Review of Performance Factors of Emotional Speaker Recognition System: Features, Feature Extraction Approaches and Databases

[1]Satish Kumar Das, [2]Uttpal Bhattacharjee, [3]Amit Kumar Mandal

•

Department of Computer Science & Engineering
Rajiv Gandhi University, Arunachal Pradesh, India
[1]satish.das@rgu.ac.in
[2]utpal.bhattacharjee@rgu.ac.in
[3]amitkumar.tu@gmail.com

Abstract

Emotion is a conscious mental reaction accompanied by physiological and behavior changes in human body.In speaker authentication system, emotional state of the speaker plays a vital role. Recently, the field of speaker recognition in emotional context attracts more and more attention of many research focuses. However, to implement more realistic and intelligent emotional speaker recognition system it is interesting to study this system under real life conditions. Speech emotion recognition is a system in which speech signals are processed to classify the embedded emotions. In recent past, speaker emotion recognition has gained a lot of attention from different researchers as it has many applications. In this regards, study of prior works is useful for further research in the field of speaker verification in emotional context. So, performance and reliability of Emotional Speaker Recognition System depend on the proper selection of features to characterize different emotional states, feature extraction approaches and databases. In this paper we briefly discuss about different features, feature extraction approaches and emotion recognition and speaker verification databases.

Keywords: speech emotion recognition, features, feature extraction approaches, databases

I. Introduction:

The human voice is used to expresses the emotions. The speech has potential to be an essential mode of interaction with the computer. Recognition of emotion from human speech is an active research area in signal processing. This is due to the enormous possibilities in the field of human-machine interaction [1].Speech emotion recognition is using in different fields from medical scienceto business, from entertainment to interactive systems.

A speech signal is a logical arrangement of sounds from which our brain acquires the information and knowledge. But how a machine interprets the human speech signals and gains information and knowledge from is in the heart of speech recognition system. The basic goal of speech recognition is to offer an interaction between a person and a machine.

In this paper we are trying to present comprehensive literature review of various features, feature extraction techniques and datasets for emotion recognition and speaker verification. In our previous paper [2], we tried to present comprehensive literature review of various classifiers for speech emotion recognition. The paper is not to compare and evaluate the works by different

Speech Emotion Recognition Classifiers: A Literature Review

Amit Kumar Mandal, Satish Kumar Das, Utpal Bhattacharjee

Abstract

- Automatic speech emotion recognition plays a significant role in human–computer interaction. The basic goals of the automatic emotion recognition are to understand and retrieve emotions from individual's speech utterance. The main key issues of speech emotion recognition systems are to prepare an appropriate dataset, selection of suitable feature sets, and design of a proper classification method. The selected feature vectors are stored in the database and later fed to classifier to detect the emotions by comparing the vectors from the trained data and test data vectors. In this paper we review the various classification techniques based on support vector machines, Gaussian Mixture Models, Hidden Markov Models, Artificial Neural Network, k-Nearest Neighbors and Naïve Bayes. We also try to present some literary survey work to show the use of classifiers by different authors.

Keywords *- Speech emotion classifier; SVM; GMM; HMM; ANN; KNN; Naïve Bayes*