International Journal of Education and Science Research
Review E-ISSN 2348-6457
Volume-3, Issue-2 April- 2016
www.ijesrr.org Email- editor@ijesrr.org

Interactive Voice Communication System

Shailendra Kumar, Hitesh Kumar, Aman Sharma
Department of Information Technology
SRM University, NCR Campus
Modinagar

ABSTRACT:
Interactive Voice Communication is a current research topic in the field of Human Computer Interaction with widely ranged applications. The speech features such as, Mel Freq. cepstrum coefficients is extracted from speech utterance. The Support Vector Machine (SVM) is used to classify various emotional states such as anger, happiness, sadness, neutral behavior, from a collection of data of speech collected from various emotional drama sound tracks. This feature is used for classification of emotions. It gives 93.69% classification accuracy for Gender independent case 94.73% for male and 100% for female speech. Automatic Emotion Recognition (AER) can be done in two ways, either by speech or by our facial expressions.
Automatic classification of a speaker's affective state is one of the major challenges in community of signal processing, since Human-Computer interaction can be improved and give insights into the nature of emotions from psychologicalway of perception. The control of amplitude & frequency of sound production influences strongly the affective voice content we propose to take advantage of the inherent speech modulations and the use of instant amplitude and feature that are derived from frequency for efficient emotion recognition. Such
Features can further increase the performance of the widely used spectral linguistic information.

I. INTRODUCTION:

Automatic Speech Emotion Recognition is a very recent research topic in the Human Computer Interaction (HCI) field. Computers have become an integral part of our lives, and the need has rise for more common communication interface between humans and computers. To do this, a computer would have to be able to perceive its current situation and response a bit different depending on that perception. The process involves an understanding to a user's emotional condition also To make the human-computer interaction more natural, it would be beneficial to give computers that it is able to recognize emotional situation the same manner as human does. Automatic Emotion
Recognition (AER) is done in two ways, either by voice or by face expressions. In the field of HCI, Voice is primary to the objectives of an emotion recognition system, as are facial expressions and gestures. We have focused on both features that are classification and extraction. Low-level outline of figure of linguistic, quality of voice and articulation information have been used to extract high-level functional that describe the emotional content of a sentence. Diverse time scales of frame and turn level have been also combined to recognize emotion from voice. Subsequent layers of binary classifiers were further fused in a hierarchical framework for emotion classification and emotions were introduced in order to identify emotional properties of ambiguous utterances. In the context of continuous tracking of affective states, a long short-term memory neural network also been proposed. Although speech emotion recognition has main focus on the extraction of linguistic and spectral features, we have attempted to use the modulation properties of speech signals to recognize human affective states.
Speech is taken as a high-level mode to communicate with intention and emotions. The Support Vector Machine is used for classification for emotion recognition. The Support Vector Machine is used for purpose of regression and classification. It performs classification by constructing hyper planes of N-dimensions that

optimally separate the data into several categories. The classification is being achieved by a linear or non-linear separation surface in the input feature space of the datasets. Its main idea is to transform the very firstinput set to a high dimension features space by using a kernel functionality, and then achievinggreat classification in this new feature space.

In this paper, in section 2 Requirement of voice Emotion Recognition is given, in section 3 SER Feature Extraction Technique is given, and in section IV we have discussed about Emotion Database.

II. REQUIREMENT OF SER:

Speech Emotion Recognition In Human-Robotic Interfaces: robots can be used as teaching by any human to interact with humans and recognizing the behavior of human. So that robotically made pet, should have ability to understand not just only command spoken by anyone, but also other details, like it as the health and emotion conditions of its human commander and can act accordingly. In smart call-centers, SER helps in detecting potential problems that begins from not being satisfactory course of interaction. A frustrated customer is offered the help of human operators typically or some reconciliation strategies. In intelligent speaking tutor system, detecting and adaptation of student's emotion is considering to be an very important strategies for ending the performance gap of humans and computer personnel as students emotions can show their impact performance and learning.

2. KEYWORDS:

Speech Emotion, Emotion Recognition, SVM, MFCC, Emotion Verification, Emotion Classification, AM-FM Features, Speech Analysis, Human-Computer Interaction.

3. SER FEATURE EXTRACTION:

There are several features extracted for classifying speech affect such as energy level, pitch level, format frequency etc. All are linguistic features. In commonlinguistic features are primary indicator of speaker's emotional state. In feature extraction process two features are extracted Mel Frequency Cepstrum Coefficient (MFCC).

Pre-processing - In the pre-processing level first every sign all is de-noised by soft-limen the wavelet coefficients, and since the silent part of the signals don't carry any useful details, those parts including the leading and trailing edges are eliminated by limen the energy of the signal. The signal are divided into different frames using a Hamming window of length 23 ms MFCC Log Mel Spectrum Mel Spectrum Magnitude Spectrum FFT Mel Filter bank & Frequency

Wrapping Log DCT Windowing Framing Preprocessing Emotional Speech Framing: It is a process of segmenting the speech samples obtained by the analogue to digital conversion method (ADC), into the small frames obe segmented into quasi-stationary frames, and enables Fourier Transforming of the speech signal. It is just because, voice signal is known to exhibit quasi-stationary behavior within the short period of 20-39ms.

Windowing- Windowing level is meant to window each individual frame, in order to make mineralizing the signal discontinuations at the starting and the finish of each frame. FFT: Fast Fourier Transform (FFT) algorithms are ideally used for evaluation of the frequency voice spectrum. FFT converts each frame of N sample starting from the domain of time into the domain of frequency. Mel Filter bank and Frequency wrapping- The Mel filter bank comprises of overlapping triangular filters with frequencies determined by the centre frequencies of the two adjacent filters. The filters have centre frequencies that are linearly facedand fix bandwidth on the Mel scale. Logarithm Taken- The logarithm has the affect of change multiplications into additions. Thus, this step simply converts the multiplication of the magnitude in Fourier transformation into additions.

Take Discrete Cosine Transform: It is used to make the filter energy vectors. Because of this step, the information of the filter energy vector is compacted into the first number of components & shortening the vector to numbers of component.

FEATURE LABELLING:

In Feature labeling each extracted details are stored in a collected data along with its class labels. Though the SVM is binary classifier it could be used for classifying multiple classes. Each feature is associated with its class label e.g. angry, happy, sad, neutral, fear. Output Emotion Class SVM Classification SVM Training Feature Labelling Feature extraction Input Emotional Voice Feature Selection : The performing of a pattern recognition way highly dependendent on the discriminate ability of the features. Selection of the most relevant sub-set from an original descriptive set, we can increase the performance of the classifier and on another hand decreasing the computational complexities. We are using the forward selection method for every single binary classifier in our way in order to selecting the more effective subset of features. At each step the variable which increase the performances of the classifier the more, is being added to the feature subset. The recognition of human emotion is always an essential a pattern recognizing problem. We have been using LS-SVM (description in next section) as a classifier in the research. Since we have been dealt with multi-class classificationl problems, we extend our two-class support vector classification methodol to multi-class problems. There are different way for multi-category SVM : a) one-against-all b) one-against-one (pairing wise) are the more popular ones.

4. THE EMOTION DATABASE:

An important issue to be considered in the evaluating of emotional voice recognizer is the degree of common-ness of the collected data used to check its performance. Wrong conclusion may be established if a low-quality collection of data is used. Moreover the designing process of the data is important in a critical way to the classification tasks being taken. For example, the condition being classified may be infant-directed - soothing and prohibition, or adult-directed, joy and anger. In other databases, the classification task is to detect stress in voice. The classification task is also classified by the number and type of expression in included in the database. Databases can be divided into three type : expressions, is obtained by asking an actor to speak with a predefined emotion. life systems (for example call-centers) reports are being used for labelling control process.

Various types of databases are suitable for different functions. In our work we have decided to create a database of type one, challenging and appealing applications in this framework which consists in sensing human body motion to gather context information about people actions, correlated to emotional voice.

The interaction among human being and computer will be more smooth if computers are able to take and answer to human that are not a verbal communication such as emotions. Thusmany approaches have been proposed to check human emotions based on facial expressions or voice, relative a define amount of work has been done to combine these two and other, modality to accelerate accuracy and robust nature of the expression recognition system. This paper analyzes the strengths and the limitations of systems based only on face expressions or detailed knowledge of acoustics. It also tells two approaches used to fuse these two modality that are decision level and feature level integration systems using a database recorded from a user, four emotions were classified as sad behavior, angriness, happy behavior, and neutral side By the use of markers on the user's face, detailed facial motions were taken and stored with motion capture in conjunctions with number of simultaneous voice recordings. The resultof this reveals that the system based on face expression gave muchbetter performances than the way based on just acoustical information for the emotions considered. It also shows the complementary of the two modality & that where the two modalities are combined, the performance and the robustness of the emotion recognition system generally improve measurably.

Information from multiple resources can be consolidated in three distinct stages

(i) Feature extraction stage
(ii) Match score stage
(iii) Decision stage.

While fusion at the match score and decision levels are extensively used, fusion at the feature level is a relatively a problem.

5. REFRENCES:

1. Voice Communication System at Wikipedia.com
2. David B. Roe, Jay G. Wilpon, for the National Academy of Sciences- 2002-Technology & Engineering
3. Dr.K.V.K.Prasad , Embedded/Real-Time Systems: concepts, Design and Programming-The Ultimate Refrence, Dream Tech Press, 2004
4. Martyn Mallick, Mobile and Wiresless Design Essentials,WileyDreamtech India pvt ltd., 2003
5. Jochen Schiller, Mobile Communication, Addision Wesley, 2000
6. Thedore.S.Rappaport, Wireless Communication-Principles and practice,Prentice Hall, second edition, 2010