Rule Based System for Recognizing Emotions Using Multimodal Approach

Preeti Khanna
Information System
SBM, SVKM's NMIMS
Mumbai, India

Sasikumar, M.
Director, R & D
Center for Development of Advance Computing, CDAC
Mumbai, India

***Abstract*—Emotion is assuming increasing importance in human computer interaction (HCI), in general, with the growing feeling that emotion is central to human communication and intelligence. Users expect not just functionality as a factor of usability, but experiences, matched to their expectations, emotional states, and interaction goals. Endowing computers with this kind of intelligence for HCI is a complex task. It becomes more complex with the fact that the interaction of humans with their environment (including other humans) is naturally multimodal. In reality, one uses a combination of modalities and they are not treated independently. In an attempt to render HCI more similar to human-human communication and enhance its naturalness, research on multiple modalities of human expressions has seen ongoing progress in the past few years. As compared to unimodal approaches, various problems arise in case of multimodal emotion recognition especially concerning fusion architecture of multimodal information. In this paper we will be proposing a rule based hybrid approach to combine the information from various sources for recognizing the target emotions. The results presented in this paper shows that it is feasible to recognize human affective states with a reasonable accuracy by combining the modalities together using rule based system.**

Keywords—Human Computer Interaction (HCI); Multimodal emotion recognition; Rule based system; Emotional state; Modalities.

I. INTRODUCTION

With the increasing role of computer system in society, HCI has become an integral part of our daily lives. In today's scenario, computers are not only used to perform tasks, but also to learn, communicate, explain, argue, debate, observe and also design. The major concern of HCI now is the need to improve the interactions between humans and computers through justifications and explanations. Thus we observe a significant growth of new forms of 'natural' and 'indirect' interfacing. HCI is experimenting with alternate input mechanisms and multimodal input mechanisms through speech, gesture, posture and facial expression. These help in substituting the largely impersonal devices such as a keyboard and a mouse, for a non-tech savvy.

One of the significant ingredients which could enhance the interaction between human and computer is emotions. Emotions play a vital role in the communication among human beings. However so far, emotions have not played a substantial role in HCI. Incorporating the emotions in HCI is a challenging task. Research studies have been undertaken to investigate and develop various approaches and technology to incorporate emotions in HCI. Some of the recent trends with this respect focus on how a computer can automatically detect the emotional state of a user and then adapt its behaviour accordingly. There is increasing research interest and various applications along these lines.

Some of the prominent areas include e-commerce, help desks, customer support, e-learning, etc. For example, an emotion-aware interface can enhance the sensitivity of an automatic tutor which can adjust the content of the tutorial and speed and style at which it is delivered. As helper/assistant robots (e.g. AIBO) are becoming common toys, often aimed at helping elderly persons in their day to day tasks, the ability to relate to their emotions becomes something of paramount importance. Computer games may adapt playing conditions to the emotional level of the players. Surveillance is another application domain in which the reading of emotions may lead to better performance in predicting the future actions of subjects. In this way, the emotion driven technology can enhance the existing systems for the identification and prevention of terrorist attacks in public places. Certainly not all computes need to pay attention to emotions, or have emotional abilities. Some machines are useful as rigid tools, and it is fine to keep them that way.

The issue of enhancing HCI with emotion raises a number of questions. What are the sources of information that a machine can use to decode the emotional state of the user? What kind of information (emotional cues) are available from these sources? How does one use these sources to estimate the emotional state? What are the emotional states of interest for us from the perspective of enhancing HCI? How to combine multiple modalities? Does the performance of the multimodal emotion recognition for a specific set of target emotions depend on type of fusion model? In what ways can a machine use the knowledge of the user's emotional state to modify its behaviour? There is a plethora of existing work that bears on one or more of these questions.

The paper begins by defining problem domain regarding multimodal emotion recognition. Section II discusses the complete framework of our rule based system. This approach is based on certainty factor i.e. the MYCIN approach. Section III talks about the overall framework of emotion recognition independent of any modalities. Then we explained our approach of rule based multimodal emotion recognition using the case scenarios of 'facial expressions' in section IV. We did few experiments on multimodal data and tested on our rule

based system which is explicitly mentioned in section V. We conclude the paper by summarizing the results and consider some challenges facing the researchers in this area.

The Problem Domain for Multimodal Emotion Recognition

Humans recognize emotion, fusing information from multiple sources: speech signals, facial expressions, gesture, bio-signals and others. Inadequacies of unimodal recognition systems drive the need to go for multimodal recognition. In literature, some attempts like [1], [2], [3] and [4] have considered the integration of information from facial expressions and speech. This paper explores how to combine the information from various sources (e.g. facial expression [24], speech [21, 22, 23] and others [20]) to achieve better recognition of emotional state using rule based approach.

There are two broad approaches to design of a multimodal recognition engine: feature based and decision based. Feature level fusion involves simply merging the features of each modality into a single feature vector. In this method of fusing, all the features are mixed together irrespective of their nature and type. For example, feature can be position of some feature points on the face or the prosodic features of a speech signal. Feature sets can be quite large as we will see later. This high cardinality can result in soaring computational cost for this fusion approach [5]. Decision level fusion is based on the fusion of decisions from each modality where the input coming from each modality is processed independently and these unimodal recognition results are combined at the end [6]. This fusion has advantage of avoiding synchronization issues over the feature level fusion. Decision level fusion ignores possible relationships between features coming from different modalities. Several works [7], [8] and [9] have discussed multimodal fusion; in particular [10] discusses many issues and techniques of multimodal fusion.

Finding an optimal fusion type for a particular combination of modalities is not straightforward. A good initial guess can be based on the knowledge of the interaction and synchronization of those modes in a natural environment. Hybrid fusion attempts to combine the benefits of both feature level and decision level fusion methods, may be a good choice for some multimodal fusion problems. However, based on existing knowledge and methods, how to model multimodal fusion for target set of emotions is still an open problem. We propose a hybrid approach for multimodal fusion. This model is based on modeling each modality through a set of rules. In this process of formulation of rules, feature analysis plays a very important role. These rule sets were tested and listed later in our running example of facial expression.

II. Rule Based System: Base for Our Hybrid Model

A rule based system consists of *if-then rules*, a bunch of *facts*, and an *interpreter* controlling the application of the rules. A simple *if-then* rule has the form 'if *x* is *A*, then *y* is *B*'. The if-part of the rule, '*x* is *A*', is called the *antecedent* or *premise*, while the then-part of the rule, '*y* is *B*', is called the *consequent* or *conclusion*. When the premise is known to hold in a scenario, the conclusion can be drawn. This is the normal interpretation of a rule. One of the major strength of rule based representation is its ability to represent various uncertainties. Uncertainty is inherently part of most human decision making. This uncertainty could arise from various sources like incomplete data or domain knowledge used being unreliable. So the if – then rules is often represented like 'If A, B, C ----> then D, with certainty X', where X represents the degree of belief or confidence in the rule [11].

A. Approaches for Handling Uncertainty

To handle these uncertainties, there are two broad approaches - those representing uncertainty using numerical quantities and those using symbolic methods. In numerical approaches, one models the uncertainty by numbers and provides some algebraic formulae to propagate these uncertainty values to the conclusions. These approaches are useful for handling the issues related to "unreliable or inaccurate knowledge". For example, Bayesian reasoning [12], Evidence theory [13] and Fuzzy set approaches [14] are numerical models. On the other hand, symbolic characterization of uncertainty is mostly aimed at handling incomplete information, e.g., Assumption Based Reasoning [15], Default Reasoning [16] and Non-monotonic Logic [17]. For example, if there is not enough information available, the system makes assumptions that can be corrected later, when more information is received.

In our domain, the basic problem is that there are hardly any features or feature combinations which can infer any emotion to complete certainty. Therefore, we concentrate on numerical approaches for handling the uncertainty. We have adopted the '*Confirmation Theory'* as used in MYCIN approach [12] to deal with uncertainty in our domain. This approach works well with rule based representation of domain knowledge.

B. Reasoning with Certainty Factors (CF): The MYCIN Approach

Shortliffe and Buchanan [12] developed the Certainty Factor (CF) model in the mid-1970s for MYCIN, an expert system for the diagnosis and treatment of infections of the blood. Since then, the CF model has been widely adopted for uncertainty management in many rule based systems. Each rule is assigned CF by domain experts. This is meant to represent the uncertainty of the rule. Higher CF indicates that the conclusion can be asserted with higher confidence when the conditions are true. Similarly every fact in the domain is also given CF indications how confident one is in that. Reference [12] intended a CF to represent the change in belief in a hypothesis given some evidence. In particular, a CF between 0 and 1 means that the person's belief in *h* given *e* increases, whereas a CF between -1 and 0 means that the belief decreases. A value of +1.0 indicates absolute belief and -1.0 indicates absolute disbelief. The method generally used to propagate the measure of uncertainty in the antecedents and the uncertainty attached to the rule to the conclusions being derived is briefly explained below. This propagation is done in two steps [11].

- The different antecedents in the rule, in general, have different values of uncertainty attached to them. Some formula is required to combine these measures and provide a consolidated uncertainty number. This option

considers the strength of the weakest link in a chain as the strength of the chain. This is defined as:

CFantecedents = {minimum of CFs of all antecedents} (1)

- Then this measure (uncertainty for the set of antecedents) is combined with the measure of uncertainty attached to the rule to give a measure of uncertainty for the conclusion of the rule.

*CF of the conclusion from rule = {CF associated with rule R1} * {CFantecedents}, provided CFantecedents >= threshold* (2)

It can be seen that the CF obtained for a conclusion from a particular rule will always be less than or equal to the CF of the rule. This is consistent with the interpretation of the CF used by MYCIN, that is, the CF of a rule is the CF to be associated with the conclusion if all the antecedents are known to be true with full certainty. In a typical rule based system, there may be more than one rule in the rule base that is applicable for deriving a specific conclusion. Some of them will not contribute any belief to the conclusion, because CF of antecedents has a CF less than the threshold. The contributions from all the other rules for the same conclusion have to be combined. For MYCIN model, initially CF of a conclusion is taken to be 0.0 (that is, there is no evidence in favour or against) and then as different rules for the conclusion fires, the CF gets updated. MYCIN uses a method that incrementally updates the CF of the conclusion as more evidence for and against is obtained. Let CFold be the CF of the conclusion so far, say, after rules R1, R2,...Rm have been fired. Let CF*in* be the CF obtained from firing of another rule Rn. The new CF of the conclusion (from rules R1, R2……...Rm and Rn), CFnew, is obtained using the formulae given below.

*CFnew = CFold + CFin * (1 - CFold)*
when (CFold, CFin >0) (3)

*CFnew = CFold + CFin * (1 + CFold)*
when (CFold, CFin < 0) (4)

CFnew = (CFold + CFin) / (1 – min (|CFold|, |CFin|))
otherwise (5)

We adopt this calculus in our model and explained later with a running example in section IV. Before that we first discuss the overall framework of emotion recognition system.

III. Framework For Emotion Recognition

The overall conceptual framework for emotion recognition includes pre processing, feature extraction, feature analysis and selection of the features, formulation of rules and measuring performance to classify the target emotional states. We will explain each of these in brief as below. We will use facial expressions as the running example to illustrate these stages, etc. The framework remains same across all modalities [20, 21, 22].

A. Pre Processing and Feature Extraction

The first step is pre processing. The objective of this step is to make the input data in a standard format and suitable for extracting the desired features. Usual preprocessing steps include size normalization of the frontal image, noise removal from speech signal etc.

The next step is feature extraction. We need to identify useful features from each of these input sources (pre-processed input data – image, audio and others). For example, location of feature points such as eyes, eye corners, eyebrows, eyebrow corner, mouth corners, upper and lower lip, nose and nostrils, etc. are important for facial expression analysis. The work in this step involves identifying relevant features and formulating algorithms to extract these features from their respective input data.

C. Feature Analysis and Selection

Once the basic feature set is ready, the next step is analysis of each of these features. The question, 'how does each of the features vary with the emotion' needs to be answered here. Usually every feature doesn't contribute to the same extent to recognize different emotional states. Thus feature analysis and selection is an important step. The case of facial expression mentioned in this paper illustrates feature analysis and selection process in detail later.

D. Formulation of Rules

If-then rules are one of the most common forms of knowledge representation used in various domains. Systems employing such rules as the major representation paradigm are called rule based systems.

To design the rules for classifying emotions, all the relevant features needs to be studied in more detail to see its ability to distinguish between different target emotional states. Influential and useful features can be used to define rules. This approach remains broadly same across different modalities and is as follows:

1) Feature analysis has been done for each feature to see its ability to distinguish among the target emotional states, and accordingly useful features were shortlisted.

2) Rules are formed using each of these features for different target emotional states. A feature may yield one or more rules. Generally these rules have the form: if feature F1 has value less than or greater than T1 and feature F1 has value less than or greater than T2 then conclude emotion = e1. For each rule, the cut- off points T1 and T2 for a given emotion class is taken to be the approximate average of the value of that class with its immediate neighbor emotional class.

3) *Corresponding to each rule, we associate CF values for each emotional class. These values of CFs are decided on the conditions mentioned in Table 1.*

TABLE I. DEFINING CERTAINTY FACTOR (CF) FOR RULES

Range of the CF	*CF Values*	*Belief and Disbelief*	*Indicated by*
Greater than 0.2 and up to 0.4	0.3	High evidence	High Inter class distance
Greater than 0.1 and up to 0.2	0.2	Moderate evidence	Medium Inter class distance
Equal to 0.1	0.1	Low evidence	Low Inter class distance

This heuristic has been arrived at based on empirical studies of the various feature graphs and behavior of the CF calculus. There may be multiple rules associated with each feature. Multiple rules, when they fire simultaneously (based on values of different features) may saturate the values of CF associated with them. To minimize this possibility, we have chosen relatively lower range of CF values. Given our observation that most features do not provide a high degree of discrimination for any of the emotions, a high value did not appear justified for any individual feature. The chosen range also allows the CF value to climb steadily to a high range, when there are many features supporting an emotion. The rules may point to a specific emotional state or a set of emotional states. If the distance of an emotion with its neighboring emotion is found to be less than 5– 6% of the entire spread (overall range) for that features value, then these emotions are grouped as a subset. Allocation of the values of CF to these classes is done based on the following rules, derived based on analysis of the emotion profile.

High Interclass Distance: If the interclass distance of an emotional class (either singleton or non-singleton) with its neighbors (left side and right side) is more than 15% of the entire spread for that feature, then the chances of a confusion with the neighboring class is low and hence the CF value associated with this class for that feature is 0.3.

Medium Interclass Distance: If the interclass distance of a emotional class (either singleton or non-singleton) with its neighbors (left side and right side) is in between 6-15% of the entire spread for that feature, then the CF value associated with this class is 0.2.

Low Interclass Distance: If the interclass distance of a emotional class (either singleton or non-singleton) with its neighbors (left side and right side) is less than 6% of the entire spread for that feature, then the CF value associated with this class is 0.1.

The exercise is done for the modalities like facial expression and speech. The next section will discuss in detail one of the case scenario for facial expression beginning with databases to the rules formulation.

IV. CASE STUDY FOR FACIAL EXPRESSION

We illustrate the process with a concrete example of emotion recognition from facial expressions. We used standard database, Cohn-Kanade (CK) [18] of the static images, where individuals are constrained to look straight at the camera and they are photographed with single colored background and illumination conditions do not vary drastically. Therefore, preprocessing issues are not a concern here. We utilize 184 images from 57 subjects. We have 32 female and 25 male subjects for the emotional states of neutral, anger, happy, fear, sad and disgust.

A. Feature Extraction

We have used the geometric features for emotion recognition and defined the model as a point-based model. The frontal view face model [19] is composed of many elements like mouth, nose, eyes and brows that could be used for analysis (Fig. 1 and Table 2). We used a set of 18 points in the frontal view image and using these points we defined a total of 21 features (f3, f4, f5, f6., f7, f8, f9, f10, f11, f12, f13, f14, f15, f16, f17, f19, f20, f21, f22, f23, f24 as shown in Fig. 1, mostly in the form of inter-point distances. For example, the feature f3 is the distance between left eye outer corner, A to left eyebrow outer corner, E.

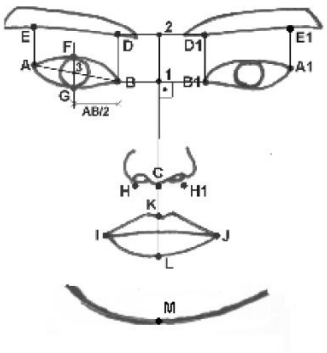

Fig.1. Facial points of frontal-view [19]

Similarly feature f4 (symmetrical to f3) is the distance between right eye outer corners, A1 to right eyebrow outer corner, E1. Each of these points has been extracted from the image. All the distances were computed. For example, mouth width is the distance between the tips of the lip corner. Similarly lip thickness, distance between left eye to left eyebrow, etc., were computed.

The distances are compiled into an output file (.xls) that is used for further analysis. All these distances were obtained for different emotions including the neutral state for all subjects. Facial expressions are often characterized by variation of a feature from its value in the neutral state, rather than its absolute value in a given state. Therefore, we standardize these features w.r.t their neutral value. These parameters were normalized in the following manner:

$$\textit{Normalized Value} = (\textit{Measured Value} - \textit{Neutral State Value}) / \textit{Neutral State Value} \quad (6)$$

B. Feature Analysis

As discussed earlier all features might not be useful in forming the rules. Individually each of these has to be analyzed. For example, the feature, lip distance (horizontal distance- f16 and vertical distance- f17) could be seen as varying with emotions (Fig. 2 and Fig. 3).

TABLE II. FEATURES OF THE FACIAL POINTS OF THE FRONTAL VIEW [19]

Features	Feature Description
f3	Distance AE
f4	Distance A1E1
f5	Distance 3F, 3 is the centre of AB (See Fig. 1)
f6	Distance 4F1, 4 is the centre of A1B1 (See Fig. 1)
f7	Distance 3G
f8	Distance 4G1
f9	Distance FG
f10	Distance F1G1
f11	Distance CK, C is 0.5HH1 (f0)
f12	Distance IB
f13	Distance JB1
f14	Distance CI
f15	Distance CJ
f16	Distance IJ
f17	Distance KL
f19	Image intensity in circle (r(0.5BB1), C(2)) above line (D, D1)
f20	Image intensity in circle (r(0.5BB1), C(2)) below line (D, D1)
f21	Image intensity in circle (r(0.5AB), C(A)) left from line (A, E)
f22	Image intensity in circle (r(0.5A1B1), C(A1)) right from line (A1, E1)
f23	Image intensity in the left half of the circle (r(0.5BB1), C(I))
f24	Image intensity in the right half of the circle (r(0.5BB1), C(J))
Total	21 Features

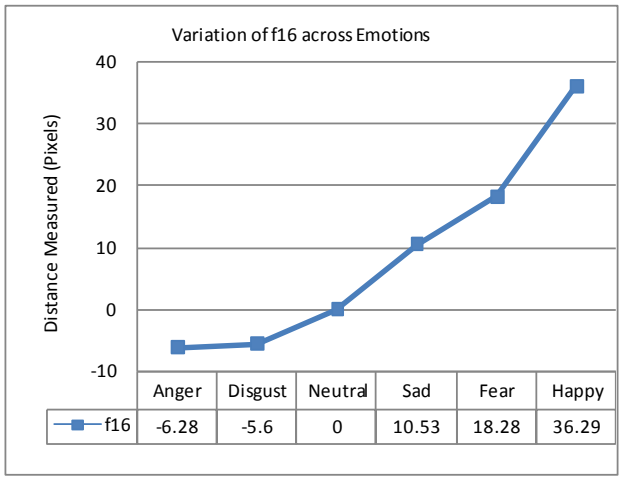

Fig.2. Variation of feature f16 (horizontal lip distance) across emotions

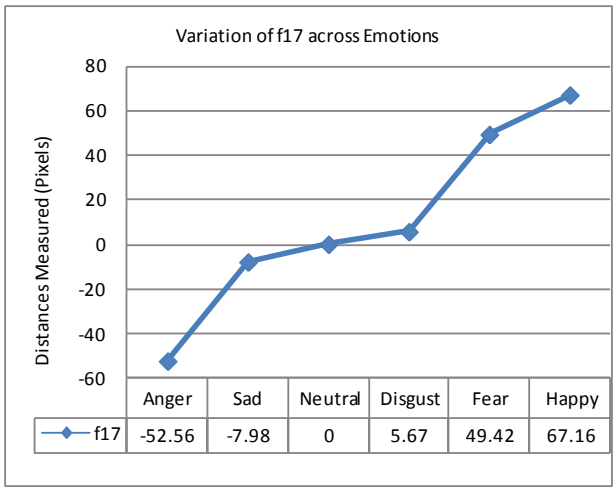

Fig.3. Variation of feature f17 (vertical lip distance) across emotions

We did these analyses using the individual features (f3 to f24) to see how each of these is varying across emotion. We found that eleven features (i.e. f3, f4, f9, f10, f11, f12, f13, f14, f15, f16, and f17) show more significance variation across the considered emotional states among all twenty one features. Also it is found that all the symmetrical pairs of features (like left eye vertical distance, f9 and right eye vertical distance, f10) do not always contribute to the same level to distinguish between the same set of emotions. The lip movement (horizontal lip distance, f16 and vertical lip distance, f17) provides good separation between 'happy', 'sad', 'fear' and 'neutral' emotions, but doesn't differentiate between 'anger' and 'disgust'. To validate these separations between emotional states, rules structure will be formed.

C. Formulation of Rules

As discussed earlier all features might not be useful in forming the rules. Individually each of these has to be analyzed. For example, the feature, lip distance (horizontal distance- f16 and vertical distance- f17) could be seen as varying with emotions (Fig. 2 and Fig. 3). From the trend of feature f16 (Fig. 2), it is seen that the emotions 'neutral', 'sad', 'fear' and 'happy' are distinguishable individually, whereas the emotions, 'disgust' and 'anger' are found to be close together (as the distances with its neighbour are found to be in the range of 5- 6% of the entire spread). Depending on the interclass distances of these classes CFs have been allocated (Table 1) and rules have been formed. For each rule (of the type if – then), the cut off point (i.e., upper limit and lower limit) belonging to the emotion class is taken to be the average of the value of that class with its immediate emotional class. For example, for 'sad' emotion the cut off points to be considered are 5 and 14, forming the singleton class and due to high inter class distances the CF values is to be considered as 0.3 (see Table 3). Similarly, the feature f17 also varies across emotions (Fig. 3). It is observed that 'neutral' along with 'disgust' is forming a non-singleton class while rest of the emotions is acting as singleton classes. Depending on distances between these classes, CFs has been allocated and rules have been formed. We found a total of five conditions each for the feature f16 and feature f17 to classify emotions. Examples of rules (Rule 1 and Rule 2) are shown below.

Example Rule 1: *Using dist_horizontal_lip (F16) for emotion identification*

(i) if (dist_horizontal_lip <= -3)
CFDis=0.2; CFAng=0.2;
(ii) if ((dist_horizontal_lip > -3) && (dist_horizontal_lip <= 5))
CFNeu=0.3;
(iii) if ((dist_horizontal_lip > 5) && (dist_horizontal_lip <= 14))
CFSad=0.3;
(iv) if ((dist_horizontal_lip > 14) && (dist_horizontal_lip <= 27))
CFFear=0.3;
(v) if (dist_horizontal_lip > 27)
CFHap=0.3;

Example Rule 2: *Using dist_vertical_lip (F17) for emotion identification*

(i) if (dist_vertical_lip < -30)
CFAng=0.3;
(ii) if ((dist_vertical_lip < -3) && (dist_vertical_lip >= -30))
CFSad=0.2;
(iii) if ((dist_vertical_lip < 27) && (dist_vertical_lip > -3))
CFNeu=0.3; CFDis=0.3;
(iv) if ((dist_vertical_lip >= 27) && (dist_vertical_lip < 58))
CFFear=0.3;
(v) if (dist_vertical_lip >= 58)
CFHap=0.3;

Such kind of exercise is done for each of the selected features extracted from the face. These features are f3, f4, f9, f10, f11, f12, f13, f14, f15, f16 and f17. Symmetrical pair of features like (f3, f4), (f9, f10), (f12, f13) and (f14 and f15) do not vary in the same way across different emotions and hence the resulting rules may differ. In the formulation of rules, we considered each of these features individually. Total of 11 rules have been formed for emotion identification using facial static images.

D. Recognizing Emotions from Facial Expressions using Rules

These rules have been tested on the database (CK database of facial expression) and final value of CF has been computed corresponding to each of the 6 emotional states. The emotion with the highest value of final CF is considered and counted against the expected emotion class for each image for all the subjects. For example, Table 3 shows the computed values of CF labelled as CF_SAD, CF_NEU, CF_ANG, CF_HAPPY, CF_FEAR and CF_DISGUST corresponding to all the six emotions - sad (S), neutral (N), anger (A), happy (H), fear (F) and disgust (D).

TABLE III. EXAMPLES OF COMPUTED VALUES OF CF USING RULES FROM FACE FOR FEMALE SUBJECT

Updated Value of CF computed using rules for respective emotion							
Subjects	**Actual Emotion**	**CF_ Sad**	**CF_ Neu**	**CF_ Ang**	**CF_ Happy**	**CF_ Fear**	**CF_ Disgust**
s1	S	**0.83**	0.56	0.30	0.36	0.72	0.00
s1	N	0.00	**0.97**	0.00	0.00	0.30	0.00
s1	A	0.10	0.37	**0.91**	0.20	0.50	0.51
s1	H	0.30	0.30	0.30	**0.82**	0.36	0.30
s1	F	0.78	0.37	0.00	0.20	**0.84**	0.00
s1	D	0.00	0.00	0.85	0.51	0.20	**0.87**

A row in this table indicates an input image of an individual subject in a particular emotional state (subjects labelled as 1). Each subject has been tested across emotions. Final outcome for the same is indicated in these CF values under the six columns labelled from CF_SAD to CF_DISGUST. For example, row 3 corresponds to subject-1 in 'angry' state; the table shows the maximum value of CF under the emotion class of 'anger' (0.91) showing correct identification. Similarly, the maximum value of CF for the subject-1 (row 6) is 0.87 and is for the target emotion of disgust. Though the value belonging to 'anger' is coming close to this value, we are considering the highest value of CF to identify the target emotion associated with the input image. Hence, the computed emotion matches with the 'predicted emotion' which is 'disgust' in this case and 'anger' in the previous case. Similarly computed value of CF has been analyzed for each of the emotions. The overall correctness of recognizing emotions using rule based approach in a unimodal system from facial expression is found to be 86.43% (i.e., out of 184 images, 159 images are correctly recognized). The recognition rates are found to be 80% (93 images correctly recognized out of 112) and 88.89% (64 images correctly recognized out of 72) for female and male subjects respectively.

V. WORKING WITH OUR OWN MULTIMODAL DATA

As seen from literature humans recognize emotion by fusing information from multiple sources: speech signals, facial expressions, gesture, bio-signals and others. Inadequacy of unimodal recognition systems provides the basis to go for multimodal recognition. We extend emotion recognition for multimodal data based on our rule based model. This model is based on preparing a set of rules derived from the individual modalities. The rules are mixed together independent of the modality into a single group. In order to test, we propose to include the data source as facial expressions with speech. The database used in the experiments consists of audio samples and static frontal images of different people (graduate students in the age group of 21 to 28 years). Total of 11 subjects participated in our experiment (5 female and 6 male). Each of these subjects was told to read a single sentence under four emotional states (anger (A), happy (H), sad (S) and neutral (N)). For the process of inducing the desired emotional state, individual subjects were shown a small video clipping of 2-3 minute corresponding to each of the four emotional categories. During this, facial expression was captured by the digital camera. The subjects chosen in our experiment don't wear 'glasses' and males don't have 'beard' on their face – this made the analysis easier. We have total of 20 images with utterances (5 each of 'anger', 'happy', 'sad', and 'neutral') of female and 23 images (6 each of 'anger', 'sad' and 'neutral' but 5 is of 'happy') with utterance of male subjects. The compiled set of rules for speech and facial expression was run against this dataset. We now discuss the results obtained, and compare with the performance of the same when using facial expression and speech alone.

a) Results using Facial Expression: Unimodal Approach

The average emotion recognition rate of the system using our own database is found to be 65% (for female subjects), 65.21% (for male subjects) and 67.44% overall. The emotion 'sad' is the best recognized and has 82% recognition rate overall. But this is not true with male subjects. 'Anger' is hard to distinguish from others and hence having the least accuracy.

b) Results using Speech: Unimodal Approach

The average emotion recognition rate of the system was found to be 55% (for female subjects), 62.5% (for male subjects) and 56.6% overall. It has been observed that the emotion 'happy' is hard to recognize both in female as well as in male subjects. The emotion 'sad' shows reasonably good recognition rate for male as well as female subjects.

c) Results when combining Facial Expression and Speech: Bimodal Approach

The average emotion recognition rate of the bimodal emotion recognition system (adding the two sets of rules together) using rules is found to be 75% (for female subjects), 65.21% (for male subjects) and 67.44% (overall). It has been observed that overall performance has increased by combining the inputs from speech signal and facial image in case of gender independent as well as gender dependent scenario.

VI. CONCLUSION

We presented a rule based approach for multimodal emotion recognition, which provides an elegant method for the design of multimodal recognition of emotion. We have formulated a multimodal recognition framework built around if-then rules using certainty factors to capture uncertainty of individual features. Multimodal emotion recognition performs better than unimodal emotion recognition system. Emotions such as 'anger' and 'sad', which was hard to recognize with facial expression yields better result when combined with speech modality. To the best of our knowledge, this approach has not been tried in the literature. This technique appears to be simple and effective for this problem. There are a number of avenues for extending this work. A more realistic evaluation with large data and more modalities is, perhaps, the most important. At present, we have used the Confirmation theory as used in MYCIN approach [12]. One of the major concerns against the use of certainty factor is that they have no sound theoretical basis; though, they often work well in practice. We allocated the values of CF to the emotional classes based on heuristic rules as defined in section III. These have been derived based on the analysis of the individual features across different emotions. In this work, we have ignored the possibility of having more than one emotional state at a time. We also would like to investigate alternative uncertainty models like the Dempster-Shafer Theory. Demspter Shafer theory provides more flexibility in assigning belief to various subsets of emotions.

The databases used for the expression analysis are all based on subjects who "performed" a series of different expressions. There is a significant difference between expressions of a spontaneous and of a deliberate nature. Without a database of spontaneous expressions, the expression analysis system cannot be robust enough. This database issue is common for all the modalities. The multimodal data fusion for emotion recognition remains an open challenge as several problems still persist, related to finding optimal features, integration and recognition. Completely automated multimodal emotion recognition system is still at the preliminary phase, shows very limited performance and is mostly restricted to the lab environment.

REFERENCES

[1] L.S. Chen, T.S. Huang, T. Miyasato, and R. Nakatsu "Multimodal Emotion/Expression Recognition", in Proceedings of the 3rd International Conference on Face and Gesture Recognition, pp.366-371, 1998.

[2] De Silva and Ng, "Bimodal Emotion Recognition", Automatic Face and Gesture Recognition, in IEEE International Conference, pp. 332 – 335, 2000.

[3] N. Sebe, I. Cohen, T. Gevers, and T.S. Huang, "Emotion Recognition Based On Joint Visual and Audio Cues", Pattern Recognition, International Conference on, vol. 1, pp. 1136–1139, 2006.

[4] Z. Zeng, Jilin Tu, Liu, Huang, Pianfetti, Roth and Levinson, "Audio-Visual Affect Recognition", IEEE Transactions on multimedia, 9 (2), pp. 424-428, 2007.

[5] B.V. Dasarathy, "Sensor Fusion Potential Exploitation Innovative Architectures and Illustrative Approaches," in Proceeding of IEEE vol. 85, pp. 24–38, 1997.

[6] C. Busso, Z. Deng, and S. Yildirim, "Analysis of Emotion Recognition using Facial Expressions, Speech and Multimodal Information", in

Proceedings of ACM 6th International Conference on Multimodal Interfaces, pp. 205-211, 2004.

[7] Corradini, A., Mehta, M., Bernsen, N. and J.-C. Martin. Multimodal input fusion in human-computer interaction on the example of the on-going nice project. In Proceedings of the NATO-ASI conference on Data Fusion for Situation Monitoring, Incident Detection, Alert and Response Management, Yerevan (Armenia), August 2003

[8] H. Liao. "Multimodal Fusion", Master's thesis, University of Cambridge, July 2002.

[9] S. Kettebekov and R. Sharma, "Understanding Gestures In Multimodal Human Computer Interaction", International Journal on Artificial Intelligence Tools, 9(2), pp. 205-223, 2000.

[10] R. Sharma, V. Pavlovic, and T. Huang. "Toward Multimodal Human Computer Interface", In Proceedings of the IEEE, 86(5), pp. 853-860, 1998.

[11] M. Sasikumar, S. Ramani, S.M. Raman, K.S.R. Anjaneyulu, and R. Chandrasekar, "Rule Based Expert Systems – A Practical Introduction", Narosa Publishers, 2007.

[12] E.H. Shortliffe. and B.G. Buchanan, "A Model of Inexact Reasoning in Medicine", Mathematical Biosciences, vol. 23, pp. 351-379, 1975.

[13] J. Gordon and E.H. Shortliffe, "The Dempster-Shafer Theory of Evidence", in [Buchanan and Shortliffe, 1984] pp. 272-292, 1984.

[14] C.V. Negoita, "Expert Systems and Fuzzy Systems", Benjamin/Cummings, 1985.

[15] J.A. Doyle, "Truth Maintenance System", Artificial Intelligence, vol. 12, pp. 231-272, 1979.

[16] R. Reiter, "A Logic for Default Reasoning", Artificial Intelligence, vol. 13, pp. 81-132, 1980.

[17] D. McDermott and J. Doyle, "Non-monotonic Logic I", Artificial Intelligence, vol. 13, pp. 41-72, 1980.

[18] T. Kanade, J. Cohn, and Y. Tian, "Comprehensive Database for Facial Expression Analysis", in Proceedings of the International Conference on Automatic Face and Gesture Recognition, pp. 46-53, 2000.

[19] M. Pantic and L.J.M. Rothkrantz, "Automatic Analysis Of Facial Expressions: The State Of The Art", IEEE Trans. Pattern Analysis and Machine Intelligence, 28(12), pp. 2037-2041, 2000.

[20] Preeti Khanna and M. Sasikumar, "Recognizing Emotions from Keyboard Stroke Pattern", International Journal of Computer Applications, 11(9): December, 2010.

[21] Preeti Khanna and M. Sasikumar,"Recognizing Emotions from Human Speech", Think Quest 2010, International Conference on "Contours of Computing Technology in association with Springer Publications, March 2010.

[22] Preeti Khanna and M. Sasikumar, "Application of Vector Quantization in Emotion Recognition from Human Speech", Springer Series in Communications in Computer and Information Science (CCIS), ICISTM – pp. 118-125, 2011.

[23] 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, Pt. 2, Springer-Verlag, pp. 71–99, 2011.

[24] S. V. Ioannou, A. T. Raouzaiou, V. A. Tzouvaras, T. P. Mailis, K. C.Karpouzis, and S. D. Kollias, "Emotion recognition through facial expression analysis based on a neurofuzzy network," in Neural Networks, vol. 18 (4), Elsevier, pp. 423–435, 2005.