International Information and Engineering Technology Association
Advancing the World of Information and Engineering

Traitement du Signal
Vol. 37, No. 3, June, 2020, pp. 413-423
Journal homepage: http://iieta.org/journals/ts

Study the Influence of Gender and Age in Recognition of Emotions from Algerian Dialect Speech

Horkous Houari*, Mhania Guerti

Laboratory Signal and Communication, Ecole National Polytechnique, Algiers 16200, Algeria

Corresponding Author Email: houari.horkous@g.enp.edu.dz

https://doi.org/10.18280/ts.370308

Received: 19 January 2020
Accepted: 20 April 2020

Keywords:
ADED, emotion, HNR, KNN, LDA, recognition, speech, SVM

ABSTRACT

Speech emotions Recognition is a very interesting area of research. In this work, the influence of gender and age on the speech emotions recognition in Algerian Dialect is studied. And on the other hand, the influence of speech emotion types on the classification of gender and age is also studied. An Algerian Dialect Emotional Database (ADED) is used in this work. ADED database is exploited for extracting the features that used in the systems of recognition and classification. These features are the statistic values of pitch and intensity, unvoiced frames, jitter, shimmer, HNR and MFCCs parameters. Analyzes based on gender and age are made to detect the influence of the four emotions on the parameters extracted. A parallel classifier composed of three classifiers, Support Vector Machines (SVM), K-Nearest Neighbor (KNN), and Linear Discriminant Analysis (LDA) is used in the recognition and classification systems. The results obtained show us that the performance of emotions recognition systems is influenced by gender and age i.e. the distinction between each gender class and each age interval in the recognition systems improves the performance compared to the systems without distinction. It was showed in the results also that the classifications of gender classes and age intervals were strongly influenced by the type of emotion.

1. INTRODUCTION

Speech is one of the natural means of human communication. It considers as a rich source of information like; physiological state, gender and emotional states. So, speech becomes one of the important sources that used for emotions recognition. Speech emotions recognition (SER) has appeared as one of the important and fundamental research areas. SER has potentially extensive applications in several areas. SER is very useful for human-machine interaction applications such as intelligent tutoring system, robots, telephone banking, emotion recognition in call center, sorting of voice mail, in-car board system, lie detection, and computer games [1]. SER has been exploited by medical doctors; emotional content of the patient has been used as a diagnostic tool for various disorders [2, 3]. In psychology and psychiatry fields, SER has been widely used to detect many diseases such as automatic detection of moods like fatigue, depression and anxiety [4]. Analyses and comparison of speech features have been performed in people diagnosed with depression [5].

Algeria has suffered with a civil war caused the appearance of many psychiatric and psychological diseases still remain at present resulting from different emotions. So, this work is based on the recognition of emotions in the Algerian dialect speech. Our results can be exploited in the field of psychotherapy and psychology. In this work, we focus on recognition of speech emotions, and classification of gender and age under emotional speech in Algerian dialect. Thus, the influence of gender classes and age intervals on speech emotion recognition is studied. On the other hand, the influence of emotion types on classification of gender and age is also studied. Algerian Dialect Emotional Database (ADED) is used in this paper. There are four types of emotion in ADED database: fear, anger, sad and neutral state.

Algerian Dialect belongs to the Maghreb dialect, and it is very different from the Arabic dialects of the Middle East. The Algerian Dialect is influenced by French, Berber and Turkish [6]. Many works have been performed on this dialect [7-10]. A system for identifying dialects based on words and sentences was presented. The Algerian dialect has been applied with other Arabic dialects [7]. An approach that identifies Arabic Algerian dialects is constructed; this approach based on prosodic speech information [8].

Systems of SER are developed by extracting features from emotional speech databases. The first studies concentrated on prosodic features such as energy, duration, pitch and their derivatives [11, 12]. Voice quality parameters like harmonic noise ratio (HNR), shimmer and jitter have been widely investigated to recognize emotions in speech [13-17]. Spectral features have been used for discriminating the emotions, among the most exploited spectral parameters are the MFCCs parameters [18, 19]. Using different combinations of features in SER systems improved the performance in many works. Parameters of pitch, intensity, jitter, shimmer, MFCCs and formants have been considered to analyze the performance of some databases in the field of emotions recognition [20]. Spectral, prosody and voice quality features have been studied for recognizing emotions on different types of corpus [21]. Classification and recognition systems are based on classifiers models. Several classifiers are explored to develop the SER systems. Among the most used classification techniques are Support vector machines (SVM), K-Nearest Neighbor (KNN), Artificial Neural Network (ANN), Gaussian Mixture Models (GMMs) and Hidden Markov Models (HMM) [22-31].

This work is structured as follows. In the next section some related works in SER are presented. Our methodology is explained in section 3. Whereas, the Algerian Dialect emotional database (ADED) is described in section 4. In section 5, extraction of parameters and analyzes based on gender and age are made to detect the influence of the four emotions on the parameters extracted. Experimental results are discussed in section 6. Finally, our conclusion is presented.

2. RELATED WORKS

SER systems have been defined as methodologies that process and classify speech signals to detect emotions [32]. For each system of SER, there are different types of emotional speech databases that built. According to these databases there are different speech features extracted. These lasts have been used by classification models and techniques to recognize the specific emotions. In this section, databases, features and classifiers used for emotions recognition are discussed in brief.

The development of a database is a requirement to build systems of emotions recognition. So, to assess the performance of SER systems, it is essential to construct an appropriate database. In the SER, databases have been studied in three parts which are acted, elicited and natural speech emotional databases [32]. Among the most famous databases are: Berlin database of emotional speech (Emo-DB), Polish emotional speech database, and Danish (DES) database. Emo-DB contains around 500 utterances spoken in seven emotional states [33]. The Polish emotional speech database consisting of 288 speech segments in six emotions [34]. DES database contains five emotional states: neutral, angry, happy, sad and surprise [35]. There are many databases in different languages such as Chinese emotional speech corpus in Mandarin language, Italian emotional speech database (EMOVO) in Italian language, Japanese emotional speech database (Keio-ESD) in Japanese language, RECOLA speech database in French and Turkish emotional speech database (TURES) in Turkish language [32]. There are a few emotional databases in Arabic speech. Tunisian dialect database [36], this database was registered by professional Tunisian actors. To recognize sentiment in the natural Arabic, speech emotional corpus was constructed, this corpus contains three emotions: happiness, angry and surprise [37]. An Emirati speech database (ESD) was built to exploit in SER, and this database was constructed by local Emirati speakers [38]. A sample data of Algerian dialect containing audio-visual recordings was presented [39].

In literature, numerous features have been explored in SER. Early studies focused on prosodic features, so there are many studies which focused on different aspects of the prosodic features. There is a correlation between the prosodic features and the emotion [40, 41]. Different acoustic features including fundamental frequency, energy and duration have been studied to classify five emotional states in speech [12]. Statistic values of pitch and energy have been used for discriminating emotions in speech [42]. Voice quality features are widely used in the field of SER. The correlation between the voice quality features and the emotional states of the speech is very strong [43]. Jitter and shimmer features have been proposed for recognizing emotions in English and Hindi speech [44]. Voice quality features and prosody features have been exploited for discriminating four emotions from Chinese natural emotional speech. The system performance has been increasing if the voice quality features were added to the system when compared to the system that used only the prosodic features [45]. Spectral parameters are among the most used parameters in SER. Spectral features like MFCCs parameters and linear prediction cepstral coefficients (LPCCs) have been used to discriminate emotions in speech [18]. In the field of SER, MFCCs parameters have performed better compared to the pitch features [46]. To improve the performance of SER systems, combinations of different features are also used in numerous works. Combination of features extracted from the shape of speech signal has been exploited for recognizing emotions in speech. These features are MFCCs parameters, spectral centroid, spectral skewness, and spectral pitch chroma [47]. To detect emotions in speech, comparison of the accuracy performance of different parameters as fundamental frequency, formants, energy, MFCC, LPC (Linear Prediction Coefficients), and PLP (Perceptual Linear Prediction) coefficients has been performed [48]. In automatic SER, feature as duration, energy, pitch, spectrum, MFCC, PLP, NHR (Noise to Harmonic Ratio), HNR, shimmer, jitter, Wavelets, teager operator have been used [49].

To recognize speech emotions, many machine learning algorithms and classifiers have been applied to perform this task. SVM classifier has been widely exploited to classify emotions in speech. The SVM techniques are based on a supervised learning mechanism that finds an optimal hyperplane for linearly separable patterns, which testing and training dataset before to classification [50]. To recognize emotions in Emo-DB and Chinese emotional database, different combinations of the features have been compared by using SVM as classifier [51]. SVM classifier has been established to obtain the best accuracy for classifying emotions in three corpuses, German, English and Polish [52]. ANN classifier is strongly used for several types of classifications. Different acoustic features have been investigated to recognize emotional states in speech by using a neural network classifier [26]. There are particular types of neural networks named convolutional neural networks (CNNs) which have been used in SER [53, 54]. Studies in [55, 56] indicated that KNN classifier is simplest and very useful in the field of SER systems. KNN classifier [55] was applied to differentiate the anxiety emotion in Emo-DB. GMMs have been successfully used in classification problems including the recognition of speech emotions [29, 57]. Spectral features have been employed to recognize emotions from speech signals by using GMM classifier [58]. Many previous works have been focused on HMMs as a classifier in SER systems [31, 58]. HMMs have been used as classifier in recognition system to classify six emotional states from speech including anger, happiness, joy, fear, sadness, disgust [58]. In the field of SER, different classifiers including, GMM, KNN, ANN and Fisher's linear discriminate analysis (FLDA) have been compared. The recognition accuracies showed that FLDA gave the higher performance [59]. In the field of SER on emotional Chinese speech, SVM classifier gave higher performance compared to other classifiers which are linear discriminant classifiers, radial basis function neutral network and KNN [22]. In many works, a number of classifiers combined to improve the performance. Hybrid classifier composed of Gaussian mixture model and deep neural network (GMM-DNN) has been proposed. This hybrid classifier has been compared with SVM and multilayer perceptron (MLP) classifiers. The result indicated that the performance of the hybrid classifier is higher than SVMs and MLP classifiers performance [38].

3. METHODOLOGY

Our aim in this work is to study the impact of the age and gender on emotions recognition in the Algerian dialect speech. And on the other hand, the influence of emotion types on gender and age classification is studied. So emotional segments of Algerian dialect collected from Algerian movies are used as database (ADED) in the recognition and classification systems. To achieve the purpose, our methodology is to divide the ADED database according to gender i.e. male speech segments and female speech segments. The same process is applied on the ADED database but this time according to age. The database is divided into two intervals based on the actors age, speech segments of actors aged 18 to 35 years and speech segments of actors aged 36 to 60 years. Then, statistics values of prosodies features such as pitch and intensity, voice quality parameters such as unvoiced frames, jitter, shimmer, HNR and MFCCs parameters are extracted from the speech segments. After features extraction, there is the classification step. A parallel classifier system composed of SVM, KNN and LDA techniques are used as classifier in this step. The scheme of the parallel classifier is illustrated in Figure 1. The recognition rate of the systems is the maximum recognition rate between the three classifiers (KNN, SVM and LDA) in each system. The parallel classifiers are widely used in the fields of recognition and classification of speech emotions [60, 61]. And the performance of these classifiers is better compared to the individual classifiers.

Our work is divided into two parts, to study the gender and age influence in the first part, emotions recognition systems are constructed to recognize the four types of emotions in gender distinction, in age distinction and in the entire database without distinction. The response in the form of recognition of various emotions presented in the ADED database are obtained and studied for their recognition rate. The emotions recognition scheme of the first part is shown in Figure 2. A comparison is made between the results of the different recognition systems. To study to influence of emotion types on the gender and age classification in the second part, we built a gender and age classification systems which can classify the gender classes (male, female) and age intervals (18 to 35 years, 36 to 60 years) respectively under each type of emotion. The response in the form of classification of gender and age are obtained and studied for their accuracy. Scheme of gender and age classification systems under different emotions is shown in Figure 3. A comparison is made between the results of the different classification systems.

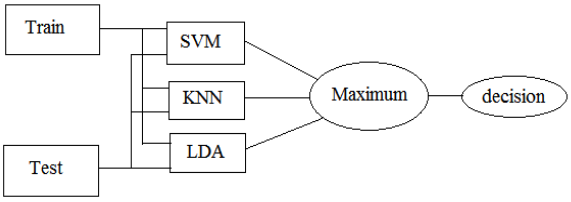

Figure 1. Scheme of the parallel classifier

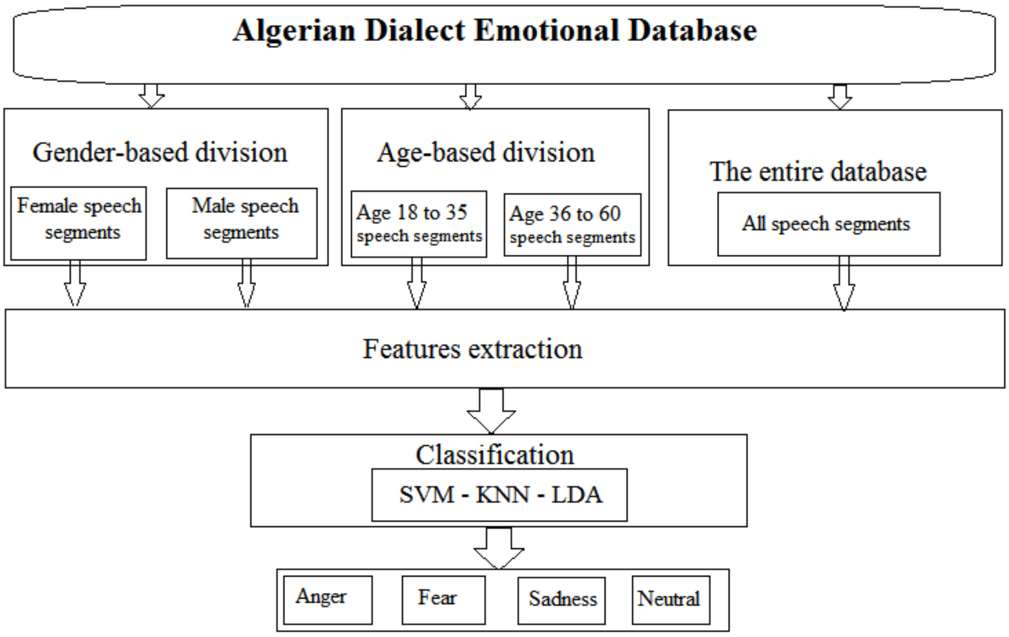

Figure 2. Scheme of emotions recognition systems based on gender and age

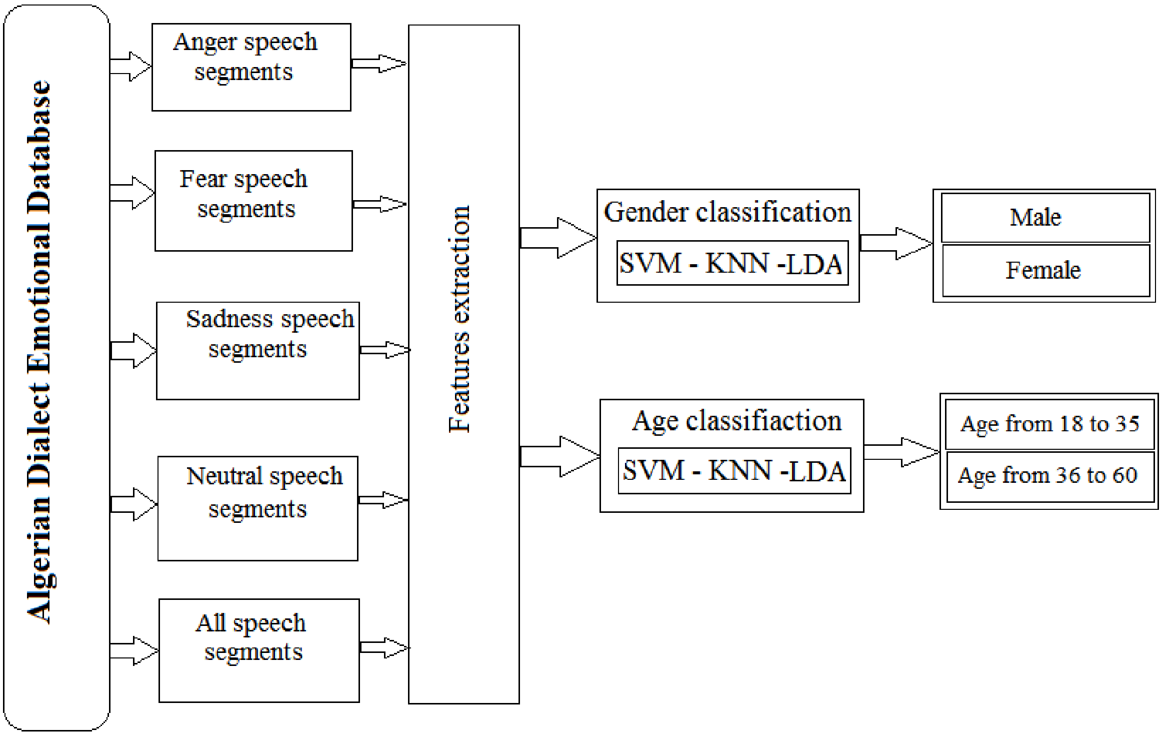

Figure 3. Scheme of gender and age classification under different emotions

4. ALGERIAN DIALECT EMOTIONAL DATABASE

The development of the SER systems is heavily influenced by the quality of the emotional speech databases [62]. As we mentioned earlier, the purpose of our work is concentrated on the Algerian dialect speech. So, in this paper an Algerian Dialect Emotional Database (ADED) is used. This database was constructed from six famous movies in Algerian Dialect. These movies describe the civil war crises in addition to the following period. The ADED database contains four emotional states including fear, anger, sad and neutral. To define the emotional classes, annotation scheme is defined by XML (eXtensive Mark-up Language) under ANVIL. Anvil is a tool for the annotation of audiovisual material [63, 64]. The description of the audio content is carried out in the context of the sequence with the help of the video support. The sequence contains the description situational and emotional content. Each sequence is split into segments. 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 48000 kHz. The speech segments included in this database are expressed by 32 actors (16 males and 16 females) from different regions of country and different ages between 18 and 60 years. 18 actors aged from 18 to 35 years, and the others actors aged from 36 to 60 years. The numbers of segments for each emotion are showed in Table 1.

Table 1. Number of segments for each emotion

Emotions	Segments number
Fear	52
Neutral	48
Anger	52
Sad	48
The sum	200

To know the content of the Algerian dialect emotional database, some Algerian Dialect sentences in the Algerian dialect emotional database are shown in Table 2. The pronunciation of the represented sentences and their equivalents in standard Arabic and English are also showed. According to the Table 2, it is remarked that there are standard Arabic words, words in native dialect and some words in French language.

5. FEATURES EXTRACTION

In this section, the emotional speech is analyzed with respect to different features. So, analyzes are done to detect the influence of the four emotions fear, anger, sad and neutral on the features extracted. These analyze are based on gender classes and age intervals. The features extracted are the statistic values of prosodies features such as: mean of pitch (meanP), maximum of pitch (maxP), minimum of pitch (minP), range of pitch (rangeP) and similar parameters of intensity (meanI, maxI, minI and rangeI), voice quality parameters such as: unvoiced frame (unvF), jitter, shimmer, HNR and MFCCs parameters. Range i.e. the difference between the maximum and minimum. The features are extracted by Praat software [65], except the MFCCs parameters are extracted by MATLAB software. To analyze the influence of emotions on the features extracted based on gender classes and age intervals, the averages values of features extracted of each emotion of male and female are given in Table 3. Table 4 illustrates the averages values of features extracted of each emotion in two age intervals, the first interval contains the parameters extracted from speech segments of actors aged 18 to 35 years and the second interval from 36 to 60 years.

Table 2. Some of Algerian Dialect sentences in the ADED

Emotions	Sentences in English	Sentences in Arab standard	Sentences in Algerian dialect	Sentences pronunciation
Fear	- Don’t be afraid! - Be quick!, be quick! - No, no!	- لا تخافي! - أسرع أسرع! - لا لا!	-ما تخافيش ! - أغصب أغصب! - لا لا !	-Ma tkhafich -Aghssab!, aghssab! - Lala!
Anger	- I do not like. - Now you heard me! - I'm not your brother.	- لا أحبها. - الآن سمعتني! - لست أخاك.	- ما نبغيهاش. - دركا سمعتني! - مانيش خوك .	- Ma nabghihach - Dorka smaattni! - Manich khouk
Sad	.- I wished death. - I was watching her until I slept - When we rode in the car.	- تمنيت الموت. - كنت أشاهدها حتى أنام - عندما ركبنا في السيارة.	- تمنيت الموت. - كنت نقابلها حتى تديني عيني. - كي ركبنا في طوموبيل.	-Tmaniit el maout. -Kont nkablha hatta tadini aini. - Ki rkabna fi tomobil.
Neural	- You can’t come with us. -I called you for Selma - Threat messages I received	- لا تستطيع المجيء معنا. - اتصلت بك من أجل سلمى - رسائل التهديد التي وصلتني	- ماتطيقش تجي معانا. - عيطتلك على جال سلمى - ليلاتخ دو مونا ص ألي وصلوني	-Mattikch tji maana. - Ayattlak ela jal Selma -Lilatkh de menace ali wasslattni.

Table 3. Averages values of the parameters extracted with Praat software concerning male and female gender

Gender	Male				Female			
Parameters	**Fear**	**Anger**	**Neutral**	**Sad**	**Fear**	**Anger**	**Neutral**	**Sad**
meanP (Hz)	219.49	**226.36**	157.18	113.20	**314.51**	305.66	230.10	231.35
maxP (Hz)	313.44	**334.56**	230.17	207.71	**448.02**	438.89	324.16	337.45
minP (Hz)	139.33	114.37	99.22	**81.36**	176.92	**140.24**	145.91	143.09
rangeP (Hz)	174.11	**219.79**	130.95	126.35	271.10	**298.65**	178.24	194.36
meanI (dB)	69.24	**71.88**	69.64	62.37	70.38	70.77	**71.07**	63.21
maxI (dB)	74.45	**77.93**	76.00	69.40	75.57	76.87	**78.13**	70.37
minI (dB)	56.40	57.24	52.87	**45.95**	57.65	53.91	49.03	**41.65**
rangeI (dB)	18.05	20.69	23.12	**23.45**	17.92	22.95	**29.11**	28.72
unvF (%)	30.67	22.86	25.38	**33.35**	20.35	24.84	23.15	**32.53**
Jitter (%)	**3.93**	3.50	2.92	3.82	2.92	2.92	2.86	**3.53**
Shimmer (%)	**18.09**	16.83	14.24	17.55	**14.85**	14.17	12.40	14.35
HNR (dB)	6.19	6.12	**7.41**	5.78	7.63	8.19	**9.91**	8.34

Table 4. Averages values of the parameters extracted with Praat software concerning the age intervals

Age interval	Age interval of 18 to 35 years				Age interval of 36 to 60 years			
Parameters	**Fear**	**Anger**	**Neutral**	**Sad**	**Fear**	**Anger**	**Neutral**	**Sad**
meanP (Hz)	196.41	**288.49**	196.44	210.38	234.69	**244.56**	184.76	188.05
maxP (Hz)	**410.41**	407.49	269.70	322.97	368.00	**369.61**	266.89	284.30
minP (Hz)	173.51	129.10	**118.87**	135.31	135.15	127.71	122.38	**118.39**
rangeP (Hz)	236.70	**278.39**	160.73	187.66	232.84	**241.90**	144.52	165.91
meanI (dB)	**70.47**	70.17	69.45	55.61	68.59	**72.70**	71.14	67.75
maxI (dB)	75.85	**76.31**	76.16	63.41	74.68	**78.68**	77.79	74.46
minI (dB)	56.75	54.47	50.56	**34.90**	56.83	56.68	51.66	**48.15**
rangeI (dB)	19.11	21.84	25.60	**28.51**	18.03	22.00	26.13	**26.32**
unvF (%)	22.66	23.14	22.85	**43.31**	**28.08**	24.91	25.87	25.96
Jitter (%)	3.03	3.15	2.89	**3.68**	**3.98**	3.24	2.89	3.56
Shimmer (%)	15.20	15.61	13.58	**15.72**	**18.29**	15.12	13.21	14.99
HNR (dB)	7.73	7.89	**8.32**	7.59	5.41	6.41	**8.78**	7.59

5.1 Pitch

Pitch (F0) is the frequency at which the vocal cords are opening and closing [66]. Pitch is a very important parameter for emotions classification in speech [12, 42]. The analysis of the speech emotions is done according to gender and age. From Table 3, the first remark that appeared, the average of statistic values of pitch in female gender are higher than the statistic values of pitch in male gender in all the emotions studied. In female gender, it is observed that the pitch statistic values of fear emotion as meanP and maxP are higher compared to the other states. The anger emotion has the wide range of pitch. For male gender, it is noted that the average of statistic values of anger pitch is higher compared to the other states. The sadness emotion has the lowest value of pitch. To analyze the age influence, from Table 4, it is remarked that anger emotion has the higher meanP and the wide rangeP in age interval of 18 to 35 years. In the same interval, it is observed that the fear state has the highest maxP and the neutral state has the lowest value of pitch. In the second interval of age, it is noted that the anger emotion has higher statistic values of meanP, maxP and rangeP compared to the others emotions and the lowest value of pitch is given by sadness emotion.

5.2 Intensity

Intensity provides a representation that reflects amplitude variation of speech signals [32]. Thus, intensity can also be considered as one of the distinguishing features of emotions in SER systems [12, 42]. In male gender, it is observed in Table 3 that the anger emotion has the highest average value of meanI and maxI. The sadness emotion has the wide range of intensity. In female gender, it is remarked that the average of statistic values of intensity of neutral state as meanI, maxI and rangeI are higher compared to the other emotions. The sadness state has the lowest value of intensity. Concerning the age influence, according to Table 4, it is remarked that anger state has the highest value of maxI in the two intervals. The lowest value of minI and the wide value of rangeI are observed in sadness emotion.

5.3 Unvoiced frames

The speech signals can be segmented into voiced and unvoiced frames. For SER systems, voiced and unvoiced segments are investigated in many previous works [67]. According to Table 3, it is noticed that the sadness emotion has the grand number of unvF while fear emotion has the lowest number in both male and female gender. For age influence, it is remarked in Table 4 that the sadness emotion has the grand number of unvF in the age interval of 18 to 35 years but in the second interval fear emotion has the grand number of unvoiced frames.

5.4 Jitter, shimmer and HNR

Jitter can be defined as variations of the fundamental period. Shimmer; similarly, to jitter, shimmer can be defined as variations of the speech waveform amplitude [68]. The variation of pitch period and the energy variation depend on the emotional state [69]. HNR measures the ratio between the energy of the harmonic part and the energy of the rest of the signal [68]. It is remarked in Table 3 that the jitter value in the fear state is slightly higher compared to the others states in male gender. But in female gender, the sadness emotion has the highest value of jitter. Fear emotion has the highest value of shimmer in the two genders. Neutral state has the highest HNR value in both genders. Concerning the age influence, in Table 4 it is observed that sadness emotion has the highest values of jitter and shimmer in the first age interval (18 to 35 years) but in the second interval fear emotion has the highest jitter and shimmer values. HNR value is higher in neutral state in the two age intervals.

5.5 Mel frequency cepstral coefficients (MFCCs)

MFCCs parameters are widely used to improve the performance and to develop the SER systems. MFCCs parameters are used in this work because these parameters were widely used in the recognition and classification of emotions in speech based on gender and age and they were used in the classification of gender and age under speech emotions [70-73]. 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 [74]. MFCCs parameters belong to the family of the cepstrals descriptors which are accurate representations of short time power spectrum of a sound. These parameters are extracted from a frame of around 20 ms because the variation in speech signal within 20 ms is ignorable. The sampling frequency used was 8 kHz. In this work 20 MFCCs are used in the system of classification. The MFCC computation consists the following Bloks: Pre-emphasize, Hamming window, FFT, Triangular band-pass filter, Logarithm and discrete cosine transformation (DCT) [75].

6. EXPERIMENTS AND RESULTS

Classification results are presented in this section. As we mentioned earlier, our work is divided into two parts, and the purpose of the first part is to study the gender and age influence in the recognition of fear, anger, sadness and neutral emotions in ADED database. In the second part, the purpose is to study to influence of emotion types to identify the gender classes and age intervals in the ADED database. The parameters that extracted in the previous section are used in the systems of recognition and classification. These parameters are the statistic values of prosodies parameters (pitch and intensity), voice quality parameters (unvoiced frames, jitter, shimmer and HNR) and MFCCs parameters. These systems are based on parallel classifier composed of three classifiers, KNN, SVM and LDA as mentioned in section of methodology. For each speech segment, the classifier is trained with the extracted features. These features are input into the classifier as features vectors. 60% of the segments of database are used as training set, and 40% of the segments are used as test set. The Experiments are made by Matlab software.

6.1 Influence of gender and age in emotions recognition

In this part is the influence of gender and age on the recognition of fear, anger, sadness and neutral emotions in ADED database is studied.

6.1.1 Influence of gender

Three experiments are made to detect the gender influence on the SER. In the first experiment, the entire database (ADED) is used without gender distinction. Male speech segments of ADED database are used in the second experiment while in the third experiment female speech segments of ADED database are used in the system of recognition. The experiments results are shown in Table 5. By comparison the results, it is remarked that the accuracy of systems that used only male or female speech database has enhanced compared to the accuracy of system that used the entire database speech. A slight improvement of recognition rate is observed in female speech. It is noted that the recognition rate of neutral state is higher when using the entire database but when using the male speech segments, sadness emotion gives us the higher recognition rate while recognition accuracy of fear and neutral states are higher in female speech segments. By comparing the results with other works which used different databases: Beihang University mandarin emotion speech database (BHUES) [76] and Berlin and SmartKom databases [77]. The same remark is observed i.e. the accuracy of systems that used only male or female speech database is higher compared to the accuracy of system without gender distinction.

6.1.2 Influence of age

To detect the age influence on the emotions recognition in Algerian dialect emotional database, three experiments are performed. The entire database is used without age distinction in the first experiment. In the second experiment, speech segments belonging to the age interval from 18 to 35 years are used in the system of recognition. Speech segments located in the age interval between 36 and 60 years are used in the third experiment. Table 6 illustrates results of the experiments. It is observed that the average value of the correct recognition is higher in the two age intervals compared to the average value in the system which used the entire database. In the age interval from 18 to 35 years, it is remarked that anger and sadness emotions have higher recognition rate but in the second age interval the accuracy recognition of fear emotion is very higher compared to others emotions. There is confusion between fear and anger emotions and between sadness and neutral in the three experiments. The similarity and the convergence between some parameters of speech lead to this confusion, especially, the statistics value of pitch is higher in fear and anger, while the convergence in the range of intensity and number of frames in sadness and neutral state.

6.2 Influence of speech emotions type in gender and age classification

In this part the influence of emotion types on gender and age classification in the ADED database is studied.

6.2.1 Influence of speech emotions type in gender classification

The accuracies of gender classification under four emotions are shown in the Table 7. According to the result of Table 7, it is observed a low classification rate when the entire database is used in the system of classification but maximum accuracy is obtained when used only the same type of emotion. The accuracy is maximum when using a single type of emotion because there are large differences between the parameters values in the two genders, male and female in each type of emotion, especially in terms of statistic values of pitch feature and vocal quality parameters. Gender classes were classified under four emotions, angry, happy, calm and sad [78]. By comparing our results to their results, the accuracy in our system is better.

6.2.2 Influence of speech emotions type in gender classification

The accuracies of age classification under four emotions are illustrated in Table 8. It is remarked higher accuracies when used only the same type of speech emotion. However, a low accuracy is obtained in the system that used the entire database (ADED). The best recognition rate is noted under fear emotion. The classification accuracy of 18 to 35 years class is higher than the accuracy of 36 to 60 years class in neutral and sadness states while it is the opposite case under anger state. Maximum accuracy is remarked in fear emotion because under this emotion there are clear differences between the speech parameters of the two age intervals, in particular, in terms of pitch, unvoiced frames, shimmer and HNR. Three ages of young, adult and senior were identified under different emotions [78]. It remarked that the identification of age groups was influenced by the type of emotion. The performance of our classification system is higher comparing to their results.

Table 5. Confusion matrices for emotions recognition with gender influence

	All speech segments of ADED				**Male speech segments of ADED**				**Female speech segments of ADED**			
	Fear	**Anger**	**Neutral**	**Sad**	**Fear**	**Anger**	**Neutral**	**Sad**	**Fear**	**Anger**	**Neutral**	**Sad**
Fear	79.17%	10.42%	6.25%	4.16%	90.91%	9.09%	0%	0%	95.45%	4.55%	0%	0%
Anger	4.16%	83.33%	10.42%	2.08%	4.55%	77.27%	13.63%	4.55%	13.64%	81.81%	4.55%	0%
Neutral	2.08%	2.08%	91.67%	4.16%	4.55%	9.09%	81.81%	4.55%	0%	4.55%	95.45%	0%
Sad	8.33%	2.08%	14.58%	75%	4.55%	0%	0%	95.45%	0%	0%	22.73%	77.27%
Average	**82.29%**				**86.36%**				**87.50%**			

Table 6. Confusion matrices for emotions recognition with age influence

	All speech segments of ADED				**Age interval from 18 to 35 years**				**Age interval from 36 to 60 years**			
	Fear	**Anger**	**Neutral**	**Sad**	**Fear**	**Anger**	**Neutral**	**Sad**	**Fear**	**Anger**	**Neutral**	**Sad**
Fear	79.17%	10.42%	6.25%	4.16%	84.21%	10.53%	0%	5.26%	100%	0%	0%	0%
Anger	4.16%	83.33%	10.42%	2.08%	0%	94.74%	5.26%	0%	15.79%	68.42%	0%	15.79%
Neutral	2.08%	2.08%	91.67%	4.16%	5.26%	5.26%	78.95%	10.53%	0%	0%	89.47%	10.53%
Sad	8.33%	2.08%	14.58%	75%	0%	0%	5.26%	94.74%	0%	5.26%	5.26%	89.47%
Average	**82.29%**				**88.16%**				**86.84%**			

Table 7. Confusion matrices for gender classification under four emotions

	All speech segments of ADED		**Fear speech segments**		**Anger speech segments**		**Neural speech segments**		**Sadness speech segments**	
	Male	**Female**	**Male**	**Female**	**Male**	**Female**	**Male**	**Female**	**Male**	**Female**
Male	89.28%	10.72%	100%	0%	100%	0%	100%	0%	100%	0%
Female	12.50%	87.50%	0%	100%	0%	100%	0%	100%	0%	100%
Average	**88.39%**		**100%**		**100%**		**100%**		**100%**	

Table 8. Confusion matrices for ager classification under four emotions

	ADED		**Fear speech segments**		**Anger speech segments**		**Neural speech segments**		**Sadness speech segments**	
	18 to 35	**36 to 60**	**18 to 35**	**36 to 60**	**18 to 35**	**36 to 60**	**18 to 35**	**36 to 60**	**18 to 35**	**36 to 60**
18 to 35	83.93%	16.07%	100%	0%	95.45%	4.55%	100%	0%	100%	0%
36 to 60	17.86%	82.14%	0%	100%	0%	100%	4.55%	95.45%	4.55%	95.45%
Average	**83.03%**		**100%**		**97.72%**		**97.72%**		**97.72%**	

7. CONCLUSION

In this work, we presented a study that evaluated the influence of gender and age on SER. On the other hand, influence of emotion types on gender and age classification was also studied. The Algerian dialect emotional database (ADED) was used for the features extraction. These lasts were the statistic values of pitch and intensity parameters, unvoiced frames, jitter, shimmer, HNR and MFCCs parameters. An analysis was made to detect the influence of the four emotions: fear, anger, sadness and neutral on the parameters extracted. This Analysis was based on gender classes and age intervals. Our experiment was divided into two parts, the influence of gender classes and age intervals for recognizing emotions was studied in the first part. However, in the second part the influence of emotions type for classifying the gender classes and age intervals was studied. A parallel classifier composed of three classifiers, KNN, SVM and LDA was used in the recognition and classification systems.

The experimental results demonstrated that building recognition systems with gender distinction gives a better performance rather than having one recognition system for both genders. The same remark was noted on the age intervals. In male gender, the recognition rate of sadness emotion was better compared to the other emotions but in female gender, the recognition rates of fear and neutral emotions were the best. It was showed in the results also that the classifications of gender classes and age intervals were influenced by the type of emotion. The accuracy of gender classes classification was very high under all the emotions included. In age interval, the accuracy of classification was very high under fear emotion.

The accuracy of the result can help psychotherapists and psychologists to detect people suffering from diseases caused by the emotions. In the future, the present work can be extended for other emotions and other parameters for emotions recognition in Algerian dialect speech.

REFERENCES

[1] Ramakrishnan, S., El Emary, I.M.M. (2011). Speech emotion recognition approaches in human computer interaction. Telecommunication Systems, 52(3): 1467-1478. https://doi.org/10.1007/s11235-011-9624-Z

[2] France, D.J., Shiavi, R.G., Silverman, S., Silverman, M., Wilkes, M. (2000). Acoustical properties of speech as indicators of depression and suicidal risk. IEEE Transactions on Biomedical Engineering, 47(7): 829-837. https://doi.org/10.1109/10.846676

[3] Polzehl, T., Schmitt, A., Metze, F., Wagner, M. (2011). Anger recognition in speech using acoustic and linguistic cues. Speech Communication, 53(9-10): 1198-1209. https://doi.org/10.1016/j.specom.2011.05.002

[4] Pantic, M., Pentland, A., Nijholt, A., Huang, T.S. (2007). Human computing and machine understanding of human behavior: A survey. Lecture Notes in Computer Science, 47-71.

[5] France, D.J., Shiavi, R.G., Silverman, S., Silverman, M., Wilkes, M. (2000). Acoustical properties of speech as indicators of depression and suicidal risk. IEEE Transactions on Biomedical Engineering, 47(7): 829-837. https://doi.org/10.1109/10.846676

[6] Menacer, M.A., Mella, O., Fohr, D., Jouvet, D., Langlois, D., Smaïli, K. (2017). Development of the Arabic Loria automatic speech recognition system (ALASR) and its evaluation for Algerian dialect. Procedia Computer Science, 117: 81-88. https://doi.org/10.1016/j.procs.2017.10.096

[7] Lichouri, M., Abbas, M., Freihat, A.A., Megtouf, D.E.H. (2018). Word-level vs sentence-level language identification: Application to Algerian and Arabic dialects. Procedia Computer Science, 142: 246-253. https://doi.org/10.1016/j.procs.2018.10.484

[8] Bougrine, S., Cherroun, H., Ziadi, D. (2018). Prosody-based spoken Algerian Arabic dialect identification. Procedia Computer Science, 128: 9-17. https://doi.org/10.1016/j.procs.2018.03.002

[9] Bougrine, S., Cherroun, H., Ziadi, D., Lakhdari, A., Chorana, A. (2016). Toward a rich Arabic speech parallel corpus for Algerian sub-dialects. The 2nd Workshop on Arabic Corpora and Processing Tools, Valencia, Spain, pp. 2-10.

[10] Djellab, M., Amrouche, A., Bouridane, A., Mehallegue, N. (2016). Algerian modern colloquial Arabic speech Corpus (AMCASC): Regional accents recognition within complex socio-linguistic environments. Language Resources and Evaluation, 51(3): 613-641. https://doi.org/10.1007/s10579-016-9347-6

[11] Lee, C.M., Narayanan. S.S. (2005). Toward detecting emotion in spoken Dialogs. IEEE Transactions on Speech and Audio Processing, 13(2): 293-303. https://doi.org/10.1109/TSA.2004.838534

[12] McGilloway, S., Cowie, R., Douglas-Cowie, E., Gielen, S., Westerdijk, M., Stroeve, S. (2000). Approaching automatic recognition of emotion from voice: A rough benchmark. ISCA Workshop Speech Emotion, Northern Ireland, pp. 207-212.

[13] Lalitha, S., Madhavan, A., Bhushan, B., Saketh, S. (2014). Speech emotion recognition. International Conference on Advances in Electronics, Computers and Communications (ICAECC), Bangalore, pp. 1-4. https://doi.org/10.1109/ICAECC.2014.7002390

[14] Batliner, A., Seppi, D., Steidl, S., Vogt, T., Wagner, J., Devillers, L., Vidrascu, L., Amir, N., Kessous, L., Aharonson, V. (2007). The relevance of feature type for the automatic classification of emotional user states: low level descriptors and functionals. Proceedings of Interspeech, Antwerp, Belgium, pp. 2253-2256.

[15] Kostoulas, T., Ganchev, T., Lazaridis, A., Fakotakis, N. (2010). Enhancing emotion recognition from speech through feature selection. Lecture Notes in Computer Science, 6231: 338-344. https://doi.org/10.1007/978-3-642-15760-8_43

[16] Wu, C.H., Liang, W.B. (2011). Emotion recognition of affective speech based on multiple classifiers using acoustic-prosodic information and semantic labels. IEEE Transactions on Affective Computing, 2(1): 10-21. https://doi.org/10.1109/t-affc.2010.16

[17] Schuller, B., Müller, R., Eyben, F., Gast, J., Hörnler, B., Wöllmer, M., Rigoll, G., Höthker, A., Konosu, H. (2009). Being bored? Recognising natural interest by extensive audiovisual integration for real-life application. Image and Vision Computing, 27: 1760-1774. https://doi.org/10.1016/j.imavis.2009.02.013

[18] Sreenivasa, K., Shashidhar, R., Koolagudi, G. (2013). Emotion recognition using vocal tract information. Springer Briefs in Electrical and Computer Engineering. Springer-Verlag, New York.

[19] Sheikhan, M., Gharavian, D., Ashoftedel, F. (2011). Using DTW neural-based MFCC warping to improve emotional speech recognition. Neural Computing and Applications, 21(7): 1765-1773. https://doi.org/10.1007/s00521-011-0620-8

[20] Koolagudi, S.G., Murthy, Y.V.S., Bhaskar, S.P. (2018). Choice of a classifier, based on properties of a dataset: case study-speech emotion recognition. International Journal of Speech Technology, 21(1): 167-183. https://doi.org/10.1007/s10772-018-9495-8

[21] Li, Y., Chao, L., Liu, Y., Bao, W., Tao, J. (2015). From simulated speech to natural speech, what are the robust features for emotion recognition? International Conference on Affective Computing and Intelligent Interaction (ACII), Xi'an, China, pp. 368-373. https://doi.org/10.1109/acii.2015.7344597

[22] Yu, C., Tian, Q., Cheng, F., Zhang, S. (2011). Speech emotion recognition using support vector machines. Advanced Research on Computer Science and Information Engineering, 1(20): 215-220. https://doi.org/10.1007/978-3-642-21402-8_35

[23] Pan, Y., Shen, P., Shen, L. (2012). Speech emotion recognition using support vector machine. International Journal of Smart Home, 6(2): 101-107.

[24] Han, Z., Wang, J. (2017). Speech emotion recognition based on Gaussian kernel nonlinear proximal support vector machine. Chinese Automation Congress (CAC). Jinan, China, pp. 2513-2516. https://doi.org/10.1109/cac.2017.8243198

[25] Song, Y., Huang, J., Zhou, D., Zha, H., Giles, C.L. (2007). IKNN: Informative K-nearest neighbor pattern classification. European Conference on Principles of Data Mining and Knowledge Discovery, Warsaw, Poland, pp. 248-264. https://doi.org/10.1007/978-3-540-74976-9_25

[26] Dai, K., Fell, H.J., MacAuslan, J. (2008). Recognizing emotion in speech using neural networks. Telehealth and Assistive Technologies, 31: 38-43.

[27] Firoz, S.A., Raji, S.A., Babu, A.P. (2009). Automatic emotion recognition speech using artificial neural networks with gender dependent databases. International Conference on Advances in Computing, Control, & Telecommunication Technologies, Trivandrum, India, pp. 162-164. https://doi.org/10.1109/ACT.2009.49

[28] Luengo, I., Navas, E., Hernez, I., Snchez, I. (2005). Automatic emotion recognition using prosodic parameters. Interspeech, Lisbon, Portugal, pp. 493-496.

[29] Tang, H., Chu, S.M., Hasegawa-Johnson, M., Huang, T.S. (2009). Emotion recognition from speech via boosted Gaussian mixture models. In IEEE International Conference on Multimedia and Expo, New York, USA, pp. 294-297. https://doi.org/10.1109/icme.2009.5202493

[30] Feraru, S.M., Schuller, D. (2015). Cross-language acoustic emotion recognition: an overview and some tendencies. International Conference on Affective Computing and Intelligent Interaction (ACII), Xi'an, China, pp. 125-131. https://doi.org/10.1109/acii.2015.7344561

[31] Womack, B.D., Hansen, J.H.L. (1999). N-channel hidden Markov models for combined stressed speech classification and recognition. IEEE Transactions on Speech and Audio Processing, 7(6): 668-677. https://doi.org/10.1109/89.799692

[32] Akçay, M.B., Oğuz, K. (2019). Speech emotion recognition: Emotional models, databases, features, preprocessing methods, supporting modalities, and classifiers. Speech Communication, 2682: 56-76. https://doi.org/10.1016/j.specom.2019.12.001

[33] Hassan, A. (2012). On automatic emotion classification using acoustic features. Ph.D. dissertation. Faculty of Pysical and Applied Sciences, University of Southampton.

[34] Staroniewicz, P., Majewski, W. (2009). Polish emotional speech database - recording and preliminary validation. Lecture Notes in Computer Science, 42-49. https://doi.org/10.1007/978-3-642-03320-9_5

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

[36] Meddeb, M., Hichem, K., Alimi, A. (2015). Speech emotion recognition based on Arabic features. 15th International Conference on Intelligent Systems Design and Applications (ISDA15), Marrakesh, Morocco, pp. 46-51. https://doi.org/10.1109/isda.2015.7489165

[37] Klaylat, S., Osman, Z., Hamandi, L., Zantout, R. (2018). Emotion recognition in Arabic speech. Analog Integrated Circuits and Signal Processing, 96: 337-351. https://doi.org/10.1007/s10470-018-1142-4

[38] Shahin, I., Nassif, A.B., Hamsa, S. (2019). Emotion recognition using hybrid Gaussian mixture model and deep neural network. IEEE Access, 7: 26777-26787. https://doi.org/10.1109/access.2019.2901352

[39] Dahmani, H., Hussein. H., Sickendiek. B.M., Jokisch. O. (2019). Natural Arabic language resources for emotion recognition in Algerian dialect. Arabic Language Processing: from Theory to Practice, pp. 18-33.

[40] Frick, R.W. (1985). Communicating emotion: The role of prosodic features. Psychological Bulletin, 97(3): 412-429. https://doi.org/10.1037/0033-2909.97.3.412

[41] Cowie, R., Douglas-Cowie, E. (1996). Automatic statistical analysis of the signal and prosodic signs of emotion in speech. Proceeding of Fourth International Conference on Spoken Language Processing (ICSLP), Philadelphia, USA, pp. 1989-1992. https://doi.org/10.1109/icslp.1996.608027

[42] Schroder, M. (2001). Emotional speech synthesis: A review. Seventh European Conference on Speech Communication and Technology, Aalborg, Denmark.

[43] Cowie, R., Douglas-Cowie, E., Tsapatsoulis, N., Votsis, G., Kollias, S., Fellenz, W., Taylor, J.G. (2001). Emotion recognition in human-computer interaction. IEEE Signal Process, 18(1): 32-80. https://doi.org/10.1109/79.911197

[44] Jacob, A. (2016). Speech emotion recognition based on minimal voice quality features. International Conference on Communication and Signal Processing (ICCSP), Melmaruvathur, pp. 886-890. https://doi.org/10.1109/ICCSP.2016.7754275

[45] Zhang, S. (2008). Emotion recognition in Chinese natural speech by combining prosody and voice quality features. Advances in Neural Networks-ISNN 2008, 457-464. https://doi.org/10.1007/978-3-540-87734-9_52

[46] Neiberg, D., Elenius, K., Laskowski, K. (2006). Emotion recognition in spontaneous speech using GMMs. Interspeech, Pittsburgh, USA, pp. 809-812.

[47] Navyasri, R.M., RajeswarRao, R., DaveeduRaju, A., Ramakrishnamurthy, M. (2017). Robust features for emotion recognition from speech by using Gaussian

mixture model classification. Information and Communication Technology for Intelligent Systems, 2: 437-444. https://doi.org/10.1007/978-3-319-63645-0_50

[48] Kamińska, D., Sapiński, T., Anbarjafari, G. (2017). Efficiency of chosen speech descriptors in relation to emotion recognition. EURASIP Journal on Audio, Speech, and Music Processing, 2017(1). https://doi.org/10.1186/s13636-017-0100-x

[49] Batline, A., Schuller, B., Seppi, D., Steidl, S., Devillers, L., Vidrascu, L., Vogt, T., Aharonson, V., Amir, N. (2010). The automatic recognition of emotions in speech. Emotion-Oriented Systems, Springer, Berlin, Heidelberg, pp. 71-94. https://doi.org/10.1007/978-3-642-15184-2_6

[50] Singh, M.K., Nandan, D., Kumar, S. (2019). Statistical analysis of lower and raised pitch voice signal and its efficiency calculation. Traitement du Signal, 36(5): 455-461. https://doi.org/10.18280/ts.360511

[51] Pan, Y., Shen, P., Shen, L. (2012). Speech emotion recognition using support vector machine. International Journal of Smart Home, 6(2): 101-107.

[52] Alonso, J.B., Cabrera, J., Medina, M., Travieso, C.M. (2015). New approach in quantification of emotional intensity from the speech signal: Emotional temperature. Experts Systems with Applications, 42(24): 9554-9564. https://doi.org/10.1016/j.eswa.2015.07.062

[53] Zhao, J., Mao, X., Chen, L. (2019). Speech emotion recognition using deep 1D & 2D CNN LSTM networks. Biomedical Signal Processing and Control, 47: 312-323. https://doi.org/10.1016/j.bspc.2018.08.035

[54] Kim, J., Englebienne, G., Truong, K.P., Evers, V. (2017). Towards speech emotion recognition “in the wild” using aggregated corpora and deep multi-task learning. Interspeech, Stockholm, Sweden, pp. 1113-1117.

[55] Dan Zbancioc, M., Feraru, S.M. (2015). A study about the automatic recognition of the anxiety emotional state using Emo-DB. E-Health and Bioengineering Conference (EHB), Iasi, Romania. https://doi.org/10.1109/ehb.2015.7391506

[56] Kuchibhotla, S., Vankayalapati, H.D., Anne, K.R. (2016). An optimal two stage feature selection for speech emotion recognition using acoustic features. International Journal of Speech Technology, 19(4): 657-667. https://doi.org/10.1007/s10772-016-9358-0

[57] Bozkurt, E., Erzin, E., Erdem, C.E., Erdem, A.T. (2010). Use of line spectral frequencies for emotion recognition from speech. International Conference on Pattern Recognition, Istanbul, Turkey, pp. 3708-3711. https://doi.org/10.1109/icpr.2010.903

[58] Nwe, T.L., Foo, S.W., De Silva, L.C. (2003). Speech emotion recognition using hidden Markov models. Speech Communication, 41(4): 603-623. https://doi.org/10.1016/s0167-6393(03)00099-2

[59] Wang, Y., Guan, L. (2004). An investigation of speech-based human emotion recognition. In IEEE 6th Workshop on Multimedia Signal Processing, New York, USA, pp. 15-18. https://doi.org/10.1109/mmsp.2004.1436403

[60] Morrison, D., Wang, R., De Silva, L.C. (2007). Ensemble methods for spoken emotion recognition in call-centres. Speech Communication, 49(2): 98-112. https://doi.org/10.1016/j.specom.2006.11.004

[61] Milton, A., Tamil Selvi, S. (2014). Class-specific multiple classifiers scheme to recognize emotions from speech signals. Computer Speech & Language, 28(3): 727-742. https://doi.org/10.1016/j.csl.2013.08.004

[62] Koolagudi, S.G., Murthy, Y.V.S., Bhaskar, S.P. (2018). Choice of a classifier, based on properties of a dataset: case study-speech emotion recognition. International Journal of Speech Technology, 21(1): 167-183. https://doi.org/10.1007/s10772-018-9495-8

[63] Kipp, M. (2001). Anvil-a generic annotation tool for multimodal dialogue. Eurospeech, Aalborg, Denmark, pp. 1367-1370.

[64] Devillers, L., Abrilian, S., Martin, J.C. (2005). Representing real-life emotions in audiovisual data with non basic emotional patterns and context features. Lecture Notes in Computer Science, 519-526. https://doi.org/10.1007/1157354867

[65] Boersma, P., Weenink, D. (2002). Praat, a system for doing phonetics by computer. Glot International, 5(9): 341-345. https://hdl.handle.net/11245/1.200596

[66] Özseven, T., Düğenci, M. (2018). Speech acoustic (SPAC): A novel tool for speech feature extraction and classification. Applied Acoustics, 136: 1-8. https://doi.org/10.1016/j.apacoust.2018.02.009

[67] Vasquez-Correa, J.C., Garcia, N., Orozco-Arroyave, J. R., Arias-Londono, J.D., Vargas-Bonilla, J.F., Noth, E. (2015). Emotion recognition from speech under environmental noise conditions using wavelet decomposition. International Carnahan Conference on Security Technology (ICCST), Taipei, Taiwan, pp. 247-252. https://doi.org/10.1109/ccst.2015.7389690

[68] Montaño, R., Alías, F. (2017). The role of prosody and voice quality in indirect storytelling speech: A cross-narrator perspective in four European languages. Speech Communication, 88: 1-16. https://doi.org/10.1016/j.specom.2017.01.007

[69] Farrús, M., Hernando, J. (2009). Using Jitter and Shimmer in speaker verification. IET Signal Processing, 3(4): 247-257. https://doi.org/10.1049/iet-spr.2008.0147

[70] H. Kaya., Salah. A.A., Karpov. A, Frolova. O, Grigorev. A., Lyakso. E. (2017). Emotion, age, and gender classification in children's speech by humans and machines. Computer Speech & Language, 46: 268-283. https://doi.org/10.1016/j.csl.2017.06.002

[71] Mittal, T., Barthwal, A., Koolagudi, S.G. (2013). Age approximation from speech using Gaussian mixture models. 2nd International Conference on Advanced Computing, Networking and Security, Mangalore, India, pp. 74-78. https://doi.org/10.1109/adcons.2013.43

[72] Murugappan, M., Baharuddin, N.Q.I., Jerritta, S. (2012). DWT and MFCC based human emotional speech classification using LDA. 2012 International Conference on Biomedical Engineering (ICoBE), Penang, 2012, pp. 203-206. https://doi.org/10.1109/icobe.2012.6179005

[73] Chenchah, F., Lachiri, Z. (2014). Impact of gender and emotion type in dialogue emotion recognition. 2014 1st International Conference on Advanced Technologies for Signal and Image Processing (ATSIP), Sousse, Tunisia, pp. 464-467. https://doi.org/10.1109/atsip.2014.6834656

[74] Huang, Y., Meng, S., Li, X., Fan, W. (2018). A classification method for wood vibration signals of Chinese musical instruments based on GMM and SVM. Traitement du Signal, 35(2): 121-136. https://doi.org/10.3166/TS.35.121-136

[75] Zhang, G., Yin, J., Liu, Q., Yang, C. (2011). The fixed-

point optimization of mel frequency cepstrum coefficients for speech recognition. Proceedings of 2011 6th International Forum on Strategic Technology. Heilongjiang, China, pp. 1172-1175. https://doi.org/10.1109/ifost.2011.6021229

[76] Fu, L., Wang, C., Zhang, Y. (2010). A study on influence of gender on speech emotion classification. International Conference on Signal Processing Systems, Dalian, China, pp. V1-534-V1-537. https://doi.org/10.1109/ICSPS.2010.5555556

[77] Vogt, T., Andre, E. (2006). Improving automatic emotion recognition from speech via gender differentiation. Proceedings of Language Resources and Evaluation Conference (LREC 2006), Genoa, Italy, pp. 1123-1126.

[78] Chen, O.T.C., Gu, J.J., Lu, P.T., Ke, J.Y. (2012). Emotion-inspired age and gender recognition systems. 2012 IEEE 55th International Midwest Symposium on Circuits and Systems (MWSCAS), Boise, USA, pp. 662-665. https://doi.org/10.1109/mwscas.2012.6292107