Emotion Recognition and Emotion Spotting Improvement Using Formant-Related Features

D. Gharavian [1], M. Sheikhan [2]
1- Assistant Professor of EE Department, Islamic Azad University, South Tehran Branch, Iran,
and Assistant Professor of EE Department, Shahid Abbaspour University, Tehran, Iran.
Email: gharavian@pwut.ac.ir
2- Assistant Professor of EE Department, Islamic Azad University, South Tehran Branch, Iran.
Email: msheikhn@azad.ac.ir

Received: April 2010 Revised: July 2010 Accepted: August 2010

Abstract:
Emotion has an important role in naturalness of man-machine communication and computerized emotion recognition from speech is investigated by many researchers in the recent decades. In this paper, the effect of formant-related features on improving the performance of emotion detection systems is experimented. To do this, various forms and combinations of the first three formants are concatenated to a popular feature vector and Gaussian mixture models are used as classifiers. Experimental results show average recognition rate of 69% in four emotional states and noticeable performance improvement by adding only one formant-related parameter to feature vector. The architecture of hybrid emotion recognition/spotting is also proposed based on the developed models.

KEYWORDS: Emotion recognition, formants, Gaussian mixture model**.**

1. INTRODUCTION

Contributions due to the fast growth of telecom services and multimedia devices in natural communication between machines and humans are necessary [1-4]. Emotion has an important role in naturalness of man-machine communication, e.g. in speech synthesis [5-8] and automatic speech recognition (ASR) [9-12].

Recognizing emotions from speech by a machine is first investigated around the mid-1980s using statistical properties of certain acoustic features [13]. In the next decade, more complicated emotion recognition algorithms were implemented and market requirements motivated further research. For example, ASRs were trained by employing stressed speech instead of neutral in environments like aircraft cockpits [14]. Iterative algorithms estimated the acoustic features more precisely. Advanced classifiers which used timing information were proposed [15, 16]. Nowadays, research is focused on finding reliable informative features and combining powerful classifiers that improve the performance of emotion detection systems in real-life applications [17-21].

The effect of formants on improving the performance of emotion detection systems is investigated in this paper. So, by generating various supplementary features, based on the first three formants (F_1, F_2 and F_3), and concatenating them to a popular feature vector, which includes "Mel-frequency cepstral coefficients (MFCCs)", "log energy" and "their velocity (Δ) and acceleration ($\Delta\Delta$)", a new rich medium-size feature vector is proposed. Recognizing emotional states in speech is performed by using Gaussian mixture model (GMM), as well.

The rest of the paper is organized as follows: we introduce the background and related works in Section 2. The speech corpus and GMM toolkit is introduced in Section 3. The experiment design and empirical results are presented in Section 4, and finally in Section 5, we conclude the paper.

2. BACKGROUND AND RELATED WORKS

Most existing researches on emotion recognition can be summarized in the procedure shown in Figure 1.

The acoustic features of speech are extracted in the first stage. These features are the basic acoustic or linguistic features, such as pitch-related or spectral-related. In addition, some transform functions are often employed to convert the speech features between different data domains [22].

Some of the extracted features used by research groups in the recent decade are listed in Table 1.

The second stage reduces the size of features set by selecting the most relevant subset of features and

removing the irrelevant ones, or by generating few new features that contain most of the valuable speech information [29-34].

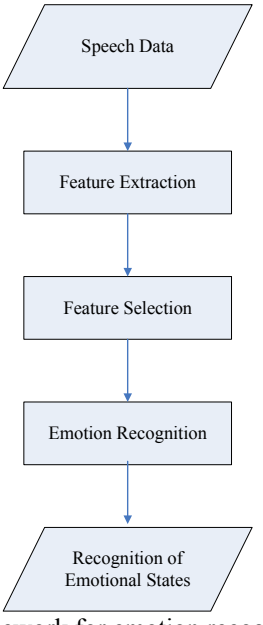

Fig. 1. Framework for emotion recognition from speech.

Table 1. Feature vectors used in emotion recognition from speech

Feature vector	Research group
Pitch, bandwidth, energy, duration, formant [23]	Petrushin (2000)
Pitch, intensity, duration [24]	Amir (2001)
Pitch, energy, duration, formant [25, 26]	Lee et al. (2003), Cai et al. (2003)
Pitch, energy [27]	Schuller et al. (2003)
Pitch, log energy, formant, MFCC [28]	Kwon et al. (2003)
Pitch, energy, formant, MFCC, vocal tract cross-section areas, speech rate [21]	Ververidis et al. (2006)
Pitch, intensity, speech rate [20]	Shami et al. (2007)
Pitch, energy, MFCC, LPC[a] [18]	Altun et al. (2009)

[a] Linear Predictive Coding

The third stage in this system is to train and build a classification model using machine learning algorithms to predict the emotional states. In 1990s, most of the emotion recognition models were based on the maximum likelihood Bayes (MLB) [31, 35] and linear discriminant classification (LDC) [31]. In the recent decade, artificial neural networks (ANNs) [21, 36, 37-39], support vector machines (SVMs) [4, 18, 20, 21, 40-42] and hidden Markov model (HMM) [21, 27, 43-46] have become popular for emotion recognition.

In this paper, GMM with 32 mixtures is used as the classifier. A feature vector which includes 12 Mel-frequency cepstral coefficients, log energy and their velocity (Δ) and acceleration ($\Delta\Delta$) coefficients, is used as the base feature vector. So, the size of this vector is 39. Then, by adding formant-related features to this base vector, new feature vectors are generated. Various GMMs are trained with these feature vectors and their recognition accuracies are evaluated in detecting emotions.

3. Database and Tools

The texts of 99 sentences from FARSDAT, the speech database of Farsi spoken language [47], are selected as base sentences in this research. Four emotional states are considered and 15 speakers uttered the base sentences with various emotional states. The speech of 10 speakers, who uttered 2520 sentences, is used for training the GMM models. Another five speakers uttered 846 sentences as test examples, too (Table 2).

Table 2. Number of training and test sentences for each emotional state

State	Number of sentences	
	Training	Test
Neutral	990	379
Anger	340	123
Happiness	690	154
Interrogative	500	190

Using HTK, MFCC coefficients and energy are calculated [48] and formant frequencies are determined by using linear prediction spectra [49]. Speech frame length is considered 25msec with 15msec overlap.

4. Experiment Design and Empirical Results

4.1. Target Models

A Gaussian mixture model is trained with the base feature vectors. We call this model, M_0. By adding formant-related parameters to the end of base feature vector, we generate 11 different models (Table 3).

To assess the effect of formants on improving the recognition accuracy, M_1-M_3 models are generated. M_4-M_6 models are generated by using the logarithm of each formant, as an additional parameter. The normalized values of formants, by subtracting the average value of formant frequency in a sentence, form the additional parameters in M_7-M_9 models. Further, the formant slope is used in M_{10}-M_{12} models, as well. This slope is calculated by using Equation (1).

To assess the effect of formants on improving the recognition accuracy, M_1-M_3 models are generated. M_4-M_6 models are generated by using the logarithm of

each formant, as an additional parameter, too. The normalized values of formants, by subtracting the average value of formant frequency in a sentence, form the additional parameters in M_7-M_9 models. The formant slope is used in M_{10}-M_{12} models, as well.

Table 3. Formant-related features in different models

Model	Features
M_0	12 MFCC+LE+Δ(12 MFCC+LE) +ΔΔ(12 MFCC+LE)
M_1	M_0+F_1
M_2	M_0+F_2
M_3	M_0+F_3
M_4	M_0+log(F_1)
M_5	M_0+log(F_2)
M_6	M_0+log(F_3)
M_7	M_0+Norm(F_1)
M_8	M_0+Norm(F_2)
M_9	M_0+Norm(F_3)
M_{10}	M_0+dF_1
M_{11}	M_0+dF_2
M_{12}	M_0+dF_3

$$d_t = \frac{\sum_{n=1}^{2} n(f_{t+n} - f_{t-n})}{2\sum_{n=1}^{2} n^2} \quad (1)$$

4.2. Evaluation of Model Efficiency

To evaluate the efficiency of M_0-M_{12} models, a GMM is considered for each emotional state. So, 52 GMMs are trained and tested in our experiments for the mentioned models.

In emotional states of anger and happiness, each part of a sentence can be affected. But in interrogative sentences, the emotion affects only the end part of a sentence. The procedure of performance evaluation for each model in an emotional state is depicted in Figure 2.

4.3. Experimental Results

The recognition rate of M_0 model for each emotional state is reported in Table 4.

Table 4. Emotion recognition accuracy using M_0 model.

Emotional state	Accuracy (%)
Neutral	63.3
Happiness	49.2
Anger	92.7
Interrogative	49.0

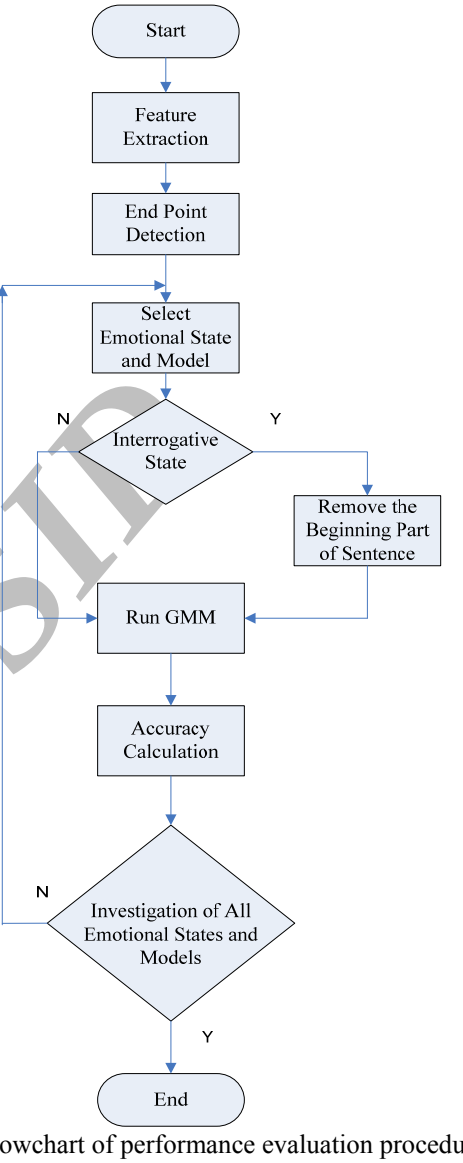

Fig. 2. Flowchart of performance evaluation procedure of each model in an emotional state.

The accuracy of this model in recognizing happiness and interrogative states is low. So, in the rest of this subsection, the effect of adding formant-related features in improving the emotion recognition accuracy is investigated.

When we use F_3 (in M_3 model), F_1 (in M_1 model), and F_2 (in M_2 model), as an additional information in the base feature vector, the performance is improved in neutral, happiness, and interrogative states, respectively (Figure 3).

The effects of logarithm, normalization, and differentiation operators on formants, when they are used as additional features in emotion recognition, are shown in Figures 4, 5, and 6, respectively.

As shown in Figure 4, the performance of M_4-M_6 models is better than M_0 model in anger-state

recognition. M_5 and M_6 models improve the recognition rate as compared to M_2 and M_3 models, too. The logarithm as a smoothing operator degrades the performance in happiness-state (Figure 4b).

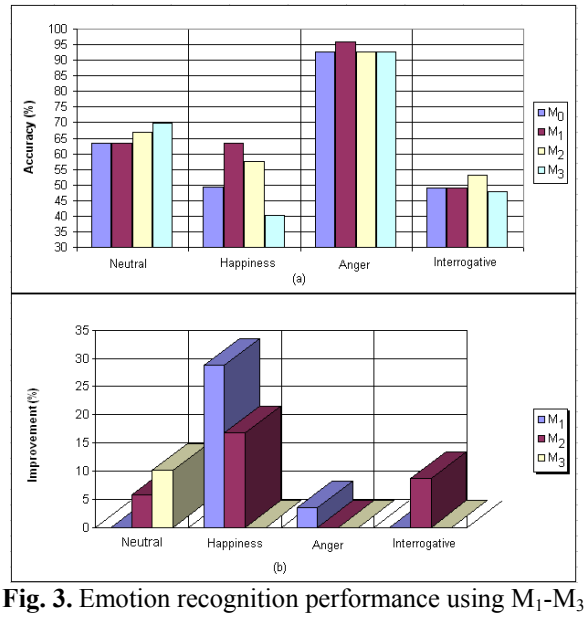

Fig. 3. Emotion recognition performance using M_1-M_3 models; a) Accuracy compared to M_0, b) Improvement.

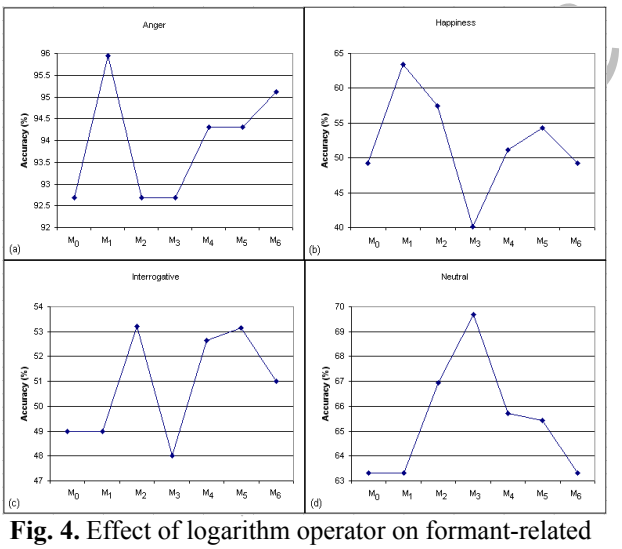

Fig. 4. Effect of logarithm operator on formant-related features in emotion-state recognition; a)Anger, b)Happiness, c)Interrogative, d)Neutral.

This operator causes the better performance of M_4 and M_6 models in interrogative-state as compared to M_1 and M_3 models (Figure 4c).

The normalization of formants degrades the performance in anger and neutral states, as shown in Figure 5. However, this operator improves the recognition rate in happiness-state (Figure 5b). The normalization of F_1 and F_3 improves the performance in interrogative-state, too (Figure 5c).

The slope of F_2 and F_3 in happiness-state and slope of F_1 and F_3 in interrogative-state are more effective than actual value of formants in emotion recognition, as well (Figures 6b and 6c).

Based on the review of results, a set of combinational features are proposed using the value of formants and their slopes (Table 5).

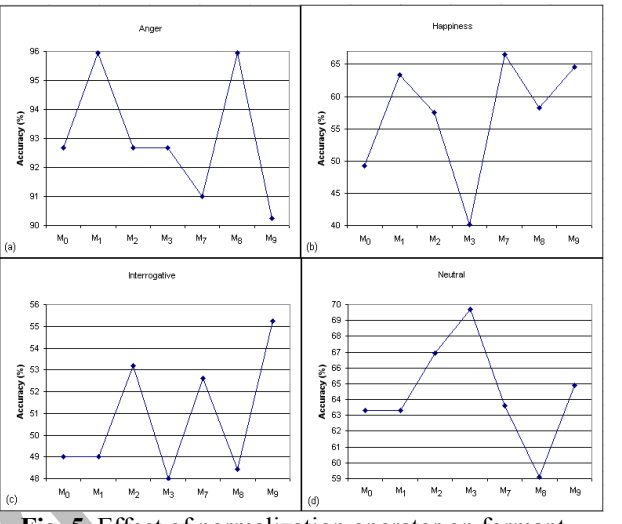

Fig. 5. Effect of normalization operator on formant-related features in emotion-state recognition; a)Anger, b)Happiness, c)Interrogative, d)Neutral.

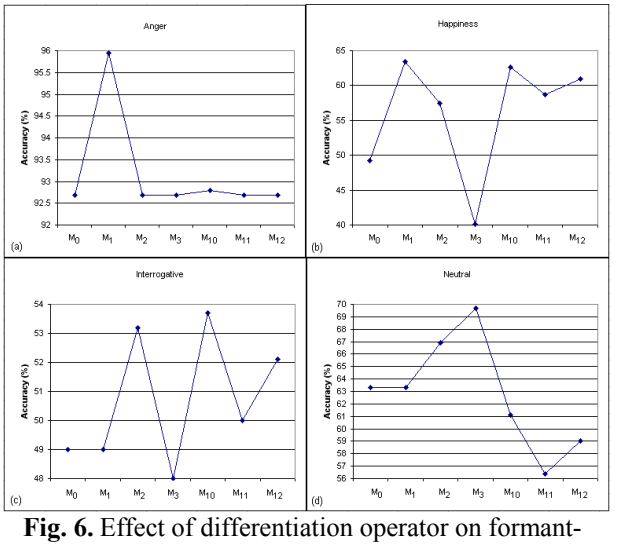

Fig. 6. Effect of differentiation operator on formant-related features in emotion-state recognition; a)Anger, b)Happiness, c)Interrogative, d)Neutral.

The performance of new-defined models (M_{13}-M_{19}) is shown in Figure 7, as compared to the base model (M_0). As shown in Figure 7, all the models of Table 5, improve the recognition rate in happiness-state. In interrogative-state, M_{14} and M_{17} models and in neutral-state, M_{13} and M_{17} models improve the performance, as well (Figures 7c and 7d).

To select the best model, the relative accuracy improvements of M_1-M_{19} models, as compared to M_0 Model, are depicted in Figure 8.

The flowchart of proposed approach is shown in Figure 9.

Table 5. Combination of formant-related features in models

Model	Features
M_{13}	$M_0+F_1+F_2+F_3$
M_{14}	$M_0+dF_1+dF_2+dF_3$
M_{15}	$M_0+F_1+F_2+dF_1$
M_{16}	$M_0+F_1+F_2$
M_{17}	$M_0+F_2+F_3$
M_{18}	$M_0+F_1+dF_1$
M_{19}	$M_0+F_2+dF_1$

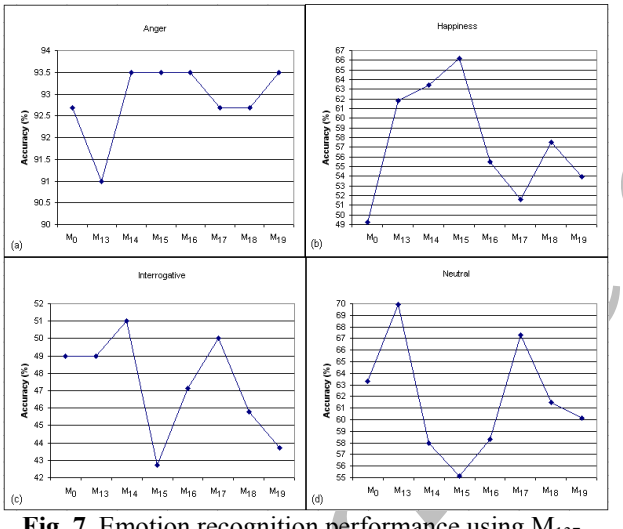

Fig. 7. Emotion recognition performance using M_{13}-M_{19} models compared to M_0 model; a)Anger, b)Happiness, c)Interrogative, d)Neutral.

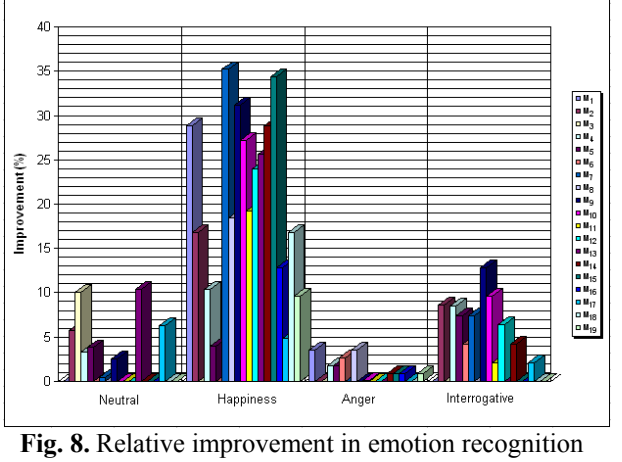

Fig. 8. Relative improvement in emotion recognition rates using M_1-M_{19} models, compared to M_0 model.

As shown in Figure 8, M_{13} for neutral, M_7 for happiness, M_1 for anger, and M_9 for interrogative state has the best performance. However, we cannot find a model with the best recognition rate for all of the emotional states. Among 19 mentioned models, four models with the better recognition rates are presented in Table 6.

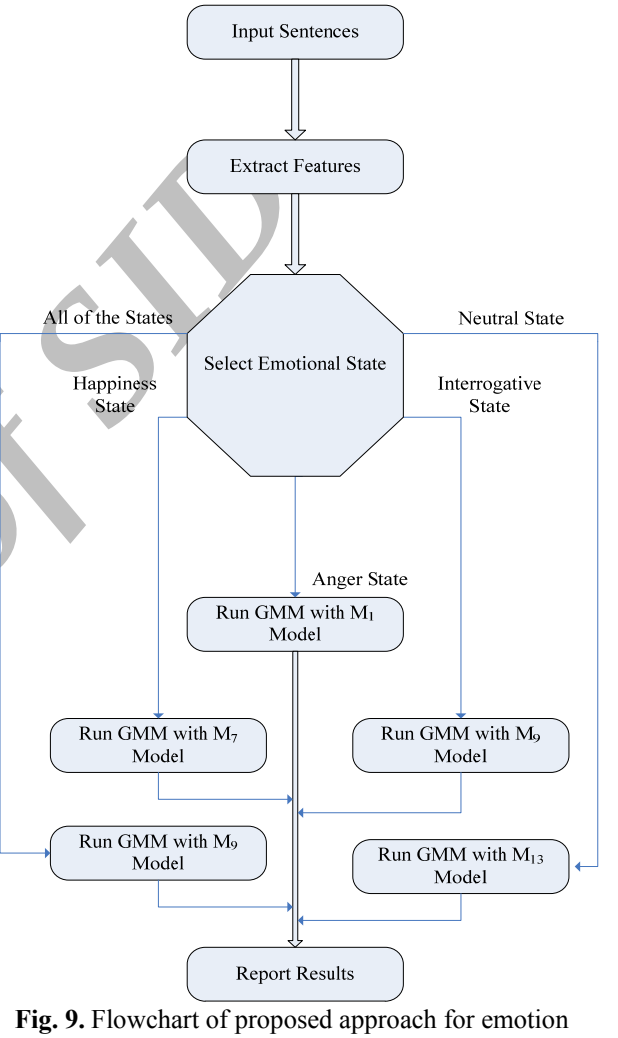

Fig. 9. Flowchart of proposed approach for emotion recognition.

Table 6. Models with better emotion recognition rates

Models	Emotional states			
	Neutral	Happiness	Anger	Interrogative
M_9	64.9	64.6	90.2	55.3
M_7	63.6	66.5	91.0	52.6
M_2	66.9	57.5	92.7	53.2
M_{10}	61.1	62.6	92.8	53.7

So, M_9 model has the best performance among others. But if we want to find a specified emotion state (emotion spotting), we can use the model with the best recognition rate for that state (e.g. M_7 for happiness-state).

5. CONCLUSION

In this paper, the effect of formant-related features on improving the performance of emotion detection systems has been investigated.

TABLE 7. PERFORMANCE OF TYPICAL SYSTEMS FOR EMOTION RECOGNITION IN THE RECENT DECADE

Emotional states	Selected features	Number of features	Classifier(s)	Recognition rate (%)
Happiness, anger, sadness, neutral [50]	Pitch and its slope, formants, MFCCs	20	SVM, ANN	71, 42
Happiness, anger, tiredness, sadness, neutral [28]	Pitch, log energy, formants, MFCCs and their Δ and ΔΔ	59	GSVM[a]	41
Happiness, anger, anxiety, fear, tiredness, disgust, neutral [51]	MFCCs, energy, and their Δ and ΔΔ	26	GMVAR[b], ANN, HMM	76, 55, 71
Happiness, anger, tiredness, sadness, disgust, fear, neutral [52]	MFCCs, log energy, and their Δ and ΔΔ	39	HMM	81
Happiness, anger, sadness, neutral [18]	Pitch, sub-band energies, MFCCs, LPC	58	Multi-class SVM	80
Happiness, anger, sadness, fear, neutral [22]	Pitch, intensity, zero crossing rate, spectral features	84	k-NN[c]	66
Anger, fear, surprise, disgust, joy, sadness [53]	V/UV, energy, pitch, VAD	86	GMM (512 mixtures)	92
Neutral, emphatic, negative [54]	Pitch, MFCCs	80	GMM (512 mixtures)	93
Happiness, anger, interrogative, neutral (proposed model)	MFCCs, log energy and their Δ and ΔΔ, formant-related features	40	GMM	69
Happiness, anger, interrogative, neutral (similar Farsi emotional speech database) [55]	MFCCs, log energy and their Δ and ΔΔ, pitch and formants-related features	52	fuzzy-ARTMAP	67

[a] Gaussian SVM
[b] Gaussian Mixture Vector Autoregressive Model
[c] k-Nearest Neighbor

To do this, various forms of the first three formants have been concatenated to a popular feature vector and the recognition rates using Gaussian mixture model have been measured. The performance of proposed system has been compared with some other emotion recognition systems (Table 7).

The accuracy of the proposed system is reported in last row of Table 7. Because of the different target emotional states and also feature sets in each research, selection of the most effective model among them is impossible. However, the proposed medium-size feature vector in this research and the performance improvement by adding only one formant-related parameter, show the effectiveness of approach in developing customized emotion recognition and emotion spotting systems.

To compare the performance of proposed model with another classifier on this database, the fuzzy ARTMAP neural network was used [55]. The fuzzy ARTMAP neural network (FAMNN) has been introduced by Carpenter *et al* [56]. The FAMNN has been successfully applied in many tasks such as data mining, remote sensing and pattern recognition. The specifications of simulated neural network are reported in Table 8. To reduce the number of input vectors to FAMNN, the average value of each feature in each sentence is used as the input in [55]. Comparison of GMM and FAMNN on similar emotional speech database shows that GMM offers better classification rates in this work.

TABLE 8. SPECIFICATIONS OF FUZZY ARTMAP NEURAL NETWORK

Specification	Value
Learning rate	1
Vigilance parameter	0.99
Number of F_0 nodes	104
Number of F_1 nodes	104
Number of F_2 nodes	3600
Training time (sec)	~13000
Number of classes	4
Number of training samples	4869
Number of test samples	504

REFERENCES

[1] C. Clavel, I. Vasilescu, L. Devillers, G. Richard and T. Ehrette, **"Fear-type emotion recognition for future audio-based surveillance systems"**, *Speech Communication*, 50, pp. 487-503, (2008)

[2] Z. Inanoglu and S. Young, **"Data-driven emotion conversion in spoken English"**, *Speech Communication* , 51, pp. 268-283, (2009)

[3] E. Leon, G. Clarke, V. Callaghan and F. Sepulveda, **"A user-independent real-time emotion recognition system for software agents in domestic environments"**, *Engineering Applications of Artificial Intelligence*, 20, pp. 337-345, (2007)

[4] D. Morrison, R. Wang and L.C. de Silva, **"Ensemble methods for spoken emotion recognition in call-centers"**, *Speech Communication*, 49, pp. 98-112, (2007)

[5] M. Sheikhan, M. Nasirzadeh and A. Daftarian, **"Design and implementation of Farsi text to speech system"**, *Journal of Engineering Faculty*, Ferdowsi University of Meshed, 17, pp. 31-48, (2005)

[6] M. Sheikhan, **"Automatic prosody generation by neural-statistical hybrid model for unit selection speech synthesis"**, *Journal of Biomedical Engineering*, 1(new), pp. 227-240, (2007)

[7] M. Sheikhan, **"Prosody generation in Farsi language"**, *In the Proceedings of International Symposium on Telecommunications*, pp. 250-253, (2003)

[8] M. Sheikhan, M. Nasirzadeh and A. Daftarian, **"Text to speech for Iranian dialect of Farsi language"**, *In the Proceedings of Second Workshop on Farsi Computer Speech*, University of Tehran, pp. 39-53, (2006)

[9] M. Sheikhan, M. Tebyani and M. Lotfizad, **"Continuous speech recognition and syntactic processing in Iranian Farsi language"**, *International Journal of Speech Technology*, 1, pp. 135-141, (1997)

[10] D. Gharavian and S.M. Ahadi, **"The effect of emotion on Farsi speech parameters: A statistical evaluation"**, *In the Proceedings of the International Conference on Speech and Computer*, pp. 463-466, (2005)

[11] D. Gharavian and S.M. Ahadi, **"Recognition of emotional speech and speech emotion in Farsi"**, *In the Proceedings of International Symposium on Chinese Spoken Language Processing*, Vol. 2, pp. 299-308, (2006)

[12] D. Gharavian, **"Prosody in Farsi language and its use in recognition of intonation and speech"**, *PhD Dissertation*, Electrical Engineering Department, Amirkabir University of Technology, Tehran (In Farsi), (2004)

[13] F.J. Tolkmitt and K.R. Scherer, **"Effect of experimentally induced stress on vocal parameters"**, *Journal of Experimental Psychology: Human Perception and Performance*, 12, pp. 302-313, (1986)

[14] J.H.L. Hansen and D.A. Carins, **"ICARUS: Source generator based real-time recognition of speech in noisy stressful and Lombard effect environments"**, *Speech Communication*, 16, pp. 391-422, (1995)

[15] D. Cairns and J.H.L. Hansen, **"Nonlinear analysis and detection of speech under stressed conditions"**, *Journal of Acoustic Society of America* , 96, pp. 3392-3400, (1994)

[16] B.D. Womack and J.H.L. Hansen, **"Classification of speech under stress using target driven features"**, *Speech Communication*, 20, pp. 131-150, (1996)

[17] H.Altun and G. Polat, **"New frameworks to boost feature selection algorithms in emotion detection for improved human-computer interaction"**, *Brain Vision and Artificial Intelligent. Lecture Notes in Computer Science* , 4729, pp. 533-541, (2007)

[18] H. Altun and G. Polat, **"Boosting selection of speech related features to improve performance of multi-class SVMs in emotion detection"**, *Expert Systems with Applications*, 36, pp. 8197-8203, (2009)

[19] C.M. Lee and S.S. Narayanan, **"Toward detecting emotions in spoken dialogs"**, *IEEE Transactions on Speech and Audio Processing*, 13, pp. 293-303, (2005)

[20] M. Shami and W. Verhelst, **"An evaluation of the robustness of existing supervised machine learning approaches to the classifications of emotions in speech"**, *Speech Communication*, 49, pp. 201-212, (2007)

[21] D. Ververidis and C. Kotropoulos, **"Emotional speech recognition: Resources, features, and methods"**, *Speech Communication*, 48, pp. 1162-1181, (2006)

[22] J. Rong, G. Li and Y.P. Chen, **"Acoustic feature selection for automatic emotion recognition from speech"**, *Information Processing and Management*, 45, pp. 315-328, (2009)

[23] V.A. Petrushin, **"Emotion recognition in speech signal: Experimental study, development, and application"**, *In the Proceedings of the International Conference on Spoken Language Processing*, pp. 222-225, (2000)

[24] N. Amir, **"Classifying emotions in speech: A comparison of methods"**, *In the Proceedings of the European Conference on Speech Communication and Technology*, pp. 127-130, (2001)

[25] L. Cai, C. Jiang, Z. Wang, L. Zhao and C. Zou, **"A method combining the global and time series structure features for emotion recognition in speech"**, *In the Proceedings of the International Conference on Neural Networks and Signal Processing*, Vol. 2, pp. 904-907, (2003)

[26] C.M. Lee and S. Narayanan, **"Emotion recognition using a data-driven fuzzy inference system"**, *In the Proceedings of the European Conference on Speech Communication and Technology*, pp. 157-160, (2003)

[27] B. Schuller, G. Rigoll and M. Lang, **"Hidden Markov model-based speech emotion recognition"**, *In the Proceedings of the International Conference on Acoustics, Speech, and Signal Processing*, Vol. 2, pp. 1-4, (2003)

[28] O.W. Kwon, K. Chan, J. Hao and T.W. Lee, **"Emotion recognition by speech signal"**, *In the Proceedings of the European Conference on Speech Communication and Technology*, pp. 125-128, (2003)

[29] R. Kohavi and G.H. John, **"Wrappers for feature subset selection"**, *Artificial Intelligence*, 97, pp. 273-324, (1997)

[30] J.B. Tenenbaum, V. de Silva and J.C. Langford, **"A global geometric framework for nonlinear dimensionality reduction"**, *Science*, pp. 2319-2323, (2000)

[31] F. Dellaert, T. Polzin and A. Waibel, **"Recognizing emotion in speech"**, *In the Proceedings of the International Conference on Spoken Language Processing*, Vol. 3, pp. 1970-1973, (1996)

[32] A. Hyvarinen, **"Survey of independent component analysis"**, *Neural Computing Surveys*, 2, pp. 94-128, (1999)

[33] H. Liu, H. Motoda and L. Yu, **"Feature selection with selective sampling"**, *In the Proceedings of the International Conference on Machine Learning*, pp. 395-402, (2002)

[34] L. Talavera, **"Feature selection as a preprocessing step for hierarchical clustering"**, *In the Proceedings of the International Conference on Machine Learning*, pp. 389-397, (1999)

[35] J. Han and M. Kamber, **Data Mining Concepts and Techniques**, *Morgan Kaufman Pub. Comp.*, (2000)

[36] C.M. Lee, S. Narayanan and R. Pieraccini, **"Combining acoustic and language information for emotion recognition"**, *In the Proceedings of the International Conference on Spoken Language Processing*, pp. 873-876, (2002)

[37] J. Nicholson, K. Takahashi and R. Nakatsu, **"Emotion recognition in speech using neural networks"**, *In the Proceedings of the International Conference on Neural Information Processing*, Vol. 2, pp. 495-501, (1999)

[38] C.H. Park, D.W. Lee and K.B. Sim, **"Emotion recognition of speech based on RNN"**, *In the Proceedings of the International Conference on Machine Learning and Cybernetics*, Vol. 4, pp. 2210-2213, (2002)

[39] C.H. Park and K.B. Sim, **"Emotion recognition and acoustic analysis from speech signal"**, *In the Proceedings of the International Joint Conference on Neural Networks*, Vol. 4, pp. 2594-2598, (2003)

[40] Z.J. Chuang and C.H. Wu, **"Emotion recognition using acoustic features and textual content"**, *In the Proceedings of the International Conference on Multimedia and Expo*, Vol. 1, pp. 53-56, (2004)

[41] S. Hoch, F. Althoff, G. McGlaun and G. Rigooll, **"Bimodal fusion of emotional data in an automotive environment"**, *In the Proceedings of the International Conference on Acoustics, Speech, and Signal Processing*, Vol. 2, pp. 1085-1088, (2005)

[42] B. Schuller, G. Rigoll and M. Lang, **"Speech emotion recognition combining acoustic features and linguistic information in a hybrid support vector machine-belief network architecture"**, *In the Proceedings of the International Conference on Acoustics, Speech, and Signal Processing*, Vol. 1, pp. 577-580, (2004)

[43] L. Bosch, **"Emotions, speech and the ASR framework"**, *Speech Communication*, 40, pp. 213-225, (2003)

[44] T.L. Nwe, S.W. Foo and L.C. de Silva, **"Speech emotion recognition using hidden Markov models"**, *Speech Communication*, 41, pp. 603-623, (2003)

[45] M. Song, J. Bu, C. Chen and N. Li, **"Audio-visual based emotion recognition- A new approach"**, *In the Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, Vol. 2, pp. 1020-1025, (2004)

[46] M. Song, C. Chen and M. You, **"Audio-visual based emotion recognition using tripled hidden Markov model"**, *In the Proceedings of the International Conference on Acoustics, Speech, and Signal Processing*, Vol. 5, pp. 877-880, (2004)

[47] M. Bijankhan, J. Sheikhzadegan, M.R. Roohani, Y. Samareh, C. Lucas and M. Tebyani, **"FARSDAT- The speech database of Farsi spoken language"**, *In the Proceedings of the Australian Conference on Speech Science and Technology*, Vol. 2, pp. 826-830, (1994)

[48] S. Young, **The HTK Book**, *Cambridge University Press*, (2001)

[49] SS. McCandless, **"An algorithm for formant extraction using linear prediction spectra"**, *IEEE Transactions on Acoustics*, *Speech and Signal Processing*, 22, pp. 135-141, (1974)

[50] F. Yu, E. Chang, Y. Xu and H. Shum, **"Emotion detection from speech to enrich multimedia content"**, *In Proceedings of the IEEE Pacific Rim Conference on Multimedia: Advances in Multimedia Information Processing*, pp. 550-557, (2001)

[51] M. Ayadi, S. Kamel and F. Karray, **"Speech emotion recognition using Gaussian mixture vector autoregressive models"**, *In the Proceedings of the International Conference on Acoustics, Speech*, *and Signal Processing*, Vol. 5, pp. 957- 960, (2007)

[52] B. Vlasenko and. A. Wendemuth, **"Tuning hidden Markov model for speech emotion recognition"**, *In Proceedings of the 33rd German Annual Conference on Acoustics*, pp. 317-320, (2007)

[53] I. Luengo, E. Navas, I. Hernaez and J. Sanchez, **"Automatic emotion recognition using prosodic parameters",** *in the proc. of the European Conf. on Speech Communication and Technology,* pp: 493-496, (2005)

[54] D. Neiberg, K. Elenius and K. Laskowski, **"Emotion recognition in spontaneous speech using GMMs",** *in the proc. of the Int. Conf. on Spoken Language Processing*, pp: 809-812, (2006)

[55] A. Nazeriyeh, **"Emotion speech recognition using prosody features by artificial neural networks in farsi language**", *MSc. Dissertation,* Islamic Azad University-South Tehran Branch (Advisor: D. Gharavian, Consultant: M. Sheikhan), (2010)

[56] G.A. Carpenter, S. Grossberg, N. Markuzon, J.H. Reynolds and D.D. Rosen, **"A neural network architecture for incremental supervised learning of analog multidimensional maps"**, *IEEE Transaction on Neural Networks*, 3:698-713, (1992)