increases with the naturalness. So at the beginning of the research on automatic vocal emotion recognition, which started seriously in the mid-90s, work began with acted speech and shifts now towards more realistic data [8][7].

Prominent examples for acted databases are the Berlin database of emotional speech [9] and the Danish Emotional Speech corpus (DES) [10] which hold recordings of 10 resp. 4 test persons that were asked to speak sentences of emotionally neutral content in 7 resp. 5 basic emotions. Induced data is for instance the SmartKom corpus and the German Aibo emotion corpus where people where recorded in a lab setting fulfilling a certain task that was intended to elicit e. g. anger or irritation in the subjects without them knowing that their emotional state was of interest. The call center communication dealt with by Devillers and colleagues is fully realistic as it is obtained from live recordings.

There are general issues consider while recording the emotion speech corpora are as follows:

1. The number of the emotions and number of the subjects who are contributing to recording this should be decided properly.
2. The database which is recorded as natural or acted helps to decide the applications provided by database and quality of database.
3. Proper contextual information is essential; as expressions mainly depend on linguistic content and its context.
4. Labelling of emotions present in the speech databases is highly subjective.
5. Size of the database matters more in speech emotion recognition for deciding properties such as scalability and reliability of the develop system.

As per the description it can be observe that, for the emotion recognition purpose 24 speech corpora are collected and for intention of synthesis 8 speech corpora's are collected. There is huge disparity among the database, in term of language, in number of emotions, number of subjects, purpose and the method of emotions database collection.

3 Review of classification model

An important task in speech emotion recognition system is selection of classifier. To perform emotion recognition from speech various types of classifier have been used. Hidden Markov Model (HMM), Bayes classifier, Support Vector Machine (SVM), Gaussian Mixtures Model (GMM), k-nearest neighbors (KNN), Artificial Neural Network (ANN) and. Maximum Likelihood Bayesian classifier etc. are the classifiers used in the speech emotion recognition system.
It is observe that Gaussian Mixture Model is more efficient over global features are extracted from the training utterances are suitable for emotion recognition from speech. It's based on expectation-maximization algorithm or Maximum a Posterior (MAP) Parameter Estimation. All the training and testing equations are based on the assumption that all vectors are independent therefore GMM cannot form temporal structure of

the training data. GMM achieved maximum efficiency of 78.77% using the accurate features of speech signal. In speaker dependent system calculated 89.12% for recognition performance using GMM and obtained typical performance of 75% using speaker independent recognition system. [11].

Hidden markov model is widely used classifier for speech application the main reason behind is its physical interconnection with the production mechanism of speech signals. HMM has achieved high accuracy for modelling acoustic and temporal information in the spectrum of speech in speech emotion recognition system. HMM having advantage that the temporal dynamics features of speech can be taken as second accessibility procedure established for optimizing the recognition framework. The process for features selection occurred as main problem in building the HMM based recognition model. Because features carries information is not enough about the emotional states, but it must be significant for the HMM structure as well. In speech emotion recognition system HMM provides higher classification accuracy for emotion recognition as compared to other classifiers. The efficiency for speech emotion recognition by using HMM classifier for the speaker independent system is observed 64.77% and for the speaker dependent it was 76.12%.

In one of the research **B**ayes classifier is adapted with genetic algorithm and sequential floating feature selection that employed ability of accurate classification. It achieved probability of correct classification in first stage related to spectral and prosodic feature with efficiency at an average rate of 67% for surprise and happiness emotional utterances [12].

Another classifier used for emotion recognition is k-nearest neighbor classifier (k-NN). It is typical form of nearest neighbor technique based classifier for random samples. The classifier can classify all the utterances in the design set properly, if "k" equals to 1; however there will be decrease in performance on the test set. K-NN classifier will succeed in achieving classification rate of 64% for four emotional states by utilizing the information of energy contours, pitch and formants etc [13].

Next one is the support vector machine (SVM) classifier is basically Transforming the original set of feature to a higher dimensional feature space by using the kernel function, which required to get optimum classification in this new feature space. SVM classifier are generally used as important applications such as classification problems and pattern recognition Hence SVM gives better classification performance over the other classifiers and due to which it is used for speech emotion recognition system. A typical SVM classifier was implemented for two class problems, but it can be use for more classes. Because of the structural risk minimization oriented training SVM is having high generalization capability. SVM has the efficiency for the speaker dependent classifications are above 80% and speaker independent classifications are 75% respectively. [14].

Other classifier proposed for the classification of emotion is an artificial neural network (ANN), which is having ability to find nonlinear boundaries for separating the emotional states. In speech emotion recognition Multilayer perceptron layer neural networks are commonly used because it has well defined training algorithm as it is relatively easy to implement. Most frequently used feed forward neural network for purpose of speech emotion recognition. The classification rate achieved by ANN based classifiers for speaker dependent recognition with accuracy of 51.19% and for speaker independent recognition with 52.87% accuracy [15].

4 Conclusion

Emotion recognition system is an important research area in today's fields. There are the several applications where speech emotion recognition can be deployed. A properly and well designed database is essential for developing the emotion recognition system. This review paper covers the recent work of speech emotion recognition for filling some important research gaps. This paper contains the review of recent works in speech emotion recognition from the points of views of emotional databases, speech features, and classification models. Some important research issues in the area of speech emotion recognition are also discussed in the paper.

IJSER

Acknowledgment

This work is supported by University Grants Commission as Major Research Project. The authors would like to thank the Department of Computer Science & IT, Dr. Babasaheb Ambedkar Marathwada University Authorities for providing the infrastructure to carry out the research.

References

[1] M. E. Ayadi, M. S. Kamel, F. Karray, "Survey on Speech Emotion Recognition: Features, Classification Schemes, and Databases,, Pattern Recognition 44, PP.572-587, 2011.

[2] Vishal B Waghmare, Ratnadeep R Deshmukh, Pukhraj P Shrishrimal "Development of Isolated Marathi Words Emotional Speech Database," International Journal of Computer Applications (0975 – 8887) Volume 94 – No 4, May 2014.

[3] Chiriacescu I., "Automatic Emotion Analysis Based On Speech," M.Sc. *Thesis, Department of Electrical Engineering, Delft University of Technology, 2009.*

[4] Ashish B. Ingale, D. S. Chaudhari "Speech Emotion Recognition", *International Journal of Soft Computing and Engineering (IJSCE) ISSN: 2231-2307, Volume-2, Issue-1, March 2012.*

[5] Vinay Kumar, Arpit Agarwal, Kanika Mittal. "Tutorial: Introduction to Emotion Recognition for Digital Images," [Technical Report] <Inria-00561918>) 2011.

[6] Nitin Thapliyal, Gargi Amoli "Speech based Emotion Recognition with Gaussian Mixture Model," International Journal of Advanced Research in Computer Engineering & Technology Volume 1, Issue 5, July 2012.

[7] Devillers, L., Vidrascu, L., Lamel, L.: "Challenges in real-life emotion annotation and machine learning based detection," Neural

Networks 18(4), 407–422 (2005)

[8] Litman, D.J., Forbes-Riley, K.:"Predicting student emotions in computer-human tutoring dialogues," In: Proceedings of the 42nd Annual Meeting of the Association for Computational Linguistics (ACL), Barcelona, Spain (2004)

[9] Burkhardt, F., Paeschke, A., Rolfes, M., Sendlmeier, W.F., Weiss, B.: "A database of German emotional speech," In: Proceedings of Interspeech 2005, Lisbon, Portugal (2005)

[10] Engberg, I.S., Hansen, A.V.: "Documentation of the Danish Emotional Speech Database (DES)," Technical report. Aalborg University, Aalborg, Denmark (1996)

[11] A. Nogueiras, A. Moreno, A. Bonafonte, Jose B. Marino, "Speech Emotion Recognition Using Hidden Markov Model," Eurospeech, 2001.

[12] Mohammad H. sedaaghi , Constantine Kotropoulos and Dimitrios Ververidis " Using Adaptive Genetic Algorithms To Improve Speech Emotion Recognition," ieee conference 2007.

[13] C. M. Lee, S. S. Narayanan, "Towards detecting emotions in spoken dialogs," IEEE transactions on speech and audio processing, Vol. 13, No. 2, March 2005.

[14] S. Emerich, E. Lupu, A. Apatean, "Emotions Recognitions by Speech and Facial Expressions Analysis," 17th European Signal Processing Conference, 2009.

[15] D. Ververidis and C. Kotropoulos, "Emotional Speech Rcognition: Resources, Features and Methods," Elsevier Speech communication, vol. 48, no. 9, pp. 1162-1181, September, 2006.

IJSER