Evaluation of Influence of Arousal-Valence Primitives on Speech Emotion Recognition

Imen Trabelsi[1], Dorra Ben Ayed[2], and Noureddine Ellouze[2]
[1]Sciences and Technologies of Image and Telecommunications, Sfax University, Tunisia
[2]Ecole Nationale d'Ingénieurs de Tunis, Université Tunis-Manar, Tunisia

Abstract: *Speech Emotion recognition is a challenging research problem with a significant scientific interest. There has been a lot of research and development around this field in the recent times. In this article, we present a study which aims to improve the recognition accuracy of speech emotion recognition using a hierarchical method based on Gaussian Mixture Model and Support Vector Machines for dimensional and continuous prediction of emotions in valence (positive vs negative emotion) and arousal space (the degree of emotional intensity). According to these dimensions, emotions are categorized into N broad groups. These N groups are further classified into other groups using spectral representation. We verify and compare the functionality of the different proposed multi-level models in order to study differential effects of emotional valence and arousal on the recognition of a basic emotion. Experimental studies are performed over the Berlin Emotional database and the Surrey Audio-Visual Expressed Emotion corpus, expressing different emotions, in German and English languages.*

Keywords: *Speech emotion recognition, arousal, valence, hierarchical classification, gaussian mixture model, support vector machine.*

Received September 3, 2015; accepted March 30, 2016

1. Introduction

With the rapid development of pattern recognition and affective computing, Speech Emotion Recognition (SER) is attracting more and more attention from the researchers. SER is developing in different directions. The developers of emotion recognizers use different types of databases of acted, induced or spontaneous emotion, different feature sets, and a wide range of pattern recognition methods. The emotions labels are described either using emotion dimension or emotion category. Many researchers especially from the field of engineering have always regarded emotion as a discrete class that differs explicitly and mutually exclusive from one emotion to another. On the contrary to the idea that emotion is a discrete class, some researchers later argued that emotion is a continuous process that will changes dynamically over time, using the multi-dimensional emotion model. Emotional dimensionality is a simplified description of the basic properties of emotional states. According to this, we selected six basic emotions of anger, joy, fear, boredom, disgust and sadness as well as neutral as emotionless state. Based on the emotions selected, the aim of this paper is to develop hierarchical-based approaches, modeling emotions on a 2D affective space. It solves N-class problems in several stages consisting of GMM/SVM-based classification. The multi-levels models were used to examine differential effects of emotional valence and arousal on the performance of SER. The evaluation presented in this work has the following parameters:

1. 7 emotions from Berlin database of Emotional speech (Emo-Db) and 6 emotions from Surrey Audio-Visual Expressed Emotion (Savee audio-Db).
2. Text-independency and speaker-independency were assumed.
3. Feature vectors were extracted on frame level. The rest of this paper is organized as follows.

Emotion modeling is briefly described in section 2. GMM N-stage classifications are presented in Section 3. The experimental results are discussed in section 4. Final conclusions are made in section 5.

2. Describing Emotional Content

Human emotions are created every time a perception of important changes in the external/internal environment appears. Basically, an emotion is a psychological state or process that functions in the management of maintaining the balance of information processes in the brain. There is a large volume of published studies available on emotional speech modeling [4]. Emotion models and descriptions are usually adopted from psychology research. Three major approaches can be distinguished [5]:

1. Categorical approach.
2. Dimensional approach.
3. Appraisal-based approach.

The categorical approach is based on research on discrete emotions [7]. According to this approach, all emotions are classified into two main categories: primary (basic) and secondary (derived) emotions [5].

Primary or basic emotions generally could be experienced by all social mammals (e.g., humans, monkeys, whales) and have particular manifestations associated with them (e.g., vocal/facial expressions, behavioral tendencies, and physiological patterns). The six basic emotions according to Ekman's theory are joy, sadness, surprise, fear, anger and disgust [4]. Secondary or derived emotions are derived from a limited number of universal and innate basic emotions, each related to a biologically scenario [4, 5]. The number of these emotions varies between six and twenty-one. However, in the dimensional approach, affective states are not independent from one another; rather, they are blended and overlapping. This idea has been supported in many current direction of speech emotion research using the multi-dimensional emotion model [14]. In this approach, the two variables that are used most often to measure emotions are: valence and arousal [11]. The valence dimension (also referred to as evaluation or pleasure) refers to how positive or negative the emotion is, and ranges from unpleasant feelings to pleasant feelings of happiness. The arousal dimension (also referred to as intensity or activity) refers to how excited or apathetic the emotion is, and it ranges from drowsiness to frantic excitement. Feelings with high arousal induces some greater sub-glottal pressure resulting in change in speech as well such as making it louder, faster and have higher average pitch etc. Terrified is a highly active emotional state and disinterested is an emotional state with low activation. At the center of the Arousal-Valence space is the point which corresponds to emotional neutrality. There's also a broad agreement that a third dimension of emotion exists, however with unclear definitions [1]. This dimension represents he dominance or power dimension. Arousal-Valence space can be modeled as illustrated in Figure 1. In this way, it is possible to characterize all emotions by their valence and arousal, and different emotional labels could be plotted at various positions on this two-dimensional plane. The valence and arousal model is a very efficient way of describing emotions and therefore represents a very popular choice for the implementation of emotional related systems. A third type of emotion model, initiated by Scherer *et al.* [14], is based on appraisal theory and introduces the role of time into the comprehension of emotions. The main assumption here is that an emotion is a reaction (internal state) to outside situations and events that are being evaluated at the cognitive level by the human. The appraisal based approach is an open research question that is rarely used in automatic affect recognition.

There exist a number of works focusing on how to map emotional expression to dimensional models. Research on affective theories showed that valence and arousal are two basic underlying dimensions of emotion. Cowie and colleagues used valence-arousal space to model and assess affect from speech [3]. Similar work was presented in [18], where the authors developed acoustics and lexical classifiers to assess the separability on arousal and valence dimensions in spontaneous speech. Later in [6], the authors evaluate the appraisal-based theory to judge emotional effects on vocal expression. In [12], the paper deals with an application for emotion detection in usability testing of software using the valence -arousal space for emotion modeling in a formal experiment. In [9], a continuous dimensional speech emotional recognition model is designed that combines acoustic and semantic features.

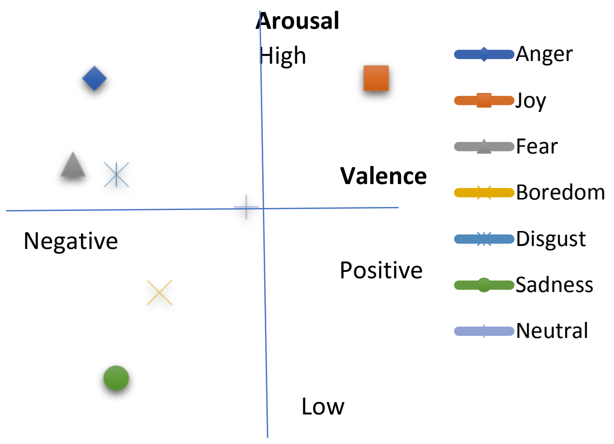

Figure 1. Arousal-valence space.

The authors show that valence is better estimated using semantic features while arousal is better estimated using acoustic features. In [10], the authors evaluate the emotion task over the nine categories including the binary valence-arousal discrimination. They further investigate the discrimination of each emotion against neutral. The results show performances for arousal and valence of up to 86.5% and for nine emotions including neutral of up to 42% unweighted average recall

3. Methodology

3.1. Speech Emotion Recognition-Based GMM

Gaussian mixture model is a probabilistic model for density estimation using a convex combination of multi-variate normal densities. Let x be a dimensional random vector that has an arbitrary distribution. A GMM describes the distribution of x defined in Equation (1).

$$P(x|\,\lambda) = \sum_{i=1}^{N} Pi\ bi\ (x) \qquad (1)$$

bi(x) are the component densities and pi the mixture weights. Each component density is a d-variate gaussian function having the form:

$$bi(x) = \frac{1}{(2\pi)^{\frac{d}{2}} \quad |\Sigma\, i|^{\frac{1}{2}}} exp\left[-\frac{1}{2}\,(x-\mu_i)^T\,\Sigma_i^{-1}(x-\mu_i)\right] \qquad (2)$$

With mean μ_i and covariance matrix $\sum_i$.

3.2. Speech Emotion Recognition -Based Support Vector Machine

Currently, Support Vector Machine (SVM) is one of the most powerful discriminative classifiers [13, 15]. SVM is a two-class classifier that maximizes the distance between nearest points of the two classes. Kernel function *K(x, y)* constructs a mapping into a high-dimensional feature space.

$$f(x) = \sum_{i=1}^{N} \propto_i \; y_i K(x, x_i) + d \tag{3}$$

$$K(x, y) = \Phi(x)^t \Phi(y) \tag{4}$$

Where y_i is the ideal output of the class (1 or 1), d is a constant, x_i is support vectors trained by the optimization algorithm and Φ is a mapping from the input space to a possible infinite-dimensional space. The kernel function has to be constrained by Mercer conditions to assure the convergence of the SVM training.

3.3. GMM Super Vector Kernel for Speech Emotion Recognition

First, a Universal Background Model (UBM) is learned with multiple audio files from all training different emotions. The UBM is trained with the Expectation-Maximization (EM) algorithm [4] on its training data. From this initial model, an adapted GMM is created for each emotional class by Maximum A Posteriori (MAP) estimation, allowing for prior distribution to be incorporated in the final estimation process. This allows a detailed model to be trained when little data is available, which is often the case when a large number of parameters are estimated. Only the mean vectors are adapted while the covariance matrices and weights remain unchanged. The prior distribution for this estimation is determined by the UBM parameter and a factor τ governing the influence or relevance of the UBM on the final emotion model. From the adapted GMM, the final GMM supervector is constructed as the representation of the input utterance. The supervector of a GMM is defined by concatenating the mean of each gaussian mixture, which can be thought of as a mapping between an utterance and a high-dimensional vector. If we have a Gaussian Mixture Model (GMM) trained with multiple audio files from all training different emotions by EM algorithm.

$$p(x) = \sum_{i=1}^{N} \lambda_i \, b_i \,(x; \mu_i; \textstyle\sum_i) \tag{5}$$

Where $p(x)$ is a GMM, λi, μi and $\sum i$ are the mixture weight, mean and diagonal covariance of the GMM.

$$px(x) = \sum_{i=1}^{N} \lambda_i \, b_i \,(x; \mu_i^y; \textstyle\sum_i) \tag{6}$$

$$py(x) = \sum_{i=1}^{N} \lambda_i \, b_i \,(x; \mu_i^y; \textstyle\sum_i) \tag{7}$$

Given a emotional utterance, whether *utt(x)* or *utt(y).* Two GMMs, *px(x)* and *py(x)*, are trained by the two utterances by MAP estimation of *p(x)* allowing for prior distribution to be incorporated in the final estimation process. This allows a detailed model to be trained when little data is available, which is often the case when a large number of parameters are estimated. Only the mean vectors are adapted while the covariance matrices and weights remain unchanged. The Kullback-Leibler divergence distance between them is:

$$\mathrm{Kl}(\mathrm{px(x)}||\mathrm{py(x)})) \leq \sum_{m=1}^{M} w_m \, \mathrm{KL}(\mathrm{N}(\mathrm{x}|\mu_m^x \textstyle\sum_m)||\mathrm{N}(\mathrm{x}|\mu_m^y \textstyle\sum_m)\,) \tag{8}$$

Considering diagonal covariances, expanding the distance between two Gaussian *D* $(\mu^x{}_m, \sum_i)$ *and N* $(\mu_y{}^m, \sum_i)$; the closed form of Equation (8) is:

$$\mathrm{D}(\mu_m^x, \mu_m^y\,) = \frac{1}{2} \sum_{m=1}^{M} w_m (\mu_m^x \;-\; \mu_m^y)^t \textstyle\sum_m^{-1} (\mu_m^x \;-\mu_m^y) \tag{9}$$

$$K(\lambda_{X\,,} \lambda_Y\,) = \sum_1^M (\sqrt{w_m} \textstyle\sum_m^{\frac{-1}{2}} \mu_m^X)^T \; (\sqrt{w_m} \textstyle\sum_m^{-1/2} \mu_m^y) \tag{10}$$

$$= \emptyset_{NSL}\,(X) A_{NSL}^{-1} \, \emptyset_{NSL}(Y) \tag{11}$$

In Equation (10), $\mu^x{}_m$ is the supervector of the utterance x, expanded to the GMM supervector space. The feature space represents a simple diagonal scaling using $(\sqrt{wm} \sum^{-(1/2)}{}_m)$ of the GMM supervectors space. As a consequence, it satisfies the Mercer conditions. We get the final SVM model to classify emotional utterances.

4. The Framework of the Proposed Hierarchical Affective Speech Recognizer

A hierarchical recognition methodology based on the GMM/SVM is adopted. This hybrid system is used in each level in order to evaluate the accuracy performance of the emotional states. The basic idea here is that emotional states can, in the first stage, be categorized into some broad and rough emotional groups. Therefore, 2-stage and 3-stage hierarchical classification systems are proposed. To illustrate this, we built three systems with different configurations: According to arousal dimension, the emotions in speech can be divided into 3 states: High arousal level state, low arousal level state and neutral level state. High arousal emotions include anger, happiness, fear and disgust whereas low arousal emotions consist of boredom and sadness. Then, in the second stage, each broad emotional class can then be further classified into final discrete states. Our hierarchical (two-stage) classification scheme is presented in Figure 2. According to valence dimension, the emotions are divided into 3 states: Positive valence state, negative valence state and neutral level state (Figure 3). Then, in the second stage, the negative valence state is classified into 5 emotional classes: anger, boredom, disgust, fear and sadness.

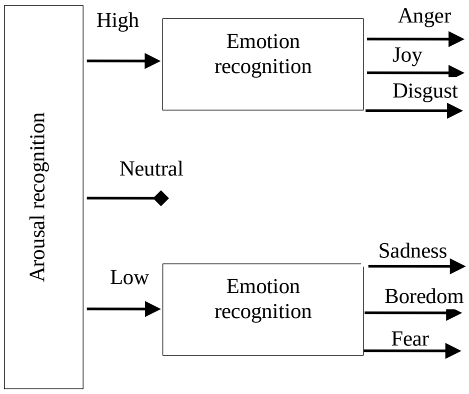

Figure 2. Proposed classification framework based on arousal prim
itive.

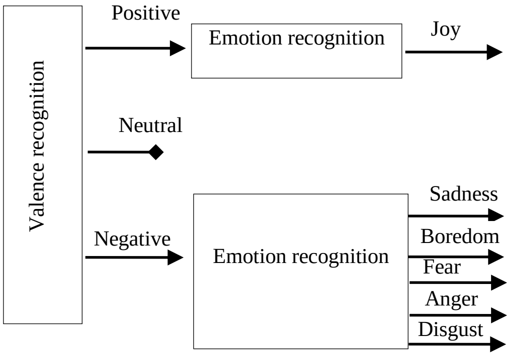

Figure 3. Proposed classification framework based on valence primitive

5. Speech Emotion Recognition Experiments and Results

5.1. Description of the Used Speech Corpora

EMO-DB [2] is recorded by speech workgroup leaded in the anechoic chamber of the Technical University in Berlin. It is a simulated open source speech database. In this database, ten professional native German actors (5 female and 5 male) simulated 7 emotions, producing 10 utterances. 5 utterances are short, while the remaining 5 are long. The emotions are: anger, boredom, disgust, fear, happiness, sadness, and neutral. This emotional speech corpus is probably the most often used database in the context of emotion recognition from speech, and also one of the few for which some results can be compared.

Savee audio-Db was recorded from four native English male speakers, reading sentences in a controlled environment in seven emotions: anger, disgust, fear, happiness, neutral, sadness and surprise. The text material consisted of 15 phonetically-balanced TIMIT sentences per emotion: 3 common, 2 emotion-specific and 10 generic sentences that were different for each emotion. The 3 common and 2 emotion specific sentences were recorded in neutral emotion, which resulted in 30 sentences for neutral emotion. This database consists of 120 utterances per actor, which resulted in 480 sentences in total. Table 1 shows the properties of Emo-DB and Savee datasets.

Table 1. Properties of Emo-Db and savee where N and L indicates, number of utterances and length of utterances in seconds, respectively.

Emotion	N (EmoDb)	N (Savee)	L (EmoDb)	L (Savee)
Anger	127	60	335.3	222,8
Boredom	81	-	180.6	-
Disgust	46	60	251.2	237,2
Fear	69	60	225	224,8
Happy	71	60	154.2	228,2
Neutral	79	120	154.1	432,6
Sad	62	60	186.3	268,8

5.2. System Description

The Emo-Db speech data were recorded at a sample rate of 16 kHz and a resolution of 16 bits whereas Savee audio-Db speech data were recorded at a sample rate of 44,1 kHz and a resolution of 16 bits. First, the signal is passed through a pre-processing system that normalizes amplitude, reduces the amount of noise and extracts combined MFCC parameters, where Mel filter banks are placed in [20-3000] Hz [16, 17]. MFCCs have been widely used in speech recognition because of superior performance over other features [11, 12]. In our experiments, the data in each case were divided into two sets. 70 % of the samples are used as training set and the other 30 % samples are used as testing set.

5.3. GMM/SVM N-Stage Classification Results

The first group of results consists of obtained values from arousal hierarchical classification. In arousal classification, our work differentiates the 3 arousal levels and the accuracy of 85.82% in Emo-db and 76.66 % in Savee audio-Db were obtained in the first stage Table 2. From this, we can conclude that recognizing high aroused speech seems to be easier than recognizing low aroused and neutral speech in Emo-Db. Nevertheless, in Savee audio-DB corpus, low aroused speech is recognized more easily than high aroused emotional speech.

Table 2. Arousal classification results.

Emotion	Emo-Db	Savee audio-Db
High arousal	100	72,5
Low arousal	86,04	82,5
Neutral arousal	71,42	75

In the second stage, the hierarchical structure shows a good discriminability for high aroused emotions and low aroused emotions in Emo-Db. However, it is completely confusable for high aroused emotions in Savee audio-Db corpus. Low aroused emotions are easy classified. Table 3 shows the recognition rates obtained on Emo-Db and Savee audio-Db for hierarchical arousal based classification. The second group of results consists of obtained values from valence hierarchical classification. In valence classification, positive, negative and neutral levels were recognized in the first level and the accuracy of 74.42% for Emo-Db and 67.91% for Savee audio-Db were obtained, as it can be seen in Table 4. Table 5 corresponds to the recognition rates obtained on the two studied corpora for hierarchical valence based classification.

Table 3. Recognition rates for arousal based classification.

Emotion	Emo-Db	Savee audio-Db
Anger	92,1	61,53
Boredom	75	-
Disgust	100	93,33
Fear	85,7	46,66
Joy	66,66	61,53
Neutral	71,42	75
Sadness	100	100
UAR(%)	84,41	73

Table 4. Valence classification results.

Emotion	Emo-Db	Savee audio-Db
Positive valence	76,19	50
Negative valence	73,91	71,25
Neutral valence	71,42	82,5

Table 5. Final Recognition rates for valence based classification.

Emotion	Emo-Db	Savee audio-Db
Anger	96,87	53,33
Boredom	100	-
Disgust	100	93,75
Fear	100	85,71
Joy	76,19	50
Neutral	71,42	82,5
Sadness	100	100
UAR (%)	92,06	77,54

Finally–as it can be seen from the summary results in Tables 6 and 7, realized comparison of emotion classification results for sentences from the Emo-Db and the Savee audio-Db show that results of Emo-Db are better compared to those of Savee audio-Db, due to the high speech noisy Savee audio-Db utterances. The overall recognition obtained using hierarchical arousal based classification on Emo-Db was 81.21% and 61.78 % on Savee audio-Db Corpus. For GMM valence based classification, the overall accuracy rate is 73.32 % for Emo-DB and 66.66 % for Savee audio-Db Corpus. Realized comparisons show that better results among the different employed methods and in the two corpora were achieved for the emotions of joy, disgust and sadness. It is also interesting to see, that the ability to recognize anger, disgust and sadness is even better than the accuracy furnished by human recognizers in Emo-Db.

Table 6. Comparison of employed methods for emo-Db.

Emotion	Arousal-based classification	Valence-based classification	Human recognition
Anger	92,1	81,57	96,9
Boredom	66,66	66,66	88,2
Disgust	100	84,61	87,3
Fear	80,95	52,38	86,2
Joy	66,66	76,19	83,7
Neutral	71,42	71,42	80,7
Sadness	94,73	78,94	79,6

Table 7. Comparison of employed methods for Savee audio-Db.

Emotion	Arousal-based classification	Valence-based classification
Anger	35	40
Disgust	60	60
Fear	60	75
Joy	40	50
Neutral	75	82,5
Sadness	70	60

6. Conclusions

In the field of speaker recognition, systems based on GMM/SVM often have superior performance than those based on GMM or SVM. Although this article focuses on recognizing emotions based on GMM/SMM. We investigate the feasibility of emotion recognition based on two primitive attributes valence and arousal. We used multilevel modeling to examine how emotional valence and arousal attributes influence the relationship between emotional states. We define several N-stage systems to recognize emotions from the actor recorded simulated databases, Emo-Db and Savee audio-Db. Our main difficult in this task comes from the fact that there is a lack of agreement among psychological research. In fact, modeling emotions continuously using the dimensions of arousal and valence is a complex problem as these dimensions are not universally recognized and decoded even by human observers. We note that automatic analysis results for arousal are remarkably better than those for valence. It seems that the perception of arousal is more universal than is the perception of valence. In near future, we envisage to further validate our approach by considering other datasets. We will study the distribution of the emotional utterances on the dimension of power. Furthermore, we hope to describe the emotion space model with a reasonable mathematic method in order to analysis the relationship between emotions. Emotion recognition used in real world will be expected soon.

References

[1] Baron-Cohen S., *Mind Reading: The Interactive Guide to Emotion*, Cdr, 2004.

[2] Burkhardt F., Paeschke A., Rolfes M., Sandlmeier W., and Weiss B., "A Database of German Emotional Speech," *in Proceedings of Interspeech Lisbon*, Portugal, pp. 1517-1520, 2005.

[3] Cowie R., Douglas-Cowie E., Tsapatsoulis N., Votsis G., Kollias S., Fellenz W., and Taylor J., "Emotion Recognition in Human-Computer Interaction," *IEEE Signal Processing Magazine*, vol. 18, no. 1, pp. 33-80, 2001.

[4] Dempster A., Laid N., and Durbin D., "Maximum Likelihood from Incomplete Data via the EM Algorithm," *Journal of the Royal Statistical Society*, vol. 39, no. 1, pp. 1-38, 1977.

[5] Ekman P., Friesen W., and Ellsworth P., *Emotion in the human face: Guidelines for Research and an Integration of Findings*, Elsevier, 2013.

[6] Ekman P. and Friesen W., "Head and Body Cues in the Judgment of Emotion: A Reformulation," *Perceptual and Motor Skills*, vol. 24, no. 3, pp. 711-724, 1967.

[7] Grandjean D., Sander D., and Sherer K., "Conscious Emotional Experience Emerges as a Function of Multilevel, Appraisal-Driven Response Synchronization," *Consciousness and Cognition*, vol. 17, no. 2, pp. 484-495, 2008.

[8] Haq S. and Jackson P., *Multimodal Emotion Recognition. Machine Audition: Principles, Algorithms and Systems*, IGI Global Press, 2010.

[9] Karadogan S. and Larsen J., "Combining Semantic and Acoustic Features for Valence and Arousal Recognition in Speech," *in Proceedings of IEEE 3rd International Workshop on*: *Cognitive Information Processing*, Baiona, pp. 1-6, 2012.

[10] Kamaruddin N. and Wahab A., "Human Behavior State Profile Mapping Based on Recalibrated Speech Affective Space Model," *in Proceedings of the 34th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, San Diego, pp. 2021-2024, 2012.

[11] Mehrabian A. and Russell J., *an Approach to Environmental Psychology*, MIT Press, 1974.

[12] Marchi E., Schuller B., Batliner A., Fridenzon S., Tal S., and Golan O., "Emotion in the Speech of Children with Autism Spectrum Conditions: Prosody and Everything else," *in Proceedings of 3rd Workshop on Child, Computer and Interaction*, pp. 17-24, 2012.

[13] Mittal T. and Sharma R., "Multiclass SVM based Spoken Hindi Numerals Recognition," *The International Arab Journal of Information Technology*, vol. 12, no. 6A, pp. 666-671, 2015.

[14] Sherer K., Schorr A., and Johnstone T., *Appraisal Processes in Emotion: Theory, Methods, Research*, Oxford University Press, 2001.

[15] Trabelsi I. and Ben-Ayed D., "A Multi Level Data Fusion Approach for Speaker Identification on Telephone Speech," *International Journal of Signal Processing, Image Processing and Pattern Recognition*, vol. 6, no. 2, 2013.

[16] Trabelsi I. and Ben Ayed D., "On the Use of Different Feature Extraction Methods for Linear and Non Linear Kernels," *in Proceedings of IEEE 6th International Conference on Sciences of Electronics, Technologies of Information and Telecommunications*, Sousse, pp. 1-11, 2012.

[17] Trabelsi I., Ben-Ayed D., and Ellouze N., "Improved Frame Level Features and SVM Supervectors Approach for the Recognition of Emotional States from Speech: Application to categorical and dimensional states," *International Journal of Image, Graphics and Signal Processing*, vol. 5, no. 9, pp. 8-13, 2013.

[18] Truong K. and Raaijmakers S., "Automatic Recognition of Spontaneous Emotions in Speech using Acoustic and Lexical Features," *Machine Learning for Multimodal Interaction*, Utrecht, pp. 161-172, 2008.

Imen Trabelsi received her MS degree in signal processing in 2011 from the Institute of Computer Science of Tunis (ISI-Tunisia) and PhD degree in electrical engineering with specialization in signal processing in 2015 from the University of Tunis-El Manar (Tunisia). Her main areas of interests include: speech processing, pattern recognition, machine learning, artificial intelligence and emotion recognition. She has published research papers at international journals and conference proceedings.

Dorra Ben Ayed received computer science engineering degree in 1995 from the National School Computer Science (ENSI-Tunisia), the MS degree in electrical engineering (signal processing) in 1997 from the National School of Engineer of Tunis (ENITTunisia), the Ph.D. degree in electrical engineering (signal processing) in 2003 fro m (ENIT-Tunisia). She is currently an associate professor in the computer science department at the High Institute of Computer Science of Tunis (ISI-Tunisia). Her research interests include fuzzy logic, support vector machines, artificial intelligence, pattern recognition, speech recognition and speaker identification.

Noureddine Ellouze received a PhD degree in 1977 from l'Institut National Polytechnique at Paul Sabatier University (Toulouse, France), and Electronic Engineer Diploma from ENSEEIHT in 1968 at the same university. In 1978, Dr Ellouze joined the Department of Electrical Engineering at the National School of Engineer of Tunis (ENIT-Tunisia), as Assistant Professor in Statistic, Electronic, Signal Processing and Computer Architecture. In 1990, he became Professor in Signal Processing, Digital Signal Processing and Stochastic Process. He has also served as Director of Electrical Department at ENIT from 1978 to 1983, General Manager and President of the Research Institute on Informatics and Telecommunications (IRSIT) from 1987 to 1990, President of the same Institute from 1990 to 1994. He is now Director of Signal Processing Research Laboratory (LSTS) at ENIT and is in charge of Control and Signal Processing Master degree at ENIT. Pr Ellouze is IEEE fellow since 1987, he directed multiple Master thesis and PhD thesis and published over 200 scientific papers in journals and conference proceedings. He is chief editor of the scientific journal Annales Maghrébines de l'Ingénieur. His research interests include Neural Networks and Fuzzy Classification, Pattern Recognition, Signal Processing and Image Processing applied in biomedical, Multimedia, and Man Machine Communication.