Comparative analysis of physiological signals and Electroencephalogram (EEG) for multimodal emotion recognition using generative models

Cristian A. Torres-Valencia, Hernán F. García-Arias, Mauricio A. Álvarez López and Alvaro A. Orozco-Gutiérrez

***Abstract*—Multimodal Emotion recognition (MER) is an application of machine learning were different biological signals are used in order to automatically classify a determined affective state. MER systems has been developed for different type of applications from psychological evaluation, anxiety assessment, human-machine interfaces and marketing. There are several spaces of classification proposed in the state of art for the emotion recognition task, the most known are discrete and dimensional spaces were the emotions are described in terms of some basic emotions and latent dimensions respectively. The use of dimensional spaces of classification allows a higher range of emotional states to be analyzed. The most common dimensional space used for this purpose is the Arousal/Valence space were emotions are described in terms of the intensity of the emotion that goes from inactive to active in the arousal dimension, and from unpleasant to pleasant in the valence dimension. The use of physiological signals and the EEG is well suited for emotion recognition due to the fact that an emotional states generates responses from different biological systems of the human body. Since the expression of an emotion is a dynamic process, we propose the use of generative models as Hidden Markov Models (HMM) to capture de dynamics of the signals for further classification of emotional states in terms of arousal and valence. For the development of this work an international database for emotion classification known as Dataset for Emotion Analysis using Physiological signals (DEAP) is used. The objective of this work is to determine which of the physiological and EEG signals brings more relevant information in the emotion recognition task, several experiments using HMMs from different signals and combinations of them are performed, and the results shows that some of those signals brings more discrimination between arousal and valence levels as the EEG and the Galvanic Skin Response (GSR) and the Heart rate (HR).**

I. INTRODUCTION

Emotions are the reactions or perceptions that a person has of an specific situation. This reactions are observable in the behavior of the subject and affect different biological systems in the human body [1]. Having this in mind, multimodal emotion recognition is the task of assessing the emotional state of an individual using data from different biological signals that can be obtained from the subject [2]. The number of signals that are combined in order to asses emotions determine the mode of the system, being unimodal for one data source and multimodal emotion recognition is when several sources are combined [3].

The authors C.A. Torres-Valencia, H.F. García-Arias, M.A. Álvarez and A.A. Orozco-Gutiérrez are with the Research Group of Automatic, Department of Electrical Engineering, Faculty of Electrical Engineering, Universidad Tecnológica de Pereira, Pereira, Colombia, {cristian.torres, hernan.garcia, malvarez, aaog}@utp.edu.co

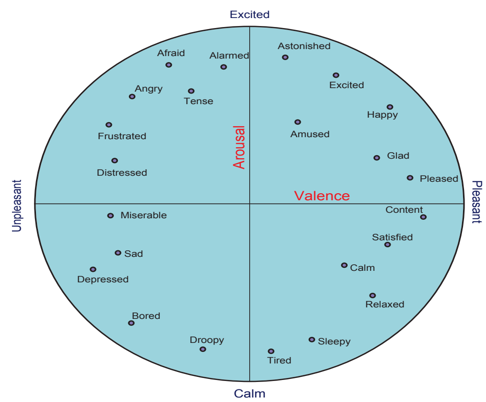


Fig. 1. Arousal/Valence space description of emotions

Emotion recognition has been an important research field in recent years, combining different areas of knowledge such as neurology, psychology and engineering. The applications of automatic emotion recognition vary from marketing, anxiety treatment to complex human-machine interfaces [4] [5]. The categories in which emotions can be discriminated are discrete classification spaces and dimensional spaces. Discrete spaces allows the evaluation of a few basic emotions as fear, happiness, sadness, defined by Eckman and are more suitable for unimodal systems [6]. Since there are several emotional states that can be assessed, a dimensional space allows the definition of a higher number of emotions from the combination of latent dimensions that allows a better understanding of the different states.

The arousal and valence space are the most used dimensional space for emotion recognition which define emotions in terms of the activation or non activation and the positiveness or negativeness of an emotional state [7]. This dimensions are usually used for describing emotions using machine learning algorithms that asses each dimension independently [7]. Classification and regression methodologies are well studied in the state of art of this field. The use of the arousal/valence space allows the description of basic emotions and several subtle emotional states that can be represented by a combination of this two dimensions [5]. Although there are some emotions that can be overlapped, the majority of emotional states can be well discriminated using this classification space. Figure 1 shows an example of the representation of emotions from the two latent dimension arousal and valence.

Generative models as Hidden Markov Models (HMM) are a machine learning algorithm that allows the representation of a dynamical signal into states that represent a particular class [8].

Since the emotions are intend to be assessed from biological signals, the HMM's are well suited to this task for capturing the dynamics of the signals that represents each emotional state. The use of HMM's with Gaussian mixtures allows the use of probabilistic theory in order to adjust the parameters of an HMM into a specific classification problem.

In recent years, the study of emotions has derived in the construction of few databases that allows the development of systems for emotional analysis, one of the database widely used for multimodal emotion recognition systems is the "Database for emotion analysis from physiological signals (DEAP)" [9]. This database allows the study of emotional states in the arousal/valence space from physiological signals and electroencephalogram.

The aim of this work is to develop a methodology for the study of multimodal emotion recognition in the arousal/valence space by using HMM as the machine learning algorithm from different combination of biological signals available in a well known database for emotion description.

II. PREVIOUS WORK

Generative models as HMM's has been widely used in the emotion recognition field from speech signals. In [10], a methodology for emotion recognition is developed combining the features extracted from the signal using wavelets to train HMM's for determining the emotional states of five discrete categories. Even the work was developed using only Chinese speech recordings the results show an effective performance and speed in the emotion recognition.

In 2012 a coupled HMM for audio-visual content analysis of emotions was developed by the definition of an adaboost classifier of HMM's. Expectation Maximization (EM)was used for the learning task and the system allows the recognizing of two emotional states effectively [11]. A similar work was presented by Chung-Hsien et. al. in 2013, present the use of HMM's in unimodal and bimodal emotion recognition form video and audio signals. The strategy propose the alignment of audio and video signals before the classifications is performed [12]. This work uses the semaine database that contains recordings of posed and naturalistic emotional experiments.

Other works that involves the use of HMM's for emotional speech recognition are reported. A methodology that includes the search of optimal features from the audio signals before classification using HMM's was presented in [13]. Another work that compares the use of HMM and SVM over the recognition of discrete emotions is presented in [14]. The comparative analysis of the results shows a more effective performance of the HMM against the SVM. Unimodal approaches are developed using HMM's from video signals in [15] and [16] by the analysis of facial expressions and fiducial points of the face.

On the multimodal emotion recognition field, several works has been developed by extracting some features from the biological signals that can be used for the training of a classifier or a regressor. This features can achieve emotion recognition in dimensional states in comparison with unimodal or bimodal systems which are developed from audio-visual signals and only are suitable for discrete emotional states. Statistical analysis is the most common feature extraction used in multimodal emotion recognition. Works like [17] presents a stage of feature extraction and feature selection in order to achieve an efficient performance of classification using SVM's for a reduced set of features. The majority of multimodal emotion recognition works are developed using EEG and other physiological signals as in [2] [18] [19].

III. MATERIALS AND METHODS

For this work all the algorithms are developed using MATLAB and the toolbox to work with HMM is the Hidden Markov Model (HMM) Toolbox for Matlab by Kevin Murphy [20].

A. Database

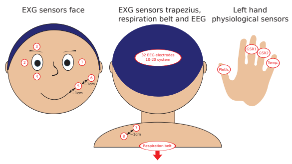
The database used in this work for the development of the methodology for emotion recognition is the Database for emotion analysis from physiological signals (DEAP) [9]. The DEAP database contains recordings of emotion elicitation experiments from 32 subjects. The emotion elicitation experiment induces an emotional state in a subject by watching a except from a video that has been previously categorized in one emotional state. While the subject is watching the video several signals are recorded [9]. The duration of the videos presented to each subject is one minute.

The full set of signals stored in DEAP database are the Electroencephalogram (EEG), electromyography (EMG), electrooculogram (EOG), galvanic skin response (GSR), Temperature (Temp), Respiration pattern (Resp), plethysmography (HR), audio and video signals. The signal recordings are stored with the information of the levels of arousal and valence for each realization of the experiment. Figure 2 shows an example of the recording methodology of the signals in an emotion elicitation experiment in the DEAP database [9].

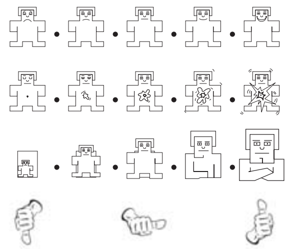
When a subject is watching one of the videos, a tagging system is used for assigning the corresponding arousal, valence, dominance and liking levels to the signal recordings in order to store the complete data in the database. Figure 2(b) shows the scheme used for capturing the subjects evaluation for the observed video. Figure 2(a) shows the location of the different sensors that allows the recording of the signals to be stored. Finally, figure 2(c) shows an example of one subject with the acquisition system ready for the experiment. The signals are preprocessed and stored with the labels into .mat files in order to work with them in MATLAB.

B. Hidden Markov Models

Hidden Markov models have been widely used for many classification and modeling problems. Perhaps the most common application of HMM is in speech recognition. One of the main advantages of HMMs is their ability to model non-stationary signals or events [14] [8]. Dynamic programming methods allow one to align the signals so as to account for the non-stationarity. However, the main disadvantage of this approach is that it is very time-consuming since all of the stored sequences are used to find the best match [8]. The HMM finds an implicit time warping in a probabilistic parametric fashion. It uses the transition probabilities between the hidden states and learns the conditional probabilities of



(a) Sensor location [9]



(b) Emotion tagging [9]

(c) Adquisition system [9]

Fig. 2. DEAP database experiment

the observations given the state of the model. In the case of emotion expression, the signal is the measurements of the facial motion. This signal is non-stationary in nature, since an expression can be displayed at varying rates, with varying intensities even for the same individual. An HMM is given by the following set of parameters:

$$\lambda = (A, B, \pi) \tag{1}$$

$$a_{ij} = P\left(q_{t+1} = S_j | q_t = S_i\right), \; 1 \leq i, j \leq N \tag{2}$$

$$B = \{b_j\left(O_t\right)\} = P\left(O_t | q_t = S_j\right), \; 1 \leq j \leq N \tag{3}$$

$$\pi_j = P\left(q_1 = S_j\right) \tag{4}$$

where A is the state transition probability matrix, B is the observation probability distribution and π is the initial state distribution. The number of states of the HMM is given by N. It should be noted that the observations (O_t) can be either discrete or continuous, and can be vectors. In the discrete case, B becomes a matrix of probability entries (Conditional Probability Table), and in the continuous case, B will be given by the parameters of the probability distribution function of the observations (normally chosen to be the Gaussian distribution or a mixture of Gaussians). Given an HMM there are three basic problems that are of interest. The first is how to efficiently compute the probability of the observations given the model. This problem is related to classification in the sense that it gives a measure of how well a certain model describes an observation sequence. The second is how, given

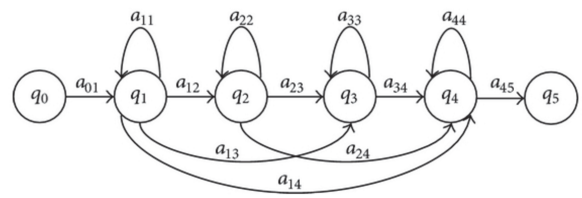


Fig. 3. Left to right HMM

a set of observations and the model, to find the corresponding state sequence in some optimal way [8]. The third one is how to learn the parameters of the model λ given the set of observations so as to maximize the probability of observations given the model. This problem relates to the learning phase of the HMMs. A comprehensive tutorial on HMMs is given by [8]. The scheme of the left to right HMM implemented in this work from the state transition matrix A is presented in figure 3.

C. Procedure

In order to develop a reliable methodology for emotion recognition, the first step is the division of the database from the information of the two dimensions, arousal and valence. The different realizations of the elicitation experiments of the database are divided in low and high arousal, low and high valence. Since the levels of the arousal and valence dimensions goes from 1 to 9, the low section is defined from 1 to 5 and the high portion of the data goes from 5 to 9. Each of this sets of signals are used for the training of a HMM, one for each level of arousal and valence.

Different HMM's are trained using few signal combinations in the multimodal emotion recognition approach. Each HMM is tested using crossvalidation by setting the 80% of the signals as training and the other 20% for test. This procedure is repeated 10 times to give an statistical validation to the obtained results.

This work presents the analysis of combination of biological signals as the EEG, EOG, EMG, and physiological signals. The results are compared in order to determine the better combination of biological signals into the emotion recognition problem. The combination of signals tested in this methodology are:

- EEG.
- EOG, EMG.
- EEG, EOG, EMG.
- Physiological (HR, GSR, Resp, Temp).
- EEG, physiological.
- EOG, EMG, Physiological.
- All signals.

IV. EXPERIMENTAL RESULTS

The results obtained for the classification using HMM's are presented in this section. This results correspond to the case where a single HMM is trained for one of the classes of arousal or valence. Table shows the results for the different combination of signals in the multimodal emotion recognition approach.

Table I shows the different classification accuracy (CA) obtained for different signal combinations of the database.

TABLE I. CA FOR THE AROUSAL DIMENSION

Arousal Dimension	
Signals	Classification Accuracy [%]
EEG	55.00 ± 4.5
EOG/EMG	56.25 ± 5.3
EEG/EOG/EMG	52.50 ± 4.6
Physiological	55.00 ± 3.9
EEG/Physiological	75.00 ± 3.7
EOG/EMG/Physiological	62.50 ± 4.3
All	56.25 ± 5.5

It can be seen for the arousal dimension in table I that the best CA is obtained when the EEG and the physiological signals are combined with a 75.00% of accuracy. Table II shows the CA for the valence dimension using the proposed methodology. It can be seen that the best results are obtained from the EEG signal and the combination of EEG, EMG and EOG.

TABLE II. CA FOR THE VALENCE DIMENSION

Valence Dimension	
Signals	Classification Accuracy [%]
EEG	58.75 ± 3.8
EOG/EMG	57.50 ± 4.3
EEG/EOG/EMG	58.75 ± 4.2
Physiological	57.50 ± 3.9
EEG/Physiological	57.50 ± 4.1
EOG/EMG/Physiological	53.75 ± 3.5
All	57.50 ± 4.5

From this results it can be seen that the data directly obtained from the body can actually classify different emotional states. Even the CA for the different experiments are not optimal, the use of the data as it was captured denotes a high complexity for the HMM's to model.

V. CONCLUSIONS AND DISCUSSION

The results shows that emotion recognition can be achieved using a multimodal approach from biological signals in order to recognize emotional states. In this work, feature extraction form the signals is not performed and the results suggest that the use of generative models as HMM's could been effectively used for this task.

From the results of the different sets of combined signals, it can be conclude that the combination of EEG with the physiological signals as GSR, HR, Temp and Resp, brings a higher discriminative information between different emotional states characterized in terms of arousal and valence. Other combination of signals brings closer accuracy rate demonstrating the capabilities of multimodal approach.

A correct feature extraction in combination with generative models could bring a higher performance in the classification of emotional states. A differentiation of more arousal and valence levels is the next step in the development of the methodology for emotion recognition based in generative models.

Future work could be for assessing emotional states in arousal/valence space by combining generative models as HMM's with discriminative models as SVM's which have been demonstrated as the state of art learning algorithm in several problems.

ACKNOWLEDGMENT

The authors of this work would like to thank the "Universidad Tecnológica de Pereira" for providing the resources needed to develop this research.

REFERENCES

[1] K. Scherer, "What are emotions? and how can they be measured?" *Social Science Information*, Jan. 2005. [Online]. Available: http://ssi.sagepub.com/cgi/content/abstract/44/4/695

[2] T. Pun, M. Pantic, and M. Soleymani, "Multimodal emotion recognition in response to videos," *IEEE Transactions on Affective Computing*, vol. 3, no. 2, pp. 211–223, 2012.

[3] S. Jerritta, M. Murugappan, R. Nagarajan, and K. Wan, "Physiological signals based human emotion recognition: a review," in *Signal Processing and its Applications (CSPA), 2011 IEEE 7th International Colloquium on*, 2011, pp. 410–415.

[4] S. Tokuno, G. Tsumatori, S. Shono, E. Takei, G. Suzuki, T. Yamamoto, S. Mituyoshi, and M. Shimura, "Usage of emotion recognition in military health care," in *Defense Science Research Conference and Expo (DSR), 2011*, 2011, pp. 1–5.

[5] Z. Zeng, M. Pantic, G. I. Roisman, and T. S. Huang, "A survey of affect recognition methods: audio, visual and spontaneous expressions," in *ICMI '07: Proceedings of the 9th international conference on Multimodal interfaces*. New York, NY, USA: ACM, 2009, pp. 126–133. [Online]. Available: http://dx.doi.org/10.1145/1322192.1322216

[6] P. Ekman, "Universals and cultural differences in facial expressions of emotions," vol. 19, pp. 207–283, 1972. [Online]. Available: http://www.paulekman.com/wp-content/uploads/2009/02/Universals-And-Cultural-Differences-In-Facial-Expressions-Of.pdf

[7] A. Hanjalic and L.-Q. Xu, "Affective video content representation and modeling," *Multimedia, IEEE Transactions on*, vol. 7, no. 1, pp. 143–154, 2005.

[8] L. Rabiner, "A tutorial on hidden markov models and selected applications in speech recognition," *Proceedings of the IEEE*, vol. 77, no. 2, pp. 257–286, Feb 1989.

[9] S. Koelstra, C. Muhl, M. Soleymani, J.-S. Lee, A. Yazdani, T. Ebrahimi, T. Pun, A. Nijholt, and I. Patras, "Deap: A database for emotion analysis ;using physiological signals," *IEEE Transactions on Affective Computing*, vol. 3, no. 1, pp. 18–31, 2012.

[10] H. Zhiyan and W. Jian, "Speech emotion recognition based on wavelet transform and improved hmm," in *Control and Decision Conference (CCDC), 2013 25th Chinese*, May 2013, pp. 3156–3159.

[11] K. Lu and Y. Jia, "Audio-visual emotion recognition with boosted coupled hmm," in *Pattern Recognition (ICPR), 2012 21st International Conference on*, Nov 2012, pp. 1148–1151.

[12] C.-H. Wu, J.-C. Lin, and W.-L. Wei, "Two-level hierarchical alignment for semi-coupled hmm-based audiovisual emotion recognition with temporal course," *Multimedia, IEEE Transactions on*, vol. 15, no. 8, pp. 1880–1895, Dec 2013.

[13] R. Bock, D. Hubner, and A. Wendemuth, "Determining optimal signal features and parameters for hmm-based emotion classification," in *MELECON 2010 - 2010 15th IEEE Mediterranean Electrotechnical Conference*, April 2010, pp. 1586–1590.

[14] Y.-L. Lin and G. Wei, "Speech emotion recognition based on hmm and svm," in *Machine Learning and Cybernetics, 2005. Proceedings of 2005 International Conference on*, vol. 8, Aug 2005, pp. 4898–4901 Vol. 8.

[15] G. Lv, S. Hu, Y. Fan, and M. Qi, "Visual emotion recognition based on dynamic models," in *Signal Processing, Communication and Computing (ICSPCC), 2013 IEEE International Conference on*, Aug 2013, pp. 1–4.

[16] Y. Guo and H. Gao, "Emotion recognition system in images based on fuzzy neural network and hmm," in *Cognitive Informatics, 2006. ICCI 2006. 5th IEEE International Conference on*, vol. 1, July 2006, pp. 73–78.

[17] C. Torres, A. Orozco, and M. Alvarez, "Feature selection for multimodal emotion recognition in the arousal-valence space," in *Engineering in Medicine and Biology Society (EMBC), 35th Annual International Conference of the IEEE*, July 2013, pp. 4330–4333.

[18] K. Takahashi, S. Namikawa, and M. Hashimoto, "Computational emotion recognition using multimodal physiological signals: Elicited using japanese kanji words," in *Telecommunications and Signal Processing (TSP), 2012 35th International Conference on*, 2012, pp. 615–620.

[19] M. Nicolaou, H. Gunes, and M. Pantic, "Continuous prediction of spontaneous affect from multiple cues and modalities in valence-arousal space," *Affective Computing, IEEE Transactions on*, vol. 2, no. 2, pp. 92–105, April 2011.

[20] K. Murphy, "Hidden Markov Model (HMM) Toolbox for Matlab," urlhttp://www.cs.ubc.ca/˜murphyk/Software/HMM/hmm.html.