International Journal of Computing and Digital Systems
ISSN (2210-142X)
Int. J. Com. Dig. Sys. 10, No.1 (Apr-2021)

http://dx.doi.org/10.12785/ijcds/100154

Recognition of Anger and Neutral Emotions in Speech with Different Languages

Houari Horkous[1] and Mhania Guerti[1]

[1]*Laboratory signal and communication, Ecole National Polytechnique, Algiers 16200, Algeria*

Received 23 Jul.2020, Revised 17 Sep. 2020, Accepted 29 Oct. 2020, Published 21 Apr. 2021

Abstract: Speech emotion recognition is a very interesting area. It has board applications in man-machine interaction. In this work, the influence of speech features on the recognition of anger and neutral emotions in different languages is studied. And on the other hand, the influence of anger and neutral emotions for classifying male and female gender in different languages is also studied. Four databases in different languages are used to achieve our purpose. These databases are Algerian Dialect Emotional Database (ADED), Berlin Database of Emotional Speech (EMO-DB), Sharif Emotional Speech Database (ShEMO) and Crowd-sourced Emotional Multimodal Actors Dataset (CREAMA-D). The databases are exploited for extracting the features that used in the recognition and classification systems. The features extracted are the pitch, intensity, formants and MFCCs (Mel Frequency Cepstral Coefficients) parameters. The results obtained show us that the use a combination of features improve the performance of recognition in all the databases. It was showed also in the results that the classification of gender classes is influenced by the type of emotion and the language of databases.

Keywords: Emotion**,** ADED, EMO-DB, ShEMO, CREAMA-D, Pitch, Iintensity, Formants, MFCCs

1. Introduction

Emotion has an important role in human communication. The human-machine interface becomes more significant if the machines can recognize the emotions. Recognition of emotion can be done in different ways: facial expressions, brain signals, speech, etc. Speech emotion recognition (SER) has wide applications. SER has board applications in man-machine interaction such as intelligent tutoring system, lie detection, telephone banking, in call center, robots, sorting of voice mail and computer games [1]. SER has been exploited in the medical field [2-3]. In the field of psychology, SER had many applications for example: detection of moods [4], discriminating depressed speech [5] and analyzing human behavior [6].

SER system has been defined as the processes that classify speech to detect emotions [7]. For each system of SER, there are different emotional databases have been created. From these databases, many speech features have been extracted. These features have been exploited to recognize and classify the specific emotions. Numerous databases have been used in the system of SER. Among the most used databases are the Berlin database (Emo-DB) [8], the Danish (DES) database [9], the database of Polish emotional speech [10] and SAVEE database [11]. Extraction of the speech features is an important step in SER systems. Very large number of works has been concentrated on the prosodic features such as pitch, intensity and duration and the early studies in this field have been focused on these features [12-14]. Spectral features have been widely used for recognition of emotions in speech, and the most exploited spectral parameters are the MFCCs [15-16]. In many researches, voice quality features such as HNR, jitter, shimmer have been used in SER systems [17-19]. Combination of different features has been used to improve the performance of SER [20-21]. Several classifiers have been applied for SER such as Support Vector Machines (SVM) [22], K-nearest Neighbors (KNN) [22-24], Gaussian Mixture Models (GMMs) [25], and Deep Neural Network (DNN) [24].

In this paper, the influence of speech features on the recognition of anger and neutral emotions in different languages is studied. On the other hand, the influence of anger and neutral emotions for classifying the gender is also studied. So the contribution of this work is to compare the recognition of emotions and the classification of the gender in the Algerian dialect with other languages. But we based only on anger and neutral

E-mail: houari.horkous@g.enp.edu.dz, mhaniag@yahoo.fr

states. Anger has semantic, conceptual, and empirical links to psychopathology [26]. Anger occurs in response to perceived threats or injustices when there is someone or something to blame [27]. To achieve the purpose of our work, four databases in different languages are used: Algerian Dialect Emotional Database (ADED), Berlin Database of Emotional Speech (EMO-DB), Sharif Emotional Speech Database (ShEMO) and Crowd-sourced Emotional Multimodal Actors Dataset (CREAMA-D). Classical speech features including prosodic features (pitch and intensity), formants and MFCCs parameters are used in the systems of recognition and classification in this work. The systems are based on SVM as classification technique.

This document is structured as follows. Some of related works are discussed in the second section. Our methodology is explained in the third section, emotional speech databases, speech features and classifier used in this work are described in this section. Results of experiment are discussed in the fourth section. We finish by conclusion.

2. RELATED WORKS

In the SER systems, there are different types of emotional speech databases have been built and developed. Many speech features have been extracted and evaluated to improve the performance of recognition. And several classifiers have been used to classify the emotions [28]. In this section some databases, features and classifiers used in SER are discussed in brief.

Collection of databases is important for the SER systems. So to assess the performance of SER systems, it is essential to build an appropriate database. There are famous databases in the SER field such as EMO-DB contains around 535 utterances spoken in seven emotional states [8]. Danish (DES) database contains five emotions: neutral, angry, happy, sad and surprise [9]. The database of Polish emotional speech composed of 288 speech segments in six emotions [10]. Surrey audio-visual expressed emotion (SAVEE) database composed of speech utterances in seven emotions recorded in native English [11]. There are many other databases in different languages for example: ShEMO database in Persian language [29], CREAMA-D in English language [30], CEMO in French language [31], IITKGP-SEHSC in Indian language [32], emotional speech database in Slovene and English languages [33], Turkish emotional speech database (TURES) in Turkish language, speech database (Keio-ESD) in Japanese language, Italian emotional speech database (EMOVO) in Italian language [7] and emotional speech database in Malayalam language [34]. There are a several emotional speech databases in Arabic speech. Egyptian Arabic speech emotion (EYASE) includes 579 utterances expressed in four basic emotions: Anger, happiness, neutral and sadness [35]. Emirati speech database (ESD) was constructed by local Emirati speakers [36]. Tunisian emotional database was built by professional Tunisian actors. This database contains five types of emotions including happy, anger, sadness, fear and neutral [37]. Moroccan emotional database (MEDB) was created from broadcasts on the YouTube channel [38]. To recognize sentiment in natural Arabic speech, database contains utterances recorded in three emotions: happiness, angry and surprise is constructed [39].

In literature, several features extracted from speech have been exploited to develop and to improve the performance of systems of SER. The first studies focused on the prosodic features. These features have been divided in three types: pitch, duration and intensity [12-14]. The statistical values of prosodic parameters have been used to discriminate the emotions from speech [40]. Energy and duration are useful features in SER field [41]. Spectral features have been strongly investigated in the SER systems. The Linear Prediction (LP) and MFCCs parameters have been used to classify speech emotions [42], [43]. The MFCC and DWT (Discrete Wavelet Transform) algorithms have been exploited for extracting features in the classification system of different emotions [44]. Voice quality parameters have been widely used in many systems of emotion recognition [17-19]. HNR, jitter and shimmer with other spectral parameters have been applied in SER [17]. Different emotions have been recognized using jitter, shimmer, or their combination by using multiple classifiers [45]. Combinations of different features have been used in the SER systems to improve the performance of recognition. Fundamental frequency (F_0), log of energy, formants, energy in Mel and MFCCs parameters are exploited to classify speech emotions [46]. Different combinations of features such as: pitch, intensity, formants, MFCCs, wavelet and long-term average spectrum have been used to classify four emotions: anger, happiness, neutral and sadness [35]. The first three formants and pitch features have been investigated to improve the performance in SER system [47]. Many features such as MFCC, spectral centroid, spectral skewness, and spectral pitch chromas have been used [48]. Combinations of different types of features: pitch, intensity, jitter, formants and MFCCs parameters have been investigated to analyze the performance on different databases in recognition of speech emotions [49]. To recognize speech emotions, MFCCs, short time energy and pitch features are exploited for recognizing speech emotions in Malayalam language [34].

Classifiers are an essential and important part in the recognition and classification systems. There are numerous classifiers which have been used for the task of SER. Among the most classifiers used in the previous researches were HMM, SVM, ANN, GMM and KNN. Studies in [50-51] indicate that previous works have exploited the HMM as classifier in the SER systems. To classify speech emotions, Emo-DB database and SVM classifier have been used to assess the performance of classification [52]. Four emotional states: neutral, happy, sad and anger have been recognized by using ANN classifier [53]. GMM classifier has given a good performance for classifying emotions in speech [48], [54]. GMM achieved 92% accuracy when used to recognize emotions in speech [55]. KNN classifier has been widely applied in SER systems [56-57]. KNN is simple classifier to differentiate the emotion of anxiety from other emotions [56]. In previous works, several classifiers designed as hybrid classifiers, multiple classifiers or ensemble classifiers. A hybrid classifier (GMM-DNN) constructed from two classifiers which were the GMM and DNN has been exploited in SER system [36]. For gender driven speech emotion recognition, Hybrid classifier composed of HMM and SVM classifiers have been proposed to improve the performance [58]. Different classification techniques have been compared to develop the systems of SER. To classify emotions in speech, different classifiers have been compared. These classifiers are KNN, SVM, linear discriminant analysis (LDA) and Regularized Discriminate Analysis (RDA). Results obtained indicated that the best performance has been given by RDA [57]. The performances of ANN and SVM classifiers have been compared. Results indicated that the ANN classifier gave high performance around to 88.4% and 78.2 % for the SVM classifier [34]. Several classifiers: GMM, ANN, K-means clustering and VQ (vector quantization) have been compared to recognize emotions. From results, it has been concluded that the performance of classifier depends on the type of database [49]. A hybrid classifier (GMM-DNN) in [36] gave better performance when compared with SVM and MLP (multilayer perception) classifiers [37]. The performance of KNN classifier has been compared with SVM classifier. The SVM classifier performed better than KNN for classifying speech emotion from EYASE database [35]. Different classification technique including KNN, SVM, MLP, Random Forest Classifier (RFC) and Convolutional Neural Networks (CNN) have been investigated in SER by using EMO-DB and Ravdess databases [59].

3. METHODOLOGY

Our work is divided into two parts. The aim of the first part is to study the influence of different features on the recognition of anger and neutral emotions in four databases with different languages. The scheme that describes the first part is illustrated in Fig.1. To achieve the aim of the first part, four databases in different languages: ADED, EMO-DB, ShEMO and CREAMA-D are exploited in the system of recognition. Only the audio files of anger and neutral states of the database are used in this work. Different speech features including prosodic features (pitch and intensity), formants and MFCCs parameters are extracted from the audio files. After features extraction there is the classification step. In this step different feature sets are formed to identify the feature sets that give the higher performance of recognition in each database. SVM technique is used as classifier in the system of recognition. The response in the form of recognition of anger and neutral emotions is obtained and studied for their recognition rate. The influence of anger and neutral emotions for identifying the gender classes in the four databases is studied in the second part. Fig.2 illustrates the scheme that describes the second part. To achieve the purpose of this part, the same databases, speech features and classifier of first part are used. But in this part the system classifies the gender classes (male and female) under anger and neutral states in each database. The response in the form of classification of gender is obtained and studied for their accuracy.

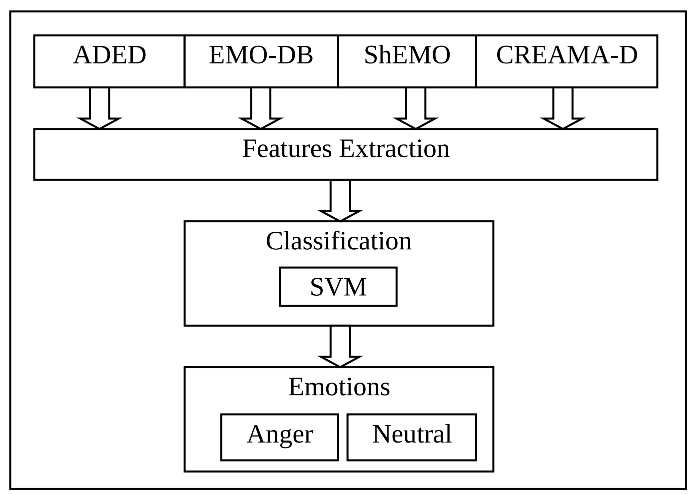


Figure 1. Scheme of the first part

A. *Emotional speech databases*

Algerian Dialect Emotional Database (ADED) is constructed from six famous movies in Algerian dialect. Algerian dialect (AD) is a mixture between the standard Arabic and the native dialect. AD was impacted by French, Spanish, Berber and Turkish languages, and this influence was caused by the long period of colonization [60]. The ADED database contains four emotions: fear, anger, sadness and neutral. This database composed of 200 speech segments of duration ranging from 0.2 s to 3 s and these segments are collected at sampling frequency of 44.1 kHz. The speech segments included in this database are expressed by 32 actors (16 males and 16 females) from different ages between 18 and 60 years. The numbers of segments of each emotion are illustrated in Table I.

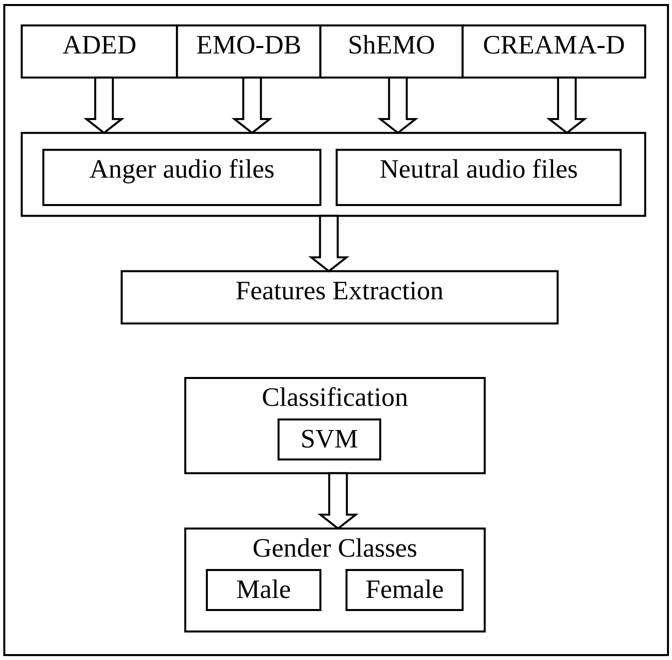


Figure 2. Scheme of the second part

TABLE.I NUMBER OF SEGMENTS OF EACH EMOTION IN ADED DATABASE

Emotions	**Segments number**
Fear	52
Neutral	48
Anger	52
Sadness	48
Total	200

Berlin database of emotional speech (EMO-DB) is a German database on emotional speech registered by the Technical University of Berlin [8]. The EMO-DB database included seven types of emotion: happiness, boredom, fear, disgust, anger, sadness, and neutral. In this database, there are 535 audio files were recorded by 5 males and 5 females aged 21 to 35 years. The numbers of audio files of each emotion are shown in Table II.

TABLE.II NUMBER OF AUDIO FILES OF EACH EMOTION IN EMO-DB DATABASE

Emotions	**Audio files number**
Fear	69
Neutral	79
Anger	127
Sadness	62
Happiness	71
Disgust	46
Boredom	81
Total	535

Sharif Emotional Speech Database (ShEMO) is a database for Persian language. This database comprises 3000 utterances expressed by 87 native Persian (31 females, 56 males). ShEMO database classified into five basic emotions: surprise, happiness, sadness, fear, anger and the neutral state [29]. The numbers of utterances of each emotion are illustrated in Table III.

TABLE.III NUMBER OF UTTERANCES OF EACH EMOTION IN SHEMO DATABASE

Emotions	**Utterances number**
Fear	38
Neutral	1028
Anger	1059
Surprise	225
Happiness	201
Sadness	449
Total	3000

Crowd-sourced Emotional Multimodal Actors Dataset (CREMA-D) composed of facial and vocal emotional expressions in English language [30]. This database composed of 7442 audio files expressed in six basic emotions including happiness, sadness, fear, disgust, anger and neutral. The audio files in CREMA-D database were recorded by 91 actors (48 male and 43 female) aged 20 to 74 years. The numbers of audio files of each emotion are shown in Table IV.

TABLE.IV NUMBER OF AUDIO FILES OF EACH EMOTION IN CREMA-D DATABASE

Emotions	**Audio files number**
Fear	1271
Neutral	1087
Anger	1271
Sadness	1271
Happiness	1271
Disgust	1271
Total	7442

B. *Features extraction*

Extraction of parameters is an important step in the SER systems. The features extracted in this work are the statistical parameters of prosodic features (pitch and intensity), formants and MFCCs parameters. Table V illustrates the features extracted in this work. The statistics values of prosodic features and formants are extracted by PRAAT software [61]. The MFCCs parameters are extracted by MATLAB software.

Prosodic features are acoustic parameters calculated from the speech segments. These features include pitch, intensity and duration. The pitch defined as the glottal waveform, it is produced from the vibration of the vocal folds [62]. Intensity generally models the loudness of a sound as perceived by the human ear [63]. Pitch and intensity are strongly correlated with the emotional states [64]. The statistical parameters of the pitch and intensity are used in this work. These parameters are: mean maximum, minimum and standard deviation of pitch, mean, maximum and minimum of intensity.

TABLE.V FEATURES EXTRACTED IN THIS WORK

Features extracted	
Prosodic features	Mean of pitch Maximum of pitch Minimum of pitch Standard deviation of pitch Mean of intensity Maximum of intensity Minimum of intensity
Formants	Formant1 Formant2 Formant3 Formant4
MFCCs	13 MFCCs

Formants are one of vocal tract features. They correspond to the resonance frequencies of system of the human vocal tract [65]. The vocal tract resonances can be changed by modifying the position of the tongue or the jaw [66]. The positioning of the formants is influenced by the type of emotion [67]. The first four formants, formant1, formant2, formant3 and formant4 are used in this work as shown in Table V.

MFCCs belong to the family of the cepstral descriptors. MFCCs are the parameters that exploit the different perception of the human ear of frequency signals. This is due to their ability to imitate human auditory perception mechanism [68]. The MFCC computation consists the following Bloks: Pre-emphasize, Hamming window, FFT, Triangular band-pass filter, Logarithm and discrete cosine transformation (DCT) [69]. The sampling frequency used was 8 kHz. In this work 13 MFCCs are used in the systems of recognition and classification as presented in Table V.

C. Support Vector Machine

Support vector machine (SVM) is a linear classifier**.** It is a supervised learning process of two steps training and testing [70]. SVM are constructed by mapping the training patterns into space of a higher dimensional feature, the points separated by using a hyper-plane [71]. SVM achieved higher performance en compared with other classifiers in many works. Some of these works are presented in Table VI.

The performance given by SVM classifier has been higher compared to KNN classifier in Egyptian Arabic speech emotion classification [35]. To classify four emotions, SVM performed better compared to KNN (K-nearest neighbor) and NN (Neural Network) classifiers [72]. Several classifiers, SVM, Random Forest (RF), Naïve Bayes (NB) and Neural Network (NN) have been used for classifying Indonesian emotion speech. These classifiers have been compared for their performance. The highest performance was obtained by the SVM classifier [73]. SVM has been used in their application for regression SVR (Support Vector Regression). SVR has been compared to a rule-based Fuzzy Logic and Fuzzy KNN classifiers. The best results have been given by SVR classifier [74]. SVM, KNN and ANN classifiers have been used to recognize the speech emotions in Tamile language speaker. Results obtained indicated that the accuracy of recognition of SVM and ANN was 85.7% and the accuracy of KNN was 66.67% [75]. The SVM classifier performed better than Recurrent Neural Networks (RNN) classifier for classifying speech emotions by using Berlin emotional database [76]. SVM achieved the highest classification rate (92.86%) when compared with other classifiers such as ANN, NB, KNN for classification emotions in speech [77].

TABLE VI. SOME WORKS THAT HAVE MADE COMPARISON BETWEEN CLASSIFIERS IN FIELD OF SER

The work	The compared classifiers
[35]	SVM and KNN
[72]	SVM, KNN and NN
[73]	SVM, RF, NB and NN
[74]	SVR, Fuzzy Logic and Fuzzy KNN
[75]	SVM, ANN and KNN
[76]	SVM and RNN
[77]	KNN, SVM, ANN and NB

4. EXPERIMENTS AND RESULTS

Our work is divided into two parts. The purpose of the first part is to study the influence of speech features on the recognition of anger and neutral emotions in four databases with different languages. In the second part, the purpose is to study to influence of emotion types (anger and neutral) to identify the gender classes in the previous databases. As mentioned in the previous section, the features used in the system of recognition are the statistical parameters of pitch and intensity, formants and MFCC parameters. The system of recognition is based on SVM classifier. The features are input into the classifier as features vectors. 60% of the segments of databases are used as training set, and 40% of the segments are used as test set. Table VII present the number of speech segments of each database used in the experiments. The Experiments are made by MATLAB software.

TABLE VII. NUMBER OF SPEECH SEGMENTS OF EACH DATABASE USED IN THE EXPRIMENTS

Databases	Anger	Neutral
ADED	52	48
EMO-DB	72	72
ShEMO	72	72
CREMA-D	72	72

A. The influence of speech features on the recognition of emotions

In this part, several experiments are performed to study the influence of the features extracted on the recognition of anger and neutral emotions in different databases. Different features sets are formed to identify the best features sets that give the higher recognition accuracy in each database. The results of experiments are shown in Table VIII. Comparisons between the performances of each features set on each database are illustrated in the Fig.3, 4, 5 and 6. It is observed that when using only the prosodic features, the recognition rate is higher in EMO-DB and CREMA-D databases but the recognition rate is decrease in ADED and ShEMO databases. Lower recognition rates are remarked when using only the formants features in all databases. Acceptable recognition rates are noted when only the MFCCs parameters are used in EMO-DB and ShEMO databases compared to ADED and CREMA-D databases. The performance of recognition is increased when using different combination of features. The best recognition rates are obtained when combination of prosodic, formants and MFCCs parameters is used in all the databases. Fig.3, 4, 5 and 6 ensure the pervious remarks. Table IX and X show the recognition rates obtained with different sets of feature in each database for male gender and female gender respectively. According to Table IX and X, the same remarks are noted compared to the experiments that used the databases without gender distinction. The higher performance is obtained when used combination of features (prosodic, formants and MFCCs) in each database. It is concluded according to the results that the performance of recognition is improved when using features combination of prosodic, formants and MFCCs parameter in the different database used in this work.

In previous works, the combination of different features improves the performance of recognition of speech emotions in databases with different languages. A combination of prosodic (energy and pitch) and spectral features (MFCCs) performed better than features (prosodic or spectral features) for recognizing emotions in Berlin and Spanish emotional speech databases [78]. Different features such as pitch, intensity, formants, MFCCs, wavelet and long-term average spectrum have been used to recognize emotions in Egyptian Arabic speech emotion. Results obtained showed that the combinations of features have achieved higher performance than systems used the same types of features [35]. In Mandarin emotional database, the performance of system that using a combination of both spectral features and prosodic features is higher than using only spectral or prosodic features [79].

TABLE VIII. THE RECOGNITION RATES OBTAINED WITH DIFFERENT FEATURES SETS IN EACH DATABASE.

Parameters	**ADED**	**EMO-DB**	**ShEMO**	**CREMA-D**
Prosodic features	78.12%	93.75%	78.47%	93.05%
Formants	61.45%	79.17%	79.17%	69.44%
MFCCs	63.54%	89.58%	84.37%	71.53%
Prosodic + Formants	78.12%	93.75%	86.45%	93.05%
Prosodic + MFCCs	83.33%	95.83%	88.54%	95.14%
Formants + MFCCs	80.21%	96.87%	83.33%	93.75%
Prosodic + Formants + MFCCs	**85.42%**	**97.92%**	**90.97%**	**95.14%**

TABLE IX. THE RECOGNITION RATES OBTAINED WITH DIFFERENT FEATURES SETS IN EACH DATABASE FOR MALE GENDER

Parameters	**ADED**	**EMO-DB**	**ShEMO**	**CREMA-D**
Prosodic features	79.54%	100%	87.50%	91.97%
Formants	63.63%	73.91%	62.50%	68.05%
MFCCs	79.54%	91.67%	84.72%	75.00
Prosodic + Formants	88.63%	100%	87.50%	91.66%
Prosodic + MFCCs	81.81%	100%	94.44%	93.05%
Formants + MFCCs	81.81%	91.67%	77.78%	86.11%
Prosodic +Formants + MFCCs	**100%**	**100%**	**97.22%**	**94.44%**

TABLE X. THE RECOGNITION RATES OBTAINED WITH DIFFERENT FEATURES SETS IN EACH DATABASE FOR FEMALE GENDER.

Parameters	**ADED**	**EMO-DB**	**ShEMO**	**CREMA-D**
Prosodic features	88.63%	98.61%	87.50%	94.94%
Formants	59.09%	84.72%	76.39%	70.83%
MFCCs	70.45%	97.22%	70.83%	70.83%
Prosodic + formants	90.91%	98.61%	94.44%	94.44%
Prosodic + MFCCs	97.73%	100%	93.05%	98.61%
Formants + MFCCs	79.54%	97.22%	83.33%	76.39%
Prosodic + formants +MFCCs	**100%**	**100%**	**95.83%**	**100%**

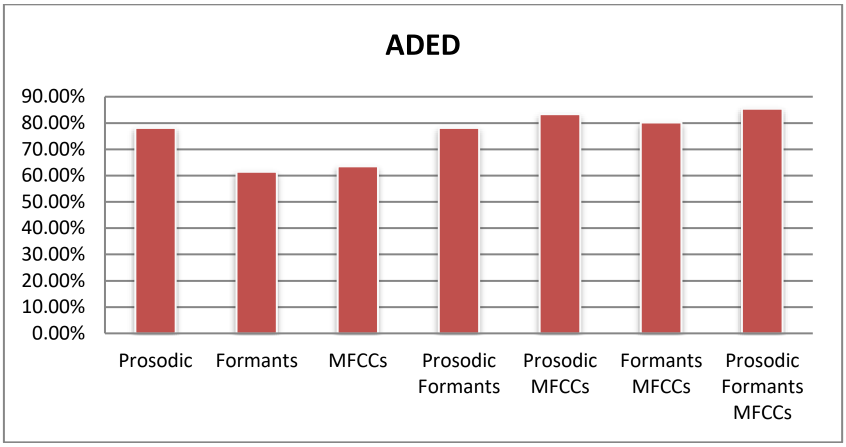


Figure3. Comparisons between the performances of each features set in ADED database

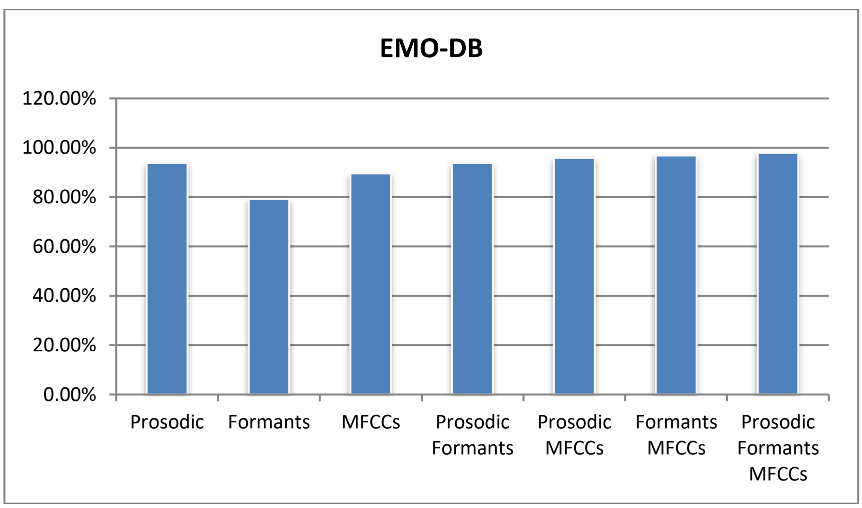


Figure4. Comparisons between the performances of each features set in EMODB database

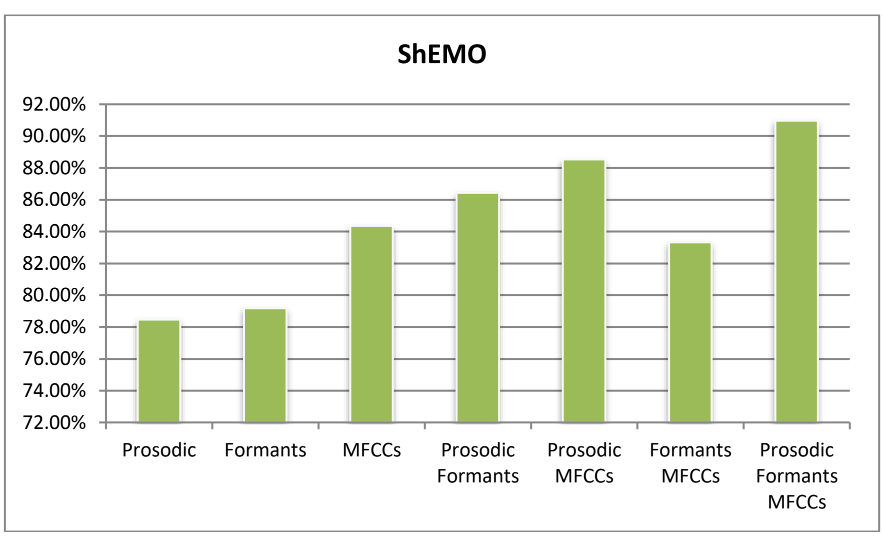


Figure5. Comparisons between the performances of each features set in ShEMO database.

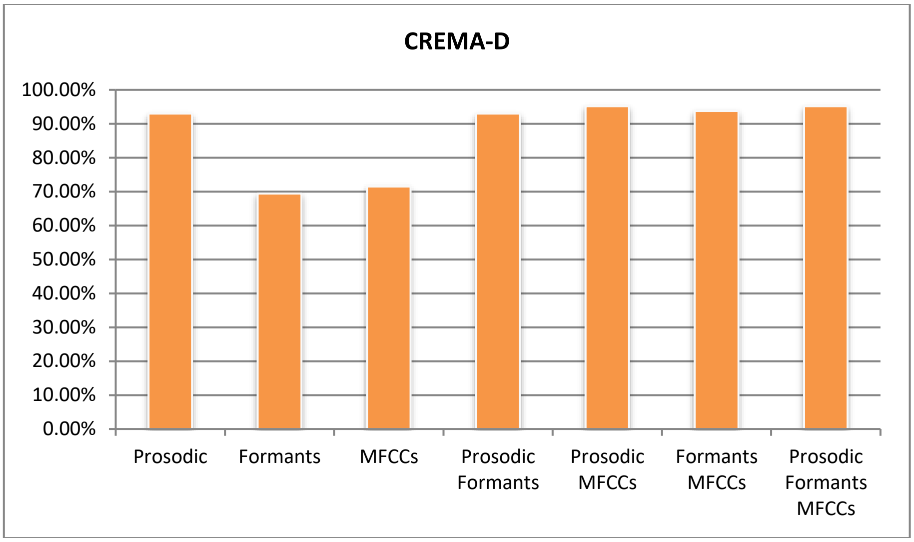


Figure6. Comparisons between the performances of each features set in CREMA-D database.

B. *Influence of emotion types to identify the gender classes*

In this part, many experiments are performed to study the influence of anger and neutral emotions to identify the gender classes in different databases with different languages. Features combination of prosodic, formant and MFCCs parameters is used in the systems of classification in this part. The results of experiments are shown in Table XI. Fig.7 illustrated a comparison between the accuracies of gender classification under anger and neutral emotions in different databases. In ADED database, it is remarked that the classification rate is maximum (100%) under neutral emotion but this classification rate is slightly lower (97.82%) under anger emotion. The opposite case is observed in EMO-DB database, the classification rate is maximum (100%) under anger emotion. A maximum classification rate (100%) is noted under the two types of emotions anger and neutral in ShEMO database. In CREMA-D, a maximum accuracy is obtained under anger emotion but the accuracy is lower (86.11%) in neutral emotion. It is concluded that the performance of the gender classification is influenced by the type of emotion and the language of database.

The gender classification was impacted by the emotion types in different databases in previous work. Child emotional speech database in Russian language was presented [80]. This database contains three emotional states (comfort, discomfort, neutral). The classification of gender was influenced by the emotional state. In the discomfort state, the classification rate of male is higher compared with female gender. But in the two other states the classification rate of female gender was higher. The authors in [81] classified the gender classes under four emotions, angry, happy, calm and sad. It was remarked that the gender classification was influenced by the type of emotion.

5. CONCLUSION

In this work, we presented a study that evaluated the performance of some speech features on the recognition of anger and neutral emotions in different databases with different languages. The impact of anger and neutral emotions to classify the gender classes in different databases was also studied. ADED, EMO-DB, ShEMO and CREAMA-D databases were exploited for extracting the features concerning the emotions treated. The features extracted were the statistical parameters of pitch and intensity, formants and MFCCs parameters.

The experimental results demonstrated that the systems that used a combination of prosodic, formants and MFCCs parameters performed compared to the performance of systems that used individual features (prosodic, formants or MFCCs) in each type of database. It was showed in the results also that the classification of gender classes was influenced by the type of emotion and the language.

In the future, the present work can be extended to enhance our results by including more speech features, more databases and more emotional states.

TABLE XI. CONFUSION MATRICES OF GENDER CLASSIFICATION UNDER ANGER AND NEUTRAL EMOTIONS IN DIFFERENT DATABASES.

Emotions		**Anger**		**Neutral**	
ADED	Gender	Male	Female	Male	Female
	Male	100%	0%	100%	0%
	Female	4.35%	95.65%	0%	100%
	Average	**97.82%**		**100%**	
EMO-DB	Gender	Male		Male	Female
	Male	100%	0%	100%	0%
	Female	0%	100%	97.22%	94.44%
	Average	**100%**		**97.22%**	
ShEMO	Gender	Male		Male	Female
	Male	100%	0%	100%	0%
	Female	0	100%	0%	100%
	Average	**100%**		**100%**	
CREMA-D	Gender	Male		Male	Female
	Male	100%	0%	83.33%	16.67%
	Female	0%	100%	11.11%	88.89
	Average	**100%**		**86.11%**	

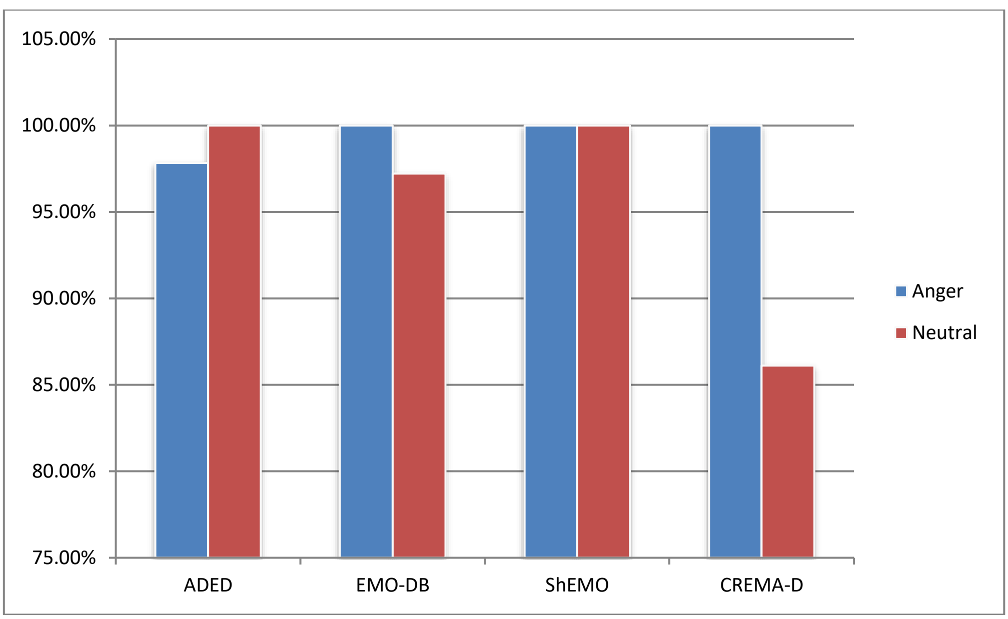


Figure 7. A comparison between the accuracies of gender classification under anger and neutral emotions in different databases.

REFERENCES

[1] S. Ramakrishnan and I. M. M. El Emary, " Speech emotion recognition approaches in human computer interaction," Telecommunication Systems, vol. 52, no. 3, pp. 1467–1478, 2011

[2] D. J. France, R. G. Shiavi, S. Silverman, M. Silverman, M. Wilkes, M, "Acoustical properties of speech as indicators of depression and suicidal risk," IEEE Transactions on Biomedical Engineering, vol. 47, no. 7, pp. 829–837, 2000.

[3] R. Prasath and T. Kathirvalavakumar, "Identifying Psychological Theme Words from Emotion Annotated Interviews,"Mining Intelligence and Knowledge Exploration, LNAI 8284, pp. 728–739, 2013.

[4] M. Pantic, A. Pentland, A. Nijholt, T. S. Huang, "Human computing and machine understanding of human behavior: A survey," In Proceedings Eighth ACM Int'l Conf, Multimodal Interfaces, Springer, pp. 239–248, 2006

[5] E. Moore, M. A. Clements, J. W. Peifer, and L. Weisser, "Critical analysis of the impact of glottal features in the classification of clinical depression in speech," IEEE Transactions on Biomedical Engineering, vol. 55, no. 1, pp. 96–107, 2008.

[6] G. I. Roisman, J. L. Tsai, and K. S. Chiang, "The emotional integration of childhood experience: Physiological, facial expressive, and self-reported emotional response during the adult attachment interview," Developmental Psychology, vol. 40, no. 5, pp. 776–789, 2004.

[7] M. B. Akçay, K. Oğuz, "Speech emotion recognition: Emotional models, databases, features, preprocessing methods, supporting modalities, and classifiers," Speech Communication, vol. 116, pp. 56-76, 2020.

[8] F. Burkhardt, A. Paeschke, M. Rolfe, W. Sendlmeier and B. Weiss B, "A database of german emotional speech," 2005. Eurospeech, 9th European Conference on Speech Communication and Technology, Lisbon, Portugal, Sept 2005.

[9] I. Engberg, A. Hansen, "Documentation of the Danish emotional speech database DES," Center for Person Communication, Institute of Electronic Systems, Alborg University, Aalborg, Denmark. September, 1996.

[10] P. Staroniewicz, W. Majewski, "Polish Emotional Speech Database – Recording and Preliminary Validation," Lecture Notes in Computer Science, pp. 42–49, 2009

[11] P. Jackson, S. Haq, and J. D. Edge, "Audio-visual feature selection and reduction for emotion classification," In: Proc. Int'l Conf. on Auditory-Visual Speech Processing. pp. 185–90, 2008.

[12] T. L. New, S. W. Foo, L. C. De Silva, "Classification of stress in speech using linear and nonlinear features," In: Proceedings of IEEE international conference acoustics, speech, and signal processing, Hong Kong, China, vol.3, pp 9–12, May 2003.

[13] B. Schuller, G. Rigoll, and M. Lang, M. "Hidden Markov model-based speech emotion recognition," In: Proceedings of international conference on multimedia and expo, IEEE, Baltimore, MD, USA, pp 401–404, July 2003.

[14] Y. Wang, S. Du, Y. Zhan, " Adaptive and optimal classification of speech emotion recognition," In: Proceedings of 4th international conference on natural computation, IEEE, Jinan, China, pp 407–411, Oct 2008.

[15] K. R. Sreenivasa, G. Shashidhar, Koolagudi, "Emotion Recognition Using Vocal Tract Information," Springer Briefs in Electrical and Computer Engineering. Springer-Verlag New York, 2013.

[16] M. Sheikhan, D. Gharavian and F. Ashoftedel, F, "Using DTW neural–based MFCC warping to improve emotional speech recognition," Neural Computing and Applications, vol. 21, no. 7, pp. 1765–1773, 2011.

[17] B. Schuller, A. Batliner, D. Seppi, S. Steidl, T. Vogt, J. Wagner, L. Devillers, L. Vidrascu, N. Amir, L. Kessous, and V. Aharonson, "The relevance of feature type for the automatic classification of emotional user states: low level descriptors and functionals," In: Proceedings of INTERSPEECH, pp. 2253–2256, 2007.

[18] T. Kostoulas, T. Ganchev, A. Lazaridis, and N. Fakotakis, "Enhancing Emotion recognition from speech through feature selection, lecture notes in artificial intelligence, vol 6231, pp 338–344, 2010.

[19] C. H. Wu, W. B. Liang, "Emotion recognition of affective speech based on multiple classifiers using acoustic- prosodic information and semantic labels," IEEE Transactions on Affective Computing, vol. 2, no. 1, pp. 10–21, 2011.

[20] S. R. Bandela, T. K. Kumar, "Emotion Recognition of Stressed Speech Using Teager Energy and Linear Prediction Features," IEEE 18th International Conference on Advanced Learning Technologies (ICALT), Mumbai, India, pp. 422-425, July 2018.

[21] B. Schuller, A. Batliner, S. Steidl, D. Seppi, "Recognising realistic emotions and affect in speech: State of the art and lessons learnt from the first challenge," Speech Communication, vol. 53, pp. 1062–1087, 2011.

[22] D. Kamińska, T. Sapiński, and G. Anbarjafari, "Efficiency of chosen speech descriptors in relation to emotion recognition," EURASIP Journal on Audio, Speech and Music Processing, vol.1 , 2017.

[23] C. M. Lee, S. S. Narayanan, and R. Pieraccini, "Recognition of negative emotions from the speech signal. In Proceedings of Automatic Speech Recognition and Understanding workshop, IEEE, Madonna di Campiglio, Italy ,pp. 240–243, Dec 2001.

[24] K. Tarunika, R. Pradeeba, and P. Aruna, "Applying Machine Learning Techniques for Speech Emotion Recognition," 2018 9th International Conference on Computing, Communication and Networking Technologies (ICCCNT), IEEE, Bangalore, India, July 2018.

[25] P. Patel, A. Chaudhari, R. Kale, and M. A. Pund, "Emotion Recognition from Speech with Gaussian Mixture Models & Via Boosted GMM, International Journal of Research In Science & Engineering, vol . 3, no. 2, pp. 47-53, 2017.

[26] R. W. Novaco, "Anger and psychopathology," In International handbook of anger, Springer, New York, NY, pp. 465-497, 2010.

[27] J. R. Averill, "Studies on anger and aggression: Implications for theories of emotion," American Psychologist, vol. 38, no.11, pp. 1145–1160, 1983.

[28] S. G. Koolagudi andK. S. Rao, "Emotion recognition from speech: a review," International Journal of Speech Technology, vol. 15, no. 2, pp. 99–117, 2012.

[29] O. Mohamad Nezami, P. Jamshid Lou, andM, Karami, , "ShEMO: a large-scale validated database for Persian speech emotion detection," Language Resources and Evaluation, vol. **53**, pp. 1–16, 2018.

[30] R. Singh, H. Puri, N. Aggarwal, and V. Gupta, " An Efficient Language-Independent Acoustic Emotion Classification System," Arabian Journal for Science and Engineering, vol.45, pp. 3111–3121, 2019.

[31] L. Vidrascu, and L. Devillers, "Real-life emotions in naturalistic data recorded in a medical call center," In 1st International Workshop on Emotion: Corpora for Research on Emotion and Affect, Genoa, Italy, pp. 20–24, 2006.

[32] K. S. Rao, G. Koolagudi, S. G, "Identification of Hindi dialects and emotions using spectral and prosodic features of speech," Systemics, Cybernetics, and Informatics, vol. 9, no. 4, pp. 24–33, 2011.

[33] D. C. Ambrus, "Collecting and recording of an emotional speech database," Advances in speech technology, pp. 239-244, 2000.

[34] T. M. Rajisha, A. P. Sunija, and K. S. Riyas, “Performance Analysis of Malayalam Language Speech Emotion Recognition System Using ANN/SVM,” Procedia Technology, vol. 24, pp. 1097–1104, 2016.

[35] L. Abdel-Hamid, “Egyptian Arabic Speech Emotion Recognition using Prosodic, Spectral and Wavelet Features,” Speech communication, May 2020.

[36] I. Shahin, A. B. Nassif, S. Hamsa, S, “Emotion Recognition Using Hybrid Gaussian Mixture Model and Deep Neural Network,” IEEE Access, vol. 7, pp. 26777–26787, 2019.

[37] M. Meddeb , K. Hichem, and A. Alimi, “ Speech Emotion Recognition Based on Arabic Features,” in 15th international conference on Intelligent Systems design and Applications (ISDA15), IEEE, Marrakesh, Morocco, December 2015.

[38] A. Agrima, A. Farchi, L. Elmazouzi, I. Mounir, and B. Mounir, “Emotion recognition from Moroccan dialect speech and energy band distribution,” International Conference on Wireless Technologies, Embedded and Intelligent Systems (WITS), IEEE, Fez, Morocco, April 2019.

[39] S. Klaylat, Z. Osman, L. Hamandi, R. Zantout, “Emotion recognition in Arabic speech,” Analog Integrated Circuits and Signal Processing , vol. 96, pp. 337–351, 2018.

[40] M. Schroder, “Emotional speech synthesis: A review”, 7th European Conference on Speech Communication and Technology, 2nd INTERSPEECH Event, Aalborg, Denmark, September 2001.

[41] B. Heuft, T. Portele, M. Rauth, “Emotions in time domain synthesis,” In Proceedings, fourth international conference on Spoken Language, IEEE, Philadelphia, PA, USA, vol. 3, pp. 1974–1977, Oct1996.

[42] A. Chauhan, S. G. Koolagudi, S. Kafley, K. S. Rao, “Emotion recognition using LP residual,” In Proceedings of the 2010 IEEE students technology symposium,Kharagpur, India, pp. 255-261, April 2010.

[43] R. B. Lanjewar, and M. Swarup, “Implementation and comparison of speech emotion recognition system using gaussian mixture model (GMM) and K- Nearest Neighbor (K-NN) Techniques,” Procedia Computer Sci, vol. 49, pp. 50–57, 2015

[44] S. T. Saste, S. M. Jagdale, “Emotion recognition from speech using MFCC and DWT for security system,” International Conference of Electronics, Communication and Aerospace Technology (ICECA), IEEE, Coimbatore, India, April 2017.

[45] A. Jacob, “Speech emotion recognition based on minimal voice quality features,” International Conference on Communication and Signal Processing (ICCSP), IEEE, Melmaruvathur, India April 2016.

[46] O. Kwon, K. Chan, J. Hao, T. Lee, “Emotion recognition by speech signals,8th European Conference on Speech Communication and Technology, Geneva, Switzerland, September 1-4, 2003

[47] D. Gharavian, M. Sheikhan andF. Ashoftedel, “Emotion recognition improvement using normalized formant supplementary features by hybrid of DTW-MLP-GMM model,” Neural Computing and Applications, vol. 22, no. 6, pp. 1181–1191, 2012.

[48] M. Navyasri, R. RajeswarRao,A. DaveeduRaju, M. Ramakrishnamurthy, “Robust Features for Emotion Recognition from Speech by Using Gaussian Mixture Model Classification,”Information and Communication Technology for Intelligent Systems (ICTIS), vol. 2,pp.437-444, 2017.

[49] S. G. Koolagudi, Y. V. S, Murthy and S. P. Bhaskar, “Choice of a classifier, based on properties of a dataset: case study-speech emotion recognition,” International Journal of Speech Technology, vol.21, no. 1, pp. 167–183, 2018.

[50] B. D. Womack andJ. H. Hansen, J. H, “L.N-channel hidden markov models for combined stressed speech classification and recognition” IEEE Transactions on Speech and Audio Processing, vol. 7, no. 6, pp. 668–677, 1999.

[51] S. M. Feraru, and D. Schuller, “Cross-language acoustic emotion recognition: an overview and some tendencies,” International Conference on Affective Computing and Intelligent Interaction (ACII), IEEE, Xi'an, China, pp. 125–131, Sept 2015.

[52] T. Seehapoch, and S. Wongthanavasu, “Speech emotion recognition using support vector machines,” In 5th international conference on Knowledge and smart technology (KST), IEEE, Chonburi, Thailand, pp. 86–91, Feb 2013.

[53] S. A. Firoz, S. A. Raji, and A. P. Babu, “Automatic emotion recognition speech using artificial neural networks with genderdependent databases,” 2009 International Conference on Advances in Computing, Control, and Telecommunication Technologies, IEEE, Trivandrum, Kerala, India, pp. 162–164, Dec 2009.

[54] Z. Zeng, M. Pantic, G. I. Roisman, T. S. Huang, “A survey of affect recognition methods: Audio, Visual, and spontaneous expressions,” IEEE Trans. on Pattern Analysis and Machine Intelligence, vol. 31, pp. 39-58, 2009.

[55] I. Luengo, E. Navas, I. Hernez, and I. Snchez, “Automatic emotion recognition using prosodic parameters,” In INTERSPEECH, Eurospeech, 9th European Conference on Speech Communication and Technology, Lisbon, Portugal, pp. 493–496, September 2005.

[56] M. Dan Zbancioc, S. M. Feraru, “A study about the automatic recognition of the anxiety emotional state using Emo-DB,” E-Health and Bioengineering Conference (EHB), IEEE, Iasi, Romania. Nov 2015.

[57] S. Kuchibhotla, H. D. Vankayalapati, R. S. Vaddi, and K. R. Anne, “An optimal two stage feature selection for speech emotion recognition using acoustic features,” International Journal of Speech Technology, vol. 19, no. 4, pp. 657–667, 2016.

[58] P. P. Ladde and V. S. Deshmukh, “Use of Multiple Classifier System for Gender Driven Speech Emotion Recognition,” International Conference on Computational Intelligence and Communication Networks (CICN), IEEE, Jabalpur, India, pp. 713-717, Dec 2015.

[59] J. S. Deusi and E. I. Popa, “An Investigation of the Accuracy of Real Time Speech Emotion Recognition,” International Conference on Artificial Intelligence, Cambridge, UK, pp.336-349, Dec 2019.

[60] M. A. Menacera, O. Mellaa , D. Fohra , D. Jouveta , D. Langloisa , K. Smailia, Development of the Arabic Loria Automatic Speech Recognition system (ALASR) and its evaluation for Algerian dialect,” 3rd International Conference on Arabic Computational Linguistics, Dubai, United Arab Emirates , vol. 117, pp. 81–88, 2017

[61] P. Boersma, D. Weenink, “Praat, a system for doing phonetics by computer,” Glot Int, vol. 5, no. 9, pp. 341–345, 2002.

[62] D. Ververidis andC. Kotropoulos, “Emotional speech recognition: resources, features, and methods,” Speech Communication, vol. 48, pp. 1162–1181, 2006.

[63] A. Batliner, B. Schuller, et al., “The automatic recognition of emotions in speech,” in: R. Cowie, C. Pelachaud, P. Petta (Eds.), Emotion-Oriented Systems, Springer,Berlin, Heidelberg, pp. 71–94, 2010.

[64] A. Meftah, S.A. Selouani, Y. A. Alotaibi, “Preliminary Arabic speech emotion classification,” International Symposium on Signal Processing and Information Technology (ISSPIT), IEEE, Noida, India,pp. 179–182, Dec 2014.

[65] A. Hassan, “On Automatic Emotion Classification Using Acoustic Features”, Faculty of Physical and Applied Sciences, University of Southampton , 2012.

[66] K. N. Stevens, “Scientific substrates of speech production”, Introduction to Communication Sciences and Disorders,” Singular, San Diego, CA, pp. 399- 437, 1994.

[67] M. Goudbeek, J. P. Goldman and K. Scherer, “Emotion dimensions and formant position,”10th Annual Conference of the International Speech Communication Association, Brighton, United Kingdom, Sept 2009.

[68] Y. Huang, S. M. Xiaoshuang, W. Fan, “A classification method for wood vibration signals of Chinese musical instruments based on GMM and SVM,” Traitement du Signalvol. 35, no. 2, pp. 121-136, 2018.

[69] G. Zhang , J. Yin , Q. Liu C. Yang,“The Fixed-Point Optimization ofMel Frequency Cepstrum Coefficients for Speech Recognition,” Proceedings of 2011 6th International Forum on Strategic Technology, IEEE,Harbin, Heilongjiang, Harbin. Aug 2011.

[70] A. Joshi, “Speech Emotion Recognition Using Combined Features of HMM & SVM Algorithm,” IJARCSSE, vol. 3, no. 8, 2013.

[71] C. W. Hsu, C. C. Chang, and C. J. Lin, “A practical guide to support vector classification,” Technical report, Department ofComputer Science, National Taiwan University, 2003.

[72] F. Yu, E. Chang, Y. Q. Xu and H. Y. Shum, “Emotion Detection from Speech to Enrich Multimedia Content,”Advances in Multimedia Information Processing, pp.550-557, 2001.

[73] J. Gondohanindijo, E. Noersasongko, Pujiono, Muljono, A. Z. Fanani, Affandy, and R. S. Baskuri, “Comparison Method in Indonesian Emotion Speech Classification,” International Seminar on Application for Technology of Information and Communication (iSemantic), IEEE, Semarang, Indonesia, pp. 230-235, Sept 2019.

[74] M. Grimm, K. Kroschel and S. Narayanan, “Support Vector Regression for Automatic Recognition of Spontaneous Emotions in Speech,” International Conference on Acoustics, Speech and Signal Processing, IEEE, Honolulu, HI, USA,pp.1085-1088, April 2007.

[75] V. Sowmya and A. Rajeswari, “Speech Emotion Recognition for Tamil Language Speakers,” Machine Intelligence and Signal Processing, pp.125-136, 2020.

[76] L. Kerkeni, Y. Serrestou, K. Raoof, M. Mbarki, M. A. Mahjoub, andC. Cleder, “Automatic Speech Emotion Recognition using an Optimal Combination of Features based on EMD-TKEO”., Speech Communication, 2019.

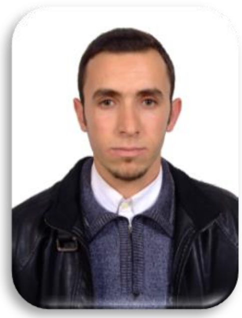
[77] S. Demircan, and H, Kahramanli, H. , “Application of fuzzy C-means clustering algorithm to spectral features for emotion classification from speech,” Neural Computing and Applications, vol. 29, no. 8, pp. 59–66, 2016.

[78] S. Kuchibhotla, H. D. Vankayalapati, R. S. Vaddi, and K. R. Anne, “A comparative analysis of classifiers in emotion recognition through acoustic features,” International Journal of Speech Technology, vol.17, no .4, pp. 401–408, 2014.

[79] . Zhou, Y. Sun, J. Zhang and Y. Yan, “Speech Emotion Recognition Using Both Spectral and Prosodic Features,” 2009 International Conference on Information Engineering and Computer Science, 2009.

[80] H.Kaya, A. A. Salah, A. Karpov, O. Frolova, A. Grigorev, and E. Lyakso, “Emotion, age, and gender classification in children’s speech by humans and machines,” Computer Speech & Language, vol. 46, pp. 268–283, 2017.

[81] O. T. C. Chen, J. J. Gu, P. T. Lu and J . Y. Ke,” Emotion-inspired age and gender recognition systems,” International Midwest Symposium on Circuits and Systems (MWSCAS), IEEE, Boise, ID, USA, pp.662-665, Aug 2012.



Houari Horkous PhD student, ENP Algiers, He received the engineering in 2013 from National Polytechnic School of Algiers (Algeria), is a researcher in Signal and Communications Laboratory, National Polytechnic School of Algiers .His research area include speech emotion recognition in the Algerian dialect.

Mhania Guerti is currently a Professor in the Department of Electronics and the Director of the Signal and Communications Laboratory, National Polytechnic School of Algiers (Algeria). She received her MSc in 1984, from the ILP Algiers in collaboration with the CNET-Lannion (France). She got her PhD from ICP-INPG (Grenoble France), in 1993. She is specialized in Speech and Language Processing. He has supervised a lot of students in Master and PhD degree. Her current research interests include the areas of speech processing, acoustics and audiovisual systems.