Detection and Classification of the Basic Emotions Using a Multimodal Approach for Emotions Detection

Chaka Koné and Cécile Belleudy
LEAT laboratory
University of Nice Sophia Antipolis
CNRS UMR 7248
France, Sophia Antipolis
Email: ckone@unice.fr,belleudy@unice.fr

Imen Tayari Meftah
REGIM laboratory
University of Sfax
Sfax, Tunisia
Email: imentayari@gmail.com

Nhan Le-thanh
I3S Laboratory
University of Nice Sophia Antipolis
CNRS UMR 7271
Sophia Antipolis, France
Email: nhan.le-thanh@unice.fr

***Abstract*—Negative emotions (anxiety, fear, anger, and grief) may affect physical health and the quality of life. Indeed, people with depression experience severe and prolonged feelings of negative emotions like sadness, anger, disgust and fear. On one hand, this paper presents a new method for the fusion of signals for the purpose of a multimodal recognition of eight basic emotions, on the other hand, it present a classification of these basic emotions in three emotional classes, namely, neutral, positive and negative emotions which are using physiological signals. After constructing an emotion data base during the learning phase, we apply the recognition algorithm on each modality separately. Then, we merge all these decisions separately by applying a decision fusion approach to improve recognition rate. The experiments show that the proposed method allows high accuracy emotion recognition. Indeed, we get a recognition rate of 81.69% under some conditions.**

Keywords*— *Signal fusion method; basic emotions; multimodal detection; physiological signals.

I. Introduction

Historically, emotions had a great impact on our behavior, our feelings and we are constantly trying to manage our emotions as well as the people that surround us, in order to live together in harmony. Indeed, emotions enable us to communicate with our environment but also to adapt, to innovate, to succeed, and to flourish.

A lot of research based on video application or speech analysis [9][10] (EMOTIENT, Eurospeech, Nice Speech) has emerged to analyze emotions, with the aim, amongst other, to provide a real-time, aggregate view of users feelings and in general to identify customer dissatisfaction. The solution proposed in this article targets the healthcare domain in that it monitors biological signals, but in a non-intrusive manner for the benefit of patients.

In the future, emotion detection tests will be very challenging because they constitute a key point to analyze the impact of all medical treatments, and the resulting device market will probably be substantial. We should also note that neurodegenerative diseases are characterized by the presence of cognitive and behavioral disturbances, gradually resulting in a loss of autonomy for the realization of acts of the daily routine. These disturbances can fluctuate over time, making the evaluation sometimes difficult with the usual clinical tools. Non-pharmacological approaches for their daily care must be favoured, as pharmacological approaches are time-consuming and involve a significant financial cost as they require a qualified professional entourage who are regularly sought. In this context, the use of new Information and Communication Technologies (ICT) can improve both clinical assessment techniques, but also non-medicated therapeutic techniques. Monitoring patients requires, amongst other, being able to assess their emotional state. The information produced is an aid for the diagnosis and the applicability of medical treatments.

Indeed, new technologies benefiting people's health have emerged and have allowed, for example, developing the bases of affective computing, defined by Rosalind Picard in 1995 [8]. This is an area which aims to study the interactions between technology and emotion to give machines the ability to understand, to interpret our emotions, or even express emotions. Affective computing offers many advantages such as the battle against depression, interactive games, E-Learning, etc.

Our goal is to collect the physiological signals of a person under different conditions of real life to detect emotions automatically. We propose a method for a multimodal detection of emotions using physiological signals. The paper is structured as follows. In Section 1, a brief state of the art on the multimodal recognition of emotions and different methods to merge signals is described. In Section 2, all the steps of the proposed methods are explained in detail; in Section 3, a comparison is made between our results and those obtained in the state of the art; finally, conclusion and future work are reported in Section 4.

II. State of the art

Emotions detection systems are based on three fundamental steps: acquisition of the signals, features

 ISBN: 978-1-61208-424-4

extraction of these signals and the emotions detection. Many works were focused on the emotions detection using facial expressions, vocal expressions or physiological signals [14][15][16]; however, fewer studies are focused on the multimodal recognition [2] of emotions. The use of a multimodal approach allows not only enhancing the recognition rate but it gives more strength to the system when one of these modalities is acquired in a noisy environment [3]. In theory, there are three methods [12] to merge the signals from various sensors: fusion at the signal level (fusion of physiological signals), feature level fusion (fusion of features) and decision level fusion (fusion of decisions) [4][5].

1) Fusion of signals [4]: This fusion is performed on raw data directly (as shown in Figure 1) from each physiological signal sensor; it can be applied only when signals are similar in nature and have the same temporal resolution. This technique is, therefore, rarely used on account of the difficulty of merging the different signals and the noise due to the sensor's sensitivity.

Figure 1. Signals level fusion

2) Fusion of features [5]: This fusion method (see Figure 2) is most frequently used; it aims at forming a multimodal vector from features vectors extracted for each sensor. It has the advantage of requiring only a single learning phase and the resulting feature vector is multimodal. Otherwise, all descriptors must be synchronized and the treatment is done without taking into account the interactions that may exist between the various parameters.

Several techniques have been proposed in the literature

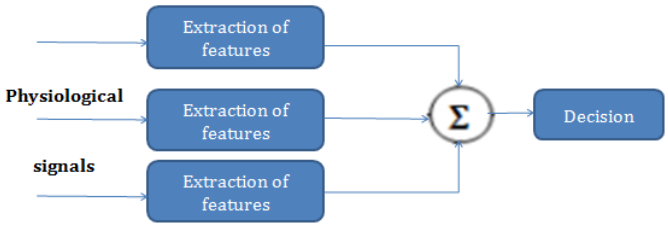

Figure 2. Characteristic level fusion.

[6][7], notably the concatenation of unimodal features vectors, the method of K-Nearest Neighbors (KNN) [17], the Analysis Principal Component (ACP) [8] etc. Fission and Fusion based Hilbert Huang Transform (HHT) features consist in the extraction from Intrinsic Mode Function (IMF) by applying the Empirical Mode Decomposition (EMD) [13].

o Fission: We extract 2 features: instantaneous frequency and amplitude using the Hilbert transform on each IMF.

o Fusion: We determine the use of weighted mean frequency, and the mean instantaneous frequency.

This technique of fusion based HHT features (fusion of 4 physiological signals) has allowed to have a 62% recognition rate while MIT's hybrid method Sequential floating forward search - Fisher projection (SFFS-FP) allowed to have an 83% recognition rate. Nevertheless, the SFFS-FP method allows to determine emotions during a fixed time interval, therefore it does not permit an instantaneous detection. The SFFS-FP method, based on a scenario, calculates the characteristics of a given emotion, and then reduces the number of characteristic data, which enables emotions detection over a long period but not instantaneously. Unlike MIT's method that does not permit an instantaneous emotions detection, our method allows automatic and instantaneous detection of emotions.

In the following section, we focus our study on the instantaneous detection of emotions.

3) Fusion of decisions: After having classified separately from each sensor signals, this is a way to merge these different decisions in order to obtain a global vision of the emotion. Unlike features level fusion, this fusion technique is independent of the nature of the low level features used for decision making [4].

The most used and most intuitive technique is to take

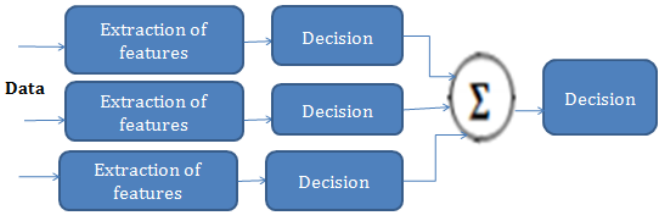

Figure 3. Decision level fusion.

into account for the vote all modalities (each sensor: ElectroMyoGram (EMG), ElectroCardioGram (ECG), Galvanic Skin Response (GSR),Blood Volune Pulse (BVP)), and to choose the decision expressed by the maximum of modalities; Figure 3 shows the steps for this technique.

III. Methodology

In this section, we present a new multimodal and automatic method of emotions recognition based on the fusion of the above decisions. Our method is divided into two major phases, namely: the Learning phase and the Detection phase.

A. Learning phase

This phase consists of four steps (signal splitting, filtering, feature extraction, creation of the basis for learning) in order to provide a learning base which will then be used in the detection phase for the automatic detection of emotions. Figure 4 shows the synoptic of the learning phase.

 ISBN: 978-1-61208-424-4

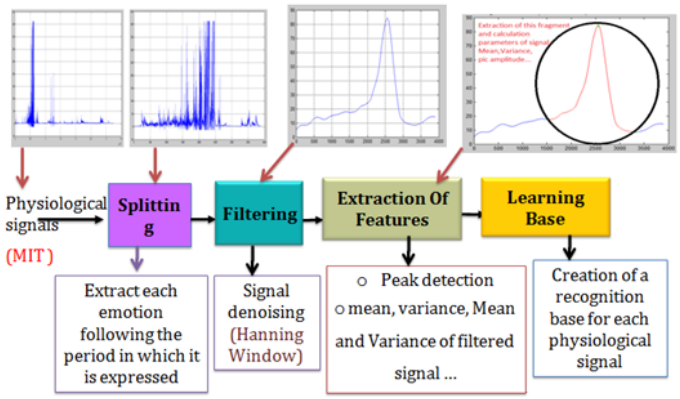

Figure 4. Synoptic of the learning phase.

1) Signal splitting: In this step, after having acquired the physiological signal (here we use the physiological signals provided by MIT [1]), we isolate the part of the signal corresponding to a given emotion because we have information on the period in which each of the eight emotions is expressed. Therefore, this step divides the input signal into eight portions of signal corresponding to eight emotions.

2) Filtering: After having isolated the signal, we filter it to remove the noise of the useful signal, which will facilitate the extraction of the features. We have opted for the convolution method for filtering, which consists in convoluting the signal in the spatial domain with different filters (for which we chose the Hanning filter [8]). This method is less computationally expensive in calculations.

3) Extraction of features: For each isolated and filtered signal, we proceed to the detection of peaks, which is done by calculating the gradient of the signal and then, finding the sign changes in the gradient, because it is rare to find points in discrete signals where the gradient is zero. A maximum is shown by the passage of a positive gradient to a negative gradient, a minimum by the passage of a negative gradient to a positive gradient. To detect and isolate a peak, our method should detect a minimum followed by a maximum followed by a minimum. Once a peak is isolated, we calculate a feature vector composed of five features that are: the mean, the variance, the mean of the filtered signal, the variance of the filtered signal, and the amplitude of the peak.

4) Creation of the Learning phase: After extraction of the features vectors, we create a learning base for each modality. Thus, at the end of this step, we get 4 learning bases which are made of 30 vectors for each emotion, resulting in a total of 240 features vectors. 40% of the signals available for each modality were used for the creation of the learning base and the remaining 60% were used for detection (test).

B. Detection phase

Our research is based on that of Imen [8], which has developed a new vector method to represent emotions. Therefore, each emotion is written as a linear combination of the 8 basic emotions (B) we considered. Indeed, each emotion e can be written as: (B)= (No emotion, anger, hate, grief, love, romantic love, joy and reverence)

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

$$(e) = \alpha_1 * NoEmotion + \alpha_2 * Anger + ... + \alpha_8 * Reverence \quad (2)$$

$$(e) = (\alpha_1 \ \ \alpha_2 ... \ \ \alpha_8)_B \quad (3)$$

where $(\alpha_1, \alpha_2, ..., \alpha_8)$ are the probabilities of the feature vector extracted belonging to each emotional class of our base. This phase consists of two steps. The first step consists in the extraction of features, requiring the same steps as in the learning phase, without going through the splitting step since in this phase, there is no information beforehand on the period at which every emotion is expressed.

The remainder of our process will be based on this features extraction step. It is necessary to detect a peak (an emotional activity) before pass to classification step. Indeed, thanks to this condition on the necessity of detecting an emotional activity, our method allows an instantaneous recognition of emotions. The second step is classification, the purpose of which is to predict the emotional class of the features vector extracted using our learning base, which was developed in the learning phase. We opted for the classification using the K-Nearest Neighbors (KNN) algorithm, which is based on the calculation of the Euclidean distance between the extracted feature vector, the emotional class of which is to be predicted, and 30 features vectors, which are found in our learning base. This allows determining the K nearest emotions in our database of extracted feature vector. Studies [8] have shown that the optimal value of K = 10 and the size of the Hanning window n = 500 enable the best detection. For example, in our algorithm, at the classification stage, among the K = 10 Nearest Neighbors, we have 6 elements belonging to the anger emotional class, 3 elements belonging to the grief emotional class and 1 element to the hate emotional class. Thus, the resulting emotional vector is:

$$(e) = 0.6 * Anger + 0.3 * Grief + 0.1 * Hate \quad (4)$$

where the coefficients (**0.6, 0.3, and 0.1**) are the probabilities of the feature vector extracted belonging to each emotional class of our base. Of this emotional vector, we keep as a final decision the most likely emotional class (the anger class in the above example).

 ISBN: 978-1-61208-424-4

1) Fusion method of signals through voting: In this section, we studied 2 voting techniques of formalisms which are (i) that consisting in calculating the vector average of the monomodal emotions vectors [19] and (ii) that consisting in making a choice among all the monomodal decisions [19].

o(i) The first technique consists in constituting a matrix (of size 4*8 because we have 4 modalities and 8 emotional classes in our case) made up of the emotional vectors for each modality. We calculate the average of this matrix, which gives us a vector average from which we choose the most probable emotional class. This technique is a better measurement of the center around which the values of the probabilities of each emotional class for each modality tend to concentrate. However, it does not allow a detection of the most probable emotional classes.

o(ii) In the second approach, starting from each monomodal vector, we take the most probable emotional class. Thus, we will have as many decisions as there are modalities (in our case, we have 4 decisions). The final decision will be the class having been decided by the maximum of modalities. This allows one side to take the most probable partial decisions for each modality, and on the other hand, it allows a measure of the central tendency as in the first technique. We opted for this technique on account of the two advantages that we have just enunciated.

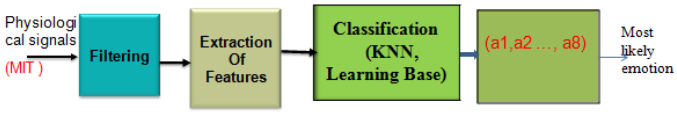

Figure 5. Synoptic of the detection phase for each modality.

Figure 5 shows the different steps in the detection phase for each of the modalities before the mixing step (fusion of decisions).

Our objective being thereafter to put these algorithms in mobile devices which do not have a great memory size, it is thus necessary to set up simple algorithms. That is the reason why we chose this fusion approach on the decisions level. It is simpler and more intuitive than the two others. For example, if we decide on the anger class for the EMG modality, the grief class for the EDA modality, the anger class for the BVP modality and the hate class for the RESP modality, in the step of fusion, we will choose the emotional class of anger which in this case is the class voted by the maximum of modalities.

IV. Results

For these results, we use as data base the signals provided by the MIT [1]. This physiological data collection, the process of which is well described in [1], was carried out on an actor during 32 days for a period of 25 minutes per day, with a sampling frequency of 20 Hz. For this collection, four physiological sensors were used: sensor for the blood volume and pulse (BVP), the pace and volume of respiration (RESP), the electromyography (EMG), and the galvanic skin response (GSR). During this collection, eight emotions were taken into account, which are "no emotion, anger, hate, grief, platonic love, romantic love, joy and reverence" and every emotion was maintained for three to five minutes per day. The results obtained by our algorithm when the unimodal recognition of emotions approach is used are grouped on the histogram below. This approach allows having a mean recognition rate of 57.24%.

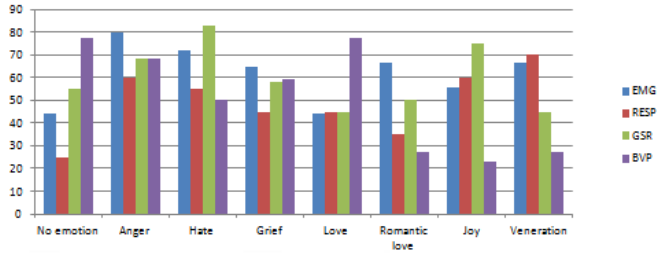

Figure 6. Monomodal recognition rate.

As observed in Figure 6, certain emotions are better detected with certain modalities than others. Indeed, the BVP modality allows to better detect the "no emotion" and "love" emotions, while the GSR modality better detects the 'hate' and 'joy' emotions. The EMG modality rather allows a better detection of the emotions "anger" and "romantic love". This characteristic of modalities is very important because it will allow putting weight on each of the modalities, depending on whether it can better detect an emotion or not for the purpose of a more efficient detection. Subsequently, we have expanded our method to the multimodal approach to increase the emotion recognition rate. Indeed, this multimodal approach allowed having a recognition rate of 81.69%, which is a considerable improvement of the recognition rate compared to the monomodal approach which presented a recognition rate of 57.24%.

Table I. RESULT OF MULTIMODAL APPROACH

Emotion	*Accuracy in % (EMG,GSR,BVP,RESP)*
No Emotion	78.89
Anger	90.11
Hate	88.89
Grief	74.45
Love	72.25
Romantic Love	71.18
Joy	83.33
Veneration	94.44

The results grouped in the Table 1 show a good average recognition rate. Furthermore, we note that our method allows to detect each of the eight emotions with a good recognition rate, where the minimum of 71.18% is obtained for the emotion "Romantic Love "and the maximum is obtained for the emotion" Veneration "with

 ISBN: 978-1-61208-424-4

a good classification rate of 94.44%. Table 2 allows doing a comparison between our results and the different results of the methods of the state of the art that allow an instantaneous detection of emotions. The method we have

Table II. COMPARISON OF RESULTS

Methods	*Good recognition rate in %*
Method of Kim [11]	61.2
Fusion based HHT features [13],[8]	62
Baseline [8]	71
Proposed Method	81.69

proposed allows a better classification of emotions than all the other methods found in table 2. Statistics demonstrate that approximately 150 million people suffer from a major depressive disorder at any moment, and almost a million commit suicide each year [18]. In fact, neurodegenerative diseases like Alzheimer, that are considered like a cognitive deficiency are caused and/or are source of negative emotions. However, detecting and classifying emotions depending on whether they are negative or not can be used as a means of preventing depression and on the other hand, it may also be used to help health professionals who work with people with cognitive deficiencies. Based on this important observation, we then decided to put in place a strategy of classification of the emotional vector in three emotional classes using the multidimensional model mathematical proposed by Imen [17]. The three emotional classes we have are:

- Neutral emotions : No emotion, Reverence
- Positive emotions : Love, Romantic love, Joy
- Negative emotions: Anger, Hate, Grief

As we mentioned in the previous section, every emotion can be written as a linear combination of other emotions, and using the algebraic representation method of emotions described in [17], from the representation of an emotional vector, representation of the basic emotions is described by a vector containing a single coefficient non zero. For example

$$E_{grief} = (0\ 0\ 0\ \alpha_4\ 0\ 0\ 0\ 0)_B \quad (5)$$

$$E_{joy} = (0\ 0\ 0\ 0\ 0\ 0\ \alpha_7\ 0)_B \quad (6)$$

where $\alpha_4 \neq 0$ and $\alpha_7 \neq 0$

By combining these two vectors of basic emotions, one can get a complex emotion defined as an emotion composed of more than one emotion.

$$E_{combination} = E_{grief} \oplus E_{joy} = (0\ 0\ 0\ \alpha_4\ 0\ 0\ \alpha_7\ 0)_B \quad (7)$$

So, based on this algebraic representation, our classification method is described as follows: For each modality, after detecting an emotional vector which consists of 8 eight basic emotions that each of them will be represented in the form of the (5), combining these vectors to have three emotional classes in the following way:

$$E_{NeutralEmotion} = E_{NoEmotion} \oplus E_{Reverence} \quad (8)$$

$$E_{PositiveEmotion} = E_{Love} \oplus E_{RomanticLove} \oplus E_{joy} \quad (9)$$

$$E_{NegativeEmotion} = E_{Anger} \oplus E_{Hate} \oplus E_{Grief} \quad (10)$$

Subsequently, we use the decisions level fusion proposed for a multimodal classification of emotions. As we note

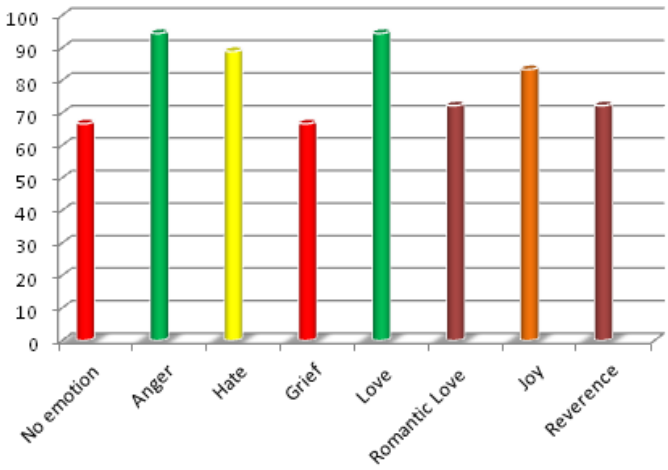

Figure 7. Classification recognition rate.

on Figure 7, our method allows us to have a satisfactory classification rate of emotions. The minimum rate of 66.67% correct classification is obtained for the emotions (**No emotion and grief**) while we obtained a maximum of 94.44% for emotions (**Anger and Love.**)

V. Conclusion and perspectives

We have developed a novel method of multimodal recognition of emotions based on the processing of physiological signals. The physiological signals of 4 modalities were used for the recognition of 8 basic emotions. A new method for multimodal recognition based on the fusion decision level has been defined and developed. On the other hand, a classification method of emotions in three emotional classes has been proposed. The different results show a marked improvement in the recognition rate of emotions. In our future work, on one hand, we will study physiological signals acquisition platforms in order to generate our own recognition base and on the other hand set up a complete system from the acquisition of physiological signals for the detection of emotions. Moreover, this system will allow creating a more appropriate recognition base for a wide range of people.

 ISBN: 978-1-61208-424-4

References

[1] J. Healey. Wearable and automotive systems for the recognition of affect from physiology. Media Arts and Sciences. Cambridge, Massachusetts Institute of Technology. Doctor of Philosophy, 2000.

[2] N. Sebe, E. Bakker, I. Cohen, T. Gevers and T. S. Huang. Bimodal emotion recognition. In Proceedings of the 5th International Conference on Methods and Techniques in Behavioral Research, 2005, August.

[3] M. Pantic and Leon J. M. Rothkrantz. Toward an Affect-Sensitive Multimodal Human-Computer Interaction. In Proceedings of the IEEE, 2003, 91(9) pp. 1370-1390.

[4] H. Hamdi, Multimodal platform for the recognition of emotions through the analysis of physiological signals: application to the simulation of job interviews. Modeling and Simulation. University of Angers, 2012. French.

[5] R. Sharma, V. Pavlovic and T. S. Huang. Toward multimodal human-computer interface. Proceedings of the IEEE,1998, 86(5), pp. 853-869.

[6] P. Teissier, P. Escudier and J.L. Schwartz, Automatic Processing of spoken language -2: speech recognition, multimodal speech chapter.2002, Hermes, Paris, pp. 141–178.

[7] Q. Zhi, M. N. Kaynak, K. Sengupta, A. D. Cheok, and C. C. Ko. Hmm modeling for audio-visual speech recognition. In In Proceedings of the IEEE International Conference on Multimedia and Expo (ICME '01), 2001, pp. 136.

[8] I. T. Meftah, Modeling, detection and annotation of emotional states using a multi-dimensional vector space. Artificial Intelligence. University of Nice Sophia Antipolis, 2013. French. <NNT : 2013NICE4017>.

[9] P.Niedenthal, J. Halberstadt, J. Margolin and A. Innes-ker, Emotional state and the detection of change in facial expression of emotion in European journal of social psychology March 9 2000, volume 30, issue 2.

[10] Busso, S. Lee and S. Narayanan, Analysis of Emotionally Salient Aspects of Fundamental Frequency for Emotion Detection in IEEE transactions on audio, speech and language processing , May 4 2009, vol.17 no. 4.

[11] K. H. Kim, S. W. Bang and S. R. Kim, Emotion Recognition System Using Short-Term Monitoring of Physiological Signals in Medical and Biological Engineering and Computing 2004, vol. 42 February 17 2004, pp. 419-427.

[12] J. Wagner, E. Andre, F. Lingenfelser and J. Kim, Exploring fusion methods for multimodal emotion recognition with missing data in Affective computing IEEE Transactions on vol. 2, issue 4, 12 January 2012, pp. 206-218.

[13] C. Zong and M. Chetouani, Hilbert Huang transform based Physiological signals analysis for emotion recognition in signal processing and information technology(ISSPIT), IEEE International Symposium on, December 14 2009, pp. 334-33.

[14] E. Monte-Moreno, M. Chetouani, M. Faundez-Zanuy and J. SoleCasals, Maximum likelihood linear programming data fusion for speaker recognition. Speech Communication, 51(9): 2009. 68, pp. 820–830.

[15] A. Mahdhaoui and M. Chetouani. Emotional speech classification based on multi view characterization. In Pattern Recognition (ICPR), 2010 20th International Conference IEEE, August 2010, pp. 4488-4491.

[16] A. Mahdhaoui, Social Signals Analysis for Modeling the interaction face to face.Signal and Image Processing. Pierre and Marie Curie University- Paris VI, 2010. French.

[17] I. M. Tayari, N. L. Thanh and C. Ben Amar, Detecting depression using a multidimensional model of emotional states: Global Health 2012, the first international conference on global health challenges, ISBN: 978-1-61208-243-1, pp. 101-107.

[18] E. Van't Hof, P. Cuijpers, W. Waheed and D. J. Stein, "Psychological treatments for depression and anxiety disorders in low- and middle- income countries: a meta-analysis." African Journal of Psychiatry, 2001, pp. 200–207.

[19] S. Hoch, F. Althoff, G. McGlaun and G. Rigoll. Bimodal fusion of emotional data in an automotive environment. In Acoustics, Speech, and Signal Processing, March 2005. Proceedings.(ICASSP'05). IEEE International Conference on (Vol. 2, pp. ii-1085).

 ISBN: 978-1-61208-424-4