International Journal of Emerging Technology and Advanced Engineering
Website: www.ijetae.com (ISSN 2250-2459, ISO 9001:2008 Certified Journal, Volume 6, Issue 11, November 2016)

Effect of Negative Emotions on the Fundamental Frequency and Formants

Laila Elmazouzi[1], Ilham Mounir[2], Karim Tahiry[3], Abdelmajid Farchi[4], Elmustapha Louragli[5]
[1,2]*LAPSSII Laboratory, Graduate School of Technology, University Cadi Ayyad, Safi, Morocco*
[3,4,5]*IMII Laboratory, Faculty of Sciences & Technics, University Hassan First, Settat, Morocco*

***Abstract*-- Speech signal provides rich information about the emotional state of the speaker. Therefore, emotion recognition of speech has become one of the research themes in the speech processing and applications based on human-computer interaction. This paper provides an experimental study and examines the detection of negative emotions such as fear and anger regarding the neutral emotional state. The dataset is collected from recorded speech in arabic moroccan dialect. Our goal is firstly to study the effects of emotion on the acoustic characteristics chosen namely the fundamental frequency F0 and the first four formants F1, F2, F3 and F4, and secondly to compare our results with previous works. We also investigate the influence of phonemes on the relevance of these characteristics in the detection of emotion. Finally we examine the influence of the speaker gender in this study. For this purpose, we performed classification tests using the WEKA software by three algorithms. We found that F0 has the best recognition rate regardless of phonemes and gender of the speaker. In addition, F2 has a recognition rate that almost reaches the one of F0 in the case of plosive phonemes.**

***Keywords*-- Emotion, Classification, Speech Processing, Fundamental frequency, Formant.**

I. INTRODUCTION

Since the first studies on human behavior, emotions have attracted the interest of researchers in neuroscience and psychology. Recently, this field of research is also emerging in the information processing sciences. Emotions are very present in our lives, they accompany us in any situation. Their detection plays a significant role in the improvement of the human interaction. Several researches were conducted in this direction. The tools on which they are based, are of order physiological, behavioral (e.g [4], [5], [8]) or vocal (e.g [17]). Our objective is to examine the relevance of certain characteristics of the vocal signal in the detection of the negative emotions such as anger and fear with respect to neutral state. We take into consideration the fundamental frequency F0 and the first four formants F1, F2, F3, and F4. We also investigate the effect of phonemes and speaker gender.

In the literature (e.g [6], [11], [12], [13], [19] etc.) similar results were reported regarding accoustic patterns that charaterize different emotional states. The main correlates presented in these papers linked to the present study are :

- *Anger :* The mean of the fundamental frequency F0 is increased (e.g [15], [18], [13]). F1 is raised (e.g [20]).
- *Fear:* Fear is difficult to detect and identify (e.g [2]). It is characterised by the strong increase of the mean of the fundamental frequency (F0) (e.g [2], [18]). However, compared with anger, the mean F0 is lower (e.g [15]). F1 is raised (e.g [20]).

Currently several studies use emotions classification systems. These systems are based on learning methods due to their ability to learn, from a sufficient quantity of acoustic data, properties of each state of emotion (e.g [1], [7], [10], [14], [16],[21]).

In the same context and in order to further verify our hypotheses about the introduced parameters (formants, gender speaker, phonemes) also in practice, we applied them within machine learning algorithms.

II. MÉTHODOLOGY

For studying the changes in acoustic features induced by different emotionnal situations, we used data from 108 speakers (54 males and 54 females). All the subjects are moroccan native speakers in the age group of 21- 23. They were asked to perform the words from moroccan dialect « Safi » (which means ”enough”) and « Maymkench » (which means ”impossible”). Each word was pronounced 3 times under different emotional states, namely fear, neutral and anger.

For the recording, a mobile phone was used and kept at a distance about 15cms away from the mouth. The experiments were conducted in laboratory conditions. The features were extracted from the recorded speech by the aid of the software Praat (e.g [3]). The speech data was digitized with a 8 kHz sampling rate.

International Journal of Emerging Technology and Advanced Engineering
Website: www.ijetae.com (ISSN 2250-2459, ISO 9001:2008 Certified Journal, Volume 6, Issue 11, November 2016)

For each utterance, we extracted the fundamental frequency F0 and the first four formants (F1, F2, F3, F4). The relevance of each feature in determining emotions were estimated by three classifiers: SVM (support vector machines), RN (artificial neural networks) and J48 (tree decision) using the WEKA software (e.g [9]).

The prepared datasets have from one to five attributes (according to studied cases) including F0, F1, F2, F3 and F4. The class attribute indicates the emotional state of the person and is labelled as F (fear), N (neutral) and A (anger) .

Tests were performed on the same data used during training with « stratified cross validation » ; more specifically 10- fold crossvalidation. The system was trained on 90 % of the database and tested on the remaining 10%. This was repeated 9 times with different percent split from 10% to 90% with 10% step.

III. Results

The effect of emotionnal state on the fundamental frequency and each formant was first studied. In figures 1,2 and 3, we plotted the results for three subjects (females). The y axes represent the values of the features under consideration.

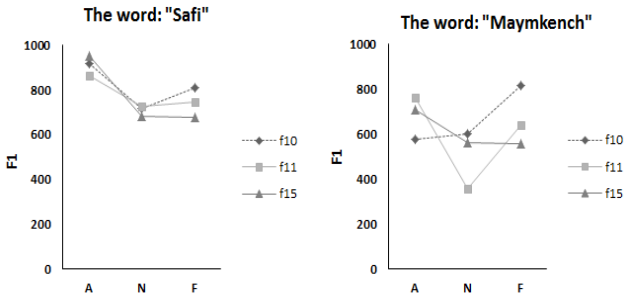

Fig1. Formant F1 for three speakers (female 10, female 11, female 15) for the words « Safi» and « Maymkench ».

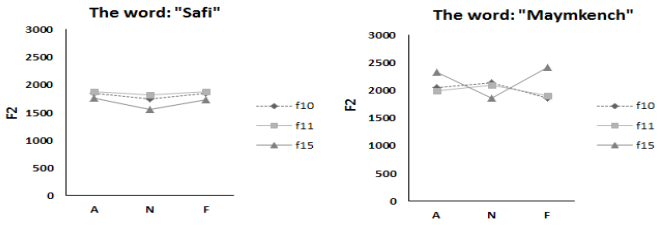

Fig2. Formant F2 for three speakers (female 10, female 11, female 15) for the words « Safi» and « Maymkench ».

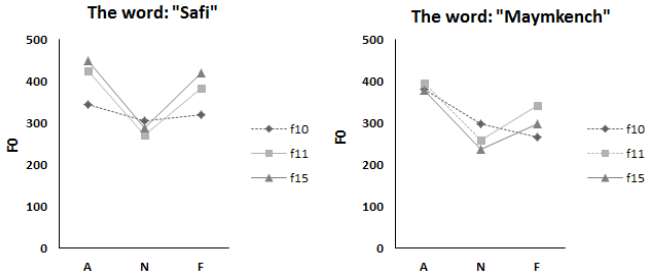

Fig3. F0 mean for three speakers (female 10, female 11, female 15) for the words « Safi» and « Maymkench ».

Comparing the words "Safi" and "Maymkench" Figure.1, 2 and 3, it may be noticed that the emotional state of a person affects the formants F1, F2 and F0 values. Moreover, the frequency range is increased for anger and fear in comparison with that of the neutral. The same findings were reported in [2], [13], [15], [18], [20].

The values obtained for F3 and F4 independently of phonemes and the speaker gender show no general tendency.

To confirm these results we used the classification algorithms provided by WEKA on all the data we have collected for the corpus. We choosed three algorithms namely neural networks (Perceptron multilayer (RN)), support vector machines (SMO) and decision trees (J48). Tests were firstly performed separetely for females and males with 10-fold crossvalidation.

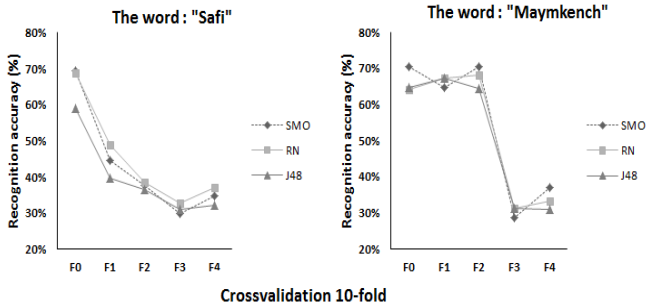

Fig 4. Recognition accuracy for the three classifiers provided for the fundamental frequency F0 and the first formants F1, F2, F3 and F4 for the words : « safi » and « maymkench » (females)

From fig 4, we remark that, for the two words, the fundamental frequency F0 gives the best recognition accuracy. SMO and RN offer the best performance with an average percentage of recognition ranging between 60% and 70, 5%.

International Journal of Emerging Technology and Advanced Engineering
Website: www.ijetae.com (ISSN 2250-2459, ISO 9001:2008 Certified Journal, Volume 6, Issue 11, November 2016)

However, it should be noted that there is a remarkable difference in the F1 and F2 recognition rates for words "safi" and "maymkench". This can be explained by the effect of the phoneme. The formants F3 and F4 seem not to be affected by emotional states.

In order to check the effectiveness of the model, the data was run nine times with the training data percentage varying from 10% to 90% with 10% step. The accuracy should not decrease with changing the training set.

3.1. Recognition accuracy analysis :

By changing the size of the training set, we noted that:

- The recognition accuracy of the first formant F1 is higher for the word "Maymkench" (from 56% to 70%) comparing to the word “Safi” (from 37% to 50%) (Fig.5).
- For the word “Safi” the behavior of the three classifiers is similar for F1 and F2 : the performance of RN and J48 increases while the train size increases. The recognition accuracy provided by SMO varies slightly and decreases around 70% ((Fig.5) and (Fig.6)).
- For the word "Maymkench" The classifiers behave differently according to the first and second formant. For F1, we observe a slight decrease in recognition accuracy of SMO and RN around 70% while the rate of J48 maintains its increase with respect to the train size increase. For F2, the three classifiers present a small decrease around 70% ((Fig.5) and (Fig.6).
- The rate of recognition of the third formant F3 for the word "Safi" (25% to 38%) and the word "Maymkench" (20% to 31%) is very low. There is no noticeable changes in the behavior of the three algorithms (Fig.7).
- For F4, the rate is very low for both words "Safi" (24% to 44%) and "Maymkench" (11% to 38%) with a great variability through the train size changes (Fig.8).
- For F0, the recognition rate is high for the word "Safi" (50% to 71%) and the word "Maymkench" (50% to 81%) with identical behavior for the three classifiers which maintain their growth as the size of the training set increases (Fig.9). The algorithm RN provides the best performance.

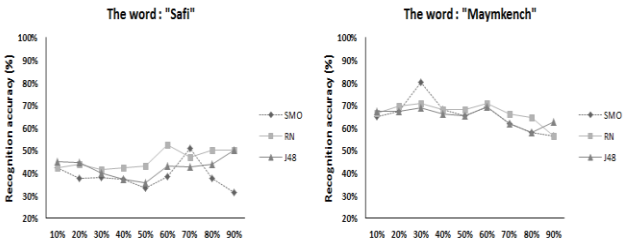

Fig 5. Recognition accuracy for the first formant F1 of three classifiers for different train size (females)

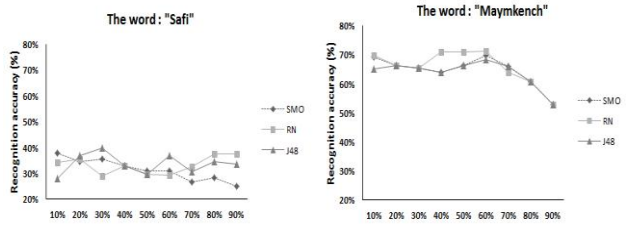

Fig 6. Recognition accuracy for the first formant F2 of three classifiers for different train size (females)

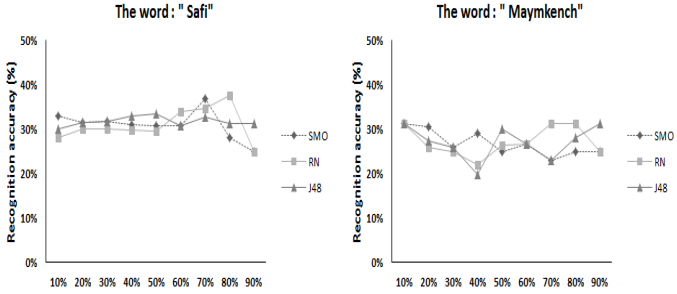

Fig 7. Recognition accuracy for the formant F3 of three classifiers for different train size (females).

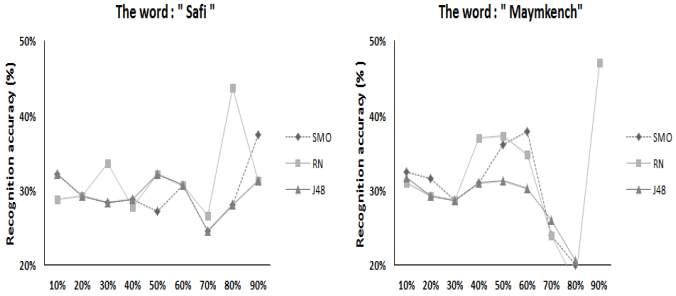

Fig 8. Recognition accuracy for the formant F4 of three classifiers for different train size (females).

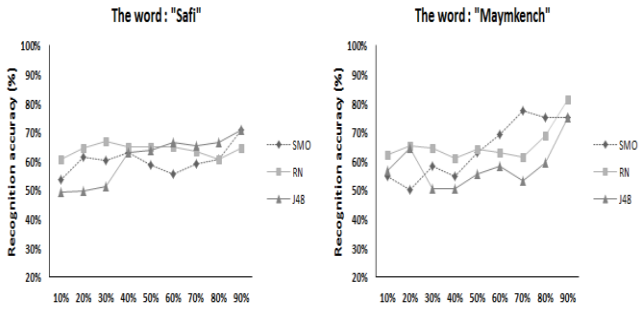

Fig 9. Recognition accuracy for the fundamental frequency F0 of three classifiers for different train size (females).

International Journal of Emerging Technology and Advanced Engineering
Website: www.ijetae.com (ISSN 2250-2459, ISO 9001:2008 Certified Journal, Volume 6, Issue 11, November 2016)

3.2. Area under the ROC cuve :

Looking at the values provided by the analysis of the area under the ROC curve (Fig.10), we noted that :

- The highest AUC values are observed for the word "Maymkench" during the neutral emotional state followed by the anger.
- The most relevant feature that caracterises emotion is F0 followed by F1and F2.
- F1 and F2 achieve better performance for the word "Maymkench" comparing to the word "Safi".
- The algorithms RN and SMO are more competitive in comparison with the algorithm J48.
- F3 and F4 seem not to be affected by emotional state.

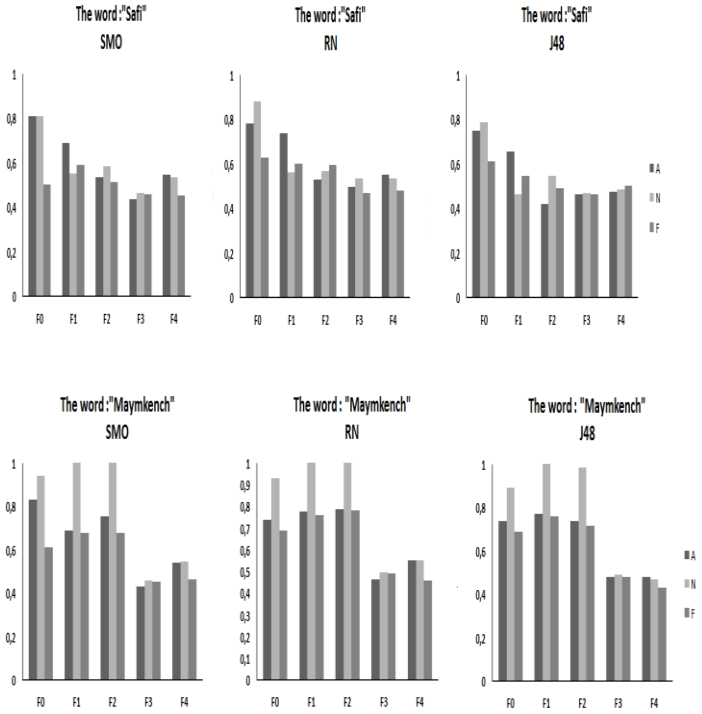

Fig 10. Area under ROC curve of the three classifiers according to fundamental frequency F0 and the four formants F1, F2, F3, and F4. (females).

IV. CONCLUSION

In this article, we studied the effect of emotion on the fundamental frequency F0 and the first four formants. We found that F0 presents the best recognition rates followed by F1 and F2. Our results are consistent with previous works (e.g [2], [13], [15], [18], [20]). We also highlighted the effect of phoneme on the relevance of these characteristics in detecting emotions.

Indeed, we have found that recognition rate of F2 is similar to F0 one for the word "Maymkench".

Moreover, the data we have collected for men, give a lot of variability in the rate of recognition for all indices except for F0, where a similar behavior to females, was observed. The difficulty of expressing the feeling of fear by men has affected the quality of data.

Analysis of the area under the ROC curve confirmed the results. Also the best detected emotional state is the neutral state for both words and for the three algorithms. Comparing the performance of three algorithms shows that RN and SMO are more competitive compared to J48.

REFERENCES

[1] Albornoz, E.M. Milone, D.H. and Rubner, H.L. 2011. Spoken emotion recognition using hierarchical classiffiers. Computer Speech and Language , 25(3), 556-570.

[2] Banse,R. and Scherer, K..R.1996. «Acoustic Profiles in Vocal Emotion Expression». Journal of Personality and Social Psychology. . vol. 70, no 3, p. 614-636.

[3] Boersma, P. 2001. Praat, a system for doing phonetics by computer. Glot Int ., 5(9/10), 341-345.

[4] Busso, C. and Shrikanth, S. N. 2007. «Interrelation between Speech and Facial Gestures in Emotional Utterances: A Single Subject Study». IEEE Transactions on Audio, Speech, and Language Processing. vol. 10, no 20, p. 1-16.

[5] Busso. C, and Shrikanth, S. N. 2007. «Joint Analysis of the Emotional Fingerprint in the Face and Speech: A Single Subject Study». In International Workshop on Multimedia Signal Processing (MMSP) (Chanée, Grèce, Octobre). IEEE, p. 43-47.

[6] Cowie, Roddy, Cowie,E.D. Tsapatsoulis, N. Votsis, G. Kollias, S.Fellenz, W. et Taylor, G.j.2001. «Emotion Recognition in Human-Computer Interaction». IEEE Signal Processing Magazine. vol. 18, no 1, p. 32-80.

[7] Davletcharova, A. Sugathan, S. Abraham, B. Alex Pappachen James.2015 . Detection and Analysis of Emotion From Speech Signals Procedia Computer Science58, 91 – 96.

[8] Ekman, P. 1984. «Expression and the Nature of Emotion». In Approaches to Emotion, Klaus R. Scherer et Paul Ekman, p. 319-343. Hillsdale, États-Unis: Lawrence Erlbaum Associates.

[9] Hall, M. Frank, E. Holmes, G. Pfahringer, B. Reutemann, P. and Witten, I.H. 2009. The weka datamining software: An update. SIGKDD Explorations ,11(1).

[10] Hartmann, K.. Siegert, I. Philippou-Hubner, A.Wendemuth , A.2013. Emotion Detection in HCI: From Speech Features to Emotion Space 12th IFAC Symposium on Analysis, Design, and Evaluation of Human-Machine Systems August 11-15. Las Vegas, NV, USA.

[11] Johnstone, T. and Scherer, K.R. 2000. «Vocal Communication of Emotion». In Handbook of Emotions, Michael Lewis et Jeannette M. Haviland-Jones, p. 220-235. New-York, États-Unis: Guilford.

[12] Juslin, P. N. And Laukka,P. 2003. «Communication of Emotions in Vocal Expression and Music Performance: Different Channels, Same Code». Psychological BulIetin. vol. 129, no 5, p. 770-814.

[13] Murray, I.R. and Arnott, J.L. 1993. «Toward the Simulation of Emotion in Synthetic Speech: A Review of the Litterature on Human Vocal Emotion». Journal of the Acoustical Society of America. vol. 93, no 2, p. 1097-1108.

[14] Oudeyer, P. 2003. The production and recognition of emotions in speech: Features and algorithms. Int'l Journal of Human-Computer Studies, 59(1-2):157–183.

[15] Paeschke, A. and Sendlmeier, W. 2000. Prosodic characteristics of emotional speech: Measurements of fundamental frequency movements. In SpeechEmotion, 75-80.

[16] Philippou-Hubner, D. Vlasenko, B. Bock, R., and Wendemuth, A. 2012. The performance of the speaking rate parameter in emotion recognition from speech. Proc. Of IEEE ICME, 248-253.

[17] Scherer, K. R.1986. «Vocal Affect Expression: A Review and a Model for Future Research». Psychological Bulletin. vol. 99, no 2, p. 143-165.

[18] Scherer, K.R.1995. How emotion is expressed in speech and singing. In Proc. of 1995 ICPhS, 90-96. Stockholm.

[19] Stibbard, R. 2001. «Vocal Expression of Emotions in Non-Laboratory Speech: An Investigation of the Reading/Leeds Emotion in Speech Project Annotation Data». Reading, Royaume Uni, Linguistics and Applied Language Studies, University of Reading, 245p.

[20] Vlasenko, B. Prylipko, D. Philippou-Hubner, D. and Wendemuth, A. 2011. Vowels formants analysis allows straightforward detection of high arousal acted and spontaneous emotions. In Proc. of INTERSPEECH 2011, 1577-1580. Florence, Italy.

[21] Vogt, T. and Andre, E. 2005.Comparing feature sets for acted and spontaneous speech in view of automatic emotion recognition. In IEEE, editor, Int'l Conf. Multimedia and Expo 474–477.