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

http://dx.doi.org/10.12785/ijcds/100125

Recognition of Emotions in the Algerian Dialect Speech

Houari Horkous[1] and Mhania Guerti[1]

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

Received 29 May 2020, Revised 30 Jul. 2020, Accepted 10 Aug. 2020, Published 8 Feb. 2021

Abstract: Emotion recognition from speech has been an important and challenging research. The performance of speech features and influence of emotions number on the systems of emotion recognition in the Algerian dialect speech are studied in this work. To achieve this aim, an emotional database of the Algerian dialect (ADED) is built. This database contains four emotions: anger, fear, sadness and neutral. Extraction of features is an important step in the recognition of speech emotion. The features extracted in this study are: pitch, intensity, duration, unvoiced frames, jitter, shimmer, HNR (Harmonic Noise Ratio), formants and MFCCs (Mel Frequency Cepstral Coefficients). Each recognition system needs a classifier, so our system is based on the KNN (K-Nearest Neighbor) method of classification. Different results are obtained, the higher recognition rates are given when using a combination of features including pitch, intensity, duration, unvoiced frames, jitter, shimmer, HNR and the MFCCs parameters. And the performance is influenced by the number of emotions included: 82.29 % when using the four emotions of ADED database, 84.02% when three emotions fear, anger and neutral are recognized and 87.50% when only the fear and neutral emotions are used in the recognition system.

Keywords: ADED, Pitch, Intensity, Duration, Unvoiced frames, Jitter, Shimmer, HNR, Formant, MFCC, KNN

1. Introduction

Speech emotion recognition (SER) is a very interesting area. It has numerous applications. SER was exploited in fields of psychology and psychiatry such as: analyzing human behavior [1], detection of moods [2], discriminating depressed speech [3], and analyzing emotions on the interview [4]. SER has board application in man-machine interaction applications such s intelligent tutoring system, robots, telephone banking, in-car board system, in call center, lie detection, sorting of voice mail and computer games [5].

Algeria suffered a civil war that lasted more than ten years. This war has caused the spread of fear, stress, sadness, anger and other emotions. These emotions have led to the appearance of many psychological diseases still remain at present. This work is based on the recognition of emotions in the Algerian dialect speech. The results obtained can be used in the psychology fields. In this work, an Algerian dialect emotional database (ADED) is constructed and used in the systems of emotion recognition. The Algerian dialect belongs to the Maghreb dialect. It is different from the Arabic dialects of the Middle East, because it is strongly influenced by other languages such as French, Italian, Turkish and Spanish, and this influence was caused by the long period of colonization [6]. Many works have focused on the characteristics and the sources of the Algerian dialect, others that have studied the sources. The influence of the French language on the Algerian dialect has been studied in [7]. A parallel corpus which included the Algerian dialect was built [8]. Speech corpus which named KALAM'DZ specialized in the Algerian Arabic sub-dialects has been built [9].

Extraction of features is important step for SER. In this field, many speech features have been used. The first works have been concentrated on the prosodic features such as pitch, intensity and duration [10-12]. Other features like voice quality, spectral and cepstral features have been exploited [13-16]. Spectral and prosodic parameters combination has been used for recognizing emotions from speech [17-18]. The classification is a fundamental stage in the SER systems. Different classifiers have been exploited for SER such as Gaussian mixture models (GMMs) [19], K-nearest Neighbors (KNN) [20-22], Support Vector Machines (SVM) [20] and Deep Neural Network (DNN) [22].

Our work is organized as follows. Some related works are presented in section two. The methodology is explained in section three. Section four describes the Algerian dialect emotional database (ADED). Extraction features used in this work and analysis of the emotions

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

influence on the features extracted is presented in section five. In Section six, experiments and results are shown. Finally, we finish by conclusion.

2. RELATED WORKS

Expressing an emotion involves numerous of physical and physiological changes. Emotions have been classified as perceiving, expression and felting emotion [23]. Emotions can be recognized by different ways such as facial expressions, brain signals, speech, etc. Systems of emotion recognition from speech have been defined as methodologies that classify the emotional speech [24]. Some databases, features and classifiers exploited in SER are presented briefly in this section.

There are many databases have been built and used to develop the SER systems. Databases which are oriented towards emotion recognition differ according to several criteria: the language, the number of emotions, the nature of the records (acted, natural or spontaneous), public or commercial databases, etc. Among the most applied databases is the emotional Berlin database (Emo-DB). This database comprises seven emotional states, anger, sadness, boredom, happiness, disgust, fear and neutral [25]. There are other famous databases, Danish database contains five emotions, neutral, angry, happy, sad and surprise [26]. Database of Polish emotional speech comprises six emotional states [27]. There are many other databases in different languages for example: the SAVEE database recorded in native English [28], KISMET [29] in American English language, emotional speech database in Slovene and English language [30], IITKGP-SEHSC in Indian language [31] and CEMO in French language [32]. In Arabic language there are a few databases in the field of SER. Natural Arabic speech database is built, and this database comprises three emotions: happiness, angry and surprise [33]. Tunisian actors were employed to build Tunisian dialect database which contains five emotions: happy, anger, fear, sadness and neutral emotions [34]. There are other databases have been built in Arabic language, emotional database (MEDB) in Moroccan [35], Emirati speech database (ESD) [36] and Egyptian Arabic speech emotion (EYASE) [37].

In the systems of SER, several speech features have been used. The first works of SER concentrated on the prosodic parameters [10-12]. The statistical values of prosodic parameters have been used to discriminate the emotions from speech [38]. In [39], pitch, energy, duration and formant have been exploited to classify speech emotions by using different classifiers. The MFCCs parameters have been widely used in SER systems [40-41]. In this field, MFCCs parameters gave better performance compared to pitch parameter [42]. HNR, jitter and shimmer with other spectral parameters have been applied for SER [13]. Jitter and shimmer parameters have been exploited to detect emotions in English and Hindi speech [43]. Combinations of prosodic, spectral and voice quality features have been used to analyze the performance on different datasets. The prosodic features represented by pitch and intensity, voice quality parameter represented by jitter and shimmer, and spectral features represented by MFCCs parameters and formants [44]. The formants and pitch features have been used in system of recognition of speech emotions [45]. Fundamental frequency, energy, formants and MFCC, LPC (Linear Prediction Coefficients) and PLP (Perceptual Linear Prediction) features have been exploited for recognizing speech emotions [20]. For recognition of speaker emotion [46], many features have been used such as pitch, jitter, shimmer, formants, MFCC and duration, and two databases, English and German databases have been applied.

After extraction of features, the last step of the emotions recognition systems is the classification. Several classifiers have been investigated for developing SER systems such as GMM, KNN, SVM, HMM, ANN, FLDA. Gaussian mixture models (GMM) classifier has been widely used for SER systems, in these area acoustic features have been modeled by this technique [42]. GMM achieved 92% accuracy when used to recognize emotions from speech [47]. KNN is a simple classification technique that used in the system of SER [48]. KNN classifier has been exploited with SVM algorithm for recognizing emotions in two different databases of Polish emotional speech [20]. To classify speech emotions, Emo-DB was exploited as database to assess the performance of recognition and SVM has been used as a classifier [49]. Hidden Markov Models (HMM) classifier had given precise results in the field of SER [50]. Artificial Neural Network (ANN) classifier has applied for recognizing four emotional states in the speech [51]. Different classifiers have been compared for their performance in the SER systems. Table I shows some works that have made comparisons between classifiers in SER field. The performance of KNN classifier has been compared to SVM algorithm [20]. Four classification techniques, GMM, ANN, K-means clustering and VQ (vector quantization) have been compared for recognizing emotions. From results, it has been concluded that the performance of classification technique depends on the type of database [44]. Fisher's Linear Discriminant Analysis (FLDA) classifier gave the best recognition rate when has compared with different classification algorithms such as GMM, ANN, KNN, and maximum likelihood classifier (MLC) in the SER field [52]. A hybrid classifier was built from two classifiers which were Gaussian mixture model and deep neural network (GMM-DNN). This classifier was compared with SVM and MLP (multilayer perceptron) classifiers in the SER system. It is indicated that the performance of the hybrid classifier was better than SVM and MLP classifiers [36]. GMM and KNN classifiers have been used and compared in the recognition system of six emotions: anger, sadness, happiness, surprise, neutral and fear [53]. 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 [54]. A feature selection method for the SER has been proposed. SVM, MLP and KNN classifiers have been exploited and compared for performing the experiments [55]. Four classification techniques, linear discriminant analysis (LDA), regularized discriminant analysis (RDA), SVM and KNN have been applied in SER systems. The databases of Berlin and Spanish emotional speech have been exploited to assess the performance. Results obtained indicated that RDA gave the best performance [56]. To identify emotions in the Basque speech database. Three classification experiments have been made. In the first experiment, GMM classifier was used with spectral features; in the second experiment, SVM classifier was used with prosodic features, and in the last experiment, GMM classifier was used with prosodic features. The classifier of the first experiment gave the best result [47]. The SVM classifier performed better than Recurrent Neural Networks (RNN) classifier for classifying speech emotions by using Berlin emotional database [57]. Several classifiers such as ANN, NB, KNN, and SVM have been used for classification emotions from speech. The highest classification rate (92.86%) was obtained by KNN and SVM classifiers [58].

TABLE I. SOME WORKS THAT HAVE MADE COMPARISONS BETWEEN CLASSIFIERS IN FIELD OF SER

The work	The compared classifiers
[20]	KNN and SVM
[44]	GMM, ANN, VQ and K-means clustering
[52]	FLDA, GMM, ANN, KNN and MLC
[36]	GMM-DNN, SVM and MLP
[53]	KNN and GMM
[54]	SVM, NN, RF and NB
[55]	SVM, MLP and KNN
[56]	LDA, RDA, SVM and KNN
[47]	GMM and SVM
[57]	RNN and SVM
[58]	KNN, SVM, ANN and NB

3. METHODOLOGY

The methodology in this work is as follows: Algerian dialect emotional database (ADED) is built from emotional segments collected from famous Algerian movies. This database including four emotional states: anger, fear, sadness and neutral. Then, different features are extracted from the speech segments of the ADED database. These features are: pitch, intensity, duration, unvoiced frames, jitter, shimmer, HNR, formants and MFCCs parameters. Next, the extracted features are input into a classifier to identify the different emotions. This step based on KNN technique which used as classifier. Fig. 1 illustrates the scheme of our system of recognition. To detect and to assess the appropriate features for recognizing emotions in the Algerian dialect speech, Different feature sets are formed to detect the feature sets that give higher performance. In the same time the influence of the number of emotions included in the system of recognition is studied. The response in the form of recognition of various emotions presented in the ADED database are obtained and studied for their recognition rate.

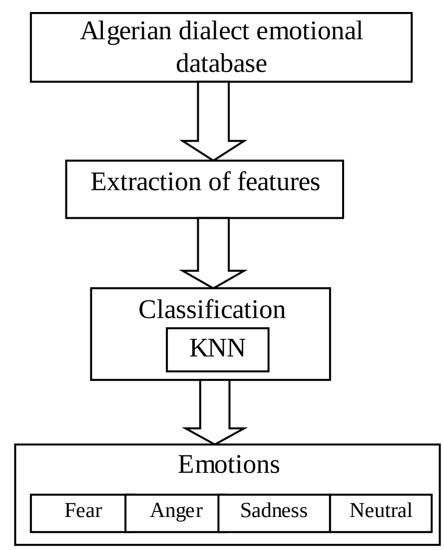


Figure 1. Scheme of the recognition system

4. ALGERIAN DIALECT EMOTIONAL DATABASE

As we mentioned earlier, an Algerian dialect emotional database (ADED) is created and used in this work. This database constructed from six famous movies in Algerian dialect. These movies describe the civil war of years (1992-2000) in addition to the period that followed. The titles of these films are: The repentant, Rachida, Bab el oued city, The doors of the sun, Fugitive, The price of the dream. The ADED database includes 32 actors (16 males and 16 females) from different regions of Algeria and different ages (from 18 to 60 years). This database contains 200 segments of duration ranging from 0.2 s to 3 s, and these segments are collected at sampling frequency of 44.1 kHz. The ADED database contains four emotions: fear, anger, sadness and neutral. Table II describes the number of segments for each emotion.

Some Algerian Dialect sentences belong in ADED are showed in Table III. The pronunciation of the represented sentences and their equivalents in standard Arabic and English are also showed in Table III. According to the sentences, some words in standard Arabic, some words in French language and the rest in the native dialect.

TABLE II. NUMBER OF SEGMENTS FOR EACH EMOTION

Emotions	Number of segments
Fear	**52**
Neutral	**48**
Anger	**52**
Sad	**48**
The sum	**200**

5. FEATURES EXTRACTION

Extraction of features is an essential and fundamental step in the SER system. In this section, the features extracted from the ADED database are presented and analyses are made to study the influence of the different emotions on the features extracted. These features are the statistical values of pitch (mean, maximum, minimum and range) and similar features of intensity, duration, unvoiced frames, jitter, shimmer, HNR, formants (formant 1 and formant 2) and MFCCs parameters. The range value equals the difference between the maximum value and minimum value. These parameters are extracted by PRAAT software [59], except the MFCCs parameters are extracted by MATLAB software. The averages of statistical values of the parameters extracted by PRAAT software for the four emotions, fear, anger, neutral and sadness are presented in Table IV.

TABLE III. SOME ALGERIAN DIALECT SENTENCES IN THE ADED DATABASE

Emotions	Sentences in English	Sentences in Arab standard	Sentences in Algerian dialect	Sentences pronunciation
Fear	- Don't be afraid! -Be quick!, be quick! -Remove your hands.	- لا تخافي! - أسرع أسرع ! - انزع يديك.	- ما تخافيش ! - أغصب أغصب! - نحي يدك.	-Ma tkhafich! - Aghssab!, aghssab! -Nahi yadak
Anger	-that's enough Djamila! -You will know the hijab. - I'm not your brother.	- يكفي يا جميلة! - ستعرفين الحجاب. - لست أخاك.	- خلاص جميلة! - تولي تعرفي الحجاب. - مانيش خوك.	- Khlass Djamila! - Twali taarfi el hijab. - Manich khouk.
Sadness	-There is no more pity. -Why? Why? -When we rode in the car.	- لم تبقى فيها رحمة. - لماذا؟ لماذا؟ - عندما ركبنا في السيارة.	- ما بقاتش فيها رحمة. - علاش؟ علاش؟ - كي ركبنا في طوموبيل.	- Ma bkattch fiha rahma. - Alach ? alach ? - ki rkabna fi tomobil.
Neutral	-You can't come with us. -Do not let me sleep. -I received threat messages.	- لا تستطيع المجيء معنا. - لا يتركني أنام. - رسائل التهديد التي وصلتني.	- ماتطيقش تجي معانا. - ما يخلينيش نرقد. - ليلاتخ دو مونا ص ألي وصلوني.	- Mattikch tji maana. - Ma ykhalinich nargod. -Lilatkh de menace ali wasslattni.

TABLE IV. AVERAGES OF STATISTICAL VALUES OF THE PARAMETERSEXTRACTED BY PRAAT SOFTWARE

Parameters	Fear	Anger	Neutral	Sadness
Mean of pitch (Hz)	**279.79**	269.09	190.60	196.89
Max of pitch (Hz)	**398.84**	387.36	273.25	299.61
Min of pitch (Hz)	163.18	131.87	**120.62**	125.09
Range of pitch(Hz)	235.65	**255.50**	152.62	174.52
Mean of intensity(dB)	69.97	**71.29**	70.30	62.69
Max of intensity(dB)	75.16	**77.35**	76.98	70.09
Min of intensity(dB)	57.20	55.45	51.11	**42.90**
Range of intensity(dB)	17.96	21.91	25.86	**27.18**
Duration(s)	0.88	1.34	1.28	**1.51**
Unvoiced frames (%)	24.12	23.92	24.36	**32.82**
Jitter (%)	3.29	3.19	2.89	**3.61**
Shimmer (%)	**16.03**	15.40	13.40	15.28
HNR (dB)	7.10	7.23	**8.55**	7.59
Formant1(Hz)	**715.89**	665.40	626.35	655.81
Formant2(Hz)	**1814.45**	1763.86	1755.50	1807.67

The fundamental frequency or pitch is very useful feature in recognition and classification of emotions from speech [38-39]. According to Table IV, it is noted that the statistical values of pitch: mean, max, min and range are different between the four emotions treated. It is observed that fear and anger emotions have the highest averages of mean of pitch and the highest averages of max of pitch, while the neutral state has low pitch mean. The sadness and neutral emotions are associated with a low range of pitch. Fig. 2, 3, 4 and 5 show the pitch contours of each emotion treated. When these contours are compared, it is remarked that the pitch varies in a visible way, this variation changes from emotion to other. Anger and fear emotions show more rise and falls in pitch contour compared to sadness and neutral states.

Intensity belongs to the prosodic parameters [38], and it provides a measure of the voice loudness. It is remarked in Table IV that the statistical values of intensity differ between the different emotions. Sadness emotion has a low value of intensity and has a broad range compared to the other states. The anger emotion has a higher max value of intensity.

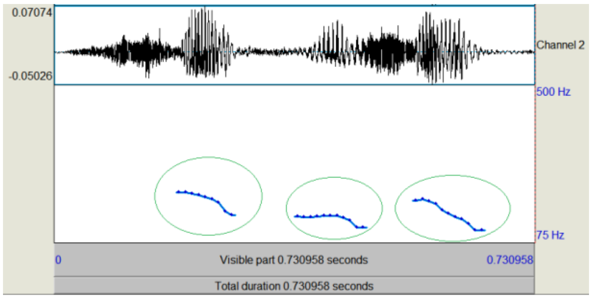


Figure 2. Contour of pitch of fear emotion

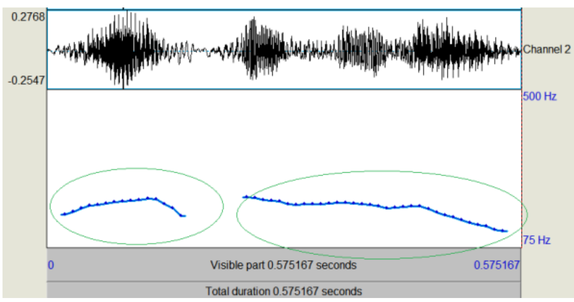


Figure 3. Contour of pitch of anger emotion

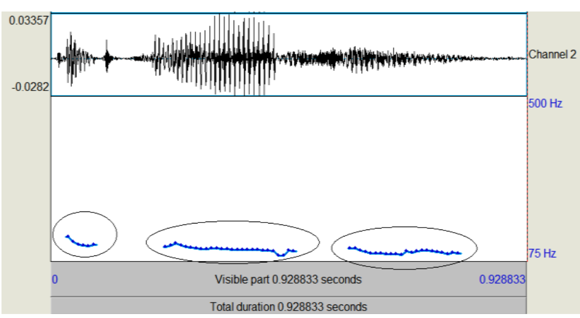


Figure 4. Contour of pitch of sadness emotion

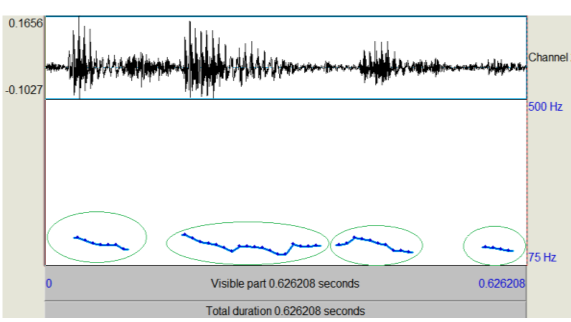


Figure 5. Contour of pitch ofneutral emotion

Duration is a prosodic parameter frequently used in the emotional speech field. It is observed in Table IV, each type of emotion having a different average duration compared to other emotions. Duration of fear emotion segment is very short. But duration concern the sadness state is long. The speech signals can be divided into voiced and unvoiced frames. The number of unvoiced frames has been used in many systems of emotion recognition [60]. According to Table IV, it can be seen that the number of unvoiced frames change between the emotions treated. Sadness emotion has the grand number of unvoiced frames, anger emotion has the lowest value, and the number of unvoiced frames is moderate in the other states.

Jitter, shimmer and HNR have been used to improve and to develop the system of SER [13], [43]. Jitter and shimmer have been defined respectively as variations of the fundamental period and variations of the speech waveform amplitude. The HNR gives the ratio of the harmonic part energy on the energy of the rest of the signal [61]. It is remarked in Table IV that the values of jitter, shimmer and HNR differ between the different types of emotions. The value of jitter is slightly higher in the sadness state, fear emotion has the highest value of the shimmer and HNR value is low in fear, anger and sadness emotion compared with the neutral state. The formants have been used as descriptors of the emotions in speech and they have been widely exploited in the SER systems [20], [45]. In theory, there is infinity of formants but the first four formants are widely exploited in practice. In this work we only use the first two formants. It is noted in Table IV that the values of formants1 and formants2 are higher in fear emotion than the other emotions, and the neutral state has the lowest values of formants. Fig. 6, 7, 8 and 9 show the contours of formants of each emotion treated. It is observed that the contour of formants varies according to the emotions. The anger emotion used their maximum power muscles but the neutral state used their minimum power.

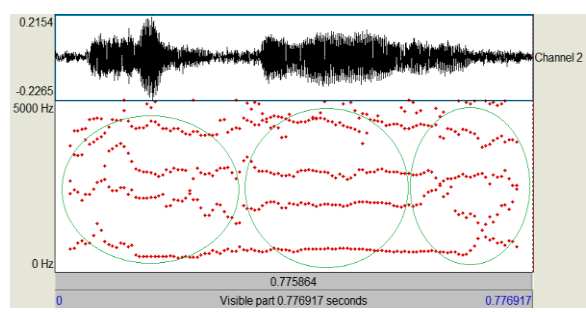


Figure 6. Contours of formants of fear emotion

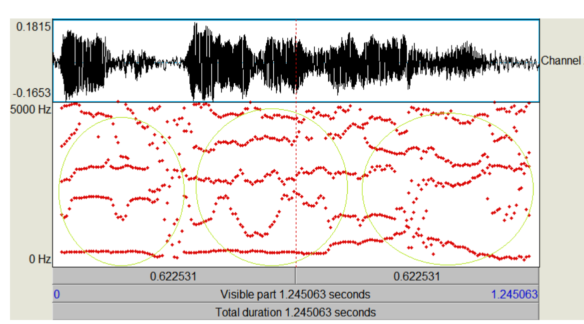


Figure 7. Contours of formants of anger emotion

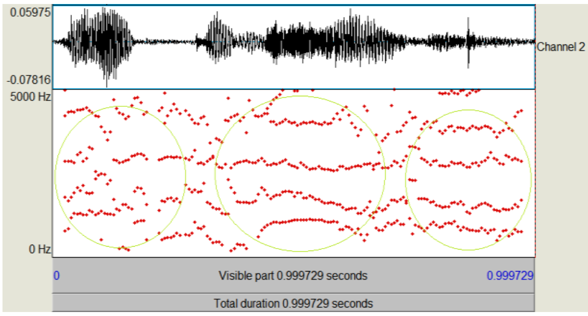


Figure 8. Contours of formants of sadness emotion

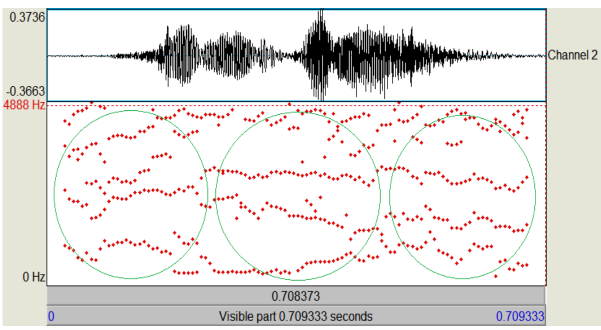


Figure 9. Contours of formants of neutral emotion

MFCCs parameters have been strongly exploited to develop and improve the performance of SER systems [20], [44-45]. These parameters are extracted from a frame of around 20 ms. The MFCC computation consists the following blocks: pre-emphasize, hamming window, FFT, triangular band-pass filter, logarithm and discrete cosine transformation (DCT) [62]. The sampling frequency used is 8 kHz. Fig. 10 illustrates an MFCC form of four emotions fear, anger, sadness and neutral. This figure is obtained by MATLAB software. In this work 20 MFCCs are used in the system of recognition.

It is remarked on Fig. 10 that the forms of MFCC of each emotion are varied according to the emotion. This difference which exists between MFCCs forms allows us used MFCCs parameters in the system of recognition of speech emotions.

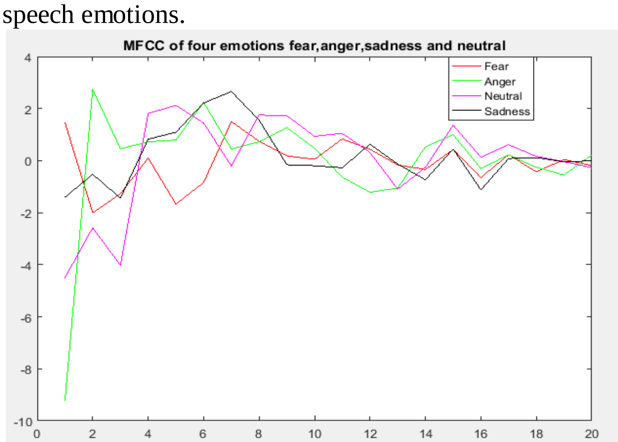


Figure 10. MFCC forms of four emotions.

6. EXPERIMENTS AND RESULT

To embody the system of emotion recognition, the features that extracted from the Algerian dialect emotional database are exploited in this section. As mentioned early the features are: mean of pitch (meanP), maximum of pitch (maxP), minimum of pitch (minP) and range of pitch (rangeP), and similar parameters of intensity (meanI, maxI, minI and rangeI), duration, unvoiced frames (unvF), jitter, shimmer, HNR, formants and MFCCs parameters. These features are used as features vectors in the systems of recognition. Different sets of features are formed as illustrated in the Table V. In our systems, K-Nearest Neighbors (KNN) technique is used as classifier. To achieve the performance of KNN, the database divided into training and testing groups. The K value determines the nearest neighbors number to the classified vector [63]. In our recognition system, the K value is equal to 3. This value is experimentally chosen in order to obtain the highest classification results.

This section is divided into three stages. In the first stage the recognition of fear emotion in relation to neutral emotion is performed. In the second stage, the state of anger is inserted with the first two emotions. Finally, the sadness emotion is added to the system in the third stage. The Experiments are made by MATLAB software.

Results of the first stage are presented in Table VI, and from these results we can study the recognition performance of fear and neutral state according to the features used. It is observed that the global recognition rate increase (87.50%) when jitter, shimmer, HNR, unvoiced frames and the MFCCs parameters are added in the system, but when formants parameters are inserted in the system of recognition, it is remarked that recognition rate decrease (82.42%). When only the MFCCs parameters are used, we obtain a low recognition rate (80.21%). The confusion matrix of the higher recognition rate is shown in Table VII.

TABLE V. DIFFERENT SETS OF FEATURES

Sets	Features
Set1	MFCCs
Set2	meanP, maxP, minP, rangeP, meanI, maxI, minI, rangeI, duration
Set3	meanP, maxP, minP, rangeP, meanI, maxI, minI, rangeI, duration, unvF, Jitter, Shimmer, HNR
Set4	meanP, maxP, minP, rangeP, meanI, maxI, minI, rangeI, duration, unvF, Jitter, Shimmer, HNR, Formants
Set5	meanP, maxP, minP, rangeP, meanI, maxI, minI, rangeI, duration, unvF, Jitter, Shimmer, HNR, MFCCs

TABLE VI. RECOGNITION RATES OF THE FIRST STAGE

Set of features	Fear	Neutral	Global recognition rate
Set1	87.50%	72.92%	80.21%
Set2	81.25 %	91.67%	86.46%
Set3	81.25%	93.75%	87.50%
Set4	72.92%	91.92%	82.42%
Set5	81.25%	93.75%	**87.50%**

TABLE VII. CONFUSION MATRIX OF THE HIGHER RECOGNITION RATE OF THE FIRST STAGE

Emotion	Fear	Neutral
Fear	**81.25 %**	18.75%
Neutral	6.25%	**93.75%**

In the second stage, the anger emotion is included in the system. The results are presented in Table VIII. It is observed that the global recognition rates decrease when compared to the first stage, this decrease is explained by the addition of another emotion (anger) in the system. The same observations of the first stage are seen, it is noticed that the higher recognition rate (84.02%) is obtained when the MFCCs parameters are used with the other parameters. It is remarked a low recognition rate if only used the parameters MFCCs (78.47%). The formants parameters have a very weak performance. Table IX shows the confusion matrix of the higher recognition rate in the second stage. Noted that the neutral state has a good recognition rate comparing with the other emotions.

TABLE VIII. RECOGNITION RATES OF THE SECOND STAGE

Set of features	Fear	Anger	Neutral	Global recognition rate
Set1	83.33%	79.17%	72.92%	78.47%
Set2	79.17 %	79.17%	87.50%	81.95%
Set3	81.25%	77.08%	87.50%	81.95%
Set4	81.25%	77.08%	89.58%	82.63%
Set5	83.33%	77.08%	91.66%	**84.02%**

TABLE IX. CONFUSION MATRIX OF THE HIGHER RECOGNITION RATE OF THE SECOND STAGE

Emotion	Fear	Anger	Neutral
Fear	**83.33 %**	8.33%	8.33%
Anger	8.33%	**77.08%**	14.58%
Neutral	4.17%	4.17%	**91.66%**

The sadness emotion is included in the third stage. Table X shows the results achieved. The same remarks of the two previous stages are noted, a good global recognition rate is obtained if there is a combination of MFCCs parameters with the other parameters (82.29%). A weak recognition rate is remarked when only the MFCCs parameters are used. The recognition rate is decrease if formants parameters are inserted in the system. Recognition rates of each emotion reduced compared to the two following stages, the neutral emotion has a higher recognition rate between the four emotions but the sadness had a low recognition rate. To ensure this remark, the confusion matrix for the higher recognition rate is showed in Table XI.

TABLE X. RECOGNITION RATES OF THE THIRD STAGE

Set of features	Fear	Anger	Neutral	Sadness	Global recognition rate
Set1	89.58%	83.33%	79.16%	68.75%	80.20%
Set2	81.25%	87.5%	87.5%	70.83%	81.77%
Set3	81.25%	81.25%	91.66%	72.91%	81.77%
Set4	83.33%	77.08%	87.5%	60.41%	77.08%
Set5	81.25%	81.25%	91.66%	75.00%	**82.29%**

TABLE XI. CONFUSION MATRIX OF THE HIGHER RECOGNITION RATE OF THE THIRD STAGE

Emotions	Fear	Anger	Neutral	Sadness
Fear	**81.25 %**	8.33%	6.25%	4.17%
Anger	6.25%	**81.25%**	10.42%	20.83%
Neutral	2.08%	2.08%	**91.67%**	4.17%
Sadness	8.33%	2.08%	14.58%	**75.00%**

A comparison between the higher recognition rates of each stage is illustrated in Fig. 11. The previous remarks are proven in this figure. The performance of recognition is reduced if the number of emotions increases in recognition systems. The recognition rate is 87.50% if there are two emotions fear and neutral. When the anger state is added to the recognition system, the recognition rate decreases around 84.02%. When sadness emotion is added to the other three states in the system, the recognition rate up to 82.29%. This decrease is caused by the emotions that share almost the same values of some parameters: the fear and anger emotion have higher average value of pitch and pitch peak, the emotions of sadness and neutral are associated with a low range of pitch value, the unvoiced frames is moderate in neural and fear states, HNR value is low in fear and anger emotion.

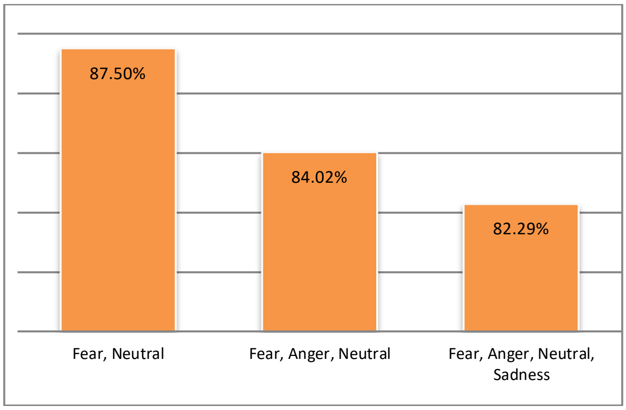


Figure 11. Comparison between the higher recognition rates of each stage

To recognize emotions in Egyptian Arabic speech [37], systems that used combinations of different features have achieved higher recognition rates than systems that used the same types of features. Different features have been used in this work such as: pitch, intensity, formants, MFCCs, wavelet and long-term average spectrum. In the same work, the author concluded that the recognition rates have been influenced by the number of emotions in the system. The performance was higher when classifying three emotions, anger, sadness and neutral than when classifying four emotions, anger, sadness, neutral and happiness. The same observationsare noted in our work that the higher performance is given by a combination of features. And the performance of the recognition system is influenced by the number of emotions. To recognize emotions in Berlin and Spanish emotional speech databases, the system that used a combination of prosodic (energy and pitch) and spectral features (MFCCs) gave better performance when compared with performance of the system that used individual features (prosodic or spectral features) [56]. These results correspond to the results obtained in our work, i.e. the performance of using features combination is better than use individual features.

7. CONCLUSION

This work was focused on the recognition of emotions from Algerian dialect speech. The ADED database was built and exploited to extract the features concerning the emotions treated. Several experiments were performed to evaluate the performance of features extracted on the recognition systems.

The obtained results showed us that the used of MFCCs parameters with the other parameters gave recognition rate very important. And formants parameters had a weak performance on the recognition system. The higher recognition rate was of 87.50% when only the fear and neutral emotions were recognized, respectively 84.02% when the anger state was introduced in the system of recognition, and 82.29% when the sadness emotion was inserted in the system.

It was concluded from the result obtained that the combination of pitch, intensity, duration, unvoiced frames, jitter, shimmer, HNR and the MFCCs parameters improve the performance of system to recognize emotions in the Algerian dialect emotional database. It was also concluded that the performance of recognition was influenced by the number of emotions included in the system. The acceptable recognition percentages obtained for recognizing the four emotional states, fear, anger, sadness and neutral justify that the parameters extracted from the speech can help in psychology research.

For future works, other emotions such surprise and disgust and other speech features can be including in our emotion recognition system for certain applications.

REFERENCES

[1] 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.

[2] 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

[3] E. Moore, M. A. Clements, J. W. Peifer, andL. 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.

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

[5] 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.

[6] [http://www.tlfq.ulaval.ca/axl/europe/danemark.htm.

[7] M. A. Menacera, O. Mellaa , D. Fohra , D. Jouveta , D. Langloisa , K. Smaila, 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.

[8] S. Harrat, K. Meftouh, M. Abbas, W. K. Hidouci, K. Smaïli,"An Algerian dialect: Study and Resources," International journal of advanced computer science and applications, vol. 7, no. 3, pp.384-396, 2016.

[9] S. Bougrine, A. Chorana, A. Lakhdari, and H. Cherroun,"Toward aWeb-based Speech Corpus for Algerian Arabic Dialectal Varieties," Proceedings of The Third Arabic Natural Language Processing Workshop (WANLP), Valencia, Spain, pp. 138–146,April 2017.

[10] 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.

[11] 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.

[12] 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.

[13] B. Schuller, A. Batliner, D. Seppi, S. Steidl, T. Vogt, J. Wagner, L. Devillers, L. Vidrascu, N. Amir, L. Kessous, andV.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.

[14] B. Schuller, R. Müller, F. Eyben, J. Gast, B. Hörnler, M. Wöllmer, G. Rigoll, A. Höthker, H. Konosu, "Being bored? Recognising natural interest by extensive audiovisual integration for real-life application,"Image and Vision Computing, vol. 27, no. 12, pp. 1760-1774, 2009.

[15] 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.

[16] 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.

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

[18] 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.

[19] 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.

[20] 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.

[21] 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.

[22] K. Tarunika, R. Pradeeba, andP. 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.

[23] D. Chaudhary, N. P. Singh, and S. Singh,"A Survey on Autonomous Techniques for Music Classification based on Human Emotions Recognition," International Journal of Computing and Digital Systems, vol. 9, no. 3. 2020.

[24] 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

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

[26] 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.

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

[28] 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.

[29] C. Breazeal and L. Aryananda,." Recognition of affective communicative intent in robot-directed speech," AutonomousRobots, vol. 12, pp. 83–104, 2002.

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

[31] 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.

[32] L. Vidrascu, andL. 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.

[33] 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.

[34] 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.

[35] A. Agrima, A. Farchi, L. Elmazouzi, I. Mounir, andB. 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.

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

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

[39] M. Swain, A. Routray, P. Kabisatpathy, and J. N. Kundu,"Study of prosodic feature extraction for multidialectal Odia speech emotion recognition," IEEE Region 10 Conference (TENCON),IEEE, Singapore, Singapore, pp. 1644-1649, Nov 2016.

[40] M. Shami and W. Verhelst, "An evaluation of the robustness of existing supervised machine learning approaches to the classification of emotions in speech," Speech Commun,vol. 49, no .3, pp. 201–212, 2007.

[41] N. A. Zaidan, M. S. Salam, M. S, "MFCC Global Features Selection in Improving Speech Emotion Recognition Rate," Advances in Machine Learning and Signal Processing, vol. 387, pp. 141–153, 2016.

[42] D. Neiberg, K. Elenius, and K. Laskowski, "Emotion recognition in spontaneous speech using GMMs," In INTERSPEECH 2006 - ICSLP, Pittsburgh, Pennsylvania, pp. 809–812, September 2006

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

[44] 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.

[45] D. Gharavian, M. Sheikhan, F. 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.

[46] D. Bitouk, R. Verma, and A. Nenkova, “Class-level spectral features for emotion recognition,” Speech Commun, vol. 52, no. 7, pp. 613–625, 2010.

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

[48] M. W. Bhatti, Y. Wang and L. Guan, “A neural network approach for human emotion recognition in speech,” In proceedings of IEEE International Symposium on Circuits and Systems,IEEE, Vancouver, BC, Canada, vol. 2, pp. 181-184, May 2004.

[49] T. Seehapoch, andS. 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.

[50] S. M. Feraru, andD. 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.

[51] 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.

[52] Y. Wang, andL. Guan, “An investigation of speech-based human emotion recognition,” In IEEE 6th workshop on multimedia signal processing, IEEE, Siena, Italy,pp. 15–18, Oct 2004.

[53] R. B. Lanjewar, andM. 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

[54] 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.

[55] T. Özseven, “novel feature selection method for speech emotion recognition,” Applied Acoustics, vol. 146, pp. 320–326, 2019.

[56] L. Kerkeni, Y. Serrestou, K. Raoof, M. Mbarki, M. A. Mahjoub, and C. Cleder Autom, “Speech Emotion Recognition using an Optimal Combination of Features based on EMD-TKEO,” Speech Communication, vol. 114, pp. 22-35, 2019.

[57] 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.

[58] 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.

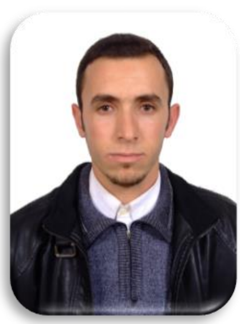
[59] P. Boersma, D. Weenink, “Praat, a system for doing phonetics by computer,” Glot Int, vol. 5, no. 9, pp. 341–345, 2002.

[60] A. Vuppala, and S. R. Kadiri, “Neutral to anger speech conversion using non-uniform duration modification,” 9th International Conference on Industrial and Information Systems (ICIIS), IEEE, Gwalior, India,Dec 2014.

[61] R. Montaño, andF. Alías, F, “The role of prosody and voice quality in indirect storytelling speech: A cross-narrator perspective in four European languages,” Speech Communication, vol. 88, pp. 1–16, 2017.

[62] G. Zhang, J. Yin, Q.Liu, and C. Yang,“The Fixed-Point Optimization of Mel Frequency Cepstrum Coefficients for Speech Recognition,” Proceedings of 2011 6th International Forum on Strategic Technology, IEEE, Heilongjiang, Harbin, pp. 1172-1175. Aug 2011.

[63] H. Mustafa, A. Abdulla, and A. Hadeel, “Predicting Software Faults Based on K-Nearest Neighbors Classification,” International Journal of Computing and Digital Systems. vol. 8 . no. 05 .2019



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 includes 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.