Emotion Recognition based on signal processing

Shreekant Vishwas Marwadi
Graduate Student – CS5760
Department of Electrical and Computer Engineering
Michigan Technological University
Houghton, Michigan
Email: svmarwad@mtu.edu

***Abstract*- Human emotions are the most crucial feature of human interaction which defines the real state of human which grant as a natural way of interaction. Recognizing the user's emotional state will improve the interface between human and computers. This would make life easier to convey the information to computers in a more effective way. This system can be used in fields such as education, automation, medical etc. We can classify this system as user dependent and independent emotional recognition systems. In the implementation of this systems, researchers have achieved maximum classification rate in user dependent system but a lower rate for user independent. The efficient emotion-recognition method needs a large number of data samples and advance signal processing techniques to improve the accuracy of user independent emotional recognition system. In this paper, I have reviewed current research and its challenges on emotional recognition using signal processing techniques.**

***Index Terms*-** Emotions, Human Computer Interaction, Signal Processing.

I. INTRODUCTION

As computers are becoming smarter by every passing day by day and their involvement increasing in human life. Expressing emotions is a most natural way of human to interact with each other. There for using emotions to interact with a computer will improve Human Computer Interaction. Emotions play important role in understanding the feelings of human and meaning of spoken words, as the meaning of same spoken word can be different depending on facial expression. The ability of computers to understand this emotional states of the user and respond with appropriate actions is one of the major research areas in Human-Computer Interaction (HCI). This empowers computers and ease human-computer interaction and give more meaning to it. For example When you are using computer to watch online lectures or any other important kinds of stuff and you felt sleepy and your computer wakes you with alarm or in another case when you felt sleepy while watching some entertaining stuff (not so important) then instead of waking you the computer go into sleep mode to save power this will really give different meaning to human-computer interaction.

Human emotions are generally expressed in different modalities and combinations of them. Missing of one of the modality can change the whole meaning of expressed emotion. For example, hidden expression of anger behind fake smiling face can sometimes even confuse human. We can classify human emotions in different categories like acted emotions, spontaneous emotions etc. these emotions also depend on genders, age group, cultural diversity etc. Because of this vast field of different emotions many times even to human misinterpret emotion of another human. There for it is very difficult but important to create such computer system which understands human emotions accurately.

Many uni-model emotion recognition systems are implemented using different features like speech, gesture, image processing (face image) etc. these features are easy to implement using feature extractions there for these are popular. But these features depends on gender, age and cultural diversity there for these are not universal and lacks recognition accuracy. Lightning conditions, noisy sound, a low-resolution camera, accessories on the face like goggles and caps harden the real-time implementation of these features. These uni-model methods are successfully implemented in other areas for face recognition and speaker recognition etc. therefor by making effective research, these methods can be improved. Using Bi-modal/ Multimodal emotion recognition systems based on a combination of uni-model recognition systems can improve the results but it will also increase the complexity of the system.

In this paper, I present a review of current research in emotion detection using signal processing in specifying to speech and image signal processing. The role of this review is to describe the theories of emotion detection based its background to look into possible advancement in future.

II. BASIC EMOTIONS

Emotion recognition was first studied in 70's during this period P. Ekam found that happy, disgust, anger, fear, sad and surprise are universally accepted emotions and does not depend on cultural diversity [1]. Later in 1999, he added amusement, contempt, contentment, embarrassment, excitement, guilt, pride and achievement, relief, satisfaction, sensory pleasure and shame into basic emotions. In 2003 Plutchik defines eight basic emotions which are anger, fear, sadness, disgust, surprise, anticipation, acceptance and joy and he stated that all other emotions can be formed by these basic emotions.

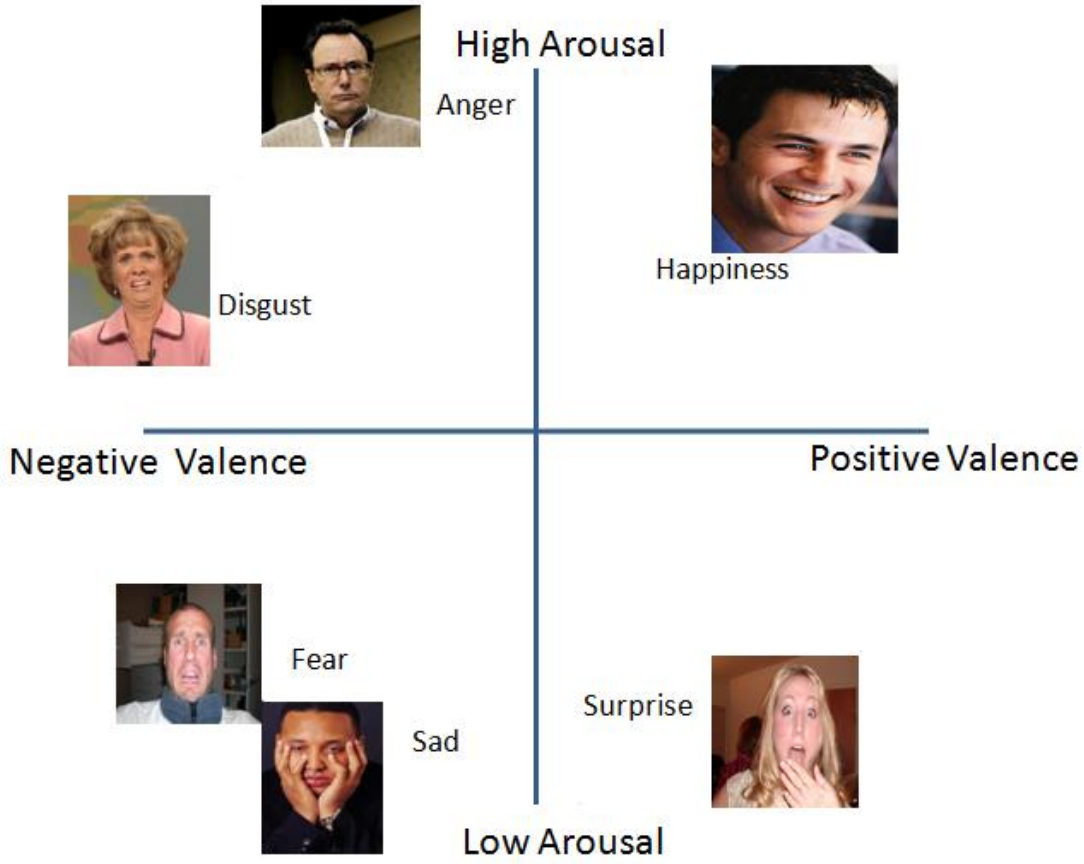

Fig.1 Basic emotions on valence-arousal 2-D model [2]

III. EMOTIONS RECOGNITION

Speech Emotion recognition

Speech is a most significant way of communication in human. To implement machine to understand emotions in a speech the machine should be able to recognize the human voice. The research was done on speech recognition techniques and effective models are developed using techniques like Mel-frequency spectrum coefficient (MFCC). But however, we are still away from the natural interface between human and computers as it efficiently recognizes the human voice but does not understand the emotional state behind it.

M. El Ayadi has given a peer review of on database, features and classification schemes of speech emotion recognition. One of the important issue speech emotion recognition is database availability. Most of the database are private and have limitations for assessing the performance of emotion recognition.[3]

The characteristics of common emotional speech database are given bellow:

Characteristics of common emotional speech databases.

Corpus	Access	Language	Size	Source	Emotions
LDC Emotional Prosody Speech and Transcripts [78]	Commercially available[a]	English	7 actors × 15 emotions × 10 utterances	Professional actors	Neutral, panic, anxiety, hot anger, cold anger, despair, sadness, elation, joy, interest, boredom, shame, pride, contempt
Berlin emotional database [18]	Public and free[b]	German	800 utterances (10 actors × 7 emotions × 10 utterances + some second version) = 800 utterances	Professional actors	Anger, joy, sadness, fear, disgust, boredom, neutral
Danish emotional database [38]	Public with license fee[c]	Danish	4 actors × 5 emotions (2 words + 9 sentences + 2 passages)	Nonprofessional actors	Anger, joy, sadness, surprise, neutral
Natural [91]	Private	Mandarin	388 utterances, 11 speakers, 2 emotions	Call centers	Anger, neutral
ESMBS [94]	Private	Mandarin	720 utterances, 12 speakers, 6 emotions	Nonprofessional actors	Anger, joy, sadness, disgust, fear, surprise
INTERFACE [54]	Commercially available[d]	English, Slovenian, Spanish, French	English (186 utterances), Slovenian (190 utterances), Spanish (184 utterances), French (175 utterances)	Actors	Anger, disgust, fear, joy, surprise, sadness, slow neutral, fast neutral
KISMET [15]	Private	American English	1002 utterances, 3 female speakers, 5 emotions	Nonprofessional actors	Approval, attention, prohibition, soothing, neutral
BabyEars [120]	Private	English	509 utterances, 12 actors (6 males + 6 females), 3 emotions	Mothers and fathers	Approval, attention, prohibition
SUSAS [140]	Public with license fee[e]	English	16,000 utterances, 32 actors (13 females + 19 males)	Speech under simulated and actual stress	Four stress styles: Simulated Stress, Calibrated Workload Tracking Task, Acquisition and Compensatory Tracking Task, Amusement Park Roller-Coaster, Helicopter Cockpit Recordings
MPEG-4 [114]	Private	English	2440 utterances, 35 speakers	U.S. American movies	Joy, anger, disgust, fear, sadness, surprise, neutral
Beihang University [43]	Private	Mandarin	7 actors × 5 emotions × 20 utterances	Nonprofessional actors	Anger, joy, sadness, disgust, surprise
FERMUS III [112]	Public with license fee[f]	German, English	2829 utterances, 7 emotions, 13 actors	Automotive environment	Anger, disgust, joy, neutral, sadness, surprise
KES [65]	Private	Korean	5400 utterances, 10 actors	Nonprofessional actors	Neutral, joy, sadness, anger
CLDC [146]	Private	Chinese	1200 utterances, 4 actors	Nonprofessional actors	Joy, anger, surprise, fear, neutral, sadness
Hao Hu et al. [56]	Private	Chinese	8 actors × 5 emotions × 40 utterances	Nonprofessional actors	Anger, fear, joy, sadness, neutral
Amir et al. [2]	Private	Hebrew	60 Hebrew and 1 Russian actors	Nonprofessional actors	Anger, disgust, fear, joy, neutral, sadness
Pereira [55]	Private	English	2 actors × 5 emotions × 8 utterances	Nonprofessional actors	Hot anger, cold anger, joy, neutral, sadness

[a] Linguistic Data Consortium, University of Pennsylvania, USA.
[b] Institute for Speech and Communication, Department of Communication Science, the Technical University, Germany.
[c] Department of Electronic Systems, Aalborg University, Denmark.
[d] Center for Language and Speech Technologies and Applications (TALP), the Technical University of Catalonia, Spain.
[e] Linguistic Data Consortium, University of Pennsylvania, USA.
[f] FERMUS research group, Institute for Human-Machine Communication, Technische Universität München, Germany.

Table.1 [3]

Most of these databases are gives age-dependent emotions and do not consider child emotions.

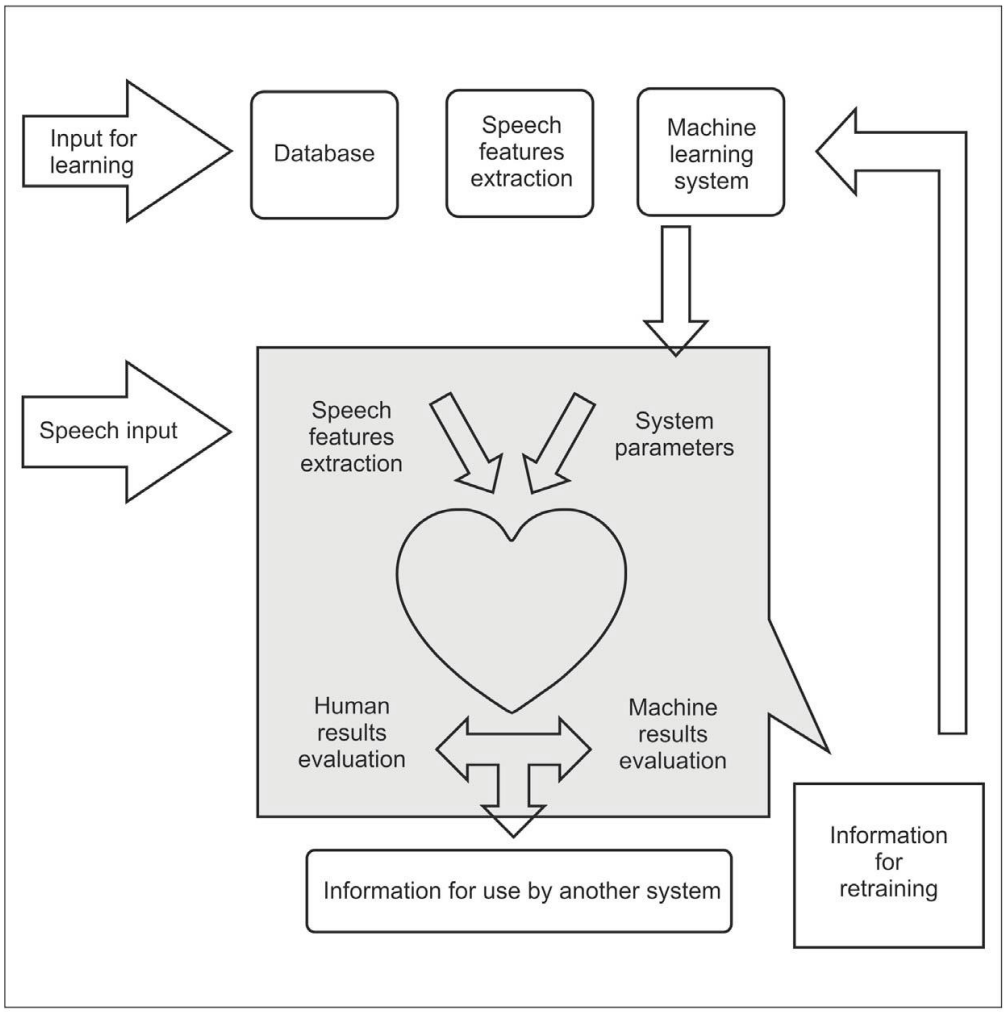

Fig.2 conceptual generic system for speech emotion recognition [4]

Speech features for extracting emotions:

Extracting of speech features relevant to the emotions are an important stage in which research is required. There is many features of speech like pitch, energy of signal, cepstral coefficients, zero crossing etc. from which we need to find those features from which we can extract maximum information related to emotions. Comparison of feature extraction approaches explained in [3]. The speech signal is a non-stationary signal there for finding features are difficult in this case so researchers divide the entire signal into small sections called frames and consider them as stationary to extract features. Global features are calculated over the entire signal and it is statistics of all the speech extracted from an utterance there for they are preferred over local features (pitch, energy etc.) which are calculated from each frame.

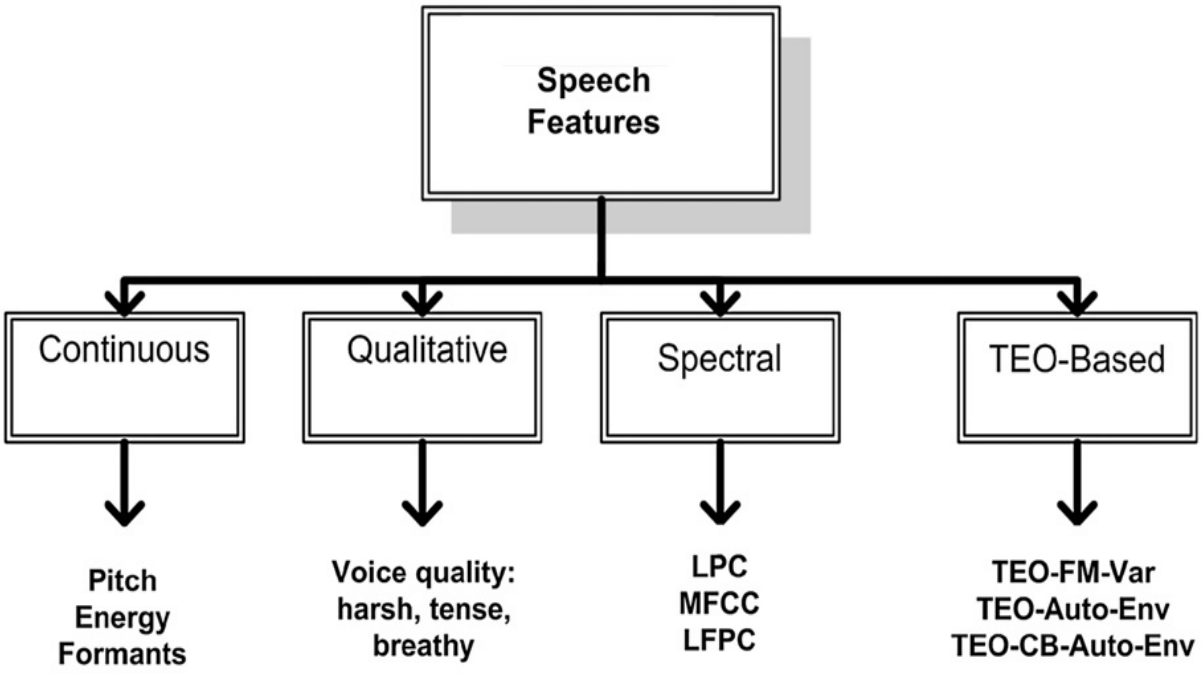

Fig.3 categories of speech features [3]

Acoustic analysis:

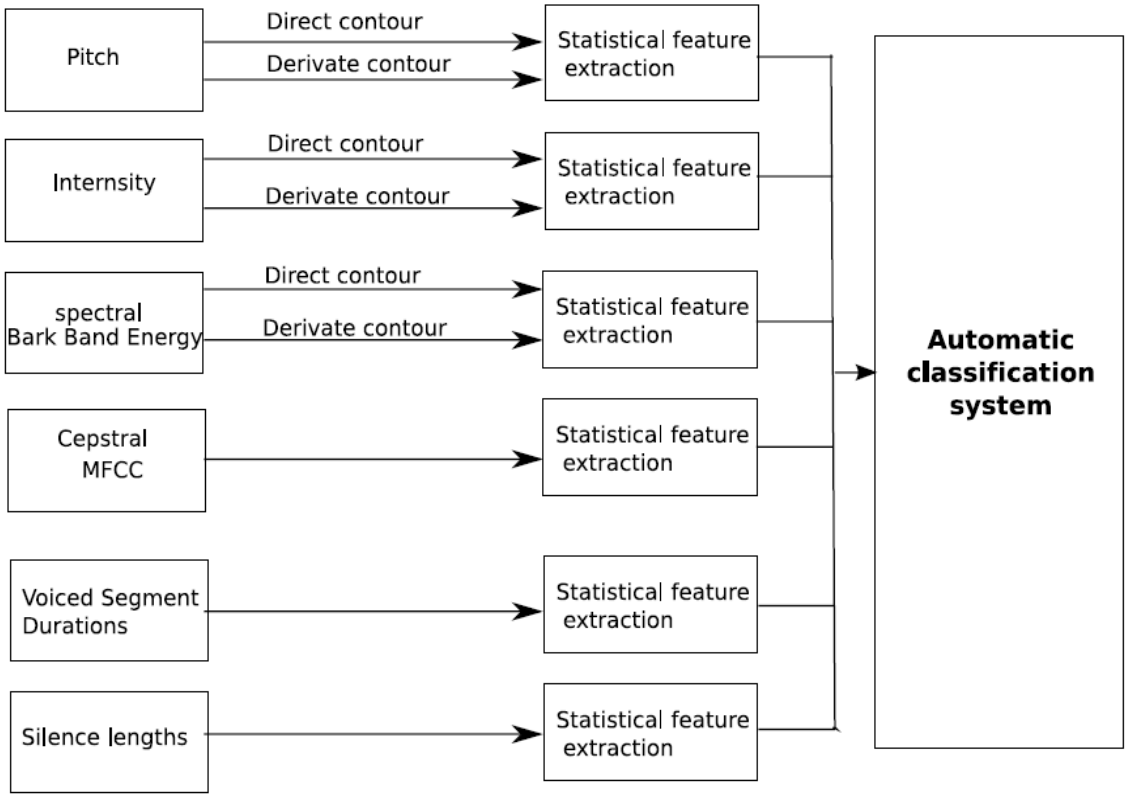

Fig.4 Speech feature extraction[7]

There are a number of feature sets that we can use for emotion detection such as intensity, pitch, MFCC, Bark spectral bands, voiced segment characteristics and pause length. We extract the statistical features from these which is summarized in above figure.4

Combining acoustic and Linguistic information:

In many situations Linguistic contains such as of spoken words contains the major part of emotions related to that word. In order to combine Linguistic information with acoustic, it is necessary to recognize the sentence of spoken utterance there for a particular language model is required to constrain the possible sentence in that particular language. In a study done in [3], it is shown that average recognition accuracy using only acoustic features is 74.2% and using only Linguistic features is 59.6% whereas using fusion it can increase till 92%.

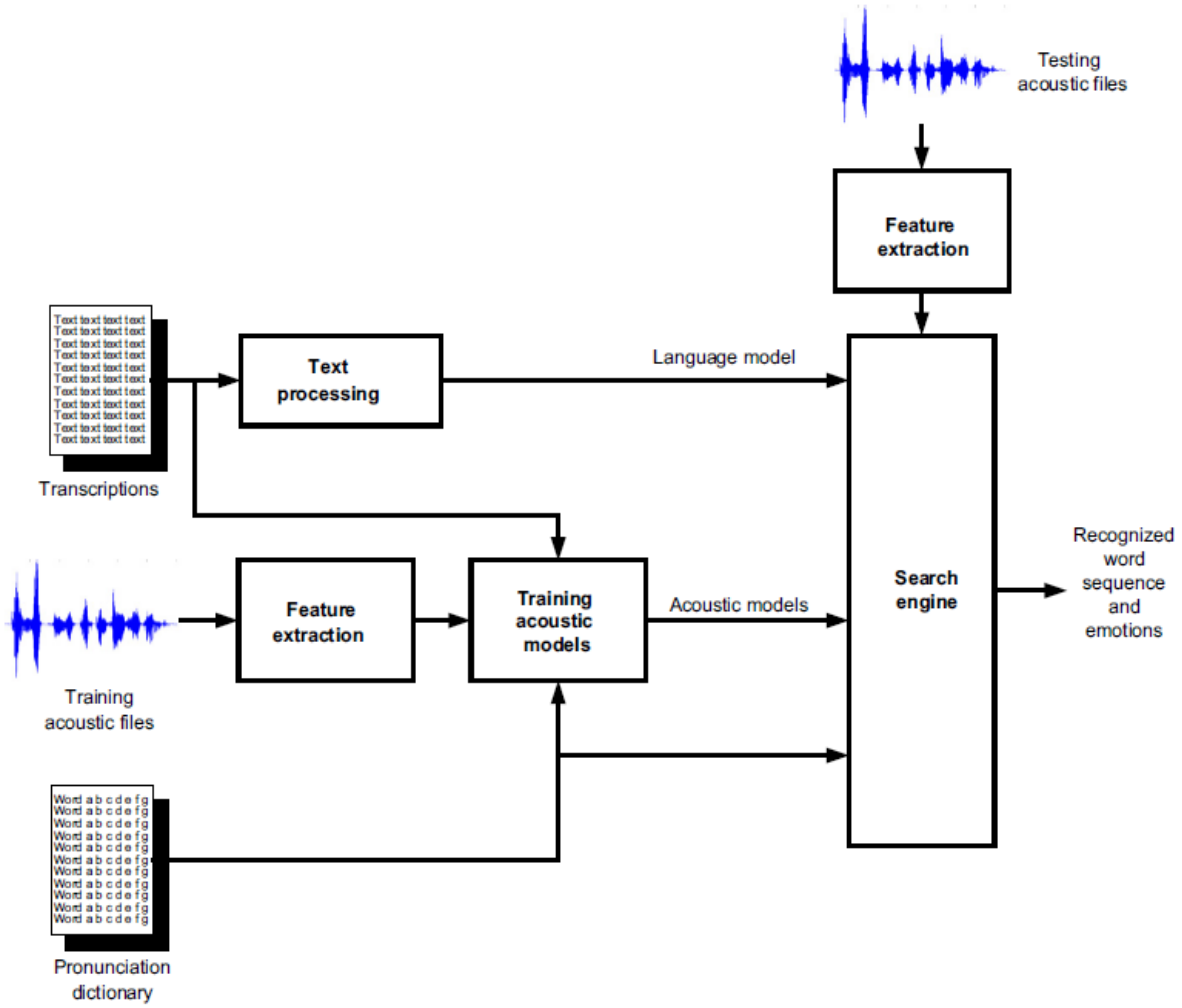

Fig.5 The architecture of a speech emotion recognition system combining acoustic and linguistic information [3]

Current research and future scope:

G. Wen has introduced Random Deep Belief Networks for recognizing emotions from speech. He believes that this technique will give improved results over other methods like hidden Markov model(HMM), Gaussian Mixture Model(GMM), Support Vector Machine (SCM) etc. which are confronted with complicated decision boundary of classification.[6]

In this research also he has stated that this system will give better results for the improved database. In Speech processing, there are many features which we can explore with greater effect if we have improved database which is currently a major concern in speech emotion recognition.

Facial Emotion Recognition

Facial expressions play most important role in human interactions. Humans communicate with each other using verbal or even in the non-verbal way. Since to implement effective human-computer interaction computer should understand non-verbal communication which is communicating with facial expressions.
Facial emotion detection can be classified mainly in three stages first face recognition, second feature extraction, and classification of recognizing emotions. Face recognition algorithm is successfully implemented efficiently using eigenvector face detection technique, the Viola/Jones Face Detector or using Neural network techniques etc. Once the face is detected the features such as eyes, nose, lips eyebrows etc. can be extracted. Some methods which can be used for feature extraction are Principal component analysis(PCA), Linear Discriminant Analysis(LDA), Gabor Wavelet, Discrete cosine transform, Dual tree complex wavelet transform etc.

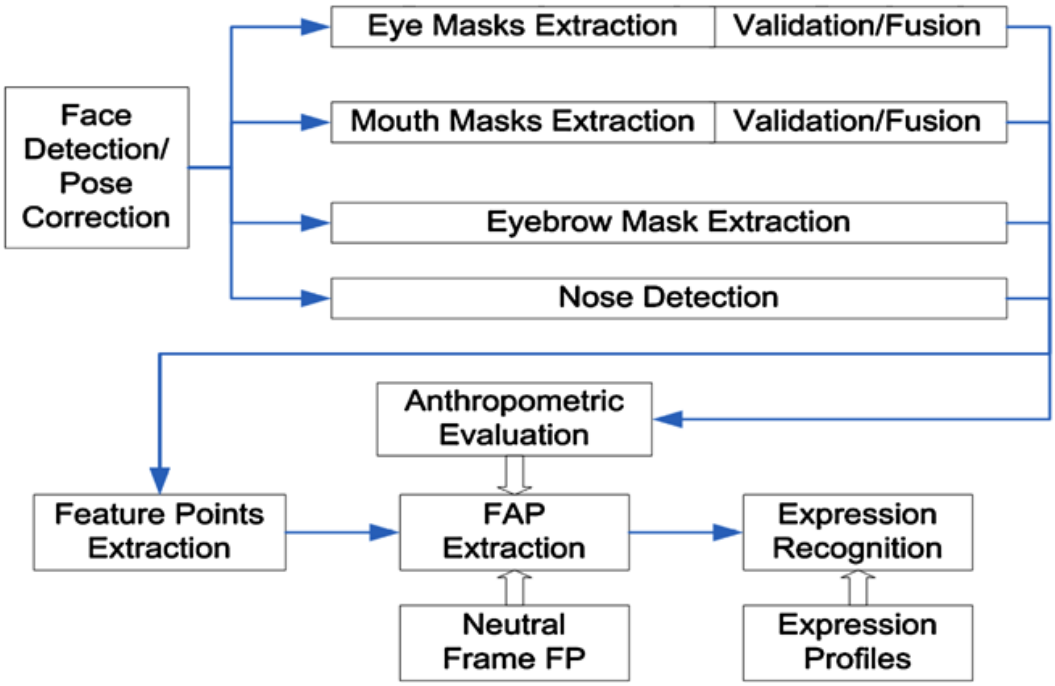

Fig.6 Face feature extraction [7]

In above face feature extraction Viola/jones face detector is used and after detecting a rectangular face boundary region is segmented into coarse facial feature areas those features boundaries need to be extracted. And each facial feature fused together which produce final mask the masking is done at Anthropometric evaluation using Anthropometric criteria. FAP (Facial Animation Parameters) measurement are calculated from Neutral Frame (Frame where participants expression found to be neutral) Feature points. The confident level of FAP then derived from FP and FAPs along with this confidence level detect desire facial expression [7].

Current Research and future work

G. Kalal in his recent research paper has developed new feature extraction technique which is a hybrid combination of 2D-PCA and 2D-LDA.
2D- linear Discriminant Analysis is used to overcomes the problem which occurs while using 1D-LDA, which are misclassification of the same face when presented with a different background. And 2D-PCA is used to laminated the correlation between images.[5]

The scenario for the database for facial emotion recognition is better than that of speech emotion recognition. Recently S. Happy made a database for facial emotion recognition which includes spontaneous expression database. [10]

Inventions of Realsence cameras will improve the performance of facial recognition systems. We can try different hybrid combinations of feature extraction methods and try to make the system more generic rather than conditional which depends on age or cultural diversity.

Multimodal Emotional recognition:

Humans generally express their Emotions through several modalities such as facial expression, spoken words with different tones and body gestures. Implementation of such multimodal emotion recognition system is very complex as we have to combine three different systems as well as implementing a decision maker which will then make a decision about the emotion based output of these three systems. A major problem in the implementation of such systems is the unavailability of the database.

L. Kesseus has built a multimodal Emotional recognition system where they used acoustic analysis for speech emotion recognition (system illustrated in fig.4), face emotion recognition (system illustrated in fig.6) and Body feature extraction. They used Bayesian classifier (BayesNet) software provided by Weka. The simple estimator is used for estimating the conditional probability table of the Bayesian network once the structure has been learned. The approach used for decision-level fusion is best probability approach i.e. selecting an emotion that received the highest probability in three models. Overview of this model is given in fig.7 [7]

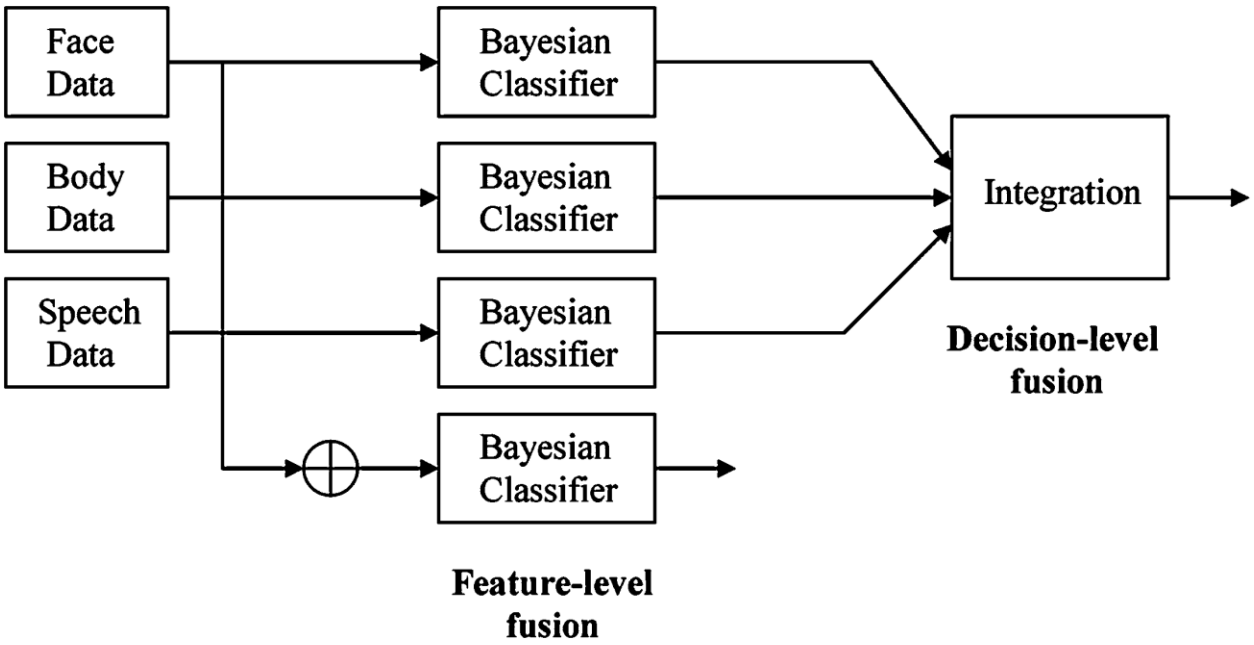

Fig.7 Multimodal Emotional Recognition [7]

Observations:

- The overall performance of system improved using multimodal Recognizer over uni-model.
- This model is useful when some modality features are missing or in noisy condition when transmitted signal get corrupted even though one or two modalities are absent still this model will give the emotion. [7]

Issues and Future work:
This study is done in less data set (only 10) and recorded in some controlled environment, therefore, its real-time implementation is still questionable.
An availability of universal data can make future work more reliable.

IV. CURRENT REAL WORLD

:) Affectiva

Affectiva is an emotion measurement technology company which has developed a way for computers to recognize human emotions based on facial cues or physiological responses. [9]
This is online web-based technology which recognizes five emotions happy, angry, disgust, contempt and surprise emotions very effectively.

V. CONCLUSION

Emotion Detection has large scope for research, from considering basic emotion detection to complex emotions. Improvement in uni-model signal processing emotion-recognizers. Multimodal emotion detection has many flows which can be minimized over the time.

I suggest using acoustic and linguistic combination model for speech emotion recognition instead of only acoustic and S. Kalal's hybrid feature extraction technique for facial recognition in multimodal recognition. As these techniques give improved results for uni-model using them in multimodal may increase the overall performance of the system.

REFERENCES

[1] P. Ekman, "Universals and cultural differences in facial expressions of emotion," in Proc. Nebraska Symp. Motivation, vol. 19, pp. 207–283, 1971

[2] Jerritta S, "Physiological Signal Based Human Emotion Recognition: A review" 2011 IEEE 7th International Colloquium on Signal Processing and its applications.

[3] M. El Ayadi, "Survey on speech emotion recognition: Feature, classification schemes, and database." 2011 ELSEVIER pattern Recognition

[4] S. Lugovic, "Techniques and Applications of Emotion Recognition in Speech" MIPRO 2016, May 30 - June 3, opatijia, Croatia.

[5] S. Kamal, "Facial Emotion Recognition for Human-Computer Interactions using hybrid feature extraction technique."

[6] G. Wen, "Random Deep Belief Networks for Recognizing Emotions from speech signals."

[7] L. Kessous, G. Castellano, and G. Caridakis, "Multimodal emotion recognition in speech-based interaction using facial expression, body gesture, and acoustic analysis," Journal on Multimodal User Interfaces, vol. 3, pp. 33-48,2009.

[8] https://en.wikipedia.org/wiki/Emotion_recognition

[9] https://en.wikipedia.org/wiki/Affectiva

[10] S L Happy, "The Indian Spontaneous Expression Database for Emotion Recognition" IEEE Transaction on affective computing VOL.8, No.1, Jan-Mar 2017t