

# RESEARCH PAPER ON AUTOMATIC REAL-TIME SPEECH EMOTION RECOGNITION USING SUPPORT VECTOR MACHINE

Ms. Talvinder Kaur Bedi, Mr. Pankaj Sharma

<sup>1</sup>M. Tech Scholar, Department of CSE, Sachdeva Institute of Technology, Mathura, Agra, Uttar Pradesh, India

<sup>2</sup>HOD, Department of CSE, Sachdeva Institute of Technology, Mathura, Agra, Uttar Pradesh, India

\*\*\*

**ABSTRACT:-** In this paper we present a system for Real-time feeling affirmation from talk using Support Vector Machine (SVM) as a gathering method. Customized Speech Emotion Recognition (ASER) is a best in class examination zone in the field of Human Computer Interaction Intelligence (HCII). Human emotions can be perceived from their talk banner by removing a part of the talk acoustic and prosodic features like pitch, Mel repeat Cepstral Coefficient (MFCC) and Mel Energy Spectrum Dynamic Coefficient (MEDC). Here SVM classifier is used to mastermind the emotions as irritation, fear, fair-minded, awful, shock, chipper and exhaustion. UGA and LDC datasets are used for detached examination of sentiments using LIBSVM parcel limits. With this examination the machine is arranged and planned for perceiving emotions logically talk.

Modified Speech Emotion Recognition (SER) is a recurring pattern inspect subject in the field of Human Computer Interaction (HCI) with wide extent of employments. The talk features, for instance, Mel Frequency cepstrum coefficients (MFCC) and Mel Energy Spectrum Dynamic Coefficients (MEDC) are expelled from talk explanation. The Support Vector Machine (SVM) is used as classifier to amass assorted energetic states, for instance, shock, satisfaction, hopelessness, fair, fear, from Berlin enthusiastic database. The LIBSVM is used for game plan of sentiments. It gives 93.75% request precision for Gender free case 94.73% for male and 100% for female talk.

**Watchword:** Support Vector Machine, Speech Signal, Experimentation, Emotion Analysis.

Controller (PDC)

## I. INTRODUCTION

Programmed Speech Emotion Recognition is an ongoing exploration subject in the Human Computer Interaction (HCI) field. As PCs have turned into a necessary piece of our lives, the need has ascended for an increasingly regular correspondence interface among people and PCs. To accomplish this objective, a PC would need to have the option to see its current circumstance and react diversely relying upon that observation. Some portion of this procedure includes understanding a client's passionate state. To make the human-PC association progressively common, it is useful to enable PCs to

perceive enthusiastic circumstances a similar route as human does.

Programmed Emotion Recognition (AER) should be possible in two different ways, either by discourse or by outward appearances. In the field of HCI, discourse is essential to the targets of a feeling acknowledgment framework, as are outward appearances and signals. Discourse is considered as an amazing mode to speak with aims and feelings. In the ongoing years, a lot of research has been done to perceive human feeling utilizing discourse data [1], [2]. Numerous specialist investigated a few characterization techniques including the Neural Network (NN), Gaussian Mixture Model (GMM), Hidden Markov Model (HMM), Maximum Likelihood Bayes classifier (MLC), Kernel Regression and K-closest Neighbors (KNN), Support Vector Machine (SVM) [3], [4]. The Support Vector Machine is utilized as a classifier for feeling acknowledgment.

The SVM is utilized for arrangement and relapse reason. It performs order by developing a N-dimensional hyperplanes that ideally isolates the information into classes. The arrangement is accomplished by a direct or nonlinear isolating surface in the info highlight space of the dataset. Its principle thought is to change the first info set to a high-dimensional element space by utilizing a part capacity, and afterward accomplish ideal order in this new element space. A Berlin Emotional database [5] is utilized for highlight extraction and preparing SVM. The Berlin database of passionate discourse was recorded at the Technical University, Berlin. The database contains discourse with acted feelings in German language. It contains 493 expressions of 10 proficient on-screen characters five guys and five females who talked 10 sentences with sincerely unbiased substance in 7 unique feelings.

The feelings were wut (outrage), langeweile (fatigue), ekel (sicken), angst (dread), freude (joy), trauer (pity) and nonpartisan passionate state. Utilizations of Speech Emotion Recognition incorporate mental analysis, insightful toys, lie identification, learning condition, instructive programming, and discovery of the passionate state in phone call focus discussions to give criticism to an administrator or a boss for checking purposes.

### 1.1 Objective of research work:

The objective of research work are:

1. To perform broad writing survey of discourse feeling acknowledgment.
2. To investigation and examination the different component, database, AI calculation of discourse feeling acknowledgment.
3. The primary goal of this postulation is to utilize Support Vector Machine (SVM) classifier to order seven unique feelings joy, outrage, pity, fatigue, nauseate, unbiased, dread.

#### 1.2 Application of discourse feeling acknowledgment

Uses of feeling characterization dependent on discourse have just been utilized to encourage connections in our everyday lives.

For instance

1. In call focuses apply feeling characterization to organize eager clients.
2. A notice framework has been created to identify if a driver shows outrage or forceful feelings.
3. For separation learning, to identifying understudies feeling auspicious and making proper move can improve the nature of instructing.
4. Feeling detecting has likewise been utilized in conduct concentrates sound highlights have been broadly investigated in both the time space and the recurrence area.

## II. LITERATURE REVIEW

To do this work, broad writing has been experienced. In this area, we present in detail various papers which have been mulled over. They give an unmistakable image of the exploration which is going on accentuation has been put on writing identified with discourse feeling acknowledgment utilizing SVM.

**Wootae Lim et al., (2017):** examined that with fast improvements in the structure of profound engineering models and learning calculations, strategies alluded to as profound learning have come to be generally utilized in an assortment of research zones, for example, design acknowledgment, grouping, and sign preparing. Convolutional Neural Networks (CNNs) particularly show exceptional acknowledgment execution for PC vision undertakings.

**Pavitra Patel et al., (2017):** explored that discourse has a few trademark highlights, for example, instinctive nature and proficient, which makes it as appealing interface medium. It is conceivable to express feelings and demeanors through discourse. In human machine

interface application feeling acknowledgment from the discourse sign has been momentum theme of research.

**Prajakta P. Dahake et al., (2016):** explored that in human PC collaboration, discourse feeling acknowledgment is having an essential impact in the field of research. Human feelings comprise of being furious, upbeat, pitiful, appall, unbiased.

**Ritu D.Shah and Dr. Anil. C.Suthar (2016):** In this paper system for feeling acknowledgment from discourse sign is displayed. Some of sound highlights are expelled from discourse sign to dissect the highlights and conduct of discourse.

**Kunxia Wang et al., (2015):** As of late, considers have been performed on congruity attributes for discourse feeling acknowledgment.

**Rahul B.Lanjewar et al., (2015):** examined that the family relationship among man and machines has turned into another pattern of innovation with the end goal that machines currently need to react by thinking about the human passionate dimensions.

**S. Lalitha et al., (2015):** explored that feeling acknowledgment from discourse causes us in improving the viability of human-machine association. This paper exhibits a technique to recognize appropriate highlights in DWT area and improve great precision.

**Qirong Mao et al., (2014):** explored that a fundamental method for human passionate conduct understanding, discourse feeling acknowledgment (SER) has pulled in a lot of consideration in human-focused sign preparing.

**Monica Feraru and Marius Zbancioc (2013):** examined an improved adaptation of the old style KNN calculation which partners to every parameter from the highlights vectors loads as indicated by their exhibition in the arrangement procedure.

**Thapanee Seehapoch and Sartra Wongthanavasu (2013):** considered that programmed acknowledgment of passionate states from human discourse is a flow examine theme with a wide range. In this paper an endeavor has been made to perceive and group the discourse feeling from three language databases, in particular, Berlin, Japan and Thai feeling databases.

**Yixiong Pan et al., (2012):** contemplated that Speech Emotion Recognition is an ongoing exploration territory in the field of Human Computer Interaction (HCI). Here we perceive three passionate states: upbeat, dismal and unbiased.

**Mansour Sheikhan et al., (2012):** researched that acknowledgment of passionate discourse, the exhibition of programmed discourse acknowledgment (ASR) frameworks is debased fundamentally.

**Simina Emerich and Eugen Lupu (2011):** contemplated the acknowledgment of the inner passionate condition of an individual assumes a significant job in a few human-related fields.

**MoatazEl Ayadi and Mohamed S.Kamel (2011):** explored that expanding consideration has been coordinated to the investigation of the passionate substance of discourse signals, and consequently, numerous frameworks have been proposed to distinguish the enthusiastic substance of an expressed articulation.

**Iulia Lefter et al., (2010):** considered potential outcomes for upgrading the consensus, versatility and vigor of feeling acknowledgment frameworks by consolidating databases and by combination of classifiers.

**Aditya Bihar Kandali et al., (2008):** explored a technique dependent on Gaussian blend model (GMM) classifier and Mel-recurrence cepstral coefficients (MFCC) as highlights for feeling acknowledgment from Assamese talks.

**Hao Hu et al., (2007):** researched that discourse feeling acknowledgment is a difficult yet significant discourse innovation. In this paper, the GMM super vector based SVM is connected to this field with otherworldly highlights.

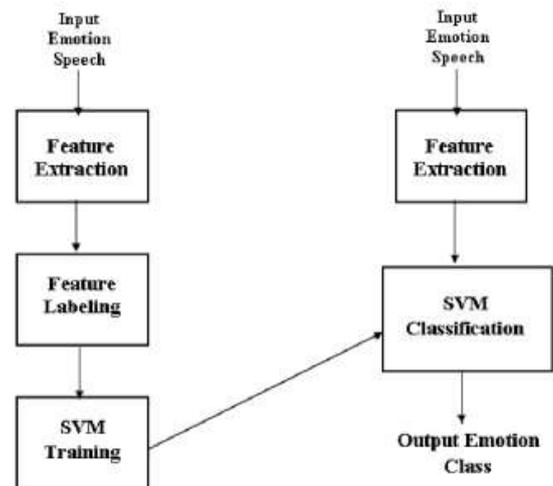
**Suja P et al., (2005):** Human face conveys significant data about an individual's enthusiastic condition. In this paper a methodology for outward appearance acknowledgment utilizing wavelet change for highlight extraction and neural system classifier for five fundamental feelings is proposed.

**Schuller et al., (2004):** They acquaint a novel methodology with the blend of acoustic highlights and language data for a most hearty programmed acknowledgment of a speaker's feeling. Seven enthusiastic states are grouped during the work. Initially a model for the acknowledgment of feeling by sound highlights is introduced.

**Tin Lay Nwe et al., (2003):** examined that feeling order of discourse flag, the well known highlights utilized are insights of crucial recurrence, span of quiet, vitality form and voice quality.

### III. SYSTEM IMPLEMENTATION

The significance of feelings in human-human connection gives the premise to analysts in the building and



**Figure 1. Speech Emotion Recognition System.**

software engineering networks to create programmed ways for PCs to perceive feelings. As appeared in fig. 1 the contribution to the framework is a .wav record from Berlin Emotion Database that contains passionate discourse expression from various enthusiastic classes. After that highlights extraction procedure is completed. In highlight extraction process two highlights are extricated MFCC [6], [7] and MEDC [8]. After that the extricated highlights and their comparing class marks are given as contribution to the LIBSVM classifier. The yield of a classifier is a name of a specific feeling class. There are all out five classes furious, pitiful, cheerful, unbiased and dread. Each name speaks to comparing feeling class.

**3.1 Feature Extraction** In past works a few highlights are extricated for ordering discourse influence, for example, vitality, pitch, formants frequencies, and so on all these are prosodic highlights. As a rule prosodic highlights are essential marker of speaker's passionate state. Here in highlight extraction process two highlights are separated Mel Frequency Cepstral Coefficient (MFCC) and Mel Energy range Dynamic coefficients (MEDC).

Fig. 1 demonstrates the MFCC include extraction process. As appeared in Fig. 2 include extraction procedure contains following advances: • Preprocessing: The nonstop time signal (discourse) is inspected at examining recurrence. At the primary stage in MFCC highlight extraction is to help the measure of vitality in the high frequencies. This preemphasis is finished by utilizing a channel.

• Surrounding: it is a procedure of fragmenting the discourse tests acquired from the simple to computerized change (ADC), into the little casings with the time length inside the scope of 20-40 ms. Confining empowers the non-stationary discourse sign to be portioned into semi stationary edges, and empowers Fourier Transformation of the discourse signal. It is on the grounds that, discourse sign is known to show semi

stationary conduct inside the brief timeframe time of 20-40 ms.

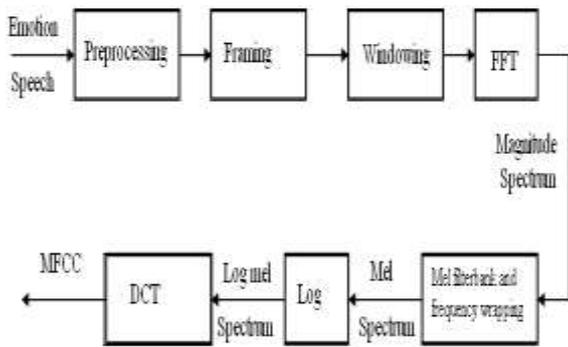


Figure 2. MFCC feature extraction

- Windowing: Windowing step is supposed to window every individual frame, that allows you to decrease the signal discontinuities at the start and the cease of every frame.
- FFT: fast Fourier transform (FFT) algorithm is ideally used for evaluating the frequency spectrum of speech. FFT converts each frame of N samples from the time area into the frequency domain.
- Mel Filterbank and Frequency wrapping: The mel filter financial institution [8] includes overlapping triangular filters with the cutoff frequencies decided by means of the middle frequencies of the two adjoining filters. The filters have linearly spaced centre frequencies and fixed bandwidth on the mel scale.
- Take Logarithm: The logarithm has the effect of converting multiplication into addition. therefore, this step clearly converts the multiplication of the magnitude within the Fourier remodel into addition
- Take Discrete Cosine transform: it's far used to orthogonalise the filter out energy vectors. because of this orthogonalization step, the statistics of the clear out power vector is compacted into the first number of additives and shortens the vector to variety of components. Another feature Mel energy spectrum Dynamic coefficients (MEDC) is likewise extracted. It is extracted as follows: the magnitude spectrum of each speech utterance is anticipated the usage of FFT, then input to a financial institution of 12 filters equally spaced on the Mel frequency scale. The logarithm mean energies of the filter out outputs are calculated  $E_n(i), i= 1.....N$ . Then, the primary and 2nd variations of  $E_n(i)$  are calculated. MEDC feature extraction method. The MEDC function extraction system includes following steps proven in figure three:

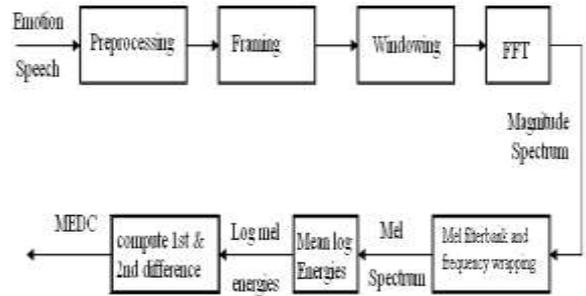


Figure 3. MEDC feature extraction

- Preprocessing, Framing, Windowing, FFT & Mel filterbank and Frequency wrapping techniques of MEDC feature extraction are same as MFCC feature extraction.
- Take logarithmic mean of energies: in this method a median log of each filter energies is calculated. This mean cost represent strength of man or woman filter out in a filterbank.
- Compute 1st and 2d distinction: The very last Mel electricity spectrum dynamics coefficients are then received with the aid of combining the primary and second variations of filter out energies.

### 3.2 Feature Labeling

In feature labeling each extracted feature is stored in a database along with its magnificence label. Although the SVM is binary classifier it may be also used for classifying more than one training. Each function is related to its magnificence label e.g. angry, glad, unhappy, impartial, fear.

### 3.3 SVM Classification

In general SVM is a binary classifier, but it could also be used as a multiclass classifier. LIBSVM [9], [10] is a maximum extensively used device for SVM category and regression evolved by means of C. J. Lin. Radial basis feature (RBF) kernel is utilized in education phase. Gain of the usage of RBF kernel is that it restricts schooling statistics to lie in specific barriers. The RBF kernel nonlinearly maps samples into a higher dimensional space, so it, not like the linear kernel, can handle the case whilst the relation between class labels and attributes is nonlinear. The RBF kernel has much less numerical problems than polynomial kernel.

## IV. SVM CLASSIFIER

The identity of emotion-related speech capabilities is extraordinarily difficult task. Guide Vector gadget is used as a classifier to classify unique emotional states which include anger, unhappiness, worry, happy, boredom. SVM is simple and green set of rules which has a excellent type performance as compared to other classifiers. SVM are the popular learning method for class, regression and different getting to know responsibilities. SVM has a higher classification performance on a small quantity of schooling samples. However we're lacking in recommendations on choosing

a better kernel with optimized parameters of SVM. There may be no uniform pattern used to the choice of SVM with its parameters and kernel feature with its parameters. The paper proposed strategies approximately choosing optimized parameters and kernel feature of SVM.

The system of the machine is as follows:

STEP1: Extracting speech emotion feature from utterances.

STEP2: the main mission in optimized procedure is to improve the type accuracy rate of the SVM.

STEP3: After optimizing method, the device trains an optimized version used to categories.

STEP4: The machine gives a category end result (elegance label or recognition price) approximately check samples.

The principal precept of SVM is to set up a hyperplane because the selection surface maximizing the margin of separation among negative and fantastic samples. For this reason SVM is designed for two class pattern type. More than one pattern classification problems can be solved the use of a combination of binary support vector machines.

**V. EXPERIMENTATION AND RESULTS**

**Emotion Classification:** in this project we recognize the temper of a consumer via their voice on the basis in their temper and classify into classifier:

- Anger
- Fear
- Disgust
- Happiness
- Sadness
- Neutral
- Boredam

Berlin Emotion database consists of 406 speech files for 5 emotion instructions. Emotion instructions Anger, sad, satisfied, neutral, worry are having 127, 62, seventy one, 79 and 67 speech utterance respectively. The LIBSVM is trained on MFCC and MEDC feature vectors the use of RBF and Polynomial kernel features. The LIBSVM is used to check those characteristic vectors. The experimentation is executed by way of varying value values for RBF kernel and degree values for Polynomial kernel. Each gender unbiased and gender dependent experiments are executed. Using RBF kernel at value fee c=4, it gives reputation charge of ninety three. Seventy five% for gender independent case, 94.seventy three% for male and a hundred% for girl speeches. The popularity charge the usage of Polynomial kernel at degree d=4 is 96.25% gender impartial, 97.36% for male and a hundred% for lady speeches. The Confusion matrices using RBF kernel gender unbiased, male and

woman are shown in table 1, 2 and three. Desk four, five and 6 indicates Confusion matrices using Polynomial kernel gender impartial, male and lady.

**Table 1. Confusion matrix of the RBF LIBSVM classifier (Gender Independent)**

Emotion	Emotion Recognition (%)				
	Angry	Sad	Happy	Neutral	Fear
Angry	100	0	0	0	0
Sad	0	100	0	0	0
Happy	0	0	100	0	0
Neutral	0	6.25	0	93.75	0
Fear	0	0	30.76	0	69.24

**Table 2. Confusion matrix of the RBF LIBSVM classifier (Male)**

Emotion	Emotion Recognition (%)				
	Angry	Sad	Happy	Neutral	Fear
Angry	100	0	0	0	0
Sad	0	100	0	0	0
Happy	16.66	0	83.34	0	0
Neutral	0	0	0	100	0
Fear	0	0	0	14.85	85.15

**Table 3. Confusion matrix of the RBF LIBSVM classifier (Female)**

Emotion	Emotion Recognition (%)				
	Angry	Sad	Happy	Neutral	Fear
Angry	100	0	0	0	0
Sad	0	100	0	0	0
Happy	0	0	100	0	0
Neutral	0	0	0	100	0
Fear	0	0	0	0	100

**Table 4. Confusion matrix of the Polynomial LIBSVM classifier (Gender Independent)**

Emotion	Emotion Recognition (%)				
	Angry	Sad	Happy	Neutral	Fear
Angry	100	0	0	0	0
Sad	0	100	0	0	0
Happy	0	0	100	0	0
Neutral	0	0	0	100	0
Fear	7.69	0	15.18	0	76.92

**Table5. Confusion matrix of the Polynomial LIBSVM classifier (Male)**

Emotion	Emotion Recognition (%)				
	Angry	Sad	Happy	Neutral	Fear
Angry	100	0	0	0	0
Sad	0	100	0	0	0
Happy	0	0	100	0	0
Neutral	0	0	0	100	0
Fear	0	0	14.28	0	85.72

**Table 6. Confusion matrix of Polynomial LIBSVM classifier (Female)**

Emotion	Emotion Recognition (%)				
	Angry	Sad	Happy	Neutral	Fear
Angry	100	0	0	0	0
Sad	0	100	0	0	0
Happy	0	0	100	0	0
Neutral	0	0	0	100	0
Fear	0	0	0	0	100

**VI. CONCLUSION**

As generation evolves, hobby in human like machines will increase. Technological gadgets are spreading and user delight increases significance. An herbal interface which responds in keeping with user wishes has emerged as feasible with affective computing. The important thing problem of affective computing is feelings. Any studies that's related with detection, popularity or generating an emotion is affective computing. Person satisfaction or un-pleasure might be detected with any emotion popularity gadget. Besides detection of person satisfaction, such systems will be used to discover anger or frustration. In such cases, user will be constrained like riding a automobile. In emotion detection responsibilities, speech or face emotion detections are the maximum popular ones. smooth get right of entry to to stand or speech facts made them very famous. Speech includes a rich set of facts. In human to human conversation, via speech facts is conveyed. Acoustic a part of speech includes important info approximately feelings. MFCC are used for the feature extraction .algorithm with the SVM's ordinary performance is examined.

**REFERENCE:**

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

[2] Shikha Gupta<sup>1</sup>, Jafreezal Jaafar<sup>2</sup>, Wan Fatimah wan Ahmad<sup>3</sup> and Arpit Bansal<sup>4</sup> J. Representative Maxwell, "Highlight extraction utilizing mfcc" Signal and Image Processing : An International Journal (SIPIJ) Vol.4, No.4, August 2013

[3] N.Murali Krishna<sup>1</sup>,P.V. Lakshmi<sup>2</sup>,Y. Srinivas<sup>3</sup> J.Sirisha Devi<sup>4</sup>," Emotion Recognition utilizing 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 and SVM Algorithm International Journal of Advanced Research in Computer

Science and Software Engineering Research Paper Volume 3, Issue 8, August 2013.

[5] Eslam Mansour mohammed<sup>1</sup>, Mohammed Sharaf Sayed<sup>2</sup>," LPC and MFCC Performance Evaluation with Artificial Neural Network for Spoken Language Identification ," International Journal of Signal Processing, Image Processing and Pattern Recognition Vol. 6, No. 3, June,2013

[6] Cowie, R., Douglas-Cowie, E., Tsapatsoulis, N., Votsis, G., Kollias, S., Fellenz, W., and Taylor, J. G., Emotion acknowledgment in human-PC connection, IEEE Signal Processing magazine, Vol. 18, No. 1, 32-80, Jan. 2001. [2] D. Ververidis, and C. Kotropoulos, Automatic discourse characterization to five passionate states dependent on sexual orientation data, Proceedings of the EUSIPCO2004 Conference, Austria, 341-344, Sept. 2004.

[7] Christopher. J. C. Burges, An instructional exercise on help vector machines for example acknowledgment, DataMining and Knowledge Discovery, 2(2):955-974, Kluwer Academic Publishers, Boston, 1998.

[8] T. D., and Narayanan, S. S. (2010), "Acoustic element examination in discourse feeling natives estimation" Interspeech .

[9]. Programmed acknowledgment of discourse feeling utilizing long haul spectro-fleeting highlights. Advanced Signal Processing, 1-6.

[10]. C J., and Tao, J. (1997), "Beginning with susas: a discourse under mimicked and real pressure database". Euro Speech, 4, 1743-1746.

[11]. Kamran Soltani and Raja Noor Ainon (2002), "Discourse feeling location dependent on neural systems" In ninth International Symposium on Signal Processing and its Applications, 1 4244-0779-6/07, IEEE, 2007.8

[12]. Thapanee Seehapoch and Sartra wongthanavasu (2013), "Discourse Emotion Recognition utilizing Support Vector Machine".

[13]. E. Dellandrea and Dou (2005) "Highlights extraction and choice for enthusiastic discourse grouping". IEEE Conference on Advanced Video and Signal Based Surveillance (AVSS), pp.411-416, Sept 2005.

[14]. T.- L. Pao et al; (2006), "Mandarin passionate discourse acknowledgment dependent on SVM and NN", Proceedings of the eighteenth International Conference on Pattern Recognition (ICPR'06), vol. 1, pp. 1096-1100, September 2006.