International Journal of Engineering and Advanced Technology (IJEAT)
ISSN: 2249-8958 (Online), Volume-1 Issue-5, June 2012

Speech Emotion Recognition by using SVM-Classifier

Vaishali M. Chavan, V.V. Gohokar

Abstract*— Automatic emotion recognition in speech is a current research area with a wide range of applications in human-machine interactions. This paper uses the support vector machine (SVM), to classify five emotional states: anger, happiness, sadness, surprise and a neutral state. The classification performance of the selected feature subset was done with that of the Mel frequency cepstrum coefficients (MFCC), Periodicity Histogram and Fluctuation Pattern. Within the method based on SVM, a new method by using Multi-class SVM is used as a classifier. Experiments were conducted on the Danish Emotion Speech (DES) Database. The recognition rates by using SVM classifier were 68 %, 60 %, 55.40 % and 60 % for Linear, Polynomial, RBF, and Sigmoid Kernel Function respectively. The recognition rates by Multiclass SVM using Linear, Polynomial, RBF and Sigmoid kernel function for Danish database for Periodicity Histogram are 64.77 %, 78.41 %, 79.55 % and 78.41% respectively.

Keywords*— Emotion recognition, Mel frequency cepstrum coefficients (MFCC), Support Vector Machine.

I. INTRODUCTION

Speech emotion recognition is an important part in emotion recognition. Accurate detection of emotion from speech has clear benefits for the design of more natural human- machine speech interfaces or for the extraction of useful information from large quantities of speech data. It is also becoming more and more important in computer application fields as health care, children education, etc. In speech-based communications, emotion plays an important role. Sometimes, playing an even bigger role than the logical information included in the speech. Research has long been done on emotion in the fields of psychology and physiology. More recently it is the subject of attention by engineers. Its most important application is in intelligent human-machine interaction. In today's human-machine interaction systems, machines can recognize "what is said" and "who said it" using speech recognition and speaker identification techniques. If it is equipped with emotion recognition techniques, machines can also know "how it is said" to react more appropriately, and make the interaction more natural. Other applications of automatic emotion recognition include psychiatric diagnosis, intelligent toys, and lie detection. [1] Today there are many different algorithms which were used for various signal processing applications. In this paper the emotions from the speech signal are recognized by considering the features of the speech signal by using SVM classifier. While classifying different emotions, several features like Periodicity Histogram, Fluctuation Pattern and MFCC (Mel Frequency Cepstral Coefficient) is used. [17] The data samples of speech are separated into two groups. The first group is used for training the data samples and the second group is used for testing purpose. The SVM classifier is used to classify different emotions from these testing data samples. The data samples which were used for testing purpose is to be compared with the data samples which is already trained. This comparison gives the detection of emotion from these data samples. In this project, Danish Emotional Database is used which gives the satisfactory detection of emotions. [26, 30]

.

II. SPEECH EMOTION RECOGNITION SYSTEM

The structure of the speech emotion recognition system studied in this paper is depicted in Figure 1. Like the typical pattern recognition system, it contains five main modules: emotional speech input, feature extraction, feature labelling, SVM training and classification and recognized emotion output. A feature selection module is part of the SVM classifier.

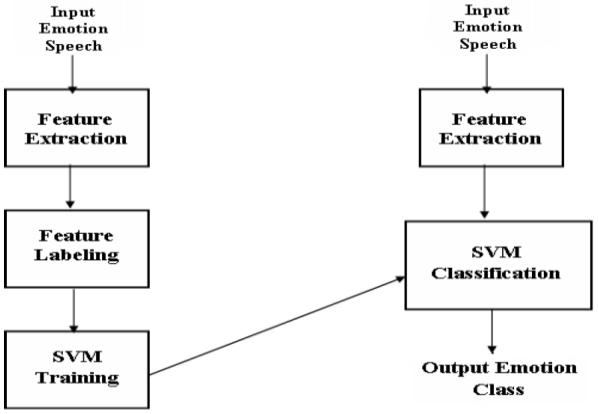


Fig.1: Structure of Speech Emotion Recognition System

A. Feature Extraction

In feature extraction process three features are extracted MFCC, fluctuation pattern and periodicity histogram. The continuous time signal (speech) is sampled at sampling frequency. At the first stage in MFCC feature extraction is to boost the amount of energy in the high frequencies. This pre-emphasis is done by using a filter. Then framing or segmenting of the speech samples are obtained from analog to digital conversion (ADC), into the small frames with the time length within the range of 20-40 msec.

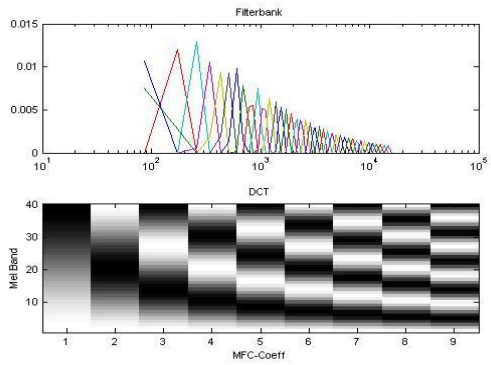
Manuscript published on 30 June 2012.
* Correspondence Author (s)
Vaishali M. Chavan, Electronics & Telecommunication Engg., Amravati University/ S. S. G. M. C. E, Shegaon
V. V. Gohokar, Electronics & Telecommunication Engg., Amravati University/ S. S. G. M. C. E, Shegaon

© The Authors. Published by Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP). This is an open access article under the CC-BY-NC-ND license http://creativecommons.org/licenses/by-nc-nd/4.0/

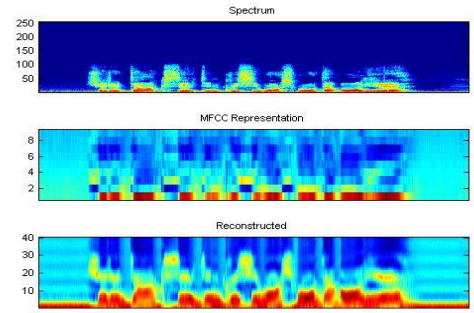
Framing enables the non-stationary speech signal to be segmented into quasi-stationary frames, and enables Fourier Transformation of the speech signal. It is because, speech signal is known to exhibit quasi-stationary behaviour within the short time period of 20-40 msec. Windowing is the another step which is meant to window each individual frame, in order to minimize the signal discontinuities at the beginning and at the end of each frame. Then Fast Fourier Transform (FFT) [11] algorithm is used for evaluating the frequency spectrum of speech. FFT converts each frame of N samples from the time domain into the frequency domain. The mel filter bank consists of overlapping triangular filters with the cut-off frequencies determined by the centre frequencies of the two adjacent filters. The filters have linearly spaced centre frequencies and fixed bandwidth on the mel scale. The logarithm has the effect of changing multiplication into addition. Therefore, this step simply converts the multiplication of the magnitude in the Fourier transform into addition. Take Discrete Cosine Transform to orthogonalise the filter energy vectors. Because of this orthogonalisation step, the information of the filter energy vector is compacted into the first number of components and shortens the vector to number of components.

First of all, take one input audio sample. Then convert its frequency scale to bark scale (as it is more close to human frequency interpretation). As a spreading function we can use ‘terhardt’ or ‘modified terhardt’. Using this bark scale representation, the frames of the input is converted into 'sone' units. (So it will be easy to differentiate between the actual speech and noise). Using this sone values, we can plot a histogram (it can be work as a feature, but we don't use this as it does not differentiate the emotions properly).

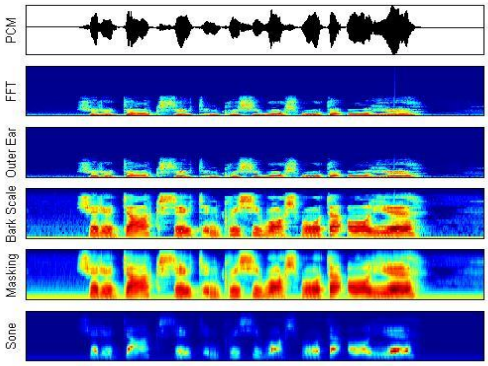
Again using this sone values and its distribution we can design a fluctuation pattern (it can work as a feature which gives good results). Again using this sone values and its distribution we can design a periodicity histogram (it can also work as a feature).



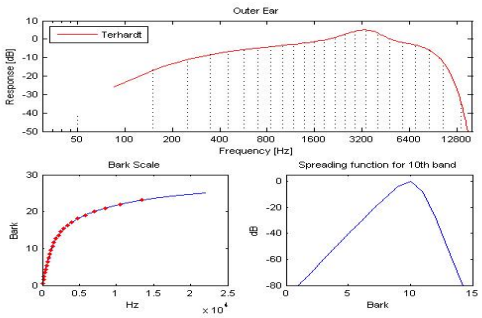
(a)



(b)



(c)



(d)

Fig. 2: Waveforms of extracted features (a: Filtered signal; b: MFCC & Spectrum of speech signal, c: Converted Sone signal, & d: Frequency Response of signal)

B. SVM training and classification

Now once we have a set of features available with us we can take use of the classifiers to distinguish between the different emotional states. Now we take use of the multiclass SVM classifier for classification purpose. First we train the classifier with some inputs of different emotional states. After training the classifier, we can use it for recognizing the new given input. The process of SVM training contains labelling of extracted features and training SVM. In the training process, every extracted feature has to assign an associated class label. The SVM is trained according to this labelled feature. The SVM kernel functions are used in the training process of SVM. The result can be improved if we use all the above features properly.

Published By:
Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP)
© Copyright: All rights reserved.

III. EXPERIMENTAL STUDY

A. Emotional Speech Database

The emotional speech database used in this study is the Danish Emotional Speech (DES) Database which includes expressions by four actors familiar with radio theatre, two male and two female. The speech is expressed in five emotional states: anger, happiness, neutral, sadness and surprise. In this study, the speech signal was re-sampled to 16 kHz, and the silence segments at the beginning and the end of the speech were cut out artificially. Total 88 samples were used for the conduction of emotion recognition process.

B. Experimental Results

The MFCC, Fluctuation Pattern and Periodicity Histogram features are extracted. Every feature value must have corresponding label of belonging class. As SVM is binary classifier the class labels are $\{+1, -1\}$. In this work SVM using linear kernel function is implemented for non separable data. The test results are calculated by using implemented SVM as well as Multiclass SVM. In Multiclass SVM [28] results are taken by using Fluctuation Pattern, Periodicity Histogram and MFCC feature for all the Kernel Functions. The overall percentage accuracy is obtained by using Multiclass SVM. Whereas for SVM classifier the particular emotion from the selected test file is to be obtained.

Table 1: Confusion Matrix for SVM Classifier (Linear Kernel Function)

Emotion	*Anger*	*Fear*	*Happy*	*Normal*	*Sad*
Anger	*61.53*	*6.67*	*17.78*	*2.22*	*2.22*
Fear	*15.38*	*66.67*	*20*	*6.67*	*8.89*
Happy	*15.38*	*15.56*	*57.78*	*1.11*	*2.22*
Normal	*2.56*	*4.44*	*2.22*	*74.44*	*8.89*
Sad	*5.12*	*6.67*	*2.22*	*15.56*	*77.78*

Table 1 shows confusion matrix of implemented SVM for Danish emotion speech utterance using one-to-one multiclass method. The Sad emotion gives maximum 77.78% and Happy gives minimum 57.78% recognition rate. The overall recognition accuracy for Danish emotion is 68% by using Linear Kernel Function.

Table 2: Confusion Matrix for SVM Classifier (Polynomial Kernel Function)

Emotion	Anger	Fear	Happy	Normal	Sad
Anger	51.28	8.89	15.56	7.78	2.22
Fear	30.77	75.56	44.44	7.78	11.11
Happy	10.26	8.89	35.56	3.33	2.22
Normal	2.56	2.22	2.22	63.33	8.89
Sad	5.13	4.44	2.22	17.78	75.56

Table 2 shows confusion matrix of implemented SVM for Danish emotion speech utterance using one-to-one multiclass method. The Sad and Fear emotion gives maximum 75.56% and Happy gives minimum 35.56% recognition rate. The overall recognition accuracy for Danish emotion is 60% by using Polynomial Kernel Function.

Table 3: Confusion Matrix for SVM Classifier (RBF Kernel Function)

Emotion	Anger	Fear	Happy	Normal	Sad
Anger	46.15	4.44	13.33	0	0
Fear	2.56	48.89	0	0	0
Happy	10.26	2.22	42.22	0	0
Normal	0	0	0	42.22	2.22
Sad	41.02	44.44	44.44	57.78	97.78

Table 3 shows confusion matrix of implemented SVM for Danish emotion speech utterance using one-to-one multiclass method. The Sad emotion gives maximum 97.78% and Happy and Normal gives minimum 42.22% recognition rate. The overall recognition accuracy for Danish emotion is 55.40% by using RBF Kernel Function.

Emotion	Anger	Fear	Happy	Normal	Sad
Anger	97.45	71.11	80.00	33.33	66.67
Fear	2.56	24.44	0	0	0
Happy	10.26	4.44	20.00	0	0
Normal	0	0	0	30.00	2.22
Sad	0	0	0	3.33	31.11

Table 4: Confusion Matrix for SVM Classifier (Sigmoid Kernel Function)

Table 4 shows confusion matrix of implemented SVM for Danish emotion speech utterance using one-to-one multiclass method. The Angry emotion gives maximum 97.45% and Happy gives minimum 20% recognition rate. The overall recognition accuracy for Danish emotion is 60% by using Sigmoid Kernel Function.

Table 5: For Multiple Class SVM Test Method

Sr. No.	Kernel Function	Fluctuation Pattern	Periodicity Histogram	MFCC
1.	Linear	62.5%	64.77%	54.55%
2.	Polynomial	34.09%	78.41%	61.36%
3.	Radial Basis Function	39.77%	79.55%	39.77%
4.	Sigmoid	14.77%	78.41%	14.77%

Figure 6: Graph for Multiple Class SVM Test Method

Published By:
Blue Eyes Intelligence Engineering
and Sciences Publication (BEIESP)
© Copyright: All rights reserved.

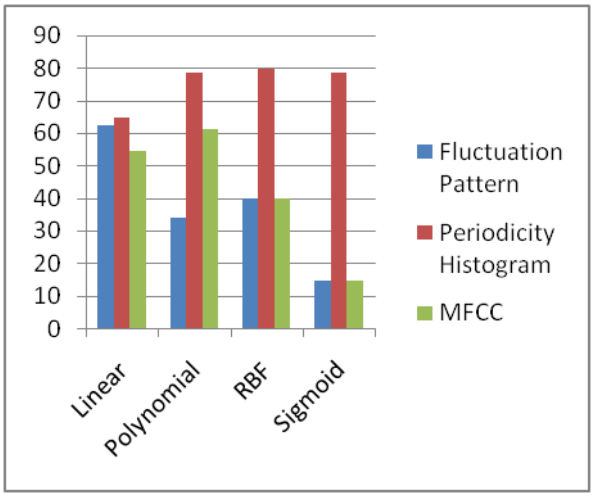
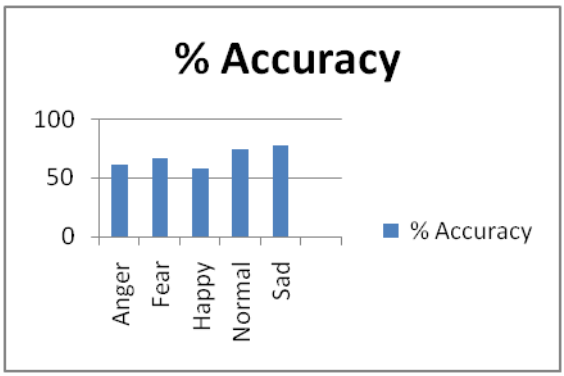


Figure 5 & 6 shows the table & graph of Multiclass SVM Test Method. The graph shows Kernel functions on X-axis and Percentage accuracy on Y- axis. It gives the Percentage accuracy on the basis of four kernel functions by using three features i.e. Fluctuation Pattern, Periodicity Histogram and MFCC.

Figure 7: Graph for Linear Kernel function



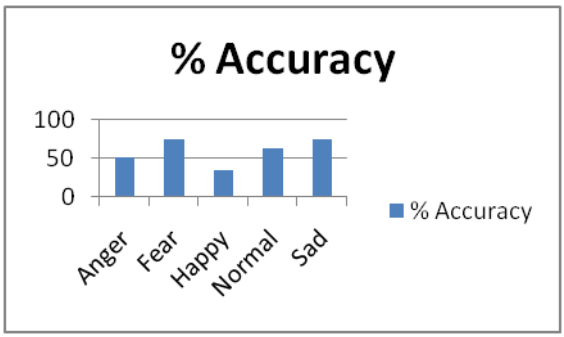


Figure 8: Graph for Polynomial Kernel function

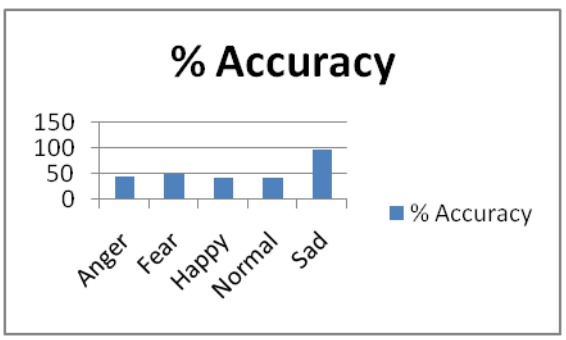


Figure 9: Graph for RBF Kernel function

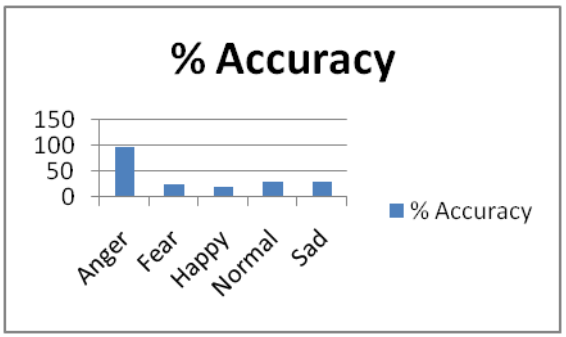


Figure 10: Graph for Sigmoid Kernel function

IV. CONCLUSION

From experimentation and results, it is proved that system is speaker and text independent. It is also observed that results from SVM and Multiclass SVM are much better. The recognition rates by implemented SVM classifier are 68 %, 60 %, 55.40 % and 60 % for Linear, Polynomial, RBF, and Sigmoid Kernel Function respectively. The recognition rates by Multiclass SVM using Linear, Polynomial, RBF and Sigmoid kernel function for Danish database for Periodicity Histogram are 64.77 %, 78.41 %, 79.55 % and 78.41% respectively. From results it is observed that maximum confusion occurs while recognizing happy and anger emotions for Danish emotion utterances. Total time required is 15 seconds for training data set of 25 values and 1 second for testing each value for SVM implementation, whereas total time required for testing 88 values by using Multiclass SVM is 43 sec. In future, this time period required for the testing purpose can be reduced and hence the efficiency can be improved.

REFERENCES

[1] Proceedings of the Fourth International Conference on Machine Learning and Cybernetics, Guangzhou, 18-21 August 2005, "SPEECH EMOTION RECOGNITION BASED ON HMM AND SVM" by YI-LIN LIN, GANG WEI.
[2] "HIDDEN MARKOV MODEL-BASED SPEECH EMOTION RECOGNITION" by Björn Schuller, Gerhard Rigoll, and Manfred Lang, Institute for Human-Computer Communication, 2003 IEEE.
[3] "Timing Levels in Segment-Based Speech Emotion Recognition", by Björn Schuller and Gerhard Rigoll of Institute for Human-Machine Communication in 2006.
[4] "A NEURAL NETWORK APPROACH FOR HUMAN EMOTION RECOGNITION IN SPEECH", by Muhammad Waqas Bhatti, Yongjin Wang and Ling Guan in 2006.
[5] "Comparison Between Fuzzy and NN Method for Speech Emotion Recognition", by Aishah Abdul Razak, Ryoichi Komiya, Mohamad Izani Zainal Abidin, in 2005 IEEE.
[6] 2005 IEEE International Workshop on Robots and Human Interactive Communication, "Fuzzy Emotion Recognition in Natural Speech Dialogue", by Anja Austermann, Natascha Esau, Lisa Kleinjohann and Bernd Kleinjohann.
[7] "Emotion Recognition on the Basis of Human Speech", by Zygmunt Ciota Technical University of Lodz, Department of Microelectronics and Computer Science.
[8] SICE-ICASE International Joint Conference 2006, Oct. 18-21, 2006 in Bexco, Busan, Korea, "Robust Speech Emotion Recognition Using Log Frequency Power Ratio", by Kyung-Hak Hyun, Eun-Ho Kim and Yoon-Keun Kwak.
[9] "Speech Emotion Recognition Based on Rough Set and SVM", by Jian Zhou, Guoyin Wang, Yong Yang, Peijun Chen, 2006 IEEE.

Published By:
Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP)
© Copyright: All rights reserved.

[10] "GMM SUPERVECTOR BASED SVM WITH SPECTRAL FEATURES FOR SPEECH EMOTION RECOGNITION", by Hao Hu, Ming-XingXu, and Wei Wu, Tsinghua University, Beijing, 2007 IEEE.
[11] "SPEECH EMOTION RECOGNITION USING GAUSSIAN MIXTURE VECTOR AUTOREGRESSIVE MODELS" by Moataz M. H. El Ayadi, Mohamed S. Kamel, and Fakhri Karray, Pattern Analysis and Machine Intelligence Lab, Electrical and Computer Engineering, University of Waterloo,2007 IEEE.
[12] IMACS Multi-conference on "Computational Engineering in System Applications"(CESA), October 4-6, 2006, Beijing, China." Emotion-detecting Based Model Selection for Emotional Speech Recognition", by Y. C. Pan, M. X. Xu, L. Q. Liu, P. F. Jia.
[13] "Speech Emotion Recognition in E-learning System Based on Affective Computing" by Wu Li, Yanhui Zhang, Yingzi Fu, Third International Conference on Natural Computation (ICNC 2007).
[14] 2007 IEEE International Conference on Control and Automation FrBP-27 Guangzhou, CHINA - May 30 to June 1, 2007. Research on "Speech Emotion Recognition in E- learning By Using Neural Networks Method", by Qian Zhang, Yan Wang, Lan Wang and Guoqiang Wang, 2007 IEEE.
[15] 16th IEEE International Conference on Robot & Human Interactive Communication August 26 - 29, 2007 / Jeju, Korea, "Speech Emotion Recognition Using Eigen-FFT in Clean and Noisy Environments", Eun Ho Kim ,Kyung Hak Hyu, Soo Hyun Kim and Yoon Keun Kwak, 2007 IEEE.
[16] "Speech Emotion Recognition using Auditory Cortex", Abdul Wahab, Chai Quek, and Sussan De, 1-4244-1340-0/07/$25.00 c-2007 IEEE.
[17] "SPEECH EMOTION VERIFICATION SYSTEM (SEVS) BASED ON MFCC FOR REAL TIME APPLICATIONS", by Norhaslinda Kamaruddin and Abdul Wahab,
[18] "Adaptive and Optimal Classification of Speech Emotion Recognition", by Ying Wang, Shoufu Du , Yongzhao Zhan. 2008 IEEE, DOI - 10.1109/ICNC.2008.713.
[19] "Efficient Speech Emotion Recognition Based on Multisurface Proximal Support Vector Machine", by Chengfu Yang, Xiaorong Pu, Xiaobin Wang, 2008 IEEE
[20] "Speech Emotion Recognition Using Canonical Correlation Analysis and Probabilistic Neural Network", by Ling Cen, Wee Ser, Zhu Liang Yu. 2008 Seventh International Conference.
[21] 2009 World Congress on Computer Science and Information Engineering, "Multi-Level Speech Emotion Recognition based on HMM and ANN" by Xia Mao, Lijiang Chen, Liqin Fu, 2008 IEEE.
[22] "Recognition of emotions in speech by a hierarchical approach", by Zhongzhe Xiao and Emmanuel Dellandrea and Liming Chen, Weibei Dou, 2009 IEEE.
[23] "Speech emotion recognition using both spectral and prosodic features", by Yu Zhou and Yanqing Sun and Jianping Zhang and Yonghong Yan, 2009 IEEE.
[24] 2009 International Conference on Advances in Computing, Control, and Telecommunication Technologies, "Automatic Emotion Recognition from Speech using Artificial Neural Networks with Gender-Dependent Databases", by Firoz Shah. A, Raji Sukumar. A, Babu Anto. P, 2009 IEEE.
[25] 2010 International Conference on Measuring Technology and Mechatronics Automation, "Speech Emotion Recognition Based on Principal Component Analysis and Back Propagation Neural Network", by Sheguo Wang,Xuxiong Ling,Fuliang Zhang,Jianing Tong, 2010 IEEE
[26] "SVM - MLP - PNN Classifiers on Speech Emotion Recognition Field -A Comparative Study", by Theodoros Iliou, Christos-Nikolaos Anagnostopoulos, 2010 IEEE. DOI 10.1109/ICDT.2010.8
[27] "Speech Emotion Recognition Based on Data Mining Technology", by Ying SHI, Weihua SONG, 2010 IEEE.
[28] 2010 International Conference on Electrical and Control Engineering, "Fuzzy multi-class support vector machine based on binary tree in network intrusion detection", by Lei L, Zhi-ping GA ,Wen-yan Dinl.
[29] "Speech emotion recognition using segmental level prosodic analysis", by Shashidhar G. Koolagudi, Nitin Kumar and K. Sreenivasa Rao, 2011 IEEE.
[30] 2011 International Conference on Electronic & Mechanical Engineering and Information Technology, "Automatic Speech Emotion Recognition Using Support Vector Machine", by Peipei Shen, Zhou Changjun, Xiong Chen.
[31] 2011 International Conference on Image Information Processing (ICIIP 2011) "Statistical Estimation of Emotions in Speech Notes", by Featured Term Analogy, Sita Kumari K., Suhasini S., Zaheer Parvez Shaik Mohd.
[32] Sixth International Conference on Spoken Language Processing (ICSLP 2000), "EMOTION RECOGNITION IN SPEECH SIGNAL: EXPERIMENTAL STUDY, DEVELOPMENT, AND APPLICATION", by Valery A. Petrushin.
[33] "HIDDEN MARKOV MODEL-BASED SPEECH EMOTION RECOGNITION" by Björn Schuller, Gerhard Rigoll, and Manfred Lang, 2003 IEEE.

First Author: I have done my B.E. (in Electronics & Telecommunication Engineering) from S.S. G. M. C. E., Shegaon. From the same college I am pursuing my M. E. (in Digital Electronics).
Second Author: I am V. V. Gohokar from Electronics & Telecommunication Department. I am the Vice Principal at S. S. G. M. C. E. Shegaon.

Published By:
Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP)
© Copyright: All rights reserved.