# Multimodal approach for emotion recognition using an algebraic representation of emotional states

Imen Tayari Meftah[1,2] and Nhan Le Thanh[1]
[1]*INRIA and University of Nice Sophia Antipolis*
*Sophia Antipolis, France*
*Email: tayari@i3s.unice.fr, nhan.le-thanh@unice.fr*

Chokri Ben Amar[2]
[2]*University of Sfax and ENIS*
*REGIM laboratory*
*Sfax, Tunisia*
*Email: chokri.benamar@enis.rnu.tn*

***Abstract*—Emotions play a key role in human-computer interaction. They are generally expressed through several ways (e.g. facial expressions, speech, body postures and gestures, etc). In this paper, we present a multimodal approach for the emotion recognition that integrates information coming from different cues and modalities . It is based on a formal multidimensional model using an algebraic representation of emotional states. This multidimensional model provides to represent infinity of emotions and provide powerful mathematical tools for the analysis and the processing of these emotions. It permits to estimate the human emotional state through combining information from different modalities (e.g. facial expressions, speech, body postures and gestures, etc) in order to allow more reliable estimation of emotional states. Our proposal permits to recognize not only the basic emotions (e.g., anger,sadness, fear) but also different types of complex emotions like simulated and masked emotions. Experimental results show how the proposed approach increase the recognition rates in comparison with the unimodal approach.**

***Keywords*-multimodal approach, emotion recognition, algebraic representation, multidimensional model.**

## I. Introduction

The use of emotion in computers is becoming an increasingly important field for human-computer interaction. Indeed, affective computing is becoming a focus in interactive technological systems and more essential for communication, decision-making and behavior. 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 [1]. 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) [2], Emotion Annotation and Representation Language (EARL) [3]) 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.

Real life emotions are often complex and people express their emotions through multiple modalities, such as their face, their speech and their body [4], [5]. 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 multi-modal approach for emotion recognition. Our contribution focuses on 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, which combines Electroencephalography (EEG) with Electro Dermal Activity (EDA).

The remainder of this paper is organized as follows. In Section 2, we give some related psychological and linguistic theories of emotion. In Section 3, we describe our model and we conclude in Section 4.

## II. Human Emotion Research

### A. Emotions

There is an apparent lack of consensus and uniformity within the scientific community on what emotions are and how we can represent them. The word emotion comes from the 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 [6] thought emotions to be innate, universal and communicative qualities. Ekman [7], [8], Izard [9], Plutchik [10], [11], Tomkins [12] and MacLean [13] 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. 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) [14]. Representing emotional states in technological environments is necessarily based on some representation format. Ball and Breese [15] 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 [16] 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 [17], [18] we have proposed an algebraic

978-0-7695-4911-8/12 $26.00 © 2012 IEEE
DOI 10.1109/SITIS.2012.84

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.

*1) Emotion Recognition From Audio And Speech:* The research for audio-based emotion recognition mostly focuses on global-level prosodic features such as the statistics of the pitch and the intensity [19]. 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 [20]. These features are generally classified using a Gaussian Mixture Model (GMM) and a continuous Hidden Markov Model (cHMM) respectively [21].

*2) Emotion Recognition Via Facial Expressions:* 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 [22]. Cohen et all [23] have proposed an architecture of HMMs for automatically segmenting and recognizing human facial expression from video sequences. Zeng et all [24] have explored methods for detecting emotional facial expressions occurring in a realistic human conversation setting.

*3) Emotion Recognition From Physiological Signals:* Many affective computing systems make use of physiological sensors to recognize humain emotions. Main works are those of [25], [26], [27], [28]. Picard et al. [25] 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 [29] 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.

*4) Multi-modal Emotion Recognition:* Multi-modal emotion recognition is currently gaining ground [30], [31]. Indeed, there are many researchs in the field of multi-modal emotion recognition. Gunes and Piccardi [32], 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 all 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 [33].
Several researchers have been interested in the fusion of information from facial expressions and speech. For example, Datcu and Rothkrantz [34] 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 all [35] 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 propose a new method of recognizing emotional states from physiological signals. Our proposal uses signal processing techniques to analyze physiological signals. It permits to recognize not only the basic emotions (e.g., anger, sadness, fear) but also any kind of complex emotion, including simultaneous superposed or masked emotions. The data used for this study comes from the data collected in the MIT Media Lab: Affective Computing Group [36]. 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

Our approach consists of two main modules: the training module and the recognition module (c.f. Figure 1) . In the First module, our algorithm extracts the features of emotion from the data to generate an emotion training data base. Then in the second step, we apply the k-nearest-neighbor classifier to assign the predefined classes to instances in the test set. Let us here presents the two modules of the proposed emotion recognition method.

*1) Training module-Features extraction:* 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 (c.f. Figure 2). Healey's original data was sampled at a rate of 20 samples per second, creating a digital version of the signal [36]. The signal processing for each sensor, includes isolation of each emotion, smoothing, peak detection and features extraction (c.f. Figure 2).

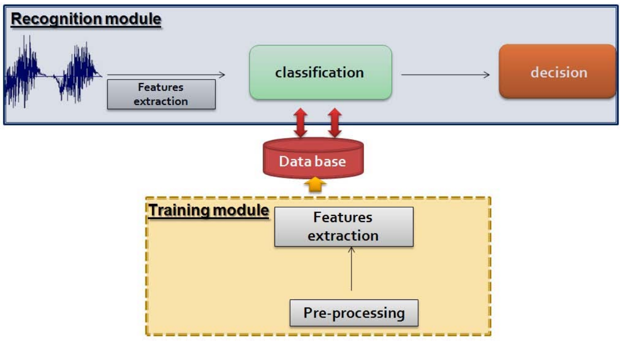


Figure 1. The global scheme of the proposed method of recognition

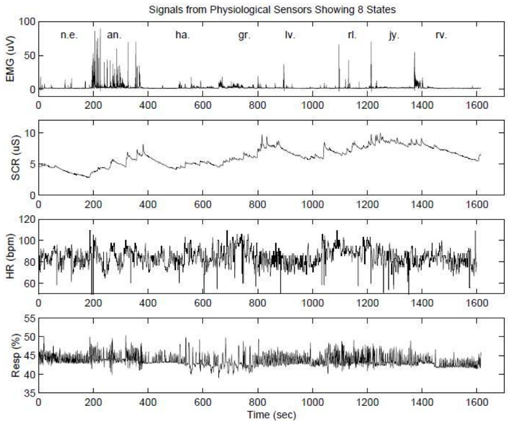


Figure 2. An example of a session data collected from four sensors [36]

The global scheme of the features extraction is given by Figure 4. 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 (e.g., anger) happened). Let A designates the samples taken from any one of the eight emotions and any one of the four sensors (e.g., 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

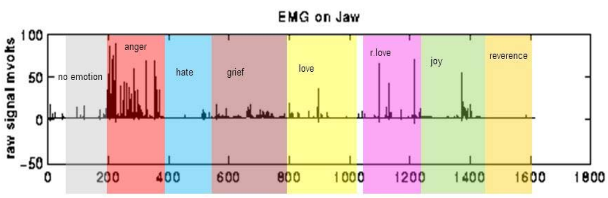


Figure 3. Segmentation of the signal by emotion

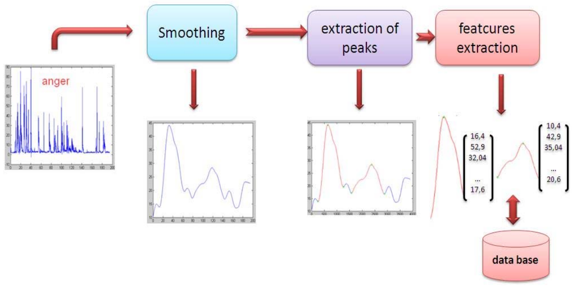


Figure 4. The global scheme of the features extraction module

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 represents a smoothed signal, e.g., a represents the smoothed signal A. Then we calculate the gradient of the result signal. Let the bar symbol represents the gradient of the smoothed signal.

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

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 second-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 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.

*2) 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 classe. Infact, K-Nearest Neighbor (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

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

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

$$e^* = \sum < n_k^e(x), e > e \quad (3)$$

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: $(e*) = 0.8anger + 0.2grief$. 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 [17] by applying the transformation matrix. Therefore, the result is a 8 component vector representing the detected emotion.

### B. Multimodal Emotion Recognition

In the last years, automatic multimodal recognition of human emotions has gained a considerable interest from the research community. By taking into account more sources of information, the multimodal approaches allow for more reliable estimation of the human emotions. 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 [18] 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 [11] definition of basic emotions which is a very intuitive and easy model including the idea that complex emotions are obtained by mixing primary ones [17]. 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.

$$(e) = \sum_{i=1}^{8} \langle E, u_i \rangle u_i \quad (4)$$

thus, $(e) = \alpha_1 Joy + \alpha_2 sadness + \alpha_3 trust + .. + \alpha_7 Surprise + \alpha_8 anticipation$

Where $\alpha_i$ 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:

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

Where $\alpha_i \in [0, 1]$ represents the intensity of the respective basic emotion. More the value of $\alpha_i$ 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 [17]. This addition is defined as the maximum value of coefficients (term by term). We have chose the maximum instead of the classic sum because the intensity must not exceed 1 and because of the specificity of each modality. For example, when we have two modalities, each one will give a vector with different 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:

$$E^{'} = E_{1u} \bigoplus E_{2u} = \max(\lambda_i, \lambda_i^{'}) for 0 \leq i \leq 8 \quad (5)$$

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

$$E_{love} = E_{Joy} \bigoplus E_{trust}$$

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

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. Due to its generic representation, 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. 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. The Figure 5 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 $V_1$ is given by the facial modality and the vector $V_2$ is given by the physiological modality. The final emotion vector $V_f$ is given by the addition of this two vectors using equation 5. 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.

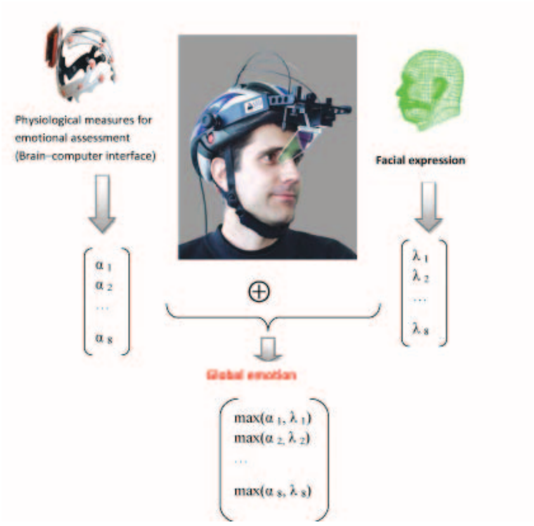


Figure 5. Multi-modality emotion recognition system

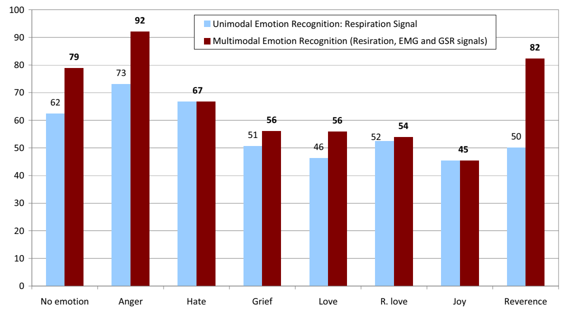


Figure 6. The classification rates using multimodal and unimodal approach

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

## IV. Experimental Results

Figure 6 compares the results of the unimodal approach using the RESPIRATION signal and the proposed multimodal method combining The RESPIRATION, EMG AND GSR signals. It can be seen, that the multimodal approach improves the accuracy percentage of detection. Indeed, we increase, for example, the recognition accurancy of anger, grief, reverence and love respectively from 73%, 51%, 50% and 46% to 92%, 56%, 82% and 56%.

Table I
CLASSIFICATION RESULTS

| Method | Classification rate |
|---|---|
| Kim's method (three physiological signals) | 61.2% |
| HHT-based, fusion based (fusion for four bio signals) | 62% |
| The proposed method | 66.34% |

Table I shows the classification rates for the proposed multimodal method, the HHT-based fusion based method [37] and the Kim's method [38]. Using three physiological signals to classify four emotions, Kim's method achieved 61.2% correct classification. The HHT-based fusion based method gives an accurancy percentage of 62%. The proposed method give the highest recognition accuracy.

## V. Conclusion

The current paper has proposed a multimodal approach for the emotion recognition that takes into account more sources of information (physiological signals, facial expressions, speech, etc). This approach is based on an algebraic representation of emotional states using multidimensional vectors. It provides a powerful mathematical tools for the analysis and the processing of emotions. Thanks to its vector addition operator, it permits to integrate information from different modalities 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 10% compared with the unimodal approach.

## 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

[1] R. Cowie, E. Douglas-Cowie, N. Tsapatsoulis, G. Votsis, S. Kollias, W. Fellenz, and J. G. Taylor, "Emotion recognition in human-computer interaction," *Signal Processing Magazine, IEEE*, vol. 18, pp. 32–80, Jan. 2001. [Online]. Available: http://dx.doi.org/10.1109/79.911197

[2] *EmotionML, an upcoming standard for representing emotions and related states*. Springer, 2011.

[3] Humaine, "Humaine emotion annotation and representation language (earl): Proposal," http://emotion-research.net/projects/humaine/earl/proposal, juin 2006.

[4] W. V. Friesen and P. Ekman, *Unmasking the Face: A Guide to Recognizing Emotions from Facial Clues*. Malor Books, 2009.

[5] K. Scherer, "Analyzing emotion blends," A. Fischer, Ed. ISRE Publications, 1998, pp. 142–148.

[6] C. Darwin, *The expression of the Emotions in Man and Animals*, 3rd ed. Oxford University Press Inc, 1872.

[7] P. Ekman and R. J. Davidson, *The nature of emotion : Fundamental questions*. Oxford University Press, New York, 1994.

[8] P. Ekman, *Basic emotions*, I. T. Dalgleish and T. Power, Eds. Sussex, U.K.: John Wiley & Sons, Ltd., 1999.

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

[10] R. Plutchik, *The Emotions: Facts, Theory and a New Model*, ser. Studies in psychology. Random House, New York, 1962.

[11] ——, *Emotion, a psychoevolutionary synthesis*. Harper and Row, New York, 1980.

[12] S. Tomkins, "Affect as amplification: some modifications in theory," *Theories of emotions, vol. 1, New York, Academic Press.*, pp. 141–165, 1980.

[13] P. D. Maclean, *Cerebral evolution of emotion*, handbook of emotions ed. Guilford Press, New-York, 1993.

[14] J. Russell, "A circumplex model of affect," *Journal of Personality and Social Psychology*, no. 39, pp. 1161–1178, 1980.

[15] G. Ball and J. Breese, "Modeling the emotional state of computer users," 1999.

[16] J. M. Lopez, I. Cearreta, N. Garay-Vitoria, K. López de Ipiñae, and A. Beristain, "A methodological approach for building multi-modal acted affective databases," in *Engineering the User Interface*. Springer London, 2009, pp. 1–17.

[17] I. Tayari Meftah, N. L. Thanh, and C. Ben Amar, "Towards an algebraic modeling of emotional states," in *Internet and Web Applications and Services (ICIW), 2010 Fifth International Conference on*, may 2010, pp. 513 –518.

[18] I. Tayari and N. Le Thanh, "Sharing Emotional Information Using A Three Layer Model," in *The Sixth International Conference on Internet and Web Applications and Services(ICIW 2011)*, IARIA. Maarten, Netherlands Antilles: Xpert Publishing Services, 2011. [Online]. Available: http://www.iaria.org/conferences2011/ICIW11.html

[19] X. H. Le, G. Quénot, and E. Castelli, "Recognizing emotions for the audio-visual document indexing," in *In Proceedings of the Ninth IEEE International Symposium on Computers and Communications*, 2004, pp. 580–584.

[20] C. Busso, Z. Deng, S. Yildirim, M. Bulut, C. M. Lee, A. Kazemzadeh, S. Lee, U. Neumann, and S. Narayanan, "Analysis of emotion recognition using facial expressions, speech and multimodal information," in *in Sixth International Conference on Multimodal Interfaces ICMI 2004*. ACM Press, 2004, pp. 205–211.

[21] M. Meghjani, F. Ferrie, and G. Dudek, "Bimodal information analysis for emotion recognition," in *In Proceedings of the IEEE, Workshop on Applications of Computer Vision (WACV)*, Utah, USA, December 2009.

[22] G. Littlewort, M. S. Bartlett, L. P. Salamanca, and J. Reilly, "Automated measurement of children's facial expressions during problem solving tasks." IEEE, 2011, pp. 30–35.

[23] I. Cohen, A. Garg, and T. S. Huang, "Emotion recognition from facial expressions using multilevel hmm," in *in In Neural Information Processing Systems*, 2000.

[24] Z. Zeng, Y. Fu, G. I. Roisman, Z. Wen, Y. Hu, and T. S. Huang, "Spontaneous emotional facial expression detection," *Journal of Multimedia*, vol. 1, pp. 1–8, 2006.

[25] R. W. Picard, *Affective computing*. MIT Press, 1997.

[26] L. Li and J.-h. Chen, "Emotion recognition using physiological signals," in *Advances in Artificial Reality and Tele-Existence*, ser. Lecture Notes in Computer Science, Z. Pan, A. Cheok, M. Haller, R. Lau, H. Saito, and R. Liang, Eds. Springer Berlin, 2006, vol. 4282, pp. 437–446.

[27] J. Anttonen and V. Surakka, "Emotions and heart rate while sitting on a chair," in *CHI '05: Proceedings of the SIGCHI conference on Human factors in computing systems*. New York, NY, USA: ACM, 2005, pp. 491–499.

[28] A. Haag, S. Goronzy, P. Schaich, and J. Williams, "Emotion recognition using bio-sensors: First steps towards an automatic system," in *Affective Dialogue Systems, Tutorial and Research Workshop, ADS 2004, Kloster Irsee, Germany*, 2004.

[29] O. Villon and C. L. Lisetti, "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, September 2006.

[30] M. Pantic and L. J. M. Rothkrantz, "Toward an affect-sensitive multimodal human-computer interaction," *Proceedings of the IEEE*, vol. 91, no. 9, p. 1370, 2003.

[31] J. Kim, E. André, M. Rehm, T. Vogt, and J. Wagner, "Integrating information from speech and physiological signals to achieve emotional sensitivity," in *Interspeech 2005 - Eurospeech*, Lisbon, Portugal, 4-8 September, 2005, pp. 809–812.

[32] H. Gunes and M. Piccardi, "Bi-modal emotion recognition from expressive face and body gestures," *Journal of Network and Computer Applications*, 2007.

[33] B. Schuller, M. K. Lang, and G. Rigoll, "Multimodal emotion recognition in audiovisual communication." IEEE, 2002, pp. 745–748.

[34] D. Datcu and L. Rothkrantz, "Multimodal recognition of emotions in car environments," in *Driver Car Interaction & Interface 2009*, Praag and Czech Republic, nov 2009.

[35] M. Wollmer, A. Metallinou, F. Eyben, B. Schuller, and S. S. Narayanan, "Context-sensitive multimodal emotion recognition from speech and facial expression using bidirectional lstm modeling," in *In Proceedings of InterSpeech*, Makuhari, Japan, 2010.

[36] J. Healey, "Wearable and automotive systems for the recognition of affect from physiology," Ph.D. dissertation, Massachusetts Institute of Technology, 2000.

[37] R. priya Muthusamy, "Seminar paper: Emotion recognition from physiological signals using bio-sensors," fev 2012, pp. 334– 339.

[38] *Development of person-independent emotion recognition system based on multiple physiological signals*, vol. 1, 2002.