iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 6 **Issue:** V **Month of publication:** May 2018

DOI: http://doi.org/10.22214/ijraset.2018.5170

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 6.887
Volume 6 Issue V, May 2018- Available at www.ijraset.com

An Efficient Approach for Emotion Detection from Speech Using Neural Networks

Komal Rajvanshi[1], Ajay Khunteta[2]
[1]*PG Student,* [2]*Associate Professor Department of Computer Science, Poornima College of Engineering, Jaipur, Rajasthan*

Abstract*: *In this paper we have proposed an efficient algorithm for detecting different human emotions like happy, sad, angry, shocked etc. from speech signal. We have analyzed this algorithm in MATLAB simulation tool to extract relevant speech features including MFCC, signal energy and statistical parameters. This analysis is performed to improve the emotion detecting capability of algorithm from speech signal.
Keywords*: *Emotion, Auditory, Texture, Speech, Affective computing

I. INTRODUCTION

People contact with their condition utilizing numerous di□erent detecting components. Human body gathers all of this information assesses and generates an emotional state in response. Emotions play a pivotal role in the entire human experience and they have a colossal e□ect on the human basic leadership process. They are critical piece of human social relations and take part in imperative life choices. Emotions are principal for humans, affecting recognition and everyday activities for example communication, learning and basic leadership. They are communicated through speech, facial expressions, motions and other non-verbal information.

With the advancement in technology, human machine communication has gone beyond straightforward legitimate calculations. As the spread of technological devices builds, the requirement for high level human-computer communication requires more natural interfaces. Assigning human like properties to computers like watching, interpreting and producing a□ective features is called a□ective computing [1]. A□ective computing empowers computers to recognize human emotional state and react to their users as indicated by their necessities. With a specific end goal to do this, the computer first perceives the emotion, and then produces a reaction in view of its preset qualities. This requires an interdisciplinary comprehension, including cognitive science, psychology, human science and software engineering. A□ective computing expects to enhance human computer interaction and adjust machine reactions as indicated by human requirements. In recent years, a□ective computing has been utilized to assess users' pleasant/unpleasant state during communication. Recognizing an unpleasant state during the particular task and interfering the procedure is possible with real time a□ective systems. In human computer communication, the principle task is to keep users' level of fulfillment as high as would be prudent. A computer with a□ective properties could recognize the users' emotion and could build up a counter reaction to expand user satisfaction.

II. TYPES OF EMOTION

Acoustic part of speech conveys important information about emotions. Each emotion has special properties that influence us to remember them. Main work of a speech emotion detection algorithm is to recognize the emotional state of a speaker from speech. General automatic speech emotion recognition algorithm made out of two sections which are feature extraction and classification stages. Prosodic and spectral features of speech are the most prominent features that are utilized in speech emotion detection algorithm. Inflection, stop, stress, pitch and beat are prosodic feature examples. Spectral features examine frequency segments of speech signal. Utilized classification algorithm changes from algorithm to algorithm. Gaussian mixture models, hidden Markov model, Support vector machines and neural networks are the most famous classification algorithms utilized in speech emotion recognition process. Fig. 1 represents the location of six common basic emotions in the Schlosberg space.

Automatic speech emotion research generally centers around the determination of right feature set and to recognize in which emotion which features are more a□ective. Our brain examines the input from our auditory system. In fig. 1 each emotion can be represented as a linear combination of valence evaluation), arousal (or activation) and potency (or power). Valence describes how positive or negative an emotion is; arousal calculates the degree of excitement or involvement of the individual in the emotional state; potency represents the strength of the emotion.

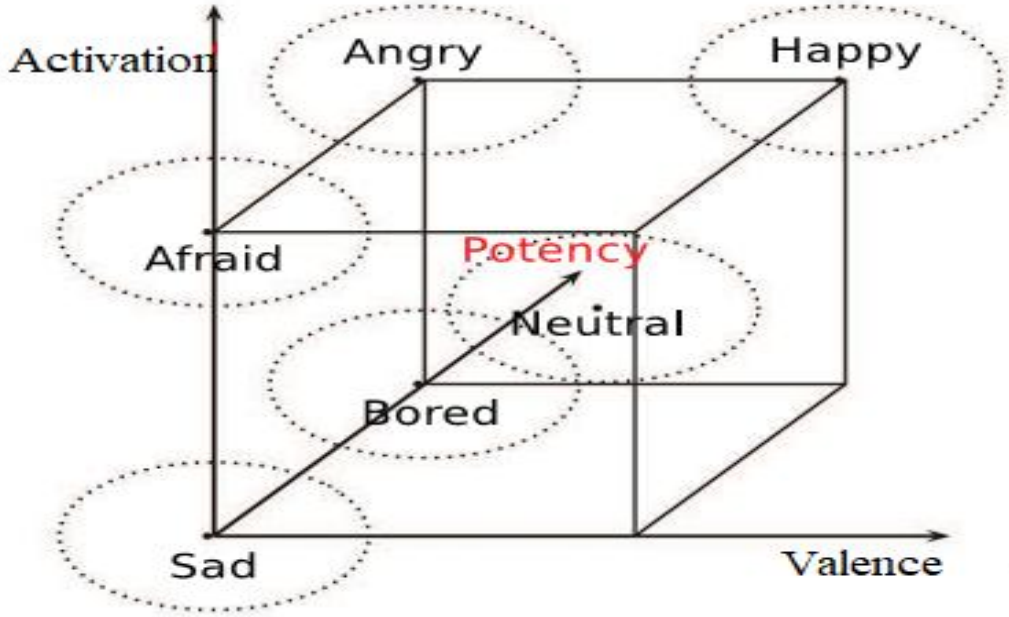


Fig. 1 Three dimensional emotion space

Since emotions are the conclusion of highly subjective experiences, it is difficult to find uniform rules or universal models to demonstrate them. Roughly speaking, there exist two theories in the psychological literature, depending on whether we take emotions discrete or continuous. The discrete approach involves in designating basic emotional classes. Ekman for instance described seven basic emotions such as happiness, sadness, anger, anxiety, disgust, neutral and boredom.

III. DATABASE

Databases play a vital role for automatic emotion recognition as the rest statistical methods are learned using examples. The databases used till know in the research acted, Elicited and real life or natural emotions. As the naturalness of the database increases the complexity also increases. So at the beginning of the research on automatic vocal emotion recognition, which started actually in the mid-90s, work began with acted speech and shifts now towards more realistic data. The most popular examples of acted database are Berlin Database of emotional speech which comprised of 5 male and 5 female actresses and the Danish Emotional speech corpus (DES).

Russian database consists of ten pronounces sentences from61 speakers (12 male 49female) of age group 16-28years expressing six emotions viz., Happy, sad, angry, fear, neutral and disgust. The example of Induced database is SmartKom corpus and the German Aibo emotion corpus without knowing the people that their emotions are being recorded. The call center communication by Devillers and et al. is obtained from live recordings and is an example of real emotional database. Other examples includes Surrey Audio Visual Expressed Emotion (SAVEE)which comprised of 4 male actors expressing 7 different emotions. The Speech Under Simulate and Actual Stress (SUSA) database of 32 speakers where speech was recorded in both simulated stress and actual stress.

IV. FEATURES SELECTION

Many different speech feature extraction methods have been proposed over the years. Methods are distinguished by the ability to use information about human auditory processing and perception, by the robustness to distortions, and by the length of the observation window. Due to the physiology of the human vocal tract, human speech is highly redundant and has several speaker-dependent features, such as pitch, speaking rate and accent. An important issue in the design of a speech emotion recognition system is the extraction of suit-able features that efficiently characterize different emotions. Although there are many interesting works about automatic speech emotion detection there is not a silver bullet feature for this aim. Since speech signal is not stationary, it is very common to divide the signal in short segments called frames, within which speech signal can be considered as stationary. Human voice can be considered as a stationary process for intervals of 20-40 (ms). If a feature is computed at each frame is called local, otherwise, if it is calculated on the entire speech is named global. There is not agreement in the scientific community on which between local and global features are more suitable for speech emotion recognition.

Coherently with the wide literature in the field, in this paper a set of 182 features has been analysed for each the recorded speech signal, including: Mean, variance, median, minimum, maximum and range of the amplitude of the speech Mean, variance, minimum, maximum and range of the formants; Energy of the Bark sub-bands Mean, variance, minimum, maximum and range of the Mel-Frequency Cepstrum Coefficients Spectrum shape features Centre of Gravity, Standard Deviation, Skewness and Kurtosis; Mean and standard deviation of the glottal pulse period, jitter local absolute, relative average perturbation, difference of difference period and (–ve) point period perturbation quotient. The spectral features are also known as vocal tract, segmental or system features. Spectral features include formants MFCCs, LPCCs, and perceptual linear prediction coefficients (PLPCs). In order to recognise anger, happy, boredom, sad and neutral emotions combination of PLP, RASTA, LPCC and MFCC and log frequency is used. A number of features can be extracted using feature extraction techniques. In order to achieve higher accuracy features should be selected wisely. There are various feature selection algorithm present some of them are Forward selection and backward selection. In forward selection there is linear loss function to which a feature is added that provides the least error. Whereas in the backward selection all features are selected at first instance and then the feature that minimises loss function is removed.

After feature extraction and feature selection the next step is to choose a suitable classifier. As classifier also contributes in accuracy of emotions recognised. There are number of classifiers available namely HMM, GMM, ANN, SVM etc. Combination of classifier can also be used making a hybrid model. Each one has some pros and cons over the other. Some are good at large database, some are good for some relevant features. Following are some of the commonly used classifiers. The Hidden Markov model has been widely used in the literature but its classification property is not upto the mark. Accuracy with above 70% Recognition rate was obtained by using Hidden Markov Model. Artificial Neural Network is also popularly used for emotion recognition from speech A three layered(two hidden and one output) feed forward neural network is used. Another popular technique is Support Vector Machine(SVM). SVM creates an hyperplane in high or infinite space for classification. An overall accuracy of 94.2% is achieved using isolated SVM.

V. PROPOSED WORK

The experimentation performed over MATLAB2013a running on Intel Core I5 processor operating at 5.5GHz. Windows 7 is the basic platform to execute MATLAB commands from higher level to lower level. In this work we have investigated the performance of neural network model on the features extracted from the audio files. Basicaly three categories of audio files are incorporated in this research work, related to sad, happiness and neutral moods of person. Various relevant features are extracted from each audio files including statistical energy, skewness, MFCC, entropy etc.

The experimental result of machine learning experimental is shown in fig.2.

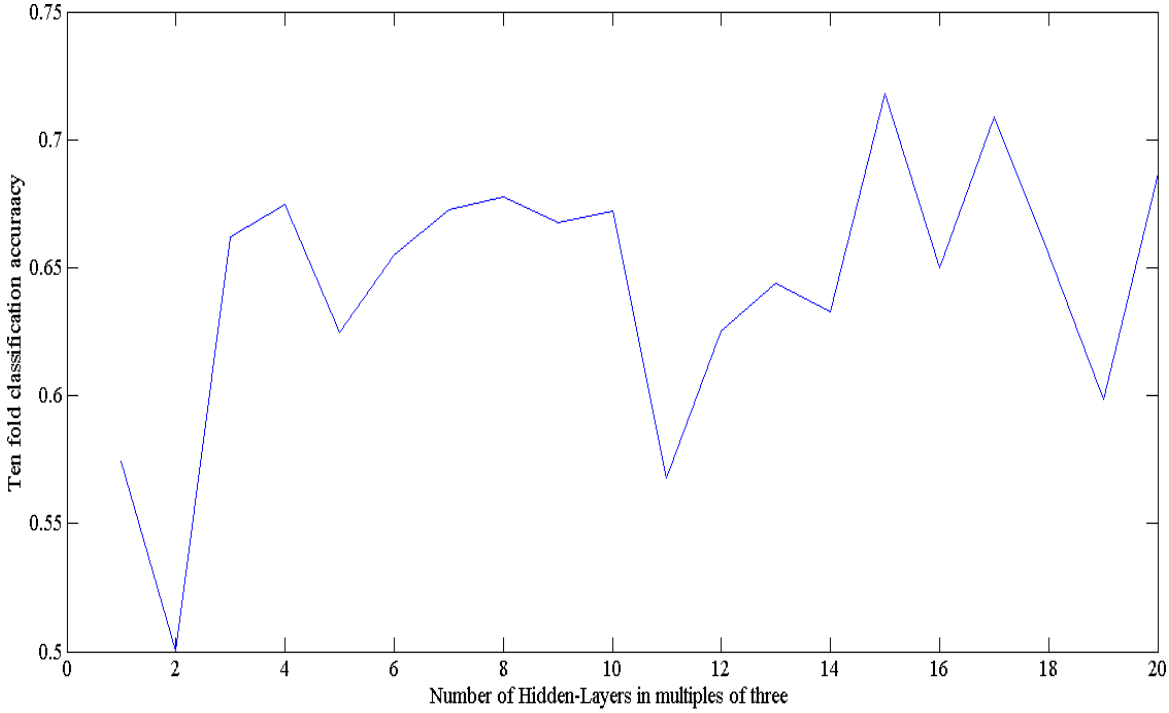


Fig. 2 Output waveform

The dataset of this research activity involves 121 instances of happiness, 117 instances of sadness and the rest 116 instances of neutral mood. The extracted features set comprised of overall 20 features.

The complexity of decision space is investigated by varying various numbers of neurons in the single hidden layers of Vanilla Artificial neural Network model.

VI. CONCLUSION

The results show that with the employment of a features selection algorithm, a satisfying recognition rate level can still be obtained also reducing the employed features and, as a con-sequence, the number of operations required to identify the emotional contents. we can conclude that the maximum achieved accuracy is 72 percents at 45 numbers of neurons in the hidden layer.

REFERENCES

[1] Dellaert, F., Polzin, T., Waibel, A.: “Recognizing emotion in speech”, In: Proceedings of ICSLP, Philadelphia, USA Automatic Recognition of Emotions from Speech 89 1996.

[2] Devillers, L., Vidrascu, L., Lamel, L.: “Challenges in real-life emotion annotation and machine learning based detection,” Neural Networks, 18(4), 407–422 (2005).

[3] 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).

[4] 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).

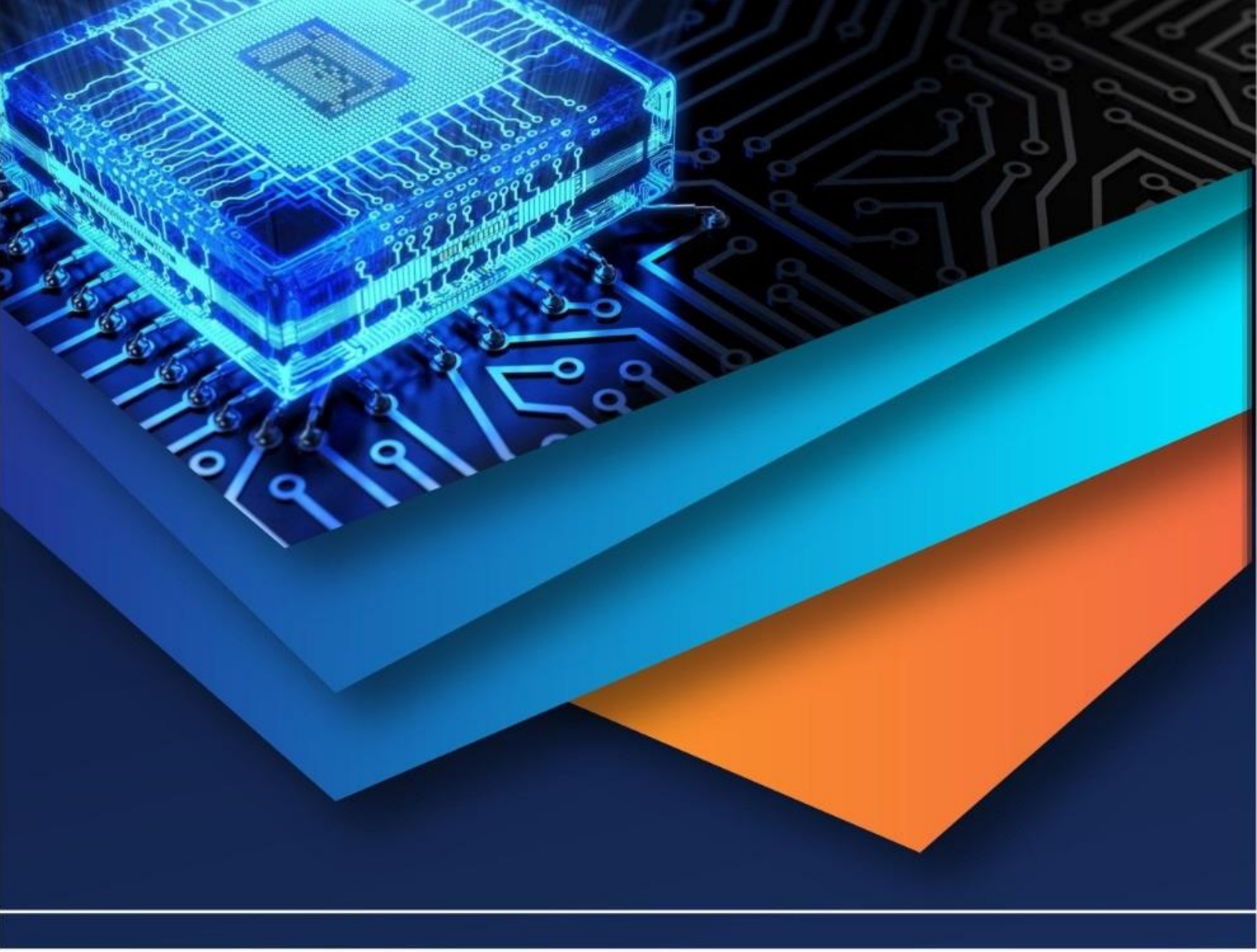
[5] Engberg, I.S., Hansen, A.V.: “Documentation of the Danish Emotional Speech Database (DES),” Technical report. Aalborg University, Aalborg, Denmark (1996).

[6] Schiel, F., Steininger, S., T¨urk, U.: “The SmartKom multimodal corpus at BAS,” In: Proceedings of the 3rd Language Resources & Evaluation Conference (LREC) 2002, Las Palmas, Gran Canaria, Spain, pp. 200–206 (2002).

[7] Batliner, A., Hacker, C., Steidl, S., N¨oth, E., D’Arcy, S., Russell, M., Wong, M.: “You stupid tin box” - children interacting with the AIBO robot: A cross-linguistic emotional speech corpus,” In: Proceedings of the 4th International Conference of Language Resources and Evaluation LREC 2004, Lisbon, pp. 171–174 (2004).

[8] M. You, C. Chen, J. Bu, J. Liu, and J. Tao, “Getting started with susas: a speech under simulated and actual stress database,” EuroSpeech, vol. 4, pp. 1743–1746, 1997.

[9] S. Haq, P. Jackson, and J. Edge, “Audio visual feature selection and reduction for emotion classification,” AVSP, pp. 185–190, 2008.



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY