UNIVERSITY OF UDINE

Department of Electrical, Management and Mechanical Engineering

PhD in Industrial and Information Engineering

CYCLE XXVII

PhD THESIS

A system for recognizing human emotions based on speech analysis and facial feature extraction: applications to Human-Robot Interaction

Supervisor:

Professor Alessandro Gasparetto

By:

Mohammad Rabiei

DEC-2014

A system for recognizing human emotions based on speech analysis and facial feature extraction: applications to Human-Robot Interaction

Foremost, I would like to express my deepest thanks to my **Professor Alessandro Gasparetto** for allowing me to conduct this research under his auspices. I am especially grateful for his confidence and the freedom that he gave me to do this work. Without a coherent and illuminating instruction, this thesis would not have reached its present form.

I wish to thank to **my family** for their encouragement. Without the support of all members of my family, I would never finish this thesis and I would never find the courage to overcome all these difficulties during this work. My thanks go to **my parents** for their confidence and their love during all these years.

I extend my sincere thanks to all members of the **Department of Electrical, Mechanical and Managerial Engineering (DIEGM) of University of Udine, (ITALY)** and all those who contributed directly or indirectly to the dissertation.

I am deeply grateful to all **members of the jury** for agreeing to read the manuscript and to participate in the defense of this thesis.

I also want to thank the all my friends and my colleagues in the laboratory for their kindness and help. Finally, I express my sincere gratitude to who kindly agreed to recordings their speech and facial emotion in the emotion recognition system.

To my parents

For unconditionally providing their love, support, guidance and encouragement throughout my education.

CONTENTS

Figure

Table

Abstract

With the advance in Artificial Intelligence, humanoid robots start to interact with ordinary people based on the growing understanding of psychological processes. Accumulating evidences in Human Robot Interaction (HRI) suggest that researches are focusing on making an emotional communication between human and robot for creating a social perception, cognition, desired interaction and sensation.

Furthermore, robots need to receive human emotion and optimize their behavior to help and interact with a human being in various environments.

The most natural way to recognize basic emotions is extracting sets of features from human speech, facial expression and body gesture. A system for recognition of emotions based on speech analysis and facial features extraction can have interesting applications in Human-Robot Interaction. Thus, the Human-Robot Interaction ontology explains how the knowledge of these fundamental sciences is applied in physics (sound analyses), mathematics (face detection and perception), philosophy theory (behavior) and robotic science context.

In this project, we carry out a study to recognize basic emotions (sadness, surprise, happiness, anger, fear and disgust). Also, we propose a methodology and a software program for classification of emotions based on speech analysis and facial features extraction.

The speech analysis phase attempted to investigate the appropriateness of using acoustic (pitch value, pitch peak, pitch range, intensity and formant), phonetic (speech rate) properties of emotive speech with the freeware program PRAAT, and consists of generating and analyzing a graph of speech signals. The proposed architecture investigated the appropriateness of analyzing emotive speech with the minimal use of signal processing algorithms. 30 participants to the experiment had to repeat five sentences in English (with durations typically between 0.40 s and 2.5 s) in order to extract data relative to pitch (value, range and peak) and rising-falling intonation. Pitch alignments (peak, value and range) have been evaluated and the results have been compared with intensity and speech rate.

The facial feature extraction phase uses the mathematical formulation (Bézier curves) and the geometric analysis of the facial image, based on measurements of a set of Action Units (AUs) for classifying the emotion. The proposed technique consists of three steps: (i) detecting the facial region within the image, (ii) extracting and classifying the facial features, (iii) recognizing the emotion. Then, the new data have been merged with reference data in order to recognize the basic emotion.

Finally, we combined the two proposed algorithms (speech analysis and facial expression), in order to design a hybrid technique for emotion recognition. Such technique have been implemented in a software program, which can be employed in Human-Robot Interaction.

The efficiency of the methodology was evaluated by experimental tests on 30 individuals (15 female and 15 male, 20 to 48 years old) form different ethnic groups, namely: (i) Ten adult European, (ii) Ten Asian (Middle East) adult and (iii) Ten adult American.

Eventually, the proposed technique made possible to recognize the basic emotion in most of the cases.

Keywords: Action Units (AUs), Emotion recognition, Facial expression, Human-robot Interaction, Speech analysis.

CHAPTER 1

Related work on speech emotion recognition

And

Facial features extraction

Human-robot interaction has been a topic of science fantasy stories even before any robots existed. Currently uses of human-robot are increasing, costs are going down and sales are growing up. In addition capabilities have been improved specially, in the safest physical communication. However, emotion is fundamental for human’s communication and perception in everyday activities. Although, we are still far from to have an effective emotional interaction between human and robot.

The system for emotion detection refers to the quantitative analysis of verbal and non-verbal communication; focus on behavior, gestures and facial expressions. The researchers in human robot interaction filed have been proposed the new freeware to increase accuracy emotional communication (verbal and non-verbal) between human and Robot.

The core part of this work, introduce the new methodology and program that uses for analysis of emotional states in speech communication, facial features extraction, emotion classification and evaluating the quality of emotion recognition system. By means of designing the new model for emotion recognition system we have been mixed the theory of sound signals and facial features recognition system. We have controlled the implementation of system with new results to create highly optimized emotion states.

In the emotion recognition system have been bused two experiments for classification of emotion stats. Experiment 1 proposed the algorithms for speech emotion recognition. The model have been used (pitch, formant, intensity, speech rate and voice quality) plots. Also, in Experiment 2 the proposed algorithm increases the accuracy of facial emotion detection and recognition system. Then, for improvement the accuracy of classification, the results with fusion of pairwise (speech emotion recognition and facial features extraction).

Finally, the proposed system have been developed on emotion recognition for different ethnic groups with new hybrid emotion recognition software.

In this work the capture 1 divided in three sections. Section 1 gives a brief overview of the previous work on speech emotion recognition, section 2 describes some of the related work on facial features extraction and section 3 gives a short overview of the hybrids model on emotion recognition system.

1.1. Speech emotion recognition

With the advance in human-robots interaction robot have been used in society such as toys , service, security guards, teachers, e-learning, diagnostic tool for therapists, search and rescue [1]. Speech emotion recognition has also been used in call center applications and mobile communication [2]. Some works tried to incorporate spoken dialogue system technology and service robots. Psychologists believe that faces to face communication and analysis of speech sounds (words and sentences) are considered as one of the prompt methods for human emotional interaction.

We are still far from having an emotional interaction between human and robot because the machine does not have sufficient intelligent to understand human voice, speech and basic emotional state of the speaker. In human–human communication process, determination of a speaker's states is more easier than to assessment of communication in human machine interaction, because partner's in speech process constantly adapt their manner of speaking, based on the culture, mood, gender, mother- tongue age and pre emotion.

Emotion is conveying significantly by verbal and non-verbal signals (body posture, gestures and facial expressions) expressed for determining the conversation. Psychologists believe that for emotional interaction between human and machine, a system must be developed to understand the human emotions in various environments.

However in Encyclopedia emotion means "A mental positive or negative state that present in the life and consistent responses to internal or external events". Although, the definition of 'emotion' in psychologist article is "what is present in most of life but absent when people are emotionless"; this is the concept of pervasive emotion [3].

The human to human interaction consists of two channels, the implicit and explicit channel. The task of speech emotion recognition is very challenging, because it is not clear which speech features are the most suitable, in different cultures and manner to distinguish basic emotions. It is very difficult to determine the boundaries between these portions of basic emotions, sometimes they have slightly overlap with each other. nevertheless, beside the emotional states, the voice and speech signal can conveys a rich source of information about a speaker's such as; physiological state, age, sex/gender and regional background.

The speech signal is the fastest, efficient and natural method of communication between humans and robot for this reasons numerous studies have been done on speech emotion recognition, for various purposes [4]–[7].

1.1.1. Speech processing recognition for basic emotion

Psychiatrists in the 20th century started the empirical investigation of the effect of emotion on the voice signals. The most important works done on emotion recognition through speech analysis are the famous Ekman's and Fox's models. In the early 1990s, Ekman and his colleagues have performed extensive work, which led to categorization of emotions into seven basic emotion models (sadness, surprise, happiness, anger, fear, disgust and neutral) and more emotions can be defined by mixtures of the basic emotions [8]. While Fox's is a multi-level emotional model and involving stimulus evaluation subsequent steps in the generation of emotion [9].

Davidson et al. in 1979 for increases the emotion recognition time, proposed new model to categorize emotions. In this method emotions specified in two axes (arousal and valence). As shown in Figure. 1-1 in vertical axis the arousal represents the quantitative activation and the valence in horizontal axis refers to the quality of basic emotion (positive and negative valence) [10].

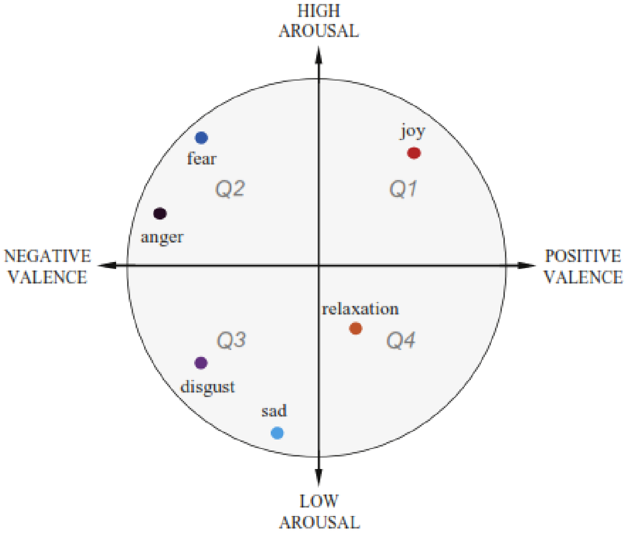


Fig. 1-1. Emotion recognition in two-dimensional.

In different comprehensive research Schlosberg proposed a three dimensional emotion space: activation (arousal), potency (power), and valence (pleasure, evaluation). Figure. 1-2 shows the locations of six common basic emotions in the three dimensional emotion space [11].

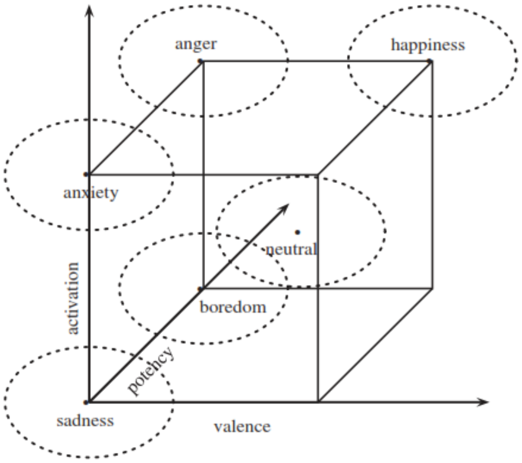


Fig. 1-2 Three-dimensional emotion space (activation, potency, and valence) and six basic emotions.

Accordingly, a lot of researches have been done on speech emotion recognition, which is defined as extracting the emotional states of a speaker from his or her speech. It is believed that speech emotion recognition can be used to extract useful semantics from speech [12]. Hence, improves the performance of speech recognition systems.

Several reviews on emotional speech analysis have been done with focused on content of speech signals and classification techniques for recognition of emotions. The most extensive work about automatic emotion recognition from speech signals was done by Yang and Lugger. They had proposed a new set of harmony features for automatic emotion recognition based on the psychoacoustic harmony perception [13]. They had estimated pitch contour of an utterance. Then, calculated the circular autocorrelation of the pitch histogram on the logarithmic semitone scale. In more comprehensive research, several approaches have been proposed to identify the emotional content of a spoken utterance [14]–[16].

Most researchers had reported emotion recognition consists of two major steps, speech feature extraction and emotion classification. However the classification theory in emotion recognition from sound signals developed well extraction of features from speech signals depends on database, algorithm and software design.

The design of speech emotion classification and recognition system have three main aspects. The first one is the choice of suitable features for speech representation. The second issue is the design of an appropriate classification scheme and the third issue is the proper preparation of an emotional speech database for an evaluating system performance [17].

Another challenging issue that helps technical researchers and social psychologists to better recognition of emotions have been designing a model that define culture dependency, regional background and environment of speaker [18].

To date, research on the verbal expression of emotion has demonstrated that many features may be involved. Whereas there has tended to be an overwhelming focus on pitch variables, pitch contour, speech rate, intensity, pausing structure, accented and unaccented syllables and duration of speech. They can provide powerful indications of the speaker's emotion [19]–[22]. Although researchers tend to pitch, spectral and formant are fundamental importance, relatively little is known about the interaction of pitch, speech rate, intensity and formant in communication.

Experimental support for the basic importance of voice quality can be found by Scherer et al. in1984. Scherer suggests that tense voice is associated with anger, joy and fear; and that lax voice (at the phonation level essentially the same as breathy voice) is associated with sadness [23]. Also, the most extensive work about sound analysis was done by Scherer. This research was based on the vocal communication of emotion and model the complete process, including both encoding (expression), transmission and decoding (impression) of vocal emotion communication. Special emphasis is placed on the conceptualization and operationalization of the major elements of the model (i.e., the speaker's emotional state, the listener's attribution, and the mediating acoustic cues) [24].

Some works tried to incorporate feature extraction and analysis the different aspects of emotions in voice [25]–[28]. Also, some researches have been improved the analysis of speech signals graph (pitch value, pitch rate and pitch range) [29]. In 1986 Scherer further asserts that ‘‘although fundamental frequency parameters (related to pitch) are undoubtedly important in the vocal expression of emotion, the key to the vocal differentiation of discrete emotions seems to be voice quality and intensity’’[23].

Numerous studies have been done on the influence of emotions on intonation patterns (more specifically F0/pitch contours) and design a new coding system for the assessment of F0 contours in emotion [30]. Nonetheless, few studies on emotions influence from sound signals have attempted to describe F0 contours (average F0, F0 level or F0 range) for different emotional expressions [31]. A number of authors have claimed that specific intonation patterns (F0/pitch and other vocal aspects) reflect specific emotions [32] [33]. However in recent year, very few empirical studies have focused on the tone sequence models. Tone sequences have been used more extensively than pitch movement algorithm for the description and analysis of linguistic intonation.

The (MSFs) system, described by Siqing et al. (2011). They proposed system (spectral features (MSFs)) for the automatic recognition of human emotion and affective information from sound signals. The features were extracted from Mel-frequency cepstral coefficients and perceptual linear prediction coefficients [34]. The features were extracted from an auditory-inspired long-term spectrum temporal representation and proposed algorithm check classification of discrete emotions and estimation of continuous emotions (e.g. valence, activation) under the dimensional framework. Also, selected features were based on frequency analysis of the temporal envelopes (amplitude modulations) of multiple acoustic frequency bins, thus capturing both spectral and temporal properties of the speech signal [34] . In the past years, researchers were focused on prosodic features, more specifically on pitch, duration and intensity and less frequently on voice quality features as harmonics-to-noise ratio (HNR), jitter, or shimmer [3].

Past researches carried out in the field of automatic recognition of emotion from speech has tended to focus on speech feature extraction, classification, robustness, evaluation, implementation and system integration. One of the most important questions in emotion recognition is how many and which kind of features must be choosing. The answer to this question is important to processing speed of system and memory requirements. The feature selection strategies have been used the most common techniques, namely; Principal Component Analysis (PCA), Linear or Heteroscedastic Discriminant Analysis (LDA), Independent Component Analysis (ICA), Singular Value Decomposition (SVD) and Non-negative Matrix Factorization (NMF) [35]. Arguably the most important phase in emotion recognition is extraction of a meaningful and reasonably set of features. So far there has not been a large-scale Linguistic features. Also, this analyzing and classification depend on the database. The feature extraction method requires some kind of division of features into basic emotion classes.

Principal Component Analysis (PCA) is based on multivariate analyses that use statistical procedure. The model have been invented in 1901. Schölkopf in 1997 used integral operator kernel functions and proposed a new method for performing a nonlinear form Principal Component Analysis (PCA) in speech emotion recognition. By the use of new functions can efficiently compute principal components in high dimensional feature spaces, related to input space by some nonlinear map [35].

Linear discriminant analysis (LDA) is closely related to regression analysis and used in machine learning, pattern recognition and statistics. In emotional speech recognition field researcher used linear discriminant analysis by means of combination of sound features. Potamianos et al. in 2003 proposed a new work on modeling of audiovisual speech feature and decision fusion combination with linear discriminant analysis (LDA). The proposed algorithm have three main subject, namely; bimodal databases, ranging from small- to large-vocabulary recognition tasks, recorded in both visually control and challenging environments [36]. The principle component analysis (PCA) method has been used for improve the accuracy of feature extraction and emotion classification. Also, linear discriminant analysis (LDA) with floating search compares the emotion features extraction.

Independent component analysis (ICA) is the computational method that use non-Gaussian signals for analyses person's that speech in a noisy room. In this system the source of sound signals must be independent each other. Attempts to generate sound recognition system have been made using Independent component analysis (ICA) by MA Casey in 2001. The proposed system used dimension log-spectral features, select robust acoustic features and a minimum entropy with hidden Markov model classifier [37].

Nonetheless, there have been a number of attempts to recognition of speech and musical instrument with Singular Value Decomposition (SVD) and Non-negative Matrix Factorization (NMF) [38]–[40]. The work by Cho et al. in 2003 proposed the system for classification of speech features. This methods was based on feature extraction from Spectrum-Temporal sounds and using the non-negative matrix factorization (NMF) [41].

A speech emotion recognition system consists of various types of classifiers. However, each classifier have its own limitations and advantages. For example in recent year, Hidden Markov Models (HMM) and Gaussian Mixture Models (GMM) generally used in the article with solid mathematical basis and classifier on emotion classification probably [42]–[44].

1.1.2. Speech recognition research in the past 5 years

Some of the high quality articles in that issue were dealing with the Automatic Speech Recognition (ASR) model, Speaker Recognition and automatic recognition of realistic emotions in speech with statistical and mathematical procedures. In more comprehensive research Yung and Lugger in 2010 proposed the Mel frequency central coefficients (MFCC) models. The system was successful for speech emotion recognition. As it shows in Figure. 1-3 all features that they used for classification of emotion divided in seven sub branches, namely energy, pitch, duration, formant, statistic, harmony and voice quality [45].

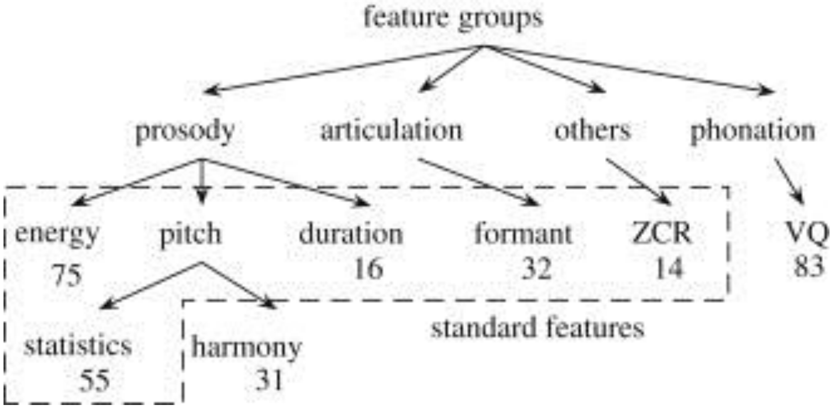


Fig. 1-3 Feature groups and number of features for speech emotion recognition.

As described above some works tried to incorporate the emotional speech with Hidden Markov Models (HMMs), Neural Networks (NNs), Support Vector Machines (SVMs), Gaussian Mixture Models (GMMs) and Meta-Classification by means of extracting certain attributes from speech.

Popular classifiers for emotion recognition such as K-Nearest Neighbor (KNN) classifiers and Linear Discriminant Classifiers (LDCs) have been used since the very first studies. K-Nearest Neighbor (KNN) divide the feature space into cells and are sensitive to outliers.

In the last few years the research in automated speech emotion recognition methods in real time with minimum redundancy – maximum relevance was steadily growing. Kukolja et al. proposed the comparative analysis of methods for physiology-based emotion estimation and the model was combination of mRMR+ KNN for real-time estimator adaptation due to considerably lower combined execution and learning time of KNN versus MLP [46].

On the other hand, a natural extension of LDCs is Support Vector Machines (SVM). If the input data have not been transformed linearly, maybe have increased or decreased the number of features and if the linear classifier obeys a maximum-margin fitting criterion, then we obtain SVM.

In obviously related work, Support Vector Machine (SVM) was done with Chen et al. in 2012. They solved the speaker independent speech recognition with new Support Vector Machine (SVM) model for six basic emotions, including happiness, anger, fear, sadness, surprise and disgust. In order to evaluate the proposed system, principal component analysis (PCA) for dimension reduction and artificial neural network (ANN) for classification were adopted to design four comparative experiments, including Fisher + SVM, PCA + SVM, Fisher + ANN, PCA + ANN [47].

As noted in the speech emotion recognition research, small data sets are in general, can better handle by discriminative classifiers [48]. The most used non-linear discriminative classifiers were likely to be Artificial Neural Networks (ANNs) and decision trees. Decision hyper planes learned with Artificial Neural Networks might become very complex and depend on the topology of the network (number of neurons), on the learning algorithm.

With the new approach in speech emotion recognition systems Stuhlsatz et al. proposed the new model for Generalized Discriminant Analysis (GerDA). This model was based on Deep Neural Networks (DNNs) by means of feature extraction and classification [49]. The GerDA was able to optimize the acoustic features in real time and used simple linear classification.

The essence of speech emotion analysis is evaluating the emotional speech in different environments. Gao et al. used Hidden Markov Models (HMMs) to evaluate the emotional recognition performance [50]. Among dynamic classifiers, HMM have been used widely in speech and speaker recognition system. Le and Mower Provost in 2013 proposed the new model to hybrid classifiers which used Hidden Markov Models (HMMs) to capture the temporal property of emotion and DBNs to estimate the emission probabilities [51].

Wollmer et al. in 2010 designed the system for sensitive artificial listener (SAL) in human-robot communication. The proposed algorithms have been used linguistic and acoustic as well as long range contextual. The main system components are hierarchical dynamic Bayesian network (DBN) for detecting linguistic keyword features and long/short-term memory (LSTM). Recurrent neural networks use for emotional history to predict the affective state of the user [52].

The sound signals transmits multiple layers of human information. Most studies related to description of emotion in speech. They have been focused on fundamental frequency (f0), speech duration, speech rate and energy (intensity of the voice and amplitude) distribution [53]–[56]. Guzman et al. done a study by means of the influence emotional expression in spectral energy. they have been suggested that expression of emotion impacts and the spectral energy distribution. Also, they discovered that emotional state by a breathy voice quality (sadness, tenderness and eroticism) present a low harmonic energy above 1 kHz, and emotional states (anger, joy and fear) by high harmonic energy greater than 1 kHz [57].

There are a number of issues in automatically detection on emotional content from human voice. Origlia et al. in 2013 attempts to analysis of speech for emotion recognition progresses with extraction of acoustic properties in a real time. Feature extraction method built on the basis of a phonetic interpretation of the concept of syllables for the emotions transmission [58].

Furthermore, in an obviously related way, some researchers have been represented a major advance in terms of conceptualizing the neural processes in emotional speech. In doing so, Iredale et al. examined the neural characteristics of emotional prosody perception with an exploratory event [59].

In order to advance the functional model of vocal emotion, they have been examined ERP correlates of affective prosodic processing perception. Finally they proposed model of vocal emotion perception for (happiness, angry and neutral).

Some of the high quality articles in the emotion recognition were dealing with the spontaneous speech analysis. The most extensive work in this field was done by Cao et al. They research was based on ranking approach for emotion recognition. It also incorporates the intuition that each utterance can express a mix of possible emotion and that considering the degree to which emotion was expressed [60]. The proposed system combined standard SVM classification and multi-class prediction algorithm by means of identify the emotion in spontaneous dialog with high accuracy in recognition. In an obviously related work they have been used arousal, valance, power and word usage in emotion recognition for predicting dimensions in spontaneous interactions. For the analysis of acoustics they found that corpus-dependent bag of words approach with mutual information between word and emotion dimensions [61].

With the development of machine learning, digital signal processing and various computer applications in real-world some researchers focused on EEG-based emotion recognition. Wei Wang et al. found that classification from EEG data with machine learning have optimal result on emotion recognition. They proposed a system for removing the noise in speech signals and used EEG features for emotion classification [62].

There have been several researches on emotion classification system in the field of Natural Language Processing (NLP) [63]. For instance Li and Xu have done an extensive work on the role of social networks on people's emotion. Finally, they proposed a model for emotion recognition from personal sound and music on social networks.

Influence of culture in speech emotion has been developed in human interaction and speech communication system. Human emotion recognition dependent on pre emotion state, culture and environment of communication place. Attempts have been made to understand the degree of culture dependency were in human interaction. Most of researchers believe that some common acoustical characteristics have similar emotion across different culture. Kamaruddin et al. in 2012 used Mel Frequency Cepstral Coefficient, (MFCC) method, classified with neural network (Multi-Layer Perceptron (MLP)) and fuzzy neural networks (Fuzzy Inference System (ANFIS). They used Generic Self-Organizing Fuzzy by means of shown the role of cultural dependency for understanding speech emotion [64].

1.1.3. Databases for speech emotion recognition system

Database is one of the important issues to the training and evaluation of the emotion recognition system. Database also can be used to assess the system performance. In some cases used low quality database led to incorrect results in conclusion. Moreover, the limitations of the emotional speech and design of the database is an important to the classification of basic emotion [65]–[67].

According to some studies emotional databases divided into two types, namely; speech scenario and type of speakers [68], [69]. The speech scenario cover the emotions in context and connected to speakers that produce specific emotions. However, the types of speakers relate to different ethnic groups with different culture in intonation and accent when the speakers read the sentences.

Accordance with past research to use speech database we have two recording collection emotional database namely, acted and realistic emotions. Williams and Stevens found that acted emotions tend to be more exaggerated than real ones [70].

The emotional real voices are usually stimulated and recorded by nonprofessional or professional speakers. In fact, nonprofessional speakers are invited to produce emotional speech databases. Also, they exaggerated in the specific emotions such as: sadness, happiness, anger and surprise.

Almost all the existing emotional speech databases do not well enough in the quality of the voice analysis. Also, the quality and the size of database is necessary to recognition of the emotion rate [71].

Another challenge is the lack of available emotional speech databases for public use among the researchers on speech emotion recognition. Thus, there are very few benchmark databases that can be shared among researchers [72].

Table. 1-1 summarizes characteristics of a collection on emotional databases. It is based on the human and commonly used in speech emotion recognition.

Table. 1-1 English Speech Databases for speech recognition.

Identifier	Emotional content	Emotion elicitation methods	Size	Nature of material	Language
Reading-Leeds database [73] (2001)	Range of full blown emotions	Interviews on radio/television	Around 4 ½ hours material	Interactive Unscripted Discourse	English
France et al. [72] (2000)	Depression, suicidal state, neutrality	Therapy sessions & phone Conversations.	115 subjects: 48 females 67 males.	Interactive Unscripted Discourse	English
Campbell CREST database, ongoing [74] (Campbell 2002-2003)	Wide range of emotional states	Record the social spoken interactions throughout the day	Target - 1000 hrs over 5 years	Interactive unscripted Discourse	English Japanese Chinese
Capital Bank Service and Exchange Customer Service [75] (2004)	Mainly negative - fear, anger, stress	Call centre human-human interactions	Unspecified (still being labelled)	Interactive Unscripted Discourse	English
DARPA Communicator corpus [76] (2002)	Frustration, annoyance	Human machine dialogue system	Recordings interactions with a call centre	Users Called Systems	English
KISMET [65] (2009)	Approval, attention, prohibition, soothing, neutral	Nonprofessional actors	1002 utterances, 3 female speakers, 5 emotions	Interactive Unscripted	American English
FERMUS III [76] (2004)	Anger, disgust, joy, neutral, sadness, surprise	Automotive environment	2829 utterances, 7 emotions, 13 actors	Interactive Unscripted	German, English
MPEG-4 [77] (2000-2009)	Joy, anger, fear, disgust, sadness, surprise, neutral	U.S. American movies	2440 utterances, 35 speakers	Interactive Unscripted	German, English
Fernandez et al. [78] (2000, 2003)	Stress	Verbal responses to maths problems in driving context	Data reported from 4 subjects	Unscripted Numerical	English
Mc Gilloway [79] (1999)	Anger, fear, happiness, sadness, neutrality	Contextualised acting: subjects asked to read passages for each emotional state	40 subjects reading 5 passages	Non interactive and scripted	English
Belfast structured Database [80] (2000)	Anger, fear, happiness, sadness, neutrality	Occurring emotion in the Belfast Naturalistic Database	50 subjects reading 20 passages	Non interactive and scripted	English
Pereira [81] (2000)	Anger (hot), anger (cold), happiness, sadness, neutrality	Acted	2 subjects reading 2 utterances	Scripted (emotionally neutral, 4 digit number)	English
Yacoub et al. (2003) (data from LDC) [82]	15 emotions Neutral, hot anger, happy, sadness, disgust, …, contempt	Acted	2433 utterances from 8 actors	Scripted	English

1.1.4. Modelling and software for categories of speech emotion recognition

Speech emotion recognition can have interesting applications in many fields of sciences, such as medical areas, psychology and human-robot interaction. Although, technologies have been developed to improve the effectiveness of communication system, affective high-level human interaction with robots is still far from ideal.

The psychological and engineering approach has been modelled speech emotional system based on categories or dimensions. For example phonetic processing field is strong to focus on categorical perception. In the case of dimensions, it is foremost the question how many and which dimensions we should assume: Traditionally, arousal and valence are modelled, with or without a third category power/dominance/control [3]. Of course, dimensional modelling can be more or less continuous and more than one dimension. The performance of automatic systems for recognition of emotions based on speech analysis is still weak for spontaneous speech. There is evidence, from human interaction experiments, that language models for dialog can be improved by using additional sources of information and by improving the modeling of acoustic and prosodic features.

Most acoustic features that used in speech recognition can be divided into spectral and prosodic categories. Prosodic features have been shown to deliver recognition, including pitch value, pitch peak, pitch range, intonation, accent, mute and rate of speech. Spectral features convey the frequency content of the speech signal. The spectral features are usually extracting over short frame duration. In addition we can express energy features such as low-frequency and high-frequency domain in some kinds of verbal interaction. Spectral features convey the frequency content of the speech signals. Also, Spectral can express energy features such as low and high frequency domain and improve emotion recognition in some kinds of human robot interaction [84].

Many speech analyses program are found as well but two open source software packages that development in the speech recognition. We used software PRAAT tools for feature extraction and Support Victor Machine (SVM) for classification of emotion. Also, the Munich open-source Emotion Recognition Toolkit (openEAR) is free platform in emotion recognition from speech and a similar initiative. The EmoVoice toolkit is a comprehensive framework for real-time recognition of emotions from acoustic properties of speech between speech analysis researchers, but it is not an open source program [83].

In section 2 we extensively explain about PRAAT and open EAR software for recording speech and extraction of features on sound signal graph.

1-2 Facial feature extraction

Humans belong to various ethnic groups with different attributes of facial features (shape, color and size). Also, they have diverse emotion expressions, depending on culture, age and gender.

Over the last decade, system for facial emotion expression has become an active research filed in different areas such as: human robot interaction, marketing analysis, facial nerve grading in medicine, social network control and new computer game. Facial expressions reflect of physiological signals and mental activities in social interaction.

Emotion on face to face interaction convey significantly by implicit and non-verbal signals (body posture, gestures and facial expressions) expressed for determining the spoken message.

According to a new study in behavioral, culture is a huge factor in determining the facial expressions. For instance, Americans people tend to look to the mouth for emotional cues, whereas Japans tend to look to the eyes [85].

Facial expressions are one of the important ways in humans and animals to conveying social information in nonverbal communication. Each emotion expression corresponds to a different motion of the facial muscles. Humans can adopt a facial expression and different emotion in each case. There are two brain pathways associated with facial expression namely: involuntarily (neural in the brain) or voluntarily (socially conditioned in the brain). But in the brain neural mechanisms and muscles are responsible for controlling the different expression in each emotion.

Facial emotion expression have been considering as one of the universal and prompt methods for human communication. People view and understand facial expressions in the social situations around them. Face Recognition generally involves two stages: firstly, is searched to find any face in the image (face Detection) and secondly, is detected, processed face and compared the results to a database of known faces (face Recognition). Finally, system decided base on sets of information and rules.

Facial expressions are generate by contractions of facial muscles such as: eyebrows, eyes, eye lids, lips, mouth and wrinkles of the noise. In facial expressions the lips and eyes are often as an important component for emotion recognition. Typical changes of muscular activities are brief, lasting for a few seconds, but rarely more than 5s or less than 250ms [86].

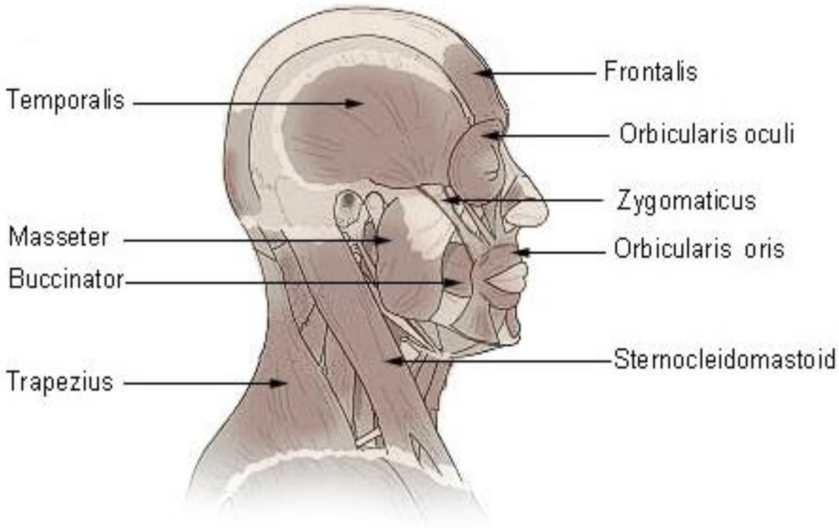
The reasons for this interest in facial research and analysis are multiple that namely: face tracking, face detection and face recognition in different area of the sciences.

In this section we would like to gives a short overview of the previous work in facial expressions and introduce the most prominent facial expression analysis methods, facial classification method and facial motion and facial recognition software and data base.

1-2-1 Facial feature extraction for basic emotion

Facial expressions have been studied for more than 150 years. In the 1870s, Charles Darwin wrote a first book about emotion and facial expressions [87]. Darwin was particularly interested in the functions facial expression as evolutionarily important for survival. He looked at the functions of facial expression in terms of the utility of expression in the life on the animal and in terms of specific expressions in species. Darwin deduced that animals were communicating feelings of different emotional states with specific facial expressions. He further concluded that communication was important for the survival of animals in group-dwelling species; the skill to effectively communicate or interpret another animal's feelings and behaviors would be a principal trait [88].

A century later, in the early 1970s, Ekman and his colleagues have performed extensive work, which led to categorization of facial expressions [89]. Also, Ekman was a pioneer of the Facial Action Coding System (FACS) model [90]. In Figure 1-4 have been shown the role of facial muscles on facial expression (FACS).



.

Fig. 1-4 The role of muscles in facial expression in (FACS).

The Ekman system was common standard to systematically categorize the physical expression of emotions. It has proven useful to psychologists and game animators.

In the past 20 years, facial expression analysis has been increasing interest in more comprehensive research such as robotic, computer vision and machine learning. As noted in this research, facial expression is one of the most important channels of nonverbal communication and human behavior (about emotion). This fact motivated researchers to design new algorithm and system in facial features detection and expressions.

The essence of facial features expression analysis is shape or appearance information. They must be extracted from an image and normalized. Thus, they used for classification of emotion in data base. Finally most of the systems have been used in different training algorithm for detecting human facial coder and evaluated accuracy of the system.

Furthermore, number of sources were useful in learning about the facial features expression and connected to non-verbal communication, verbal communication, mental states and psychological activity. As you can see in Figure 1-5 emotions are not the only source of facial emotion expression.

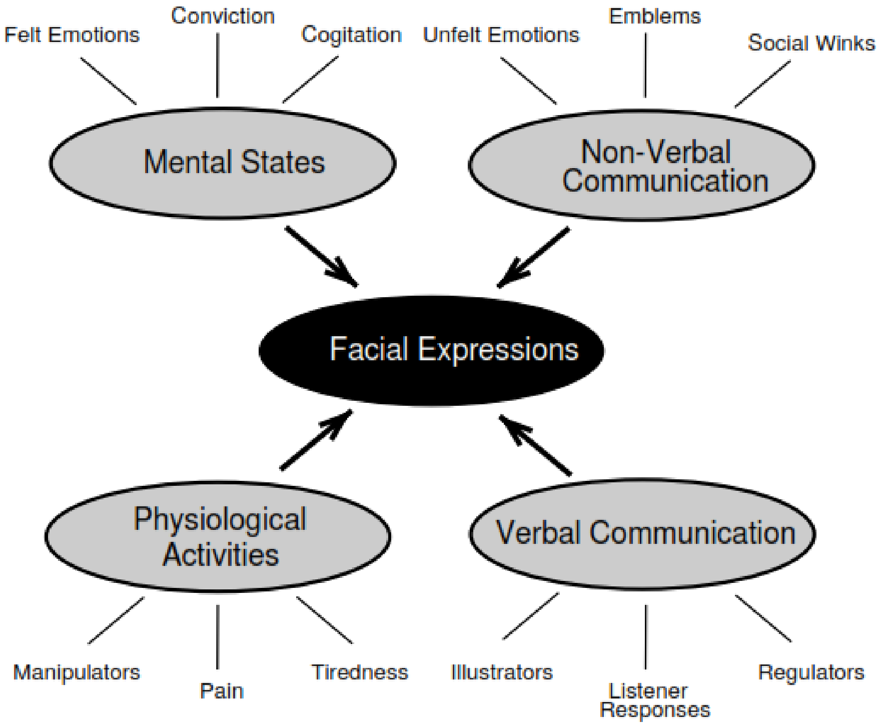


Fig. 1-5 Different source of facial expressions.

Arguably the first survey of the field facial expression recognition published in 1992 with Samal et. al. and has been followed continues with new approach in this field [86], [91]–[97].

We can be categorically divided facial features to transient and intransient base on actions of muscles. As shown in Figure 1-6 this category can help us for evaluating different model for facial features extraction [94].

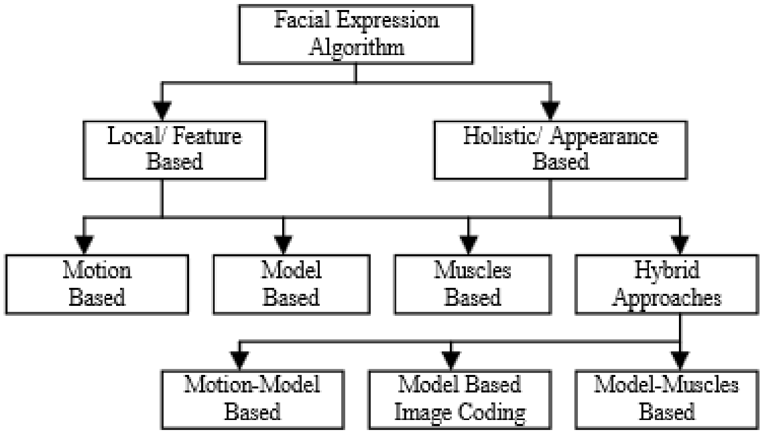


Fig. 1-6 Distinction of feature extraction and representation.

Furthermore, in an obviously related way, as you can see in Table 1-2 we overview of methods that used in facial expression analysis and computer vision community. In this table we divided the methods base on model base and image base.

Table. 1-2 Methods that used in facial expression analysis.

Deformation extraction	**Local methods**	**Holistic methods**
Model base	Two view point-based models [98]	Active appearance model [99]–[101]
	Geometric face model [102]	Point distribution model [103]
		Label graphs [104]–[106]
Image base		
	PCA + Neural networks [107], [108]	Neural network [109]
	Intensity paroles [110]	Gabor wavelets [97]
	High gradient components [112]	

Accordance with past research, feature extraction methods can be focused on deformation of faces or motion the facial muscles. Facial features expression and discreet emotion recognition from an face images has been promising approaches in past 10 years, particular in Action Unites (AUs) and Facial Action Code System (FACS) fields.

In more comprehensive researches to approach the facial expression recognition and emotion detection, researcher proposed the Action Unites (AUs) for classification and detection of the basic emotions [113]. For Example Table 1-4 lists an Action Unites for definition of each facial feature to synthesize emotional facial expressions through systematic manipulation of facial action units.

Table. 1-4 Action Units included in the facial features detection.

AUs	Description
1	Inner brow raiser
2	Outer brow raiser
4	Brow lower
6	Cheek raiser
7	Lid tightener
10	Upper lip raiser
12	Lip corner puller
15	Lip corner depressor
17	Chin raiser
18	Lip pucker
25	Lips part
26	Jaw drop

During the past decades, various methods have been proposed for facial emotion recognition algorithm. Significant fundamental work on facial features detection was done by Hefenbrock et. al in 2011. They proposed the Standard Viola & Jones face detection and recognition algorithm [114].

Numerous methods have been developed on face detection. Most of these techniques emphasize statistical learning. As you can see in Figure 1-7 the Viola–Jones face detector used facial point detection method.

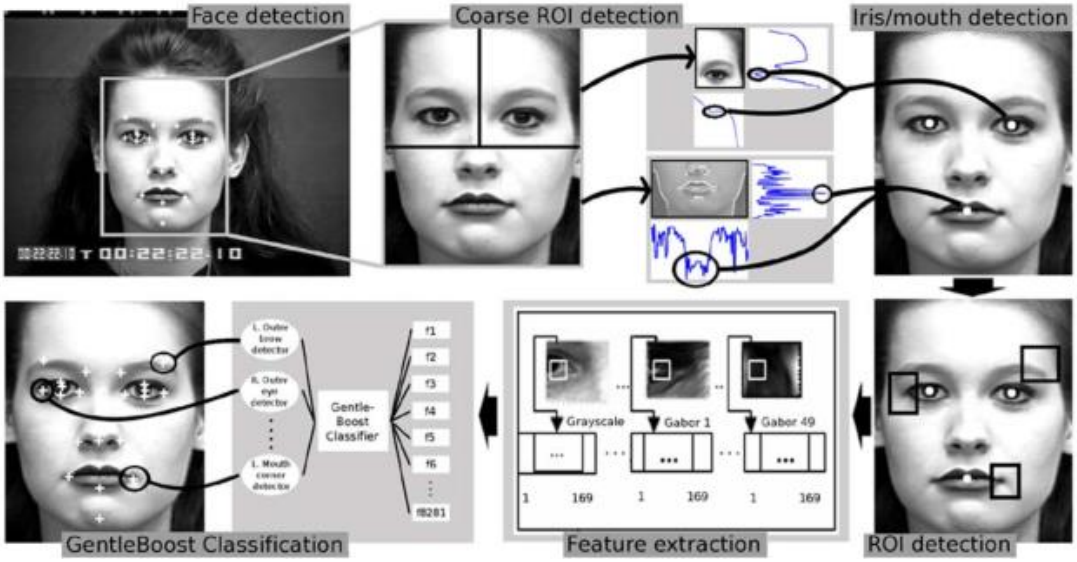


Fig. 1-7 Outline of the facial point detection method based on Viola–Jones face detector algorithm.

The essence of this approach was nearest with our idea. In this method the features are using Principal Component Analysis (PCA) for patterns appearance descriptors. As classifier they employ standard Support Vector Machines (SVMs) basis function kernel. Figure 1-8 have been given an overview of the baseline system's approach [115].

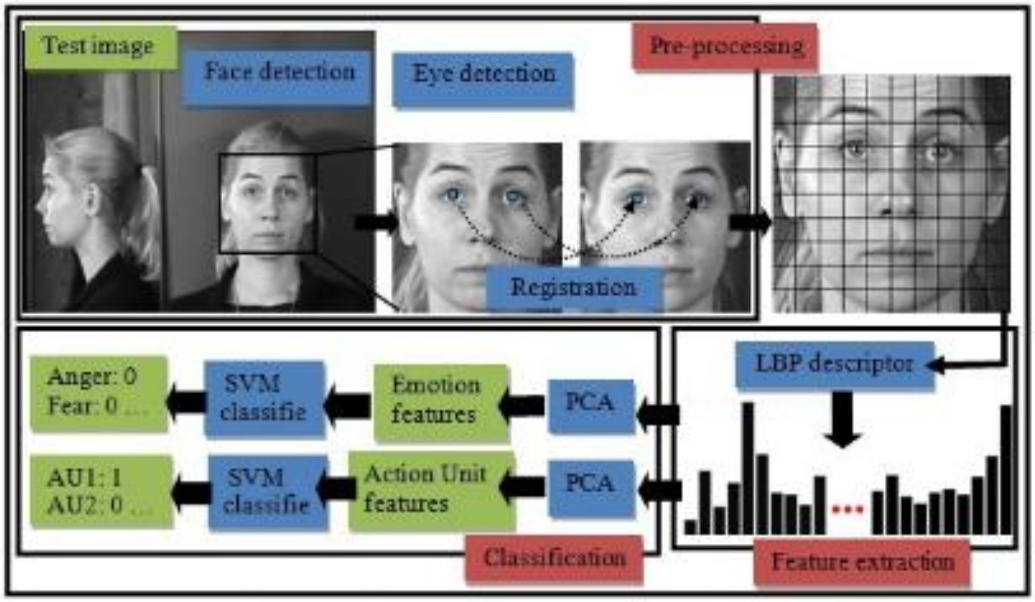


Fig. 1-8 Overview of the system based on PCA and SVM for detection of Action Units and emotions.

Furthermore, in an obviously related way, Maalej et al. completed this model and used the geometric features. They defined on the landmark points around the facial features (eyes, eye-brow and mouth) to represent the face images and then conducted the emotion recognition with various classifiers [115]. In more comprehensive research Zheng et al. used landmark points to represent the facial images based on Bayes error [116].

As mention above two main streams in the current research on analysis of facial expressions affect (emotion) consider facial muscle action detection and measurements Facial Action Code System [117][118].

Facial Action Code System associates facial expression changes with movement of the muscles. Additionally, as Cohen et al. shown in the Figure 1-9 , their program defines action descriptors and belonging to the upper or the lower face [119]. Facial Action Code system also provides the rules for the recognition of different Action Unites for classification of the results. In particular, Action Unites and Action Code System attracted the interest of computer vision and robotic researchers.

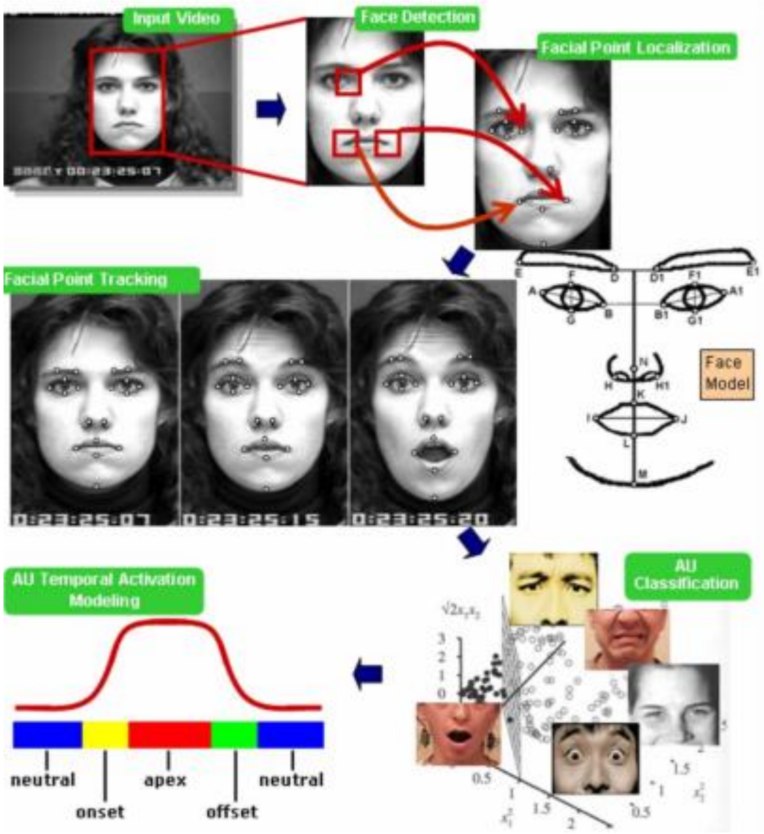


Fig. 1-9 Outline of the upper or the lower face system for recognition of AUs.

Exceptions from this overall state of the art in the facial analysis field include studies on 3D facial features expression and real time emotion recognition with hidden Markov models (HMMs), neural network (NN), Fuzzy Logic, Discrete Cosine Transform, Dynamic Bayesian networks and others [120]–[122].

For instance, Hoang Thai et al. proposed a solution for Facial Expression Classification using Principal Component Analysis (PCA) and Artificial Neural Network (ANN). For example Figure 1-10 show the main functionalities on Facial Action Coding System, Neural Network and Robot vision for facial emotion recognition [123].

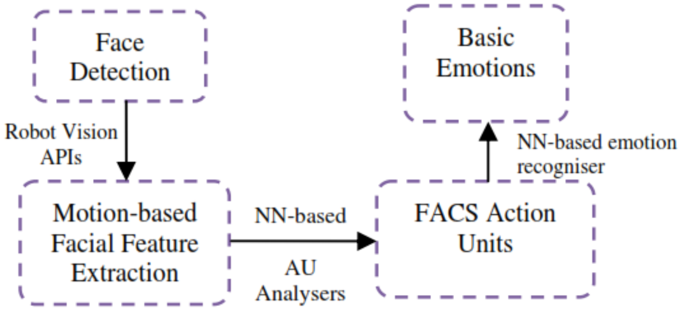


Fig. 1-10 The main functionalities for facial emotion recognition.

A number of related researches which identified the mapping between Action Units and emotional facial behaviors were also well established. Facial Action Coding System Affect Interpretation Dictionary, consider only emotion-related facial actions [124]. Ekman proposed the new method in (Facial Action Coding System techniques) for measurement the muscles movement and facial emotion detection. Figure 1-11 show the point and measurement the definition of distances principal points in the face.

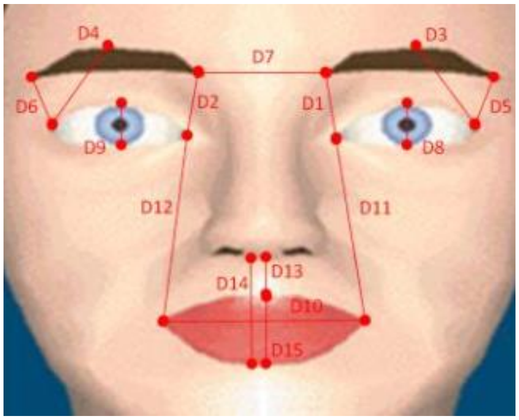


Fig. 1-11 FDPs used for recognizing facial expression and definition of distances.

An important functionality of user friendly interface robot will be the capacity to perceive and understand the human emotions as communicated emotionally with human by facial expressions. In more comprehensive research has been done on recognition and generation of emotions for social and humanoid robots [125].

The expressive behavior of robotic faces is generally connected to human robot interaction. In more comprehensive research instead of using mechanical actuation, another approach to facial expression is mechanical movement, computer graphics and animation techniques (see Fig 1-12) [126].

Vikia, for example, has a 3D rendered face of a woman based on different code of facial expressions [127]. Vikia's face is graphically rendered with many degrees of freedom and are available for generating facial expressions.

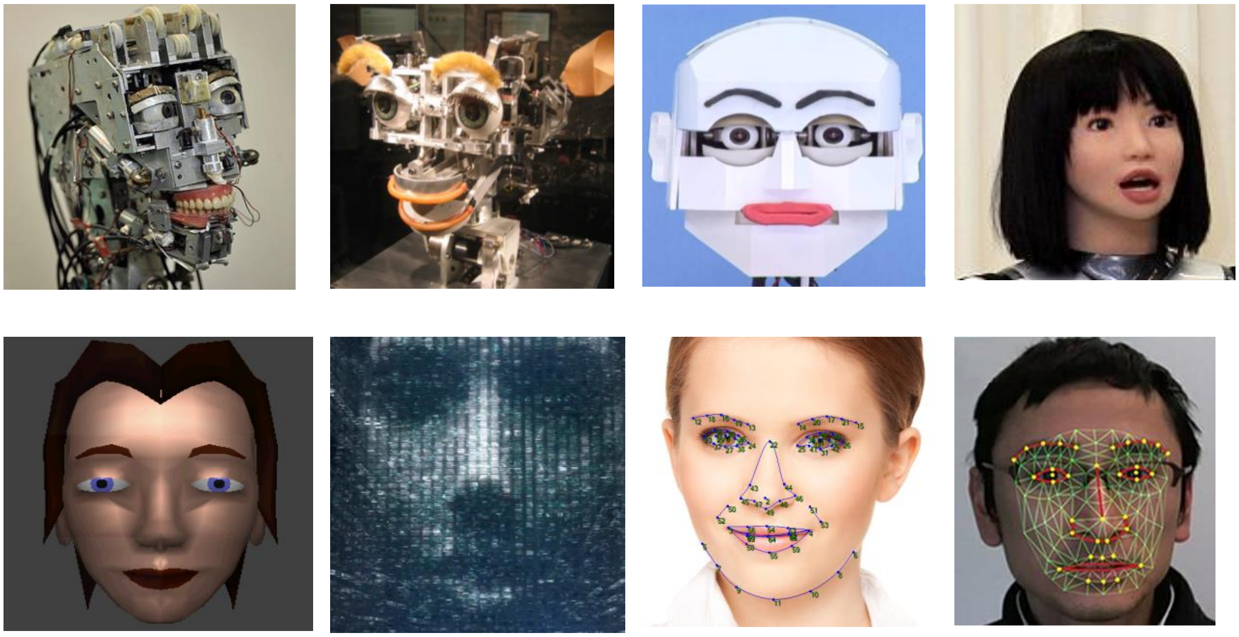


Fig. 1-12 Vikia computer generator face (first in the second row) and different face robots (University of Tokyo).

Obviously in a new filed of researches on human robot interaction, Zhang and Shrkay examined how the surrounding emotional context (congruent or incongruent) influenced users' perception of a robot's emotion. They founded when there is a surrounding emotional context, people will be better at recognizing robot emotions and they suggested that the recognition of robot emotions can be strongly affected by a surrounding context [128].

1-2-2 Facial feature extraction on basic emotion in the past 5 years

In recent years many researches have been done on emotions recognition in the laboratory such as facial expressions [129] , texts emotion recognition [130], slides emotion recognition [131], movies actors emotion recognition [132] and music signal emotion recognition [133] [134]. Among this filed, facial emotion recognition is the most important topic for researchers.

In more comprehensive research for extracted emotion from facial image, Majumder et al was using the geometric facial features for emotion recognition. They used Kohonen Self-Organizing Map (KSOM) to classify the features data into six basic facial expressions. The features data first of all clustered with KSOM, then the cluster centers used to train the data for recognition of the basic different emotions [135].

Wan and Aggarwal in 2014 did more comprehensive research on metric learning. They focused on issues that were still under addressed in the spontaneous facial expression recognition field. In comparative experiments they showed spontaneous facial expressions tend to have overlapping geometric and appearance features, making it difficult to find effective spontaneous expression recognition [136].

Zhang et al. inspired many researchers by proposing Hidden Markov Models and a new methodology for automatically recognize emotions based on analysis of human faces from video clip [118].

Significant fundamental work on facial features detection with Neural Network was done by Caridakis et al. They proposed an extension of a Neural Network adaptation procedure, for emotion recognition and training from different emotions. Their results shown that emotion recognition accuracy is improved by using Neural Network. After training and testing on a particular subject, the best-performing network is adapted using prominent samples from discourse with another subject, so as to adapt and improve its ability to generalize [137].

Action Units (AUs) represent the movements of individual facial muscles. Therefore, the combination of AUs produce facial appearance changes and meaningful facial expression [138]. Numerous studies have been done on emotion recognition by extracting Action Units (AUs) for frontal facial images extracted from video clip frames [118].

When an image sequence is presented to a facial expression recognition system, it is necessary to detect the facial regions as a preliminary pre-processing step. There are several methods which can be used to achieve this task. Valstar et al. employed probabilistic and statistical techniques, that used face image sequences for automatic classification of AUs [139].

The most extensive work about facial expression was done by Kharat and Dudul. Their research exploited and combined various feature extraction techniques such as Discrete Cosine Transform (DCT), Fast Fourier Transform (FFT) and Singular Value Decomposition (SVD) to extract facial features [140].

However, some researchers had reported geometrical facial feature point positions in 3D facial. Also, few studies uses Support Vector Machine (SVM) for classifies expressions in six basic emotional categories. Yurtkan and Demirel proposed a feature selection procedure base on Support Vector Machine (SVM) for improved facial expression recognition utilizing 3-Dimensional (3D) [141]. The system designed based on classifiers in two classes with 15 couple of emotion. As shown in Figure 1-13 linear kernel function and Support Vector Machine (SVM) have been used for classifiers basic emotion.

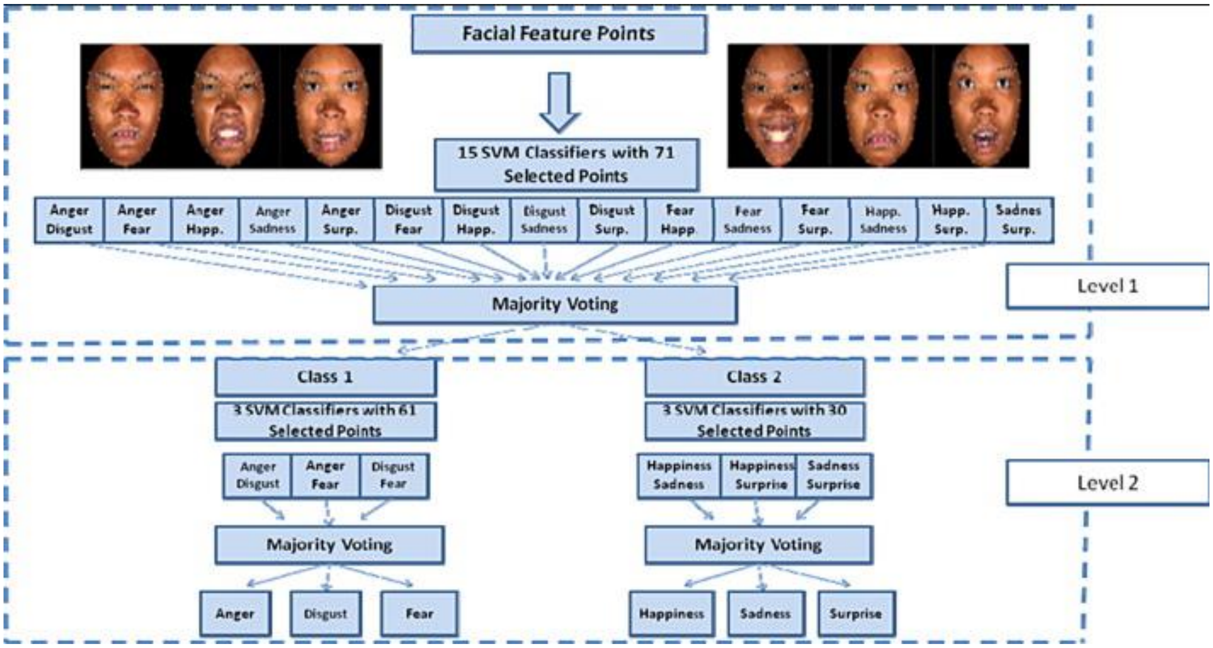


Fig. 1-13 SVM classifier system used for facial expression recognition.

Khan and Bhuiyan worked on the facial feature expressions, especially eyes and lips, which were extracted and approximated using Bézier curves. They defined the relationship between the motion of features and the change of expression. In this research, pictures of 200 individuals have been analyzed and classified into neutral and four basic emotions [142].

Gomathi et al. used Multiple Adaptive Neuron Fuzzy Inference System (MANFIS) [143]. They used Fuzzy logic system for classification of basic emotion. Crosier et al. proposed a method to segment the facial image into three regions. They first used a Local Binary Pattern (LBP), then extracted texture features and finally, they built a histogram descriptor [143]. McDuff et al. introduce an open-source tool that analyzed naturalistic combinations of dynamic face and head movement across large groups of people [144].

There have been many advances reported towards automatic facial expression recognition from 2D static images or 3D in real time [145]–[147].

Recent work on facial expression analysis in video was done by Sebe et al. They proposed a method that measured the distances between frames in 2D image. In image sequence, landmark facial features such as the eye and mouth corners are selected interactively and put mesh for each one [148]. Finally, the shape of the mesh can be changed by changing the locations of the control points and with data in database system can control the basic emotion in real time.

In most of this research, after the algorithm has detected the face, facial feature expressions are extracted and classified into a set of facial actions [149], [150].

Many techniques for facial expression recognition have been proposed, Wang et al. in 2009 extracted trajectories of the feature points contain both rigid head motion components and non-rigid facial expression motion components. The proposed system used the combination of hidden Markov models (HMMs), Facial Action Coding System (FACS) and Action Units (AUs) [151].

Ahmed Khan et al. propose a generalized automatic recognition of emotion in facial expressions for low spatial resolution in facial images and extended the results in human–computer interactions applications [152]. Chia-Te Liao et al. proposed a new technique based on graphical representation of face for emotion detection [153].

Some researchers had reported the system that uses the dynamic information of the facial features extracted from video sequences and outperforms techniques based on recorded static images. Hui Fang proposed a method for facial features extraction based on existing work are reliant on extracting information from video sequences and employ either some form of subjective threshold of dynamic information.

The innovation of this system was attempt to identify the particular individual frames for behavior [154]. Also, they proposed to warp the landmarks defined in Figure 1-14 to each face image after dense correspondences between sets of images are built.

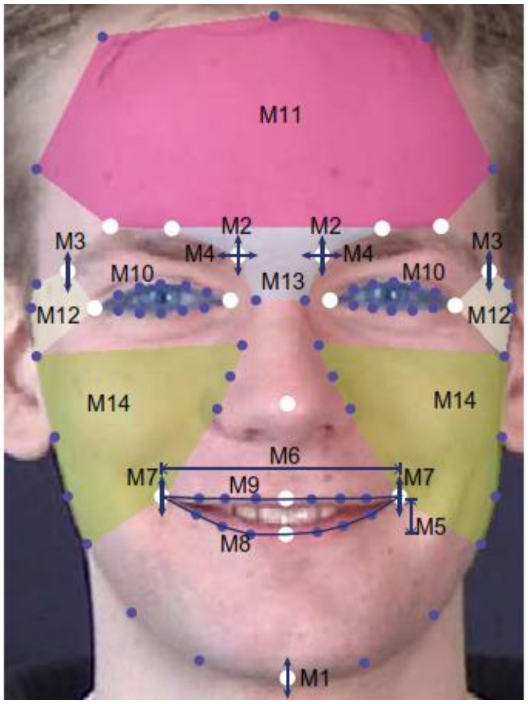


Fig. 1-14 Example of landmarks, geometric features and texture regions.

Attempts to facial emotion recognition have also been investigated the possibility to detect the three emotions happy, angry and sadness in video sequences by applying a tracking algorithm by Besinger et al. in 2010. They results shown that the used point tracking algorithm separately applied to the five facial image regions can detect emotions in image sequences [155].

1-2-3 Databases for facial feature extraction system

One of the most important aspects of developing any new recognition or detection system is the choice of the database that will be used for testing the new system. If a common database is used by all the researchers, then testing the new system, comparing it with the other state of the art systems and benchmarking the performance becomes a very easy and straightforward job. Researchers often do report on the accuracy of their proposed approaches using their training database or number of popular facial expression databases.

However, building such a ‘common’ database that can satisfy the various requirements of the problem domain and become a standard for future research is a difficult and challenging task. Therefore, the problem of a standardized database for face expression recognition is still an open problem. But as duplicated in Table 3 we compare different data base for choosing one of them for evaluating the new hybrid system.

When have been compared the face recognition system, face expression recognition poses a very unique challenge in terms of building a standardized database. There was a lot of facial recognition data base in internet library now such as: UMB database of 3D occluded faces, Vims Appearance Dataset (VADANA) for facial Analysis, MORPH Database (Craniofacial Longitudinal Morphological Face Database), Long Distance Heterogeneous Face Database (LDHF-DB), Photo Face for face recognition using photometric stereo, YouTube Faces Database, YMU (YouTube Makeup) Dataset, VMU (Virtual Makeup) Dataset, MIW (Makeup in the "wild") Dataset, 3D Mask Attack Database (3DMAD), McGill Real-world Face Video Database and Siblings DB Database. But the human machine interaction environment sometimes needs to define new database for recognition an emotion.

From the above discussion it is quite apparent that the creation of a database that will serve everyone’s is a very difficult job. However, there have been new databases created that contain spontaneous expressions, frontal and profile view data, 3D data, data under varying conditions of occlusion, lighting [156].

Table 1-5 Summaries of some of the facial expression databases that have been used in the past few years

Identifier	Emotional content	Emotion elicitation methods	Size	Nature of material
The AR Face Database [157] **The Ohio State University, USA**	Smile, anger, scream neutral	Posed	154 subjects (70 male, 56 female) 26 pictures per person	1 : Neutral, 2 Smile, 3 : Anger, 4 : Scream , 5 : left light on, 6 : right light on, 7 : all side lights on, 8 : wearing sun glasses, 9 : wearing scarf,
The Psychological Image Collection at Stirling [158]	Smile, surprise, disgust	Posed	116 subjects Nottingham scans: 100 Nott-faces-original: 100	Contains 7 face databases of which 4 largest are: Aberdeen , Nottingham scans, Nott-faces-original, Stirling faces
The Japanese Female Facial Expression (JAFFE) [159]	Sadness, happiness, surprise, anger, disgust, fear, neutral	Posed	10 subjects 7 pictures per subject	6 emotion expressions + 1 neutral posed by 10 Japanese female models
CMU PIE Database (CMU Pose, Illumination, and Expression (PIE) database) [160]	Neutral, smile, blinking and talking	Posed for neutral, smile and blinking	68 subjects	13 different poses, 43 different illumination conditions, and with 4 different expressions.
Indian Institute of Technology [161] Kanpur Database	Sad, scream, anger, expanded cheeks and exclamation.	Posed	20 subjects	Varying facial expressions, orientation and occlusions. All of these images are taken with and without glasses in constant background; for occlusions some portion of face is kept hidden and lightning variations are considered.
The Yale Face Database [162]	Sad, happy, sleepy, surprised	Posed	15 subjects	One picture per different facial expression or configuration: centre-light, w/glasses, happy, left-light, w/no glasses, normal, right-light, sad, sleepy, surprised, and wink
Facial Expression Database (Cohn-Kanade) [163]	Joy, surprise, anger, fear, disgust, and sadness.	Posed	200 subjects	Subjects were instructed by an experimenter to perform a series of 23 facial displays that included single action units (e.g., AU 12, or lip corners pulled obliquely).
The EURECOM Kinect Face Dataset (EURECOM KFD) [164]	neutral, smile, open mouth, left profile, right profile, occluded eyes, occluded mouth	Posed	facial images of 52 people (14 females, 38 males)	facial expressions, lighting and occlusion and associated 3D data
AT&T (formerly called ORL database) [164]	Smiling / not smiling	Posed	40 subjects	10 images for each subject which vary lighting, glasses/no glasses, and aspects of facial expression broadly relevant to emotion –open/closed eyes, smiling/not smiling

1-2-4 Face detection and facial expression APIs and software

There are many systems currently implemented aiming to improve the robustness and reliability of the facial emotion recognition procedure. However, current research in facial emotion expression focuses on computer vision and machine learning approaches to improve analysis of dynamics facial detection and recognition performance. But we decided to extend the aforementioned work by extracting and labeling precise feature boundaries on a frontal image.

There have been a lot of Face Recognition program and software namely: Skybiometry Face Detection and Recognition, Face++, Face and scene recognition, OpenCV Face Recognizer, OpenBR, FaceReader, Betaface API, Hunter TrueID and Bob [165].

In recent literature about affective image classification in computer vision, most of the researcher used the same strategy. As you can see in Figure 1-15 we duplicated to the different stages on automatic facial expression analysis system [86]. Base on the Figure 1-15 we decided to write our program with C++ language by means of cover all aspect of facial emotion recognition.

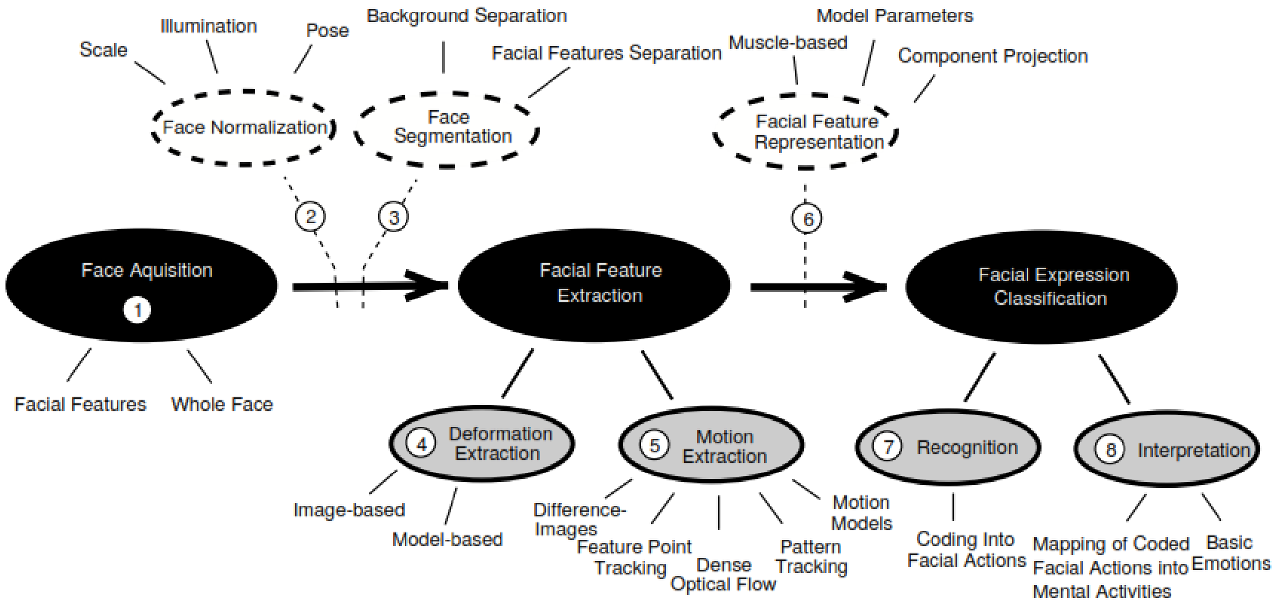


Fig. 1-15 Framework on automatic facial expression system.

Face tracking is a core component to enable the computer to "see" the computer user in a Human-Computer Interface (HCI) system. Also, Face tracking can serve as a front end to further analysis modules, such as: face detection, face expression analysis, face recognition and analyze human emotions from facial expressions. Chun Peng et al. used face tracking for affective image classification in computer vision. In the proposed system each emotion category independently and predict hard labels, ignoring the correlation between emotion categories.

The “Emotient” is software on emotion detection and accurate facial expression detection and analysis technologies. Two of the advisor of this system was the most prominent leaders in facial behavior Paul Ekman and Sejnowski. They system sets the industry standard for accuracy, with the highly precise ability to detect single-frame facial expressions of emotions [166].

The results of collaboration between Ekman and Sejnowski lead to propose a new technique for Computer Facial Expression Recognition Toolbox. CERT was an automated system for fully facial expression recognition that operates in real-time [167]. Thus, CERT automatically detects frontal faces in the video stream and codes each frame for recognition six basic emotions.

Computer Expression Recognition system was employed in pioneering experiments on spontaneous behavior, including Facial Action Coding System and 30 facial action units (AU’s) [168].

nViso is an industrial software analyze human emotions in Lausanne, Switzerland that developed a technology on facial expression [169]. The company explains that their technology can measure emotions based on automated facial expression recognition and eyes tracking. nViso's sophisticated artificial intelligence algorithms automated capture hundreds of measurement geometric of the face points and facial muscles tracking in real-time. The proposed system precisely decodes facial movements based on Ekman's Facial Action Coding System into the underlying expressed emotions. Deploy nViso technology offline or online and provides API’s for tablet, smartphone and computers for human emotion recognition. This program was named a winner of a 2013 IBM Beacon Award for Smarter Computing.

1-3 New hybrids model on emotion recognition system

With the advance in computer vision and sound analysis, robots with human shape and perception enter everyday life and start to help a human being, receive human orders and optimize their behavior. Emotion makes a proper mutual communication between human sensation and robot interaction. The most natural way to recognize basic emotions is extracting set of features from human speech, facial expression and body gesture.

Mehrabian as phycologist in 2009 proposed a new method about human emotion. In that research they found when people have communication with each other, 55% of the message is conveyed through facial expression alone, vocal cues provide 38% and the remaining 7% is via verbal cues [170].

Thus, the most new method for recognize human emotions is design a new hybrid model to extracting features in speech communication (pitch, intensity and formant), facial features extraction and attention the movement of the facial muscles. As noted in most of researches the essence of hybrid emotion recognition system is based on feature-level fusion of acoustic and visual cues.

Considering Audio-Visual modalities are already widely used portable devices designing of the accurate software on emotional communication between human and machine is more important.

Several research have been done on speech analysis and facial expression that focus on describing the feature extraction methods in classification techniques for recognition of emotions [171][172], [173].

Koda investigated a new method for human facial emotion recognition by using both thermal image processing and speech [174]. Accordingly, a lot of researches have been validated the model and software for combining speech recognition , facial expression and body gesture [175] [156].

Numerous studies have been done on emotion recognition by extracting speech analysis, Action Units (AUs) and facial appearance changes for frontal view face images extracted from video clip frames [176].

Accordance with past research we decided to extend the aforementioned work by using acoustic and phonetic properties of emotive speech with the minimal use of signal processing algorithms and extracting precise feature boundaries on a frontal image. Then, have been measured the discrepancy between the extracted features in the given image and the features extracted from a reference image, which contain the “neutral” expression of the same face. This step is based on Action Units (AUs) and

Action Code system (ACS). For each AU, an Action Code (AC) is defined; in this way, faster processing and reduction of computation time can be achieved. Finally, the emotion will recognize by using a hybrid algorithm that combining speech graph and facial features extraction.

This thesis is organized as follows: we first summarize the related studies and general description in this field. In chapter 2 we give a general description of the methodology and the hybrid algorithm. Chapter 3, we focus on the sound and facial feature extraction from the experimental tests. Then we suggest a set of rules for classification of the emotions.

Finally, experimental results of the application in the methodology are presented and discussed and suggest how we can implementation this project in the real uncertainty world.

CHAPTER 2

Ontology, Methodology and scenarios

For

Emotion recognition system

Humans belong to various ethnic groups with different accent and attributes of facial features (shape, color and size). Also, members in the same society have diverse behavior and emotion expressions, depending on culture, age and gender. Understanding the mechanism of emotion in natural scenes must be developed for a machine to understand the human emotions in various environments. Especially for human-machine interaction, we combined simple speech understanding and standard dialogue management functions with facial features expression for recognition the basic emotion in service robot. On the other hand, it is very difficult to determine the boundaries between the basic emotions (happiness, anger, fear, sadness, surprise and disgust) in human behavior.

In this work, a framework is proposed to architect human-robot interaction for a service robot based on an interactive decision making system, psychological processes underlying social perception, evaluating (sound signals) plots, mathematical formulation, cognition and action unites controlling in realistic settings (in laboratory).

Five core steps for the design of new methodology in human robot interaction was;

- Focus on ontology on human robot social interaction by means of building a hybrid model,
- Analyzing natural human communicative signals (verbal),
- Analyzing human facial expression signals (non-verbal),
- Combination of two proposed model on a robot software platform,
- Extending the new model to manage testing, learning, training and evaluation of the system accuracy.

As mentioned above and the goal of this research, the new method for recognize human emotions is extracting features in speech communication (pitch, intensity and formant) and movement of the facial muscles Action Unit System (AUs).

This capture is organized as follows: first we summarized the ontology of robotic and general description in this field. Then, designed the methodology and the algorithm on speech emotion recognition system and facial features extraction. Also, we focused on the new hybrid algorithm to recognize the emotion in communication. Finally, extracted rules by means of design a new software program from the experimental tests.

With respect to other works in the scientific literature, the methodology we proposed was very suitable for implementation in real-time systems, since the computational load is very low indeed.

2-1 Ontology for human robot interaction

The structure in the presented ontology for human-robot interaction consists of emotion recognition model. The human-robot interaction ontology explains how the knowledge of these fundamental sciences is applied in physics (sound analyses), mathematics (facial features extraction, detection and classification), philosophy theory (human behavior) and robotic science context. For instance ontology makes possible much greater inter-operability between sensing and architecture for robotic system [177]. Ontology typically serves two purposes:

1. They provide agreed unambiguous terminology for a domain, with that goal of human expression and transform their knowledge more effectively and accurately.
2. Ontology's allows developers use t background of the knowledge.

Robot needs to establish appropriate correspondences between behavior and actions in the prototypes and their counterparts. In this way the robot can perceive and categorize.

Ontological knowledge would also be used for design of robotics systems when selecting and matching components and distinguishing properties of a particular robotic system or application.

In other words, it encodes the semantics of meta-level concepts and domains of human-robot interaction. Figure 2-1 shows an architecture design of human-robot 3 in columns. The first column shows the aspect of the interaction-robot architecture, second column shows the context of the design and third column on the right shows the system-level of the architecture.

2-1-1 Ontology for concepts layer

Visual perception and information fusion is one important capability of a humanoid robot to sense concepts of its environment. Concepts are the basic mechanism for using the ontology. Implemented in complex autonomous systems in (M3= meta-meta model), and their empirical evaluations are key techniques to understand and validate concepts of intelligence. Meta-meta model represents the knowledge of the properties and relationships in the model with terminology and syntactical constructs [178]. At this level we have highest level of abstraction and contain representations of devices, environment and tasks. Beside a description of the general concepts, we will focus on aspects of perception, behavior architecture, and reinforcement learning.

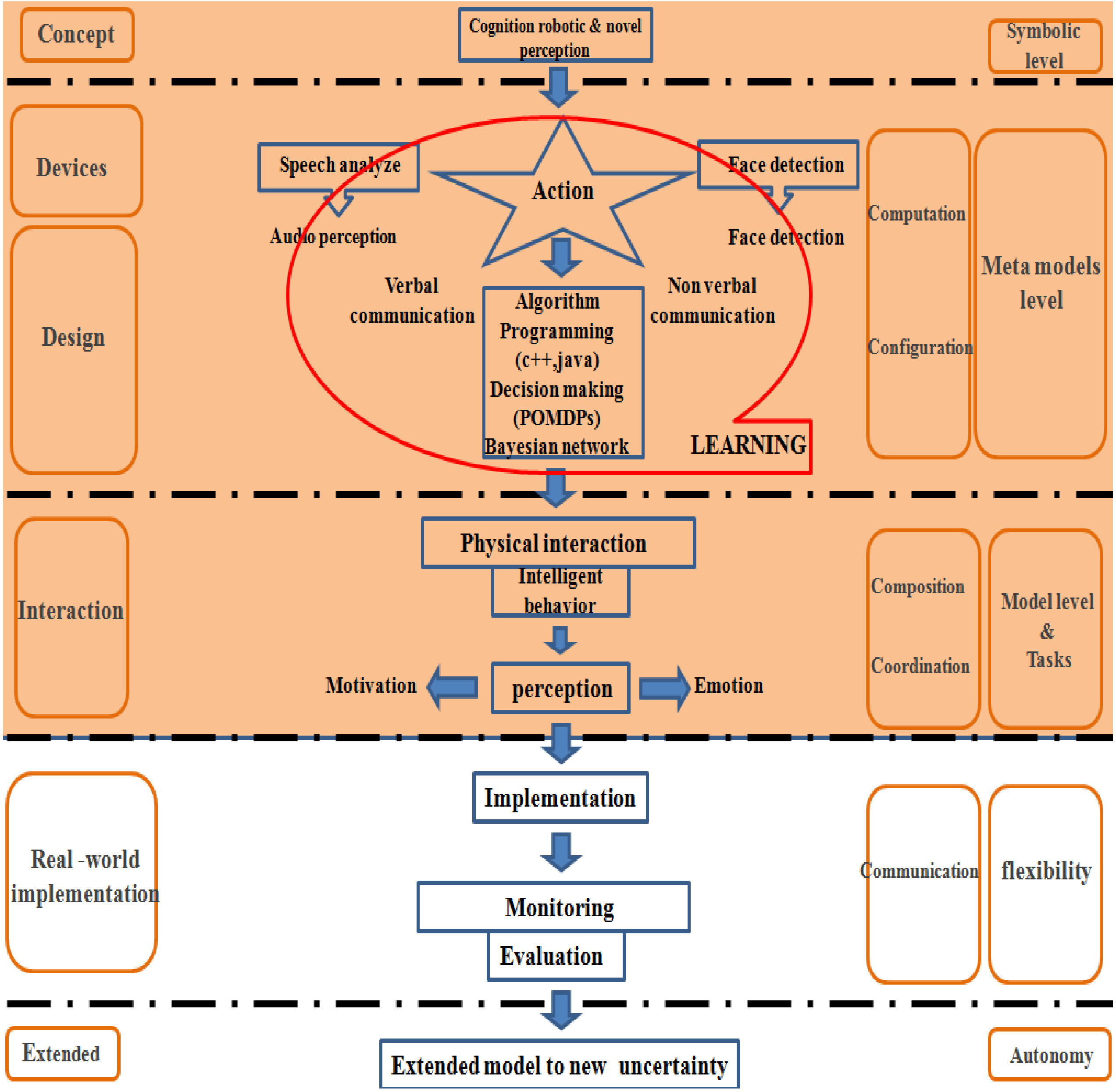


Fig. 2-1 Map of architecture, relationships between the architecture, design and ontology.

2-1-2 Ontology for device layer

New technologies in material and computer science emerged in the past few decades and enabled robotics researchers to develop realistic looking humanoid robots [179]. They underestand face to face conversations amongst humans is the most natural way of communication and recognition the emotional states.

A common theme in M2 (Meta model) implementation as device and discussion of social robot designs, their success as applications refers to the “human like” characteristics (motivation, emotions and personality) and skills (speech communication, facial expression and gestures) of the robot. Calculation and configuration of contexts define values of the properties, risk cost, number of component and relationship between the devices.

2-1-3 Ontology on speech recognizer

The speech signal is the fastest and the most natural method of communication between humans and robot. Figure 2-2 shows the speech emotion recognition architecture that combines the linguistic and acoustic models by means of checking the most probable word sequences.

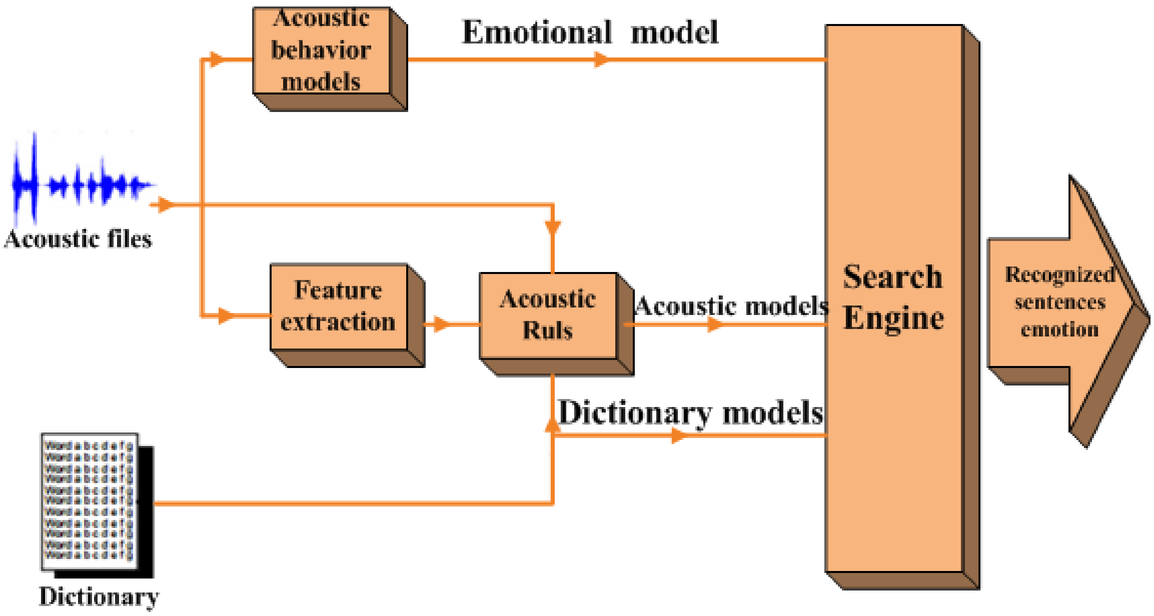


Fig. 2-2 The architecture of a speech emotion recognition engine combining acoustic features and behavior.

For the speech recognition system, a commercial engine uses a Semantic Context-Free Grammar (CFG) that specifies two types of information:

• Regular expressions that define the words and the syntactic rules that combine these words into a set of sentences called Regular language. This language specifies the utterance that the robot is able to recognize. This part of the grammar is called literal.

• Semantic attributes. This is the semantic part of the grammar.

In other words, it is possible to modify more regular expressions, that is, more recognizable words and syntactic rules, and to modify and add more semantic attributes with their values and this will allow to label new built skills.

The human speech communication consists of two channels, the explicit channel carrying the linguistic content of the conversation (‘‘what was said’’) and the implicit channel containing the so-called paralinguistic information about the speaker [180].

Our ambition is to establish a general model for the symbol-level dialogue and behavior controller of robots/agents that can engage in analyzed dialogue with humans to understand their requests and give useful information as well as perform desired emotional behaviors.

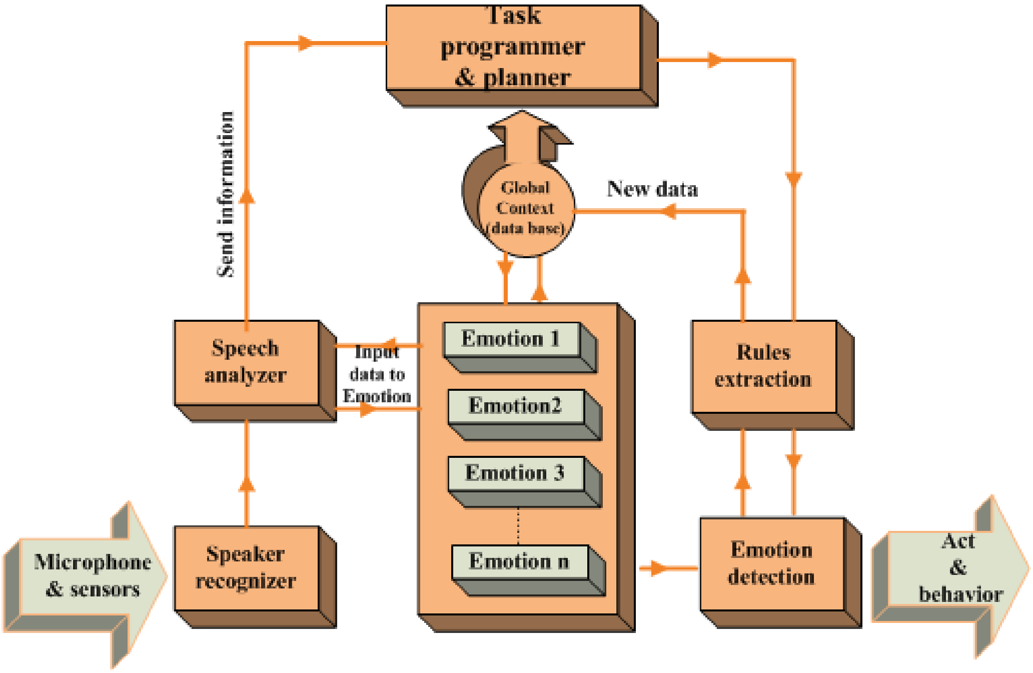


Fig. 2-3 Automatic speech recognition (ASR) and storage in global context (data base).

In case of the robot utilization, the speech and human body activity recognition as well as speech synthesis and overall robot posture considers important parameters. Figure 2-3 shows that the speech recognition system is realized by using an onboard microphone speech recognition engine, modified to deliver discrete probability distributions.

Each expert hold information on the progress of the primitive task. They classified into task type independent information. Task type independent information includes information such as which action in this primitive task is being performed, and whether the previous robot action selected, information and action storage in data base to reuse this data for future.

2-1-4 Ontology for facial detection

Face detection for different ethnic group is a fundamental problem in many computer vision applications. The challenge of detecting human faces from an image mostly comes from the variation of human faces such as races, illumination, facial expressions, face scales, head poses (off-plane rotations), face tilting (in-plane rotations), occlusions [181]. Also, environment issues such as lighting conditions, image quality, and cluttered backgrounds may cause great difficulties.

The positive and negative feature patterns are important features to face detection. However, it has some limitations. First, if an image is too blurred or lack of important facial features, we may not be able to extract the features to recognition an emotion. Another limitation is that the feature patterns may change if the poses of faces are largely changed. We proposed method to use not only facial components in terms of edges but other information (movement of muscles) to detect or recognize human faces will be considered.

2-1-5 Ontology for design layer

Design layer as (M1=Model) describe computational concepts of cognition that were successfully implemented in the domain of human- robots. Design layer development the interactions between cognitive concepts, software engineering and system implementation.

As it can be shown in Figure 2-4, some details are relevant concerning the capabilities of respective skill-activity and task-activity. Robot architecture system in this layer divided to; decision making, features extraction for classification of emotion. In this layer system must be check top to bottom algorithm to exchange tasks to skill.

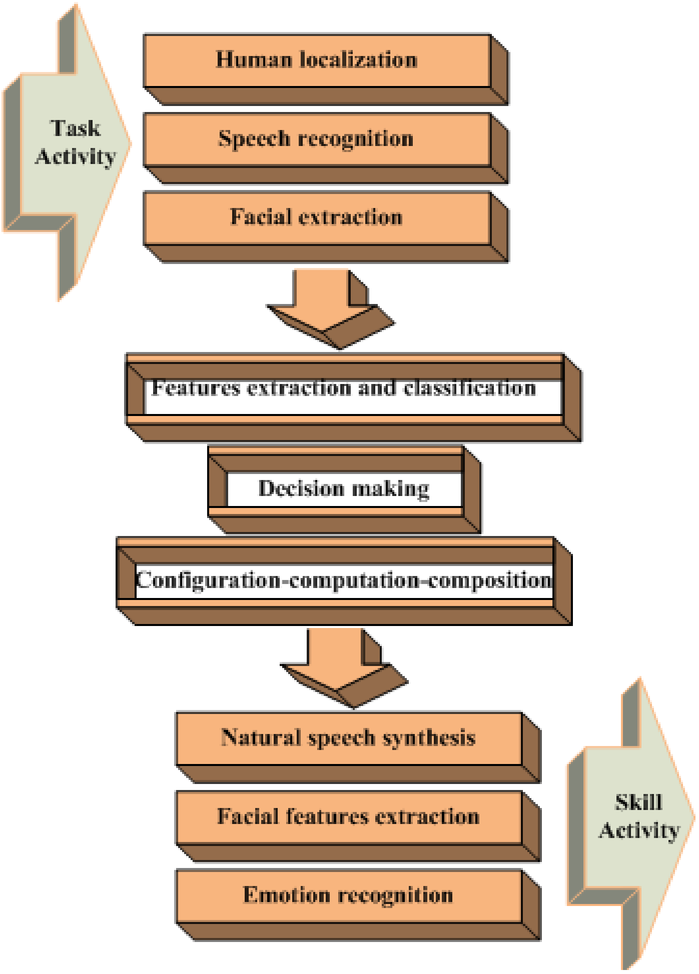


Fig. 2-4 The HRI system architecture exchange task to skill action from top to bottom.

One central property of cognitive systems is the ability to learn and improve the system continually. Also, beside a description of the general concepts of perception, behavior architecture and reinforcement learning are to be focused.

Human robot interaction system combines programming and training in order to learn constantly and make learning executable in the normal robot program [182]. Additionally, human robot interaction system needs prediction models on human's behavior and acting in the real-time environment.

Robot has to understand the human features and transfer this information to its own control system, based on a learning methodology. The domain knowledge in the current learning algorithm was formalized by two simple constraints: a range of a parameter (features behavior), and relative relationships between different parameters (different behavior) and action.

Learning consists of; abstract, semantic descriptions of manipulations, design program and algorithm extracted by a robot. In the learning part we must answers to these questions:

- How can robots realize “Human emotions”?
- How can robots perceive the speech, facial features and behavior?
- How do we model, specify, classify and control (behavior and action)?
- How can coordinate robots intelligent and human emotion in communication together?

The proposed system in behavior phase show one of the basic emotion (happiness, surprise, anger, fear, sadness, disgust and neutral). Also in different behavior phase just shows this emotion is unknown and then we can insert information about this emotion manually.

2-1-6 Ontology on interaction layer

For human-robot interaction, M0 level (implementation layer) used as concrete robotic system-level. All the knowledge and details are relevant concerning the capabilities of perceptive compounds and actuators.

Interaction layer makes contribution toward identifying and formalizing the relationship between domains of the system-levels and ontology in human-robot interaction.

The intensity of the coupling between device and environment is necessary to accomplish the task and degree of interaction. In other words, the action of the devices changes the environment and increases the interaction. As it can be seen in Figure 2-5, behavior controller receives output from sensor interpreters such as: speech recognizer, responsible for selecting utterances and facial features expression to perform.

On the other hand in human robot emotion recognition system algorithm must be developed an advanced in social communication that includes emotional behavioral and cognitive interaction.

An ontology of robotic enabling humanoid robot to used clear map and decision making in mission scenarios within uncertain environments. Also, the emotional activations and cognitive events of human-robots have important role for decision-making and long-term deliberative process planning with humans. The perceived emotion is subjective and highly dependent on pre-emotional state, environment and culture of utterances.

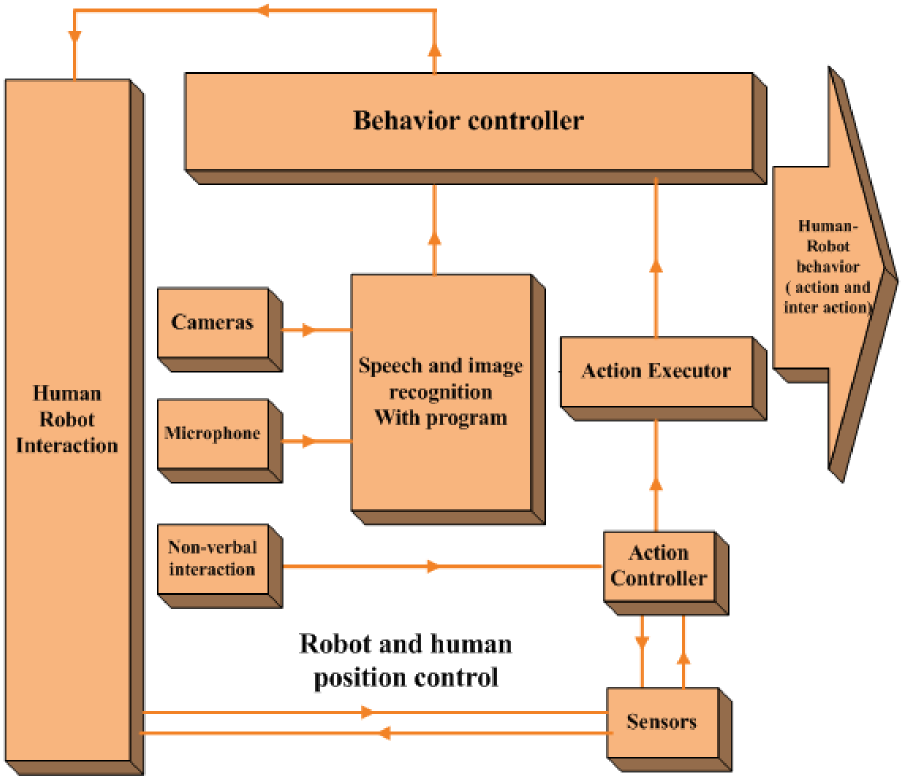

Behavior controller
Human Robot Interaction
Cameras
Microphone
Non-verbal interaction
Speech and image recognition With program
Action Executor
Human-Robot behavior (action and inter action)
Action Controller
Robot and human position control
Sensors

Fig. 2-5 Behavior controller system to connected the action and behavior.

Classical models of emotion recognition system consider the interaction aspect of emotions (for instance the emotions conveyed by the speech signals and facial expressions), but in centralist theory scientists believe that emotion is the result of a brain process.

The Model and computational architectures of emotional systems are categorized into two different families: firstly, the models devoted to expression of an emotional state and generally to the control of expressiveness. On the other side, the models devoted to the autonomous learning and Meta control.

Emotional recognition system core is discrete state-space transitions from all part of the human body that combine and express emotional state. The training of cognitive systems for such advanced applications in human robot interaction requires the development of mixed real virtual training environments.

Finally, we investigate creating large knowledge bases of model build blocks, including transition from observing human behavior by using studies coupled to programming by demonstration and future learning from observation, larger dialog and interaction methodologies. As mentioned above Real-world usually cannot be more or less hospitable to the emotional interaction between human and robot. Hence social robots are most often evaluated in the laboratory environment. Any implementation of an architecture human-robot interaction requires at least, two important choices to be made, at design time:

Which representation use to store the data and information?

Which software to use to support humans and computer to work with the stored knowledge?

Both choices come on the real hard part of the architecture process and relationship between algorithm, software program and control mechanism. We suggest that an effective and the only way these emergent capabilities can be evaluated is to take the robots in the real word with new algorithm and the proposed system did not connected directly to environment.

The system for recognition of emotion in human-robot interaction must be learned and updated continually individual skills (speech recognition and facial features extraction) and behavior (pre emotional state and non-verbal action).

2-2 Methodology on speech emotion recognition system

Emotion recognition can have interesting applications in human-robot interaction. Human-robot interaction will normally take place in the real world.

When a speaker expresses an emotion while adhering to an inconspicuous intonation pattern, human listeners can nevertheless perceive the emotional information through the pitch and intensity of speech. On the other hand, our aim is to capture the diverse acoustic cues that are in the speech signal and to analyze their mutual relationship to the speaker's several basic emotions, namely sadness (SAD), anger (ANG), surprise (SUR), fear (FEA), happiness (HAP) and disgust (DIS), based on the analysis of phonetic and acoustic properties.

An experimental methodology was set up, consisting of three different databases that built from speakers of different areas of the world. The first database includes ten European adults in the age group from 25 to 35 years (5 women and 5 men; mean age 29) from different countries of the European Union (Spain, Italy, Belgium, Romania, and France), the second group contains ten Asian (Middle East) adult speakers in the age group from 19 to 45 years (5 women and 5 men; mean age 31) and the third database contains recordings from ten American English speakers in the age group from 21 to 38 years (4 women and 6 men; mean age 28).

Six simple sentences of everyday life were chosen in learning phase, namely: "What are you doing here?"- "Are you ready?"-"Come in"-"Thank you"-"You are welcome"-"Where are you?" The participants to the experiment had to repeat these six sentences for three times with a neutral (NEU) intonation, with 1 second of interval between each sentence, in order to distinguish rising-falling intonations and pitch movements. Also, five simple sentences of everyday life were chosen for testing phase, namely: "Hello"- "Good noon"-"It's a sunny say"-"Where do you live, sir?"-"What are you doing in the street?"

Then, every participant had to repeat again three times the same six sentences, but with each one of the emotions listed above. All the sentences were recorded, thus obtaining 630 files, which were input to a dedicated program for speech analysis (Figure 2-6), which could provide the intensity, pitch (peak, range, values) alignment and speech rate for all the sentences.

The program used for speech analysis is the standard PRAAT program. In sub section 2-7-2 explain about this software completely. PRAAT is an open-source software that is extensively used by researchers in the field of speech analysis [183].

The technique for emotion recognition proposed in the speech emotion recognition system is based on two steps; namely: 1) feature extraction and 2) rules definition.

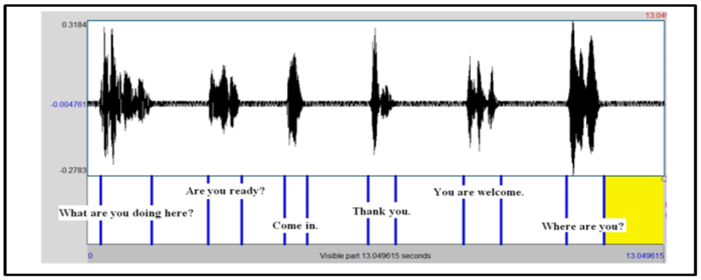


Fig. 2-6 Example of speech analysis with the PRAAT program.

A block diagram that proposed for automatic multi-level emotion recognition system can be seen in Figure 2-7.

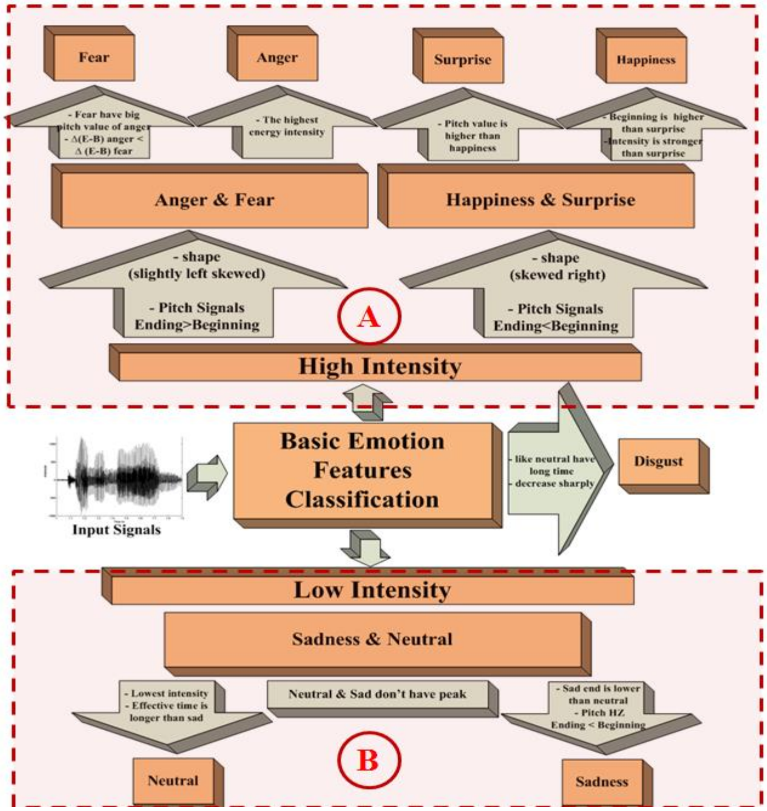

Fear
Anger
Surprise
Happiness
- Fear have big pitch value of anger
- Δ(E-B) anger < Δ (E-B) fear
- The highest energy intensity
- Pitch value is higher than happiness
- Beginning is higher than surprise
-Intensity is stronger than surprise
Anger & Fear
Happiness & Surprise
- shape (slightly left skewed)
- Pitch Signals Ending>Beginning
- shape (skewed right)
- Pitch Signals Ending<Beginning
A
High Intensity
Input Signals
Basic Emotion Features Classification
- like neutral have long time
- decrease sharply
Disgust
Low Intensity
Sadness & Neutral
- Lowest intensity
- Effective time is longer than sad
Neutral & Sad don't have peak
- Sad end is lower than neutral
- Pitch HZ Ending < Beginning
B
Neutral
Sadness

Fig. 2-7 Model of the system for speech emotion recognition.

The input of the system is the file obtained from the PRAAT software. The algorithm described in the foregoing analyzes the plots of pitch, intensity, formant and speech rate, thus recognizing if the emotion belongs to the category of “high intensity or low intensity”. If the speech falls into the “high intensity” category, it is further analyzed in order to distinguish between fear, anger, surprise and happiness. In the same way, if the speech falls into the “low intensity” emotion, it is further analyzed in order to distinguish between neutral and sad. We can use speech rate and the graphs of pitch signals in low intensity categories to distinguish between neutral and sadness emotion.

The novelty of methodology that we propose is makes minimal use of signal processing algorithms for features extraction and emotions classification. It was possible to successfully recognize the basic emotions in most of the cases.

2-3 Methodology on facial feature extraction system

Facial emotion recognition can have interesting applications in Human-Robot Interaction. Initially, researchers classified human emotions into two categories; pose-based and spontaneous expressions [184]. Pose-based expressions are artificially made by people, when they are asked to show some special emotion; on the other hand, spontaneous expression are made by people spontaneously, such as during conversations. The proposed methodology on facial features extraction consists of three steps in facial emotion recognition, namely: (i) detecting the facial region within the image; (ii) extracting the facial features (such as: the eyes, eyebrows, nose, mouth, lips, head position and etc.); and (iii) classifying the features alignment.

The proposed emotion detection algorithm is based on Action Units (AUs) and mathematical techniques that are relevant to geometrical features of the face parts. An overview of the methodology for extracting and classifying the facial emotions is shown in Figure 2-8.

In this work, in order to achieving the fully automated system for facial detection and expression needs various pose-based of 2D emotional facial images in the recording, training and evaluating phases.

By means of implementation of the proposed algorithm firstly, we used a webcam and a storage memory device for acquisition of pose-based human images. Then, the system distinguished region between facial and non-facial. By means of detecting the facial, the system used skin texture segmentation and different filter for determined border between two reigns.

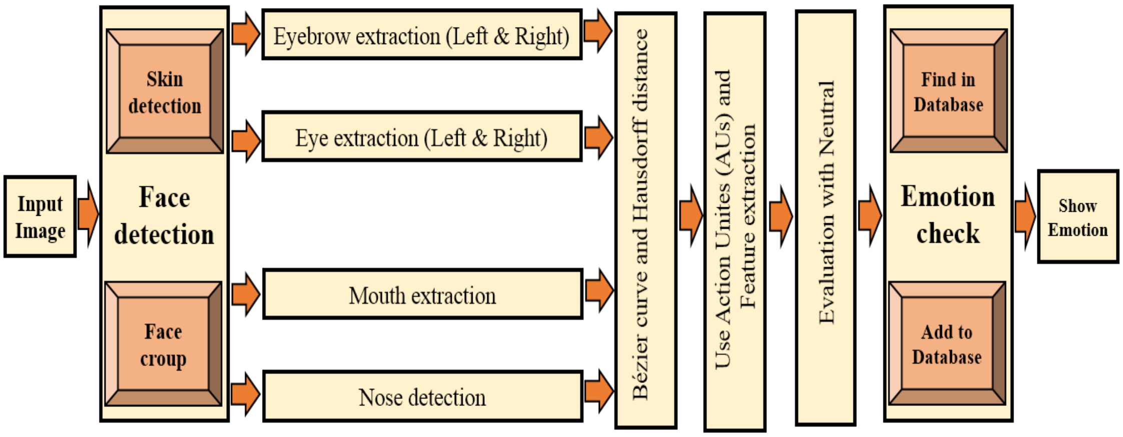


Fig. 2-8 Diagram of the proposed methodology for facial features expression.

The methodology for detection of the skin color that have been used in this project transformed the image from the RGB into black and-white pixels (skin pixels are set to white and the rest of them is set to black pixels). The next step was to extract facial features from the binary image. This could be done by histogram and facial geometric analysis, or by filtering the binary image (there are many different edge detection filter such as: Sobel, Canny edge detector, Scharr, Laplacian filters, etc) [185], [186]. For instance, in the proposed algorithm hairs are detected as a set of black continuous pixels and the skin appears when the color of the pixels changes (black to white).

In Figure 2-9, the original image is shown on the left side and the generated binary image, with Sobel filter and edge mask, is shown on the right side.

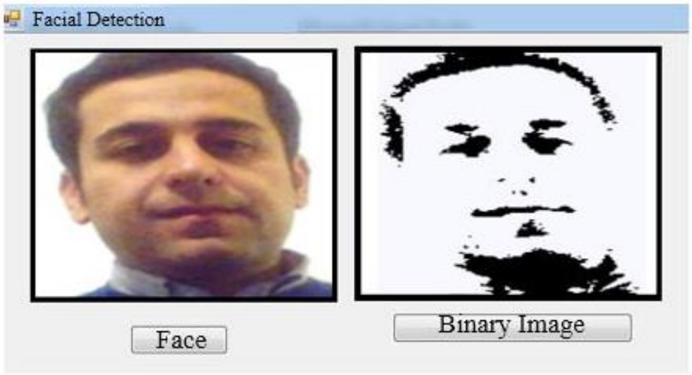


Fig 2-9 Original image (left side) and the binary image (right side)

Secondly, we focused on facial Principal Component Detection (PCD) such as: eye left/right, eyebrow left/right, mouth, lips and nose. For each of these components we determined the approximate position by framing it into a rectangular boundary. Figure 2-10 shows the results of PCD.

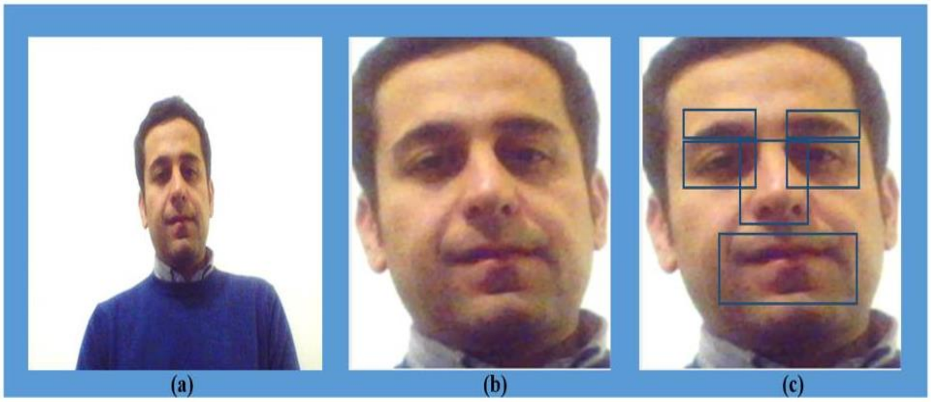


Fig 2-10 Face detection steps: (a) full frame photo (b) Zoom on face (c) feature candidate areas

Then, we interpolated the extracted facial features by means of Bézier curves and we defined the Action Code System (ACs) corresponding to each facial feature [188]. The last step of the algorithm was to recognize the emotion by analyzing the ACs and the Bézier curves. In section 2-6 we extensively explain about the Bézier curves and Support Victor Machine (SVM) for classification of the emotion in proposed system.

The facial expressions have been recognized form static image. We made a database with a set of 840 images with various basic facial emotions, and we used it as a training database. For the experimental tests like speech analysis. We used 30 individuals (15 female and 15 male, 20 to 48 years old) from different ethnic groups, namely: (i) European, (ii) Asian (Middle East) and (iii) American. This work, includes ten European adults in the age group from 25 to 35 years (5 women and 5 men; mean age 29) from different countries of the European Union (Spain, Italy, Belgium, Romania, and France), the second group contains ten Asian (Middle East) adult in the age group from 19 to 45 years (5 women and 5 men; mean age 31) and the third database contains ten American in the age group from 21 to 38 years (4 women and 6 men; mean age 28).

The efficiency of the methodology was evaluated by comparison of the results with those obtained by using the Cohn-Kanade AU-Coded Expression standard Facial database.

2-4 Hybrid algorithm on emotion recognition system

In emotion recognition field most of the researcher believe that with hybrid algorithm we can yield better result in human robot interaction. This section shows how the algorithm can be extended and decreases the time for recognition of emotion.

The proposed hybrid methodology consists of five steps in human emotion recognition, namely:

(i) analyzing human communicative signals (pitch, intensity, formant, speech rate and voice quality), (ii) detecting the facial region and extracting facial features (such as: eyes, eyebrows, nose, mouth), (iii) extending the rules to manage learning, (iv) recognizing the emotion, (v) training and learning the new emotional data in the database system.

The proposed emotion detection algorithm is based on pitch (peak, value and range) graph analysis, intensity and speech rate calculation, Action Units (AUs) and mathematical techniques that are relevant to action codes distance and geometrical features of the face parts. The algorithm described in the foregoing speech analyzes the emotion that belongs to the category of “high and low" intensity. An overview of the methodology for extracting and classifying the emotions is shown in Figure 2-11.

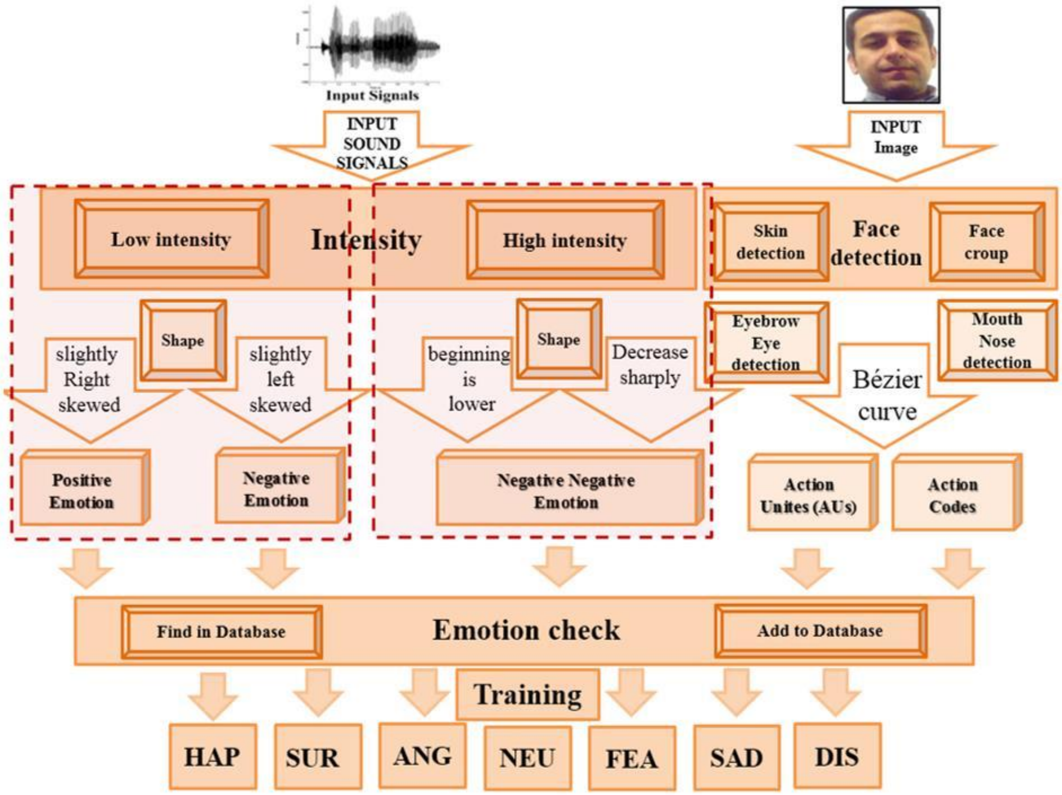


Fig. 2-11 Diagram of the hybrid system methodology.

2-5 Basic emotion Theory

However in many disciplines researchers on emotion imply different processes and meanings about emotion, they cannot agree on the same definition. The most important debate in this filed is the processes to activate emotion and the role of emotion in our daily activities and pursuits. This section describes the emotion theory behind the proposed algorithm for speech recognition and facial features extraction that used in this project.

Humans belong to various ethnic groups with different attributes of sound signals (intensity, accent and speech rate) and facial features (shape, color and size). Also, they have diverse emotion expressions, depending on culture, age and gender.

The evidence reviewed suggests that a theory that builds on concepts of basic emotions, the continual basic emotion is as a factor that influence on mind and behavior [187]. Base of this reason firstly, we have been described that each emotion expression corresponds to the different motion of the muscles. However, it is very difficult to determine the boundaries between the basic emotions muscles movement. The “neutral” emotion is used like an intermediate state when switching between two different emotions and it is core of the emotion recognition system.

Moreover, humans do not normally have two basic facial emotion at the same time, although in some cases (like fear and sadness), they can slightly overlap. Figure 2-12 shows how all the emotions have a connection with the neutral emotion, and in some cases (like fear and sadness), they can be slightly connected together.

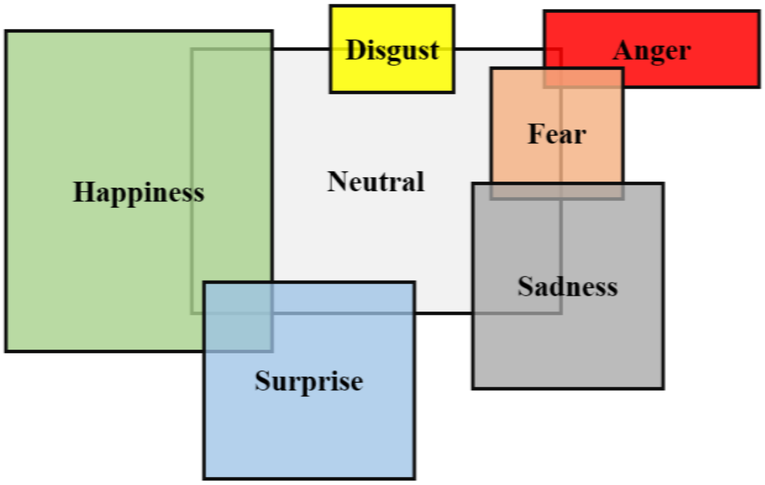


Fig. 2-12 Interaction of the basic emotions with each other.

2-6 Mathematical Methods for emotion recognition

A wide range of mathematical methods and algorithms are currently used to solve emotion recognition in human robot interaction. They recognize and import different patterns in huge data. Also, they have been used in different sciences namely; image analysis, speech analysis, person identification and character recognition. In this section we give a brief overview of the Bézier curves model and Support Victor Machine (SVM) in order to detection and classification of the basic emotions.

2-6-1 Beizer curve

The Bézier curves model is a powerful mathematical tool for constructing curves in different surfaces [188]. As mentioned above we want to convert the eyebrows, eyes and mouth features to fitted curves is the next step before recognition of the emotions. Bézier curves could approximately represented 2D facial shapes. For applying the Bézier curves, we need to generate some contour points for control information.

Zhang et al. in 1998 used ($u_{0,1}$ (t) = 1-t and $u_{1,1}$ (t) = t) as the two initial functions, by means of define the Bézier curve basis [188]. As it can be shown in Equation (1-7) and Figure 2-13, Zheng extended the model for two and four pointed.

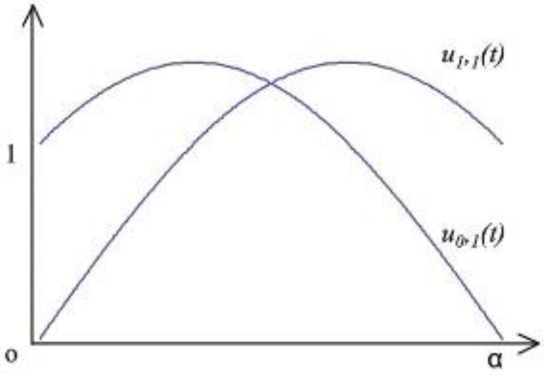


Fig. 2-13 The two initial functions bais on Bézier curves model.

$$\mathbf{u_{0,1}(t) = \sin(\alpha\text{-}t)/\sin\alpha,}$$

$$\mathbf{u_{0,1}(t) = \sin t/\sin\alpha,}$$

$$\textbf{Where } \mathbf{t \in [0, \alpha] \text{ and } \alpha \in [0,t]}$$

$$\boldsymbol{\delta}_{\mathbf{0,1}} = \left(\int_0^{\alpha} u_{0,1}(t)\,\mathrm{d}t\right)^{-1} \quad \text{and} \quad \delta_{1,1} = \left(\int_0^{\alpha} u_{1,1}(t)\,\mathrm{d}t\right)^{-1}. \qquad \text{Equation (1)}$$

Then

$$\mathbf{u}_{\mathbf{0,2}}\,\mathbf{(t) = 1-} \int_0^t \delta_{0,1} u_{0,1}(s)\,\mathrm{d}s = \big(1-\cos(\alpha - t)\big)/(1-\cos\alpha), \qquad \text{Equation (2)}$$

$$\mathbf{u}_{\mathbf{1,2}}\,\mathbf{(t) =} \int_0^t \big(\delta_{0,1} u_{0,1}(s) - \delta_{1,1} u_{1,1}(s)\big)\,\mathrm{d}s = \big(1-\cos t + \cos\alpha - \cos(\alpha - t)\big)/(1-\cos\alpha), \qquad \text{Equation (3)}$$

$$\mathbf{u}_{\mathbf{2,2}}\,\mathbf{(t) =} \int_0^t \delta_{1,1} u_{1,1}(s)\,\mathrm{d}s = (1-\cos t)/(1-\cos\alpha). \qquad \text{Equation (4)}$$

In the same way for n >2 we define the Bézier Curve method;

$$\mathbf{u}_{\mathbf{0,n}}\,\mathbf{(t) = 1-} \int_0^t \delta_{0,n-1} u_{0,n-1}(s)\,\mathrm{d}s, \qquad \text{Equation (5)}$$

$$\mathbf{u}_{\mathbf{i,n}}\,\mathbf{(t) =} \int_0^t \big(\delta_{i-1,n-1} u_{i-1,n-1}(s) - \delta_{i,n-1} u_{i,n-1}(s)\big)\,\mathrm{d}s, \qquad \text{Equation (6)}$$

$$\mathbf{u}_{\mathbf{n,n}}\,\mathbf{(t) =} \int_0^t \delta_{n-1,n-1} u_{n-1,n-1}(s)\,\mathrm{d}s. \qquad \text{Equation (7)}$$

Figure 2-14 shows the result of this formulation with four points.

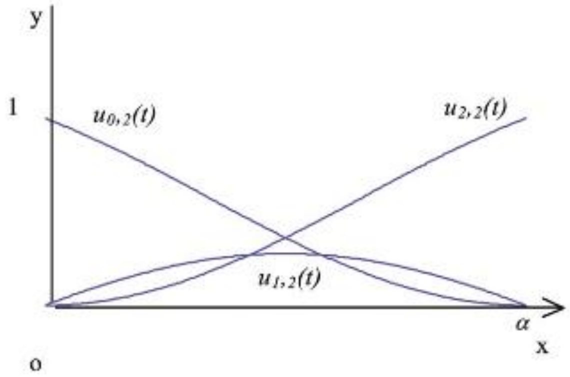


Fig 2-14 The graph of Bézier Curve with four point

In the propose system we used Zheng Bézier Curve method that lie inside its control polygon. For example, in this work we chose eight neighbor points for each facial region (left/right eye, left/right eyebrow, upper /lower). The aim of the Bézier Curve method is to interpolate a sequence of points [189]. Equation (8) represents the formula of a Bézier curve.

$$\gamma^{(t)} = \sum_{i=0}^{n} pi\, [n\,!/((i\,!\,*\,(n-1)\,!)](1-t))^{n-i} \times t^{i} \qquad \text{Equation (8)}$$

As shown in Figure 2-15, the system generates Bézier curves which can represent the principal facial regions. The Bezier curve has two anchor points (begin and end) for each of 8 points in facial features and also four control points that determine its shape of surface.

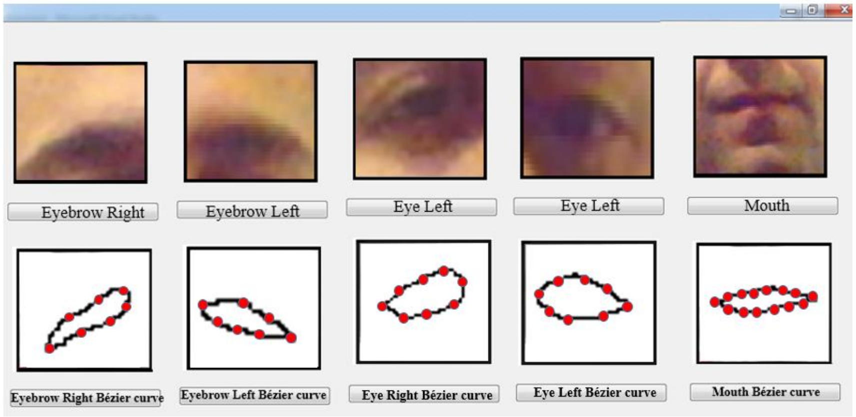


Fig. 2-15 Result of Bézier curve form five principal facial feature.

The Bézier representation of the curves can then be employed to determine the distances between the points of interest. For instance, the formula in Equation (9) can be used to calculate the distance between left and right eye (we used Hi " Eq. (8) " for right and Hj for left eye) and the formula in Equation (10) can be used to calculate the distance between the two extreme points of the mouth.

Equation (9)

$$\mathbf{Z} = \int_{i}^{n} \sum_{i=2}^{6} \ \mathbf{(\ e^{H_{\mathit{i}}}\ \ SinW\mathit{i}\ -\ e^{H_{\mathit{j}}}\ \ CosW\mathit{i}\ \ )}$$

Z = feature point distance

n = number of feature points

$$\mathbf{Z} = \int_{i}^{n} \sum_{i=1}^{2} \ \mathbf{(\ e^{H_{\mathit{i}}}\ \ SinW\mathit{i}/2\ -\ e^{H_{\mathit{j}}}\ \ CosW\mathit{i}\ /2\ )}$$

Equation (10)

Then the system save all the computed values into a database. Finally the proposed system found the nearest matching pattern with related emotion.

The proposed algorithm for facial emotion recognition is divided into two steps. The first step included Bézier curve analysis and measurement of each input image base on graphic curve. In the second step, extracted facial Action Units (AUs) and calculated the distance parameters between facial feature on input face image and normal face image.

The source code that has been used on emotion recognition system to generate the shape of eight points continuously put in the below program. The proposed shapes is not optimal. We calculated some points and stored each set of point as specific emotion in database. The system database contains index of seven kinds of emotion with different features, which are extracted based on the rules. The program have been repeated for six facial features (two eyebrows-right/left, two eyes-right/left, two lips upper/lower). If the input graphic curve did not match with the emotion in database, we added this emotion data manually in to the database when program started the training phase.

```
// Number of intermediate points between two source ones,
  // points between (x1,y1), (x2,y2), (x3,y3) and (x4,y4),
 // Then x0,y0 - the previous vertex,
// x5,y5 - the next one.
#define NUM_STEPS
void curve4(Polygon* p),

          double x0, double y0,   //Anchor1
          double x1, double y1,   //Control1
          double x2, double y2,   //Control2
          double x3, double y3,   //Control3
          double x4, double y4)   //Control4
          double x5, double y5)   //Anchor2
```

```
{
    double dx1 = x2 - x1;
    double dy1 = y2 - y1;
    double dx2 = x3 - x2;
    double dy2 = y3 - y2;
    double dx3 = x4 - x3;
    double dy3 = y4 - y3;
    double dx3 = x5 - x4;
    double dy3 = y5 - y4;

  double subdiv_step  = 1.0 / (NUM_STEPS + 1);
    double subdiv_step2 = subdiv_step*subdiv_step;
    double subdiv_step3 = subdiv_step*subdiv_step*subdiv_step;
    double subdiv_step4 = subdiv_step*subdiv_step*subdiv_step*subdiv_step;
    double subdiv_step5 = subdiv_step*subdiv_step*subdiv_step*subdiv_step*subdiv_step;

    double tmp1x = x2 - x1 * 2.0 + x3;
    double tmp1y = y2 - y1 * 2.0 + y3;
    double tmp2x = (x3 - x2)*3.0 - x1 + x4;
    double tmp2y = (y3 - y2)*3.0 - y1 + y4;
    double tmp2x = (x4 - x3)*6.0 - x2 + x5;
    double tmp2y = (y4 - y3)*6.0 - y2 + y5;

    double fx = x1;
    double fy = y1;

    double dfx = (x2 - x1)*pre1 + tmp1x*pre2 + tmp2x*subdiv_step3;
    double dfy = (y2 - y1)*pre1 + tmp1y*pre2 + tmp2y*subdiv_step3;

    double ddfx = tmp1x*pre4 + tmp2x*pre5;
    double ddfy = tmp1y*pre4 + tmp2y*pre5;

    double dddfx = tmp2x*pre5;
    double dddfy = tmp2y*pre5;

int step = NUM_STEPS;

    double xc1 = (x0 + x1) / 2.0;
    double yc1 = (y0 + y1) / 2.0;
    double xc2 = (x1 + x2) / 2.0;
    double yc2 = (y1 + y2) / 2.0;
    double xc3 = (x2 + x3) / 2.0;
    double yc3 = (y2 + y3) / 2.0;
    double xc4 = (x3 + x4) / 2.0;
    double yc4 = (y3 + y4) / 2.0;
    double xc5 = (x4 + x5) / 2.0;
    double yc5 = (y4 + y5) / 2.0;

    double len1 = sqrt((x1-x0) * (x1-x0) + (y1-y0) * (y1-y0));
    double len2 = sqrt((x2-x1) * (x2-x1) + (y2-y1) * (y2-y1));
    double len3 = sqrt((x3-x2) * (x3-x2) + (y3-y2) * (y3-y2));
    double len4 = sqrt((x4-x3) * (x4-x3) + (y4-y3) * (y4-y3));
    double len5 = sqrt((x5-x4) * (x5-x4) + (y5-y4) * (y5-y4));

    double k1 = len1 / (len1 + len2);
    double k2 = len2 / (len2 + len3);
    double k3 = len3 / (len3 + len4);
    double k4 = len4 / (len4 + len5);

    double xm1 = xc1 + (xc2 - xc1) * k1;
    double ym1 = yc1 + (yc2 - yc1) * k1;

    double xm2 = xc2 + (xc3 - xc2) * k2;
    double ym2 = yc2 + (yc3 - yc2) * k2;

    double xm3 = xc3 + (xc4 - xc3) * k2;
    double ym3 = yc3 + (yc4 - yc3) * k2;
    double xm4 = xc4 + (xc5 - xc4) * k2;
    double ym4 = yc4 + (yc5 - yc4) * k2;

   // Resulting control four points.

    ctrl1_x = xm1 + (xc2 - xm1) * smooth_value + x1 - xm1;
    ctrl1_y = ym1 + (yc2 - ym1) * smooth_value + y1 - ym1;

    ctrl2_x = xm2 + (xc2 - xm2) * smooth_value + x2 - xm2;
    ctrl2_y = ym2 + (yc2 - ym2) * smooth_value + y2 - ym2;

    ctrl3_x = xm3 + (xc3 - xm3) * smooth_value + x3 - xm3;
    ctrl3_y = ym3 + (yc3 - ym3) * smooth_value + y3 - ym3;

    ctrl4_x = xm4 + (xc4 - xm4) * smooth_value + x4 - xm4;
    ctrl4_y = ym4 + (yc4 - ym4) * smooth_value + y4 - ym4;

    while(step--)
    {
        fx   += dfx;
        fy   += dfy;
        dfx  += ddfx;
        dfy  += ddfy;
        ddfx += dddfx;
        ddfy += dddfy;
        p->AddVertex(fx, fy);
    }
}
```

2-6-2 Support Vector Machine (SVM)

Support Vector Machine (SVM) algorithm used in the analysis and classification some of original input features. Support vector machine constructs the set of points for classification [190]. The main goal to use Support Vector Machine is to find a decision boundary between seven emotion classes that is maximally far from any point in the training data. Furthermore, SVMs generalize well even when few training data are provided.

However, note that classification performance decreases when the dimensionality of the feature set is far greater than the number of samples available in the training set. The Support Vector Machine can efficiently perform a non-linear classification and separates between a set of objects having different class memberships. Moreover, we extended the linear SVM for set of non-linear data for recognition emotion. In the case, we used the SVM model that optimally separate data into seven categories (seven basic emotion classes). The proposed system implemented the Support Vector Machine for increases the emotion classification accuracy. The features sound and facial features are too complex to compute the space.

The function that performs this mapping is to transform the original training nonlinear data into a higher dimension linear mapping is the linear kernel function, because the number of features in sound signals is large [190]. The most frequently used SVM kernel function in the domain of emotion recognition in speech is the radial basis function (RBF) kernel. We used the linear kernel and RBF kernels because the training speed in speech recognition and classification of the emotion decreases. Also, the SVM parameters were determined independently of the test data. Figure.2-16 shown how the system change nonlinear surface to the linear space.

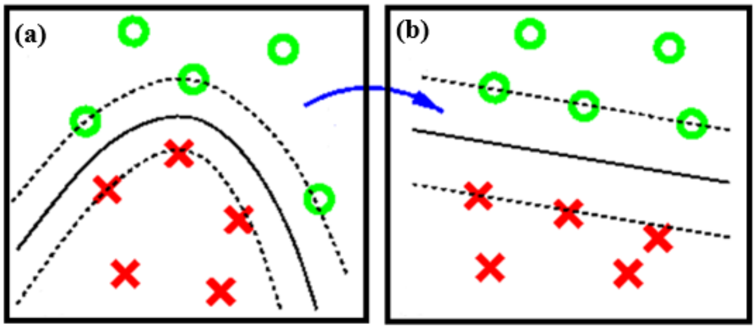


Fig. 2-16 change non liner space to linear with kernel.

a) Separation is provided by a non-linear surface b) non-linear surface to a linear surface in a feature space

2-7 Open-source toolkits for sound recognition

In recent year various open-source toolkits on speech processing field have been improved with high accuracy and performance. Also, this kind of system is able to hold whole conversations with the user (audio recording), audio file reading, features extraction, classification data, train, and improve the general acoustic models. Additionally, they can be used with a variety of Window and operating system applications. However the main window of this open-source program such as; (open EAR, PRAAT, WEKA, Simon, SPRAAK, XVoice, Speech Filling System (SFS), Open Mind Speech, EmoVoice and Babel Technologies) have been reorganized to bring the most important options together in one screen [104]. Moreover, between all open-source toolkits on speech emotion recognition only PRAAT and HTK freely available to anybody and include certain classifiers. In this work we proposed to use PRAAT for speech graph analysis.

2-7-1 Open EAR

The Munich open Affect Recognition Toolkit (openEAR) is one of the first tools on open-source speech recognition toolkits. OpenEAR is introduced as an emotion recognition and feature extraction algorithms that implemented in C++. OpenEAR, is a stable and efficient set of tools for researchers and those developing emotional aware applications, providing the elementary functionality for emotion recognition [192]. Also, openEAR freely available to anybody and can be used as an out-of-the-box emotion live affect recognizer for various domains, using pre-trained models which are included in the distribution.

Open EAR program can extracted a set of feature like (Low-Level Descriptors (LLD) and various statistical functional. In addition, we can use other information namely; signal energy, FFT-Spectrum, Mel-Spectrum, voice Quality, pitch, LPC Coefficients, formants, Spectral and time Signal for emotion recognition with Open EAR.

2-7-2 PRATT Soft ware

In order to implement features extraction, a standard computer system program, “PRAAT” used for speech analysis. PRAAT software is an open-source and flexible tool for voice sampling in the field of pitch, formant, spectrograms and intensity analysis by academic researchers [193]. PRAAT is a wonderful software package written and maintained by Paul Boersma and David Weenink of the

University of Amsterdam [194]. It can run on a wide range of operating systems, including various versions of Unix, Linux, Mac and Microsoft Windows (95, 98, NT4, ME, 2000, XP, Vista, 7, 8) [195]. PRAAT freeware computes four pitch values in one frame length. The PRAAT is an efficient and flexible tool that combines many of the recent advancements in automatic speech recognition. On the other hand this program have simple interface. The PRAAT software organizes sound file into ‘‘frames’’ for analysis, computing four pitch values within one frame length. The segmented wave files were analyzed one at a time and the pitch contours were saved in separate files. To record sound using PRAAT, pull up a recording menu which allows to choose a sampling frequency (the default, 44100 Hz, is fine for most purposes), a microphone or other sound source, and whether to record a mono or stereo sound [196].

Voice characteristics at the prosodic level, including intonation and intensity patterns, carry important features for emotional states. Hence, prosody clues such as pitch and speech intensity can be used to model different emotions and the fundamental frequency pitch contours, pitch values, pitch range, as well as the average intensity can enable one to build a classification of various emotion types. For example, high values of pitch are correlated with happiness and anger, whereas sadness and boredom are associated with low pitch values [118].

Three types of features were considered: pitch (range, value and peak), intensity (energy) and rate speech; hence, the graphs of formant, pitch and intensity were analyzed.

Pitch features are often made perceptually more adequate by logarithmic/semitone function, or normalization with respect to some (speaker specific) baseline. Pitch is a fundamental acoustic feature of speech and needs to be determined during the process of speech analysis [197]. The modulation of pitch plays a prominent role in everyday communication. Pitch extraction can influence the performance of emotion recognition.

The ***Pitch value*** of a sound is the length of the time interval when the sound signal is higher than the average. The ***pitch peak*** of a sound is the maximum intensity (peak) of the signal. The ***pitch range*** is defined as the ratio between the highest and lowest values of intensity of the sound signal.

Intensity features usually represent the loudness (energy) of a sound as perceived by the human ears, based on the amplitude in different intervals [198].

Energy is the square of the amplitude multiplied by the duration of the sound.

Voice quality is a complicated issue in itself, since there are many different measures of voice quality, mostly clinical in origin and mostly evaluated for constant vowels only, though once again standardization in this area is lacking [198].

The ***spectrum*** is characterized by formants (spectral maxima) modeling spoken content. Higher formants amplitude also represents speaker position and characteristics [199].

Non-linguistic vocalizations are non-verbal phenomena, such as breathing, mute and laughter [194].

The ***speech rate*** specifies the speaking rate in words per minute, a rate that varies somewhat by language, but is nonetheless widely supported by speech synthesizers [194].

2-5 Programming in C/C++ and open CV

In order to detect various basic emotional states, we implemented our program with Visual C# in the Visual Studio 2012 software. For connection of the webcam to the proposed hybrid algorithm, we used an open CV library and the Sobel filter. The Open CV (Open Source Computer Vision Library) is a machine learning software library [200]. This library can take advantage of multi-core processing. Also, it enables computers and robots to see and make decisions based on the input data. We used the Open CV and storage devices for loading facial image files (JPEG, PNG, GIF, TIFF, BMP …)

In this project in speech emotion recognition phase we decided to produces privet data base in training section. We used five simple sentences of everyday life. The participants to the experiment had to repeat these six sentences for three times with a neutral (NEU) intonation, with 1 second of interval between each sentence. Then, every participant had to repeat again three times the same five sentences, but with each one of the emotions listed above. All the sentences were recorded, thus obtaining 630 files, which were input to a dedicated program for speech analysis which could provide the intensity, pitch (peak, range, values) alignment and speech rate of all the sentences.

Also, in facial feature extraction we used two databases. The first database consists of 840 sequences of 30 participants, 15 female and 15 male (from 20 to 48 years old), used for training the algorithm. The neutral face and the corresponding Facial Action were manually identified in each of these sequences. All images are colorful and taken against a white homogenous background in an upright frontal position. In each sequence, the human subject displays a mimic facial expression (happiness, sadness, disgust, surprise, fear, anger and neutral).

Finally, in order to evaluate the hybrid algorithm and checking the accuracy of our method we used a comprehensive and rich emotional image database named Cohn-Kanade AU-Coded Facial Expression Database.

CHAPTER 3

Implementation the

Emotion recognition system

The previous chapter gave a description of proposed algorithm, mathematical formulation and theoretical ideas for emotion recognition on speech and facial feature extraction. This chapter discuss, how the proposed methodology is applied in practice to create a high accuracy emotion recognition software.

Extraction of features is an essential pre-processing and fundamental step in recognition of basic emotion. One of the important questions in the field of human robot interaction and emotion recognition is how many and which features must chose for automatic recognition of emotions? The answer of this question have main role to improve performance and reliability the system. Also, to obtain more efficient classification system, we must configure the processing speed emotion recognition and memory requirements.

Ideally, feature selection methods should not only reveal single or most relevant attributes. Features such as; pitch, intensity, duration, formant, voice quality, speech rate (SR), facial extraction, action unite codes (AUCs) were consist of extracted from human robot communication. The essence of speech emotion analysis is control and compare sound plots. Also, facial features emotion expression is effectively connected to movement of the facial muscles as well as to deformations of the face. An automated facial recognition system has solve two basic problems: facial feature localization and feature extraction. This task is the most complicated and time consuming on emotion recognition system.

Thus, have been proposed new method to decrees the recognition rate. Current capture discuss two sections to extract features from human robot interaction namely; speech and facial features extraction and features classification.

3-1 Speech features extraction

The speech signal is the fastest and the most natural method of communication between humans and robot. Each emotion corresponds to a different portion of the spoken utterance. However, it is very difficult to determine the boundaries between these portions. In this project we proposed an algorithm based on classification meaningful and informative set of features, such as pitch (peak, value and range), intensity, duration speech time, formant, voice quality and speech rate (SR). In Figure 3-1 we decided to analyses simple sentences with PRAAT program by means of provide the intensity, pitch (peak, range, values) alignment and speech rate in one windows.

Voice characteristics at the prosodic level, including intonation and intensity patterns, carry important features for emotional states. Hence, prosody clues such as pitch and speech intensity can be used to model different emotions. Also, the fundamental frequency pitch contours, pitch values, pitch range, as well as the average intensity can enable one to build a classification of various emotion types. As depicted in Figure 3-1, three types of features were considered: pitch (range, value and peak), intensity (energy) and rate speech; hence, the graphs of formant, pitch and intensity were analyzed.

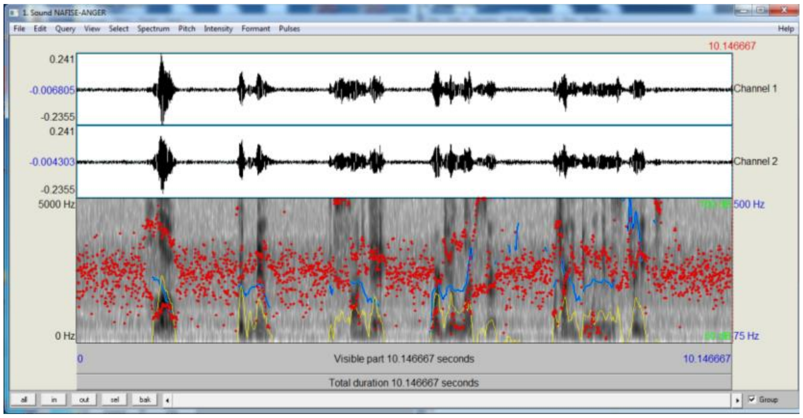


Fig. 3-1 Example of speech analysis with the PRAAT program.

In the proposed system, before started to speech analyses we know, high values of pitch are correlated with happiness and anger, whereas sadness, disgust and boredom are associated with low pitch values [31].

3-1-1 Pitch features extraction

Pitch features are often made perceptually more adequate by logarithmic/semitone function, or normalization with respect to some (speaker specific) baseline. Pitch is a fundamental acoustic feature of speech and needs to be determined during the process of speech analysis. The modulation of pitch plays a prominent role in everyday communication fulfilling very different functions, like contributing to the segmentation of speech into syntactic and informational units, specifying the modality of the sentence, regulating the speaker–listener interaction, expressing the attitudinal and emotional state of the speaker, and many others. Automatic pitch stylization is an important resource for researchers that working both on prosody and speech technologies [201].

Pitch range was considered as a necessary feature for emotion recognition. Pitch contour extraction was done using the PRAAT software. Figure 3-2 shows some pitch plots for the sentences spoken by the participants to the experiment (the clearest results were chosen, among the three repetitions made by each participant).

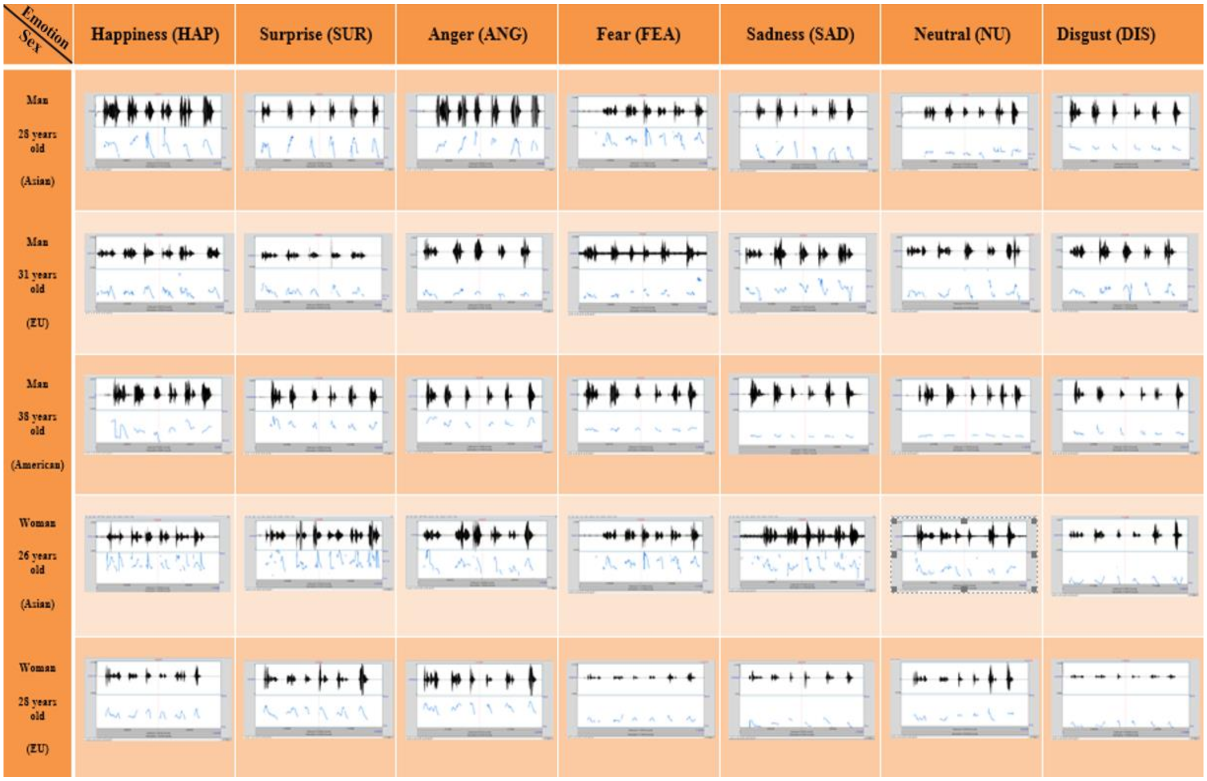


Fig. 3-2 Some pitch results from the 30 interviewed persons (Europeans, Americans, and Asians).

As depicted in Figure 3-3, 3-4, 3-5, the pitch contours under positive valence emotions (such as surprise and happiness) are similar. The value of the pitch at the end of the sentence is lower than the value at the beginning, but surprise has a bigger pitch value. We can see that the highest pitch value is for surprise and the lowest corresponds to disgust.

Also, we can see that the pitch peak under positive valence emotions is sharper among Asian speakers, while European and American speakers more or less have similar pitch contours under positive valence emotions. Happiness and anger have the highest average pitch peak for European speakers (see Figure 3-6) while sadness has the lowest pitch peak. In our experiment, we can also see that surprise and anger for Asian and American speakers have the highest average pitch peak (see Figure 3-4, 3-5, 3-7 and 3-8).

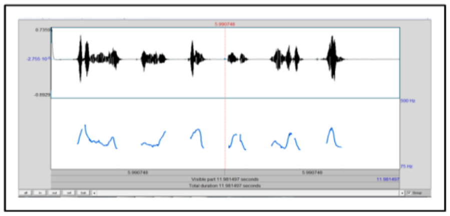
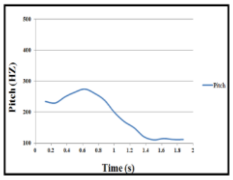
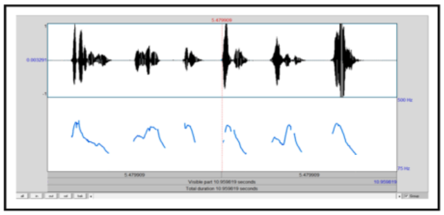
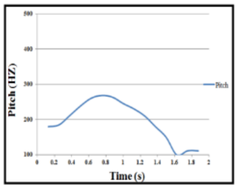
<table>
<tr><th></th><th>Pitch</th><th>Graph</th></tr>
<tr><td>Happiness</td><td></td><td>
</td></tr>
<tr><td></td><td colspan="2">Typical pitch contour of Happiness (HAP) emotion</td></tr>
<tr><td>Surprise</td><td></td><td>
</td></tr>
<tr><td></td><td colspan="2">Typical pitch contour of Surprise (SUR) emotion</td></tr>
</table>

Fig. 3-3 European pitch contours for Happiness and Surprise.

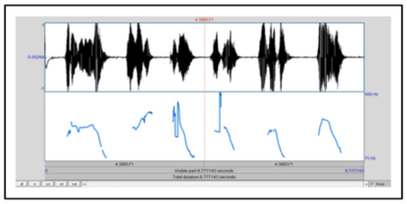
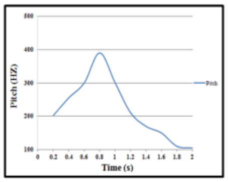
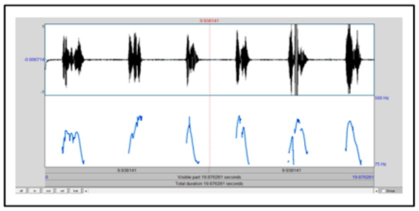
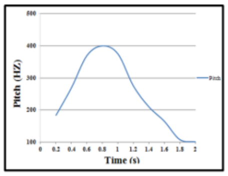
<table>
<tr><th></th><th>Pitch</th><th>Graph</th></tr>
<tr><td>Happiness</td><td></td><td>
</td></tr>
<tr><td></td><td colspan="2">Typical pitch contour of Happiness (HAP) emotion</td></tr>
<tr><td>Surprise</td><td></td><td>
</td></tr>
<tr><td></td><td colspan="2">Typical pitch contour of Surprise (SUR) emotion</td></tr>
</table>

Fig. 3-4 Asian pitch contours for Happiness and Surprise.

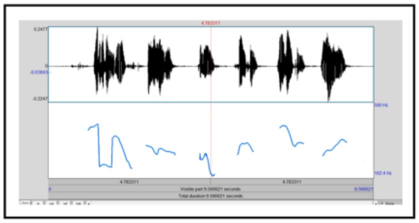
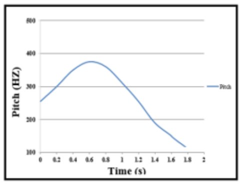
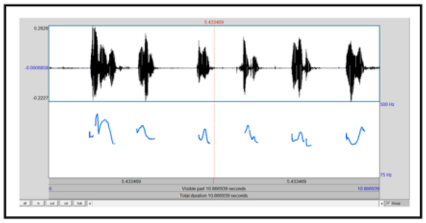
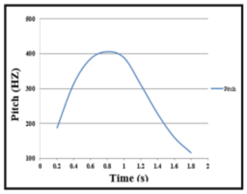
	Pitch	Graph
Happiness		
	Typical pitch contour of Happiness (HAP) emotion	
Surprise		
	Typical pitch contour of Surprise (SUR) emotion	

Fig. 3-5 American pitch contours for Happiness and Surprise.

Among the negative valence emotions, anger has the highest pitch peak (see Figure 3-6, 3-7 and 3-8). Sadness decreases sharply for Asian and American speakers, but sadness slop decreases slowly.

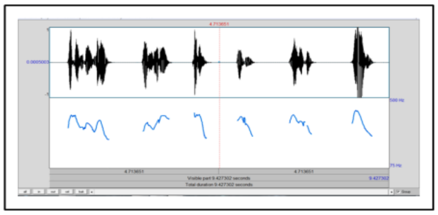
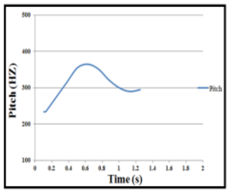
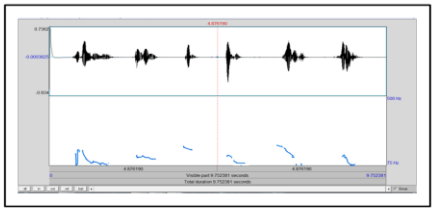
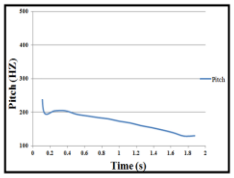
	Pitch	Graph
Anger		
	Typical pitch contour of Anger (ANG) emotion	
Sadness		
	Typical pitch contour of Sadness (SAD) emotion	

Fig. 3-6 European pitch contours for Anger and Sadness.

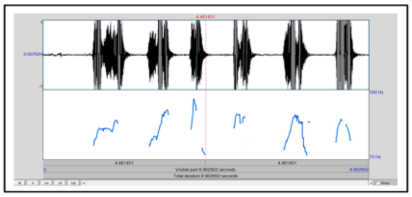
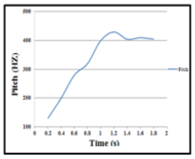
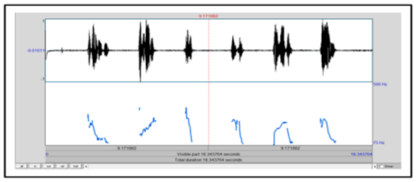
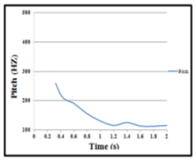
	Pitch	Graph
Anger		
	Typical pitch contour of Anger (ANG) emotion	
Sadness		
	Typical pitch contour of Sadness (SAD) emotion	

Fig. 3-7 Asian pitch contours for Anger and Sadness.

If we compare sadness and neutral for all groups of speaker (Figure 3-6, 3-7, 3-8 and 3-12) the neutral emotion does not have a distinct peak and is similar to sadness. However, sadness has lower ending pitch signals.

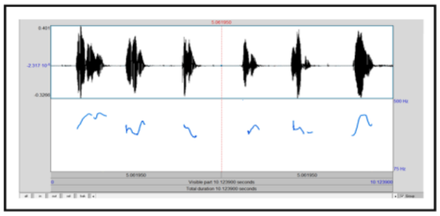
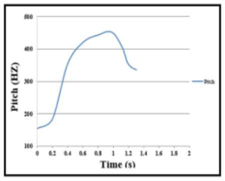
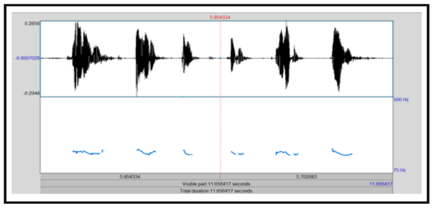
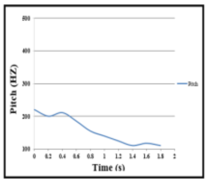
	Pitch	Graph
Anger		
	Typical pitch contour of Anger (ANG) emotion	
Sadness		
	Typical pitch contour of Sadness (SAD) emotion	

Fig. 3-8 American pitch contours for, Anger and Sadness.

Asian speakers were more sensitive to sad emotion, while the pitch graphs of Americans and Europeans were similar. Anger is associated with the highest energy for Asian and American speakers but for Asian speakers the anger slope decreases slowly, while sadness is associated with the lowest energy for Asian and European speakers.

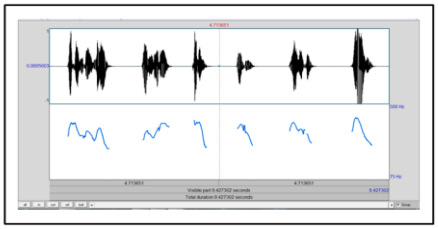
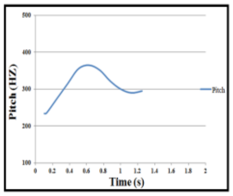
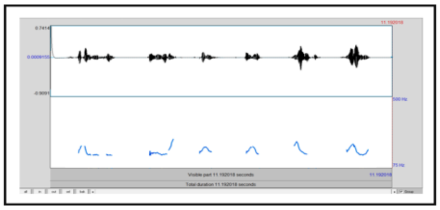
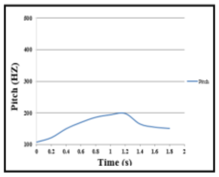
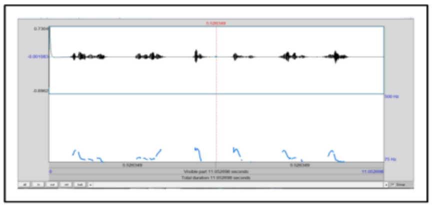
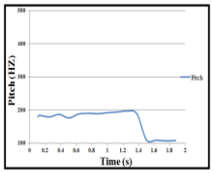
	Pitch	Graph
Anger		
	Typical pitch contour of Anger (ANG) emotion	
Fear		
	Typical pitch contour of Fear (FEA) emotion	
Disgust		
	Typical pitch contour of Disgust (DIS) emotion	

Fig. 3-9 European pitch contours for Anger, Fear and Disgust.

Even though no universal similarities are observed among negative valence emotions, similarities are noted between certain utterances under anger and fear. In Figure 3-9, 3-10 and 3-11 anger is characterized by a rising peak followed by either decrease or a leveling out of the pitch values and the utterance duration is observed to be small.

In almost all utterances under anger and fear, the pitch increases to a peak and then decreases slightly left-skewed. European and American speakers more or less have similar pitch contours under fear emotion.

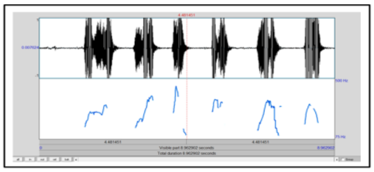
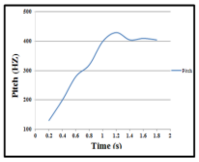
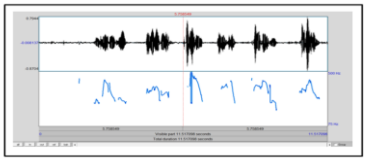
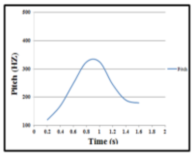
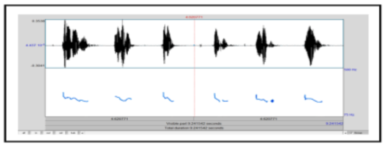
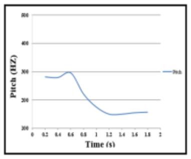
	Pitch	Graph
Anger		
	Typical pitch contour of Anger (ANG) emotion	
Fear		
	Typical pitch contour of Fear (FEA) emotion	
Disgust		
	Typical pitch contour of Disgust (DIS) emotion	

Fig. 3-10 Asian pitch contours for anger, fear and disgust.

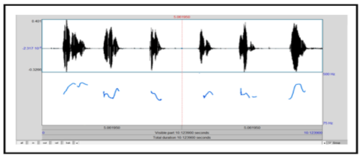
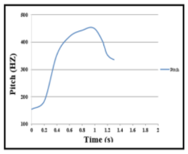
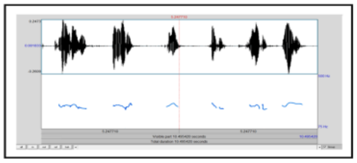
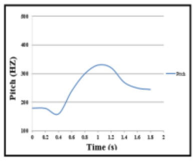
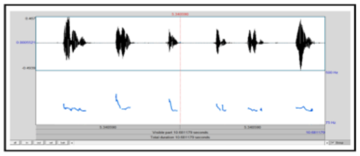
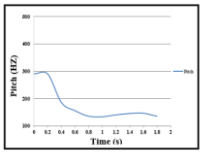
	Pitch	Graph
Anger		
	Typical pitch contour of Anger (ANG) emotion	
Fear		
	Typical pitch contour of Fear (FEA) emotion	
Disgust		
	Typical pitch contour of Disgust (DIS) emotion	

Fig. 3-11 American pitch contours for Anger, Fear and Disgust.

In Figures 3-10 and 3-11 it can be seen that the highest mean pitch values are for American speakers, while Asians have sharper pitch peaks. Pitch values and speech rate are connected together. We can see that usually the speech rate of American speakers is higher than Asian and European speakers.

As depicted in Figure 3-12, the beginning and ending of pitches in neutral emotion for Americans after rising and falling have similar frequencies. This is probably due to the fact that the mother language of American speakers was English, while Europeans and Asians (whose mother language was not English) show a bit of stress in neutral speech.

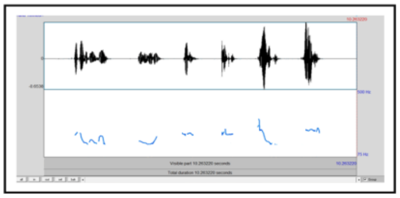
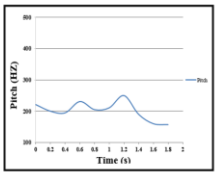
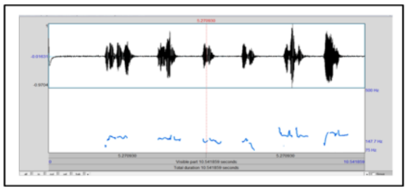
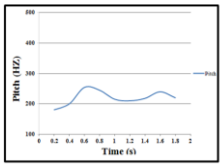
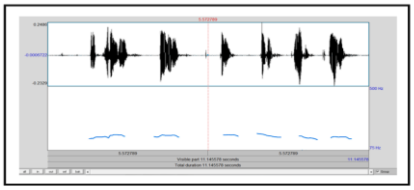
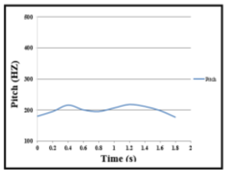
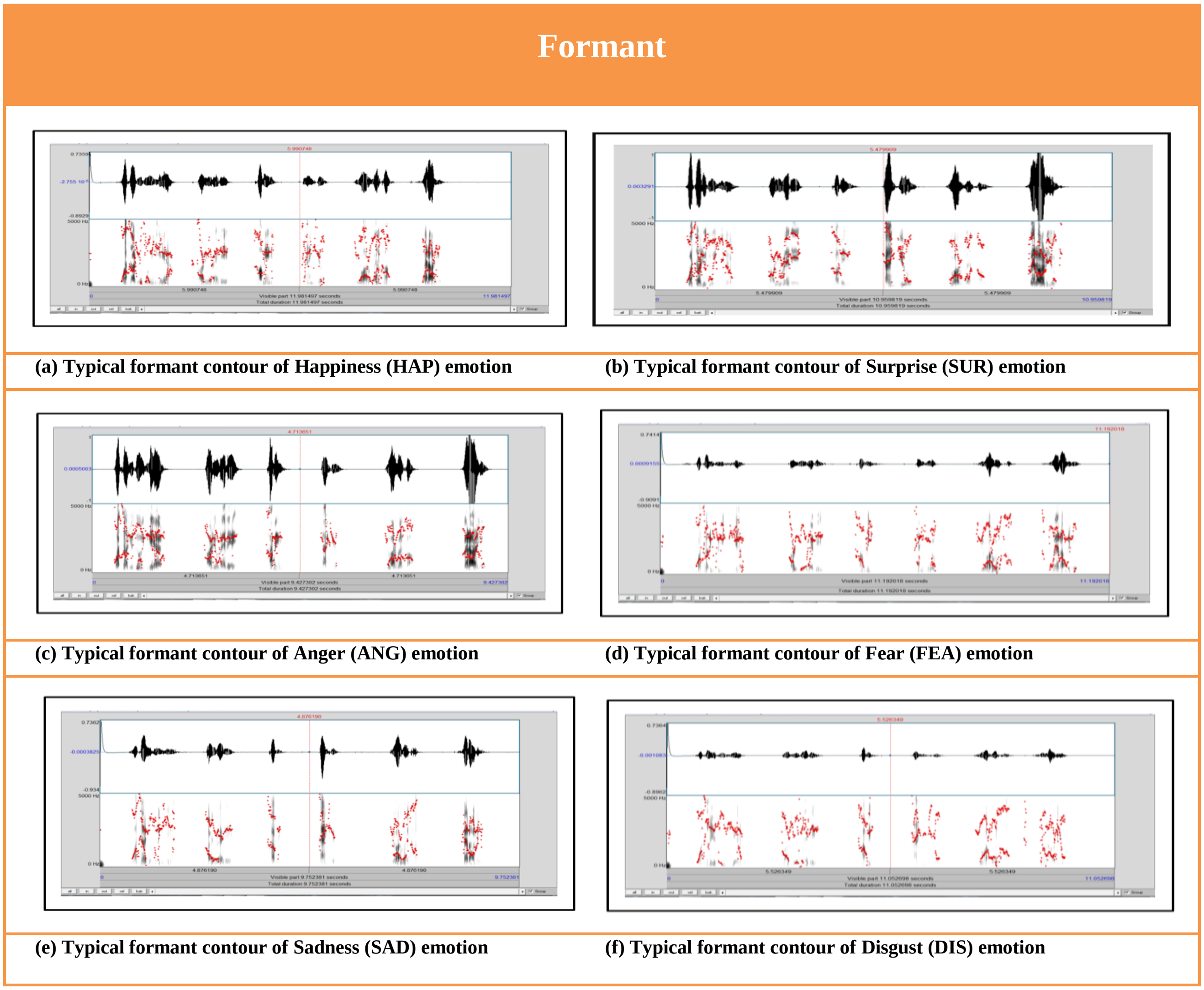
	Pitch	Graph
Neutral		
	Typical pitch contour of Neutral (NEU) emotion European	
Neutral		
	Typical pitch contour of Neutral (NEU) emotion Asian	
Neutral		
	Typical pitch contour of Neutral (NEU) emotion American	

Fig. 3-12 European, Asian and American pitch contours for Neutral.

3-1-2 Formant features extraction (speech communication)

Formants are the meaningful frequency components of human speech and contents of the vowel sounds. Formant used to be important in determining the phonetic content of speech signals. Further empirical results discussed formant used to be important in determining the phonetic content of speech signals. Also, used as identifying silence in speech recognition. We can change the position of the formants by moving around the tongue and the lip muscles so as to show the emotion in speech. In the PRAAT software the maximum value of the formant should be set to about 5000Hz for a male speaker, 5500Hz for a female speaker and even higher for children. In Figure 3-13, 3-14 and 3-15, we extracted of formant features from spectral feature vectors with PRAAT program.


Formant

(a) Typical formant contour of Happiness (HAP) emotion

(b) Typical formant contour of Surprise (SUR) emotion

(c) Typical formant contour of Anger (ANG) emotion

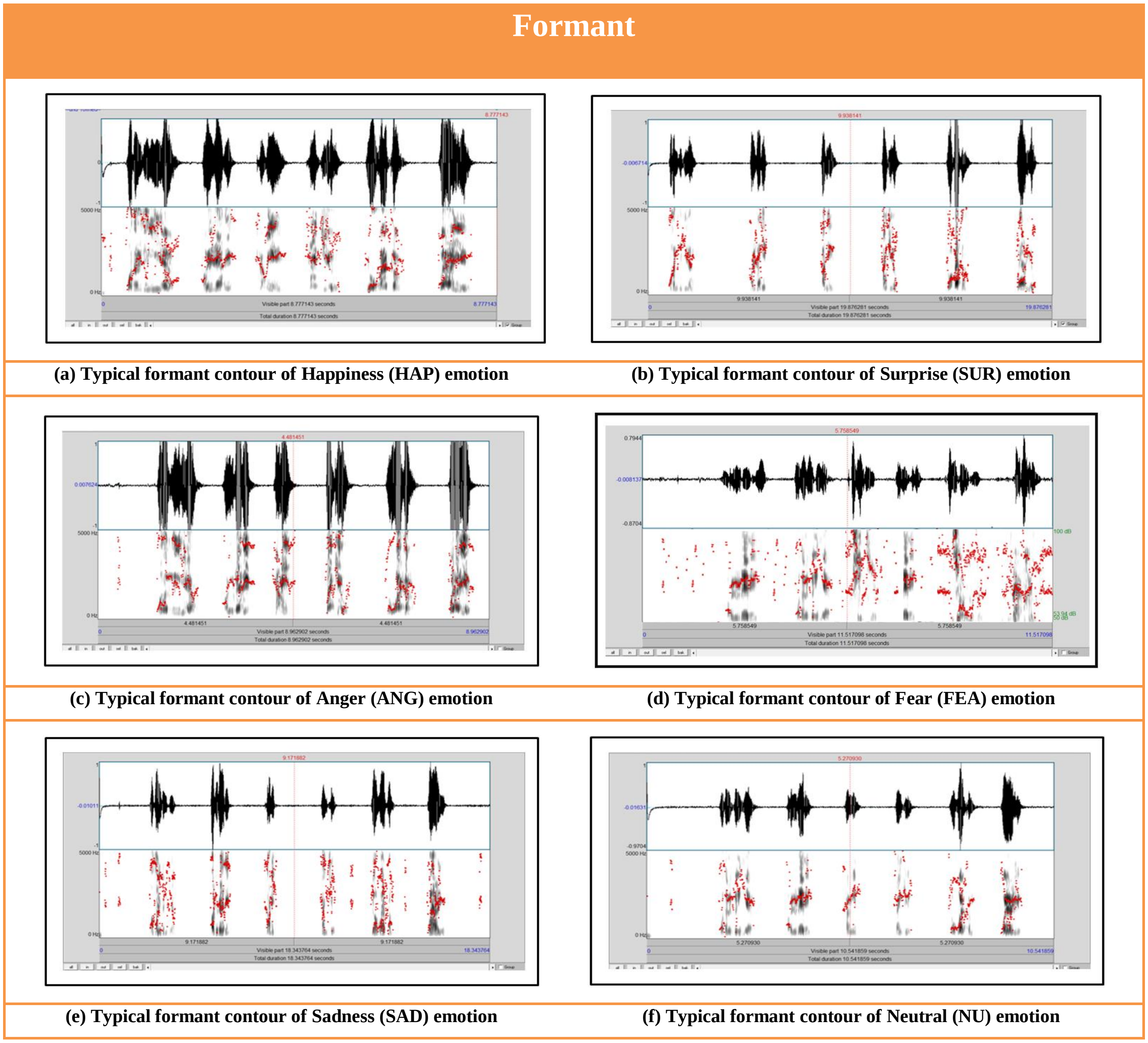
(d) Typical formant contour of Fear (FEA) emotion

(e) Typical formant contour of Sadness (SAD) emotion

(f) Typical formant contour of Disgust (DIS) emotion

Fig. 3-13 European speaker typical formant contour of basic emotions.

The plot of the formant displays the amplitude of the frequency components of the signal over time. It is generally agreed that formants are perceptually important features. It is also often acknowledged that spectral peaks (formants) should be more robust to additive noise since the formant regions will generally exhibit a large signal-to-noise ratio [202]. For most analyses of human speech, we may want to extract 5 formants per frame.



Formant

(a) Typical formant contour of Happiness (HAP) emotion

(b) Typical formant contour of Surprise (SUR) emotion

(c) Typical formant contour of Anger (ANG) emotion

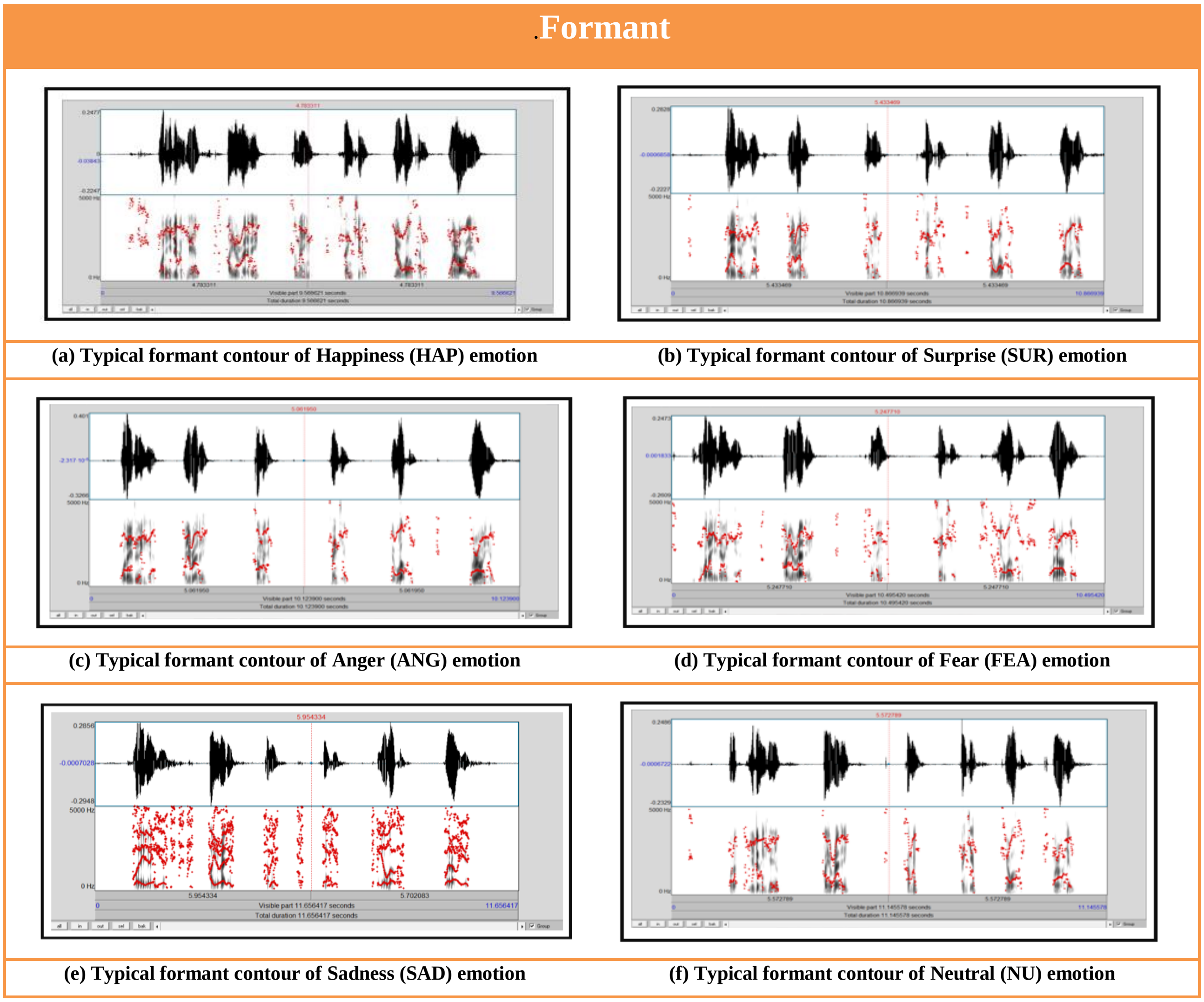
(d) Typical formant contour of Fear (FEA) emotion

(e) Typical formant contour of Sadness (SAD) emotion

(f) Typical formant contour of Neutral (NU) emotion

Fig. 3-14 Asian speaker typical formant contour of basic emotions.

As depicted in Figure 3-13, 3-14 and 3-15, the formant contour in anger and happiness for European speakers has the highest power, while we have the lowest spectral power in fear. Formant contour in Figure 3-14, 3-15 explain that anger, fear and happiness have the highest power for Asians and Americans, while we have a lot of wave and formant dots the fear plot. Asians and Europeans have the lowest spectral power in sadness, while Americans have the lowest spectral power in neutral emotion.



.Formant

(a) Typical formant contour of Happiness (HAP) emotion

(b) Typical formant contour of Surprise (SUR) emotion

(c) Typical formant contour of Anger (ANG) emotion

(d) Typical formant contour of Fear (FEA) emotion

(e) Typical formant contour of Sadness (SAD) emotion

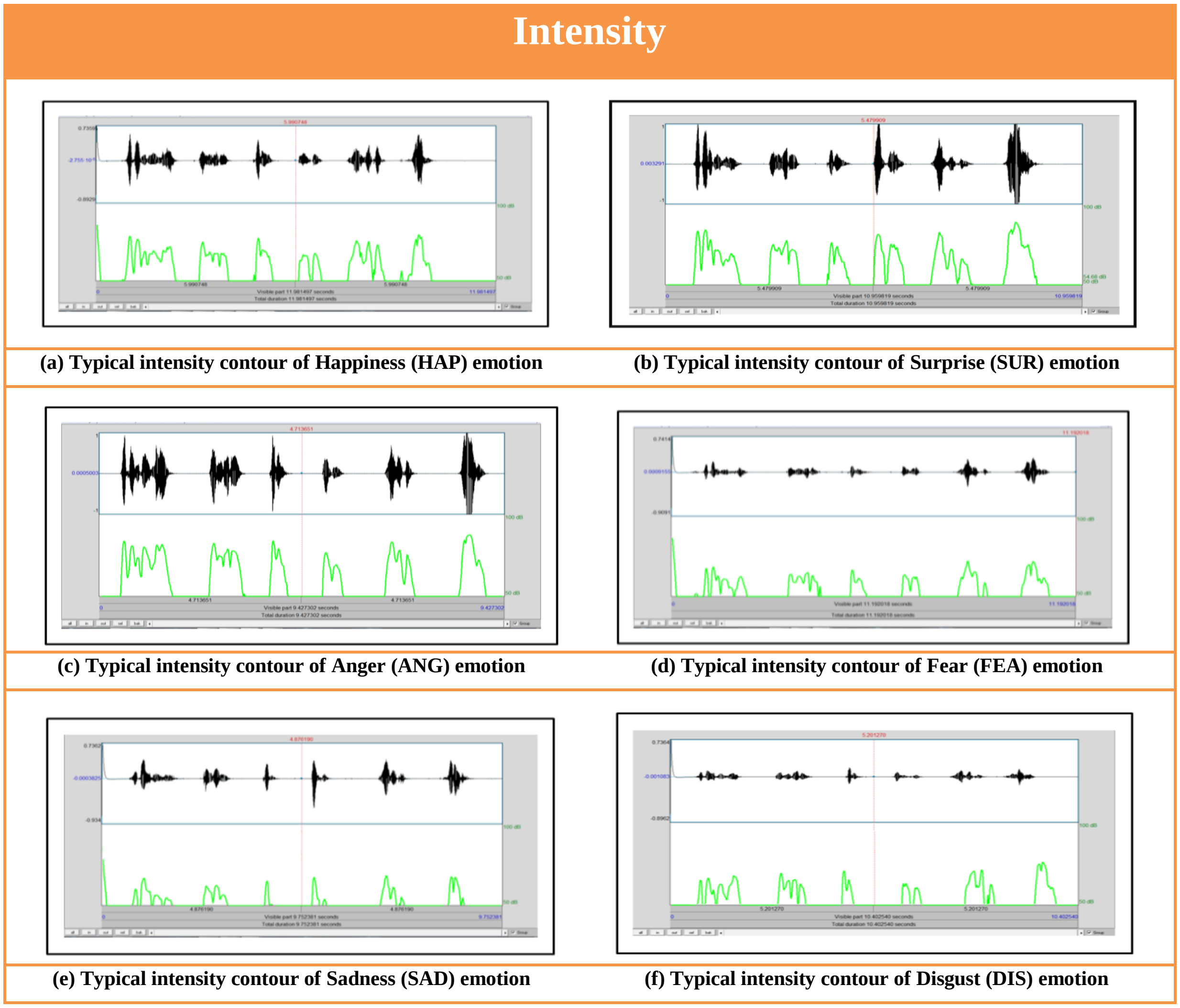
(f) Typical formant contour of Neutral (NU) emotion

Fig. 3-15 American speaker typical formant contour of basic emotions.

Finally, experiments for all ethnic groups of speakers described, formant graph information yielded significant results in the noisy conditions performance. This result show that the formant plots increases the recognition on; happiness, surprise, anger and fear emotion in real time in various environment.

3-1-3 Intensity features extraction (speech communication)

Sound or acoustic intensity is defined as the sound power and is measured in dB. The typical context in this field is the listener's location for the measurement of sound intensity. Sound intensity is a specifically defined quantity and is very sensitive to the location of the microphone. In our experiments using the PRAAT software, we put the microphone at a distance of 30cm form each participant. As expected and further shown later, this approach indeed resulted in the extraction of some meaningful intensity information that we used on emotion recognition program. In terms of intensity, as it can be seen in Figure 3-16, 3-17 and 3-18, when we have strong power on the source of sound signals, the energy and the intensity of the sound increase.



(a) Typical intensity contour of Happiness (HAP) emotion

(b) Typical intensity contour of Surprise (SUR) emotion

(c) Typical intensity contour of Anger (ANG) emotion

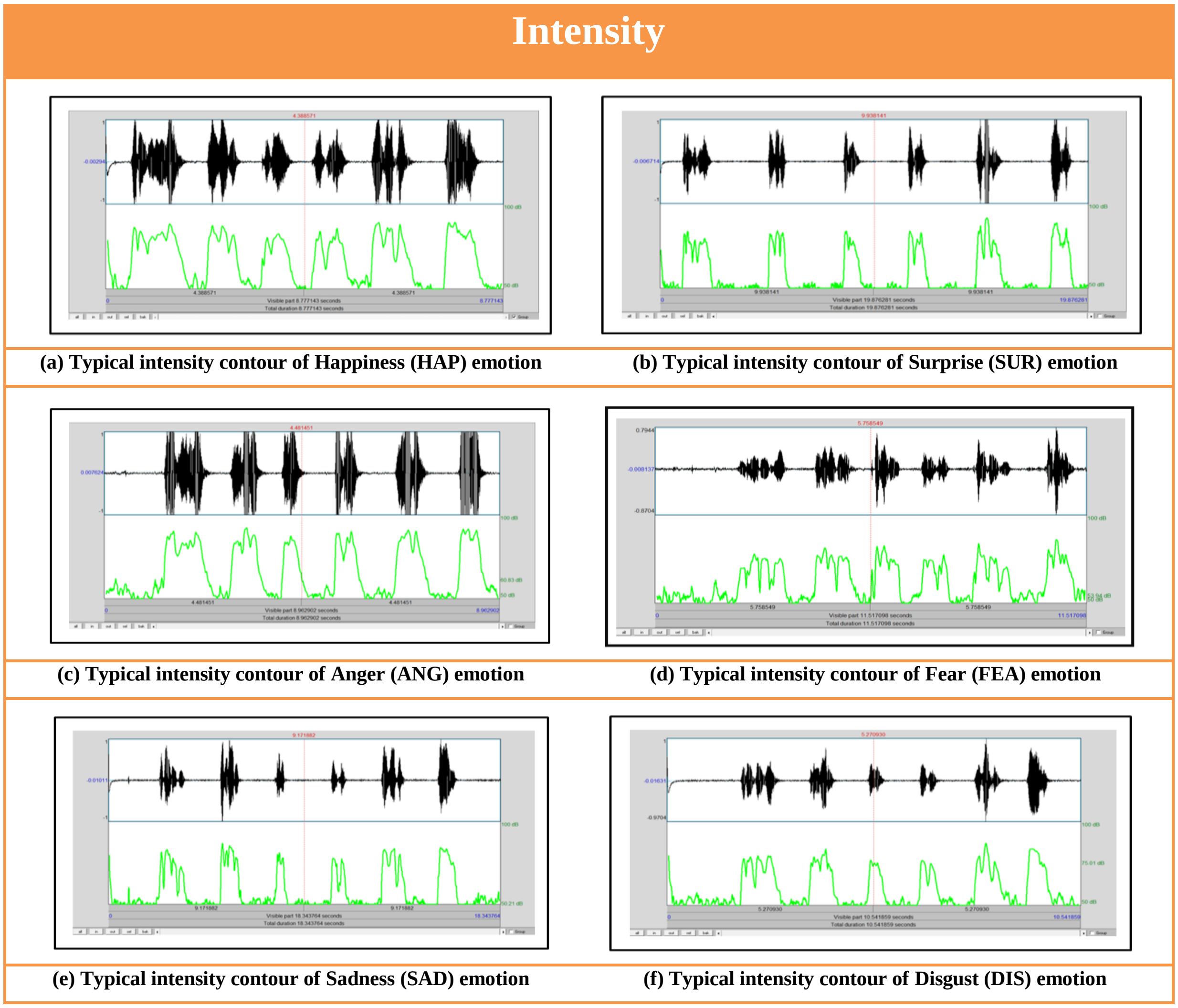
(d) Typical intensity contour of Fear (FEA) emotion

(e) Typical intensity contour of Sadness (SAD) emotion

(f) Typical intensity contour of Disgust (DIS) emotion

Fig. 3-16 European speaker typical intensity contour of basic emotions.

Anger and surprise for European speakers have the highest energy and intensity, while neutral and sadness have the lowest intensity. For Asian speakers, as it can be seen in Figure 3-17, anger and happiness have the highest energy and intensity, while fear has the lowest intensity.



(a) Typical intensity contour of Happiness (HAP) emotion

(b) Typical intensity contour of Surprise (SUR) emotion

(c) Typical intensity contour of Anger (ANG) emotion

(d) Typical intensity contour of Fear (FEA) emotion

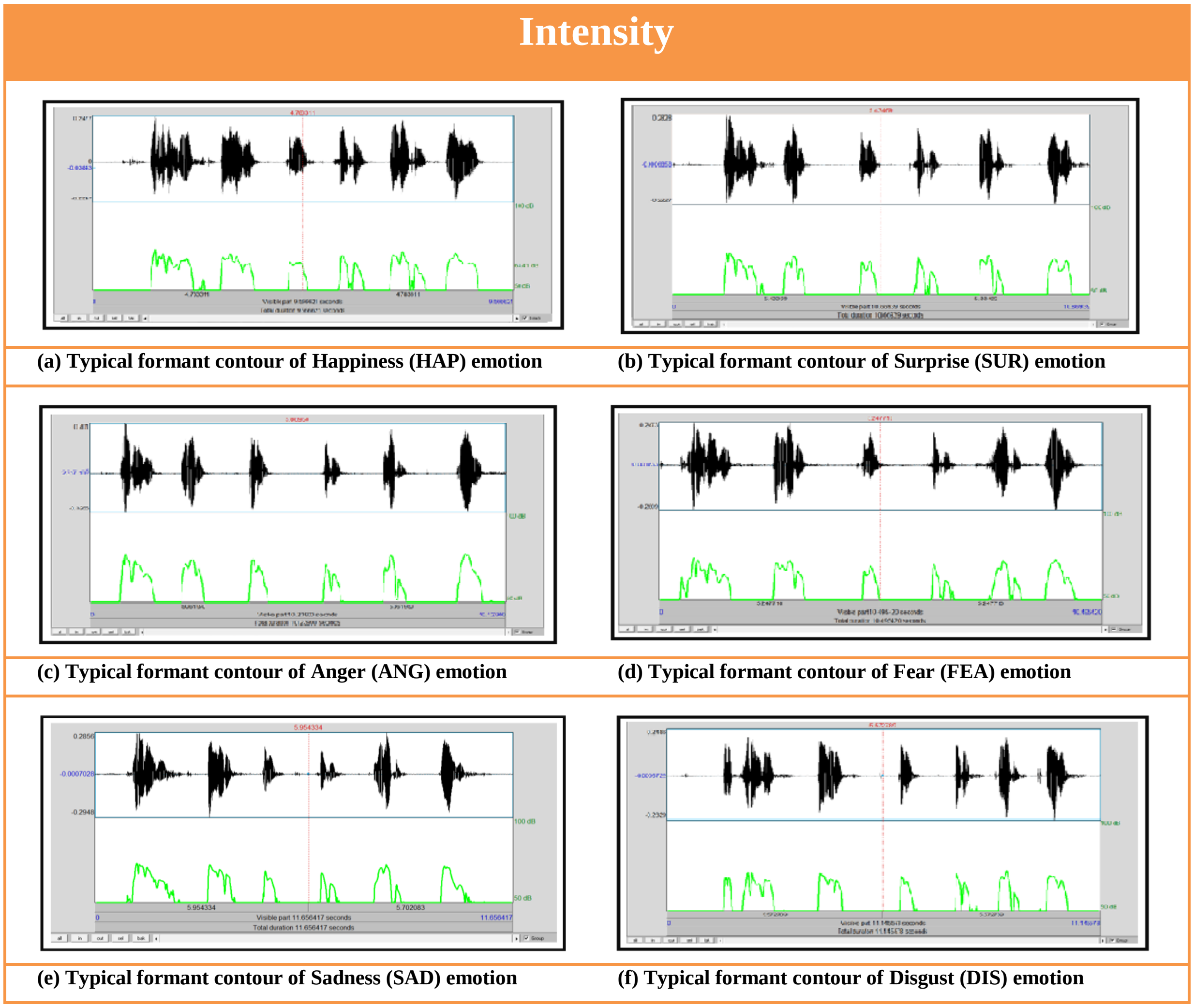
(e) Typical intensity contour of Sadness (SAD) emotion

(f) Typical intensity contour of Disgust (DIS) emotion

Fig. 3-17 Asian speaker typical intensity contour of basic emotions.

Further empirical results discussed, for American speakers, as it can be seen in Figure 3-18, anger, surprise and happiness have the highest energy and intensity, while fear has the lowest intensity.

It is straightforward to infer that the difference between results is due to the difference between the cultures to which the speakers belong. Categorizing the emotions into “high intensity” or “low intensity” can be of great help to increase and design algorithms for emotion recognition.



Intensity

(a) Typical formant contour of Happiness (HAP) emotion

(b) Typical formant contour of Surprise (SUR) emotion

(c) Typical formant contour of Anger (ANG) emotion

(d) Typical formant contour of Fear (FEA) emotion

(e) Typical formant contour of Sadness (SAD) emotion

(f) Typical formant contour of Disgust (DIS) emotion

Fig. 3-18 American speaker typical intensity contour of basic emotions.

3-1-4 speech rate features extraction

Human listeners are able to understand soft, loud, fast and slow speech. The speech rate determines the speaking rate in syllables per minute (SPM). Speech rate recognizers will have implemented for Human-Robot interaction. Speech rate is typically defined as the number of words spoken divided by the time of speech. For emotion recognition in sound signals, speech rate is an important factor as well. Human listeners are able to understand both fast and slow speech. Fast speech have been used in angry and fear emotional communications. The PRAAT software expresses speech rate in seconds, which means the time taken to pronounce the analyzed sentence. A notable result (see Table 3-1) is that anger and fear have the lowest speech rate for European speakers, meaning that the sentences pronounced with anger or fear are pronounced faster, while happiness has the lowest speech rate.

Table. 3-1 Speech rate on emotions (average of the 30 experimental tests) for European, Asian and American.

Emotion Quality	Speech rate for European speakers	Speech rate for Asian speakers	Speech rate for American speakers
Happiness (HAP)	2.0921 s	1.9457 s	1.9931 s
Surprise (SUR)	1.6439 s	1.7001 s	1.6128 s
Anger (ANG)	1.3176 s	1.8832 s	1.2204 s
Fear (FEA)	1.4863 s	1.7121 s	1.6585 s
Disgust (DIS)	1.5521 s	1.4401 s	1.5736 s
Sadness (SAD)	1.7071 s	1.3764 s	1.4750 s
Neutral (NU)	1.6889 s	1.5343 s	1.5158 s
SUM RATE	**11.489 S**	**11.616 S**	**10.482 S**

Sadness and disgust have the lowest speech rate for Asian speakers, while anger and happiness have the highest speech rate: this results is probably due to the fact that Asian people have bigger emotional reaction to happiness and anger. For American speakers anger and disgust have the lowest speech rate, while happiness and fear have the highest speech rate. In general, happiness and surprise have the highest speech rate, while anger and sadness have the lowest speech rate. Moreover, Americans have the highest speech rate. The proposed system used speech rate with standard binderies for each sentences with different emotion. Therefore, we must attention to that average speech rate for different ethnics group of speakers. For example for Asian speakers speech rate is lower than European and American. Also, speech rate change for different ages. Speech rate is one of the important factor that decrees the time and increases the accuracy of the algorithm for recognition of emotion in speech.

3-2 Facial features extraction

The proposed system for facial feature localization is based on a set of facial geometries and a set of rules for features localization. While facial feature detection is based on definition of Active Facial Shape Models (AFSMs) and Bézier curves, we defined the Facial Action Code System (FACS) corresponding to each facial feature.

Active Facial Shape Models define the shape of statistically facial features such as the eyebrows, eyes, nose, mouth and eyebrows in an image [203].

3-1-1 Eyes and Mouth Localization

Eyes and mouth localization is a very important issue for emotion recognition. This directly influences the localization of eyebrows and nose. The geometry of the eye is simple, so an effective method for eye localization can be quite easily defined. In order to find the middle position of the eye, we proposed the matrix to divide the face into a 9*6 matrix, so that the face is divided into 54 equal cells, each one made of 20*20 pixels (see Figure 3-19). The position of the right eye is generally estimated in the cell of index (4, 2). We calculated the standard distances between two eyes and the anthropometric position of the eyes, so that due to symmetry, the position of the left eye can be found easily. Then, a rectangular boundaries containing the two eyes can be generated.

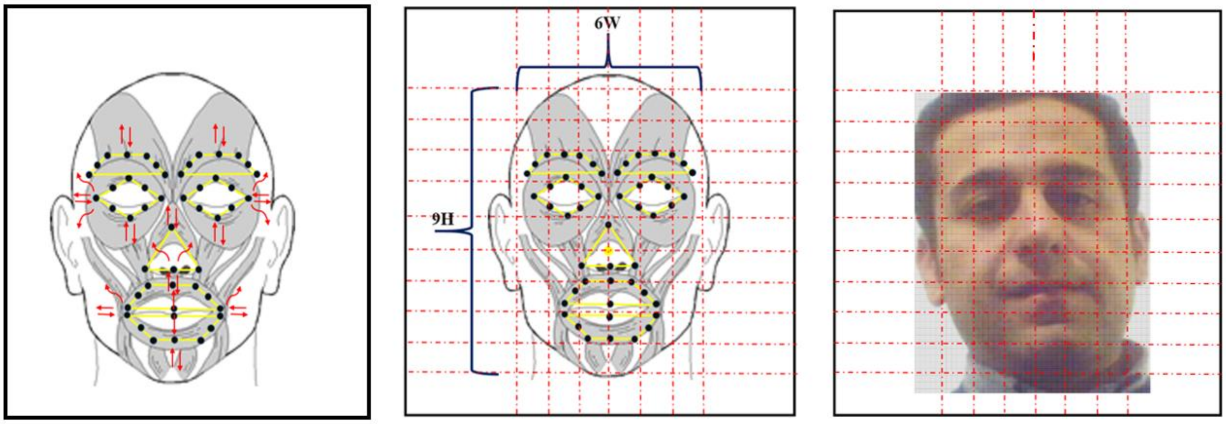


Fig. 3-19 Anthropometric of human face for features detection.

We used skin texture techniques in order to evaluate the amount of eye opening under different emotions (happiness, surprise, anger and sadness): eyelids usually are darker than skin, so they are converted into black pixels, while skin is converted into white pixels.

Mouth is another very important feature for emotion expression. Also, localization of the lips region is very important for recognizing happiness, surprise, anger, fear and sadness. The mouth area has six regions: (whole mouth, lower lip, upper lip, between lips, mouth cavity and teeth) [204]. As shown in Figure 3-19, the mouth occupies in the lower third of the face image. For estimation of mouth position, a rectangular boundary of 80*40 pixels containing the mouth is created.

Lip localization is similar to eye localization, but the system must eliminate the false lip edge and shadow. If the mouth is closed, the task is relatively easy task, but if the mouth is open, edge detection techniques (Sobel filtering) are needed.

The maximum distances between the two lips are reported as the mouth opening value. When the mouth is open teeth detecting techniques must be used. In the first step for teeth detection, we must find the center of the mouth. Then, have been counted the number of white pixels at the center of the mouth boundary, and if the number of the white pixels was equal or higher than one-third of all pixels, the system found the position of the teeth in the image.

In happiness emotion the length of the mouth stretches, while in surprise the length of the mouth decreases and instead the mouth width increases. For calculating the mouth length, the system computes the horizontal distance between the lip corners, which have the darkest color pixels in the mouth region.

3-1-2 Eyebrows and Nose Localization

Eyebrows and nose generally are detected by using facial geometries and Active Facial Shape Models (AFSMs). Inner and outer parts of the eyebrows are especially important for recognizing surprise, happiness, anger and sadness.

Eyebrows location is based on the forehead detection and is quite easy because eyebrows have a simple shape. The most important part in the algorithm is to accurately localize the eyebrows boundaries. As shown in Figure 3-20, the eyebrows rectangular boundaries have the same size and are located above of eye boundaries. They lie inside two 60*20 pixels frames.

However, in eyebrows localization shadows near the eyebrows, as well as thin or light-colored eyebrow hairs can decrease the accuracy of the system.

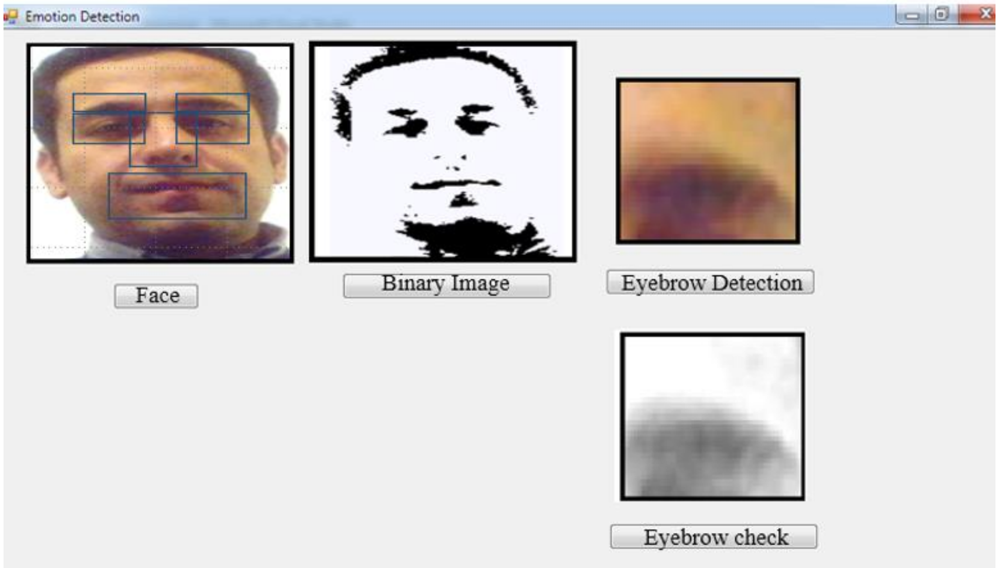


Fig. 3-20 Eyebrows detection and localization boundary.

The nose is a fixed element in the center of the face. The system detected the nose based on its geometry. For detecting the nose and the nose side wrinkles, we must select the window that contains the nose. The nose rectangular boundary lies above the upper lip and has a size of 40*40 pixels.

Table. 3-2 summarizes the procedure for localization of the face features.

Area	Location	Width	Height
Eyebrows	Top left and right part of the face below forehead	Right and Left 60 pixels	Second area from top 20 pixels
Eyes	Top left and right part of the face below eyebrows	Right and Left 60 pixels	Second area from top 20 pixels
Nose	Center part of the face upper lip	In Center 40 pixels	In Center 40 pixels
Mouth	Bottom part of the face	In Center 80 pixels	First area from bottom 40 pixels

After that, the next step of the algorithm is to extract the facial feature. As it can be seen in Figure 3-21, finally five types of features can be recognized with some accuracy.

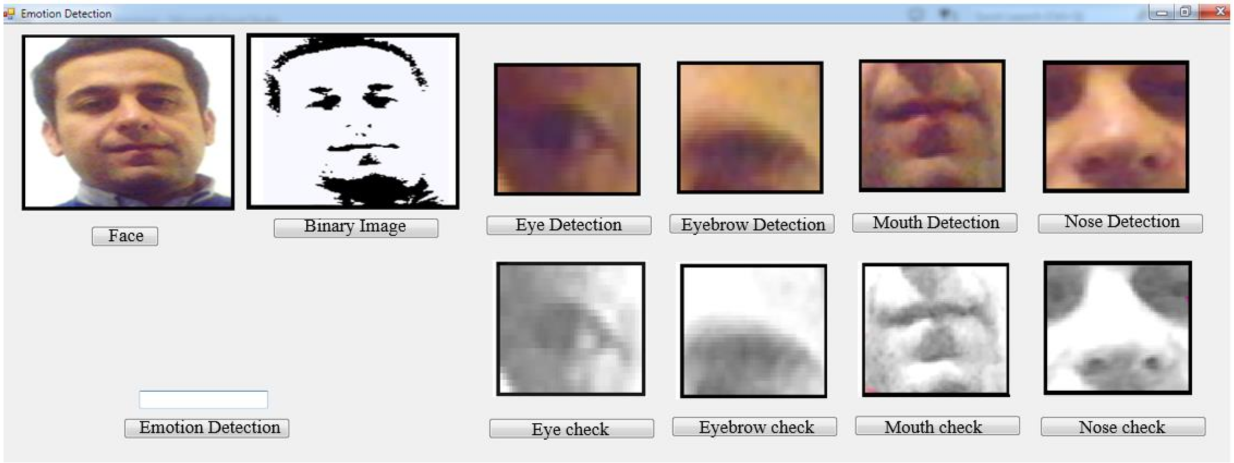


Fig. 3-21 Extraction main facial features from image.

3-3 Speech features classification system

Extraction rules for classification of emotion can extract form speech signals. In this section we wants to developing the classifier sound system. Then in capture 4, have been evaluated and tested these extracted rules and proposed system.

As mentioned above the algorithm in methodology described in the foregoing analyzes the emotion that belongs to the category of “high and low" intensity. An overview of the block diagram for extracting and classifying the multi-level emotion is shown in Figure 3-22. If the algorithm compare PRAAT plat and distinguish the speech falls into the “high intensity” category, it is further analyzed to recognition emotion between happiness, surprise, anger and fear.

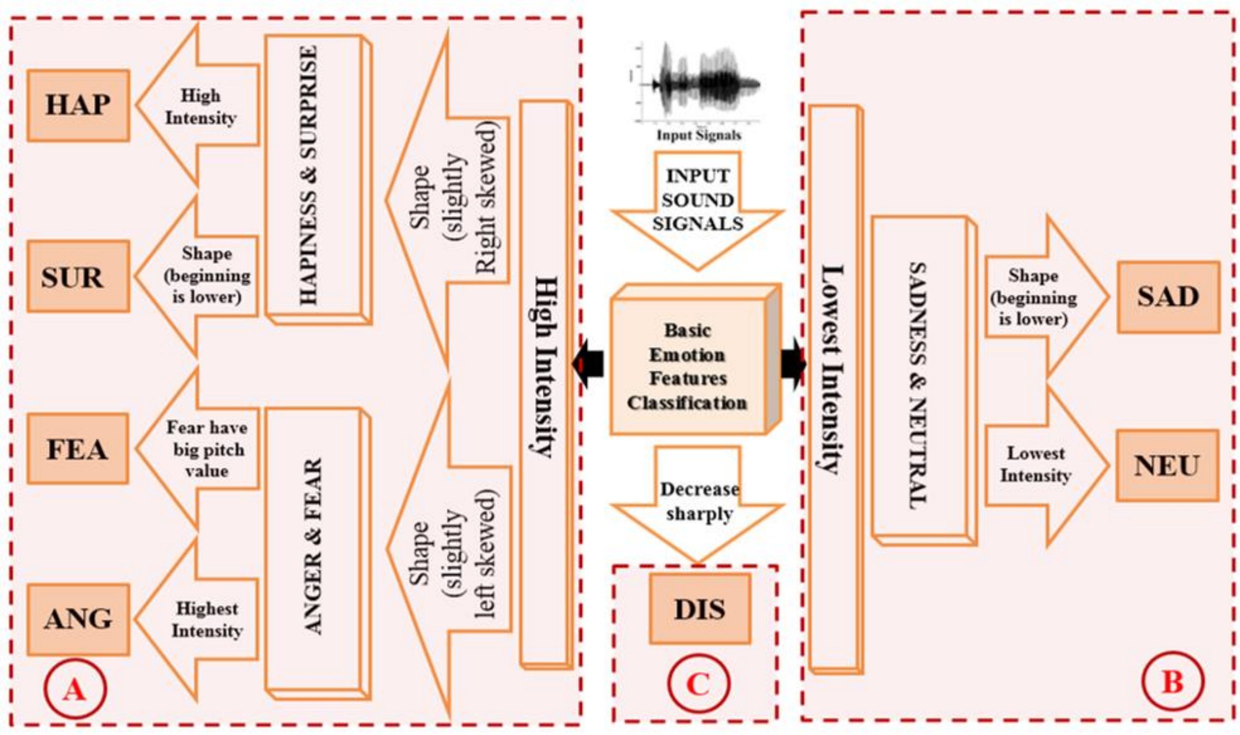


Fig. 3-22 Model of the system for speech emotion classification and recognition.

In the same way, the system check intensity graph in order to distinguish “low intensity” for control between neutral and sadness emotion. Also, we can be used the graphs of pitch and speech rate. The proposed model for distinguish between neutral and sadness control ending pitch (sadness emotion has lower ending pitch).

If the shape of signal is “left skewed” and the ending of the signal is higher than the beginning, the emotion must be fear or anger emotion. Base on rule extraction In order to distinguish between fear and anger, the algorithm must compare the intensity: anger emotion has the highest intensity, thus it is easily distinguishable. If the shape of the signal is “right skewed”, it must further analyzed in order to distinguish between surprise and happiness.

To this aim, the algorithm must check the pitch value and intensity: happiness has the highest pitch range and pitch peak while surprise has the highest pitch value. If the speech does not belong to any of the aforementioned emotions, it is classified as disgust.

3-4 Facial features classification system

Researchers in facial features expression, mostly focus on defining a universal set of features that convey emotional clues and try to develop classifiers that efficiently model these features. The system for recognition of emotion needed to present methods for discovering emotions, modeling and evaluating the results. In this section we used the mathematical formulation (Bézier curves) and the geometric analysis of the facial image, based on measurements a set of Action Units (AUs) and Facial Action Code System for classify emotion.

3-4-1 Bézier curves for facial features classification

Converting the eyebrows, eyes and mouth features to fitted curves is the next step before detecting emotions. Bézier curves can approximately represent 2D facial shapes. To apply the Bézier curves, we need to generate some contour points for control information. For example, in this paper we chose eight neighbor points for each facial region (left/right eyes, left/right eyebrows, upper /lower lips). The aim of the Bézier curve method is to interpolate a sequence of 8 proposed points.

As it can be seen in Figure 3-23, the system generates Bézier curves which can represent the principal facial regions.

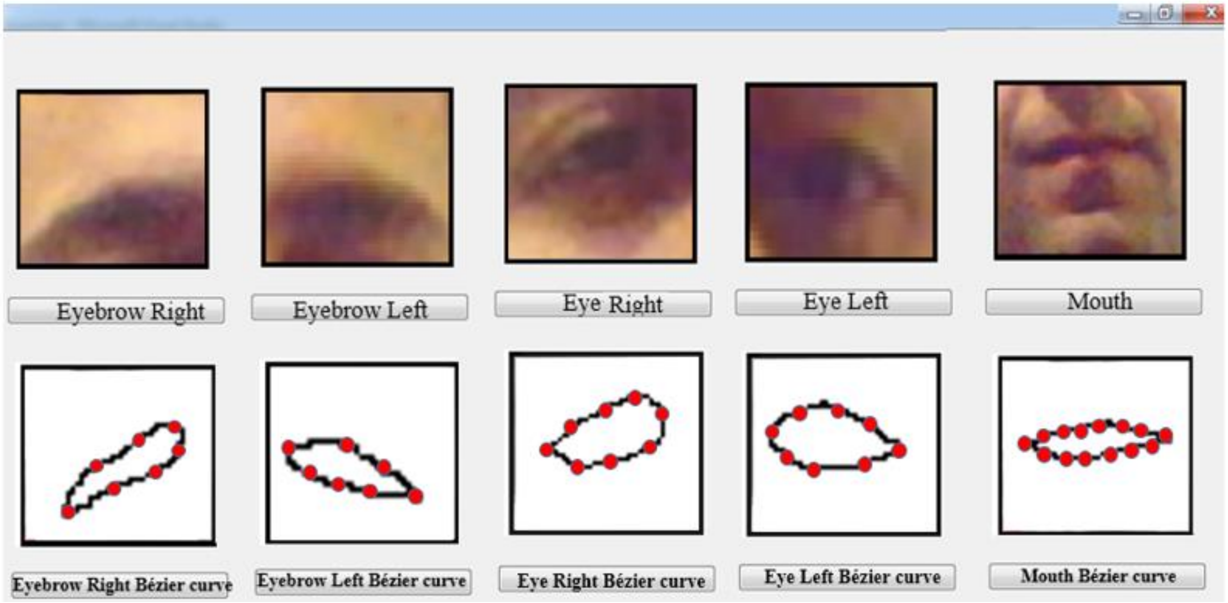


Fig. 3-23 Result of Bézier curves for five principal facial features.

The Bézier representation of the curves can then be employed to determine the distances between the points of interest. Then the system saves all the computed values into a database and finally finds the nearest matching pattern with related emotion. The system database contains index of seven kinds of emotion with different features, which are extracted based on the rules in Table 3-3. If the input graphic curve did not match with the emotion in database, we added this emotion manually to the database in the training phase.

As mentioned above the proposed algorithm for facial emotion recognition is divided into two steps. The first step included Bézier curve analysis and measurement of each input image base on graphic curve. In the second step, the algorithm extracted facial Action Units (AUs) and calculated the distance parameters between facial feature on input face image and normal face image. When the algorithm starts the second step, automatically the system begins the training section after checking the Facial Action Codes System (FACS).

Table. 3-3 Basic emotion with different features for recognizing facial expression.

EMOTION	Eyebrows	Eyes	Nose	Mouth	Lips	Teeth	Other
Happiness	Outer eyebrows corners Stretch	-	Widen	Increases length horizontally	Movement muscle that orbits the eye	Show 50% teeth	Cheeks Pushed up – Chin Rise up
Surprise	Completely raised up	Widen	-	Increases length vertically	Both open with high slope	Show 10% teeth	
Anger	Down, Inner eyebrows corners go together	Glare	-	Compress	Narrow	-	Nostril Compress
Fear	Raised and pulled together	Raised upper of the eyes	-	-	Slightly stretched horizontally	Show 30% teeth	
Disgust			Side wrinkling	-	Supper lip raised	-	Nostril deeper
Sadness	Inner corner of the eyebrows rises up	Losing focus in eyes	-	Lower mouth Corner	Slight pulling down of lip corners	-	Drooping upper eyelids

3-4-1 Action Units (AUs) for facial features classification

One of the most important way to approach facial emotion classification and recognition is use facial muscle movements. The specific application on facial coding system is to describe the motion of muscles and the combination of different components.

In this research we chose a set of 32 facial Action Units (AUs) that was useful for basic emotion expression. They are shown and described in Tables 3-4.

Table. 3-4 Basic emotions and 32 Action Units resulting from feature extraction.

Action Units (AU_s)	Feature	Description	Measurement name
0	**Neutral**		
1	**Eyebrows**	**Inner eyebrows corners Slope**	Inner Eyebrow Raiser
2			Inner Eyebrow Depressor
3		**Outer eyebrows corners Slope**	Outer Eyebrow Raiser
4			Outer Eyebrow Depressor
5		**center eyebrows**	center Eyebrow Raiser
6			Eyebrow Lower
7	**Eye**	**Eye shape**	Outer Eye Raiser
8			Outer Eye Depressor
9			Squint
10			Eyes Tightened
11		**Eyes opening**	Eyes open
12	**Nose**	**Nose shape**	Nose wrinkling
13			Nose width
14		**Nostril**	Nasolabial Deepener
15			Nostril Compress
16	**Mouth**	**Mouth shape**	Mouth open
17			Mouth close
18			Mouth Stretch
19		**Lower Lip**	Lower Lip Corner Depressor
20			Lower Lip Corner Puller
21			Lower lip to chin
22		**Upper Lip**	Upper Lip Corner Depressor
23			Upper Lip Corner Puller
24			Upper lip to nose

25		Lips shape	Lip stretcher
26			Lip Tightener
27			Lip Suck
28	Head	Rotatio head	Head up
29			Head down
30			Head tilt
31			Head turn (left\ right)
32		Forehead	Forehead wrinkling

As mentioned above, we compared the relationship among AUs, on the basis of a nearest neighbor on the current and input image. After extracting the facial features correctly, we employed the 32 Action Units suitably represented and divided into: neutral with references (AU$_0$), upper face AUs (AU$_1$_AU$_{15}$), lower face AUs (AU$_{16}$_AU$_{27}$) and head position (AU$_{28}$_AU$_{32}$).

As it can be shown in the bar diagram in Figure 3-24, we grouped the 32 AUs into five categories that contribute toward the facial emotion expression.

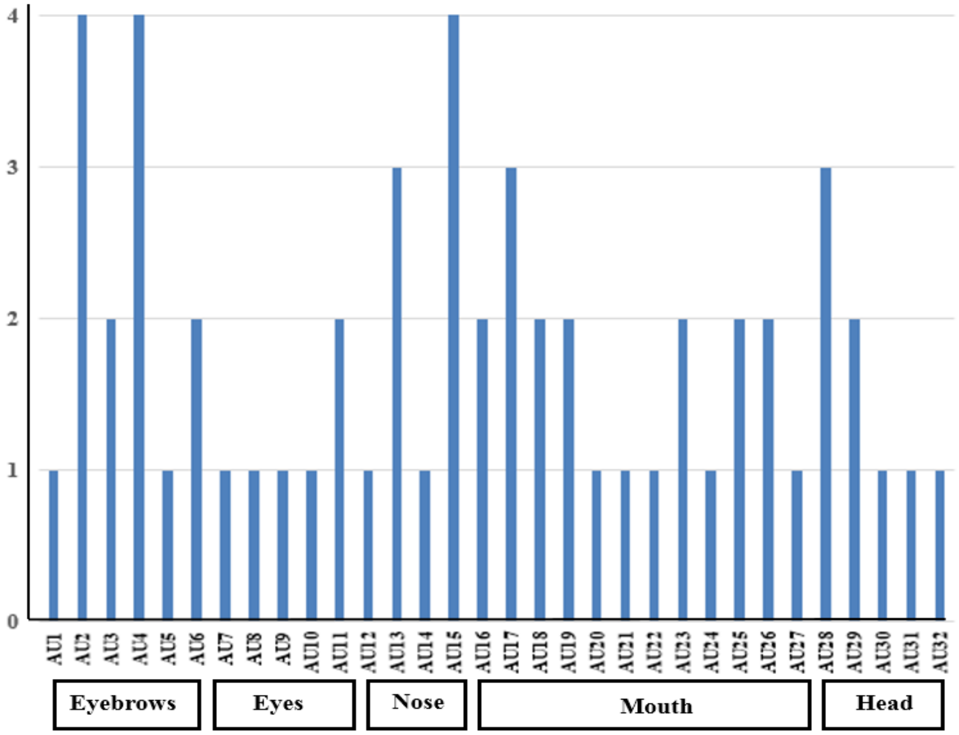
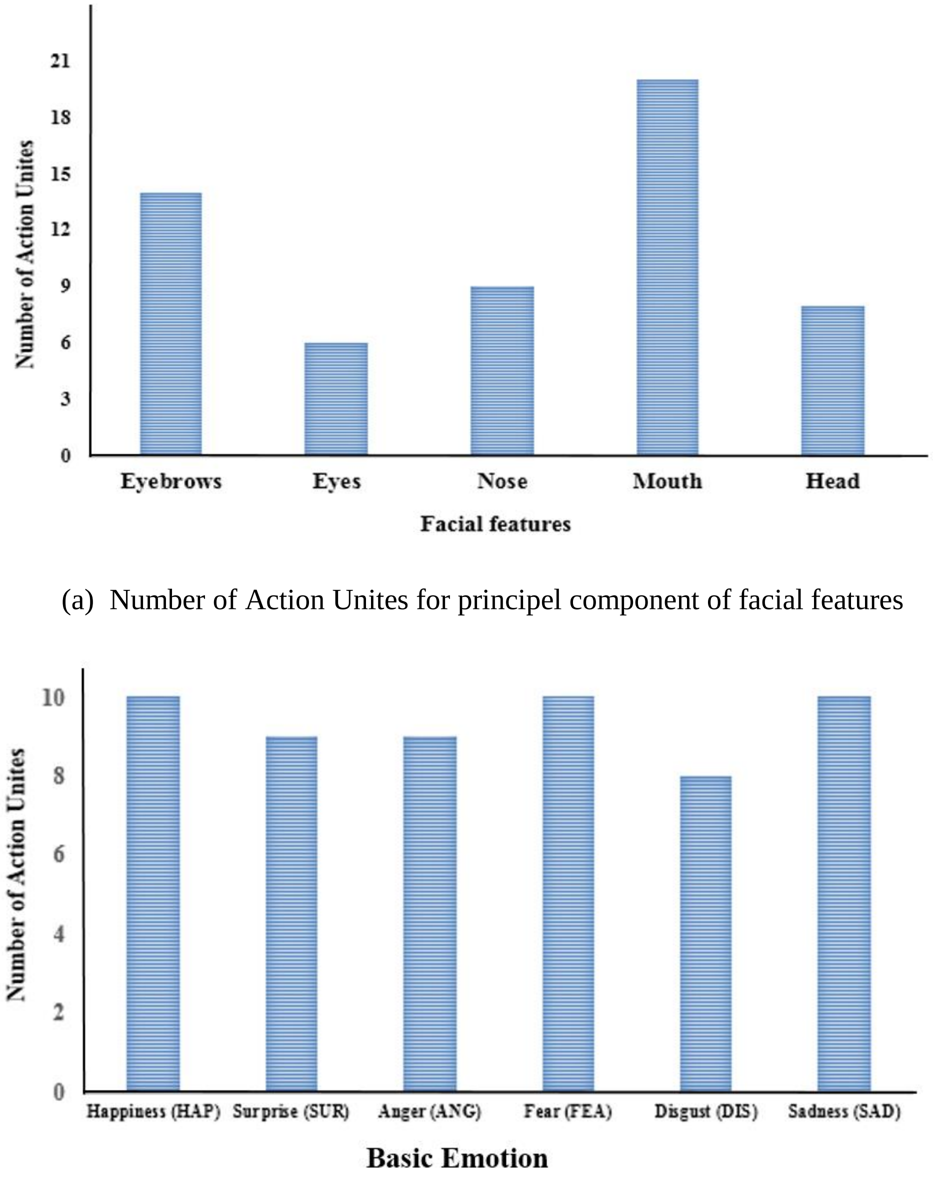


Fig. 3-24 Mutual interaction between facial features and Action Units (AUs) in the proposed system.

Figure 3-25 (a) shows the number of action unites for principel component of facial features and Figure 3-25 (b) shows the role of mouth and eyebrows expressions in different emotions; namely, the mouth with 20 and the eyebrows with 12 checks on each image. In the second place, the proposed system used nose and eyes Action Units (AUs). Additionally, we also used the head motion to increase the accuracy of the system. Head position can have significant variations between different ethnic groups.



(a) Number of Action Unites for principel component of facial features

(b) The roles of Action Unites for diffrent basic emotion

Fig. 3-25 Interaction between facial features, basic emotions and Action Units (AUs).

By increasing the performance of the algorithm, based on the role of AUs in each emotion we added a weight to each feature Action Code. Action Codes describe the same facial expression category and can be used to compare facial repertoires.

In the proposed system, the Action Codes are classified into four categories. The mouth Action Codes with “50 percent” weight, eyebrows Action Codes with “25 percent” weight, eyes Action Codes with “12.5 percent” weight and the remaining “12.5 percent” weight for nose and head position.

As illustrated in Figure 3-26, we connected AUs to extract facial features with a group of landmark points. We chose the facial matrix to determine the position of Action Codes. The facial matrix has a size of 1800*1200 pixels with 52 points on the Facial Action Codes. Also, have been selected a subset of 52 landmark points. The facial matrix has four regions: (eyebrows 16 point, eyes 16 point, nose 4 point and lips 16 point). But some of this landmarks are very close to each other.

By computing the AUs distance between the input image and normal image, we developed an algorithm with the mathematical concepts for measurement of the distance between the Action Codes.

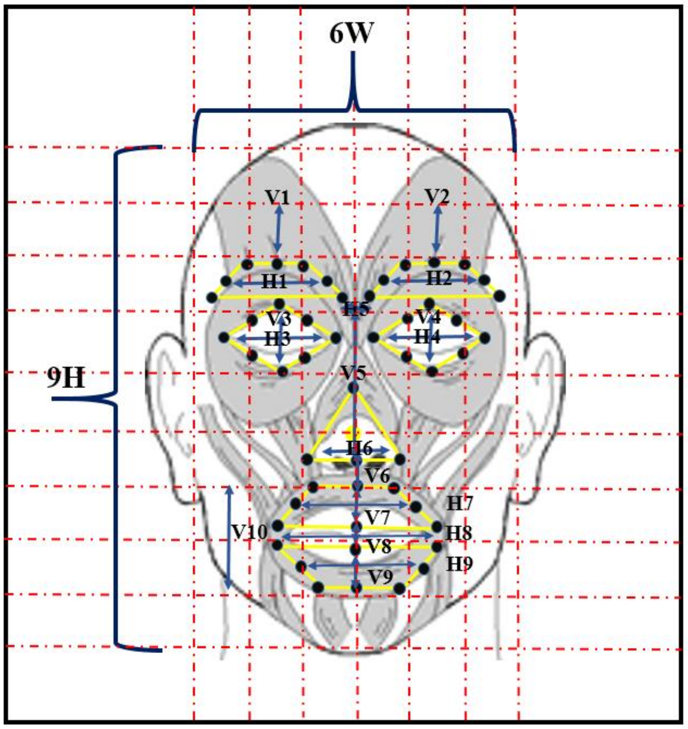


Fig. 3-26 Facial matrix and 52 landmark points (eyebrows 16, eyes 16, nose 4, lips 16).

As it can be seen in Table 3-5, the emotions are associated with the distance extracted from facial codes points.

Table. 3-5 Facial features distance measurements (15 codes) for basic emotion expression.

Code	Validation distances	Distances
D1	Distance of eyebrows top to forehead	**V1-V2**
D2	Eye height	**V3-V4**
D3	Distance of nose top to eyebrow's middle	**V5**
D4	Distance of Upper Lip to nose	**V6**
D5	Upper Lip height	**V7**
D6	Distance of Upper Lip to Lower Lip	**V8**
D7	Lower Lip height	**V9**
D8	Mouth height	**V10**
D9	Eyebrow width	**H1-H2**
D10	Eye width	**H3-H4**
D11	Distance of inner eyebrow left corner to inner eyebrow right corner	**H5**
D12	Nose width	**H6**
D13	Upper Lip width (corner of lip position)	**H7**
D14	Mouth length	**H8**
D15	Lower Lip width (corner of lip position)	**H9**

In order to compute the distance between the 15 Action Codes, the Mahalanobis algorithm and the facial geometry was used in the software program [200]. Mahalanobis distance used in facial expression classification techniques and computed of the appropriate dimension of the facial code points. The algorithm automatically measures 15 distances, respectively: D_{14}, D_8, D_6, D_{13}, D_{15}, D_7, D_5, D_{11}, D_1, D_4, D_9, D_2, D_{10} and D_3.

Different distances between the emotional facial images and normal facial image were extracted Facial Action Points (FAP). Table 3-6 shows the average distances from the corresponding values and neutral. Parameters displayed negative deviation, positive deviation or no substantial deviation from the neutral value (Table 3-6). The trend of variation of different parameters with respect to neutral values for different expressions helps in the effective emotion recognition.

Table. 3-6 shows the average distances from the corresponding values and neutral.

Validation distances	Emotion						
	NEU	**HAP**	**SUR**	**ANG**	**FEAR**	**SAD**	**DIS**
Distance of eyebrows top to forehead	0.00	-2.45	-7.20	5.35	-1.55	-0.75	-1.35
Eye height	0.00	-2.30	4.83	-1.95	-1.12	-2.43	-2.25
Distance of nose top to eyebrow's middle	0.00	-2.10	3.24	-3.56	-0.52	-0.79	1.34
Distance of Upper Lip to nose	0.00	-1.84	1.53	-1.23	-1.00	10.76	-0.32
Upper Lip height	0.00	-0.61	-0.42	-0.85	0.34	0.64	0.12
Distance of Upper Lip to Lower Lip	0.00	4.22	7.48	-0.38	1.14	-0.53	0.62
Lower Lip height	0.00	-0.43	-0.45	-0.87	0.30	0.69	0.14
Mouth height	0.00	6.85	12.38	-1.53	2.56	-1.73	1.53
Eyebrow width	0.00	2.12	-4.28	-1.85	1.26	1.64	0.52
Eye width	0.00	5.33	-7.54	-3.42	2.11	3.27	1.17
Distance of inner eyebrow left and right corner	0.00	0.72	0.50	-0.52	0.42	-0.23	0.21
Nose width	0.00	2.29	-1.14	1.37	0.74	0.54	4.58
Upper Lip width (corner of lip position)	0.00	18.55	-11.30	4.50	-3.37	5.56	3.42
Mouth length	0.00	15.33	-13.25	4.35	-4.23	5.10	3.19
Lower Lip width (corner of lip position)	0.00	17.42	-11.28	4.48	-3.47	6.25	3.56

Finally, when calculating the Action Codes distance, we solved slight overlapping between sets of very close points.

CHAPTER 4

Evaluation and rules extraction

on human emotion recognition system

4-1 Rules extraction for emotion recognition system

Accordance with past research there is great opportunity for rules extraction in complex system. We explain that, by means of to have large impact in rules extraction algorithm the system must be extract high level of generality in sound and facial observation.

Most of the rules extraction system has been used data only within narrow class (like seven basic emotion class). Researchers mostly focus on defining a universal set of rules that convey emotional clues and try to develop that efficiently of this rule on emotion recognition program.

The most important issues in rules extraction is discovering of applicable and general rules in order to transfer the optimal results on emotion recognition system or training method. The system for recognition of basic emotion needed rules from pitch, formant, intensity, speech rate, AUs and Facial features Action Code System. The proposed algorithm has led to develop an applicable rules in order to provide portable speech and facial source code to use in the emotion recognition system.

4-1-1 Rules extraction in sound signals and facial feature extraction

In this section the most important issues is discovering and developing the emotional speech and facial feature rules. For practical purposes, the important goals are discovery of which feature rules are the most informative and meaningful for recognition of emotions.

In order to use rules extraction system first we defined universal set of rules for emotion recognition and classification based on human facial features and human sound signals. Secondly we evaluated and tested these rules. Extraction rules for recognition of emotion base on speech signals (verbal interaction) can extract form pitch peak, pitch value, pitch range, intensity, formant, and speech rate.
Table 4-1 summarizes group of this rules. The importance results of this section is distinguish emotions between the different ethnic groups.

For example one of these rules is: positive valence emotion (happiness and surprise) have right-skewed pitch contours, while negative valence ones (anger and fear) have slightly left-skewed pitch contours and the ending of the signal is higher than the beginning. Neutral and sadness have the lowest ending pitch contours.

Table. 4-1 Rules for emotion classification based sound signal.

Observation	Set Of Rules
Observation 1	Happiness has highest average pitch peak and intensity, while has lowest speech rate
Observation 2	Anger has the highest pitch peak, pitch values, speech rate, intensity
Observation 3	Fear and sadness are associated with the lowest energy
Observation 4	Disgust has the lowest pitch value and decreases sharply
Observation 5	Positive valence emotion (happiness and surprise) have right- skewed pitch contours
Observation 6	Negative valence emotion (anger and fear) have slightly left-skewed pitch contours
Observation 7	Sadness and neutral have the lowest ending pitch contours

Also, we focused on defining a universal set of AUs that convey facial expression clues. Based on each Action Unit, a set of rules are generated in terms of the feature representation, as well as a few simple combination relationships among Action Units (see Table 4-2). As mentioned in capture 3 we used motion of muscles and the combination of different components (Table 3-3 and Table 3-4) with defining the new action unite for proposed system.

In emotion detection system, using more Action Units makes the recognition system more accurate.

Table. 4-2 Basic emotion recognition base on extracted facial Action Units.

EMOTION	Action Units (AU_s)
Happiness	AU_1+ AU_3+ AU_7+ AU_{13}+ AU_{16}+ AU_{18}+ AU_{20}+ AU_{23}+ AU_{25}+ AU_{28}
Surprise	AU_2+ AU_4+ AU_5+ AU_{11}+ AU_{13}+ AU_{16}+ AU_{21}+ AU_{24}+ AU_{28}
Anger	AU_2+ AU_3+ AU_{10}+ AU_{15}+ AU_{17}+ AU_{26}+ AU_{27}+ AU_{29}+ AU_{32}
Fear	AU_2+ AU_4+ AU_6+ AU_{11}+ AU_{13}+ AU_{15}+ AU_{17}+ AU_{18}+ AU_{25}+ AU_{26}+ AU_{28}
Disgust	AU_4+ AU_9+ AU_{12}+ AU_{15}+ AU_{19}+ AU_{23}+ AU_{30}+ AU_{31}
Sadness	AU_2+ AU_4+ AU_6+ AU_8+ AU_{17}+ AU_{14}+ AU_{15}+ AU_{19}+ AU_{22}+ AU_{29}

For implementation of the facial features rule extraction have been used Action Units (Table 4-2), we employed the 32 Action Units suitably represented and divided into: neutral with references (AU_0), upper face AUs (AU_1_AU_{15}), lower face AUs (AU_{16}_AU_{27}) and head position (AU_{28}_AU_{32}).

The proposed algorithm starts with the "Mouth Action Codes" step. The mouth region is divided into upper lip and lower lip part. They are especially important for recognizing surprise, happiness, anger and sadness. Accuracy of the system depends on the mouth center positions and lips corner. The most important issues in this section were the different between human ethnic groups. For example, Asian lips are thicker than European ones.

If the Action Code of the lips corner is "upper" or "higher" than the normal, the system recognizes happiness.

If the Action Code of the lips is "open" or "tighter than the normal", the system recognizes surprise or anger.

In the same way, the algorithm runs the "Eyebrows Action Codes" step. This step is important for recognizing sadness, anger, surprise and fear. The eyebrows' inner and outer corner is important for sadness and anger. The shape of eyebrows is also important for anger and fear emotion.

In the third step, the algorithm considers the "Eyes Action Codes". It is especially important for recognizing sadness, anger, surprise and happiness. If we compare happiness and surprise, eyes opening are the most important parameter. The eye shrink is useful for recognizing anger and fear emotion.

We can use the nose and head position for recognizing disgust and anger. The head landmark points are particularly helpful for describing anger and disgust. For this purpose, the algorithm must check the nose wrinkling and position of the head.

Finally, we defined a set of universal observations for emotion recognition based on the procedure of speech signals and facial feature rules extraction on emotion recognition in hybrid system.

Some examples of these observations are shown in the Table 4-3. For example one of these rules is: Anger emotion. It has the highest pitch peak, pitch values, speech rate and intensity, slightly left-skewed pitch contours (the ending of the signal is higher than the beginning), eyes gaze, Inner eyebrows corners go together, mouth compress and lips thickness.

Table. 4-3 Set of rules for emotion recognition.

Observation	Set Of Rules
Observation 1	Happiness has highest average pitch peak and intensity, while has lowest speech rate, right- skewed pitch contours, open mouth and teeth is visible.
Observation 2	Surprise has the highest pitch range and high pitch peak, right- skewed pitch contours, size of the eye is wider than normal, mouth length decreases and the mouth height increases.
Observation 3	Anger has the highest pitch peak, pitch values, speech rate and intensity, slightly left-skewed pitch contours, eyes gaze, Inner eyebrows corners go together, mouth compress and lips thickness.
Observation 4	Fear has the lowest energy, slightly left-skewed pitch contours, eyebrows raised and pulled together, mouth Slightly stretch.
Observation 5	Sadness has the lowest ending pitch contours, eyes tightening, inner eyebrows corner rises up, obliquely lowering of the lip corners.
Observation 6	Disgust has the lowest pitch value and pitch graph decreases sharply, nose side have wrinkles.
Observation 7	Neutral has lower end pitch and facial is in normal position.

4-2 Evaluation of the results of implementation system

By means of implementation and validation of the algorithm for recognition basic emotion (happiness, sadness, disgust, surprise, fear, anger and neutral) we used two databases. The system was evaluated by actor and actress under audio, visual and audio-visual conditions in the laboratory

The experimental tests for the first database consists of 630 sound file and 840 sequences 2D image on 30 individuals (15 female and 15 male, 20 to 48 years old) participant form different ethnic groups (Europeans, Middle East Asians and Americans).

Also, in order to evaluate the hybrid algorithm and checking the accuracy of our method in speech recognition and facial emotion expression phase have been used another data base. The proposed data base consists of recorded sound file with different sentences and speakers. Also, in facial emotion recognition have been used comprehensive and rich emotional image database named Cohn-Kanade facial expression database [205].

The Cohn-Kanade AU-Coded facial expression database have been used for research in facial image analysis and synthesis and for perceptual studies. Version 1, Cohn-Kanade data base includes six prototypic emotions (i.e. happiness, surprise, anger, fear, disgust, and sadness), and includes 486 sequences from 97 posers.

4-2-1 Results on speech emotion recognition system

In almost any category of emotion we have successfully identified certain characteristic features and these formed the basis for classification of different emotions. The typical pitch and intensity contours characterizing each of the basic emotions. Happiness and surprise have the highest pitch range and pitch peak. In pitch peak and intensity analysis, happiness and anger are distinguished faster than other emotions for European speakers, while in order to distinguish fear and disgust, the algorithm must check all the acoustic features.

In pitch value analysis, surprise, happiness and fear can be distinguished quicker than other emotions. For Asian speakers in pitch peak, happiness and anger are distinguished faster than other emotions. Also, anger has the highest and sadness has the lowest range of intensity for Asian speakers.

The procedure for emotion recognition from speech can be implemented using a Likert type scale (see Table 4-4), which categorizes the basic emotions based on discrete values of pitch peak, pitch range, pitch value, intensity and speech rate. Perceptual listening tests had been conducted to verify the correctness of emotional content in the recordings under each of the seven categories of emotion.

Human experts (3 persons) listened to the sample sentences and indicated the perceived emotion from a list of six emotions (apart from the neutral). The listeners rated the sound files for each emotion on a five point scale ranging from excellent to bad through very good, good and fair.

Table. 4-4 Likert type scale for emotion recognition.

MOTION QUALITY	**Difference between highest and lowest of signals**		**Top of the signal shape**		**For long time is higher than the average**		**sound power**	**Model Result**	**Total quality distinguish**					**Total Results**
	Pitch range		**Pitch peak**		**Pitch values**		**intensity (Energy)**	**SUM**	**Likert type scale**					**SUM**
Happiness	European	7	European	7	European	6	6	**925**	EX	VG	G	F	B	**EX**
	Asian	6	Asian	6	Asian	4	6		EX	VG	G	F	B	
	American	6	American	5	American	5	5		EX	VG	G	F	B	
Surprise	European	6	European	7	European	5	7	**835**	EX	VG	G	F	B	**EX**
	Asian	5	Asian	5	Asian	6	3		EX	VG	G	F	B	
	American	5	American	6	American	6	6		EX	VG	G	F	B	
Anger	European	4	European	6	European	2	7	**931**	EX	VG	G	F	B	**VG**
	Asian	7	Asian	7	Asian	2	7		EX	VG	G	F	B	
	American	7	American	7	American	3	7		EX	VG	G	F	B	
Fear	European	5	European	3	European	5	3	**732**	EX	VG	G	F	B	**G**
	Asian	4	Asian	4	Asian	7	5		EX	VG	G	F	B	
	American	3	American	4	American	5	2		EX	VG	G	F	B	
Disgust	European	2	European	4	European	1	4	**670**	EX	VG	G	F	B	**F**
	Asian	3	Asian	3	Asian	1	1		EX	VG	G	F	B	
	American	4	American	3	American	1	1		EX	VG	G	F	B	
Sadness	European	1	European	1	European	4	2	**635**	EX	VG	G	F	B	**G**
	Asian	2	Asian	2	Asian	3	4		EX	VG	G	F	B	
	American	2	American	1	American	2	4		EX	VG	G	F	B	
Neutral	European	3	European	2	European	3	1	**710**	EX	VG	G	F	B	**G**
	Asian	1	Asian	1	Asian	5	2		EX	VG	G	F	B	
	American	1	American	2	American	7	3		EX	VG	G	F	B	

For validation of the algorithm we intend to compare the accuracy and recognition rate of the basic emotions in speech analysis phase. Table 4-5 show that the anger and happiness have the highest emotion detection rate and recognition accuracy of these emotions is higher than 80%. Fear, disgust and sadness have the lowest emotion recognition rate. It appears that sadness, disgust and fear are not so easy to distinguish from speech signals; however, this is also true for humans.

Table. 4-5 Percentage of emotions recognized correctly in speech analysis phase-Part (a).

Emotion quality	Results of the algorithm part (A)	Emotion recognition rate
Happiness	92.5%	1.5393 s
Surprise	83.5%	2.0812 s
Anger	93%	1.2054 s
Fear	73%	1.5736 s
Disgust	67%	2.5434 s
Sadness	63.5%	1.7798 s
Neutral	71%	1.8901 s

As it can be seen in Table 4-6, in order to check and validate the results it appears that sadness, disgust and fear are not so easy to distinguish from speech signals; however, this is also true for humans.

Table. 4-6 Percentage of emotions recognized correctly.

Emotion quality	Results of model	Human expert
Happiness	92.5%	Excellent
Surprise	83.5%	Excellent
Anger	93%	Very good
Fear	73%	Good
Disgust	67%	Fair
Sadness	63.5%	Fair
Neutral	71%	Good

Figure 4-1 shows the pitch characteristics of the six basic emotions in a diagram that extracted from PRAAT and used it codes in the software program. This representation is complementary to those described above. We performed experimental tests to assess the effectiveness of the algorithm. System used graph of the pitch (peak, range and value), intensity and speech rate.

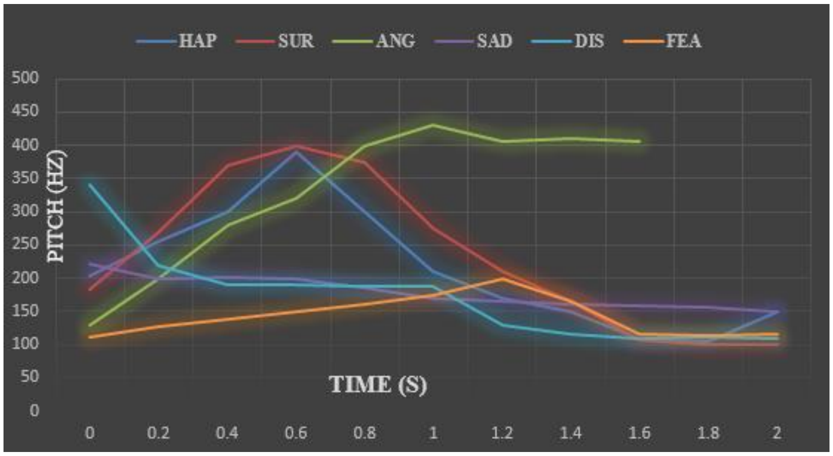


Fig. 4-1 The location of the pitch (range, peak and value) graph for six basic emotions.

Figure 4-2 Shows the characteristics of the seven basic emotions in a bar diagram. This representation is complementary to those described above.

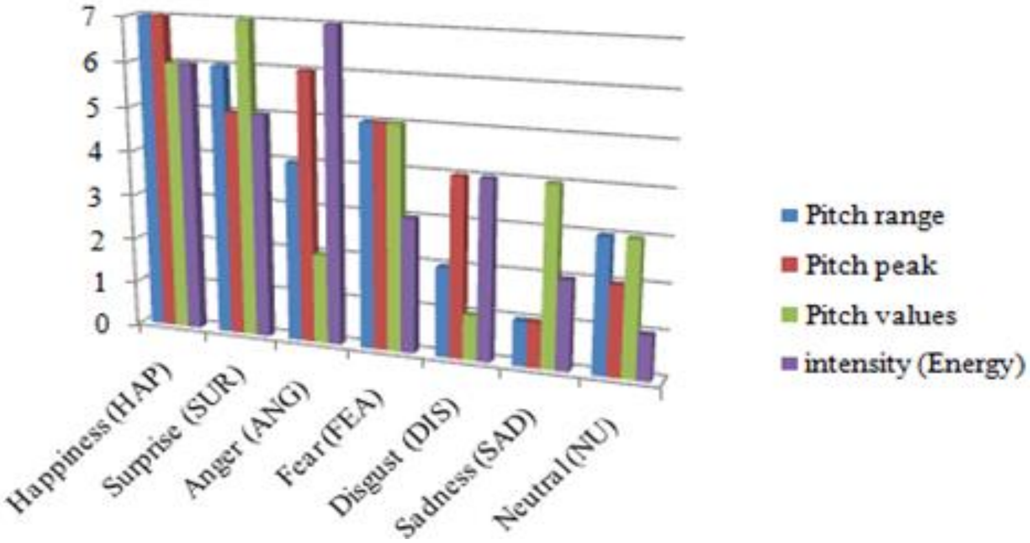


Fig. 4-2 Results of the seven basic emotions in the pitch and intensity plots.

Base on the results of chart and the environment of conversation we can filter the sound alignment in order to increases the accuracy of the system and decries the time that need for recognition of emotion.

Figures 4-3 and 4-4 show the results arranged in three-dimensional graphs using a discrete approach for the classification of emotions. For instance, Figure 4-3 shows the location of the six basic emotions in the three-dimensional graph whose axes are: the pitch peak, the pitch range and the pitch value.

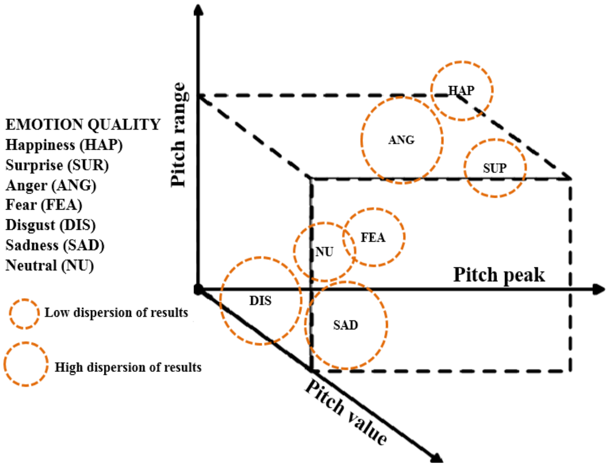


Fig. 4-3 The location of the emotions in the three-dimensional graph with axes: pitch (range, peak and value)

Figure 4-4 shows the location of the six basic emotions in the three-dimensional graph whose axes are: the total pitch score, the intensity and the speech rate. Other three-dimensional graphs can be built by selecting a different set of three features.

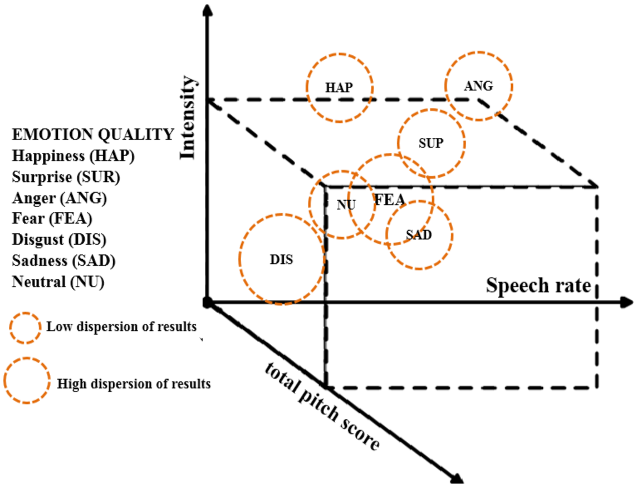


Fig. 4-4 The location of the basic emotions in the three-dimensional graph with axes: total pitch score, intensity and speech rate

From the analysis of both Figures (4-3 and 4-4), it results that when pitch range, pitch value and pitch peak are considered, happiness, surprise and anger can be distinguished faster than other emotions, and with a high degree of accuracy. In order to distinguish other emotions, we can use other features such as: intensity and speech rate, as well as the total pitch score (Figure 4-4).

As it can be shown in Table 4-7, we intend to compare the accuracy of recognition emotions, based on the sound analysis. For example happiness and anger can be distinguished faster than with a high degree of accuracy between all emotions.

Table. 4-7 The accuracy recognition of the basic emotions.

EMOTION	HAP	SUR	ANG	FEA	DIS	SAD	NEU
Happiness	**92.23**	6.44	0.00	0.00	0.00	0.00	1.33
Surprise	9.38	**83.55**	0.00	0.00	4.05	0.00	2.02
Anger	0.00	2.12	**92.76**	0.00	4.40	5.22	0.50
Fear	0.00	0.00	0.00	**73.27**	0.00	13.12	8.61
Disgust	0.00	0.00	0.00	10.11	**67.02**	11.63	11.24
Sadness	0.00	0.00	0.00	14.38	10.12	**63.48**	10.02

The software checks the result of PRAAT and rules extracted such as; graph of total pitch score, intensity and speech rate for classification of emotions. With this features, we can distinguish all the six basic emotions more easily, because the boundary between emotions are very distinct.

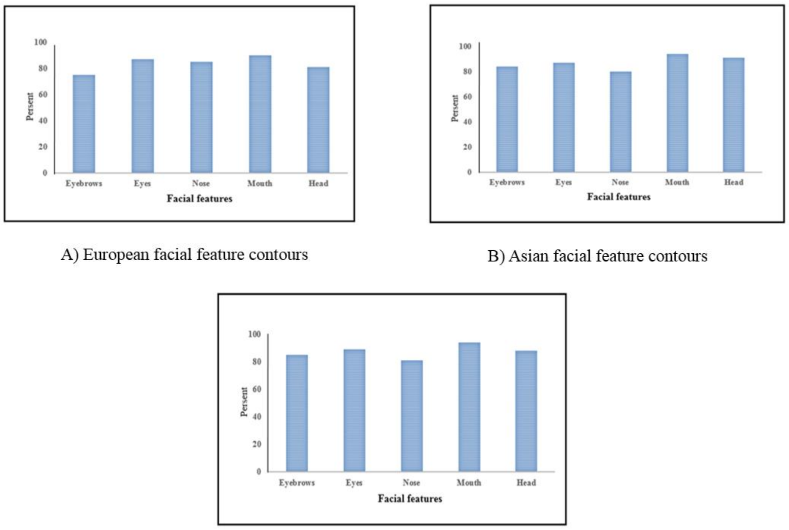
4-2-2 Results on emotion recognition, facial features expression

For implementation and validation of the proposed system, have been used 3 different images for each basic category of emotion. We performed experimental tests to assess the effectiveness of the algorithm.

All these sets of experiments were performed in the laboratory (offline validation) and used standard 2D image dataset. After completing each testing stage, we can also manually label the facial emotion expression for training the database. We performed three experiments, from these experiments we inferred universal observation to test different scenarios.

The first experiments have been started for facial detection from different groups testing. For practical purposes, the important outcome of this section is the localization and detection of meaningful facial features. We selected 30 individuals (15 female and 15 male, 20 to 48 years old), belonging to different ethnic groups, namely: (i) European, (ii) Asian (Middle East) and (iii) American.

The results of localization of facial features for 15 participants show that Americans have the highest and Europeans has the lowest accuracy in localization of facial features. Also, the results of this exploratory study in Figure 4-5 show that among all facial features, eyebrows detection and head position have the highest accuracy detection rate for Asians and Americans, while Europeans have the highest accuracy detection rate for eyes and mouth



A) European facial feature contours

B) Asian facial feature contours

C) American facial feature contours

Fig. 4-5 Accuracy in detection of facial features for Europeans, Asians and Americans.

If we compare nose and head detection for European, Asian and American participants, more or less system have similar results (81%-89% accuracy). In our experiment, we can also see that in the categorization and classification of the facial features, mouth and eyes detection are two important parameters for our proposed system. Also, the accuracy of the emotion detection system directly depends on the detection of the areas of the mouth and the eyes.

As it can be seen in Table 4-8, we intend to compare the accuracy of detection for several facial features, namely: head localization, eyebrows, eyes, nose and mouth, based on the analysis of the properties of the three ethnic groups. If we compare facial detection for all groups of participants (Table 4-8), detection of eyebrows has the lowest accuracy for Europeans. For Asians, the lowest accuracy lies in nose detection, while mouth detection has the highest accuracy in all groups. Also, universal similarities are observed among head and eyes detection, so detection of mouth and eyes is the most important step for any ethnic group.

Table. 4-8 Facial detection accuracy for (European, Asian and American).

Ethnic	Classification	Head localization	Eyebrows	Eyes	Nose	Mouth
European	**Head**	**100.00**	0.00	0.00	0.00	0.00
	Eyebrows	0.00	**75.33**	15.17	10.50	0.00
	Eyes	0.00	9.16	**87.04**	3.80	0.00
	Nose	0.00	4.36	5.22	**85.24**	5.18
	Mouth	0.00	0.00	0.00	9.89	**90.11**
Asian (Middle East)	**Head**	**100.00**	0.00	0.00	0.00	0.00
	Eyebrows	0.00	**84.17**	7.25	8.58	0.00
	Eyes	0.00	8.89	**86.63**	4.38	0.00
	Nose	0.00	3.24	7.59	**80.36**	8.45
	Mouth	0.00	0.00	0.00	5.72	**94.28**
American	**Head**	**100.00**	0.00	0.00	0.00	0.00
	Eyebrows	0.00	**87.22**	0.95	10.83	0.00
	Eyes	0.00	10.99	**88.59**	3.52	0.00
	Nose	0.00	1.07	5.46	**87.54**	5.93
	Mouth	0.00	0.00	0.00	6.68	**93.32**

The results of this implementation show that for emotion recognition in facial features, mouth and eyes are an important factor as well. As it can be seen in Table 4-9, in order to check and validate the results, have been used 1000 facial images in different databases (Cohn-Kanade, The Iranian Face Database (IFDB) and internet photo (facebook)). It is straightforward to infer that the head detection has the highest result in the proposed system with 100% accuracy.

Table. 4-9 Facial expression detection accuracy in the proposed system.

True/Classification	Head localization	Eyebrows	Eyes	Nose	Mouth
Head	**100.00**	0.00	0.00	0.00	0.00
Eyebrows	0.00	**91.3**	3.8	4.9	0.00
Eyes	0.00	5.5	**92.7**	1.8	0.00
Nose	0.00	0.8	4.2	**90.5**	4.5
Mouth	0.00	0.00	0.00	4.6	**95.4**

The mouth and eyes are meaningful facial features components. However, this is also true for another features (Mouth (95.4), Eyes (92.7), Eyebrows (91.3) and Nose (90.5)). This was validated by preliminary experiments that produced 92.5% accuracy in facial features detection in the proposed system.

The second and third experiments have been implemented for the proposed hybrid algorithm on two databases: namely, the training database and the Cohn-Kanade database.

The system checks the localization and detection of several facial features. As it can be shown in Table 4-10, have been intend to compare the accuracy of recognition of the basic emotions, based on the analysis of ethnic properties.

The results show that anger and surprise for Asians have the highest accuracy recognition rate among different ethnic groups. Also, it can be seen that for Europeans and Americans, disgust has the highest accuracy rate.

As depicted in Table 4-10, the lowest facial feature expressions accuracy for European participants is happiness with 91.2%, for Asians is disgust with 90.1% and Americans is sadness with 89.5%. Also, for Asian (Middle East) participants, the facial expression accuracy is higher than for Europeans and Americans.

In Table 4-10, it can be seen that, the second phase of algorithm, which used anthropometrics of face, Bézier curves and Action Units, reached in average an accuracy of 94.4% without training phase.

Table. 4-10 Facial expression recognition accuracy for Europeans, Asians and Americans.

Ethnic	EMOTION	Happiness	Surprise	Anger	Fear	Disgust	Sadness	Neutral
European	**Happiness**	**91.2**	8.8	0.00	0.00	0.00	0.00	0.00
	Surprise	3.9	**96.1**	0.00	0.00	0.00	0.00	0.00
	Anger	0.00	0.00	**94.6**	0.00	5.4	0.00	0.00
	Fear	0.00	0.00	0.00	**93.2**	0.00	3.1	3.7
	Disgust	0.00	0.00	0.00	0.00	**97.1**	1.6	1.3
	Sadness	0.00	0.00	0.00	4.3	1.1	**93.4**	1.2
Asian (Middle East)	**Happiness**	**96.1**	3.9	0.00	0.00	0.00	0.00	0.00
	Surprise	1.1	**98.9**	0.00	0.00	0.00	0.00	0.00
	Anger	0.00	0.00	**97.3**	0.00	2.7	0.00	0.00
	Fear	0.00	0.00	0.00	**92.7**	0.00	4.3	3.0
	Disgust	0.00	0.00	0.00	0.00	**90.1**	5.7	4.2
	Sadness	0.00	0.00	0.00	2.1	2.1	**95.4**	0.4
American	**Happiness**	**95.6**	4.4	0.00	0.00	0.00	0.00	0.00
	Surprise	3.2	**96.8**	0.00	0.00	0.00	0.00	0.00
	Anger	0.00	0.00	**96.7**	0.00	3.3	0.00	0.00
	Fear	0.00	0.00	0.00	**93.1**	0.00	0.4	6.5
	Disgust	0.00	0.00	0.00	0.00	**98.6**	0.9	0.5
	Sadness	0.00	0.00	0.00	1.6	1.6	**89.5**	7.3

In order to evaluate our system and to perform the training phase, have been used the Cohn-Kanade database. The results of these experiments in hybrid algorithm are shown in Table 4-13. As mentioned above, the accuracy rate of the results of this study in facial detection was 92.2% (Table 4-8).

The speed of emotion recognition is an important factor in human robot communication. A notable result (see Table 4-11) is that disgust and anger have the highest recognition speed for Europeans, Asians and Americans. The meaning of this result is that when the proposed algorithm does not check the eyebrows and eyes Action Units and uses just facial Action Shape Models, the speed of emotion recognition increases.

However, Table 4-11 show that the emotion recognition result for disgust is two times faster than fear. Moreover, fear and happiness have the lowest emotion detection rate, but recognition accuracy of these emotions is higher than 91%. Sadness and fear have the lowest emotion recognition rate in Asian individuals. This result is probably due to the fact that Asian people have lower emotional reaction to sadness and fear. On the other side, anger and disgust have the highest emotion detection rate for Americans and Europeans.

Table. 4-11 Emotion recognition rate (average of the 15 experimental tests) for Europeans, Asians and Americans.

Emotion Quality	Emotion recognition average rate for (European, Asian and American)
Happiness	1.13 s
Surprise	0.87 s
Anger	0.75 s
Fear	1.21 s
Disgust	0.62 s
Sadness	0.97 s

Emotion recognition in human interaction normally needs two or three seconds. The proposed system can distinguish all the six basic emotions faster and more easily, because the boundaries between emotions are more distinct and the features are classified in a set of categories.

4-3 training of the emotion recognition system

The training phase is one of the main challenges of any emotion recognition program. The accuracy of the system for emotion recognition depends on precise feature labeling in the training phase.

The training phase enables the automatic system to detect the emotion with new data. Also, it can be used as a learning algorithm to manage and predict facial image database. Training is based on classified data in the same emotional subject, extracted from new facial image.

The project is evaluated from two perspectives. Firstly, calculate the accuracy in quality of detection and emotion recognition and secondly, high run-time performance evaluates the training and classification system.

Hence, the training phase produces a facial emotion classification considerably speeding up the recognition process.

The training phase consists of defining emotional features and save the new emotion landmarks data in training database. However, sometimes the system has to manually declare and save information in the training database.

The training phase runs similarly to the data. Namely, the proposed algorithm for classification of basic emotions first normalizes the facial image. Then the algorithm based on Table 4-2 determines position of Bézier Curve points and extracted Facial Action Codes (ACs). Finally, our algorithm checks the similarity between new data and references data for basic emotional states (happiness, anger, fear, sadness, surprise, disgust and neutral).

It is noted that when some of the features are not save in the training database maybe system detected more than one emotion concluded. In order to have effective and correct training process on emotion recognition in proposed system, training process managed 3 facial image data for each basic emotion subject.

Table4-12 shows the accuracy of facial features algorithm with training phase reached 95.6% for all six basic emotions. Finally, if we added neutral features as an emotion into Action Units and corresponding data in training phase, the accuracy in emotion recognition increases, while the speed of the algorithm decreases.

Table. 4-12 Facial expression recognition accuracy using the Cohn-Kanade database.

EMOTION	Happiness	Surprise	Anger	Fear	Disgust	Sadness	Neutral
Happiness	**95.9**	4.1	0.00	0.00	0.00	0.00	0.00
Surprise	1.6	**98.4**	0.00	0.00	0.00	0.00	0.00
Anger	0.00	0.00	**97.2**	0.6	2.2	0.00	0.00
Fear	0.00	0.00	0.00	**94.7**	0.00	1.2	4.1
Disgust	0.00	0.00	0.00	0.00	**95.7**	1.5	2.8
Sadness	0.00	0.00	0.00	2.6	1.9	**91.8**	3.7

4-4 Evaluation of the results on emotion recognition in hybrid system

In the proposed software, system combined the speech-based algorithm and facial expression analysis. Then, with the features in database checks the results of PRAAT (graph of total pitch score, intensity and speech rate) and Action codes (ACs). Finally, as you can see in Table 4-13 system can distinguish all the six basic emotions more easily, because the boundary between emotions are very distinct.

From the validation results (see Table 4-13) it appears that fear and disgust are not so easy to distinguish from hybrid system; however, they are also true for humans.

Table. 4-13 emotion recognition accuracy for Europeans, Asians and Americans with hybrid system.

Ethnic	EMOTION	Happiness	Surprise	Anger	Fear	Disgust	Sadness	Neutral
European	**Happiness**	**92.50**	7.40	0.00	0.00	0.00	0.00	0.10
	Surprise	3.35	**96.50**	0.00	0.00	0.00	0.00	0.15
	Anger	0.00	0.00	**95.60**	0.20	4.20	0.00	0.00
	Fear	0.00	0.00	0.00	**93.00**	0.00	3.25	3.75
	Disgust	0.00	0.00	0.00	0.00	**91.40**	4.00	4.60
	Sadness	0.00	0.00	0.00	5.40	1.15	**92.30**	1.15
Asian (Middle East)	**Happiness**	**97.20**	2.80	0.00	0.00	0.00	0.00	0.00
	Surprise	0.90	**99.10**	0.00	0.00	0.00	0.00	0.00
	Anger	0.00	0.00	**98.50**	0.00	1.00	0.00	0.50
	Fear	0.00	0.00	0.00	**91.60**	0.00	4.20	4.20
	Disgust	0.00	0.00	0.00	0.00	**90.10**	5.70	4.20
	Sadness	0.00	0.00	0.00	2.50	3.50	**93.70**	0.30
American	**Happiness**	**96.70**	2.30	0.00	0.00	0.00	0.00	1.00
	Surprise	2.50	**97.50**	0.00	0.00	0.00	0.00	0.00
	Anger	0.00	0.00	**96.70**	0.00	2.20	0.00	1.10
	Fear	0.00	0.00	0.00	**92.10**	0.00	1.45	6.45
	Disgust	0.00	0.00	1.40	0.00	**95.60**	1.90	1.10
	Sadness	0.00	0.00	1.00	2.85	1.30	**89.70**	5.15

The proposed algorithm is very suitable for implementation in real-time systems in outdoor environment, since the computational load is very low indeed. However, the results of this implementation show that emotions such as: Fear, disgust and sadness could be more easily detected if the hybrid algorithm proposed in this work is combined with an emotion recognition algorithm based on all of (verbal and non-verbal) features.

CHAPTER 5

Conclusion

Human emotions are reflected in body gestures, hand movement, voice and facial expressions. Even though there have been many advances in sound analysis and facial expression for emotion recognition in last few years, researchers believe that emotion is the result of a brain process. In human robot interaction individual skills and human behavior must be learned and updated continually, because there is still a long way to achieve high accuracy rates in automatic recognition of emotions.

Emotion recognition technology is still not robust enough for very demanding applications like humanoid robots. The proposed emotion recognition system is as an important tool in behavioral communication that facilitating human robot interaction (HRI).

In this work the proposed algorithm combined sound recognition results with facial expressions program. The proposed methodology in facial expressions work base on Facial Action Code System, Facial geometric and Action Units (AUs) for feature detection. Finally, we used Support Victor Machine and nearest likelihood for classified basic emotions.

In the speech emotion recognition phase have been proposed a methodology for recognition of emotions, based on different speech features, which can be employed for human-robot interaction. The features that are taken into account are prosodic features, such as: pitch, intensity, speech rate and formant.

The proposed technique is based on a first analysis of pitch graph contours (namely: pitch peak, pitch value, pitch range), followed by a second analysis of the intensity and the speech rate in the dialogue, which is considered complementary to the first analysis in order to recognize all types of emotions.

PRAAT software is an open-source and very flexible tool for voice sampling in the field of pitch (peak, range and value), formant, spectrograms and intensity analysis.

The proposed algorithm combined with open source PRAAT program used for feature detection and classified basic emotions.

When we compered all of the pitch results with PRAAT software it can be seen that the highest pitch value is for surprise and the lowest corresponds to disgust. Also, the pitch peak under positive valence emotions is sharper among Asian speakers, while European and American speakers more or less have similar pitch contours under positive valence emotions.

Happiness and anger have the highest average pitch peak for European speakers, while sadness has the lowest pitch peak. In our experiment, we can also see that surprise and anger for Asian and American speakers have the highest average pitch peak.

Among the negative valence emotions, anger has the highest pitch peak. Sadness decreases sharply for Asian and American speakers, but sadness slop decreases slowly for European speakers.

If we compare sadness and neutral for all groups of speaker the neutral emotion does not have a distinct peak and is similar to sadness; however, sadness has lower ending pitch signals. Asian speakers were more sensitive to sadness emotion, while the pitch graphs of Americans and Europeans were similar.

Anger is associated with the highest energy for Asian and American speakers but for Asian speakers the anger slope decreases slowly, while sadness is associated with the lowest energy for Asian and European speakers. In almost all utterances under anger and fear, the pitch increases to a peak and then decreases slightly left-skewed. European and American speakers more or less have similar pitch contours under fear emotion.

When we compered all of the formant results we can see that the contour in anger and happiness for European speakers has the highest power, while we have the lowest spectral power in fear. Formant contour explain that anger, fear and happiness have the highest power for Asians and Americans, while we have a lot of wave and formant dots the fear plot. Asian and European speakers have the lowest spectral power in sadness, while Americans have the lowest spectral power in neutral emotion.

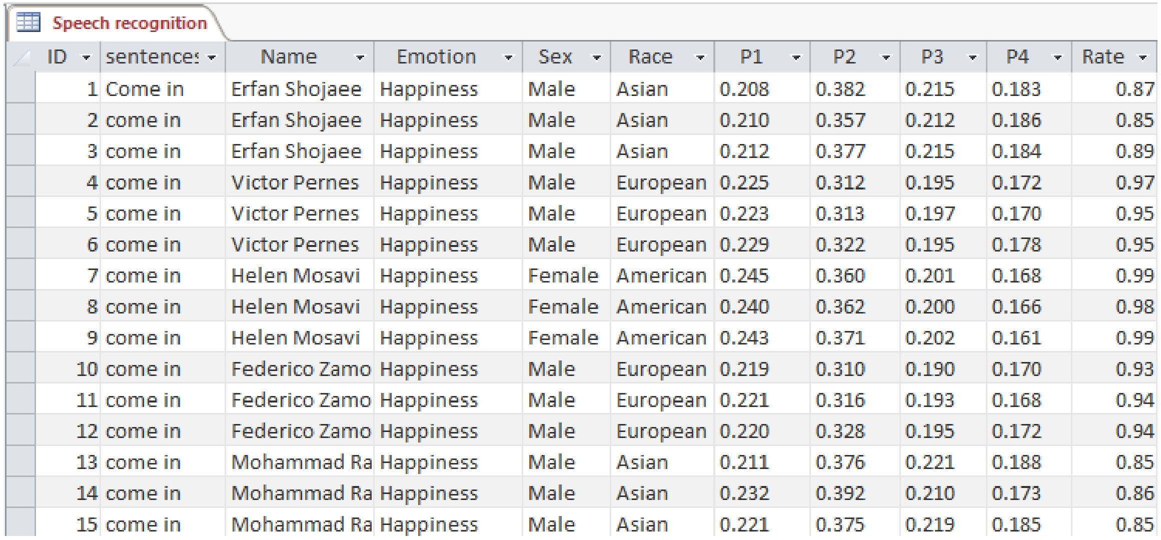
In terms of intensity, anger and surprise for European speakers have the highest energy and intensity, while neutral and sadness have the lowest intensity. For Asian and American speakers, anger and happiness have the highest energy and intensity, while fear has the lowest intensity.

A notable result in speech rate analysis is that anger and fear have the lowest speech rate for European speakers, meaning that the sentences pronounced with anger or fear are pronounced faster, while happiness has the highest speech rate. Sadness and disgust have the lowest speech rate for Asian speakers, while anger and happiness have the highest speech rate: this result is probably due to the fact that Asian people have bigger emotional reaction to happiness and anger. For American speakers anger and disgust have the lowest speech rate, while happiness and fear have the highest speech rate.

In general, happiness and surprise have the highest speech rate, while anger and sadness have the lowest speech rate. Moreover, Americans have the highest speech rate.

Pitch values and speech rate are connected together for all ethnic groups of speaker. We can see that usually the speech rate of American speakers is higher than Asian and European speakers.

One of the challenges on speech emotion analysis is the difficulty in creating a worldwide database on emotional speech recognition. Thus, we designed a database a new database for recognition on emotions, based on the sound analysis. Figure 5-1 summarizes speech recognition database for the participants to experiment (three repetitions made by each participant) the accuracy of the system.



ID	sentence:	Name	Emotion	Sex	Race	P1	P2	P3	P4	Rate
1	Come in	Erfan Shojaee	Happiness	Male	Asian	0.208	0.382	0.215	0.183	0.87
2	come in	Erfan Shojaee	Happiness	Male	Asian	0.210	0.357	0.212	0.186	0.85
3	come in	Erfan Shojaee	Happiness	Male	Asian	0.212	0.377	0.215	0.184	0.89
4	come in	Victor Pernes	Happiness	Male	European	0.225	0.312	0.195	0.172	0.97
5	come in	Victor Pernes	Happiness	Male	European	0.223	0.313	0.197	0.170	0.95
6	come in	Victor Pernes	Happiness	Male	European	0.229	0.322	0.195	0.178	0.95
7	come in	Helen Mosavi	Happiness	Female	American	0.245	0.360	0.201	0.168	0.99
8	come in	Helen Mosavi	Happiness	Female	American	0.240	0.362	0.200	0.166	0.98
9	come in	Helen Mosavi	Happiness	Female	American	0.243	0.371	0.202	0.161	0.99
10	come in	Federico Zamo	Happiness	Male	European	0.219	0.310	0.190	0.170	0.93
11	come in	Federico Zamo	Happiness	Male	European	0.221	0.316	0.193	0.168	0.94
12	come in	Federico Zamo	Happiness	Male	European	0.220	0.328	0.195	0.172	0.94
13	come in	Mohammad Ra	Happiness	Male	Asian	0.211	0.376	0.221	0.188	0.85
14	come in	Mohammad Ra	Happiness	Male	Asian	0.232	0.392	0.210	0.173	0.86
15	come in	Mohammad Ra	Happiness	Male	Asian	0.221	0.375	0.219	0.185	0.85

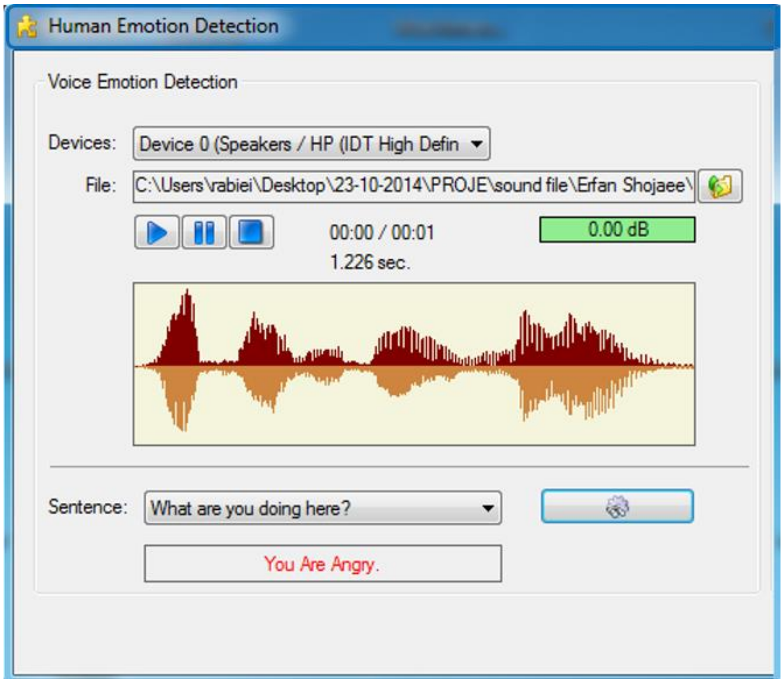
Fig. 5-1 speech recognition database for the participants to experiment.

In order to implement our algorithms for feature extraction and emotion recognition, have been defined rules based on human sound signals. With this rules and features, we can distinguish all the six basic emotions and neutral more easily, because the boundaries between emotions are very distinct.

In the presented model, emotions are first categorized in two main classes; namely high and low intensity of emotions, then a more precise distinction is performed within each category.

An experimental test, with the participation of ten European, ten Asian and ten American individuals, was set up, in order to experimentally validate the proposed methodology in the laboratory.

The results of this exploratory study show that it could be feasible to build a technique which is effective in recognizing emotions. Figure 5-2 shows the graphic interface of the software (Voice Emotion Detection) system.



Devices: Device 0 (Speakers / HP (IDT High Defin

File: C:\Users\rabiei\Desktop\23-10-2014\PROJE\sound file\Erfan Shojaee\

00:00 / 00:01

1.226 sec.

0.00 dB

Sentence: What are you doing here?

You Are Angry.

Fig. 5-2 Graphical interface of speech emotion recognition program.

In order to check and validate the results it appears that happiness and anger can be distinguished faster than and with a high degree of accuracy between all emotions in different ethnic groups.

In this experiment, the accuracy of emotion detection was 78.5%. Also, the recognition rate of the proposed system was 0.97 s.

The novelty in the speech emotion recognition phase is that system work based on phonetic and acoustic properties of emotive speech with the minimal use of signal processing algorithms.

Thus, the results of this study provide a better understanding on the manifestation and production of basic emotions and can be useful for the purpose of analysis and synthesis of emotional speech for technical researchers, social psychologists and human-robot interaction.

In the second phase, we have presented an approach for recognition of emotions, based on facial expression analysis, for possible implementation results in Human-Robot Interaction systems.

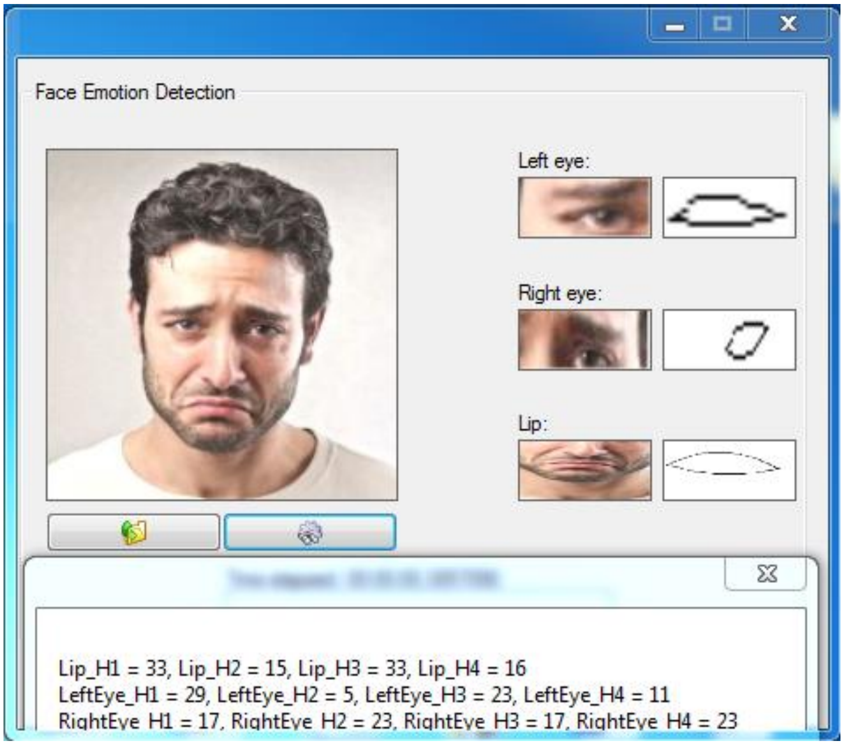
Even though there have been many advances in facial emotion detection in last few years, there is still a long way to achieve high accuracy rates in automatic recognition on emotions. Facial recognition technology is still not robust enough for very demanding applications like humanoid robots. Facial emotion recognition systems need to analyze the facial expression regardless of ethnics, culture and gender.

The proposed technique is based on a first analysis of face localization and facial features detection, followed by a second stage is features extraction and emotion recognition, which is considered complementary to recognize all types of emotions.

The proposed system is based on a facial feature extraction algorithm, which determines face localization, features detection, facial muscle movements (Action Units), Bézier curves and Facial Action Codes calculation. The system is implemented by geometric analysis of the facial image, based on measurements and classification a set of AUs. We have presented a new method to locate 32 Action Units, 52 Action points and 15 facial code distances.

An experimental test, we selected 30 individuals (15 female and 15 male) participant, belonging to different ethnic groups, namely: (i) European, (ii) Asian (Middle East) and (iii) American., was set up in order to experimentally validate the proposed methodology.

As it can be shown in Figure 5-3, we extracted the proposed principal points based on Bézier curves and Action Units for universal basic emotions from different ethnic group images.



```
Lip_H1 = 33, Lip_H2 = 15, Lip_H3 = 33, Lip_H4 = 16
LeftEye_H1 = 29, LeftEye_H2 = 5, LeftEye_H3 = 23, LeftEye_H4 = 11
RightEye H1 = 17, RightEye H2 = 23, RightEye H3 = 17, RightEye H4 = 23
```

Fig. 5-3 extract the facial distances based on universal basic emotions.

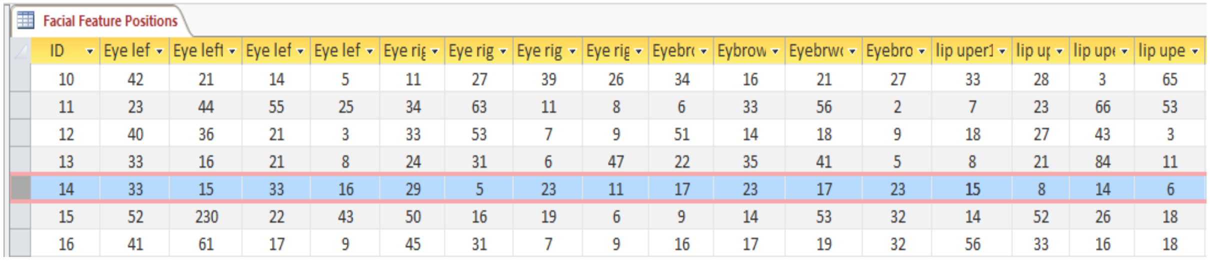
The result of localization of facial features for participants shows that Americans have the highest and Europeans has the lowest accuracy in localization of facial features. Also among all facial features, eyebrows detection and head position have the highest accuracy detection rate for Asians and Americans, while Europeans have the highest accuracy detection rate for eyes and mouth detection.

In our experiment, we can also see that in the categorization and classification of the facial features, mouth and eyes detection are two important parameters for our proposed system. Also, the accuracy of the emotion detection system directly depends on the detection of the areas of mouth and eyes.

The results of this implementation show that for emotion recognition in facial features, mouth and eyes are an important factor as well. However, this is also true for another feature (Mouth (95.4), Eyes (92.7), Nose (90.3) and Eyebrows (91.2)). This was validated by preliminary experiments that produced 92.5% accuracy in facial detection in the proposed system.

For recognizing facial expression in proposed software, have been used two database for maintenance facial features, namely: personnel list and facial features position table.

Facial features Position table holds information of Action Units suitably represented and divided into: eyes right/left, eyebrows right/left, nose, lips and head of 30 participants for all of emotional states. Figure 5-4 contains the reference number of 32 Action Codes for seven basic emotions.



ID	Eye lef	Eye left	Eye lef	Eye lef	Eye rig	Eye rig	Eye rig	Eye rig	Eyebr	Eybrow	Eyebrw	Eyebro	lip uper1	lip up	lip upe	lip upe
10	42	21	14	5	11	27	39	26	34	16	21	27	33	28	3	65
11	23	44	55	25	34	63	11	8	6	33	56	2	7	23	66	53
12	40	36	21	3	33	53	7	9	51	14	18	9	18	27	43	3
13	33	16	21	8	24	31	6	47	22	35	41	5	8	21	84	11
14	33	15	33	16	29	5	23	11	17	23	17	23	15	8	14	6
15	52	230	22	43	50	16	19	6	9	14	53	32	14	52	26	18
16	41	61	17	9	45	31	7	9	16	17	19	32	56	33	16	18

Fig. 5-4 Facial feature database (Access) for 32 Action Units.

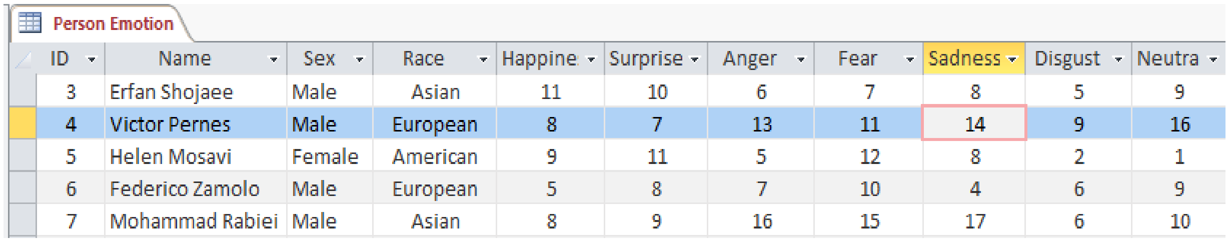
We defined a set of universal observations for emotion recognition based on evaluation of the procedure of facial emotion extraction.

Emotions like anger, surprise and sadness can be directly recognized, based on the shape of the eyes and eyebrows position. Estimation of the lips and mouth is especially important for recognizing happiness and surprise emotion. In surprise, the mouth length decreases and the mouth height increases. In some situations, the lips thickness in associated to anger (the lips result thinner than normal), and sadness is associated with the obliquely lowering of the lip corners.

Fear and sadness have similarity in sets of Action Codes. But in most cases the inner eyebrows corner rises up in sadness. Also, the outer eyebrows corner in fear is higher than in sadness. The location of the eyebrows slope in sadness is usually greater than in fear.

Fear and disgust have the same value for group of features. While disgust occurs, nose side wrinkles appear. Wrinkles up the nose usually express displeasure or disgust and appear along the lateral nose boundaries.

We defined person emotion table based on Facial Features Position table for emotion recognition. As specified in Figure 5-5 we recorded the person emotion information for seven basic emotions. The proposed software have been used K-nearest neighbor's technique for comparing the data with the training database. This technique was tested in the evaluation of the proposed system. For example, number 14 in Figure 5-5 is connected to sadness emotion information for Victor Pernes. As mentioned above row 14 Figure 5-4 was explained the position of 32 facial action codes in sadness emotion.



ID	Name	Sex	Race	Happine	Surprise	Anger	Fear	Sadness	Disgust	Neutra
3	Erfan Shojaee	Male	Asian	11	10	6	7	8	5	9
4	Victor Pernes	Male	European	8	7	13	11	14	9	16
5	Helen Mosavi	Female	American	9	11	5	12	8	2	1
6	Federico Zamolo	Male	European	5	8	7	10	4	6	9
7	Mohammad Rabiei	Male	Asian	8	9	16	15	17	6	10

Fig. 5-5 Interaction between Facial feature database and person emotion.

In order to evaluate our algorithms for feature extraction and emotion recognition, we have performed experiments with the Cohn-Kanade facial image database. In this experiment, the accuracy of facial detection rate was 92.2%. Also, the accuracy of the proposed system with training stage was 94.4%.

In the proposed model, if we added the neutral emotion in our Action Codes and data in training database, the accuracy of the system meaningfully increases, while the time for loading algorithm increases.

The research shows that using the hybrid algorithm can yield better results for emotion recognition. Thus, we intend to combine the proposed algorithm with a technique based on speech analysis, in order to design a hybrid technique for emotion recognition, to be employed in HRI system.

Finally, we make contribution toward design a hybrid technique for emotion recognition. In order to design system for emotion recognition. We combined algorithm based on speech analysis with a technique for facial features extraction. As it can be shown in Figure 5-6, we intend to fusion the database for emotion recognition system in hybrid system.

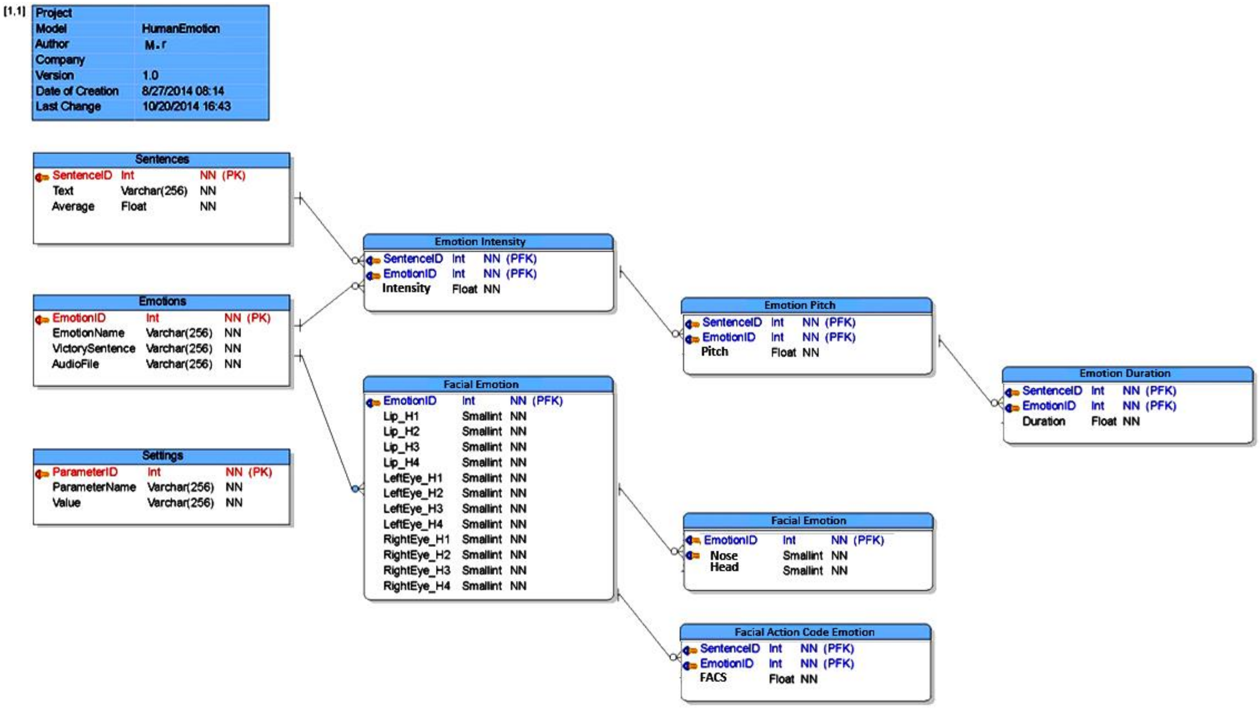
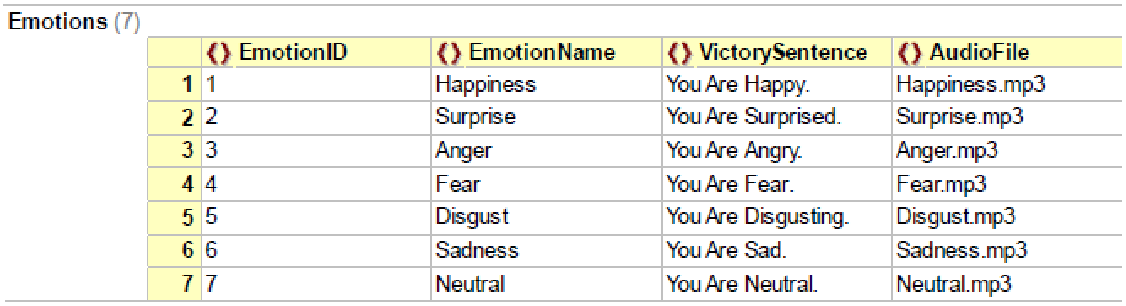


Fig. 5-6 Fusion of database for emotion recognition system in hybrid system.

As depicted in Figure 5-6, the proposed system is done in 2 stages: the first stage is analysis of speech signals namely; pitch peak, pitch value, pitch range, intensity and the speech rate (duration). The second stage is based on facial expression analysis such as; (Action Units, Facial Action Codes and Bézier curves) in order to recognize all types of emotions.

The hybrid system for selected human emotion, comparing the data base on the minimum difference and K-nearest neighbors technique. Available parameters are calculated and compared with the value specified in the database. If the algorithm results was equal with database, program show the emotion like text file and connected to database for read the emotional states (Figure 5-7). Otherwise system cannot be detected the emotion and wrote "Unknown Emotion".



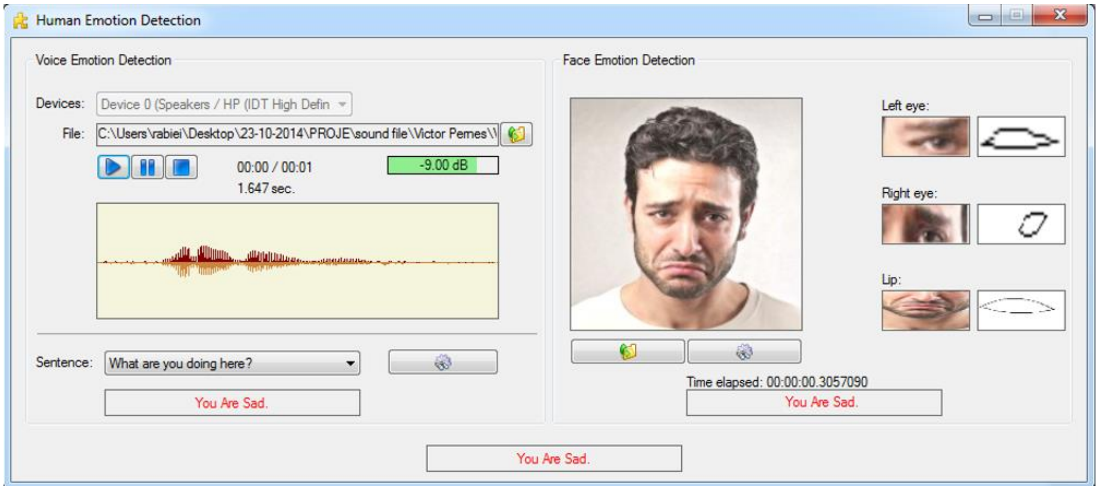
Emotions (7)

	{} EmotionID	{} EmotionName	{} VictorySentence	{} AudioFile
1	1	Happiness	You Are Happy.	Happiness.mp3
2	2	Surprise	You Are Surprised.	Surprise.mp3
3	3	Anger	You Are Angry.	Anger.mp3
4	4	Fear	You Are Fear.	Fear.mp3
5	5	Disgust	You Are Disgusting.	Disgust.mp3
6	6	Sadness	You Are Sad.	Sadness.mp3
7	7	Neutral	You Are Neutral.	Neutral.mp3

Fig. 5-7 Sample of database for play audio file in emotion recognition system.

However, the challenge of detecting human faces from an image mostly comes from the variation of human faces such as races, face scales and environment issues such as lighting conditions, image quality, and cluttered backgrounds may cause great difficulties. The proposed algorithm can perform well under a large variation of facial image quality in the normal light, for different ethnic groups of people.

In Figure 5-8 we show that the software package with hybrid algorithm for emotion recognition system. This applet provides the following options to users: divice for recording a sound, file for browsing the sound file and facial image from hard disk, (play- stop) button and two windows for show facial image and speech signals emotions.



Devices: Device 0 (Speakers / HP (IDT High Defin

File: C:\Users\rabiei\Desktop\23-10-2014\PROJE\sound file\Victor Pemes\\

00:00 / 00:01

1.647 sec.

-9.00 dB

Sentence: What are you doing here?

You Are Sad.

Time elapsed: 00:00:00.3057090

You Are Sad.

You Are Sad.

Fig. 5-8 Software package with hybrid algorithm on emotion recognition system.

With respect to other works in the scientific literature, the methodology we propose in this paper uses on speech and facial expression by means of emotion recognition. In this experiment, the accuracy of hybrid algorithm for emotion detection rate was 2.53 s. Thus, the proposed algorithm is very suitable for implementation in real-time systems, since the computational load is very low indeed. Also, the accuracy of the proposed system with 19 subjects (three times six basic emotion with neutral) and training stage was 94.30% for all of seven basic emotions. Moreover, evaluating and comparing the performance of different software and algorithm on human emotion recognition (present system) have been shown in Table 5-1.

Table 5-1 Evaluation of proposed system with another literature report.

Authors	Subject tested	Speech recognition tested	Images tested	Percentage accuracy
Edwards [206]	22		200	74%
Kobayashi and Hara [207], [208]	15		90	85%
Pantic and othkrantz [98]	8		246	94%
Huang and Huang [209]	15		90	75%
Hong et al [104]	25		175	81%
Zhang [106]	10		213	91%
Lyons et al [210]	10		193	92%
Picard et al [211]	8	850		81%
Rani et al [212]	5	480		86%
Leon et al [213]	7	380		80%
Kim and Andre [214]	4	560		96%
Van den Brock [215]	10	860		62%
Soleymani [216]	3	390		78%
Present study	19	630	840	94.3%

The present study demonstrated the development and the application of human robot interaction base on seven basic emotion types from speech recognition and facial expressions. Emotion recognition in human interaction normally needs two or three seconds. In the proposed system, we can distinguish all the basic emotions faster and more easily, because the boundaries between emotions are more distinct and the features are classified in a set of categories.

Even though there have been many advances in sound analysis and facial expression for emotion recognition in last few years, scientists believe that emotion is the result of a brain process.

In future work, we intend to combine the proposed algorithm with a technique based on gesture and thermal human body analysis in the noisy environment.

Recently, Thermal Infrared (TIR) imaging with standard GSR (temperature values of two infrared cameras) was used, to examine fear, happiness, sadness and joy emotion. Thus, for the future work, it will important to develop a consistent methodology to integrate both facial features image, gesture and Thermal Infrared (TIR) image sequences on expression of emotional state system.

Thermal IR imaging, allows to recording of the cutaneous temperature through the measurement of the spontaneous thermal irradiation of the body based on the average heat signature of pixels in a ROI.

We will use thermal Regions of Interest (t-ROIs) on the facial to be developed in invisible image and uncontrolled environment, in order to design a new technique for emotion recognition, to be employed in Human-Robot Interaction. The most important features for this technique, namely: nose or nose tip, the surrounding the eyes, upper lip, muscle and forehead. This will overcome problems across studies related to facial emotion recognition.

Bibliography

Bibliography

[1] D. J. France, R. G. Shiavi, S. Silverman, M. Silverman, and D. M. Wilkes, “Acoustical properties of speech as indicators of depression and suicidal risk,” Biomed. Eng. IEEE Trans., vol. 47, no. 7, pp. 829–837, 2000.

[2] J. Ma, H. Jin, L. T. Yang, and J. J.-P. Tsai, “Ubiquitous Intelligence and Computing: Third International Conference, UIC 2006, Wuhan, China, September 3-6, 2006,” Proc. (Lecture Notes Comput. Sci. Springer-Verlag New York, Inc., Secaucus, NJ, 2006.

[3] B. Schuller, A. Batliner, S. Steidl, and D. Seppi, “Recognising realistic emotions and affect in speech: State of the art and lessons learnt from the first challenge,” Speech Commun., vol. 53, no. 9–10, pp. 1062–1087, Nov. 2011.

[4] A. B. Ingale and D. S. Chaudhari, “Speech emotion recognition,” Int. J. Soft Comput. Eng. ISSN, pp. 2231–2307, 2012.

[5] S. G. Koolagudi, N. Kumar, and K. S. Rao, “Speech emotion recognition using segmental level prosodic analysis,” in Devices and Communications (ICDeCom), 2011 International Conference on, 2011, pp. 1–5.

[6] S. G. Koolagudi and K. S. Rao, “Emotion recognition from speech: a review,” Int. J. Speech Technol., vol. 15, no. 2, pp. 99–117, 2012.

[7] Y. Pan, P. Shen, and L. Shen, “Speech emotion recognition using support vector machine,” Int. J. Smart Home, vol. 6, no. 2, pp. 101–107, 2012.

[8] P. Ekman, “Are there basic emotions?,” Psychol. Rev., vol. 99, no. 3, pp. 550–553, 1992.

[9] N. A. Fox, “If it’s not left, it's right: Electroencephalograph asymmetry and the development of emotion.,” Am. Psychol., vol. 46, no. 8, p. 863, 1991.

[10] R. J. Davidson, G. E. Schwartz, C. Saron, J. Bennett, and D. J. Goleman, “Frontal versus parietal EEG asymmetry during positive and negative affect,” in Psychophysiology, 1979, vol. 16, no. 2, pp. 202–203.

[11] H. Schlosberg, “Three dimensions of emotion.,” Psychol. Rev., vol. 61, no. 2, p. 81, 1954.

[12] J. Ang, A. Krupski, E. Shriberg, A. Stolcke, S. Technology, S. R. I. International, and M. Park, “PROSODY-BASED AUTOMATIC DETECTION OF ANNOYANCE AND FRUSTRATION IN HUMAN-COMPUTER DIALOG International Computer Science Institute , Berkeley , CA 94704 , U . S . A . University of California , Berkeley , CA,” no. June, 2000.

[13] B. Ã. Yang and M. Lugger, “Emotion recognition from speech signals using new harmony features,” vol. 90, pp. 1415–1423, 2010.

[14] Y. Cheng, S.-Y. Lee, H.-Y. Chen, P.-Y. Wang, and J. Decety, “Voice and emotion processing in the human neonatal brain,” J. Cogn. Neurosci., vol. 24, no. 6, pp. 1411–1419, 2012.

[15] T. Iliou and C.-N. Anagnostopoulos, “Classification on speech emotion recognition-a comparative study,” Int. J. Adv. Life Sci., vol. 2, no. 1 and 2, pp. 18–28, 2010.

[16] S. Ntalampiras and N. Fakotakis, “Modeling the temporal evolution of acoustic parameters for speech emotion recognition,” Affect. Comput. IEEE Trans., vol. 3, no. 1, pp. 116–125, 2012.

[17] M. El Ayadi, M. S. Kamel, and F. Karray, “Survey on speech emotion recognition: Features, classification schemes, and databases,” Pattern Recognit., vol. 44, no. 3, pp. 572–587, Mar. 2011.

[18] B. Schuller, S. Steidl, A. Batliner, F. Burkhardt, L. Devillers, C. Müller, and S. Narayanan, “Paralinguistics in speech and language—State-of-the-art and the challenge,” Comput. Speech Lang., vol. 27, no. 1, pp. 4–39, Jan. 2013.

[19] M. Hamidi and M. Mansoorizade, “Emotion Recognition From Persian Speech With Neural Network,” Int. J. Artif. Intell. Appl., vol. 3, no. 5, 2012.

[20] S. G. Koolagudi and R. S. Krothapalli, “Two stage emotion recognition based on speaking rate,” Int. J. Speech Technol., vol. 14, no. 1, pp. 35–48, 2011.

[21] C.-H. Wu and W.-B. Liang, “Emotion recognition of affective speech based on multiple classifiers using acoustic-prosodic information and semantic labels,” Affect. Comput. IEEE Trans., vol. 2, no. 1, pp. 10–21, 2011.

[22] A. Gravano, R. Levitan, L. Willson, Š. Beòuš, J. B. Hirschberg, and A. Nenkova, “Acoustic and prosodic correlates of social behavior,” 2011.

[23] H. G. Wallbott and K. R. Scherer, “Cues and channels in emotion recognition.,” J. Pers. Soc. Psychol., vol. 51, no. 4, pp. 690–699, 1986.

[24] K. Scherer, “Vocal communication of emotion: A review of research paradigms,” Speech Commun., vol. 40, no. 1–2, pp. 227–256, Apr. 2003.

[25] B.-S. Kang, C.-H. Han, S.-T. Lee, D. H. Youn, and C. Lee, “Speaker dependent emotion recognition using speech signals.,” in INTERSPEECH, 2000, pp. 383–386.

[26] S. G. Koolagudi and S. R. Krothapalli, “Emotion recognition from speech using sub-syllabic and pitch synchronous spectral features,” Int. J. Speech Technol., vol. 15, no. 4, pp. 495–511, 2012.

[27] M. Lugger and B. Yang, “The relevance of voice quality features in speaker independent emotion recognition,” in Acoustics, Speech and Signal Processing, 2007. ICASSP 2007. IEEE International Conference on, 2007, vol. 4, pp. IV–17.

[28] J. Rong, G. Li, and Y.-P. P. Chen, “Acoustic feature selection for automatic emotion recognition from speech,” Inf. Process. Manag., vol. 45, no. 3, pp. 315–328, 2009.

[29] D. Ververidis and C. Kotropoulos, “Emotional speech recognition: Resources, features, and methods,” Speech Commun., vol. 48, no. 9, pp. 1162–1181, Sep. 2006.

[30] T. Bänziger, M. Morel, and K. R. Scherer, “Is there an emotion signature in intonational patterns? And can it be used in synthesis?,” in INTERSPEECH, 2003.

[31] T. Bänziger and K. R. Scherer, “The role of intonation in emotional expressions,” Speech Commun., vol. 46, no. 3–4, pp. 252–267, Jul. 2005.

[32] T. Bänziger and K. R. Scherer, “The role of intonation in emotional expressions,” Speech Commun., vol. 46, no. 3, pp. 252–267, 2005.

[33] G. zyn. Demenko and A. Wagner, “The stylization of intonation contours,” in Proceedings of Speech Prosody, 2006, pp. 141–144.

[34] S. Wu, T. H. Falk, and W. Chan, “Automatic speech emotion recognition using modulation spectral features,” Speech Commun., vol. 53, no. 5, pp. 768–785, 2011.

[35] B. Schölkopf, A. Smola, and K.-R. Müller, “Kernel principal component analysis,” in Artificial Neural Networks—ICANN’97, Springer, 1997, pp. 583–588.

[36] G. Potamianos, C. Neti, G. Gravier, A. Garg, and A. W. Senior, “Recent advances in the automatic recognition of audiovisual speech,” Proc. IEEE, vol. 91, no. 9, pp. 1306–1326, 2003.

[37] M. A. Casey, “Reduced-rank spectra and minimum-entropy priors as consistent and reliable cues for generalized sound recognition,” in Workshop for Consistent & Reliable Acoustic Cues, 2001, p. 167.

[38] D. FitzGerald, M. Cranitch, and E. Coyle, “Non-negative tensor factorisation for sound source separation,” 2005.

[39] T. Virtanen, A. T. Cemgil, and S. Godsill, “Bayesian extensions to non-negative matrix factorisation for audio signal modelling,” in Acoustics, Speech and Signal Processing, 2008. ICASSP 2008. IEEE International Conference on, 2008, pp. 1825–1828.

[40] E. Benetos, M. Kotti, and C. Kotropoulos, “Musical instrument classification using non-negative matrix factorization algorithms,” in Circuits and Systems, 2006. ISCAS 2006. Proceedings. 2006 IEEE International Symposium on, 2006, p. 4–pp.

[41] Y.-C. Cho, S. Choi, and S.-Y. Bang, “Non-negative component parts of sound for classification,” in Signal Processing and Information Technology, 2003. ISSPIT 2003. Proceedings of the 3rd IEEE International Symposium on, 2003, pp. 633–636.

[42] H. Hu, M.-X. Xu, and W. Wu, “GMM Supervector Based SVM with Spectral Features for Speech Emotion Recognition.,” in ICASSP (4), 2007, pp. 413–416.

[43] T. L. Nwe, S. W. Foo, and L. C. De Silva, “Speech emotion recognition using hidden Markov models,” Speech Commun., vol. 41, no. 4, pp. 603–623, 2003.

[44] D. Ververidis and C. Kotropoulos, “Emotional speech recognition: Resources, features, and methods,” Speech Commun., vol. 48, no. 9, pp. 1162–1181, 2006.

[45] B. Yang and M. Lugger, “Emotion recognition from speech signals using new harmony features,” Signal Processing, vol. 90, no. 5, pp. 1415–1423, May 2010.

[46] D. Kukolja, S. Popović, M. Horvat, B. Kovač, and K. Ćosić, “Comparative analysis of emotion estimation methods based on physiological measurements for real-time applications,” Int. J. Hum. Comput. Stud., vol. 72, no. 10–11, pp. 717–727, Oct. 2014.

[47] L. Chen, X. Mao, Y. Xue, and L. L. Cheng, “Speech emotion recognition: Features and classification models,” Digit. Signal Process., vol. 22, no. 6, pp. 1154–1160, Dec. 2012.

[48] P. Rani, C. Liu, N. Sarkar, and E. Vanman, “An empirical study of machine learning techniques for affect recognition in human--robot interaction,” Pattern Anal. Appl., vol. 9, no. 1, pp. 58–69, 2006.

[49] C. Meyer, F. Eyben, and T. Zielke, "DEEP NEURAL NETWORKS FOR ACOUSTIC EMOTION RECOGNITION : RAISING THE BENCHMARKS Andr ´ Dept . of Mechanical and Process Engineering , D ¨ usseldorf University of Applied Sciences , Germany Dept . of Electrical Engineering , D ¨ usseldorf University of Appl," pp. 5688–5691, 2011.

[50] H. Gao, S. Chen, P. An, and G. Su, "Emotion recognition of mandarin speech for different speech corpora based on nonlinear features," in Signal Processing (ICSP), 2012 IEEE 11th International Conference on, 2012, vol. 1, pp. 567–570.

[51] D. Le and E. M. Provost, "Emotion recognition from spontaneous speech using Hidden Markov models with deep belief networks," in Automatic Speech Recognition and Understanding (ASRU), 2013 IEEE Workshop on, 2013, pp. 216–221.

[52] M. Wollmer, B. Schuller, F. Eyben, and G. Rigoll, "Combining long short-term memory and dynamic bayesian networks for incremental emotion-sensitive artificial listening," Sel. Top. Signal Process. IEEE J., vol. 4, no. 5, pp. 867–881, 2010.

[53] R. Cowie, N. Sussman, and A. Ben-Ze'ev, "Emotion: Concepts and definitions," in Emotion-Oriented Systems, Springer, 2011, pp. 9–30.

[54] A. Batliner, B. Schuller, D. Seppi, S. Steidl, L. Devillers, L. Vidrascu, T. Vogt, V. Aharonson, and N. Amir, "The automatic recognition of emotions in speech," in Emotion-Oriented Systems, Springer, 2011, pp. 71–99.

[55] N. Escoffier, J. Zhong, A. Schirmer, and A. Qiu, "Emotional expressions in voice and music: Same code, same effect?," Hum. Brain Mapp., vol. 34, no. 8, pp. 1796–1810, 2013.

[56] I. R. Murray and J. L. Arnott, "Applying an analysis of acted vocal emotions to improve the simulation of synthetic speech," Comput. Speech Lang., vol. 22, no. 2, pp. 107–129, 2008.

[57] M. Guzman, S. Correa, D. Muñoz, and R. Mayerhoff, "Influence on spectral energy distribution of emotional expression.," J. Voice, vol. 27, no. 1, pp. 129.e1–129.e10, Jan. 2013.

[58] A. Origlia, F. Cutugno, and V. Galata, "ScienceDirect," vol. 57, pp. 155–169, 2014.

[59] J. M. Iredale, J. a Rushby, S. McDonald, A. Dimoska-Di Marco, and J. Swift, "Emotion in voice matters: neural correlates of emotional prosody perception.," Int. J. Psychophysiol., vol. 89, no. 3, pp. 483–90, Sep. 2013.

[60] H. Cao, R. Verma, and A. Nenkova, "Speaker-sensitive emotion recognition via ranking: Studies on acted and spontaneous speech," Comput. Speech Lang., pp. 1–17, Feb. 2014.

[61] H. Cao, A. Savran, R. Verma, and A. Nenkova, "Acoustic and lexical representations for affect prediction in spontaneous conversations," Comput. Speech Lang., pp. 1–15, Apr. 2014.

[62] X.-W. Wang, D. Nie, and B.-L. Lu, "Emotional state classification from EEG data using machine learning approach," Neurocomputing, vol. 129, pp. 94–106, Apr. 2014.

[63] W. Li and H. Xu, "Text-based emotion classification using emotion cause extraction," Expert Syst. Appl., vol. 41, no. 4, pp. 1742–1749, Mar. 2014.

[64] N. Kamaruddin, A. Wahab, and C. Quek, "Expert Systems with Applications Cultural dependency analysis for understanding speech emotion," Expert Syst. Appl., vol. 39, no. 5, pp. 5115–5133, 2012.

[65] C. Breazeal and L. Aryananda, "Recognition of affective communicative intent in robot-directed speech," Auton. Robots, vol. 12, no. 1, pp. 83–104, 2002.

[66] D. Morrison, R. Wang, and L. C. De Silva, "Ensemble methods for spoken emotion recognition in call-centres," Speech Commun., vol. 49, no. 2, pp. 98–112, 2007.

[67] C. M. Lee and S. S. Narayanan, "Toward detecting emotions in spoken dialogs," Speech Audio Process. IEEE Trans., vol. 13, no. 2, pp. 293–303, 2005.

[68] M. Slaney and G. McRoberts, "BabyEars: A recognition system for affective vocalizations," Speech Commun., vol. 39, no. 3, pp. 367–384, 2003.

[69] F. Burkhardt, A. Paeschke, M. Rolfes, W. F. Sendlmeier, and B. Weiss, "A database of German emotional speech.," in Interspeech, 2005, vol. 5, pp. 1517–1520.

[70] C. E. Williams and K. N. Stevens, "Emotions and speech: Some acoustical correlates," J. Acoust. Soc. Am., vol. 52, no. 4B, pp. 1238–1250, 1972.

[71] D. Ververidis and C. Kotropoulos, "A review of emotional speech databases," in Proc. Panhellenic Conference on Informatics (PCI), 2003, pp. 560–574.

[72] M. El Ayadi, M. S. Kamel, and F. Karray, "Survey on speech emotion recognition: Features, classification schemes, and databases," Pattern Recognit., vol. 44, no. 3, pp. 572–587, Mar. 2011.

[73] K. Halsey, "Reading the evidence of reading: An introduction to the Reading Experience Database, 1450--1945," Pop. Narrat. Media, vol. 1, no. 2, pp. 123–137, 2008.

[74] E. Douglas-Cowie, N. Campbell, R. Cowie, and P. Roach, "Emotional speech: Towards a new generation of databases," Speech Commun., vol. 40, no. 1, pp. 33–60, 2003.

[75] M. A. Franey, G. C. Idzorek, and T. Linstroth, "Interactive synthesized speech quotation system for brokers." Google Patents, 2004.

[76] M. A. Walker, R. Passonneau, and J. E. Boland, "Quantitative and qualitative evaluation of DARPA Communicator spoken dialogue systems," in Proceedings of the 39th Annual Meeting on Association for Computational Linguistics, 2001, pp. 515–522.

[77] B. Schuller, S. Reiter, R. Muller, M. Al-Hames, M. Lang, and G. Rigoll, "Speaker independent speech emotion recognition by ensemble classification," in Multimedia and Expo, 2005. ICME 2005. IEEE International Conference on, 2005, pp. 864–867.

[78] R. Fernandez and R. W. Picard, "Classical and novel discriminant features for affect recognition from speech.," in Interspeech, 2005, pp. 473–476.

[79] S. McGilloway, R. Cowie, E. Douglas-Cowie, S. Gielen, M. Westerdijk, and S. Stroeve, "Approaching automatic recognition of emotion from voice: a rough benchmark," in ISCA Tutorial and Research Workshop (ITRW) on Speech and Emotion, 2000.

[80] M. Schröder, R. Cowie, E. Douglas-Cowie, M. Westerdijk, and S. C. A. M. Gielen, "Acoustic correlates of emotion dimensions in view of speech synthesis.," in INTERSPEECH, 2001, pp. 87–90.

[81] C. Pereira, "Dimensions of emotional meaning in speech," in ISCA Tutorial and Research Workshop (ITRW) on Speech and Emotion, 2000.

[82] S. M. Yacoub, S. J. Simske, X. Lin, and J. Burns, "Recognition of emotions in interactive voice response systems.," in INTERSPEECH, 2003.

[83] S. G. Pulman, J. Boye, M. Cavazza, C. Smith, and R. S. De La Cámara, "'How was your day?,'" in Proceedings of the 2010 Workshop on Companionable Dialogue Systems, 2010, pp. 37–42.

[84] T. Polzehl, A. Schmitt, F. Metze, and M. Wagner, "Anger recognition in speech using acoustic and linguistic cues," Speech Commun., vol. 53, no. 9–10, pp. 1198–1209, Nov. 2011.

[85] K. Sato, M. Yuki, and V. Norasakkunkit, "A Socio-Ecological Approach to Cross-Cultural Differences in the Sensitivity to Social Rejection The Partially Mediating Role of Relational Mobility," J. Cross. Cult. Psychol., vol. 45, no. 10, pp. 1549–1560, 2014.

[86] B. Fasel and J. Luettin, "Automatic facial expression analysis: a survey," Pattern Recognit., vol. 36, no. 1, pp. 259–275, 2003.

[87] Y. Nakanishi, Y. Yoshitomi, T. Asada, and M. Tabuse, "Facial expression recognition of a speaker using thermal image processing and reject criteria in feature vector space," Artif. Life Robot., vol. 19, no. 1, pp. 76–88, 2014.

[88] L. Petrinovich, "Darwin and the representative expression of reality," Ekman P Darwin Facial Expression. A Century Res. Rev. Malor Books Los Altos. S, pp. 223–254, 2006.

[89] P. Ekman, "Darwin's Compassionate View of Human Nature," JAMA, vol. 303, no. 6, pp. 557–558, 2010.

[90] K. S. Kendler, L. J. Halberstadt, F. Butera, J. Myers, T. Bouchard, and P. Ekman, "The similiarity of facial expressions in response to emotion-inducing films in reared-apart twins," Psychol. Med., vol. 38, no. 10, pp. 1475–1483, 2008.

[91] M. Pantic, A. Pentland, A. Nijholt, and T. S. Huang, "Human computing and machine understanding of human behavior: a survey," in Artifical Intelligence for Human Computing, Springer, 2007, pp. 47–71.

[92] M. El Ayadi, M. S. Kamel, and F. Karray, "Survey on speech emotion recognition: Features, classification schemes, and databases," Pattern Recognit., vol. 44, no. 3, pp. 572–587, 2011.

[93] E. Hjelmås and B. K. Low, "Face detection: A survey," Comput. Vis. image Underst., vol. 83, no. 3, pp. 236–274, 2001.

[94] M. Murtaza, M. Sharif, M. Raza, and J. H. Shah, "Analysis of face recognition under varying facial expression: a survey.," Int. Arab J. Inf. Technol., vol. 10, no. 4, pp. 378–388, 2013.

[95] M. Pantic and L. J. M. Rothkrantz, "Automatic analysis of facial expressions: The state of the art," Pattern Anal. Mach. Intell. IEEE Trans., vol. 22, no. 12, pp. 1424–1445, 2000.

[96] M. F. Valstar, B. Jiang, M. Mehu, M. Pantic, and K. Scherer, “The first facial expression recognition and analysis challenge,” in Automatic Face & Gesture Recognition and Workshops (FG 2011), 2011 IEEE International Conference on, 2011, pp. 921–926.

[97] W. Zhao, R. Chellappa, P. J. Phillips, and A. Rosenfeld, “Face recognition: A literature survey,” Acm Comput. Surv., vol. 35, no. 4, pp. 399–458, 2003.

[98] M. Pantic and L. J. M. Rothkrantz, “Expert system for automatic analysis of facial expressions,” Image Vis. Comput., vol. 18, no. 11, pp. 881–905, 2000.

[99] A. Lanitis, C. J. Taylor, and T. F. Cootes, “Automatic interpretation and coding of face images using flexible models,” Pattern Anal. Mach. Intell. IEEE Trans., vol. 19, no. 7, pp. 743–756, 1997.

[100] I. Matthews and S. Baker, “Active appearance models revisited,” Int. J. Comput. Vis., vol. 60, no. 2, pp. 135–164, 2004.

[101] M. B. Stegmann, “ACTIVE APPEARANCE,” 2000.

[102] H. Kobayashi and F. Hara, “Facial interaction between animated 3D face robot and human beings,” in Systems, Man, and Cybernetics, 1997. Computational Cybernetics and Simulation., 1997 IEEE International Conference on, 1997, vol. 4, pp. 3732–3737.

[103] J. Wang, L. Yin, X. Wei, and Y. Sun, “3D facial expression recognition based on primitive surface feature distribution,” in Computer Vision and Pattern Recognition, 2006 IEEE Computer Society Conference on, 2006, vol. 2, pp. 1399–1406.

[104] H. Hong, H. Neven, and C. der Malsburg, “Online facial expression recognition based on personalized galleries,” in Automatic Face and Gesture Recognition, 1998. Proceedings. Third IEEE International Conference on, 1998, pp. 354–359.

[105] X. Feng, M. Pietikainen, and A. Hadid, “Facial expression recognition with local binary patterns and linear programming,” Pattern Recognit. Image Anal. C/C Raspoznavaniye Obraz. I Anal. Izobr., vol. 15, no. 2, p. 546, 2005.

[106] Z. Zhang, M. Lyons, M. Schuster, and S. Akamatsu, “Comparison between geometry-based and gabor-wavelets-based facial expression recognition using multi-layer perceptron,” in Automatic Face and Gesture Recognition, 1998. Proceedings. Third IEEE International Conference on, 1998, pp. 454–459.

[107] C. Padgett and G. W. Cottrell, “Representing face images for emotion classification,” Adv. Neural Inf. Process. Syst., pp. 894–900, 1997.

[108] B. Moghaddam and M.-H. Yang, “Learning gender with support faces,” Pattern Anal. Mach. Intell. IEEE Trans., vol. 24, no. 5, pp. 707–711, 2002.

[109] C. L. Lisetti and D. E. Rumelhart, “Facial Expression Recognition Using a Neural Network.,” in FLAIRS Conference, 1998, pp. 328–332.

[110] M. S. Bartlett, “Face image analysis by unsupervised learning and redundancy reduction,” Citeseer, 1998.

[111] M. S. Bartlett, G. Donato, J. R. Movellan, J. C. Hager, P. Ekman, and T. J. Sejnowski, “Image Representations for Facial Expression Coding.,” in NIPS, 1999, pp. 886–892.

[112] I. Cohen, N. Sebe, A. Garg, L. S. Chen, and T. S. Huang, “Facial expression recognition from video sequences: temporal and static modeling,” Comput. Vis. Image Underst., vol. 91, no. 1, pp. 160–187, 2003.

[113] E. B. Roesch, L. Tamarit, L. Reveret, D. Grandjean, D. Sander, and K. R. Scherer, “FACSGen: a tool to synthesize emotional facial expressions through systematic manipulation of facial action units,” J. Nonverbal Behav., vol. 35, no. 1, pp. 1–16, 2011.

[114] D. Hefenbrock, J. Oberg, N. Thanh, R. Kastner, and S. B. Baden, “Accelerating Viola-Jones Face Detection to FPGA-Level Using GPUs.,” in FCCM, 2010, pp. 11–18.

[115] A. Maalej, B. Ben Amor, M. Daoudi, A. Srivastava, and S. Berretti, “Shape analysis of local facial patches for 3D facial expression recognition,” Pattern Recognit., vol. 44, no. 8, pp. 1581–1589, 2011.

[116] W. Zheng, H. Tang, Z. Lin, and T. S. Huang, “A novel approach to expression recognition from non-frontal face images,” in Computer Vision, 2009 IEEE 12th International Conference on, 2009, pp. 1901–1908.

[117] M. Pantic and M. S. Bartlett, “Machine analysis of facial expressions,” 2007.

[118] Z. Zeng, M. Pantic, G. I. Roisman, and T. S. Huang, “A survey of affect recognition methods: audio, visual, and spontaneous expressions.,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 31, no. 1, pp. 39–58, Jan. 2009.

[119] F. la Torre and J. F. Cohn, “Facial expression analysis,” in Visual Analysis of Humans, Springer, 2011, pp. 377–409.

[120] O. Starostenko, R. Contreras, V. A. Aquino, L. F. Pulido, J. R. Asomoza, O. Sergiyenko, and V. Tyrsa, “A Fuzzy Reasoning Model for Recognition of Facial Expressions,” Comput. y Sist., vol. 15, no. 2, pp. 163–180, 2011.

[121] D. McDuff, R. El Kaliouby, T. Senechal, M. Amr, J. F. Cohn, and R. Picard, “Affectiva-MIT Facial Expression Dataset (AM-FED): Naturalistic and Spontaneous Facial Expressions Collected‘ In-the-Wild,’” in Computer Vision and Pattern Recognition Workshops (CVPRW), 2013 IEEE Conference on, 2013, pp. 881–888.

[122] J.-C. Lin, C.-H. Wu, and W.-L. Wei, “Error weighted semi-coupled hidden Markov model for audio-visual emotion recognition,” Multimedia, IEEE Trans., vol. 14, no. 1, pp. 142–156, 2012.

[123] L. H. Thai, N. D. T. Nguyen, and T. S. Hai, “A facial expression classification system integrating canny, principal component analysis and artificial neural network,” arXiv Prepr. arXiv1111.4052, 2011.

[124] J. F. Cohn, “Observer-Based Measurement of Facial Expression With the Facial Action Coding System,” pp. 203–221, 2005.

[125] A. Ramirez Rivera, R. Castillo, and O. Chae, “Local Directional Number Pattern for Face Analysis: Face and Expression Recognition,” Image Process. IEEE Trans., vol. 22, no. 5, pp. 1740–1752, May 2013.

[126] D. W. Massaro, Perceiving talking faces: From speech perception to a behavioral principle, vol. 1. Mit Press, 1998.

[127] A. Bruce, I. Nourbakhsh, and R. Simmons, “The role of expressiveness and attention in human-robot interaction,” in Robotics and Automation, 2002. Proceedings. ICRA’02. IEEE International Conference on, 2002, vol. 4, pp. 4138–4142.

[128] J. Zhang and A. J. C. Sharkey, “It’s not all written on the robot’s face,” Rob. Auton. Syst., vol. 60, no. 11, pp. 1449–1456, Nov. 2012.

[129] G. L. Ahern and G. E. Schwartz, “Differential lateralization for positive and negative emotion in the human brain: EEG spectral analysis,” Neuropsychologia, vol. 23, no. 6, pp. 745–755, 1985.

[130] T. Danisman and A. Alpkocak, “Feeler: Emotion classification of text using vector space model,” in AISB 2008 Convention Communication, Interaction and Social Intelligence, 2008, vol. 1, p. 53.

[131] P. J. Lang, A. Öhman, and D. Vaitl, “The international affective picture system [photographic slides],” Gainesville, FL Cent. Res. Psychophysiology, Univ. Florida, 1988.

[132] W. Hubert and R. de Jong-Meyer, “Autonomic, neuroendocrine, and subjective responses to emotion-inducing film stimuli,” Int. J. Psychophysiol., vol. 11, no. 2, pp. 131–140, 1991.

[133] K. R. Scherer, “Which emotions can be induced by music? What are the underlying mechanisms? And how can we measure them?,” J. new Music Res., vol. 33, no. 3, pp. 239–251, 2004.

[134] E. M. M. Artignoni, C. L. P. Acchetti, F. R. M. Ancini, R. O. A. Glieri, C. I. R. A. F. U. O, and G. I. N. Appi, “Active Music Therapy in Parkinson ’ s Disease : An Integrative Method for Motor and Emotional Rehabilitation,” vol. 393, pp. 386–393, 2000.

[135] A. Majumder, L. Behera, and V. K. Subramanian, “Emotion recognition from geometric facial features using self-organizing map,” Pattern Recognit., vol. 47, no. 3, pp. 1282–1293, Mar. 2014.

[136] S. Wan and J. K. Aggarwal, “Spontaneous facial expression recognition: A robust metric learning approach,” Pattern Recognit., vol. 47, no. 5, pp. 1859–1868, May 2014.

[137] G. Caridakis, K. Karpouzis, and S. Kollias, “User and context adaptive neural networks for emotion recognition,” Neurocomputing, vol. 71, no. 13–15, pp. 2553–2562, Aug. 2008.

[138] F. Dornaika, E. Lazkano, and B. Sierra, “Improving dynamic facial expression recognition with feature subset selection,” Pattern Recognit. Lett., vol. 32, no. 5, pp. 740–748, 2011.

[139] M. F. Valstar, M. Pantic, Z. Ambadar, and J. F. Cohn, “Spontaneous vs. Posed Facial Behavior: Automatic Analysis of Brow Actions,” in Proceedings of the 8th International Conference on Multimodal Interfaces, 2006, pp. 162–170.

[140] G. U. Kharat and S. V Dudul, “Human Emotion Recognition System Using Optimally Designed SVM with Different Facial Feature Extraction Techniques,” W. Trans. Comp., vol. 7, no. 6, pp. 650–659, Jun. 2008.

[141] K. Yurtkan and H. Demirel, “Feature selection for improved 3D facial expression recognition,” Pattern Recognit. Lett., vol. 38, pp. 26–33, Mar. 2014.

[142] M. I. Khan and M. A. Bhuiyan, “Facial Features Approximation for Expression Detection in Human-Robot Interface,” in Intelligent Information Hiding and Multimedia Signal Processing (IIH-MSP), 2010 Sixth International Conference on, 2010, pp. 393–397.

[143] V. Gomathi, K. Ramar, and A. S. Jeeyakumar, “Human facial expression recognition using MANFIS model,” World Acad. Sci. Eng. Technol., vol. 50, p. 2009, 2009.

[144] D. McDuff, R. El Kaliouby, K. Kassam, and R. Picard, “Affect valence inference from facial action unit spectrograms,” in Computer Vision and Pattern Recognition Workshops (CVPRW), 2010 IEEE Computer Society Conference on, 2010, pp. 17–24.

[145] G. Sandbach, S. Zafeiriou, M. Pantic, and L. Yin, “Static and dynamic 3D facial expression recognition: A comprehensive survey,” Image Vis. Comput., vol. 30, no. 10, pp. 683–697, 2012.

[146] F. Tsalakanidou and S. Malassiotis, “Robust facial action recognition from real-time 3D streams,” in Computer Vision and Pattern Recognition Workshops, 2009, pp. 4–11.

[147] G. Sandbach, S. Zafeiriou, M. Pantic, and D. Rueckert, “Recognition of 3D facial expression dynamics,” Image Vis. Comput., vol. 30, no. 10, pp. 762–773, 2012.

[148] N. Sebe, M. S. Lew, Y. Sun, I. Cohen, T. Gevers, and T. S. Huang, “Authentic facial expression analysis,” Image Vis. Comput., vol. 25, no. 12, pp. 1856–1863, Dec. 2007.

[149] N. Karpinsky and S. Zhang, “High-resolution, real-time 3D imaging with fringe analysis,” J. Real-Time Image Process., vol. 7, no. 1, pp. 55–66, 2012.

[150] F. Tsalakanidou and S. Malassiotis, “Real-time 2D+ 3D facial action and expression recognition,” Pattern Recognit., vol. 43, no. 5, pp. 1763–1775, 2010.

[151] T.-H. Wang and J.-J. James Lien, “Facial expression recognition system based on rigid and non-rigid motion separation and 3D pose estimation,” Pattern Recognit., vol. 42, no. 5, pp. 962–977, May 2009.

[152] R. A. Khan, A. Meyer, H. Konik, and S. Bouakaz, “Framework for reliable, real-time facial expression recognition for low resolution images,” Pattern Recognit. Lett., vol. 34, no. 10, pp. 1159–1168, 2013.

[153] C.-T. Liao, S.-F. Wang, Y.-J. Lu, and S.-H. Lai, “Video-based face recognition based on view synthesis from 3D face model reconstructed from a single image,” in Multimedia and Expo, 2008 IEEE International Conference on, 2008, pp. 1589–1592.

[154] H. Fang, N. Mac Parthaláin, A. J. Aubrey, G. K. L. Tam, R. Borgo, P. L. Rosin, P. W. Grant, D. Marshall, and M. Chen, “Facial expression recognition in dynamic sequences: An integrated approach,” Pattern Recognit., vol. 47, no. 3, pp. 1271–1281, Mar. 2014.

[155] A. Besinger, T. Sztynda, S. Lal, C. Duthoit, J. Agbinya, B. Jap, D. Eager, and G. Dissanayake, “Optical flow based analyses to detect emotion from human facial image data,” Expert Syst. Appl., vol. 37, no. 12, pp. 8897–8902, Dec. 2010.

[156] V. Bettadapura, “Face Expression Recognition and Analysis : The State of the Art,” pp. 1–27.

[157] A. M. Martinez, “The AR face database,” CVC Tech. Rep., vol. 24, 1998.

[158] G. B. Huang, M. Mattar, T. Berg, and E. Learned-Miller, “E.: Labeled faces in the wild: A database for studying face recognition in unconstrained environments,” 2007.

[159] Y. Wang, H. Ai, B. Wu, and C. Huang, “Real time facial expression recognition with adaboost,” in Pattern Recognition, 2004. ICPR 2004. Proceedings of the 17th International Conference on, 2004, vol. 3, pp. 926–929.

[160] T. Sim, S. Baker, and M. Bsat, “The CMU pose, illumination, and expression database,” Pattern Anal. Mach. Intell. IEEE Trans., vol. 25, no. 12, pp. 1615–1618, 2003.

[161] V. Jain and A. Mukherjee, “The Indian face database,” UR L http//vis-www. cs. umass.. edu/\ sim vidit/{I} ndian {F} ace {D} atabase, 2002.

[162] M. Minear and D. C. Park, “A lifespan database of adult facial stimuli,” Behav. Res. Methods, Instruments, Comput., vol. 36, no. 4, pp. 630–633, 2004.

[163] T. Kanade, J. F. Cohn, and Y. Tian, “Comprehensive database for facial expression analysis,” in Automatic Face and Gesture Recognition, 2000. Proceedings. Fourth IEEE International Conference on, 2000, pp. 46–53.

[164] M. Szwoch, “FEEDB: a multimodal database of facial expressions and emotions,” in Human System Interaction (HSI), 2013 The 6th International Conference on, 2013, pp. 524–531.

[165] G. P. Redei, “http://blog.mashape.com/list-of-50-face-detection-recognition-apis/.” 15-Jul-2014.

[166] J. Movellan, M. S. Bartlett, I. Fasel, G. F. Littlewort, J. Susskind, and J. Whitehill, “COLLECTION OF MACHINE LEARNING TRAINING DATA FOR EXPRESSION RECOGNITION.” 2014.

[167] G. Littlewort, J. Whitehill, T. Wu, I. Fasel, M. Frank, J. Movellan, and M. Bartlett, “The computer expression recognition toolbox (CERT),” in Automatic Face & Gesture Recognition and Workshops (FG 2011), 2011 IEEE International Conference on, 2011, pp. 298–305.

[168] T. Wu, N. J. Butko, P. Ruvolo, J. Whitehill, M. S. Bartlett, and J. R. Movellan, “Multi-Layer Architectures for Facial Action Unit Recognition.”

[169] S. Kamberi, “A Cross-Case Analysis of Possible Facial Emotion Extraction Methods that Could Be Used in Second Life-Pre Experimental Work,” J. Virtual Worlds Res., vol. 5, no. 3, 2012.

[170] J. Studies, “Institute of Judicial Studies Institute of Judicial Studies.”

[171] M. Wöllmer, A. Metallinou, F. Eyben, B. Schuller, and S. S. Narayanan, “Context-sensitive multimodal emotion recognition from speech and facial expression using bidirectional LSTM modeling.,” in INTERSPEECH, 2010, pp. 2362–2365.

[172] H. Cao, A. Savran, R. Verma, and A. Nenkova, “spontaneous conversations ც,” pp. 1–15, 2014.

[173] W. Martin, A. Metallinou, F. Eyben, and S. Narayanan, “Context-Sensitive Multimodal Emotion Recognition from Speech and Facial Expression using Bidirectional LSTM Modeling,” no. September, pp. 2362–2365, 2010.

[174] Y. Yoshitomi and others, “Facial expression recognition for speaker using thermal image processing and speech recognition system,” in Proceedings of the 10th WSEAS international conference on Applied computer science, 2010, pp. 182–186.

[175] B. Straube, A. Green, A. Jansen, A. Chatterjee, and T. Kircher, “Social cues, mentalizing and the neural processing of speech accompanied by gestures,” Neuropsychologia, vol. 48, no. 2, pp. 382–393, 2010.

[176] Y. Ji and K. Idrissi, “Automatic facial expression recognition based on spatiotemporal descriptors,” Pattern Recognit. Lett., vol. 33, no. 10, pp. 1373–1380, 2012.

[177] G. Lortal, S. Dhouib, and S. Gérard, “Integrating ontological domain knowledge into a robotic DSL,” in Models in Software Engineering, Springer, 2011, pp. 401–414.

[178] A. Juarez, C. Bartneck, and L. Feijs, “Using semantic technologies to describe robotic embodiments,” in Proceedings of the 6th international conference on Human-robot interaction, 2011, pp. 425–432.

[179] P. Jaeckel, N. Campbell, and C. Melhuish, “Towards realistic facial behaviour in humanoids-mapping from video footage to a robot head,” in Rehabilitation Robotics, 2007. ICORR 2007. IEEE 10th International Conference on, 2007, pp. 833–840.

[180] B. Yang and M. Lugger, “Emotion recognition from speech signals using new harmony features,” Signal Processing, vol. 90, no. 5, pp. 1415–1423, 2010.

[181] W.-K. Tsao, A. J. T. Lee, Y.-H. Liu, T.-W. Chang, and H.-H. Lin, “A data mining approach to face detection,” Pattern Recognit., vol. 43, no. 3, pp. 1039–1049, 2010.

[182] A. Kirsch, “Robot learning language—integrating programming and learning for cognitive systems,” Rob. Auton. Syst., vol. 57, no. 9, pp. 943–954, 2009.

[183] P. Boersma, “Praat, a system for doing phonetics by computer,” Glot Int., vol. 5, no. 9/10, pp. 341–345, 2002.

[184] J. Chen, X. Liu, P. Tu, and A. Aragones, “Learning person-specific models for facial expression and action unit recognition,” Pattern Recognit. Lett., vol. 34, no. 15, pp. 1964–1970, 2013.

[185] M. Sorci, G. Antonini, J. Cruz, T. Robin, M. Bierlaire, and J.-P. Thiran, “Modelling human perception of static facial expressions,” Image Vis. Comput., vol. 28, no. 5, pp. 790–806, 2010.

[186] W. Zhao and R. Chellappa, Face Processing: Advanced Modeling and Methods: Advanced Modeling and Methods. Academic Press, 2011.

[187] C. E. Izard, “Basic emotions, natural kinds, emotion schemas, and a new paradigm,” Perspect. Psychol. Sci., vol. 2, no. 3, pp. 260–280, 2007.

[188] Q. Chen and G. Wang, “A class of Bézier-like curves,” Comput. Aided Geom. Des., vol. 20, no. 1, pp. 29–39, Mar. 2003.

[189] Q. Chen and G. Wang, “A class of B{é}zier-like curves,” Comput. Aided Geom. Des., vol. 20, no. 1, pp. 29–39, 2003.

[190] J. Shawe-Taylor and N. Cristianini, Kernel methods for pattern analysis. Cambridge university press, 2004.

[191] J. Wagner, F. Lingenfelser, T. Baur, I. Damian, F. Kistler, and E. André, “The social signal interpretation (SSI) framework: multimodal signal processing and recognition in real-time,” in Proceedings of the 21st ACM international conference on Multimedia, 2013, pp. 831–834.

[192] F. Eyben, M. Wollmer, and B. Schuller, “OpenEAR—introducing the Munich open-source emotion and affect recognition toolkit,” in Affective Computing and Intelligent Interaction and Workshops, 2009. ACII 2009. 3rd International Conference on, 2009, pp. 1–6.

[193] N. H. de Jong and T. Wempe, “Praat script to detect syllable nuclei and measure speech rate automatically,” Behav. Res. Methods, vol. 41, no. 2, pp. 385–390, 2009.

[194] B. Kreifelts, T. Ethofer, T. Shiozawa, W. Grodd, and D. Wildgruber, “Cerebral representation of non-verbal emotional perception: fMRI reveals audiovisual integration area between voice-and face-sensitive regions in the superior temporal sulcus,” Neuropsychologia, vol. 47, no. 14, pp. 3059–3066, 2009.

[195] J. P. Goldman, “EasyAlign: An Automatic Phonetic Alignment Tool Under Praat.,” in INTERSPEECH, 2011, pp. 3233–3236.

[196] W. Styler, “Using Praat for Linguistic Research,” Univ. Color. Boulder Phonetics Lab, 2013.

[197] T. Polzehl, A. Schmitt, and F. Metze, “Salient features for anger recognition in german and english ivr portals,” in Spoken dialogue systems technology and design, Springer, 2011, pp. 83–105.

[198] M. Grimm, K. Kroschel, E. Mower, and S. Narayanan, “Primitives-based Evaluation and Estimation of Emotions in Speech,” Speech Commun., vol. 49, no. 10–11, pp. 787–800, Oct. 2007.

[199] J. F. Kaiser, “On a simple algorithm to calculate theenergy’of a signal,” 1990.

[200] S. Milborrow and F. Nicolls, “Locating facial features with an extended active shape model,” in Computer Vision--ECCV 2008, Springer, 2008, pp. 504–513.

[201] A. Origlia, G. Abete, and F. Cutugno, “A dynamic tonal perception model for optimal pitch stylization,” Comput. Speech Lang., vol. 27, no. 1, pp. 190–208, 2013.

[202] K. Weber, S. Bengio, and H. Bourlard, “HMM2-extraction of formant structures and their use for robust ASR.,” in INTERSPEECH, 2001, pp. 607–610.

[203] A. Savran, B. Sankur, and M. Taha Bilge, “Comparative evaluation of 3D vs. 2D modality for automatic detection of facial action units,” Pattern Recognit., vol. 45, no. 2, pp. 767–782, 2012.

[204] T. Fang, X. Zhao, O. Ocegueda, S. K. Shah, and I. A. Kakadiaris, “3D/4D facial expression analysis: an advanced annotated face model approach,” Image Vis. Comput., vol. 30, no. 10, pp. 738–749, 2012.

[205] S. D. Pollak, M. Messner, D. J. Kistler, and J. F. Cohn, “Development of perceptual expertise in emotion recognition,” Cognition, vol. 110, no. 2, pp. 242–247, 2009.

[206] G. J. Edwards, T. F. Cootes, and C. J. Taylor, “Face recognition using active appearance models,” in Computer Vision—ECCV’98, Springer, 1998, pp. 581–595.

[207] T. Hashimoto, S. Hitramatsu, T. Tsuji, and H. Kobayashi, “Development of the face robot SAYA for rich facial expressions,” in SICE-ICASE, 2006. International Joint Conference, 2006, pp. 5423–5428.

[208] H. Kobayashi and F. Hara, “Recognition of six basic facial expression and their strength by neural network,” in Robot and Human Communication, 1992. Proceedings., IEEE International Workshop on, 1992, pp. 381–386.

[209] C.-L. Huang and Y.-M. Huang, “Facial expression recognition using model-based feature extraction and action parameters classification,” J. Vis. Commun. Image Represent., vol. 8, no. 3, pp. 278–290, 1997.

[210] N. Yuill and J. Lyon, “Selective difficulty in recognising facial expressions of emotion in boys with ADHD,” Eur. Child Adolesc. Psychiatry, vol. 16, no. 6, pp. 398–404, 2007.

[211] R. Fernandez and R. Picard, “Recognizing affect from speech prosody using hierarchical graphical models,” Speech Commun., vol. 53, no. 9–10, pp. 1088–1103, 2011.

[212] K. Kumar, A. Yadav, and H. Rani, “Artificial Neural Network based Classification of IC through Extracting the Feature Set of IC Images using 2-Dimensional Discrete Wavelet Transform,” Int. J. Comput. Intell. Res., vol. 9, no. 2, pp. 115–119, 2013.

[213] P. L. De Leon, I. Hernaez, I. Saratxaga, M. Pucher, and J. Yamagishi, “Detection of synthetic speech for the problem of imposture,” in Acoustics, Speech and Signal Processing (ICASSP), 2011 IEEE International Conference on, 2011, pp. 4844–4847.

[214] J. Wagner, E. Andre, F. Lingenfelser, and J. Kim, “Exploring fusion methods for multimodal emotion recognition with missing data,” Affect. Comput. IEEE Trans., vol. 2, no. 4, pp. 206–218, 2011.

[215] E. L. van den Broek and J. H. D. M. Westerink, “Considerations for emotion-aware consumer products,” Appl. Ergon., vol. 40, no. 6, pp. 1055–1064, 2009.

[216] M. Soleymani, M. Pantic, and T. Pun, “Multimodal emotion recognition in response to videos,” Affect. Comput. IEEE Trans., vol. 3, no. 2, pp. 211–223, 2012.