IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 2, Issue 5, June 2022
Impact Factor: 6.252

Emotion Recognition using Python

Sahil Sharma
Student, Department of Computer Science & Information Technology
Dronacharya College of Engineering, Gurugram, India

Abstract: *Human-computer contact will be more natural if computers can perceive and respond to nonverbal communication such as emotions. Although various ways to recognizing human emotions based on facial expressions or speech have been presented, there has been relatively little effort done to merge these two modalities, as well as others, to improve the accuracy and robustness of the emotion detection system. This research examines the benefits and drawbacks of systems that rely solely on facial expressions or audio data. Facial expression recognition is a subset of facial recognition that is gaining in importance as the need for it grows.*

Keywords: Facial recognition; expression recognition; deep learning; image recognition; Facial technology; signal processing; image classification are all terms used to describe facial recognition.

I. INTRODUCTION

Nonverbal cues such as hand gestures, facial expressions, and voice tone are utilized to express feelings and provide feedback in inter-personal human communication. However, new human computer interface trends, which have evolved from the traditional mouse and keyboard to automatic speech recognition systems and special interfaces designed for handicapped people, do not fully exploit these valuable communicative abilities, resulting in less-than-natural interactions. If computers could understand these emotional cues, they could provide more personalized and appropriate assistance to people based on their needs and preferences. Psychological theory suggests that human emotions can be divided into six archetypal emotions: surprise, fear, anger, and sadness.. To transmit different feelings, the muscles of the face can be adjusted, as well as the tone and energy with which the speech is produced. By simultaneously processing information collected by ears and sight, humans can perceive these signals, even if they are softly expressed. Based on psychological research that show that visual information influences speech perception [17], it's reasonable to believe that human emotion perception follows a similar pattern. DE Silva ET AL. were inspired by these findings and conducted trials in which 18 persons were asked to evaluate emotion using both visual and acoustic information from an audio-visual database recorded from two subjects [7]. They came to the conclusion that some emotions, such as sadness, are better identified through audio. Furthermore, Chen ET AL. demonstrated that these two modalities provide complimentary information by claiming that when both modalities were examined together, the system's performance improved [4]. Although various automatic emotion identification systems have looked into using facial expressions [1],[11],[16],[21],[22] or voice [9],[18],[14] to detect human emotional states, only a few have looked into using both modalities [4],[8]. When one of these modalities is acquired in a noisy environment, it is hoped that the multi modal method will provide not just greater performance but also higher robustness [19]. Previous research mixed facial expressions with audio information either at a decision-level, where the outputs of uni modal systems are combined using appropriate criteria, or at a feature-level, where input from both modalities is combined before categorization. None of these studies, however, attempted to compare which fusion method is better for emotion identification. This article compares and contrasts the two fusion approaches in terms of overall system performance.

II. EMOTION RECOGNITION SYSTEMS

2.1 Emotion recognition by speech

There have been several techniques of recognizing emotions from speech. [6] and [19] provide extensive reviews of various approaches. Most researchers have derived utterance-level statistics using global supplemental/prosody features as their auditory cues for emotion identification. In this regard, the mean, standard deviation, maximum, and minimum of pitch contour and energy in utterances are common aspects. Using pitch-related data, Delbert ET AL. sought to identify four human emotions [9]. They used the Maximum Likelihood Bayes classifier (M LB), Kernel Regression (KR), and K-

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 2, Issue 5, June 2022
Impact Factor: 6.252

nearest Neighbors classifiers (KNN). Roy and Pent land used a Fisher linear classifier to classify emotions [20]. They identified two types of emotions using short spoken sentences: approval and disapproval. They performed multiple trials using characteristics extracted from pitch and energy measurements, with accuracy ranging from 65 to 88 percent. The fundamental drawback of these global-level acoustic properties is that they can't account for dynamic fluctuation within a single phrase. To overcome this, short-term spectral features can be used to trace dynamic fluctuation in emotion in speech in spectral changes at a local sentimental level. A Hidden Markov Model (HMM) was trained to distinguish four emotions using 13 Mel-frequency cepstral coefficients (MFCC) in [14]. To characterize the six archetypal emotions, N we ET AL. used 12 Mel-based speech signal power coefficients to train a Discrete Hidden Markov Model [18]. In both systems, the average accuracy was between 70 and 75 percent. Finally, there have been alternative techniques. Prosody information, as well as the duration of voiced and unvoiced parts, are used as acoustic features in this investigation.

2.2 Emotion Recognition by Facial Expressions

Emotions can be deduced from facial expressions. As a result, numerous ways of classifying human affective states have been developed. Unlike audio-based techniques, which use global statistics of auditory aspects, the features used are often based on local spatial location or displacement of specific points and regions of the face. [19] provides a comprehensive overview of modern emotion identification algorithms based on facial expression. Mase presented an emotion identification system based on the major facial muscle directions [16]. Optical flow was used to extract muscle movements from 11 windows physically placed in the face. The K-nearest neighbour rule was employed for categorization, with an accuracy of 80% for four emotions: happy, rage, disgust, and disgust. They categorised the six main emotions with 88 percent accuracy using a rule-based method. To extract the shape and movements of the mouse, eye, and brows, Black et al. employed parametric models [1]. They also used a similar method to [22] to build a mid- and high-level representation of facial motions that was 89 percent accurate. Tian et al. used permanent and transient facial features such as the lip, nasolabial furrow, and wrinkles to try to recognise Action Units (AU), which were discovered by Ekman and Friesen in 1978 [10]. The shapes and appearances of these features were located using geometrical models. They had a 96 percent accuracy rate. Based on parametric models of separate facial movements, Essa et al. created a system that quantified facial movements. They used an optical flow approach in conjunction with geometric, physical, and motion-based dynamic models to represent the face. They created spatial-temporal templates that may be used to recognise emotions. A recognition accuracy rate of 98 percent was attained without taking into account melancholy that was not included in their work. The extraction of facial features is done using markers in this study. Face detection and tracking algorithms are so unnecessary.

III. METHODOLOGY

Different systems based on facial expression and bimodal information are used to distinguish four emotions: sadness, happiness, rage, and neutral state. The major goal is to assess the performance of unimodal systems, identify their strengths and shortcomings, and compare different ways for fusing these disparate modalities to improve the system's overall recognition rate. The database for the studies was created by recording an actor reading 258 emotional lines. The expressive facial motion data was captured with a VICON motion capture system with three cameras (left of Figure 1) at a sampling frequency of 120Hz. An actress was requested to speak a custom phonemebalanced corpus four while wearing 102 markers on her face (right of Figure 1). The recording was made at a sampling rate of 48 kHz in a quiet room with a close-talking SHURE microphone. The technology caught the markers' movements and aligned audio at the same time. Due to the great precision with which the facial features are collected, this multimodal database is useful for extracting key indications regarding both facial emotions and speech.

Three alternative techniques were used to compare unimodal and multimodal systems, all of which used a support vector machine classifier (SVC) with 2nd order polynomial kernel functions [3]. In a prior work, SVC was employed to recognise emotions and outperformed other statistical classifiers [13][14]. It's worth noting that the only difference between the three algorithms is the features utilised as inputs, therefore it's feasible to draw conclusions about the strengths and limitations of acoustic and facial expression features for recognising human emotions. The database was trained and tested using the leave-one-out cross validation method in all three systems.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 2, Issue 5, June 2022
Impact Factor: 6.252

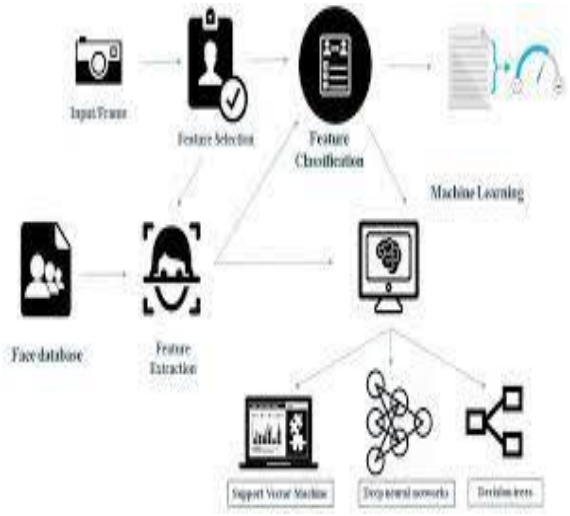

Figure 1: Data recording system

3.1 System Based on Speech

Global-level prosodic variables such as pitch and intensity statistics are the most extensively employed speech signals for audio emotion identification. As a result, using Praat voice processing software [2], the means, standard deviations, ranges, maximum values, minimum values, and medians of pitch and energy were calculated. The ratios of voiced/speech and unvoiced/speech were also calculated. A 11-dimensional feature vector for each syllable was used as input in the audio emotion identification system using the sequential backward features selection technique.

3.2 System Based on Facial Expressions

The spatial data acquired from markers in each frame of the movie is condensed into a 4-dimensional feature vector per sentence in the system based on visual information, as shown in figure 4, and then used as input to the classifier. The face expression system (illustrated in Figure 4) is discussed in the following paragraphs. The motion data is normalised after it has been captured: (1) all markers are translated so that a nose marker is the local coordinate centre of each frame, (2) one frame with a neutral and close-mouth head pose is chosen as the reference frame, (3) three roughly rigid markers (manually chosen and illustrated as blue points in Figure 1) define a local coordinate origin for each frame, and (4). Each data frame is broken down into five sections: the forehead, brow, low eye, right cheek, and left cheek area (see Figure 2). The 3D coordinates of markers in this block are concatenated to generate a data vector for each block. The amount of features per frame is then reduced to a 10-dimensional vector for each area using the Principal Component Analysis (PCA) approach, which covers more than 99 percent of the variation. The markers around the lips are not taken into account since the articulation of the speech could be mistaken for a smile, which would confuse the emotion identification system [19]..

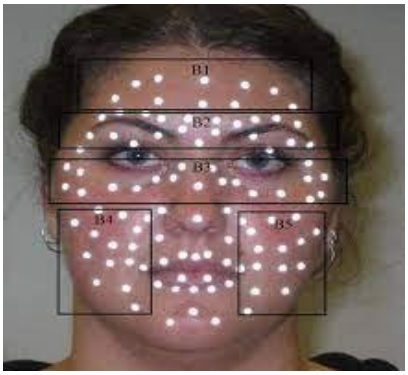

Figure 2: five areas of the face considered in this study

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 2, Issue 5, June 2022
Impact Factor: 6.252

The first two components of the low eye area vector were shown in figure 3 to show how effectively these feature vectors describe the emotion classes. As can be seen, different emotions emerge in separate clusters, hence the spatial position of these 10-dimensional characteristics space can provide vital hints.

Figure 3: First two components of low eye area vector

It's worth noting that each block obtains a 10-dimensional feature vector for each frame. This local data could be fed into dynamic models like HMM. However, for both unimodal systems, we decided to use global features at the utterance level in this research, therefore these feature vectors were preprocessed to create a low dimensional feature vector per utterance. The 10-dimensional characteristics at the frame level were categorised using a K-nearest neighbour classifier (k=3) in each of the 5 blocks, taking advantage of the fact that various emotions appear in different clusters (Figure 3). The number of frames classified for each emotion was then counted, yielding a four-dimensional vector for each block at the utterance level. These feature vectors at the utterance level make advantage of not only For example, they are categorised as joyful when happiness is expressed in more than 90% of the frames, whereas they are classified as sad when sadness is displayed in more than 50% of the frames. This information is used by the SVC classifiers, which improves the system's performance dramatically. Furthermore, because the face expression characteristics and global auditory information are not synced in this approach, they can be simply integrated in a feature-level fusion. As shown in Figure 4, each block has its own SVC classifier, making it feasible to determine which facial area provides better emotion discrimination. Before classification, the 4-dimensional characteristics vectors of the 5 blocks were also added, as seen in figure 4. The integrated facial expressions classifier is the name given to this system.

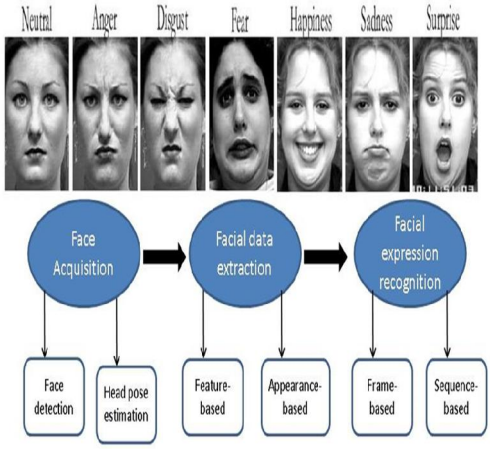

Figure 4: System based on facial expression

3.3 Bimodal System

Two alternative approaches were utilised to merge the facial expression and audio information: feature-level fusion (left of Figure 5), in which single classifiers with features from both modalities are employed; and decision-level fusion, in which each modality's outputs are mixed using specified criteria (right of Figure 5). A sequential backward feature

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 2, Issue 5, June 2022
Impact Factor: 6.252

selection technique was utilised in the first attempt to locate the features from both modalities that maximised the classifier's performance. A total of ten features were chosen. Several criteria were used in the second approach to combine the posterior probabilities of the mono-modal systems at the decisionlevel: maximum, in which the emotion with the greatest posterior probability in both modalities is selected; average, in which the posterior probabilities of each modalities are weighted equally and the maximum is selected; product, in which the posterior probabilities are multiplied and the maximum is selected; and weight, in which different wet weights are used to combine the posterior probabilities of the mono-modal systems at the

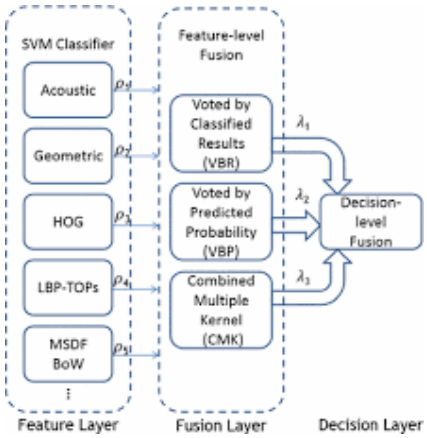

Figure 5: Features-level and decision-level fusion

IV. RESULTS

4.1 Acoustic Emotion Classifier

The confusion matrix of the emotion recognition system based on acoustic input is shown in Table 1, which highlights the system's strengths and drawbacks. This classifier's overall performance was 70.9 percent. Table 1's diagonal components demonstrate that using simply the properties of speech, all emotions can be detected with greater than 64% accuracy. Table 1 demonstrates, however, that some pairs of feelings are frequently confounded. Sadness (22%) is misclassified as a neutral condition, and vice versa (14 percent). Happiness and rage, which are sometimes confused, show the same pattern (19 percent and 21 percent, respectively). These findings support De Silva et alhuman.'s evaluations [7], and can be explained by comparable patterns in auditory parameters of these emotions [23]. Speech associated with anger and happiness, for example, has a longer utterance duration, a shorter inter-word silence, a higher pitch, and a wider range of energy levels. In neutral and sad lines, on the other hand, the energy and pitch are frequently kept at the same level. As a result, categorising these feelings is difficult.

Table 1: Confusion matrix of the emotion recognition system based on audio

	Anger	Sadness	Happiness	Neutral
Anger	0.68	0.05	0.21	0.05
Sadness	0.07	0.64	0.06	0.22
Happiness	0.19	0.04	0.70	0.08
Neutral	0.04	0.14	0.01	0.81

4.2 System Based on Facial Expressions

Table 3 illustrates the performance of facial expression-based emotion recognition systems for each of the five face blocks and the combined facial expression classifier. The cheek areas provide useful information for emotion classification, as seen in this table. It also reveals that the brows, which are commonly employed in facial emotion recognition, perform the worst. The fact that happiness is correctly identified can be explained by figure 3, which demonstrates that happiness is grouped individually in the 10-dimensional PCA spaces, making it easy to identify. Table 2 also shows that the combined facial expression classifier has an accuracy of 85%, which is greater than the majority of the five facial blocks classifiers. Take note that this database was saved.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 2, Issue 5, June 2022
Impact Factor: 6.252

Table 2: Performance of the facial expression classifiers

Area	Overall	**Anger**	Sadness	Happiness	Neutral
Forehead	0.73	0.82	0.66	1.00	0.46
Eyebrow	0.68	0.55	0.67	1.00	0.49
Low Eye	0.81	0.82	0.78	1.00	0.65
Right cheek	0.85	0.87	0.76	1.00	0.79
Left Cheek	0.80	0.84	0.67	1.00	0.67
Combined Classifier	0.85	0.79	0.81	1.00	0.81

The combined facial expression classifier is a feature level integration strategy that fuses the features of the five blocks before classification. These classifiers can also be used at the decision-making level. Table 3 demonstrates the system's performance when the face block classifiers are merged using various criteria. The outcomes are generally comparable. All of these decision-level criteria outperform the combined facial expression classifier by a little margin.

Table 3: Decision-level integration of the 5 facial blocks emotion classifiers

	Overall	Anger	Sadness	Happiness	Neutral
Majority Voting	0.82	0.92	0.72	1.00	0.65
Maximum	0.84	0.87	0.73	1.00	0.75
Averaging combining	0.83	0.89	0.72	1.00	0.70
Product combining	0.84	0.87	0.72	1.00	0.77

The confusion matrix of the combined facial expression classifier is shown in Table 4 to examine the limitations of this emotion recognition system in further depth. This classifier's overall performance was 85.1 percent. This table shows that happiness is detected with a high degree of precision. The other three emotions are roughly identified with an accuracy of 80%. Table 4 also reveals that in the arena of facial expressions, anger is confused with sadness (18%) and neutral state with happy (18%). (15 percent). Because sadness/anger and neutral/happiness can be distinguished with high accuracy in the acoustic domain, the bimodal classifier is projected to function well in the angry and neutral states. This table also demonstrates that grief and neutrality are often mistaken (13 percent). Unfortunately,

Table 4: Confusion matrix of the combined facial expression classifier

	Anger	**Sadness**	**Happiness**	**Neutral**
Anger	0.79	0.18	0.00	0.03
Sadness	0.06	0.81	0.00	0.13
Happiness	0.00	0.00	1.00	0.00
Neutral	0.00	0.04	0.15	0.81

V. DISCUSSION

Because humans employ more than one modality to perceive emotions, automatic multimodal systems are projected to perform better than automatic unimodal systems. The results of this study support this notion, since the bimodal technique improved performance by over 5% (absolute) when compared to the facial emotion recognition system. The findings demonstrate that pairs of emotions that were misclassified in one modality were easily labeled in the other. For example, the facial expression emotion classifier accurately distinguished rage from happiness, which were previously misclassified in the acoustic domain. As a result, when these two modalities were combined at the feature level, these emotions were accurately categorized. Unfortunately, sadness is mistaken for neutral in both areas, resulting in low performance. Although the feature-level and decision-level bimodal classifiers had equal overall performance, an examination of their confusion matrices revealed that the detection rate for each emotion type was vastly different. When compared to the best unimodal identification system, the facial expression classifier, the recognition rate of each emotion rose in the decision level bimodal classifier (except happiness, which decreased in 2 percent). The recognition rate of anger and neutral state increased dramatically in the feature-level bimodal classifier. Happiness, on the other hand, was recognized at a lower percentage of 9%.As a result, the ideal method for fusing the modalities will vary depending on the application. The findings of this study show that, while the system based on audio information performed worse than the

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 2, Issue 5, June 2022
Impact Factor: 6.252

facial expression emotion classifier, its characteristics contain essential information about emotions that cannot be derived from visual data. These findings are consistent with Chen et al[4] .'s observation that audio and facial expression data provide complementing information. On the other hand, it is reasonable to predict that the utilization of either aural or visual characteristics will yield some distinct emotional patterns. When the qualities of one of the emotions are similar, this redundant information is particularly useful for improving the performance of the emotion detection system. Face expressions will be extracted with a significant amount of error if a person wears a beard, moustache, or eyeglasses, for example. In that instance, audio features can be used to compensate for the visual information's limitations. Although the use of facial markers is not appropriate for real-world applications, the study reported in this paper provides vital information about emotion discrimination in various facial blocks. Although the shapes and movements of the brows have long been used to classify facial expressions, the data presented in this research reveal that this face area performs poorer than other facial areas such as the cheeks in emotion discrimination. Because just four affective states were evaluated in this study, it's likely that eyebrows were overlooked. The studies were carried out with a database based on a single female speaker, and the three algorithms were trained to distinguish her facial expressions. It is believed that the system's performance will differ if it is used to detect the emotions of other people. As a result, more data from other people is needed to guarantee that the database adequately represents the variety with which humans express emotions, which is still a work in progress. Another drawback of the method used in this study is that the visual information was obtained through the use of markers. It is not possible to tie these markers to people in real-world applications. As a result, an automatic method should be created to extract facial gestures from video without the use of markers. Using optical flow, which has been successfully used in prior research [11][16], is one method. The next stage in this research will be to develop better algorithms for combining audio-visual data and modelling the dynamics of facial expressions and speech. Acoustic information at the segmental level can be utilised to track emotions at the frame level. Other types of features that define the link between the two modalities in terms of temporal progression may also be beneficial. The relationship between face gestures and pitch and energy contours, for example, could be beneficial in identifying emotions.

VI. CONCLUSION

The merits and drawbacks of facial expression classifiers and audio emotion classifiers were investigated in this study. Some pairings of feelings are frequently misclassified in these unimodal systems. However, the findings of this paper suggest that most of these ambiguities can be resolved by using a different modality. As a result, the bimodal emotion classifier performed better than any of the unimodal systems. The feature-level and decision-level fusion techniques were compared. Both approaches had comparable total results. However, there were substantial differences in the recognition rate for various emotions. Anger and neutral state were accurately recognised by the feature-level bimodal classifier when compared to the best unimodal system, the facial expression classifier. Happiness and sadness were accurately categorised by the decision-level bimodal classifier. The optimal fusion process will thus be determined by the application. The findings of this study reveal that using audio and visual modalities, it is possible to distinguish human affective states with excellent accuracy. As a result, the next generation of human-computer interfaces may be able to detect human feedback and respond correctly and timely to changes in users' affective states, thus increasing the performance and engagement of present interfaces.

REFERENCES

[1]. Tracking and recognising rigid and non-rigid facial motions using a local parametric model of picture motion, M. J. Black and Y. Yacoob. Pages 374–381 in Proceedings of the International Conference on Computer Vision. Cambridge, MA: IEEE Computer Society, 1995.

[2]. P. Boersma and D. Weenink, Praat Speech Processing Software, University of Amsterdam's Institute of Phonetics Sciences. http://www.praat.org

[3]. A tutorial on support vector machines for pattern recognition, Burges, C. 1998, Dat Mining and Knowledge Disc., vol. 2(2), pp. 1–47.

[4]. CMultimodal human emotion / expression recognition, in Proc. of the International Conference on Automatic Face and Gesture Recognition, (Nara, Japan), IEEE Computer Society, April 1998.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 2, Issue 5, June 2022
Impact Factor: 6.252

[5]. Emotional expressions in audiovisual human-computer interaction, Chen, L.S., Huang, T.S. 2000 Multimedia and Expo ICME 2000, Volume 1: 30 July-2 August 2000, IEEE International Conference on. Pages 423-426 in volume 1

[6]. Emotion recognition in human-computer interaction. Cowie, R., Douglas-Cowie, E., Tsapatsoulis, N., Votsis, G., Kollias, S., Fellenz, W., Taylor, J.G. IEEE Signal Processing Magazine, Volume 18, Number 1, January 2001. Pages: 32 – 80

[7]. Facial Emotion Recognition Using Multimodal Information, De Silva, L. C., Miyasato, T., and Nakatsu, R. In Proceedings of the IEEE International Conference on Information, Communications, and Signal Processing (ICICS'97), Singapore, September 1997, pp. 397-401.

[8]. L.C. De Silva and P. C. Ng Recognition of bimodal emotions Face and Gesture Recognition Automatically, 2000. Proceedings. The fourth IEEE International Conference on will be held on March 28-30, 2000. Pages: 332 – 335.

[9]. Recognizing emotion in speech, F. Dellaert, T. Polzin, and A. Waibel. ICSLP 96 Proceedings, Spoken Language, 1996. Volume 3 of the Fourth International Conference on, 3-6 October 1996. 1970-1973 volume 3 pages

[10]. Facial Action Coding System: A Technique for Measuring Facial Movement, by P. Ekman and W. V. Friesen. Palo Alto, California: Consulting Psychologists Press, 1978.

[11]. Coding, analysis, interpretation, and recognition of facial expressions. Essa, Pentland, A. P. JULY 1997, IEEE Transactions on Pattern Analysis and Machine Intelligence, 19(7):757–763.

[12]. T. S. Huang, L. S. Chen, H. Tao, T. Miyasato, R. Nakatsu Man and Machine Recognize Bimodal Emotions April 1998, Proceedings of the ATR Workshop on Virtual Communication Environments (Kyoto, Japan).

[13]. Classifying emotions in human-machine spoken dialogues. Lee, C. M., Narayanan, S.S., and Pieraccini, R. 2002 Multimedia and Expo Proceedings of the ICME '02. Volume 1 of the IEEE International Conference on, August 26-29, 2002. Volume 1, pages 737-740

[14]. Emotion Recognition based on Phoneme Classes. to appear in Proc. ICSLP'04, 2004. Lee, C. M., Yildirim, S., Bulut, M., Kazemzadeh, A., Busso, C., Deng, Z., Lee, S., Narayanan, S.S.

[15]. Lee C. M., Narayanan S.S. Towards detecting emotions in spoken dialogs. IEEE Trans. on Speech & Audio Processing, in press, 2004.

[16]. IEICE Transc., E. 74(10):3474–3483, October 1991. Mase, K. Recognition of facial expression from optical flow.

[17]. Illusions and Issues in Bimodal Speech Perception, D. W. Massaro. Auditory Visual Speech Perception '98 Proceedings (pp. 21-26). Terrigal, December 1998, Sydney, Australia

[18]. Speech-based emotion classification, T. L. Nwe, F. S. Wei, and L. C. De Silva. TENCON, Electrical and Electronic Technology, 2001. Volume 1 of the Proceedings of the IEEE Region 10 International Conference on, 19-22 August 2001. Volume 1, pages 297-301

[19]. Toward an affect-sensitive multimodal human-computer interaction, M. Pantic and L.J.M. Rothkrantz. Volume 91, Issue 9 (September 2003), Proceedings of the IEEE. Page(s): 1370 – 1390.

[20]. Automatic spoken affect classification and analysis, D. Roy and A. Pentland. Proceedings of the Second International Conference on, 14-16 October 1996. Automatic Face and Gesture Recognition, 1996. Pages: 363 – 367