An HMM based Model for Prediction of Emotional Composition of a Facial Expression using both Significant and Insignificant Action Units and Associated Gender Differences

Suvashis Das
Department of Management and Information Systems Science
1603-1 Kamitomioka, Nagaoka
Niigata, Japan

Koichi Yamada
Department of Management and Information Systems Science
1603-1 Kamitomioka, Nagaoka
Niigata, Japan

ABSTRACT
The problem of emotion prediction from the face is twofold. First, it requires that the facial Action Units (AUs)[1] and their intensities are identified and second interpreting the recorded AUs and their intensities as emotions. This work focuses on developing an accurate model to predict emotions from Facial Action Coding System(FACS) coded facial image data based on a Hidden Markov Model (HMM)approach. The novelty of this work is: 1) A new and more accurate model for emotion prediction from AU data is proposed by assigning a set of N HMMs to every AU where N is the number of emotions we consider while conventional studies have assigned at most one HMM per AU or lesser like 6 emotion specific HMMs for the entire set of AUs [3-6]. Assigning N HMMs per AU takes away the errors that might creep in due to non-consideration of the insignificant or non-present AUs by calculating separately the probability contributions towards each emotion by every single AU in the entire AU set which is used later to calculate the mean probability for each emotion considering all AUs together. 2) A percentage score of each emotion that composed the face of a subject is predicted rather than to just identify the lead or prominent emotion from the maximum probability considerations as exhibited my majority of similar researches. 3) Discuss the gender differences in the depiction of emotion by the face.

General Terms
Human Computer Interaction, Psychology, Emotions, Gender Stereotypes, Facial Expressions.

Keywords
FACS, Action Units, Hidden Markov Model, Plutchik's Wheel of Emotions, Baum-Welch Algorithm, Forward-Backward Procedure, CK+ Database.

1. INTRODUCTION
Charles Darwin in his book *The Expression of the Emotions in Man and Animals* [7] wrote about the face being a representation of inner physiological reactions. Plutchik [8] gave the wheel of emotions which associates many emotions as opposites and adjacent emotions which combine to render advanced non-basic emotions. The wheel of emotions is shown in Figure 1. According to Plutchik [8] there are eight basic emotions which are universal and innate but according to P.Ekman and W.V. Friesen [1, 2] there are seven, in fact psychology researchers have put forward varied ways to represent emotions but research by P. Ekman and W.V. Friesen have been quite generalized and formulated with lot of experimentation towards the formulation. It is also evident from the analysis of images of subjects showing contempt in many facial expression database like the Extended Cohn Kanade Database (CK+) [9,10], that without the depiction of anger and/or disgust on the face an expression can be generated by facial muscles to represent contempt.

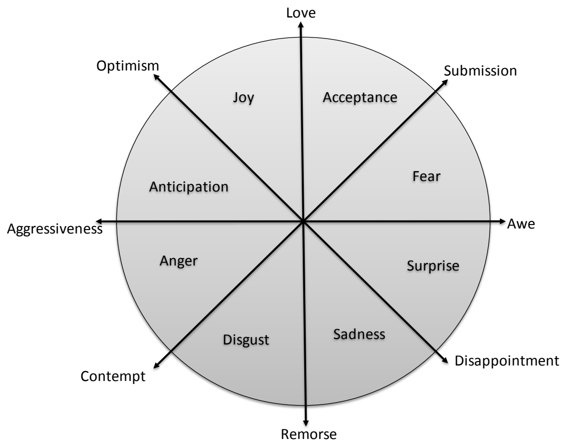

Figure 1: Plutchik's wheel of emotions

Furthermore, Matsumoto [11] and Ekman and Heider [12] have presented more evidence concluding that contempt is a universal and a basic emotion. In this paper we will assume the basic emotion set to be of anger, contempt, disgust, fear, happiness, sadness and surprise. Some of the leading ways of observing human emotion are by speech [13], facial actions [3-6] and biomedical means [14]. Our research uses the face method to detect emotion because voluntarily or involuntarily emotions are very well depicted on the human face [7]. In the process of detecting emotions from the face there are many techniques that have already been applied and proposed with varying success rates. But the problem lies with the fact that almost all researches have concentrated on identifying the significant visible changes as compared to the neutral face. It is to be noted that human face represents emotions using the entire face [15]. What we think to be significant and the

[1]Action units (AUs) represent the facial muscle movements that bring about changes to facial expressionsas defined byP.Ekman and W.V.Friesen in*Facial Action Coding System* [1, 2].

driving factor behind a particular emotion is in reality contributed also by the lack of changes in the non-significant areas of the face as compared to the neutral face. Let's say that the lip-tightener and lip-pressor muscles are both active then we can say that the subject is angry(see Figure 2).In the figure there is significant lip muscle movement and moderate eyebrow gatherer muscle movement in the right image(anger) as compared to the left image(neutral).In this case other muscle movements are insignificant. But if there are minute movements of other muscles the resulting emotion maybe different or a simultaneous non-prominent emotion might be displayed.

Figure2: (left) neutral face, (right) angry face[2]

We assume that the significant muscle movements, the insignificant muscle movements or even the muscles with no movement accounts for the representation of an emotion. Our proposal devises a way to take into consideration the insignificant muscle movements and muscle with no movements into consideration as well by assigning a set of seven HMMs for each muscle unit. To enable the study of emotions a systematic and formulated method needs to exist to decode facial muscle movements to meaningful inference. A lot of painstaking research has already been done in order to find the best possible way to decode information from the face [1, 2, 16-22]. Key among them is the Facial Action Coding System developed by P. Ekman and W. V. Friesen in 1978 [1, 2]. The facial muscular actions that render facial expressions were organized by Ekman and Friesen into Action Units. They formulated around 64 AUs whose different combinations represented the major set of atomic and complex emotions on the human face. Also some set of rules to decipher emotions are given by Cohn et al. [6] in the Extended Cohn Kanade Dataset. These set of rules may formulate a rule-based system to decode emotion from the face but it will suffer inconsistencies as these rules do not incorporate the idea of insignificant AUs. Also a more robust method needs to be found to avoid errors in emotions like contempt that can be easily confused with disgust or lower levels of anger. Cohen et al. [4] in year 2000 used multilevel HMMs to identify the six basic emotions (Surprise, Sadness, Fear, Happiness, Anger and Disgust) having a common observation vector consisting of multiple AU sequences in a single stream. The research did not consider the contributions of the non-present AUs. It was also lacking in the dealing of combination of emotions or the composition of the face in terms of emotions. This left it a step away from effective emotional interpretation. Later in the same year Pantic and Leon[22] analyzed varied techniques of deciphering emotion from AUs and concluded that classification of emotions were primarily targeted at identifying the six basic emotions and that reported results from some of the researches were of little practical value. In 2005 Azcarate et al. [23] studied facial actions and identified emotions but again the research paid little attention to handle the presence of other emotions with a lesser significance simultaneously in faces or the combination of emotions in simple words. They also did not focus much on the attained accuracy of the emotion predictions rather focus was more on how to retrieve AU codes from the face automatically. Quite a few researches have attempted the problem of classification and interpretation of facial expressions by using one HMM for each AU and putting the output through a Support Vector Machine or a Neural Network to obtain inferences about the facial expression in terms of the six basic emotions (Surprise, Sadness, Fear, Happiness, Anger and Disgust). In these works the contribution and significance of accounting for the insignificant and non-present AUs have been undermined and only the significant AU data has been fed to attain model parameter estimation and updating.There are 64 main AUs identified by Ekman and Friesen [1,2] and a few others. The 64 AUs play in combination in varying degrees of displacement from their neutral position to present the final facial expression. The facial expressions according to many researchers are representative of the atomic or basic emotions. Around 7000 combinations of action units are possible. Khademi et al. [24] in 2010 applied a combination of HMM and Neural Networks to identify combination of emotions. Few studies have also addressed the problem of emotion combinations in a facial expression but they have done it at AU level which is quite cumbersome and complex. To overcome the problems of undermining the effect of insignificant and non-present AUs and deciphering the emotional composition of a facial expression in terms of percentages of the seven basic emotions a new model needs to be proposed. To this purpose we propose a HMM based model that incorporates the insignificant AUs by assigning to each of the *M* AUs a set of*N* HMMs, each of the *N* HMMs would be representative of one atomic emotion. Here *M* is the number of AUs we consider and *N* is the number of emotions we intend to study. This type of HMM model was first proposed in [25], but it lacked the ability to predict the emotional composition of an expression as a mixture of 7 basic emotions. Extending the idea we devise a model to predict the emotion mixture for a facial expression in terms of percentages of 7 basic emotions. In the next section we discuss the importance of identifying the mixture of emotions. Also, we know that facial features generally differ between genders, which enable us to differentiate between the two genders visually. So, if we incorporate gender segmentation in our model it would possibly lead us to a better model due to the fact that separate models will be trained according to gender specific features in terms of AU intensities. Thus we use two parallel HMM models to be trained and tested selectively with male and female data respectively. Sections 4 and 5 describe our proposed model and its implementation. In Section 6 we present the results of our experiments and Section 7 presents some concluding remarks on this research.
The gender segmentation of the data after being passed through the model would render two sets of prediction results in terms of percentages of basic emotions. The emotion composition trends of the two sets of results can be compared to analyze gender differences in emotional representation on the human face. There is a lot of confusion about the existence of gender stereotype in emotional expressions on the human face. In Section 3 we introduce few researches on gender stereotype in emotion expression and in Section 6 we discuss about gender stereotype with respect to our findings in addition to the results of prediction of emotional mixtures for observations.

[2]The images belong to the CK+ database (Subject: S130) and is allowed to be published in print or online media. ©Jeffrey Cohn http://www.pitt.edu/~jeffcohn/C-K_Agree.pdf

2. DO BASIC EMOTION ACTUALLY EXIST IN REALITY?

In general, human emotion is rarely "pure" e.g. 100 percent happiness [22]. From psychological point of view, the human mind holds a continuous flow of thought processes and we all know that the human brain can be considered as a fast and vast multitasking machine. In reality the human face with the exception of deception is a portrayal of simultaneous thought processes going inside the mind. But all concurrent thought processes are not equally emphasized on the face and will result in an ordered (according to emphasis of thought processes) combination of emotions displayed on the face. Thus a combination of AUs and their respective intensities visible on the face might represent multiple emotions at the same time. Also, according to Browndyke [26] even emotional deception although effective is not perfect, in that observers can still guess the underlying emotion at greater than chance accuracy. This means that even in the case of deception the facial expression is a combination of emotions. Therefore, it is important to study the emotional composition or mixture of the face rather than just the prominent emotion for facial expressions. Thus we designed our model to predict the mixture of emotions that make up a particular facial expression. Although we do not use data for expressions representing deception, as our model has the ability to predict the mixture of emotions on the face, even for deception it can identify the underlying emotions. For training of our model it is required that we have data for atomic emotions rather than that of combinations of emotions as it is impossible to update model parameters accurately if the learning data consists of expressions reflecting multiple emotions. Let's say an observation in the learning data is representative of multiple emotions then it brings in ambiguity in the choice of emotion specific HMM to be trained with this data. This problem can be solved by using posed expressions where only one single emotion is highly prominent with almost no trace of other emotions. Due to this reason posed facial expression data is the most suitable for training our model and hence we chose the CK+ dataset [10] which contains posed facial expressions performed by trained actors. Our model outputs a percentage mixture of 7 basic emotions while the learning data consists of posed expressions.

3. GENDER DIFFERENCES IN FACIAL EXPRESSIONS AND EMOTIONS

As discussed in Section 1 our model uses gender segmented data with two parallel HMM models for training and testing and outputs two sets of gender specific emotion mixture results (See Figure 3). So we have the scope to analyze and discuss differences in emotion trends of the two genders with respect to the ground truth emotions.It is important to analyze if the difference in facial features of two genders has a great impact on their emotional expressions. In the study of facial expressions and emotions the idea of gender differences has been for long influenced by gender stereotypes. Belk & Snell [27] in 1986 followed by Hess et al. [28] in 2000 concluded that, be it different demographics, gender or different cultures there is persistent belief among all that women are more emotional than men. In 1991, Fabes & Martin [29] mentioned that gender stereotypes held true for both basic and non-basic emotions. This was supported by Fischer [30] in 1993. In the same year Grossman & Wood [31] reiterated that gender stereotype was indeed valid. The idea of women being more emotional than men has been a target for debate for many researchers. Contradicting this, Barrett et al. [32] in 1998 concluded that if the effects of gender stereotype biases are removed, the gender inequalities of emotion response are almost non-existent. In the same year another research by Robinson et al. [33] came up with a similar conclusion. In 1999 Fujita et al. [34] conducted a research on self-reported emotional experience and the results indicated that the gender stereotype actually holds. A year later 2000 Hess et al. [28] and Plant et al. [35] again attempted to establish the gender stereotype followed in 2003 by Timmers et al. [36].Shields [37] in 2003 with a vast reference to empirical research called the common belief that women are more emotional than men is a 'master stereotype'. According to Fabes & Martin [29], Grossman & Wood [31] and Shields [37] the belief of gender stereotype that women are more emotional than men is a generalized concept that ubiquitously exists with different individuals and with most of the known emotions with anger and probably pride being the only exceptions to the generalization. There is also enough empirical studies and research that suggests otherwise. In 2000, Algoe et al. [38] concluded that gender stereotypes do not exercise much influence in represented emotions. This was well supported in coming years by Hess et al. [39] in 2004 and Plant et al. [40] in the same year. Later in 2008 this view was re-established by Simon et al. [41] in his study including intensity ratings of observed emotions. In fact after studying this vivid and vast dilemma among researchers on gender differences in emotion representation this research area seems to be a challenging one. In this work we will examine the general traits of emotion composition on the face for seven basic emotions for both genders separately and compared the results. As we work in this research only with posed expressions, our conclusion will throw some light on the effect of gender differences in posed facial expressions and not natural expressions.

4. THE PROPOSED MODEL

In this work we propose a model to accurately identify the dominant emotion as well as the percentage composition of the facial expression separately for females and males. Our model is an HMM based model. The model is realized in two blocks: one for female data and one for male data. The two parts are functionally the same except for that the training and testing are done separately using gender wise segmented input data in the form of AU intensity observations. The input data consists of 64 AU intensity values (V_1 to V_M) per observation. The input is selectively passed onto the right block in both training and testing phase by a gender redirector (see Figure 3). The gender redirector is a simple gateway to the two blocks wherein the selection of the correct block to be executed is done by the gender input that is fed along with the AU inputs. This enables us to get two gender specific blocks trained and ready for testing in the testing phase with their parameters updated by only one type of gender data. The two identical HMM blocks for male (upper HMM block) and female (lower HMM block) respectively consists of $M*N$ HMMs each. In this paper,N=7 pertaining to the seven basic emotions that we are interested to study. The HMM blocks are used for training and updating of the HMM model parameters according to training set input data. In the testing phase the same blocks are used to calculate probabilities that a particular AU represent the emotionspecific HMM they are passed through. The HMM descriptions are similar to what have been previously proposed by the same authors of this paper in [25]. A set of N HMMs are assigned to each of the M AUs. This makes our model able to gather emotion information from all the AUs irrespective of their visible significance or presence.

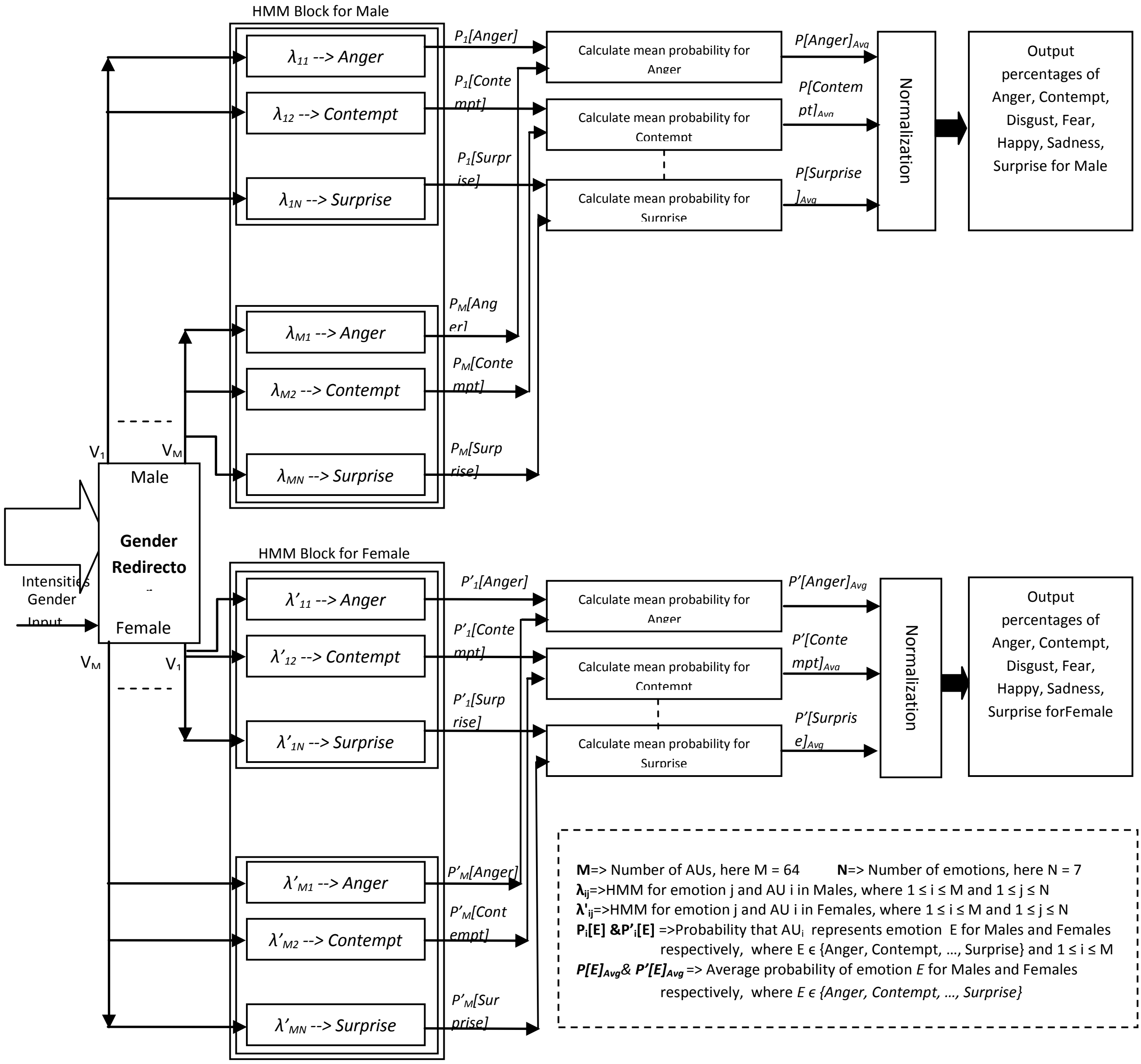

Fig 3: Block diagram of our proposed model

In our case, a set of 7 HMMs one each for Anger, Contempt, Disgust, Fear, Happy, Sadness and Surprise are assigned to each of the 64 AUs. The HMMs are denoted by $\lambda_{ij}$for male block and λ'_{ij} for female block, where$1 \le i \le M$ and $1 \le j \le N$, each corresponding to one of the seven basic emotions (λ_{i1}&λ'_{i1}Anger, λ_{i2}&λ'_{i2} → Contempt, λ_{i3}&λ'_{i3} → Disgust, λ_{i4}&λ'_{i4} → Fear, λ_{i5}&λ'_{i5} → Happy, λ_{i6}&λ'_{i6} → Sadness and λ_{i7}&λ'_{i7}→ Surprise, here $N=7$). Also, the inputs to the HMMs or the observation symbols are$L_{ir} \epsilon\ (L_{i1}, L_{i2},\ldots,L_{iR})$ are theAU_i intensities graded on a scale of 1 to R(here R = 7) where $1 \le i \le M$, $0 \le r \le R$and R is the total number of observable symbols per state inλ_i.The FACS Investigator's Guide [1, 2] grades AU intensities on a scale of A to E where A is the weakest trace of an AU and E is the most prominent trace of an AU. The CK+ database [10] grades AU intensities similar to the FACS Investigator's Guide but assigns numbers from 0 to 5 in increasing order of intensities. It adds an extra level (grade 0) for the AUs that are visible but with no intensity. For simplicity and the inclusion of the non-present condition of an AU in a facial expression we grade it from 1 to 7.

Here an intensity of 1 means no trace of a particular AU, an intensity value of 2 indicates the presence of an AU with no intensity, 3 indicates the weakest trace of the same and moving similarly up the scale an intensity value of 7 represents the most prominent presence of an AU. The parameters of the proposed HMM block according to Das & Yamada [25] is as follows:

$V_i = (V_1, V_2,\ldots,V_M)$ is the observation sequence for each observation in terms of AU intensities, where$0 \le i \le M$. $S_{ij(k)}$are the hidden states for HMMλ_i, where $0 \le k \le X$, $0 \le i \le M$, $1 \le j \le N$and X is the number of hidden states. We have experimentally determined that a value of 7 for X is optimal, by iterating with different values of X starting from 2 until 10. $A_{ij(f,g)}$is the state transition matrix for HMM λ_{ij} where$1 \le f,g \le X$, $1 \le i \le M$ and $1 \le j \le N$, is the probability of transition from previous state $S_{ij(f)}$to the next state $S_{ij(g)}$. Thus,$A_{ij(f,g)} = [q_t = S_{ij(g)} \mid q_{t-1} = S_{ij(f)}]$is the probability of $q_t = S_{ij(g)}$ given at time $t-1$, $q_{t-1} = S_{ij(f)}$where q_t is the state at time$t$, such that, $A_{ij(f,g)} \ge 0$, and $\sum A_{ij(f,g)} = 1$ for $g=1$ to X. $B_{ij(d,e)}$is the observation symbol probability distribution.$B_{ij(d,e)}= P[V_{it} = O_{ie}$at time $t \mid q_t = S_{ij(d)}]$

is the probability of observation symbol O_{ie} for current state $q_t = S_{ij(d)}$ where $1 \leq d \leq X$, $1 \leq e \leq R$. $\pi_{ij(a)} = 1/X$is the initial state distribution, where $1 \leq a \leq X$. As we use discrete data from different facial expressions the AU intensities will be present without a precursor unlike a video stream. So it is equally likely for the HMMs to start at any of the hidden states. Thus we use equal probabilities for the initial state distribution [42]. During the training phase, we update the parameters of the HMMs so as to best explain the patterns of the input vectors. For example, in Figure 3 input V_1 is fed to HMM λ_{11} which represents Anger. In this case updating the parameters means to adjust the state transition probabilities and the output probabilities so as to best match the input sequence V_1. For all the other emotion specific HMMs connected to V_1 gets updated similarly during the training phase. During the training phase each emotion specific HMM (λ_{1j}, $1 \leq j \leq N$, here N=7) of the first sub-block of the upper block gets updated by only the V_1 intensities of those expressions that belongs to the same emotion category i.e. if the HMM is labeled for anger, only those inputs from the training set that have been marked by ground truth as anger will be used to train the HMM. This essentially means during the testing phase the HMMs linked to V_1 can predict the probabilities $P_1[Anger]$, $P_1[Contempt]$, $P_1[Disgust]$, $P_1[Fear]$, $P_1[Happy]$, $P_1[Sadness]$ and $P_1[Surprise]$ that the intensity inputs in V_1 represents anger, contempt, disgust, fear, happy, sadness and surprise respectively. In a similar way in the upper block, all sub-blocks render the probabilities that V_i ($1 \leq i \leq M$) represents the emotion represented by the respective HMMs. So at this point we get M (here M= 64) probabilities for each of the N (here N= 7) emotions. A point to be noted here is that the M probabilities are statistically independent of each other given the face image, because the calculation of probability in one HMM does not require any information of the other HMMs nor AUs. The (conditional) independence could be proved directly using the concept of "d-separation" in Bayesian networks [43]. To integrate all the M probabilities for each emotion into one representative value, we find the mean probabilities for each emotion category to arrive at 7 probability values ($P[Anger]_{Avg}$, $P[Contempt]_{Avg}\ldots P[Surprise]_{Avg}$). The probabilities thus achieved would actually be indicative of the value of average chance that any AU from the entire AU set represents a particular emotion. As these probability values come from different non-mutually exclusive emotions, to calculate the percentage composition or mixture of emotions of the face concerned, we normalize these values by dividing each of the obtained mean probabilities by the sum of the mean probabilities. For example, if for a particular facial expression data, after the normalization step, we get anger = 0.50, contempt = 0.20, disgust = 0.15, fear = 0.05, happy = 0.05, sadness = 0.04 and surprise = 0.01 then we can say that the facial expression is composed of 50% anger, 20% contempt, 15% disgust, 5% fear, 5% happy, 4% sadness and 1% surprise. As the existence of one emotion does not nullify simultaneous coexistence of the other ones [22], the final output can be treated as the percentage composition or mixture of the face in terms of emotions. The entire procedure is repeated for the lower HMM block and $P'[E]_{Avg}$ for all E (where E represents any of the 7 basic emotion considered in our research) can be found for all emotion categories, which is finally normalized to predict the percentage composition of emotions. Also, the lead or prominent emotion would be the emotion category that bears the highest percentage.

5. IMPLEMENTATION

The next two sub-sections deal with the datasets, model training and model testing.

5.1 Datasets

In this research we use the CK+ database [10]. The database containsimage sequences in increasing order of intensity, starting from the neutral expression and ending in the final emotion representation or the peak expression. Total number of frames in the dataset including neutral expressions, peak expressions and intermediate frames is 10,734 across 123 different subjects, out of which 69 percent were females. Emotion data was not given for the intermediate frames and only 327 peak observations were emotion labeled for the peak expression. Under the assumption that minute changes in the intensities do not heavily affect the final depicted emotion we included intermediate frames for our research and manually selected 2749 frames comparatively closer to the peak expression than other intermediate frames. The closeness to the peak expression for intermediate frames is important so as the final emotion depicted is still visually the same and can be treated as separate observations for the corresponding emotion type.

Table 1. Gender and emotion-wise data distribution for training and testing.

Gender	Female		Male		Total
Emotion	Training	Testing	Training	Testing	
Anger	233	233	105	105	676
Contempt	36	36	16	17	105
Disgust	267	267	120	120	774
Fear	83	83	37	38	241
Happy	78	79	35	36	228
Sadness	90	91	40	41	262
Surprise	159	160	72	72	463
Total	946	949	425	429	2749

The data was partitioned gender-wise and emotion-wise. The partitioned dataset was divided into training and testing data in two equal parts, selecting observations for both training and testing in a random manner. The data distribution is shown in Table 1.

5.2 Method of Training and Testing

Once we segmented the data we started training the model. While training the model we trained the upper HMM block for male with 425 observations for male data (see Table 1). As mentioned earlier in section 4 we do not consider any bias for the start state and the HMMs are likely to start in any state.While training the upper block we trained the emotion labeled HMMs with the same emotion category observations. For example, in the male training data there are 105 observations for anger in the training set (see Table 1). So we train all the HMMs labeled with anger for all the M different AUs for each of the 105 observations. Similarly all other emotion categories for the male data were used to train the corresponding emotion specific HMMs. Also, in a similar way 946 female observations(see Table 1) was used to train the lower HMM block. In the above training process, apart from the significant AUs, the other insignificant and non-present AU intensities were also used to train the corresponding HMMs.As discussed earlier that apart from the significant AUs, the insignificant or visible AUs with no intensity and even the non-present AUs contribute to the depiction of emotion on the face, forinsignificant and non-present AUs, the HMMswere trained with intensity grade of 1 and 2 respectively, as described in model description in section 4. This became useful when we moved on to the testing phase in a way that besides the prominent emotion, the less prominent or insignificant emotions simultaneously depicted on the facial expression could be detected.

We used the Baum-Welch algorithm for parameter re-estimation[44, 45] to train the model. The Baum-Welch algorithm is a very precise and efficient way to train HMMs from known observation sequences. Once the training phase finished, we started the testing phase. The testing phase predictedprobabilities for each emotion once for each AU. Then we found the mean probability for all 64 HMMs per emotion category for each of the 7 basic emotions. Finally, we normalized the outputs to get the final composition of the observation in terms of emotion percentages. In the process of probability estimation from the HMMs corresponding to respective inputs (AUintensities), we used the Forward-Backward procedure as explained by Rabiner[44].

6. RESULTS

The success rate or accuracy is defined as the percentage of correct predictions by the model. After completing training and testing of the model we found some interesting results. Das & Yamada [25] achieved an overall average success rate of around 93%. As an extension and improvement of the model, gender segmentation has been proposed in this paper. This improvement in the model yielded better results (around 97%). The emotion-wise success results are shown in Table 2 and Table 3 shows a comparison of our method compared to other similar researches.

Table 2. Gender and emotion-wise success rate for ourmodel

Emotion	Females		Males		%Success All Genders
	No. of Obs	%Success	No. of Obs	%Success	
Anger	233	98.96	105	97.61	98.54
Contempt	36	93.24	17	94.04	93.50
Disgust	267	97.68	120	96.74	97.39
Fear	83	93.79	38	93.54	93.71
Happy	79	94.88	36	94.36	94.72
Sadness	91	98.37	41	97.78	98.19
Surprise	160	97.37	72	95.93	96.92
Overall	949	97.27	429	96.33	96.97

Table 3. Proposed Model Prediction Accuracy Compared with Other Researches

Author	Classification Method	Database Used	Accuracy
Mase[5]	k-Nearest Neighbor	Own	86%
Black et al.[16]	Rule-based	Own	92%
Mingli et al.[46]	Support Vector Machines	Own and Cohn-Kanade	85%
Otsuka & Ohya[6]	HMM	Own	93%
Cohen et al.[4]	Multilevel HMM	Own and Cohn-Kanade	83%
Our Model	M*N HMM	Cohn-Kanade	97%

From Table 3 it is evident that the proposed model achieves some improvement over existing methods of facial emotion recognition.The emotion-wise success percentage is the percentage of the observations within each emotion category for which the prominent emotion predicted by our model matched the ground truth data. Table 4shows the results for emotion-wise average percentage compositions of both prominent and non-prominent emotions.

Table 4.Emotion-wise average percentage compositions of prominent and non-prominent emotions all genders

Emotion	Anger	Contempt	Disgust	Fear	Happy	Sadness	Surprise	Total
Anger	97.43	0.28	1.83	0.22	0.08	0.05	0.11	100
Contempt	0.49	92.19	7.16	0.05	0.02	0.06	0.03	100
Disgust	0.28	5.02	93.1	0.07	0.05	1.37	0.11	100
Fear	3.75	0.17	1.91	85.66	0.09	0.13	8.29	100
Happy	0.32	0.57	0.05	0.07	95.34	0.02	3.63	100
Sadness	0.08	1.17	7.85	0.07	0.05	90.67	0.11	100
Surprise	0.2	0.04	0.07	2.42	2.2	0.03	95.04	100

Table 5.Emotion rankings compared to ground truth ranked by their average percentages in females

Emotion Ground Truth	Ranked Average Occurrences of emotions in Females						
	Rank 1	Rank 2	Rank 3	Rank 4	Rank 5	Rank 6	Rank 7
Anger	Anger (97.64)	Disgust (1.56)	Contempt (0.33)	Fear (0.25)	Surprise (0.12)	Sadness (0.09)	Happy (0.01)
Contempt	Contempt (90.27)	Disgust (9.15)	Anger (0.48)	Sadness (0.04)	Fear (0.03)	Surprise (0.02)	Happy (0.01)
Disgust	Disgust (97.9)	Contempt (1.12)	Sadness (0.62)	Anger (0.15)	Surprise (0.12)	Fear (0.08)	Happy (0.01)
Fear	Fear (89.2)	Surprise (8.31)	Anger (1.3)	Disgust (0.7)	Contempt (0.26)	Sadness (0.16)	Happy (0.07)
Happy	Happy (96.86)	Surprise (2.08)	Anger (0.68)	Contempt (0.32)	Fear (0.03)	Disgust (0.02)	Sadness (0.01)
Sadness	Sadness (95.03)	Disgust (3.53)	Contempt (1.07)	Surprise (0.14)	Anger (0.09)	Fear (0.08)	Happy (0.06)
Surprise	Surprise (95.5)	Fear (2.24)	Happy (2.01)	Anger (0.12)	Disgust (0.08)	Sadness (0.04)	Contempt (0.01)

Table 6. Emotion rankings compared to ground truth ranked by their average percentages in males

Emotion Ground Truth	Ranked Average Occurrences of emotions in Males						
	Rank 1	Rank 2	Rank 3	Rank 4	Rank 5	Rank 6	Rank 7
Anger	Anger (97.22)	Disgust (2.1)	Contempt (0.23)	Fear (0.19)	Surprise (0.1)	Sadness (0.09)	Happy (0.07)
Contempt	Contempt (94.11)	Disgust (5.17)	Anger (0.5)	Sadness (0.08)	Fear (0.07)	Surprise (0.04)	Happy (0.03)
Disgust	Disgust (88.3)	Contempt (8.92)	Sadness (2.12)	Anger (0.41)	Surprise (0.1)	Fear (0.09)	Happy (0.06)
Fear	Fear (82.12)	Surprise (8.27)	Anger (6.2)	Disgust (3.12)	Contempt (0.08)	Sadness (0.11)	Happy (0.1)
Happy	Happy (93.82)	Surprise (5.18)	Anger (0.46)	Contempt (0.32)	Fear (0.11)	Disgust (0.08)	Sadness (0.03)
Sadness	Sadness (86.31)	Disgust (12.17)	Contempt (1.27)	Surprise (0.08)	Anger (0.07)	Fear (0.06)	Happy (0.04)
Surprise	Surprise (94.58)	Fear (2.6)	Happy (2.39)	Anger (0.28)	Disgust (0.06)	Sadness (0.05)	Contempt (0.04)

After completing the testing phase by running the entire testing dataset on our model, we calculated the weighted average for all genders for each emotion category using the number of observations in each gender for that category.The overall success rate was found by calculating the weighted average for all emotion categories using the numberof observations in each category. The data from Table 4 very nearly coincides with the Plutchik's [8] wheel of emotions (see Figure 1) with a few exceptions. From the wheel of emotions and Table 4 together we can observe that after the prominent emotion, the next significant emotions are mostly neighbors on the wheel. As an example, for contempt, the next two prominent emotions are disgust and anger, which are neighbors on either sides of contempt in Plutchik's [8] wheel of emotions. Table 5 and 6 list out the prominent and non-prominent emotions that compose the expressions for the basic emotions ranked in order of their percentages. For example, in the first row in Table 5 the ground truth is anger and on rank1 is anger itself. This means that the average percentage of anger in the emotional composition across all observations of emotion type anger has been the highest. The next high is disgust and so on. We are interested to see the differences in the pattern of emotional compositions between genders and if gender stereotype really holds.

But from Table 5 and 6 we observe that except for the lowest significant emotions i.e. rank 6 and 7 there exists no difference between the two. This may be indicative that gender stereotype for emotions hold true. But in Table 4, if we look closely we can easily observe that the lowest percentages in each row are very small fractions, which means that these emotions will not be readily observed or inferred from the face and will not impact the clarity or intensity of the facial expression. So, the difference between Table 5 and 6 are really insignificant from the point of view of facial expression of emotions. This finding is in accordance to Algoe et al. [38], Hesset al. [39], Plant et al. [40] and Simon et al. [41]. So we can say that for posed facial expressions gender differences do not exist.

7. DISCUSSION

In Table 2, it can be seen that the success rate for contempt and fear categories are lower with respect to the other categories. This is due to lesser number of data available for training. For sadness as well the training dataset was not big but the success rate was still high. This is due to individual differences between subjects. The results in Table 4 do not fully coincide with the wheel of emotions due to the nature of our data but there are a lot of similarities. With the use of N HMMs for the M AUs the model gained more accuracy in predicting emotions and with the introduction of gender segmentation the accuracy was further enhanced.To validate our idea of gender segmentation and the consequent use of two parallel HMM blocks for the two genders, we tested the male HMM block with female testing data and the female HMM block with male testing data. The success results of the model when male testing data is replaced with female testing data and vice versa is shown in Table 7. From the table we see that the overall success rate is reduced by around 13 percent. So we conclude that although gender differences do not exist in case of facial representation of emotions for posed facial expressions, but by developing different models between the two genders, we can get a better model with increased accuracy of prediction. Similar to gender, there is also need to study the effects of culture, racial and ethnic differences on emotion dynamics. This could be an area for future research.

We have already discussed that human emotion is never pure thus this research holds a lot of importance in studying emotional behavior of a per

Table 7.Gender and emotion-wise success rate of the proposed model when testing data is interchanged between genders

Emotion	Females		Males		%Success All Genders
	No. of Obs	%Success	No. of Obs	%Success	
Anger	105	96.51	233	80.11	85.21
Contempt	17	83.57	36	73.35	76.63
Disgust	120	96.25	267	79.42	84.64
Fear	38	88.80	83	76.29	80.22
Happy	36	91.21	79	77.54	81.82
Sadness	41	86.16	91	78.38	80.80
Surprise	72	92.71	160	78.96	83.23
Overall	429	93.17	949	78.75	83.24

Also, this method of emotion recognition is non-intrusive and observational in nature it can be used to develop systems that can assess the mental state in real time, for instance, of a driver while driving or of a psychological patient while talking to a psychiatrist or even of a gamer playing a video game. This project is still in progress and we intend to study how emotions relate to stress which will enable us to assess instantaneous psychological stress of a person.

8. REFERENCES

[1] Ekman P. &Friesen W.V., 1978. Facial Action Coding System: Investigator’s Guide, Consulting Psychologists Press, Palo Alto, CA.

[2] Ekman P., Friesen W.V, &Hager J.C., 2002. The New Facial Action Coding System (FACS), Research Nexus division of Network Information Research Corporation.

[3] Hu T., De Silva L.C., Sengupta K., 2002. A hybrid approach of NN and HMM for facial emotion classification, Pattern Recognition Letters, Volume 23, Issue 11, pp. 1303-1310.

[4] CohenI., GargA., & HuangT. S., 2000. Emotion Recognition from Facial Expressions using Multilevel HMM, Science And Technology, Citeseer.

[5] Mase K., 1991. Recognition of facial expression from optical flow, IEICE Transactions, E74 (10), pp. 3474–3483.

[6] Otsuka T. and Ohya J., 1996. Recognition of Facial Expressions Using HMM with Continuous Output Probabilities, Proceedings International Workshop Robot and Human Comm., pp. 323-328.

[7] Darwin C., 1898. The Expression of the Emotion of man and animals, D. Appleton & Co., New York.

[8] Plutchik R., 2002. Emotions and Life Perspectives from Psychology, Biology, and Evolution, American Psychological Association,Washington DC.

[9] Kanade T., Cohn J., &Tian Y.L., 2000. Comprehensive database for facial expression analysis, Proceedings of the Fourth IEEE International Conference on Automatic Face and Gesture Recognition (FG'00), Grenoble, France, pp. 46-53

[10] Lucey P., CohnJ.F.,Kanade T.,Saragih J.,Ambadar Z.& Matthews I., 2010. The Extended Cohn-Kanade Dataset (CK+): A complete expression dataset for action unit and emotion-specified expression,Proceedings of the Third International Workshop on CVPR for Human Communicative Behavior Analysis, San Francisco, USA, pp. 94-101.

[11] Matsumoto D., 1992. More evidence for the universality of a contempt expression, Motivation& Emotion, vol. 16, pp. 363-368.

[12] Ekman P. & Heider K.G., 1988. The universality of a contempt expression: A replication, Motivation & Emotion, vol. 12, pp. 303-308.
[13] Dai K., Fell H.J. & Macauslan J., 2008. Recognizing emotion in speech using neural networks, In Proceedings of the IASTED International Conference on Telehealth/Assistive Technologies, Ronald Merrell (Ed.). ACTA Press, Anaheim, CA, USA, pp. 31-36.
[14] Wilhelm F.H., Pfaltz M.C. & Grossman P., 2006. Continuous electronic data capture of physiology, behavior and experience in real life: towards ecological momentary assessment of emotion. Interacting With Computers, vol. 18 issue. 2, pp. 171-186.
[15] Nusseck M., Cunningham D.W., Wallraven C. & Bulthoff H.H., 2008. The contribution of different facial regions to the recognition of conversational expressions, Journal of Vision.
[16] Black M. J. & Yacoob Y., 1997. Recognizing facial expressions in image sequences using local parameterized models of image motion. International Journal of Computer Vision, vol. 25 issue. 1, pp. 23–48.
[17] Essa I. A. & Pentland A. P., 1997. Coding, analysis, interpretation, and recognition of facial expressions, IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 19 issue. 7, pp. 757–763.
[18] Kimura S. & Yachida M., 1997. Facial expression recognition and its degree estimation. In Proceedings of the 1997 conference on computer vision and pattern recognition, Washington, DC, USA: IEEE Computer Society, pp.295.
[19] Hong H., Neven H. & von der Malsburg C., 1998. Online facial expression recognition based on personalized galleries. In Proceedings of the 3rd international conference on face & gesture recognition, Washington, DC, USA: IEEE Computer Society, pp. 354.
[20] Ekman P., 1999. Facial expressions, The handbook of cognition and emotion, UK: John Wiley & Sons Ltd., pp. 301–320.
[21] Fasel B. & LuettinJ., 2003. Automatic facial expression analysis: A survey, Pattern Recognition, vol. 36 issue. 1, pp. 259–275.
[22] PanticM.&RothkrantzL.J.M., 2000. Automatic Analysis of Facial Expressions: The State of the Art, IEEE Transactions on Pattern Analysis and Machine Intelligence, pp. 1424-1445.
[23] AzcarateA., HagelohF., SandeK. van de& ValentiR.,2005. Automatic facial emotion recognition, Universiteit van Amsterdam.
[24] KhademiM. et al., 2010. Recognizing Combinations of Facial Action Units with Different Intensity Using a Mixture of Hidden Markov Models and Neural Network,Proceedings of 9th International Workshop on Multiple Classifier Systems, (MCS 2010), Springer-LNCS, vol. 5997, pp. 304-313.
[25] Das S. & Yamada K., 2011. A Hidden Markov Model Based Approach to Identify Emotion from Facial Expression Using a Combination of Emotion Probability Classifier and Facial Action Unit Intensity Classifier, 12th International Symposium on Advanced Intelligent Systems, Suwon, Korea.
[26] BrowndykeJ. N., 2002. Neuropsychological factors in emotion recognition: Facial expressions,www.NeuropsychologyCentral.Com.
[27] Belk S.S. & Snell Jr. W.E., 1986. Beliefs about women: Components and correlates, Personality and Social Psychology Bulletin, vol. 12, pp. 403–413.
[28] HessU. et al., 2000. Emotionalexpressivity in men and women: Stereotypes and self-perceptions, Cognition & Emotion, vol. 14, pp. 609–642.
[29] FabesR.A. &MartinC.L., 1991. Gender and age stereotypes of emotionality, Personality and Social Psychology Bulletin, vol. 17, pp. 532–540, 1991.
[30] FischerA.H., 1993. Sex differences in emotionality: Fact or stereotype?, Feminism Psychology, vol. 3, pp. 303–318.
[31] Grossman M. & WoodW., 1993. Sex differences in intensity of emotional experience: A social role interpretation, Journal of Personality and Social Psychology, vol. 65, pp. 1010–1022.
[32] Barrett L.F., Robin L., PietromonacoP.R. & Eyssell K.M., 1998. Are women the ‘more emotional’ sex? Evidence from emotional experiences in social context," Cognition & Emotion, vol. 12, pp. 555–578.
[33] Robinson M.D., Johnson J.T. & Shields S.A., 1998. The gender heuristic and the database: Factors affecting the perception of genderrelated differences in the experience and display of emotions, Basic and Applied Social Psychology, vol 20, pp. 206–219.
[34] Fujita F. et al., 1991. Gender differences in negative affect and well-being: The case for emotional intensity, Journal of Personality and Social Psychology, vol. 61, pp. 427–434.
[35] PlantE.A., HydeJ.S., KeltnerD. & DevineP.G., 2000. The gender stereotyping of emotions, Psychology of Women Quarterly, vol. 24, pp. 81–92.
[36] TimmersM., FischerA. & MansteadA., 2003. Ability versus vulnerability: Beliefs about men’s and women’s emotional behavior, Cognition & Emotion, vol. 17, pp. 41–63.
[37] Shields A.S., 2003. Speaking from the heart: Gender and the social meaning of emotion. New York: Cambridge University Press.
[38] Algoe S.B., Buswell B.N. & DeLamater J.D., 2000. Gender and job status as contextual cues for the interpretation of facial expression of emotion – statistical data included, Sex Roles, vol. 42, pp. 183–208.
[39] Hess U., Adams Jr. R.B &Kleck R.E., 2004. Facial appearance, gender, and emotion expression, Emotion, vol. 4, pp. 378–88.
[40] Plant E.A., Kling K.C. & Smith G.L., 2004. The influence of gender and social role on the interpretation of facial expressions, Sex Roles, vol. 51, pp. 187–96.
[41] Simon D., Craig K.D., Gosselin F., BelinP. & Rainville P., 2008. Recognition and discrimination of prototypical dynamic expressions of pain and emotions, Pain, vol. 135(1-2), pp. 55-64.
[42] Crowder M., Davis M. & Giampieri G, 2005. A Hidden Markov Model of Default Interaction, Second International Conference on Credit Risk, Citeseer.
[43] Perl. J, 1988. Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference, Morgan Kaufmann Pub, San Francisco.
[44] Rabiner L.R., 1989. A tutorial on Hidden Markov Models and selected applications in speech recognition, IEEE Proceedings 77 (2), pp. 257–286.
[45] Welch L.R., 2003. Hidden Markov Models and the Baum–Welch Algorithm, The Shannon Lecture. IEEE Information Theory Society Newsletter.
[46] Mingli Song et al., 2010. Image Ratio Features for Facial Expression Recognition Application, IEEE Transactions on System, Man & Cybernetics, Part B, 40(3): 779-788.