Multimodal Approach for Emotion Recognition Using a formal computational model

Imen Tayari Meftah[1;2], Nhan Le Thanh[1] and Chokri Ben Amar[2]
[1] Wimmics, INRIA and University of Nice
[2] REGIM Laboratory, University of Sfax,Tunisia

Abstract— **Emotions play a crucial role in human-computer interaction. They are generally expressed and perceived through multiple modalities such as speech, facial expressions, physiological signals. Indeed, the complexity of emotions makes the acquisition very difficult and makes unimodal systems (i.e., the observation of only one source of emotion) unreliable and often unfeasible in applications of high complexity. Moreover the lack of a standard in human emotions modeling hinders the sharing of affective information between applications. In this paper, we present a multimodal approach for the emotion recognition from many sources of information.**
This paper aims to provide a multi-modal system for emotion recognition and exchange that will facilitate inter-systems exchanges and improve the credibility of emotional interaction between users and computers. We elaborate a multimodal emotion recognition method from Physiological Data based on signal processing algorithms. Our method permits to recognize emotion composed of several aspects like simulated and masked emotions. This method uses a new multidimensional model to represent emotional states based on an algebraic representation. The experimental results show that the proposed multimodal emotion recognition method improves the recognition rates in comparison to the unimodal approach. Compared to the state of art multimodal techniques, the proposed method gives a good results with 72% of correct.

Keywords-component- Emotion recognition, multimodal approach, multidimensional model, algebraic representation.

I. INTRODUCTION

Automatic emotion recognition is becoming a focus in interactive technological systems. Indeed, there is a rising need for emotional state recognition in several domains, such as psychiatric diagnosis, video games, human-computer interaction or even detection of lies (Cowie et al., 2001). The lack of a standard in human emotions modeling hinders the sharing of affective information between applications. Current works on modeling and annotation of emotional states (e.g., Emotion Markup Language (EmotionML) (Schroder et al., 2011), Emotion Annotation and Representation Language (EARL) (The HUMAINE Association, 2006)) aim to provide a standard for emotion exchange between applications, but they use natural languages to define emotions. They use words instead of concepts. For example, in EARL, joy would be represented by the following string '<emotion category='joy' >', which is the English word for the concept of joy and not the concept itself, which could be expressed in all languages (e.g., joie, farah, gioia). Our goal is to provide a multi-modal system for emotion recognition and exchange that will facilitate inter-systems exchanges and improve the credibility of emotional interaction between users and computers. We proposed in previous work a Three-Layer Model (Tayari Meftah, Thanh, & Ben Amar, 2011) for emotion exchange composed of three distinct layers: the psychological layer, the formal computational layer and the language layer (cf. Figure 1). In this study we enrich this model by adding a multimodal recognition module that is capable of estimating the human emotional state through analyzing and fusing a number of cues provided through humans' speech, facial expressions, physiological changes, etc.

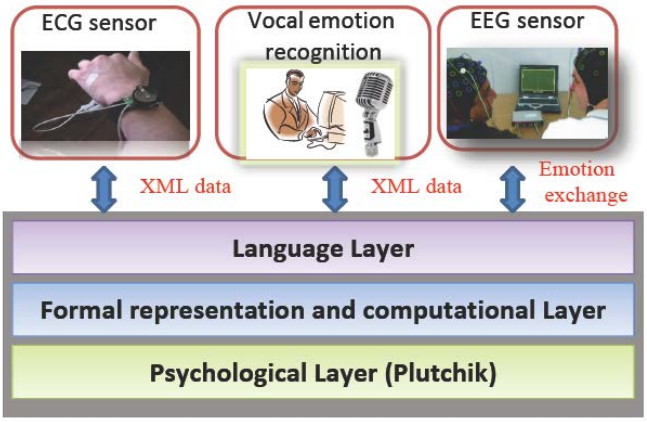

Figure 1. The three layer model

Real life emotions are often complex and people naturally communicate multimodally by combining language, tone, facial expression and gesture (Friesen & Ekman, 2009; Scherer, 1998). Indeed, the complexity of emotions makes the acquisition very difficult and makes unimodal systems (i.e., the observation of only one source of emotion) unreliable and often unfeasible in applications of high complexity. For example, a person can attempt to regulate the expression of her face to hide the true felt emotion. If we analyze only her facial expression we can found joy but he really felt angry and he tries to hide his true emotion. Thus, to improve the emotion recognition system needs to process, extract and analyze a variety of cues provided through humans' speech, facial expressions, physiological changes, etc.

In this paper, we present a new multimodal approach for emotion recognition that integrates information from different modalities in order to allow more reliable estimation of

emotional states. Our contributions concern two points. First, we propose a generic computational model for the representation of emotional states and for a better multi-modal analysis. The second point of our contribution is focused on a multi-modal biometric Emotion Recognition method. This method combines four modalities: Electromyography (EMG), Electro Dermal Activity (GSR), blood volume pulse (BVP) and respiration (cf Figure 2). For each modality we apply a new monomadal emotion recognition based on signal processing techniques.

The remainder of this paper is organized as follows. In Section 2, we give some related psychological and linguistic theories of emotion. Then, we present some of relevant work in the field of automatic emotion recognition. In Section 3, we perform emotion recognition in two stages: unnimodal and multimodal emotion recognition. firstly we describe an unimodal method of emotion recognition based on signal processing algorithm. Secondly, we describes the details of the proposed multimodal approach. Section 4 provides the experimental results. Finally, we conclude in Section 5.

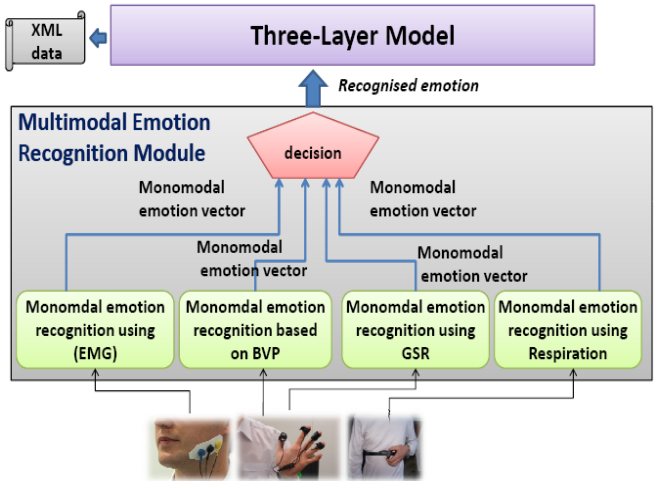

Figure 2. The global scheme of the proposed method

II. RELATED WORK

A. *Emotions*

There is no consensus among psychological and linguistic theories on emotions. The word "emotion" comes from the Latin Latin " emovere, emotum", which means movement towards the outside. An emotion is the consequence of a feeling or the grasping of a situation and generates behavioral and physiological changes. Emotion is a complex concept. Darwin (Darwin, 1872) said that emotional behavior originally served both as an aid to survival and as a method of communicating intentions. He thought emotions to be innate, universal and communicative qualities. Ekman (Ekman, 1982, 1999), Izard (Izard, 1977), Plutchik (Plutchik, 1962, 1980), Tomkins (Tomkins, 1980) and MacLean (Maclean, 1993) have developed the theory that there is a small set of basic emotions all others are compounded. The most famous of these basic emotions are the Big Six, used in Paul Ekman's research on multi-cultural recognition of emotional expressions (Ekman & Davidson, 1994).

According to research in psychology, two major approaches to affect modelling can be distinguished: categorical and dimensional approach. The categorical approach posits a finite set of basic emotions which are experienced universally across cultures. The second approach models emotional properties in terms of emotion dimensions. It decomposes emotions over two orthogonal dimensions, namely arousal (from calm to excitement) and valence (from positive to negative) (Russell, 1980).

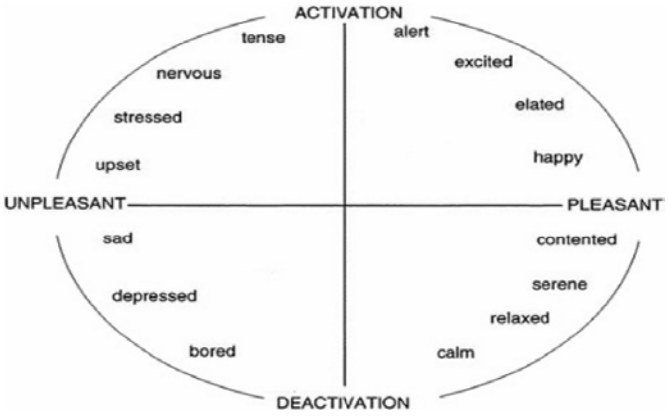

Figure 3. The dimensional approach

Representing emotional states in technological environments is necessarily based on some representation format. Ball and Breese (Ball & Breese, 1999) have proposed a model of emotional states based on the Bayesian networks, which is designed to estimate the emotional state of a user interacting with a conversational compute. López an all (Lopez, Cearreta, Garay-Vitoria, Lopez de Ipiñae, & Beristain, 2009) have proposed a model based on a generic ontology for describing emotions and their detection and expression systems taking contextual and multimodal elements into account. In earlier work (Tayari Meftah, Thanh, & Ben Amar, 2010; Tayari Meftah et al., 2011) we have proposed an algebraic model for the representation and the exchange of emotions. Our model permits to model not only the basic emotions (e.g., anger, sadness, fear) but also different types of complex emotions like simulated and masked emotions.

B. *Previous work on emotion recognition*

Emotion recognition has in the last decade shifted from a side issue to a major topic in human computer interaction. Emotions are displayed by visual, vocal, and other physiological means. This paragraph presents an overview of

research efforts to classify emotions using different modalities: audio, visual, physiological signal and multi-modal emotion recognition.

Audio-based Emotion Recognition. The research for audio-based emotion recognition mostly focuses on global-level prosodic features such as the statistics of the pitch and the intensity (Le, Quénot, & Castelli, 2004). Therefore, the statistical measures such as the means, standard deviations, ranges, maximum values, minimum values and the energy were computed using various speech processing software (Busso et al., 2004). These features are generally classified using a Gaussian Mixture Model (GMM) and a continuous Hidden Markov Model (cHMM) respectively (Meghjani, Ferrie, & Dudek, 2009).

Facial Expression Recognition. The leading study of Ekman and Friesen (1975) formed the basis of visual automatic face expression recognition. Ekman has developed a coding system for facial expressions where movements of the face are described by a set of action units (AUs). Each AU has some related muscular basis. Many researchers were inspired to use image and video processing to automatically track facial features and then use them to categorize the different expressions. Bartlett et al. have developed the Computer Expression Recognition Toolbox (CERT) that automatically extracts facial expressions from video sequences. CERT has been applied to the detection of spontaneous facial expressions of children during problem solving (Littlewort, Bartlett, Salamanca, & Reilly, 2011). Cohen et al (Cohen, Garg, & Huang, 2000) have proposed an architecture of HMMs for automatically segmenting and recognizing human facial expression from video sequences. Zeng et all (Zeng et al., 2006) have explored methods for detecting emotional facial expressions occurring in a realistic human conversation setting.

Emotion recognition from physiological signals. Many affective computing systems make use of physiological sensors to recognize human emotions. Main works are those of (Picard, 1997; Li & Chen, 2006; Anttonen & Surakka, 2005; Andreas, Haag, Goronzy, Schaich, &Williams, 2004). Picard et al. (Picard, 1997) have developed a system that can classify physiological patterns for a set of eight emotions (including neutral) by applying pattern recognition techniques. They have used Sequential Floating Forward Search SFFS and Fisher projection: FP for the selection of an optimal subset of features from physiological signals. Villon and Lisetti (Villon & Lisetti, 2006) have proposed a method and a system to infer psychological meaning from measured physiological cues, oriented toward near to real-time processing. They have introduced the notion of Psycho Physiological Emotional Map (PPEM) as the data structure hosting the emotional mapping between psychological responses and the affective space.

Multi-modal emotion recognition. Multi-modal emotion recognition is currently gaining ground (Pantic & Rothkrantz, 2003; Jonghwa Kim & Wagner, 2005). Indeed, there are many researches in the field of multi-modal emotion recognition. Gunes and Piccardi (Gunes & Piccardi, 2007), have fused facial expressions and body gestures information for bimodal emotion recognition. They have focused on facial expressions and body gestures separately and have analyzed individual frames, namely neutral and expressive frames. Their experimental results show that the emotion classification using the two modalities achieved a better recognition accuracy outperforming classification using the individual facial or bodily modality alone. Schuller et al have proposed techniques for emotion recognition by analyzing the speech signal and haptical interaction on a touch-screen or via mouse. They have classified seven emotional states: surprise, joy, anger, fear, disgust, sadness, and neutral user state (Schuller, Lang, & Rigoll, 2002).

Several researchers have been interested in the fusion of information from facial expressions and speech. For example, Datcu and Rothkrantz (Datcu & Rothkrantz, 2009) have proposed a method for bimodal emotion recognition using face and speech data. They have used hidden Markov models - HMMs to learn and to describe the temporal dynamics of the emotion clues in the visual and acoustic channels. The authors report an improvement of 18.19% compared to the unimodal performances. Also, Wollmer et al (Wollmer, Metallinou, Eyben, Schuller, & Narayanan, 2010) have proposed a multimodal emotion recognition framework based on feature-level fusion of acoustic and visual cues. They focus on the recognition of dimensional emotional labels, valence and activation, instead of categorical emotional tags, such as "anger" or "happiness".

III. EMOTION RECOGNITION

In our study, we explore the use of physiological signals for detecting affect. We elaborate an emotion recognition method from Physiological Data based on signal processing algorithm. Our method permits to recognize emotion composed of several aspects like simulated and masked emotions. The data used for this study comes from the data collected in the MIT Media Lab: Affective Computing Group (Healey, 2000). MIT's data set comprised four physiological signals, obtained from the masseter muscle (EMG), blood volume pressure (BVP), skin conductance (GSR) and respiration rate (RESP) collected over a period of 20 days, concerning eight emotions:

the neutral state, anger, hate, grief, platonic love, romantic love, joy and reverence. We performed emotion recognition in two stages: unnimodal and multimodal emotion recognition. The details of these procedures are explained in the following sections.

A. Unimodal Emotion Recognition

As shown in Figure 4, our approach is composed of two modules: training module and the recognition module. In the training module, feature vectors are extracted from emotion training patterns. In the recognition module, classification has been performed by using the K-Nearest Neighbor algorithm. The result is a 8 component vector representing the detected emotion. This vector is transformed to XML data thanks to the three layer model (Tayari Meftah et al., 2011).

Let us here presents the two modules of the proposed emotion recognition method.

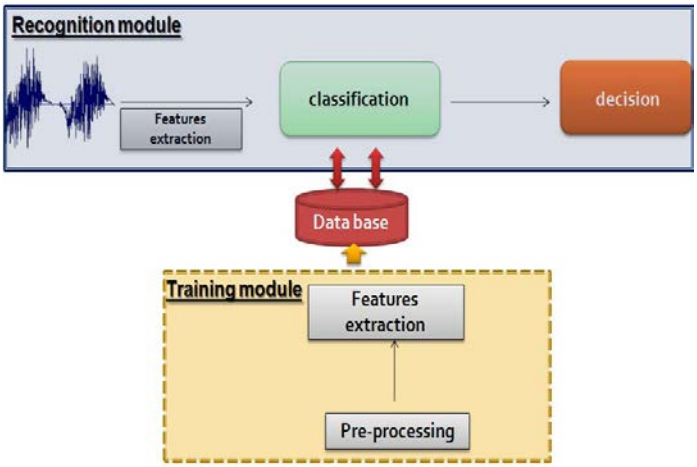

Figure 4. The global scheme of the proposed method of recognition

Training Module. This session explains the proposed method to collect training data. Our newly developed method is based on feature extraction using signal processing techniques. The data consist of 25 minutes of recording time per day over a period of 20 days. Each day include 4 signals showing 8 states in the order: the neutral state, anger, hate, grief, platonic love, romantic love, joy and reverence (cf. Figure 5).

Healey's original data was sampled at a rate of 20 samples per second, creating a digital version of the signal (Healey, 2000). The signal processing for each sensor, include isolation of each emotion, smoothing, peak detection and features extraction. The global scheme of the features extraction is given by Figure 7. Firstly, we segmented the data, according to the emotions elicited at corresponding time frames (for example, although the recording time was 25 minutes, we only used the data from the time frame when the appropriate emotion (eg anger) happened).

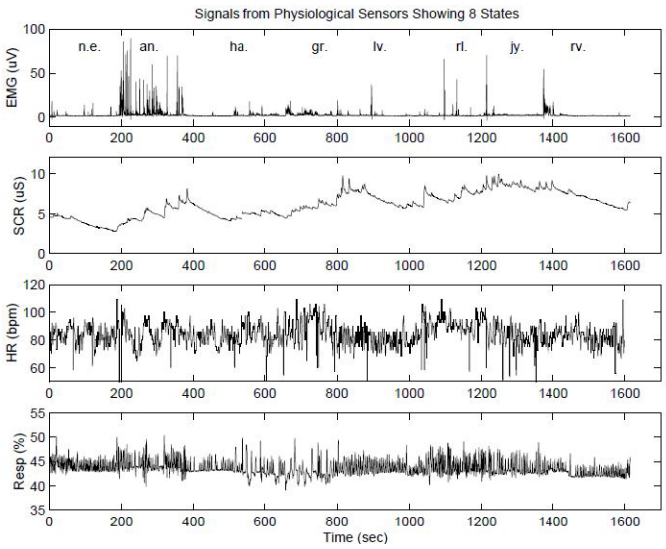

Figure 5. The An example of a session data collected from four sensors

Let A designates the samples taken from any one of the eight emotions and any one of the four sensor (eg emotion anger, sensor:EMG).We process each appropriate emotion data separately to extract 30 representative vectors for this emotion. This is done by applying 3 major steps. First, we smooth the signal to reduce its variance and facilitate the detection of its maxima and minima. Secondly, we compute the gradient of the signal and we apply the zero-crossings method to detect the peaks. Indeed, each peak represents a significant change of the affective state. Thirdly we extract features for each emotion. These steps will be more detailed in the following paragraph. Hanning window (smooth curve) has been used to smooth the signal in order to reduce the variability of the signal. Let the lower case represent a smoothed signal, eg a represents the smoothed signal A. Then we calculate the gradient of the result signal. Let the bar symbol represent the gradient of the smoothed signal.

$$\overline{a} = grad(a) \qquad (1)$$

Afterwards, we apply the detection of the zero-crossings ("PPZ method") of the signals $\overline{a}$ to detect peak. Indeed, a "zero-crossing" is a point where the sign of a function changes (e.g. from positive to negative). Therefore the zero crossing based methods search for zero crossings in a first order derivative expression computed from the signal in order to find the maximum and the minimum of the smoothed signal. Finally we calculate typical statistical values related to peak, such as mean value, standard deviation, the amplitude and the

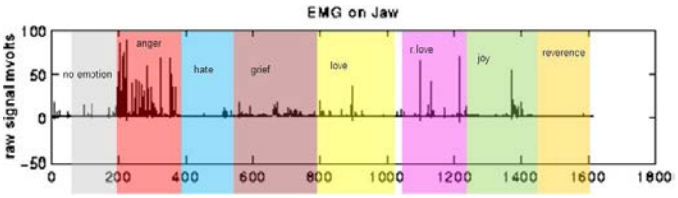

Figure 6. Segmentation of the signal by emotion

width of peak. Then we stored the data in a vector (the emotion feature vector) which corresponds to the appropriate emotion. Thus, we built an emotion training data base composed by 240 vectors representing the eight affective states.

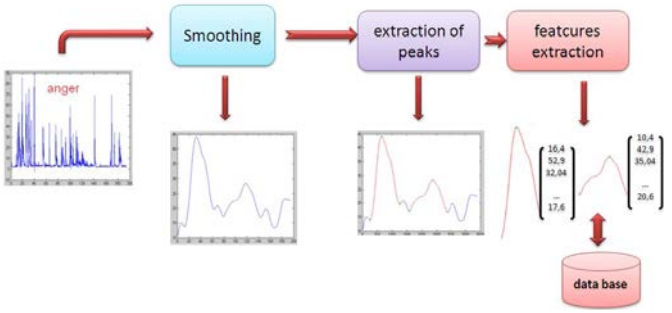

Figure 7. The global scheme of the features extraction module

Recognition Module. The recognition module consists of two steps: (i) features extraction to have test data set and (ii) classification. Test data set was done by using similar steps to the training data, except that it does not have the emotion information. However, we used the K-Nearest Neighbor algorithm (KNN) to classify an instance of a test data into an emotion class. In fact, KNN classification is a powerful classification method. The key idea behind KNN classification is that similar observations belong to similar classes. Thus, one simply has to look for the class designators of a certain number of the nearest neighbors and sum up their class numbers to assign a class number to the unknown.

In practice, given an instance of a test data *x*, KNN gives the *k* neighbors nearest to the unlabeled data from the training data based on the selected distance measure and labels the new instance by looking at its nearest neighbors. In our case, the Euclidean distance is used. The KNN algorithm finds the *k* closest training instances to the test instance. Now let the *k* neighbors nearest to *x* be $N_k(x)$ and $c(z)$ be the class label of *z*. The cardinality of $N_k(x)$ is equal to *k*. Then the subset of nearest neighbors within class $(e) \in$ the neutral state, anger, hate, grief, platonic love, romantic love, joy and reverence is:

$$N_k^e(x) = \{z \in N_k(x), c(z) = e\} \qquad (2)$$

We then normalize each $N_k(x)$ by *k* so as to represent probabilities of belonging to each emotion class as a value between 0 and 1. Let the lower case $n_k^e(x)$ represent the normalized value. The classification result is defined as linear combination of the emotional class.

$$e^* = \sum < N_k^e(x), e > e \qquad (3)$$

Thus,

$$(e^*) = n_k^{noemotion}(x)noemotion + n_k^{anger}(x)anger + n_k^{hate}(x)hate + ...n_k^{joy}(x)\,joy + n_k^{reverence}(x)reverence$$

Thus we build a probability model for each emotion class. Where $n_k^e(x)$ represents the probability of the respective emotion class. For example, if K = 10 and 8 of the nearest neighbors are from emotion class anger and the other 2 are grief , then emotion class anger has an intensity value of 0.8 ($n_{10}^{anger}(x) = 0.8$) and emotion class grief has an intensity value of 0.2 ($n_{10}^{grief}(x) = 0.2$). The classification result is defined as:

$$(e^*) = 0.8anger + 0.2grief \qquad (4)$$

Thus, our recognition method builds a probability model for each class and permits to recognize emotion composed of several aspects. Therefore we get all the information on the emotion. This representation can be transformed, therefore, to the generic computational model of emotional states defined on (Tayari Meftah et al., 2010) by applying the transformation matrix.

B. Multimodal Emotion Recognition

In general, modality fusion is about integrating all single modalities into a combined representation. Indeed, more than one modality can be combined or fused to provide a more robust estimation of the subject's emotional state.

In previous work we have proposed a multidimensional model (Tayari Meftah et al., 2011) to represent emotional states based on an algebraic representation using multidimensional vectors. It is similar to the RGB colors representation model which is based on three basic colors (Red, Green, Blue) to build all the others ones. For example, blue and yellow paints mix together to create a green pigment. In order to develop this analogy, it's necessary to define the basic emotions. For this, we adopted the Plutchik (Plutchik, 1980) definition of basic emotions which is a very intuitive and easy model including the idea that complex emotions are obtained by mixing primary ones (Tayari Meftah et al., 2010). We represent every emotion as a vector in a space of 8 dimensions where every axis represents a basic emotion defined on the Plutchik theory. We defined our base by *(B)* = (joy, sadness,

trust, disgust, fear, anger, Surprise, anticipation). Thus, every emotion *(e)* can be expressed as a finite sum (called linear combination) of the basic elements.

$$(e) = \sum_{i=1}^{8} < E, u_i > u_i \qquad (5)$$

Thus,

$$(e) = \alpha_1 joy + \alpha_2 sadness + \alpha_3 trust + \ldots + \alpha_7 surprise + \alpha_8 anticipati$$

Where α_i are scalars and u_i *(i=1..8)* elements of the basis *(B)*. Typically, the coordinates are represented as elements of a column vector E:

$$E = \begin{pmatrix} \alpha_1 \\ \alpha_2 \\ . \\ . \\ \alpha_8 \end{pmatrix}$$

Where $\alpha_i \in [0;1]$ represents the intensity of the respective basic emotion. More the value of α_i get nearer to 1, more the emotion is felt. According to the Plutchik's theory, the mixture of pairs of basic emotions resulted of complex emotion. Joy and trust for example produce the complex emotion "love". "Submission" is a mixture of trust and fear. We defined the combination between emotions as the sum of two emotion vectors (Tayari Meftah et al., 2010). This addition is defined as the maximum value of coefficients for the same emotion. For the same axis, we keep the highest one because each modality can detect better a specific emotion. For example with the heart rate modality we can detect the fear component better than the facial expression modality.

Let E_{1u} and E_{2u} be two emotional vectors expressed in the basis (B) respectively by $(\lambda_1, \lambda_2, ..., \lambda_8)$ and $(\lambda'_1, \lambda'_2, ..., \lambda'_8)$. The addition of these two vectors is defined as:

$$E' = E_{1u} \oplus E_{2u} = \max_{0 \le i \le 8}(\lambda_i, \lambda'_i) \qquad (6)$$

In this sense, the vector representing the emotion love, which is mixture of joy and trust, is defined as:

$$E_{love} = E_{joy} \oplus E_{trust}$$

$$E_{love} = \begin{pmatrix} \alpha_1 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix} \oplus \begin{pmatrix} 0 \\ 0 \\ \alpha_3 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix} = \begin{pmatrix} \alpha_1 \\ 0 \\ \alpha_3 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}$$

Where $\alpha_1 \neq 0$ et $\alpha_3 \neq 0$

Using vector addition, our model permits to combine information of two or more modalities, e.g. from audio and video. The emotion recognition is done separately for each modality and provide an output vector containing the individual confidence measurements of the unimodal classification process (cf. Figure 8). Our sensor fusion approach combines these four unimodal results to a multimodal decision thanks to the vector addition of our multidimensional model. Due to its generic representation (multidimensional model), this model provides the representation of an infinity of emotions and provides also a powerful mathematical tools for the analysis and the processing of these emotions (Meftah Tayari, Thanh, & Ben Amar, Nov). It is not limited to work with a certain input device, but supports any channel humans use to express their emotional state, including speech, mimic, gesture, pose and physiological signals. Figure 9 shows an example of the using of the add operation on application of emotion detection. On this example the detection is done using two modalities. Each modality gives an emotion vector. The vector *V1* is given by the facial modality and the vector *V2* is given by the physiological modality. The final emotion vector *Vf* is given by the addition of this two vectors using equation 6. Our model permits to combines information of two or more modalities. It can be a solution for the problem of blended and masked emotions, which lead to ambiguous expressions across modalities. For instance, if we consider a situation where the user is forced to talk with calm voice, while at the same time he use mimics to express his anger about something. if we apply an unimodal recognition with voice recognition. we can not detect the anger feeling and the result of detection will be wrong. But using facial recognition make easier the detection of his real emotion state "anger". By taking into account more sources of information, the multimodal approaches allow for more reliable estimation of the human emotions. Indeed, our model takes into account different aspects of emotions: the emotions triggered by an event (the felt emotions) and the

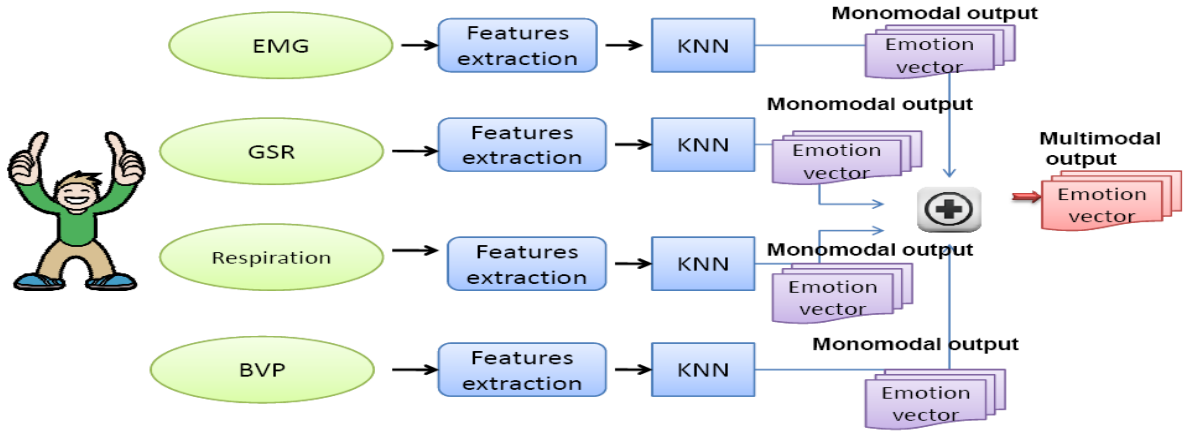

Figure 8. Multimodal approach system

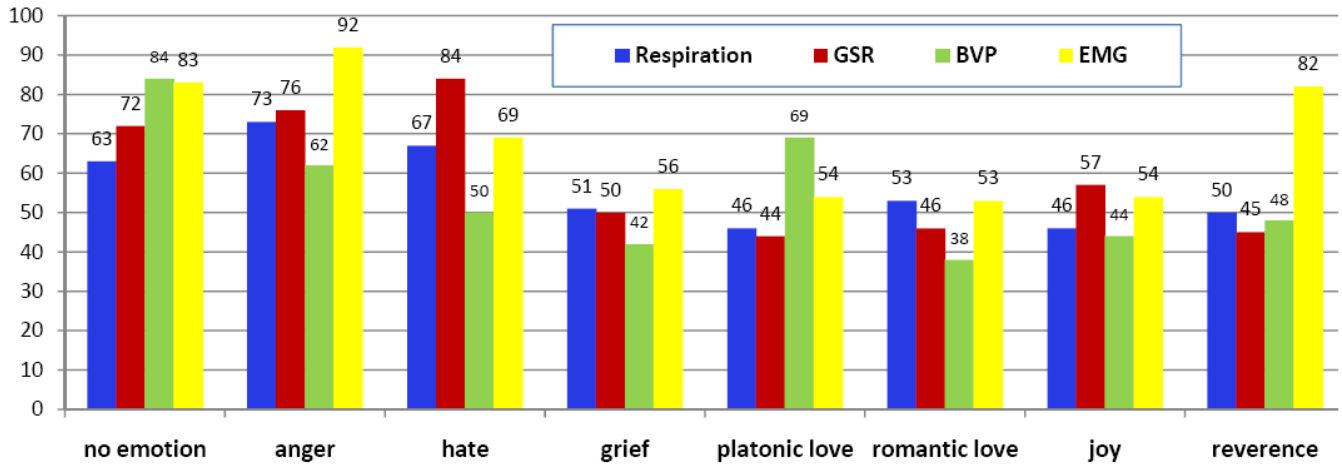

Figure 10. Classification rate of eight emotions for four bio signals

expressed emotions (the displayed ones), which may differ in real life because of the vector addition by means of vector addition operator.

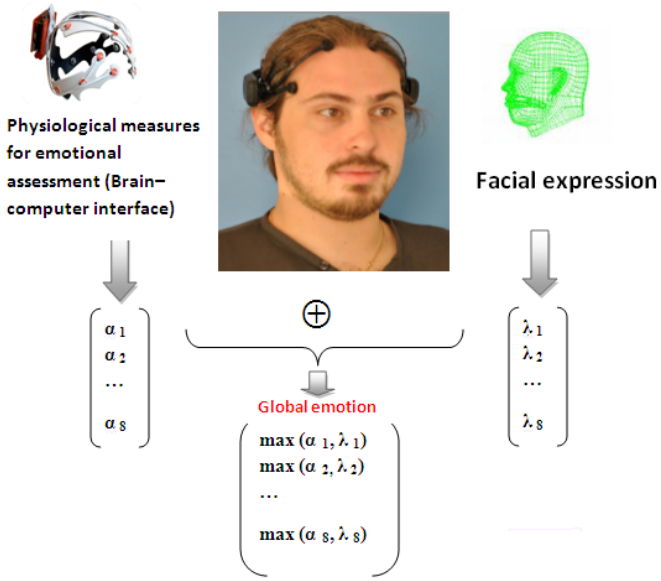

Figure 9. Multi-modality emotion recognition system

IV. EXPERIMENTS AND RESULTS

A. *Unimodal approach*

The first stage of evaluation consisted in the analysis of results coming from the unimodal analysis. Figure 10 shows the results of emotion recognition, indicating the percentage of correct classification using the Respiration and the GSR, the BVP and the EMG signal.

We can notice that the recognition rates varies from one modality to another. The analysis of the EMG signal using the proposed unimodal method, gives the best accuracy percentage of detection. Indeed, we obtained for example 83% for no emotion, 92% for anger and 82% for reverence.

Table 1 shows, for the EMG signal, the classification rates for the proposed method, the baseline method (Muthusamy, n.d.), the HHT-based method (Zong & Chetouani, 2009), Kim's method (Kim, Bang, & Kim, 2002) and SFFS method.

The proposed method is better than the baseline method, the HHT-based method and Kim's method with 64% of accuracy. The SFFS method gives the best result with 83%. Despite its high level of correct classification (83%), this method is limited in the sense that statistical features were calculated over one-day periods and local temporal variations (minutes,

hours) were not taken into account, making real-time detection difficult. The proposed unimodal method can easily detect in real time the affect states thanks to the local temporal analysis done using the extraction of features from peaks.

Table. 1 – *The classification rates of different methods using the EMG signal*

Methode	Classification rate
Baseline	39%
HHT-based, fusion based	44%
HHT-based, fission based	52%
Kim's method	61%
The proposed method	67.8%
SFFS	83%

B. Multimodaleapproach

As depicted in Figure 11, we compare the results obtained from the unimodal method based on respiration signal and the multimodal method using two (respiration+ GSR), three (Respiration+ GSR+BVP) and four modalities (Respiration + GSR+BVP+EMG). It is to be noted that the emotion recognition accuracy is improved by increasing the number of modalities. We obtained the best results by combining four modalities. In fact, fusing the multimodal data resulted in a large increase in the recognition rates in comparison to the unimodal approach.

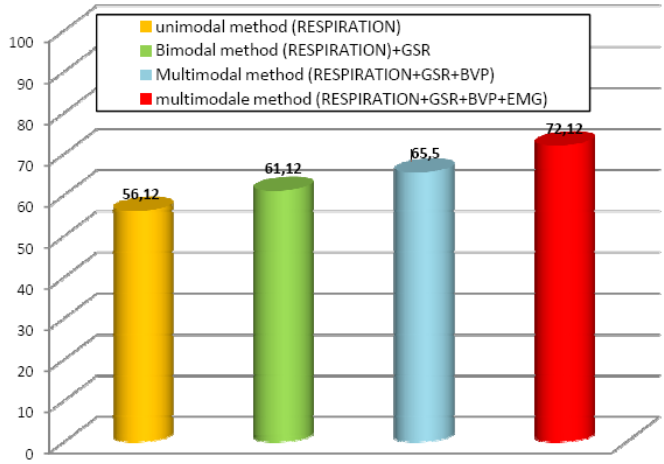

Figure 11. Comparison of results of the unimodal and multimodal approach using two, three and four modalities

A comparison of the results of the unimodal approach using the Respiration signal and the proposed multimodal method combining The Respiration, BVP, EMG and GSR signals has been shown in Figure 12 . It can be seen, that the multimodal approach improves the accuracy percentage of detection. Indeed, we increase, for example, the recognition accuracy of anger, grief, reverence and platonic love respectively from 73%, 51%, 50% and 46% to 92%, 56%, 82% and 59%.

The classification rates for the proposed multimodal method, the HHT-based fusion based method (Muthusamy, n.d.) and the Kim's method (Kim et al., 2002) has been illustrated in Table 2. Using three physiological signals to classify four emotions, Kim's method achieved 61.2% correct classification. The HHT-based fusion based method gives an accuracy percentage of 62%. The proposed method gives the highest recognition accuracy.

Table. 2 – *Classification results: rate of detection for each method*

Fusion of four bio signal	Classification rate
Kim's method	61.2%
HHT-based, fusion based	62%
SFFS-FP, MIT	83%
Baseline	71%
The proposed method	72.1%

V. CONCLUSION

In this paper we proposed a method for multimodal emotion recognition that takes into account more sources of information (physiological signals, facial expressions, speech, etc). This method is based on an algebraic representation of emotional states using multidimensional vectors. This multidimensional model provides a powerful mathematical tools for the analysis and the process of emotions. It permits to build complex emotions out of basic ones. Indeed, using vector addition, permits to combine information of two or more modalities taking into account different kinds of emotions: the felt emotion and the expressed one in order to allow more reliable estimation of emotional states. Experiments show the efficiency of the proposed multimodal approach for emotion recognition. It increased the recognition rates by more than 20% compared with the unimodal approach.

VI. ACKNOWLEDGMENT

We are pleased acknowledge Dr. Rosalind Picard and Dr. Jennifer Healey of the Affective Computing Group at the MIT for providing the experimental data employed in this research.

REFERENCES

Andreas, A. S., Haag, A., Goronzy, S., Schaich, P., & Williams, J. (2004). Emotion recognition using bio-sensors: First steps towards an. In *André et al (eds.): Affective dialogue systems, proceedings of the kloster irsee tutorial and research workshop on affective dialogue systems, lecture notes in computer science 3068, springer-verlag* (pp. 36–48).

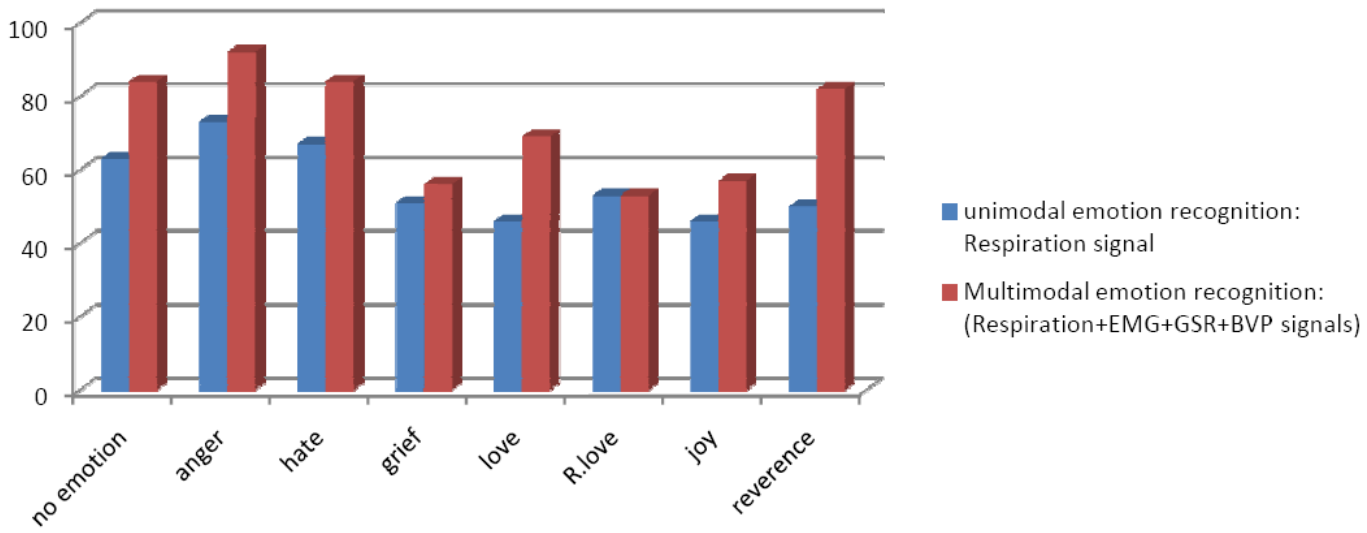

Figure 12. The classification rates using multimodal and unimodal approach

Anttonen, J., & Surakka, V. (2005). Emotions and heart rate while sitting on a chair. In *Chi '05*: *Proceedings of the sigchi conference on human factors in computing systems* (p. 491-499). New York, NY, USA: ACM.

Ball, G., & Breese, J. (1999). *Modeling the emotional state of computer users*.

Busso, C., Deng, Z., Yildirim, S., Bulut, M., Lee, C. M., Kazemzadeh, A., . . . Narayanan, S. (2004). Analysis of emotion recognition using facial expressions, speech and multimodal information. In *Sixth international conference on multimodal interfaces icmi 2004* (pp. 205–211). ACM Press.

Cohen, I., Garg, A., & Huang, T. S. (2000). Emotion recognition from facial expressions using multilevel hmm. In *in neural information processing systems*.

Cowie, R., Douglas-Cowie, E., Tsapatsoulis, N., Votsis, G., Kollias, S., Fellenz,W., & Taylor, J. G. (2001, January). Emotion recognition in human-computer interaction. *Signal Processing Magazine, IEEE, 18, 32–80*. Retrieved from http://dx.doi.org/10.1109/79.911197 doi: 10.1109/79.911197.

Darwin, C. (1872). *The expression of the emotions in man and animals* (3Sub ed.). Oxford University Press Inc. Paperback.

Datcu, D., & Rothkrantz, L. (2009, nov). Multimodal recognition of emotions in car environments. In *Driver car interaction & interface 2009*. Praag and Czech Republic.

Ekman, P. (1982). *Emotion in the human face*. Cambridge University Press, New York.

Ekman, P. (1999). *Basic emotions* (I. T. Dalgleish & T. Power, Eds.). Sussex, U.K.: John Wiley & Sons, Ltd.

Ekman, P., & Davidson, R. J. (1994). *The nature of emotion : Fundamental questions*. Oxford University Press, New York.

Friesen, W. V., & Ekman, P. (2009*). Unmasking the face: A guide to recognizing emotions from facial clues*. Malor Books.

Gunes, H., & Piccardi, M. (2007). Bi-modal emotion recognition from expressive face and body gestures. *Journal of Network and Computer Applications*.

Healey, J. (2000). *Wearable and automotive systems for the recognition of affect from physiology*. Unpublished doctoral dissertation, Massachusetts Institute of Technology.

Izard, C. E. (1977). *Human emotions* (S. Verlag, Ed.) [Book]. Plenum Press, New York.

Jonghwa Kim, M. R. T. V., Elisabeth Andr´e, & Wagner, J. (2005). Integrating information from speech and physiological signals to achieve emotional sensitivity. In *Interspeech 2005 - eurospeech* (p. 809-812). Lisbon, Portugal, 4-8 September.

Kim, K. H., Bang, S. W., & Kim, S. R. (2002). Development of person-independent emotion recognition system based on multiple physiological signals. In *Proceedings of the second joint conference and the annual fall meeting of the biomedical engineering society embs/bmes conference* (Vol. 1, p. 50- 51).

Le, X. H., Quénot, G., & Castelli, E. (2004). Recognizing emotions for the audio-visual document indexing. In *Proceedings of the ninth ieee international symposium on computers and communications* (pp. 580–584).

Li, L., & Chen, J.-h. (2006). Emotion recognition using physiological signals. In Z. Pan, A. Cheok, M. Haller, R. Lau, H. Saito, & R. Liang (Eds.), *Advances in artificial reality and tele-existence* (Vol. 4282, p. 437-446). Springer Berlin.

Littlewort, G., Bartlett, M. S., Salamanca, L. P., & Reilly, J. (2011). Automated measurement of children's facial expressions during problem solving tasks. In (p. 30-35). IEEE.

Lopez, J. M., Cearreta, I., Garay-Vitoria, N., Lopez de Ipiñae, K., & Beristain, A. (2009). A methodological approach for building multimodal acted affective databases. In *Engineering the user interface* (p. 1-17). Springer London.

Maclean, P. D. (1993). *Cerebral evolution of emotion* (Handbook of emotions ed.). Guilford Press, New-York.

Meftah Tayari, I., Thanh, N. L., & Ben Amar, C. (Nov.). Multimodal recognition of emotions using a formal computational model. In *Complex systems (iccs), 2012 international conference on* (p. 1-6).

Meghjani, M., Ferrie, F., & Dudek, G. (2009, December). Bimodal information analysis for emotion recognition. In *Proceedings of the ieee, workshop on applications of computer vision (wacv)*. Utah, USA.

Muthusamy, R. P. (n.d.). Seminar paper: Emotion recognition from physiological signals using bio-sensors. In (p. 334-339).

Pantic, M., & Rothkrantz, L. J. M. (2003). Toward an a ectsensitive multimodal human-computer interaction. *Proceedings of the IEEE, 91(9)*, 1370.

Picard, R. W. (1997). *Affective computing*. MIT Press.

Plutchik, R. (1962). *The emotions: Facts, theory and a new model.* Random House, New York.

Plutchik, R. (1980). *Emotion, a psychoevolutionary synthesis*. New York.

Russell, J. (1980). A circumplex model of affect. *Journal of Personality and Social Psychology* (39), 1161-1178.

Scherer, K. R. (1998). Analyzing emotion blends. In A. F. E.) (Ed.), (A. Fischer (Ed.) ed.). Würzburg: ISRE Publications.

Schroder, M., Baggia, P., Burkhardt, F., Pelachaud, C., Peter, C., & Zovato, E. (2011). EmotionML, an upcoming standard for representing emotions and related states. In *Proceedings of affective computing and intelligent interaction, memphis, usa*. Springer.

Schuller, B., Lang, M. K., & Rigoll, G. (2002). Multimodal emotion recognition in audiovisual communication. In (p. 745-748). IEEE.

Tayari Meftah, I., Thanh, N. L., & Ben Amar, C. (2010, May). Towards an algebraic modeling of emotional states. In *Fifth international conference on internet and web applications and services iciw'10* (p. 513 -518).

Tayari Meftah, I., Thanh, N. L., & Ben Amar, C. (2011). Sharing Emotional Information Using A Three Layer Model. In *The sixth international conference on internet and web applications and services*. Netherlands Antilles: Xpert Publishing Services. Retrieved from http://www.iaria.org/conferences2011/ICIW11.html

The HUMAINE Association. (2006, juin). Humaine emotion annotation and representation language (earl): Proposal. http://emotion-research.net/projects/humaine/earl/proposal.

Tomkins, S. (1980). Affect as amplification: some modifications in theory. *Theories of emotions*, vol. 1, New York, Academic Press., 141-165.

Villon, O., & Lisetti, C. L. (2006, September). A user-modeling approach to build user's psycho-physiological maps of emotions using bio-sensors. In *15th ieee international symposium on robot and human interactive communication, session emotional cues in human-robot interaction*. Hatfield, United Kingdom.

Wollmer, M., Metallinou, A., Eyben, F., Schuller, B., & Narayanan, S. S. (2010). Context-sensitive multimodal emotion recognition from speech and facial expression using bidirectional lstm modeling. In *In proceedings of interspeech*. Makuhari, Japan.

Zeng, Z., Fu, Y., Roisman, G. I., Wen, Z., Hu, Y., & Huang, T. S. (2006). Spontaneous emotional facial expression detection. *Journal of Multimedia*, 1, 1–8.

Zong, C., & Chetouani, M. (2009). Hilbert-huang transform based physiological signals analysis for emotion recognition. In *Signal processing and information technology (isspit), 2009 ieee international symposium on* (p. 334- 339).