Volume 5, Issue 6, June 2015 ISSN: 2277 128X

International Journal of Advanced Research in Computer Science and Software Engineering

Research Paper

Available online at: www.ijarcsse.com

Identifying Emotions of Humans from Voice

Shambhavi. S. Sheerur*, Dr. V. N Nitnaware
Department of E & TC
D.Y. Patil College of Engineering
Maharashtra, India

Abstract— Recently, the importance of reacting to the emotional state of a user has been generally accepted in the field of human-computer interaction and especially speech has received increased focus as a modality from which to automatically deduct information on emotion. So far, mainly academic and not very application-oriented offline studies based on previously recorded and annotated databases with emotional speech were conducted. However, demands of online analysis differ from that of offline analysis, in particular, conditions are more challenging and less predictable.

Real-time automatic emotion recognition from acoustic features of speech was investigated. First, offline experiments were conducted to find suitable audio segmentation, feature extraction and classification algorithms. Suitable means in this context that they should be fast and at the same time give as correct results as possible.

Keywords— FFT, PSD, MFCC feature, SVMclassifier, kernel

I. INTRODUCTION

Considering affective processes in computing is necessary both for a better and more adequate user experience as well as for problem solving in computer-internal processes. With respect to the user experience this means interaction should rely less on the traditional input devices. Keyboard and mouse in favour of more natural forms of interaction such as language or gestures,compatible with a ubiquitous and unobtrusive computing environment. Knowledge about the emotional state can help to connect angry callers of an automatic dialogue system to a human operator, to motivate a student at the right time, or to develop a fun game that is influenced by emotional expressions. However, emotion recognition from speech is a very challenging tasks.

II. RELATED WORKS ON SPEECH EMOTION

Sr.no	Name of publication	Title of paper	Work done
1	Proceedings of the ISCA(International Speech Communication and Association)ITRW on speech and emotion,2000	Databases of emotional speech	Real world emotions or acted ones
2	Speech Audio Process. 13 (2) (2005)	Toward detecting emotions in spoken dialogs, IEEE	Simulation of utterences
3	Institute of Electronics, Technical University of Lodz, Wolczanska 211/215, 90-924 Lodz, Poland	Emotion recognition in speech signal using emotion extracting Binary decision trees	Application of a binary-tree based classifier, where consecutive emotions are extracted. The method has been verified using two databases of emotional speech on German and Polish
4	IJCSI International Journal of Computer Science Issues, Vol. 8, Issue 5, No 1, September 2011	Emotion Recognition using Dynamic Time Warping Technique for Isolated Words(S2011)	DTW is utilized to recognize speaker independent Emotion recognition.DTW is used to store a prototypical version of each word in the vocabulary and compute incoming emotion with each word.
5	International Journal of Advanced Research in Computer Science and Software Engineering Volume 3, Issue 8, August 2013	Speech Emotion Recognition Using Combined Features of HMM & SVM Algorithm (2013)	To classify four emotions viz. happy, angry, sad and aggressive. Combining advantage on capability to dynamic time warping of HMM and pattern recognition of SVM. HMMs, which export likelihood probabilities and optimal state sequences, have been

			used to model speech feature sequences
6	Signal and image processing:An international journal (SIPIJ)vol.4 no.4 august 2013	Feature extraction usingMFCC (2013)	New purpose of working with MFCC by using it for Hand gesture recognition. The objective of using MFCC for hand gesture recognition is to explore the utility of the MFCC for image processing.
7	IEEE Transaction on language speech and audio processing.	Multiview Supervised Dictionary Learning in Speech Emotion Recognition (2014ieee)	The effectiveness of the proposed multiview learning techniques in using the complementary information of single views is demonstrated in the application of speech emotion recognition (SER).
8	International Journal of Signal Processing, Image Processing and Pattern Recognition Vol. 6, No. 3, June, 2013	LPC and MFCC Performance Evaluation with Artificial Neural Network for Spoken Language Identification(2013)	LPC and MFCC with ANN are among the most usable techniques for spoken language identification. LPC and MFCC methods are used for extracting features of a speech signal and ANN is used as the recognition and identification method.
9	Springer-Verlag Berlin Heidelberg	A study of emotion recognition and its applications	Focoused on distinguishing anger versur neutral speech for call center applications

III. SYSTEM DESIGN

A speech emotion recognition system in the tradition of pattern recognition systems consists of three principal parts :signal processing, feature calculation and classification.

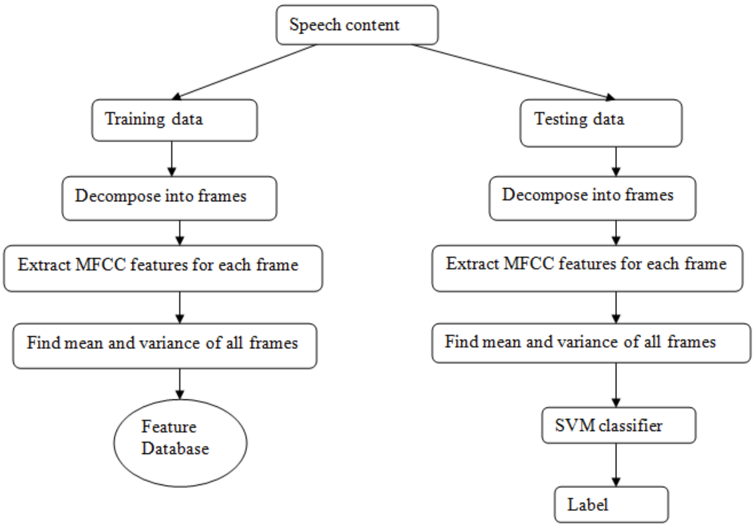

Speech content
Training data
Testing data
Decompose into frames
Decompose into frames
Extract MFCC features for each frame
Extract MFCC features for each frame
Find mean and variance of all frames
Find mean and variance of all frames
Feature Database
SVM classifier
Label

Fig1.Overall implementation of Emotion Identification using speech

A) Emotional speech input

The emotional speech input to the system may contains the collection of the acted speech data the real world speech data. After collection of the database containing short Utterances of emotional speech sample which was considered as the training samples, proper and necessary features such as prosodic and spectral features were extracted from the speech signal. These feature values were provided to the SVM for training of the classifiers. Then recorded emotional speech samples presented to the classifier as a test input. Then classifier classifies the test sample into one of the emotion from the above mentioned five emotions and gives output as recognized emotion.

The complexity of the affect recognition process increases with the amount of classes (affects) and speech descriptors used within the classifier. It is therefore crucial to select only the most relevant features in order to assure the ability of the model to successfully identify emotions, as well as increasing the performance, which is particularly significant to real-time detection.Signal processing involves digitisation of the recorded signal, potentially acoustic preprocessing such as filtering, and the segmentation of the input signal into meaningful units. While the first three steps are standard procedures, the last step is the most crucial with respect to speech emotion recognition.

The aim of the feature calculation is to find those properties of the digitised and preprocessed acoustic signal that are characteristic of emotions and to represent them in an n-dimensional feature vector. So far, there is not yet a general agreement on which features are the most important ones and good features seem to be highly data dependent.

B) Features extraction and selection

An important step in emotion recognition System through speech is to select a significant feature which carries large emotional information about the speech signal. Several researches have shown that effective parameters to distinguish a particular emotional states with potentially high efficiency are spectral features such as Mel frequency cepstrum coefficients (MFCC) and prosodic features such as pitch ,speech energy, speech rate ,fundamental frequency. Speech Feature extraction is based on smaller partitioning of speech signal into small intervals of 20 ms or 30 ms respectively known as frames. Speech features basically extracted from vocal tract , excitation source or prosodic points of view to perform different speech tasks. In this work some prosodic and spectral feature has been extracted for emotion recognition. Speech energy is having more information about e motion in speech. The energy of the speech signal provides a representation that reflects these amplitude variations here short time energy features estimated energy of emotional state by using variation in the energy of speech signal. The analysis of energy is focused on short- term average amplitude and short-term energy.

Another important feature carries information about emotion in speech is pitch. The pitch signal is also called the glottal wave-form. The pitch signal produced due to the vibration of the vocal folds , tension of the vocal folds and the sub glottal air pressure. Vibration rate of vocal cords is also called as fundamental frequency . Another features considering is a simple measure of the frequency content of a signal which is the rate at which zero crossings occur. Zero-crossing rate is a measure of number of times in a given time interval/frame such that the amplitude of the speech signals passes through a value of zero .it is one of the important spectral feature.

Features for emotion recognition are calculated from speech signals as produced by a speaker. The speech signal can be assumed to be a waveform that is momentarily periodic. As every waveform it has certain properties such as amplitude, time, and (usually superposed) frequencies that serve to characterise it and also help to distinguish between different emotions. For computational purposes, the waveform is digitised. Consequently, the rate at which samples are taken is an important characteristic of the signal. According to the Nyquist theorem, a continuous singal can be sampled without loss of information only if it does not contain frequency components above one half of the sampling rate. Although the human ear perceives frequencies up to about 20 kHz (though this ability degrades with age), only frequencies below 8 kHz are used for the perception of speech. Thus, a sampling rate of 16 kHz is sufficient for most speech processing purposes. Telephone lines usually transmit only 8 kHz sampled sound signals, which makes some phonemes sound equal though humans are able to differentiate them by means of the context.

C) Pitch

The term pitch refers to the ear’s perception of tone height . For most purposes, this is just the fundamental frequency f0, though the two terms are not identical since f0 can be measured as a property of an acoustic wave, while pitch is grounded by human perception. Pitch is a very obvious property of speech, also for non-experts, and it is often erroneously considered to be most important for emotion perception. Pitch does definitely have some importance for emotions, but it is probably not as huge as typically assumed. Generally, a rise in pitch is an indicator for higher arousal, but also the course of the pitch contour reveals information on affect. Pitch can be calculated from the time or the frequency domain. In the time domain pitch can for example be calculated from the zero-crossings rate. This method is however more suited for musical pitch detection. For speech, pitch is usually determined by looking at the maxima of the auto-correlated frequency spectrum.

D) Energy

Loudness is the strength of a sound as perceived by the human ear. It is hard to measure directly, therefore the signal energy is often used as a related feature. Energy can be calculated from the spectrum after a Fourier transformation of the original signal. However, it differs from loudness in that all existing noises add to the signal energy, while the ear perceives the loudness of speech as just that. The energy curve depends on many factors, such as phonemes, speaking style, utterance type (e. g. declarative, interrogative, exclamatory), but also on the affective state of the speaker. Again, like pitch, high energy roughly correlates with high arousal, but also variations of the energy curve give hints on the speaker’s emotion.

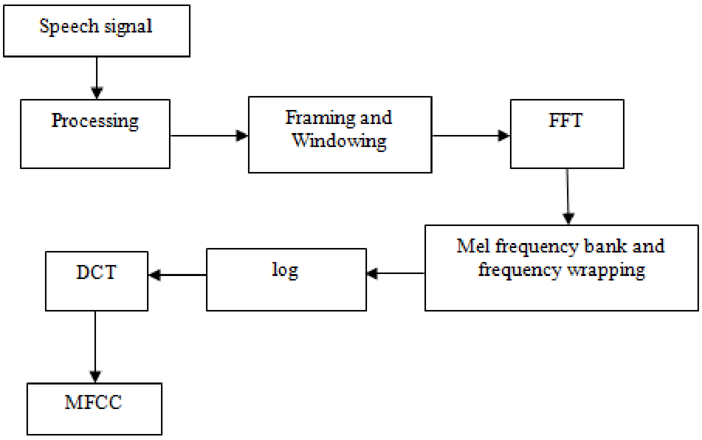


Fig2.MFCC feature extraction

E) Combination of MFCC and SVM

MFCCs plus energy of an utterance are used as the input for Support Vector Machine. Support Vector Machine (SVM) has been profoundly successful in the area of pattern recognition. In the recent years there has been use of SVM for speech recognition. Many kinds of kernel functions are available for SVM to map an input space problem to high dimensional spaces. We lack guidelines on choosing a better kernel with optimized parameters of SVM. Some kernels are better for some questions, but worse for other questions. Which is better is unknown for speech emotion recognition, thus the thesis studies the SVM classifier and proposes methods used to select a better kernel with optimized parameters. The new method we proposed in this paper can more efficiently gain optimized parameters than common methods. In order to improve recognition accuracy rate of the speech emotion recognition system, a speech emotion recognition based on optimized support vector machine is propose.

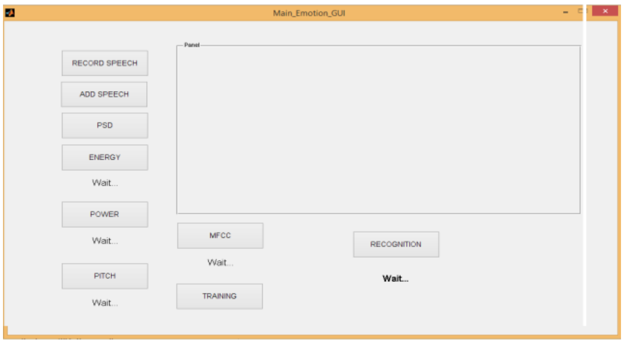
Speech emotional features extraction is a process extracting a small number of parameters from the speech signal that can be later used to represent each utterance. Speech extraction techniques include temporal analysis and spectral analysis techniques. The waveform of speech signal is used for analysis in temporal. The spectral form of speech signal is used for analysis in spectral analysis. Mel Frequency Cepstral Coefficients is a spectral analysis technique. In the recent years, MFCC feature has been widely used for not only the speaker but also speech recognition. Mel Frequency Cepstral Coefficients are set of features reported to be robust in various kinds of pattern classification tasks in speech signal It has been proven that human being perception of the frequency contents of sounds for speech signals does not follow a linear scale in the psychology studies. Therefore for each tone with an actual frequency f measured in Hz, a subjective pitch is measured on a scale called the Mel scale The Mel frequency scale is in the form of a linear frequency scale below 1000 Hz while a logarithmic scale above 1000 Hz.

Hidden Markov model (HMM) and Gaussian mixture model (GMM) using MFCC have achieved valuable results on speech emotion recognition. However, there is a problem when using GMM to recognize speech emotion states. Effective training of GMM requires a great deal of data, while collecting emotional speech utterances will cost a lot and therefore the available training data is usually scant. SVM has a better classification performance on a small amount of training samples. But we are lacking in guidelines on choosing a better kernel with optimized parameters of SVM. Some kernels are better for some questions, but worse for other questions. There is no uniform pattern used to the choice of SVM with its parameters and kernel function with its parameters. The paper proposed methods about selecting optimized parameters and kernel function of SVM.

The most important aspect of emotion recognition system through speech is classification of an emotion. The performance of the system influenced by the accuracy of classification, on the basis of different features extracted from the utterances of emotion speech samples emotions can be classified by providing significant features to the classifier.Training a dtata set to obtain a model and using the model to predict information of the testing data set.parameter selection is important for obtaining good SVM models.It is simple and efficient computation of machine learning algorithms,and is widely used for pattern rrecognition and classification problems .Under the conditions of limited training data, it can have a very good classification performance compared to other classifiers.

IV. RESULTS

Finally the main emotions such as happy or joy,anger and neutral are classified for a particular incoming speech signal. This is done using GUI matlab.



Wait...

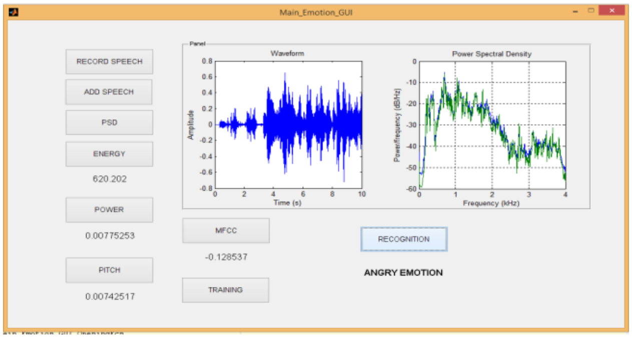
Wait...

Wait...

Wait...

Wait...

Fig3.Matlab GUI for Emotion Identification



620.202

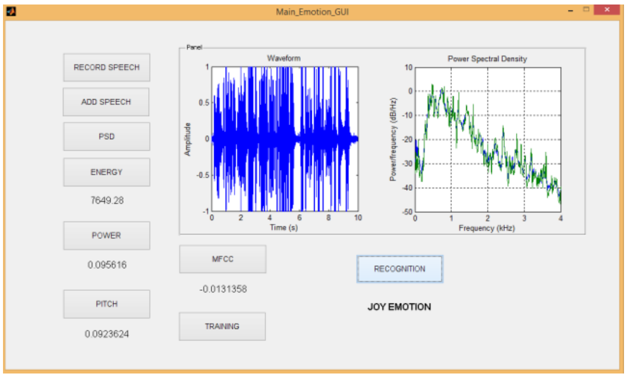
0.00775253

0.00742517

-0.128537

ANGRY EMOTION

Fig4.Emotion identified Angery emotion



ENERGY 7649.28

POWER 0.095616

PITCH 0.0923624

MFCC -0.0131358

JOY EMOTION

Fig5.Emotion Identified is Happy/Joy emotion

IV. CONCLUSION

Analysis of speech signal for emotion (neutral, angry, happiness, sad) is done using hierarchical algorithm that uses different feature extraction techniques.Neutral emotion is differentiated under the consideration of mean values of MFCC. Mean values of MFCC in other emotion are greater than mean values of MFCC obtained in case of neutral emotion.

As technology evolves, interest in human like machines increases. Technological devices are spreading and user satisfaction increases importance. A natural interface which responds according to user needs has become possible with affective computing. The key issue of affective computing is emotions. Any research which is related with detection, recognition or generating an emotion is affective computing. User satisfaction or un-satisfaction could be detected with any emotion recognition system. Besides detection of user satisfaction, such systems could be used to detect anger or frustration. In such cases, user could be restrained like driving a car. In emotion detection tasks, speech or face emotion detections are the most popular ones. Easy access to face or speech data made them very popular. Speech carries a rich set of data. In human to human communication, via speech information is conveyed. Acoustic part of speech carries important info about emotions.MFCC are used for the feature extraction .Algorithm with the SVM's Overall performance is tested.

ACKNOWLEDGMENT

I take this opportunity to express my deep heartfelt gratitude to all those people who have helped me in the successful completion of the paper.First and foremost, I would like to express my sincere gratitude towards my guide Dr.V.N Nitnaware for providing excellent guidance, encouragement.Without his valuable guidance, this work would never have been a successful one.I would like to express my sincere gratitude to our Head of the Department of Electronics & Communication Engineering, Prof.Santhosh .G. Bari and PG coordinator Prof.Vaishali.V.Thorat for their guidance and inspiration.I would like to thank our Principal Dr.V.N. Nitnaware for providing all the facilities and a proper environment to work in the college campus.

REFERENCES

[1] Mehrdad J. Gangeh, AliGhodsi,Mohamed S. Kamel,"Multiview Supervised Dictionary Learning in Speech Emotion Recognition," IEEE Transaction on audio, speech, and language processing.

[2] Shikha Gupta1, Jafreezal Jaafar2, Wan Fatimah wan Ahmad3 and Arpit Bansal4 J. Clerk Maxwell, "Feature extraction using mfcc" Signal & Image Processing : An International Journal (SIPIJ) Vol.4, No.4, August 2013

[3] N.Murali Krishna1,P.V. Lakshmi2,Y. Srinivas3 J.Sirisha Devi4," Emotion Recognition using Dynamic Time Warping Technique for Isolated Words," IJCSI International Journal of Computer Science Issues, Vol. 8, Issue 5, No 1, September 2011

[4] Aastha Joshi," Speech Emotion Recognition Using Combined Features of HMM & SVM Algorithm International Journal of Advanced Research in Computer Science and Software Engineering Research Paper Volume 3, Issue 8, August 2013

[5] A. Khan, et al., "Speech Recognition: Increasing Efficiency of Support Vector Machines," International Journal of Computer Applications vol. 35, dec 2011.

[6] L. Muda, et al., "Voice recognition algorithm Using Mel Frequency Cepstral Coefficient (MFCC) and Dynamic Time Warping (DTW) Techniques," Journal of computing vol. 2, 2010.