Use of differents Classifiers for Recognition of Fear Emotions in speech

Horkous Houari
Laboratory signal and communication
Ecole National Polytechnique
Algiers, Algeria
houarihorkous29@yahoo.fr

Guerti Mhania
Laboratory signal and communication
Ecole National Polytechnique
Algiers, Algeria
mhaniag@yahoo.fr

***Abstract*—This work consists on the automatic recognition of the emotions in the speech, because it plays a very significant role in the communication. The automatic recognition of the emotions potentially had a broad application in the Human Machine Interaction. In this work emotional speech corpus in Algerian Dialect was created for parameters extraction. The selected parameters in our study are the prosodic (pitch, intensity and duration), the unvoiced frames, jitter, shimmer ¶and cepstral parameters MFCCs (*Mel-Frequency Cepstral Coefficients*) to analyze the emotions of fear and neutral. These parameters will be used in the automatic recognition of the emotions. The system of recognition is based on the methods of classification KNN *(K-Nearest Neighbor)*, SVM *(Support Vector Machine)* and ANN *(Artificial Neural Network)*.¶ The obtained results lead us to observe that the use of MFCCs parameters gives a very acceptable rate of emotion recognition.**

Keywords—Speech emotion, Algerian Dialect, prosodic, MFCC, KNN, SVM, ANN.

I. INTRODUCTION

Human transmit several and various messages by the voice. Among these messages is the emotion, it has a very significant role in the communication of the human. The interface between the human and the machine will be more comprehensible if the machine recognized the state emotional of the human. The automatic recognition of the emotions potentially had a broad application in the human machine interaction for example: robots, emotion recognition in call center, intelligent tutoring system, In-car board system, diagnostic tool by speech therapists, Telephone banking, computer games, etc...

Therefore, our work consists on the automatic recognition of the emotion of the fear and neutral. So the goal is to extract the parameters prosodic, unvoiced frame, jitter, shimmer and MFCC to know the influence of each parameter chosen on the emotions fear and neutral, for exploiting in the emotions recognition system. The extraction of the parameters is obtained by the Praat software which is free software and easy to use and to interpret. The MFCC parameters are extracting by Matlab software. The second section presents notions on the emotion and the parameters used. ¶The third section shows the corpus used, the extraction and the analysis of the selected parameters. ¶In the fourth section, the recognition of the two emotions fear and neutral is made with the methods of K-Nearest Neighbor (KNN) ,Support Vector Machine (SVM) and Artificial Neural Network (ANN) . Finally we finish by a conclusion.

II. NOTIONS ON THE EMOTION AND THE PARAMETERS USED

We present here a short definition of the emotion and the parameters which we chose in order to characterize the target emotions of our work, emotions of the fear type and neutral.

A. Emotion

The fact of expressing an emotion implies a great number of physical and physiological changes (neuronal, activation of certain zones of the brain, increase in the rate of heartbeat, etc). The emotions exist only in reaction to events which they are external or interior. When it is expressed, it can be according to very different means: vocal, gestural, facial, physiological, etc. If only the vocal emotion is considered, that which generates sounds spoken or not, the expression of an emotion passes by physical modifications of the vocal tract and articulation, changes on the level of breathing, saliva or by words or sounds [1].

B. The parameters used

The parameters used in this work are:

Prosodic parameters

The prosodic parameters make it possible to model the accents, the rhythm, the intonation, the melody of the sentence and are thus very relevant for modeling the emotional state of the speaker [2]. The prosodic parameters are the fundamental frequency, intensity and duration.

a) The fundamental frequency or pitch: In speech, the fundamental frequency or pitch characterizes the voiced parts of the speech signal and is related to the feeling height of the voice (Figures 1). The voiced parts have a pseudo-periodic structure and on these portions, the signal is generally modeled as the sum of a periodic signal T and a white noise.The fundamental frequency is the reverse of the period T, $F_0 = \frac{1}{T}$. The Figure 1 represents the contour of fundamental frequency.

Parameters like the minimum, the maximum, the average, the variance, the range and the standard deviation of pitch are used like significant parameters for the discrimination ¶ of the emotions [3], [4], [5], [6], [7].

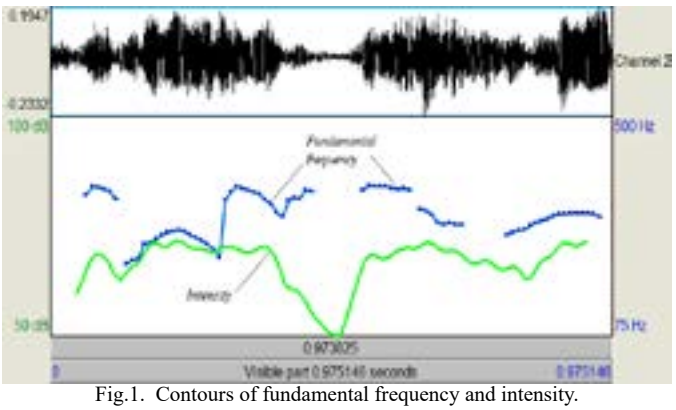


Fig.1. Contours of fundamental frequency and intensity.

b) Intensity: The intensity corresponds to the variation of the speech signal amplitude caused by a more or less strong energy coming from the diaphragm and causing a variation of the air pressure under the glottis. This descriptor makes it possible to

provide a measurement of the sound force of the voice (weak or strong). Pitch, energy, the durations and their derivatives are used for describing the emotional states that are expressed in speech [8]. The Figure 1 represents the contour of intensity.

c) Duration: Duration is the most difficult parameter to specify because nothing indicates how the system of production control or speech perception measures time. The complex relations between pitch, duration and energy parameters are exploited for detecting the speech emotions [9].

1) *Unvoiced frames*

The speech signal is broken up into a number of frames. These frames can be *voiced or unvoiced*. The voiced frames contain the prosodic parameters. ¶And the unvoiced frames contain the parameters of excitation.

2) *Jitter and shimmer*

Jitter and shimmer are measures of the cycle-to-cycle variations of the fundamental frequency and amplitude, which have been largely used for the description of pathological voice quality [10].

3) *Mel-Frequency Cepstral Coefficients*

MFCCs belong to the family of the cepstrals descriptors which base on the representation cepstral of signal. ¶The cepstre has the advantage of allowing a separation of the respective contributions of the source and vocal tract.

MFCCs are obtained while using, for the calculation of the cepstre, a nonlinear frequential scale taking account the characteristics of the human ear, the scale of the Mel frequencies.

The scale of the Mel frequencies is obtained by the following expression [2]:

$$m(f) = 2595 \log\left(1 + \frac{f}{700}\right) \qquad (1)$$

Where is the frequency in Hertz.

The MFCCs parameters are used in the speech emotion recognition [7], [11], [12], [13]. To improve the performance, some recent studies of the speech emotions recognition use the combination enter spectral and prosodic parameters. Information of F_0, the log of energy, the formants, energy in Mel and MFCCs are explored to classify the emotions [14].

III. Extraction and analyses of the selected parameters

In this section we present a definition for the corpus used thus the extraction and the analysis of the selected parameters.

A. Corpus

The corpus used is built from 8 films in Algerian Dialect, This corpus contains more than 400 segments of duration ranging from 0.2 s to 2 s, These segments includes six emotions: fear, anger, sadness, joy, surprise and neutral. Table 1 describes the number of segments for each emotion:

NUMBER OF SEGMENTS FOR EACH EMOTION

Emotion	Number of segments
fear	62
anger	98
sad	63
joy	49
surprise	39
neutral	104
The sum	415

Among the emotions that exist in the corpus we chose two emotions fear and neutral.

B. Extraction the chosen parameters

The results obtained are given in Table 2. This last ¶ illustrates the statistics values of the parameters chosen for the two treated emotions.

THE PARAMETERS EXTRACTED BY *SOFTWARE PRAAT*

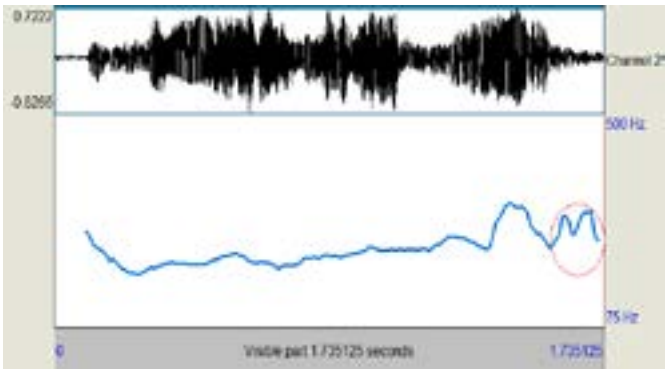
parameter	fear	neutral
Mean of pitch(Hz)	280.10	224.81
Max of pitch(Hz)	394.69	304.00
Min of pitch(Hz)	172.03	122.40
Mean of Intensity(dB)	71.00	70.50
Max of Intensity(dB)	76.43	76.91
Min of Intensity(dB)	57.09	52.70
Duration(s)	0.77	1.32
Unvoiced Frame (%)	63.25	28.78
Jitter (%)	3.20	3.28
Shimmer (%)	16.40	16.83
Range of pitch(Hz)	222.65	179.85
Range of intensity(dB)	19.34	24.02

C. Analyses

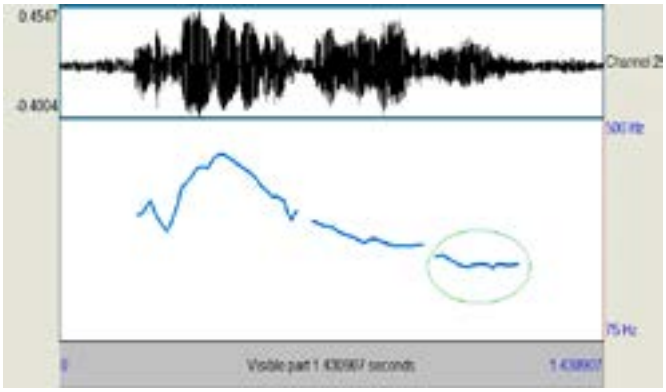
The pitch plays a very significant role, a high pitch correspondent a high frequency sound, a low pitch correspondent a low frequency sound.

We notice according to Table 2 that the statistical values of pitch (mean, max, min, range) are higher compared to the neutral state.

We notice in the Figures 2.a and 2.b during the last word, in the states of fear and neutral the pitch varies in a visible way. But this variation changes from emotion to other.



a. Emotion of fear



b. Emotion of neutral

Fig.2. Fig. 2. Contours of pitch for each emotion

We remark in Table 2 for the intensity that the fear emotion has a high minimum value and has a broad range compared to the neutral state.

It is observed that the duration concern to the emotion of fear is very short.

It is noticed that the emotion of fear has grand number of unvoiced frames. But in the neutral state the number of frames is moderate.

The values of jitter and shimmer are slightly higher in the fear state.

The classifier performance of the emotions is connects with the quality of the data. MFCC is a very powerful technique to analyze the speech signal. Figure 3 illustrates an MFCC form of two emotions fear and neutral, these results it's obtained by Matlab software.

We have remark on figure 3 that the forms of MFCC for the two emotions fear and neutral are vary according to the emotion. This difference which exists between MFCCs allows us used MFCCs parameters in the recognition and classification of speech emotions.

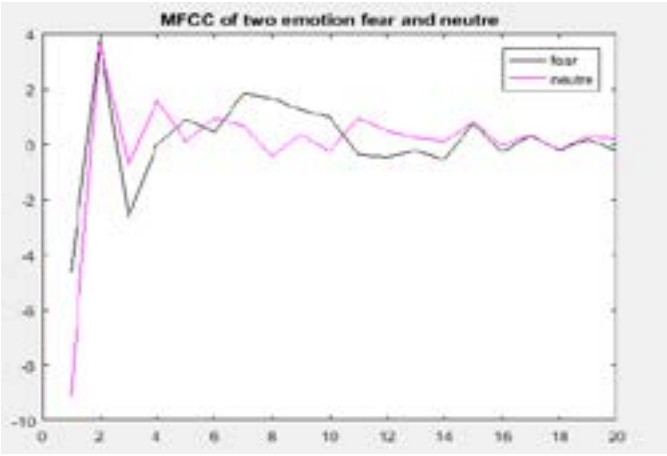


Fig.3. MFCC of two emotions fear and neutral.

IV. Results ans evaluatio

The emotions recognition systems are based on classifiers methods, of which we present a brief review on the methods of classification KNN, SVM and ANN, we describe then the classification system of the emotional states related to the fear and neutral.

A. Classifiers

K-Nearest Neighbor (KNN) is simplest and influential method of classification [15]. Emotion recognition using speech processing using k-nearest neighbor algorithm [16]. KNN used for analysis of emotion recognition system through speech signal [17].

Support Vector Machine (SVM) [18] is a classification technique that is used a lot in the field of speech emotion recognition. SVM was used for automated extraction of features from arabic emotional speech corpus[19].

Artificial Neural Network (ANN) is a classification method often used. Dai et al [20], have presented neural network and combination of feature as to recognize speech's emotions as angry, happy, neutral and sadness. Emotion recognition from Marathi speech database using adaptive artificial neural network [21].

B. The emotions Recognition

The system of recognition is realized by the KNN, SVM, ANN methods in Matlab software. For the method of KNN three values of variable k is used k=3, 5, 7. ¶The data used by this system correspond to a representation of the speech signal by the parameters chosen.

The system is divided into two parts. ¶In the first part we use only the statistic values of parameters chosen without the MFCCs parameters. In the second part we use the statistic values of the parameters chosen with the MFCCs parameters.

The statistic values of the parameters used in our system are the mean, the maximum, the minimum and the range of pitch, the mean, the maximum, the minimum and the range of intensity, the duration, unvoiced frames, jitter and shimmer.

1) without the MFCCs parameters

In this part we use only the statistic values of parameters chosen without the MFCCs parameters. The results are given in tables 3.1, 3.2, 3.3, 3.4, 3.5 and 3.6 .

TABLE .3.1 CONFUSION MATRIX FOR KNN WITH K=3

Emotion	Fear	Neutral
Fear	**85.45%**	14.55%
Neutral	12.73%	**87.27%**

TABLE .3.2 CONFUSION MATRIX FOR KNN WITH K=5

Emotion	Fear	Neutral
Fear	76.36%	13.64%
Neutral	12.73%	87.27%

TABLE .3.3 CONFUSION MATRIX FOR KNN WITH K=7

Emotion	Fear	Neutral
Fear	76.36%	13.64%
Neutral	20%	80%

TABLE .3.4 CONFUSION MATRIX FOR SVM

Emotion	Fear	Neutral
Fear	**81.81%**	18.19%
Neutral	14.55%	**85.45%**

TABLE .3.5 CONFUSION MATRIX FOR ANN

Emotion	Fear	Neutral
Fear	87.27%	12.73%
Neutral	7.18%	92.72%

TABLE .3.6 THE PERFOEMANCE OF CLASSIFICATION METHODS OF BOTH FEAR AND NEUTRAL EMOTION WITHOUT THE PARAMETERS MFCCS

Method	**Percentage**
KNN avec K=3	86.36 %
KNN avec K=5	81.81 %
KNN avec K=7	78.18%
SVM	83.63
ANN	90%
Average	83.99%

From the tables 3.1 to 3.5 we notice that the obtained results are acceptable but vary from method to another. We note also from the table 3.6 that method of ANN is the best by 90%, while Method KNN with k=7 is the lowest by 70%. The average of performance of the methods used is 83.99%.

2) with the MFCCs parameters

In this part we use the statistic values of parameters chosen with the MFCCs parameters. The results are given in tables 4.1, 4.2, 4.3, 4.4, 4.5 and 4.6.

TABLE .4.1 CONFUSION MATRIX FOR KNN WITH K=3

Emotion	Fear	Neutral
Fear	**85.45%**	14.55%
Neutral	10.91%	89.09%

TABLE .4.2 CONFUSION MATRIX FOR KNN WITH K=5

Emotion	Fear	Neutral
Fear	76.36%	23.64%
Neutral	12.73%	87.27%

TABLE .4.3 CONFUSION MATRIX FOR KNN with k=7

Emotion	Fear	Neutral
Fear	76.36%	23.64%
Neutral	18.19%	81.81%

TABLE .4.4 CONFUSION MATRIX FOR SVM

Emotion	Fear	Neutral
Fear	92.72%	7.28%
Neutral	14.55%	85.45%

TABLE .4.5 CONFUSION MATRIX FOR ANN

Emotion	Fear	Neutral
Fear	94.54%	5.46%
Neutral	10.91%	89.09%

TABLE .4.6 THE PERFOEMANCE OF CLASSIFICATION METHODS OF BOTH FEAR AND NEUTRAL EMOTION WITH THE PARAMETERS MFCCS

Method	**Percentage**
KNN avec K=3	87.27 %
KNN avec K=5	81.81 %
KNN avec K=7	79.09%
SVM	89.09
ANN	91.81%
Average	86.14%

From the tables 4.1 to 4.5 when the MFCCs parametres are added, an improvement in the performance of the system is observed. We notice in table 4.6 the performance of the methods used in this part is 86.14% that explains an improvement of performance the recognition system for each method.

V. CONCLUSION

In this work we extracted the prosodic parameters, the unvoiced frames, jitter, shimmer and the MFCCs parameters concerning the two emotions fear and neutral. The corpus used for extraction is built in Algerian dialect. An analysis is made to know the influence of the extracted parameters on the two emotions. These parameters are exploited in a classification system of the two emotions which are indicated. The obtained results show us that the used of the MFCCs parameters give classification rate very important 86.14%. The results obtained indicate that the method of ANN gives the best performance compared to the other method used in this work.

REFERENCES

[1] Marie Tahon. Analyse acoustique de la voix émotionnelle de locuteurs lors d'une interaction humain-robot, these doctorat, paris, 2012.

[2] Chloé CLAVEL. Analyse et reconnaissance des manifestations acoustiques des émotions de type peur en situations anormales, Thèse doctorat, paris, 2007.

[3] Schroder, M. . Emotional speech synthesis: A review. In *Seventh European conference on speech communication and* technology, *Eurospeech* Aalborg, Denmark, Sept. 2001.

[4] Murray, I. R., & Arnott, J. L. Implementation and testing of a system for producing emotion by rule in synthetic speech. *Speech Communication*, 1995, *16*, 369–390.

[5] Wang, Y., Du, S., & Zhan, Y. Adaptive and optimal classification of speech emotion recognition. In *Fourth international conference on natural computation , 2008,* (pp. 407–411).

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

[7] K. Wang, N. An, B.N. Li, Y. Zhang, L. Li, Speech Emotion Recognition Using Fourier Parameters, *IEEE Transactions on Affective Computing, 2015,* 6 , 69-75

[8] Cowie, R., & Cornelius, R. R. Describing the emotional states that are expressed in speech. *Speech Communication*, 2003, *40*, 5–32.

[9] Iida, A., Campbell, N., Higuchi, F., & Yasumura, M. A corpus based speech synthesis system with emotion. *Speech Communication*, 2003, *40*, 161–187.

[10] M. Farrus, J. Hernando, and P. Ejarque. Jitter and shimmer measurements for speaker recognition. *Interspeech, 2007,* pages 778- 781.

[11] Ververidis, D., & Kotropoulos, C. Emotional speech recognition: Resources, features, and methods. *Speech Communication*, 2006, *48*, 1162–1181.

[12] Iliev, A. I., Scordilis, M. S., Papa, J. P., & Falco, A. X. Spoken emotion recognition through optimum-path forest classification using glottal features. *Computer Speech and Language*, 2010, *24*(3), 445–460.

[13] [13] Bitouk, D., Verma, R., & Nenkova, A. Class-level spectral features for emotion recognition. *Speech Communication*, 2010, 52(7–8), 613–625.

[14] Kwon, O., Chan, K., Hao, J., & Lee, T. (2003). Emotion recognition by speech signals. In *Eurospeech*, Geneva (pp. 125–128).

[15] Y. Song, J. Huang, D. Zhou, H. Zha, C. L. Giles. IKNN: Informative K-Nearest Neighbor Pattern Classification, in*: European Conference on Principles of Data Mining and Knowledge Discovery*, 2007, 248–264.

[16] Anuja Bombatkar, Gayatri Bhoyar, Khushbu Morjani, Shalaka Gautam,Vikas Gupta. Emotion recognition using Speech Processing Using k-nearest neighbor algorithm. *International Journal of Engineering Research and Applications (IJERA), 2014, ISSN: 2248-9622.*

[17] Chandra Prakash, Prof. V.B Gaikwad, Dr. Ravish R. Singh Dr. Om Prakash. Analysis of Emotion Recognition System through Speech Signal Using KNN & GMM. Classifier *IOSR Journal of Electronics and Communication Engineering (IOSR-JECE), 2015, 55-61.*

[18] Mohamadally Hasan, Fomani Boris. SVM : Machines à Vecteurs de Support ou Séparateurs a Vastes Marges. Versailles St Quentin, France 2006.

[19] Mohamed Meddeb, Hichem Karray and Adel.M.Alimi. Automated Extraction of Features from Arabic Emotional Speech Corpus. *International Journal of Computer Information Systems and Industrial Management Applications.* 2016, 184-194.

[20] K. Dai, H.J.Fell, and J.MacAuslan, " Recognizing emotion in speech using neural networks,"*Telehealth and Assistive Technologies*, 2008, pp. 31-38.

[21] Raviraj Vishwambhar Darekar, Ashwinikumar Panjabrao Dhande. Emotion recognition from Marathi speech database using adaptive artificial neural network. *Elsevier Biologically Inspired Cognitive Architectures. 2018,* Pages 35-42.