IOSR Journal of Computer Engineering (IOSR-JCE)
e-ISSN: 2278-0661,p-ISSN: 2278-8727
PP 38-44
www.iosrjournals.org

Study of Affect Computing Techniques in Human Computer Interface

Dhananjay Theckedath, Dr. R.R Sedamkar
[1]*Assistant Professor, Thadomal Shahani College of engineering*
[2]*Academic &Professor, Department of Computer Engineering, Thakur college of Engineering.Mumbai, India*

Abstract: *Computer systems which are capable of emotion detection via automated devices open up new horizons in Human-Computer Interaction (HCI). Due to the inseparable link between emotions and cognition, the field of Affective Computing aspires to narrow the communicative gap between human beings and computers. This is achieved by developing systems that not only recognize but also respond to the affective states of the user. This paper presents a literature survey on the various Affect detection techniques including Facial expression analysis, Speech analysis, Multi-modality and Gaze detection. The aim of the paper is to obtain a deeper understanding of affective states, which will form the foundation for designing a data-driven Human Computer interface.*

I. Introduction

Computer systems which are capable of emotion detection by detecting non-verbal cues via automated devices open up new horizons in Human-Computer Interaction (HCI). Although these systems are still far from achieving the capacity of human perception, they are able to classify and assess user emotions through predetermined mathematical models with limited human intervention. The growth in HCI field has primarily been in the quality of interaction. The thrust in research has made new technology to become available to everyone in no time. For HCI to be successful, the design should consider many aspects of human behaviour. The intricacy of the involvement of a human in interaction with a machine is sometimes invisible compared to the ease of the interaction method itself [1]. Due to this, the degree of user activity with a machine should be thoroughly thought whie designing HCI systems. The user activity has three different levels of interaction viz. physical and cognitive. The physical aspect is related to the mechanical interaction between human and compute while the cognitive aspect deals with way the user understands the system and interacts with it. HCI also needs to be an Intelligent interaction and for that to happen identifying human expressions is the key [2].

Visual based human computer interaction is possibly the most widespread area in HCI research. Researchers are trying to tackle a variety of open ended problems considering the extent of applications where it could be used. Some of the main research areas in affect computing for Human Computer Interactions are as follows:

Facial Expression Analysis
1. Speech and non-speech audio
2. Multi modality of visual and audio
3. Gaze Detection (Eyes Movement Tracking)

Human-computer intelligent interaction (HCII) is an emerging field of science which is aimed at providing natural ways for humans to interact with computers. While basic emotions can be detected fairly accurately, it is still unclear if affective states viz. confusion, frustration, boredom etc can be detected with the same fidelity. For basic emotions the links between emotion and expression have been carefully mapped. Similar mapping for learning centered affective states is largely missing and is an open question. The relation of the automatic face analysis process to the other parts of the typical HCI system is shown in Fig 1 There may be other components such as speech and gaze tracking components that communicate with the face analysis component, so that the application receives multimodal feedback from them [3].

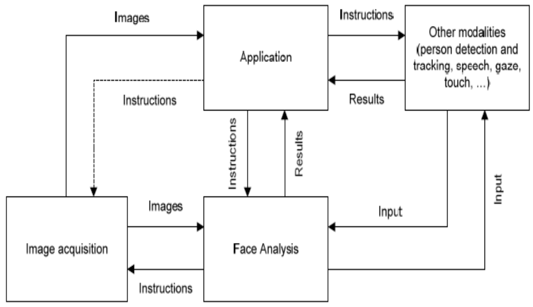

Fig 1. Block diagram of a HCI system with Multimodal feedback

One of the areas of Human Computer Interface (HCI) in which progress is being made is in a virtual classroom environment where learning typically means using a computer to deliver part, or all of a course [1]. In such a scenario, it becomes imperative to evaluate the extent to which the student has learnt and also find out ways to improve the teaching learning process. Affect detection which is centered around learning is a major component in developing educational interfaces that are capable of responding to the learning needs of students [4].

II. Literature Survey

2.1 Facial Expression Analysis

The expression of emotion is achieved through a complex combination of information produced from the body and the brain. It has been widely believed that facial expressions are indicative of mental states. However the scientific community has yet to measure and quantify the relationship between facial expressions and emotions. There have been several attempts to quantify this relationship. One such method uses objective coding schemes based on visible units of facial behavior. This is usually done by analyzing videos, frame-by-frame which are recorded during an experiment. Each frame is analyzed in a detailed manner by human coders who systematically follow descriptive rules of judgment. Systematic measurements were created by Ekman and Friesen [5], which has proved to be the bench mark for subsequent studies in Expression analysis. They introduced a new system known as the Facial Action Coding System (FACS). The Facial Action Coding System proposed by them partitions the visible effects of facial muscle activation into "Action units" (AU). In this system, each action unit is related to one or more facial muscles. The Facial Action Coding System (FACS) is a broad, anatomically based system which is used for measuring nearly all noticeable facial movements. The latest FACS system describes facial activity based on 44 unique Action Units. The FACS taxonomy is based on manually observing gray level variation between expressions in images. This as mentioned is performed by human coders who systematically follow vivid rules of judgment. Depending on which edition of FACS is used, there are 30 to 44 AUs and additional action descriptors. More than 7000 AU combinations have been observed. Because of its descriptive power, FACS has now become the gold standard for facial measurement in expression analysis. The challenge is to develop systems which automatically detect and classify learning centered affect states.

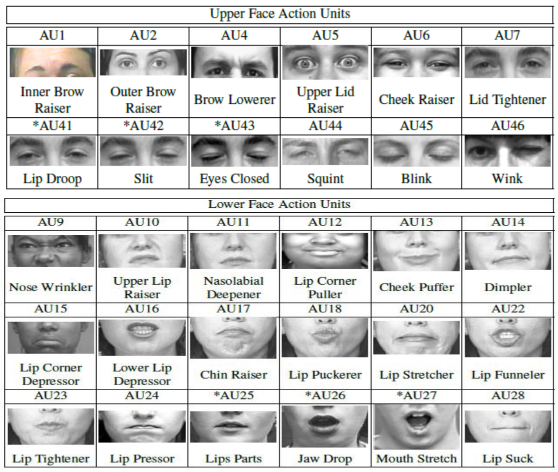

Fig. 2. Facial Action Units as proposed by Ekman and Friesen

While there has been some work carried out in detecting six basic emotions which are Happy, Sad, Surprise, Anger, Fear and Disgust , there is very little research reported on detection and classification of learning centered affective states. Affect states pertaining to learning are known to be different than the standard basic emotions. Learning centered affective states include Boredom, Confusion, Delight, Engagement and Frustration. Since thousands of anatomically possible facial expressions can be represented as combination of a few dozens of Action Units, these are ideal to describe the complexity of spontaneous facial behavior. It is however a matter of investigation whether these affective states can be detected with the same fidelity as the basic emotions, where the links between emotion and expression have been carefully mapped. Similar mapping for learning centered affective states is largely missing and is an open for further research. Due to the highly skewed class distribution of learning centered affective states, designing classifiers is a challenging task and the accuracy for detecting learning centered affect also varies widely.

Table 1: Facial Features and Classifiers Used

References	Facial Feature Detector	Classifier
Nigel Bosch [6]	FACET	C4.5, Bayes NET
Md. Eshan Hoque et al [7]	Google's facial tracker, Shore	Adaboost, Random forest
Shan et al[8]	Local Binary pattern	SVM, Adaboost
Ramkumar [9]	DNN	Naive Bayes, SVM and Decision Tree classifiers
Barlet et al [10]	Gabor wavelet	SVM+HMM,
Zia Uddin [11]	Gabor wavelet	Adaboost SVM
Petar S. Aleksic [12]	12 motion unit	Tree augmented DBN
Cohen et al [13]	Shape models, Gabor wavelet	LDC
Fasel et al [14]	Grey level intensity	NN
Gunes and Piccardi [15]	Shape features, optical flow	BayesNet
Valstar et al [16]	8 facial points	Gentle boost, SVM
Francesc Tarrés et al [17]	20 facial points	Gentle boost, SVM

Although face analysis is the main focus in HCI, other modalities are equally important and it is advantageous to use several modalities in parallel.

2.2 . Speech and Non-Speech Audio

Sound is an important input and output modality for humans. Speech and vision are the main communication channels when humans communicate face to face. Affective stimulation modulates all human communicative signals. Psychologists and linguists have identified the importance of audio cues in human affect [18]. Hence analyzing speech has a profound effect of Human computer interaction. While the attention of the other person and most of the expressions are perceived through vision, most of the explicit information and a part of the emotional information and expressions are conveyed through speech. The speaker's affective state can be inferred from the surface features of words. Speech recognition technology uses algorithms which are similar to the ones used for machine learning and pattern recognition [19]. For example, Hidden Markov models (HMMs) that are commonly used in both, speech recognition and pattern recognition. Other common techniques are neural networks and Linear Discriminant analysis (LDA). Automatic vocal affect recognition systems are trained and tested using speech data that are collected by asking trained actors to speak prescribed words with certain emotions. Most of the existing approaches to vocal affect recognition use acoustic features as classification input based on the acoustic correlation for emotion expressions. The popular features that are identified are Prosodic features and Spectral features. Prosodic features include pitch-related feature, energy-related features, and the speech rate. Similarly Spectral features include MFCC and cepstral features. There have been studies to show that pitch and energy are among those features which contribute the most to affect recognition. There has been a research shift towards the analysis of spontaneous human behavior. Classifiers such as LDA and SVM are commonly used. Table 2 provides an overview of some of the currently existing standard systems for audio-based affect recognition with respect to the utilized auditory features, classifier, and performance.

Table 2: Speech Features And Classifiers Used

References	Feature Detector	Classifier
Ang et al [20]	Prosody, LM features, position	Decision Tree
Austermann et al [21]	Prosody	Fuzzy rules
Batliner et al [22]	Prosody, POS, DA	MLP, LDA
Devilleres et al [23]	Lexical cues., prosody, spectrum, disfluency	SVM
Kwon et al [24]	Prosody, MFCC	QDA,SVM,LDA
Liscombe et al [25]	Acoustic prosodic	C 4.5 with Adaboost

2.3. Audio Visual and Affect Recognition

There is a also a renewed interest in multimodal interaction. It is often beneficial to use several modalities and channels in parallel. It is proposed that a combination of information obtained from facial expressions and speech could lead to robust affect recognition systems. Since then, a number of efforts are reported toward this direction[26]. Most of them apply the FeelTrace system which allows raters to continuously label the change of affective expressions. It was noticed that there was considerable labeling variation in the results obtained from the raters using FeelTrace due to subjectivity of audio-visual affect judgment. It was observed that some observers mainly relied on audio information to make their judgment while others relied on visual information. One way to reduce this variation is to assume that both, the facial expression as well as the vocal expression have the same coarse emotional states-positive and negative. Once this is done, FACS-based labels of facial expressions could be used as audio-visual expression labels as well.

Table 3: Multimodal Features And Classifiers Used

References	Feature Detector	Classifier
Busso et al [27]	102 markers, Prosody,	SVM
Go et al [28]	Eigenfaces, MFCC	LDA
Hoch et al [29]	Gabor features,prosody	SVM
Schuller et al [30]	AAM, prosody	SVM
Song et al [31]	54 FAP's prosody	THMM

2.4 . Gaze Detection

Eye-gaze is an input which has the potential of increasing the efficiency of the human interface system. In this system, the movement of the user's eyes provides a convenient and a natural source of input. By tracking the direction of gaze of the user, information about what the user is looking at can be obtained along with being able to partly identify the users interest level and hence the affective states[32]. Gaze estimation with head movement is a very challenging task. A simplified block diagram of a Gaze detection system is shown in Fig 3. The two standard algorithms available in literature are the: Longest Line Scanning and Occluded Circular Edge Matching algorithms. In both these algorithms, the focus is on estimating the orientation of the eyes with slight head movement.

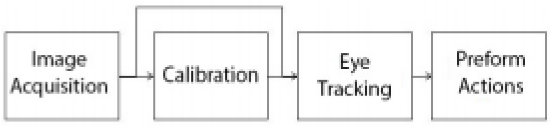

Fig. 3. Gaze detection Block diagram

The eye tracking block consists of detection of the pupil along each frame of the image sequence. Because of the calibration requirements, pupil detection is one of the most important steps. Hough circle transform is used to identify the iris and also the pupil. Once the pupil position is identified, it is possible to determine the exact point on the image screen where the user is looking. Gaze tracking involves the use of Haar Cascade classifier and Kalman filter.

III. Research Opportunities

A total of 7 papers were analysed in detail. The objective was to go through some of the recent works published in the domain of extracting affect states from facial expressions and understand the areas were further research could be carried out. Literature review was broadly classified into two groups viz. Studying papers involving various algorithms to extract features from video and papers which discuss various applications of affective computing.

TABLE 4: GAP ANALYSIS

PAPER	OBSERVATIONS
Nigel Bosch [6]	This paper deals with learning centric affect states and uses the SMOTE oversampling method for addressing the skewed class distribution. Registration of affect states is possible only for 76% of the time in the wild
Caifeng Shan et al [8]	In this paper, images are manually cropped and the LBP operators are applied on them. They have not provided a method of automatically identifying a face from the frame and cropping it

Ramkumar[9]	A theory driven approach to predict frustration of student working with an ITS is proposed to understand the causes of frustration. The results are compared to existing approaches. The paper deals only with one affective state which is frustration. Other effect states are not discussed
Zia Uddin [11]	Experiments are performed in a controlled environment with uniform lightning effects. The video obtained has a camera which focuses only on a single person and not on a group of people. Success of this method on low resolution images has not been stated
Petar S. Aleksic, [12]	The multi stream HMM facial expression system propsed utilizes stream reliability weights which achieves relative reduction of the facial expression recognition error of 44% compared to the single-stream HMM system.
Francesc Tarrés et al [17]	The main drawback of the new algorithm is the computational cost which is 6 times higher than the Eigenfaces algorithm. In the paper, it has been concluded that the main false recognition decisions were due to the strong illumination variations;
M.Yeasin etall [33]	The signatures computed from the training data set are used to train discrete hidden Markov models (HMMs) . This aids the system to learn the underlying model for each facial expression. Database consisting of 488 video sequences are used which includes 97 subjects are.

2 . Database

Having a exhaustive database with labeled data of human affective expressions is a precondition in designing automatic affect recognizing systems. Authentic affective expressions are relatively rare and short lived. They are also filled with subtle context-based changes which make it very difficult to collect. Due to the low probability of occurrence of these expressions, manual labeling of these spontaneous emotional expressions is very time consuming and also error prone. This state of affairs makes the automatic analysis of spontaneous emotional expression a very difficult task [34]. There have been efforts in the past to collect spontaneous sequences of basic and natural emotions while the participants were acting or interacting. There are a few databases which contain deliberate affective behavior. The Cohn-Kanade facial expression database is the most widely used database for facial expression recognition. The BU-3DFE database of Yin and colleagues consists of 3-Dimensional range data of six standard facial expressions which are displayed at four different levels of intensity. There is another database known as the FABO database which consist of videos of facial expressions as well as body gestures portraying posed displays of basic as well as non basic affective states. Datasets such as RU-FACS, SAL, Spaghetti, SEMAINE, MindReading etc include spontaneous affective behavior along with acted behavior. Apart from these there is the MMI facial expression database [35] which is considered to be the most comprehensive data set of facial behavior recordings to date. The MMI database contains both posed as well as spontaneous expressions of facial behavior. The recordings of deliberate facial behavior are both static images and videos. A large part of video recordings are recorded in both the frontal and the profile views of the face. The database is freely available and downloadable via the Internet. There is a simple registration process required to access the database.

Figure 2 is a graphical representation of each dataset in terms of whether it is acted or spontaneous and whether it contains basic as well as beyond basic emotions. Ideally, a dataset should contain spontaneous natural emotion for affect analysis [36].

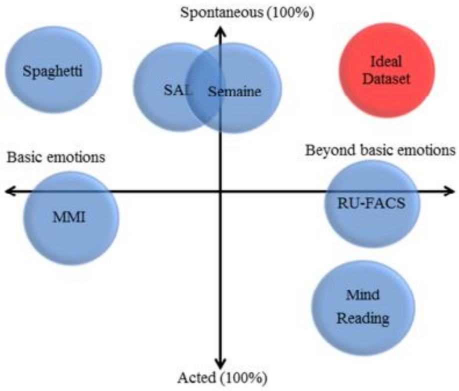

Fig. 4. Comparison of existing databases

IV. Conclusion

To achieve an effective human-computer intelligent interaction, there is a need for the computer to be able to interact with the user naturally, similar to the way human-human interaction takes place. While there has been some work carried out in detecting six basic emotions which are Happy, Sad, Surprise, Anger, Fear and Disgust , there is very little research reported on detection and classification of affective states. Affect states pertaining to learning are known to be different than the standard basic emotions. Learning centric affective states include Boredom, Confusion, Delight, Engagement and Frustration. Accuracy for learning centered affect detectors also varies widely in terms of percentage of instances correctly classified. Detecting learning centred affective states with high accuracy is still an open research problem. The goal of research would be to detect and classify learning centred affect states

References

[1] Gong Chao ; Beijing Inst. of Technol., Beijing, China. "Human-computer interaction: The usability test methods and design principles in the human-computer interface design." Aug. 2009 283 – 285.

[2] D. Te'eni, J. Carey and P. Zhang, *Human Computer Interaction: Developing Effective Organizational Information Systems*, John Wiley & Sons, Hoboken (2007).

[3] [3] Erno Mäkinen , *Face Analysis Techniques for Human-Computer Interaction*, , Academic Dissertation, Department of Computer Sciences, University of Tampere, 2007

[4] Sangeeta Kakoty, Monohar Lal, Shikhar Kr. Sarma "E-learning as a Research Area: An Analytical Approach", *International Journal of Advanced Computer Science and Applications*, Vol. 2, No. 9, 2011.

[5] Ekman, P., Friesen, W.: "*Facial action coding system: A technique for the measurement of facial movement*". Consulting Psychologists Press, Palo Alto (1978)

[6] Nigel Bosch, Sidney D'Mello Ryan Baker, et all, "Automatic Detection of Learning-Centered Affective States in the Wild" Proceeding IUI '15 Proceedings of the 20th International Conference on Intelligent User Interfaces Pages 379-388

[7] Md. Eshan Hoque "Exploring Temporal Patterns in Classifying Frustrated and Delighted Smile" *IEEE transactions onAffective omputing,vol 3, No. 3, Sept 2012*

[8] Caifeng Shan, Shaogang Gong , Peter W. McOwan ,"Facial expression recognition based on Local Binary Patterns: A comprehensive study" *ELSVIER Image and Vision Computing* 27 (2009) 803–816

[9] Ramkumar, Sridhar Iyer et all. "A theory driven approach to predict frustration in an ITS" , *IEEE transactions on learning technologies*; 6, 4; 378-388; 2013

[10] M.S. Bartlett, G. Littlewort, P. Braathen, T.J. Sejnowski, and J.R. Movellan, "A Prototype for Automatic Recognition of Spontaneous Facial Actions," *Advances in Neural Information Processing Systems*, vol. 15, pp. 1271-1278, 2003.

[11] Md. Zia Uddin, "A depth video-based facial expression recognition system utilizing generalized local directional deviation-based binary pattern feature discriminant analysis", SPRINGER APRIL 2015

[12] Petar S. Aleksic "Automatic Facial Expression Recognition using facial animation parameters and mulitstream HMMS" *IEEE Transactions on Information Forensics and Security* (Volume: 1, Issue: 1, March 2006).

[13] J.F. Cohn, L.I. Reed, Z. Ambadar, J. Xiao, and T. Moriyama, " Automatic Analysis and Recognition of Brow Actions and Head Motion in Spontaneous Facial Behavior," *Proc. IEEE Int'l Conf.Systems, Man, and Cybernetics (SMC '04),* vol. 1, pp. 610-616, 2004.

[14] B. Fasel, F. Monay, and D. Gatica-Perez, "Latent Semantic Analysis of Facial Action Codes for Automatic Facial Expression Recognition," *Proc. Sixth ACM Int'l Workshop Multimedia Information Retrieval (MIR '04),* pp. 181-188, 2004.

[15] H. Gunes and M. Piccardi, "Affect Recognition from Face and Body: Early Fusion versus Late Fusion," *Proc. IEEE Int'l Conf. Systems, Man, and Cybernetics (SMC '05),* pp. 3437-3443, 2005.

[16] M. Valstar, M. Pantic, Z. Ambadar, and J.F. Cohn, "Spontaneous versus Posed Facial Behavior: Automatic Analysis of Brow Actions," *Proc. Eight Int'l Conf. Multimodal Interfaces (ICMI '06),* pp. 162-170, 2006.

[17] Francesc Tarrés, Antonio Rama "A Novel Method for Face Recognition under partial occlusion or facial expression Variations". Presented at 47th International Symposium , ELMAR, 2005

[18] J. Hirschberg, S. Benus, J.M. Brenier, F. Enos, and S. Friedman, " Distinguishing Deceptive from Non-Deceptive Speech," Proc. Ninth European Conf. Speech Comm. and Technology (INTERSPEECH 05), pp. 1833-1836, 2005.

[19] H.J. Go, K.C. Kwak, D.J. Lee, and M.G. Chun, "Emotion Recognition from Facial Image and Speech Signal," Proc. *Int'l* Conf. Soc. of Instrument and Control Engineers, pp. 2890-2895,2003.

[20] J. Ang et al., "Prosody-Based Automatic Detection of Annoyance and Frustration in Human-Computer Dialog," Proc. Eighth Int'l Conf. Spoken Language Processing (ICSLP), 2002.

[21] A. Austermann, N. Esau, L. Kleinjohann, and B. Kleinjohann, "Prosody-Based Emotion Recognition for MEXI," Proc. *IEEE/RSJ* Int'l Conf. Intelligent Robots and Systems (IROS '05), pp. 1138-1144,2005.

[22] A. Batliner, K. Fischer, R. Hubera, J. Spilkera, and E. Noth, "How To Find Trouble in Communication," Speech Comm., vol. 40, pp. 117-143, 2003.

[23] L. Devillers and I. Vasilescu, "Real-Life Emotions Detection with Lexical and Paralinguistic Cues on Human-Human Call Center Dialogs," Proc. Ninth Int'l Conf. Spoken Language Processing (ICSLP), 2006.

[24] O.W. Kwon, K. Chan, J. Hao, and T.W. Lee, "Emotion Recognition by Speech Signals," Proc. Eighth European Conf. Speech Comm. And Technology (EUROSPEECH), 2003.

[25] J. Liscombe, J. Hirschberg, and J.J. Venditti, "Detecting Certainness in Spoken Tutorial Dialogues*,"* Proc. Ninth European Conf. Speech Comm. and Technology (INTERSPEECH), 2005.

[26] Zhihong Zeng, Maja Pantic, Glenn I. Roisman, and Thomas S. Huang, "A Survey of Affect Recognition Methods: Audio,Visual, and Spontaneous Expressions", *IEEE Transactions On Pattern Analysis And Machine Intelligence,* Vol. 31, No. 1, January 2009

[27] C. Busso et al., "Analysis of Emotion Recognition Using Facial Expressions, Speech and Multimodal Information," Proc. Sixth ACM Int'l Conf. Multimodal Interfaces *(ICMI '04)*, pp. 205-211, 2004.

[28] H.J. Go, K.C. Kwak, D.J. Lee, and M.G. Chun, "Emotion Recognition from Facial Image and Speech Signal," Proc. Int'l Conf. Soc. of Instrument and Control Engineers, pp. 2890-2895, 2003.

[29] S. Hoch, F. Althoff, G. McGlaun, and G. Rigoll, "Bimodal Fusion of Emotional Data in an Automotive Environment," Proc. 30th Int'l Conf. Acoustics, Speech, and Signal Processing (ICASSP '05*)*, vol. II, pp. 1085-1088, 2005.

[30] B. Schuller, R. Muller, B. Hornler, A. Hothker, H. Konosu, and G. Rigoll, "Audiovisual Recognition of Spontaneous Interest within Conversations," Proc. Ninth ACM Int'l Conf. Multimodal Interfaces*(ICMI '07),* pp. 30-37, 2007.

[31] M. Song, J. Bu, C. Chen, and N. Li, "Audio-Visual-Based Emotion Recognition: A New Approach," Proc. Int'l Conf. Computer Vision and Pattern Recognition (CVPR '04), pp. 1020-1025, 2004.

[32] Rafael Santos, Nuno Santos, Pedro M. Jorge, Arnaldo Abrantes, "Eye Gaze As A Human-Computer Interface", Conference on Electronics, Telecommunications and Computers – *CETC 2013*, Available online at

[33] Mohammed Yeasin etall "Recognition of Facial Expressions and Measurement of Levels of Interest From Video". *IEEE Transactions on Multimedia* Volume:8 , Issue: 32006

[34] Zhihong Zeng, Maja Pantic, Glenn I. Roisman, and Thomas S. Huang,"A Survey of Affect Recognition Methods: Audio, Visual, and Spontaneous Expressions", *IEEE Transactions On Pattern Analysis And Machine Intelligence*, VOL. 31, NO. 1, Jan 2009

[35] Michel F. Valstar, Maja Pantic, "Induced Disgust, Happiness and Surprise: an Addition to the MMI Facial Expression Database", Proc. 3rd Intern. Workshop on EMOTION 2010

[36] Md. Eshan Hoque "Exploring Temporal Patterns in Classifying Frustrated and Delighted Smile" *IEEE transactions on Affective omputing,vol 3, No. 3, Sept 2012*