

Contents list available at JMCS

Journal of Mathematics and Computer Science

Journal Homepage: [www.tjmcs.com](http://www.tjmcs.com)



## Speech Emotion Recognition Based on Learning Automata in

### Fuzzy Petri-net

Sara Motamed\*, Saeed Setayeshi, Zeinab Farhoudi, Ali Ahmadi  
*Science and Research University, Tehran, Iran*

\*[samotamed@yahoo.com](mailto:samotamed@yahoo.com)

#### Article history:

Received July 2014

Accepted August 2014

Available online August 2014

#### Abstract

This paper explores how fuzzy features' number and reasoning rules can influence the rate of emotional speech recognition. The speech emotion signal is one of the most effective and neutral methods in individuals' relationships that facilitate communication between man and machine. This paper introduces a novel method based on mind inference and recognition of speech emotion recognition. The foundation of the proposed method is the inference of rules in Fuzzy Petri-net (FPN) and the learning automata. FPN is a new method of classification which is introduced for the first time on emotion speech recognition. This method helps to analyze different rules in a dynamic environment like human's mind. The input of FPN is computed by learning automata. Therefore learning automata has been used to adjust the membership functions for each feature vector in the dynamic environment. The proposed algorithm is divided into different parts: preprocessing; feature extraction; learning automata; fuzzification; inference engine and defuzzification. The proposed model has been compared with different models of classification. Experimental results show that the proposed algorithm outperforms other models.

**Keywords:** Emotional Speech, Fuzzy Rules, Learning Automata, Mel frequency Cepstral coefficients (MFCC), Fuzzy Petri-net.

### 1. Introduction

Emotion is an important indicative of human feelings and conditions. It also shows the interaction between man and machines. In this context speaking plays a major role in showing human emotions which can hardly be switched off. Recently, many researchers have discovered important information regarding the various domains of emotional speak recognition. The robots or agents used different emotions or human-computer interfaces to explain human emotions in their relationships. Cowie and Erickson employed emotional speech recognition using different classification models [3, 4]. Also,

there are several approaches that exist such as those based on K-nearest neighborhood (KNN) [1], the Gaussian mixture model (GMM) [5], the hidden Markov Model (HMM) [2, 6], and the NN-C5.0 method [30]. Pierre-Yves, Lee and Narayanan proposed many techniques based on the classification of different emotional situations in discrete categories including happy, sad, anger, fear, surprise, and disgust [7, 8]. On the other hand, they believed one of the important things is the detection of variability between a certain emotion such as 'little sad' or 'very happy' for expanding emotion categories.

Since a single label of a discrete category cannot accurately reflect a human's interaction, many researchers used a dimensional description of human emotion by estimating it on a continuous-valued scale in a multi-dimensional space [9- 13]. As feature extraction finds the true pitch of voice need to discover human emotions, it is known as the most important and sensitive section in speech emotion recognition. Therefore, this paper uses a new method classification like mind and based on rules to label the different emotions in multi dimensional space. The proposed algorithm is divided into different parts to present a useful method to solve the uncertain problems in the dynamic environment by using reasoning rules in FPN. The Berlin dataset, as a good quality public dataset, is used. By combining the Petri-net model with fuzzy set theory, one can explicitly analyse the expert knowledge in the case of having incomplete information. This type of network is known under the name of fuzzy Petri-nets (FPN) and was proposed for the first time by Looney [14, 15]. Such an approach is appropriate for the case where a state of the modeled system corresponds to a marking of the associated FPN. The goal of this paper is to identify the speech emotion recognition using MFCC feature extraction and FPN, and learning automata based on three emotions. Section 2 discussed speech recognition feature extraction method which uses MFCC. Then fuzzy Petri-net (FNP) is applied for the classification of feature's vectors in section 3. Learning Automata, Fuzzy Rules in FPN and proposed algorithms are introduced in sections 4 and 5. Finally, test results and the conclusion are given in sections 6 and 7.

## **2. Mel Frequenc Cepstrum Coefficients (MFCC)**

One of the most important steps in emotional identification of speech recognition is extracting suitable features that identify different speech states in an efficient way. Speech features are divided in four main groups and numerous methods are introduced by researchers to obtain a powerful feature extractor. These four groups are: continuous speech, qualitative, spectral and TEO based classes [16]. In Price (2004), the speech sound and energy features are grouped under spectral classes and used in identification of speech recognition [17]. The MFCC algorithm is one of the efficient methods grouped under spectral features and is used for description of input speech.

Statistical characteristics of MFCCs depend on noise state. Recognition systems should be effective in true environmental conditions. Since true environments are noisy; therefore, a special noise-resistant system should be selected. Another reason to make speech recognition systems resistant to the environment is the different communicating channels of speaker and computer. The MFCC method is based on the human ear, and as such does not extract frequencies higher than 1 KHz [18]. Since the MFCC is effective in noisy environments and accounts for the differing communication channels it is

used to extract speech emotion recognition features. This method includes the following mathematical approaches:

### Step 1- Pre-emphasize:

The signal passes a filter of high frequency emphasis by Eq. (1), [19].

$$Y[n] = X[n] - 0.95 X[n - 1] \quad (1)$$

### Step 2- Framing:

The process of classifying speech samples from analogue conversations is changed into digital conversations in small frames of 20 to 40 ms length. The speech signal is divided in N frames. Vicinity of separated frames by M is ( $M < N$ ) where  $M=100$  and  $N=256$ , [19].

### Step 3- Hamming Window:

Windows are defined with  $W(n)$  and  $0 \leq n \leq N-1$ . N is the number of samples in each frame and  $Y[n]$  is the output signal. Input signals are shown by  $X(n)$  and windowing results are shown in Eq.s (2) and (3), [19].

$$Y(n) = X(n) \times W(n) \quad (2)$$

$$w(n) = 0.54 - 0.46 \cos\left[\frac{2\pi n}{N} - 1\right] \quad 0 \leq n \leq N - 1 \quad (3)$$

### Step 4- Fast Fourier Transform:

Each frame is changed into frequency amplitude from time amplitude with N. The Fourier Transform is used to reverse pulse complexity related to glottal  $U[n]$  and the instigation of vocal tract  $H[n]$  in time domain and is calculated by Eq. (4), [19].

$$y(w) = \text{FFT} [h(t) * X(t)] = H(w) * X(w) \quad (4)$$

### Step 5- Processing Mel Filter bank:

Cepstral coefficients are mainly obtained from the output of a set of filter banks, which suitably covers the full range of defined frequency spectrum. Generally the set of 24 buffer filters were used, which are similar to human ear performance. Filters were placed along the frequency axes variably. More filters were allocated to the part of the spectrum below 1 KHz, since they have more information about sound. Frequency responses of filter banks were simulated according to ear performance. Filtering is referred to as conceptual weighting.

Non-linear frequency analysis was used to obtain the frequency/time analysis. However, while using thin buffer filters for low frequencies can discover harmonics, they give weak primary information. Using wider band filters in higher frequencies gives a better signal analysis. In Automatic Speech

Recognition systems (ASR), the frequency span in use is similar to that of human conception. In the Mel filter bank, triangular filters are used. The central frequency for filters lower than 1 KHz is linear and then changes in a logarithmic manner for higher frequencies [16, 20]. Filter outputs are a set of their filtered spectral components: Eq. (5) indicates calculation of Mel for given frequencies in Hz [19]:

$$F(mel) = [2595 * \log_{10} [1 + f] / 700] \quad (5)$$

#### Step 6- Cos transform:

This step is a process for transferring Mel spectrum to time span using the Discrete Cosine Transform (DCT). Results are Cepstral coefficients for Mel frequency and the set of coefficients is referred to as acoustic vectors.

### 3. Fuzzy Petri-net

After extracting the desired features from the input data, they are applied to the decision making stage in order to make the appropriate decision on the specific class that the input data belongs to. In this work FPN is used as the decision making tool. Petri-nets developed by Carl Adam Petri in his Ph.D. thesis in 1962, are generally considered as a tool for the studying and modeling of systems. A Petri-net (PN) is for most a mathematical description, but it is also a visual or graphical representation of a system. The application areas of Petri-nets being vigorously investigated involve knowledge representation and discovery, robotics, process control, diagnostics, grid computation, traffic control, to name a few high representative domains [21]. Petri-nets (PNs) are a mathematical tool to describe concurrent systems and model their behavior. They have been used to study the performance and dependability of a variety of systems. Petri-nets are a directed graph that consists of two types of nodes [21]: places and transitions, where circles represent places and bars represent transitions. Each place represents an antecedent or consequent and may or may not contain a token associated with a truth degree between zero and one which speaks for the amount of trust in the validity of the antecedent or consequent. Each transition representing a rule is associated with a certainty factor value between zero and one. The certainty factor represents the strength of the belief in the rule. The relationships from places to transitions and vice versa are represented by directed arcs. The concept of FPN is derived from Petri-nets. As in looney (1998), a generalized FPN structure can be defined as having 8 parameters [14]:

FPN = (P, T, D, I, O,  $\mu$ ,  $\alpha$ ,  $\beta$ ); a finite set of places is  $P = \{p_1, p_2, \dots, p_n\}$ ; a finite set of transitions is  $T = \{t_1, t_2, \dots, t_m\}$ ; a finite set of propositions is  $D = \{d_1, d_2, \dots, d_n\}$ ;  $P \cap T \cap D = \emptyset$ ,  $|P| = |D|$ ; the input function is  $I: P \times T \rightarrow \{0,1\}$ ; the output function is  $O: T \times P \rightarrow \{0,1\}$ , is an association function  $\mu: T \rightarrow \{0,1\}$ ; and the range of mapping from transitions is  $[0,1]$ . That is, certainty factors of  $\alpha: T \rightarrow \{0,1\}$  and  $\beta: T \rightarrow \{0,1\}$  are an association function that  $\alpha$  is a mapping from places to  $[0,1]$  and  $\beta$  is a mapping from places to propositions.

### 4. Learning Model and Fuzzy Rules in FPN

#### 4.1. Variable Structure Stochastic Automata (VSSA)

The most important application of learning automata is to estimate the parameters that are, used in identification of pattern and play theory [22, 23, 31]. In the book by Narendra and Thathachar (1989), all researches up to late 1980 have been reviewed and discussed [24, 25]. Numerous examples and applications of learning automata have been presented in [26]. Another important feature of learner systems is their potential to improve efficiency with time. As we know, the objective of a learner system is to improve the function which is not totally identified [27]. Therefore an approach to the problem is to decrease the objectives of the learner system to the subject of optimization which is defined on a set of parameters which aim at finding the optimum parameters set. Reinforcement learning is a common method in finding the optimum objective. The main benefit of reinforcement learning compared with other learner methods is that it requires no information from the environment (than reinforcing signal) [24]. Stochastic learning automata are one of the reinforcement learning models which try to increase their efficiency and are divided into stochastic automata of fixed and variable structures [25].

A VSSA is defined by the quadruple  $E = \{a, b, p, T\}$  in which  $\{a_1, a_2, \dots, a_r\}$  is a set of actions (or outputs of the automaton). The output or action of an automaton is an instant of  $n$  denoted by  $a(n)$ , which is an element of the finite set  $\{a_1, a_2, \dots, a_r\}$ .  $\{b_1, b_2, \dots, b_r\}$  represents the input set or response set;  $p = \{p_1, p_2, \dots, p_r\}$  represents the action probability set; and finally  $p(n+1) = T(a(n), b(n), p(n))$  represents the learning algorithm. The following shows the operation of the automaton based on the action probability set  $p$ . The automaton randomly selects an action  $a_i$ , and performs it on the environment. This paper only focuses on the S-model of variable structure automata [24, 25]. The S-model of environment response, which is considered as a described function to improve automata, has been added to the original learning automata. The S-model is divided into three classifications and  $a, b$  are reward and penalty parameters. When  $a, b$  are equal, the algorithm is called S-LRP, When  $0 < b < a < 1$  the algorithm is called S-LR $\epsilon$ P and whenever  $b$  is zero the algorithm is called S-LRI [24, 25].

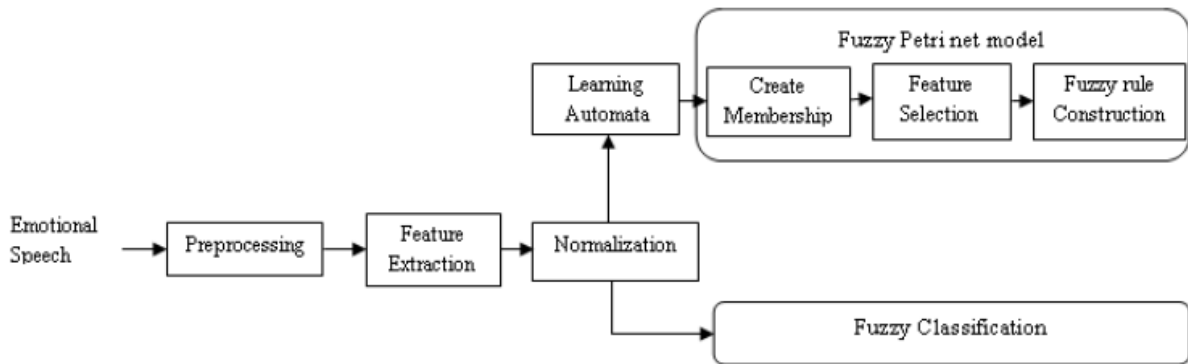
#### 4.2. Fuzzy Emotion Model:

According to [28], the Fuzzy Emotion (FE) model consists of  $n$  basic emotions  $e_1, e_2, \dots, e_n$  in a finite set. The output of the features level is the input of fuzzification. In this part, each feature vector proposed model should determine a membership. So the fuzzy membership function introduced by  $\{\mu_{FE_j} : FE \rightarrow [0, 1], j = 1, 2, \dots\}$ . And each  $FE_j := \{(e_i, \mu_{FE_j}(e_i) \mid e_i \in FE), j = 1, 2, \dots\}$  computes a fuzzy set corresponding to one emotional state  $E_j$  [28]. If  $FE, \mu_{FE_j}$  and  $FE_j$  are defined as described above, we shall use the membership vector  $(\mu_{FE_j}(e_1), \mu_{FE_j}(e_2), \dots, \mu_{FE_j}(e_n)) =: (\mu_{FE_j}(e_i))$  to denote a point in an  $n$ -dimensional space and each axis corresponds to one basic emotion  $e_i$ . Thus a membership vector  $(\mu_{FE_j}(e_i))$  corresponds to one emotional state  $E_j$  and can be interpreted psychologically as vector of emotion intensities  $(I_{e_i}) := (I_{e_1}, I_{e_2}, \dots, I_{e_n})$  [28].

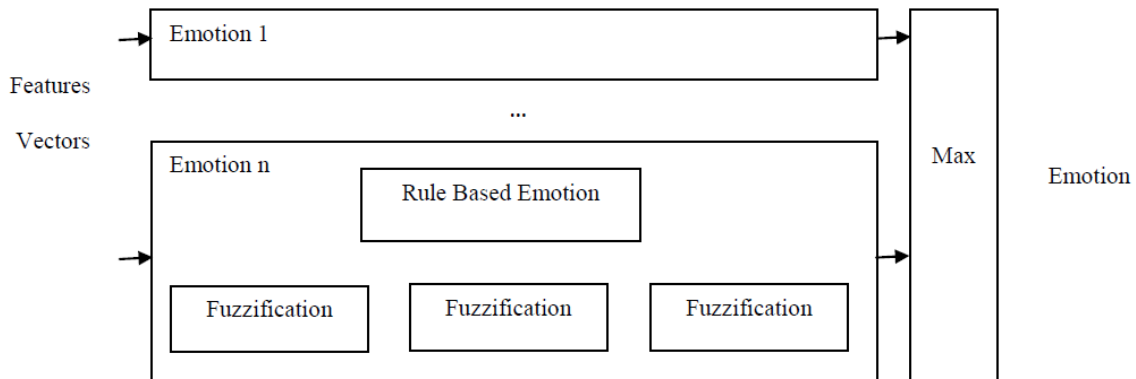
### 5. The Proposed Algorithm

This section explains the proposed algorithm with a combination of Fuzzy rules and learning automata in the Petri-net. This algorithm is a new technique to discover the rate of speech emotion recognition based on the rule's inference in FPN and the effective model of learning automata. Hence human's mind decision is an attractive and challenging topic, this paper tend to introduce a close

model like human’s mind decision. The reason to choose FPN has been its capability in extracting inferences and analyzing different rules. Learning automata is a powerful model of learning for prediction in the dynamic environment. Therefore here it has been used to adjust the membership functions for each feature vector in a dynamic environment to inference the rules, like in a mind decision. Figure 1 presents the description of the proposed algorithm with different parts including preprocessing, feature extraction, learning automata, fuzzification and defuzzification in FPN model. Figure 2 shows the action of the fuzzy structure to combine the features vectors with the operation of Max after exiting from the classification level in FPN.



**Figure 1.** Recommended system block diagram



**Figure 2.** Fuzzy structure

### 5.1. Fuzzy Rules

After feature extraction level, each basic emotion feature is introduced as an input to learning automata for getting the membership function. In other words, the process of adjusting fuzzy membership functions is used by learning automata. Then the output of this level is the input of FPN ( $e_i, i = 1, \dots, n$ ). FPN consists of different parts including fuzzification, inference engine and defuzzification. Each rule takes the fuzzified features  $f_j, j = 1, \dots, K$ , ( $K$  is 6), as input and produces a

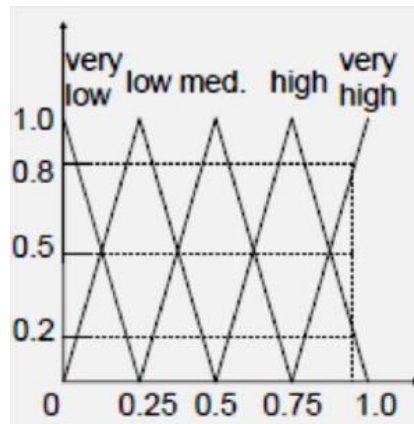
fuzzy emotion value  $l_{ei}$  as output. This part computes each feature  $f_j$  and emotion intensity  $l_{ei}$  by five triangular membership functions (very low, low, medium, high and very high) as schematically depicted in figure 3. However, the actual start and end coordinates, as well as the maximum coordinates, are generated automatically during the training phase. This diagram shows the fact the proposed algorithm can support the real time emotion recognition and it allows us to distinguish degrees to which features or emotions are presented in the current input of fuzzification. Each emotion is defined by the information below, and after that is used in each rule [28]:

**Happiness:** The mean pitch of the speaker is high, it has a high variance, the rhythm is rather fast and the contours of all syllables are rising.

**Anger:** The mean pitch is high, it has a high variance, the rhythm is fast, with little variance of phoneme durations and the pitch contours of all syllables are falling.

**Sad:** The mean pitch is low, it has a low variance, the rhythm is slow, with high variance of phoneme durations and the contours of all syllables are falling.

**Natural:** The mean pitch is high but less than happiness, the rhythm is slow, with a high variance of phoneme durations and the contours of syllables are rising.



**Figure 3.** Membership function for features and emotions

The rule set for the emotion  $e_i$  is generated by a fuzzified step. Since this approach uses only the AND connector it generates  $5^{K+1}$  rules of the following form:

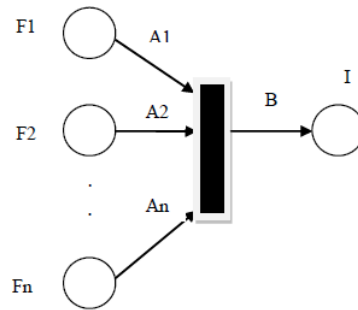
*IF  $f_1$  IS very low AND ... AND  $f_k$  IS very low THEN  $l_{ei}$  IS very high*

*IF  $f_1$  IS very low AND ... AND  $f_k$  IS low THEN  $l_{ei}$  IS very high*

*IF  $f_1$  IS very low AND ... AND  $f_k$  IS medium THEN  $l_{ei}$  IS medium*

The number of rules could be reduced, if the OR connector or rule pruning could be used. For defuzzification of emotion values we use the center of gravity method. By projecting the center of gravity method to the x-axis we calculate the corresponding emotion intensity. Hence, a three dimensional vector  $(I_h, I_s, I_a) = (\mu_E(h), \mu_E(s), \mu_E(a))$  containing the intensities of the three emotions happiness  $h$ , sad  $s$  and anger  $a$ , is generated. In order to gain emotion recognition that is determined by the Max operation. If the Max cannot be determined unambiguously, since two or more intensity values are maximal that emotion is selected which was recognized for the previous sentence. It is obvious that there is a special mapping between IF-THEN rules and FPN. Also, each IF-THEN rule can be explained as:

$R1 := IF f1 \text{ is } A11 \text{ AND}, \dots, \text{ AND } fn \text{ is } A1n \text{ THEN } I \text{ is } B1$ . Figure 4 can be expressed as the fire of each rule in Petri-net:



**Figure 4.** FPN denoted fuzzy rule

## 6. Test Result

### 6.1. Dataset

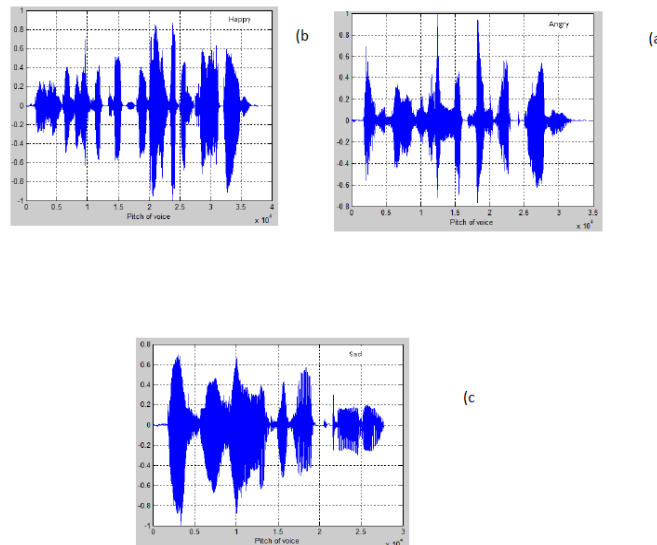
The “Berlin Database of Emotional Speech” is used to train and test the algorithm in this paper [29]. The Berlin Database consists of 535 speech samples, which contain German utterances related to emotions such as anger, disgust, fear, happiness, sad, surprise and neutral, acted by five males and five females. This database was chosen to be used because of the following reasons: (1) the quality of its recording is very good, and (2) it’s a public and popular database in emotion recognition based on the literature. This paper focused on only three main emotions from Berlin database; anger, happy and sad, in order to achieve to a higher and more accurate rate of recognition.

### 6.2. Lab Result

This paper introduced a new method of classification in speech emotion signals based on the Berlin database. Berlin is one of the public databases with good quality of voice which consists of 535 speech samples and contains German utterances. Three emotions including happy, anger and sad are studied in this paper using the proposed algorithm with four important steps. In the first step, all input emotional signals are normalized. Then the MFCC method is applied for the purpose of feature extraction. The output of the second step is the vectors of emotional speech features. For means of classification by FPN, each feature should have a suitable membership and this work was executed by SLRP (S-model of learning automata). Then the Fuzzy rule is applied on each vector followed by the last step, computing the rate of recognition by using defuzzification. Table I indicates the result of the proposed algorithm. The highest learning rate achieved is 98.50, which is related to the anger



situation. In order to show the efficiency of the proposed model, this paper used the data mining tool of OpenSMILE. Also, the proposed model has been compared with SVM, KNN, NN and MLP classifiers. The OpenSMILE toolkit is a flexible feature extractor for speech processing and machine learning applications, which is fully written in C++ [30]. Table II indicates the rate of speech emotion recognition using the MFCC feature extraction with Opensmile toolkit and proposed classification method. The result of recognition using the MFCC by the OpenSMILE toolkit and the three layers MPL methods for feature extraction and classification are respectively shown in table III.



**Figure 5.** a, b and c Show the different pitches of voice with three emotions (Happy, Anger and Sad)

Figure 5. a, b, c indicates the different pitches of voice after feature extraction step. Table I indicates the highest rate of emotional speech recognition would be 98.50 percent by using proposed algorithm for state of anger.

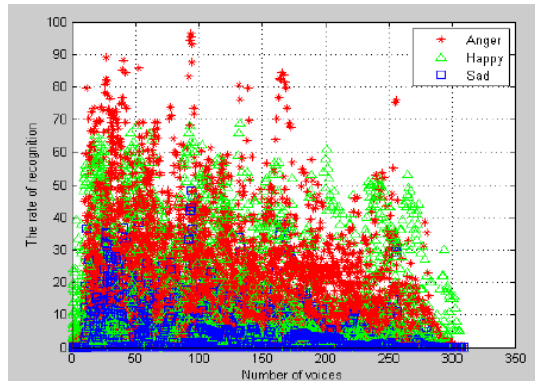
**Table I.** Speech emotion recognition using the proposed algorithm

Proposed Algorithm	Anger	Happy	Sad
Anger	98.50	-	-
Happy	-	70.01	-
Sad	-	-	49.98

Figure 6 shows the rate of recognition based on the Berlin Dataset for each of anger, happy and sad emotions. It shows the high rate of recognition given to the anger state.

Table II and table III indicate the rate of emotional speech recognition by using the MFCC's OpenSMILE feature's extraction in two different models of classification. The first one is FPN and the

second one is the MLP classification. The highest rate of recognition relates to the anger state by 90.50 and 86.0 respectively. Therefore, it can be claimed that the FPN classification gives better results than other models that are mentioned in this paper.



**Figure 6.** The rate of recognition based on the Berlin dataset in different situation (Anger, Happy and Sad)

**Table II.** Speech emotion recognition using MFCC features extraction by Open SMILE toolkit and the proposed classification

MFCC feature by using Opensmile toolkit+FPN classification	Anger	Happy	Sad
Anger	90.50	-	-
Happy	-	65.25	-
Sad	-	-	20.0

**Table III.** Speech emotion recognition using MFCC features extraction by Open SMILE toolkit and three MLP classification

MFCC feature by using Opensmile toolkit +MLP classification	Anger	Happy	Sad
Anger	86.0	-	-
Happy	-	55.0	-
Sad	-	-	12.50

Table IV shows the average rate of speech emotion recognition by using SVM, KNN, HMM and NN classifiers. This table shows that the highest rate of recognition is for the HMM classifier (89.90%). Comparing rates of recognition observed in table 4, with the results obtained from the proposed model of classification, indicates that the proposed model is more efficient than other models of classification.

**Table IV.** Speech emotion recognition using MFCC features extraction and SVM, KNN, HMM, NN classifications

Different Classification	Rate of Recognition
MFCC+SVM	85.40
MFCC+KNN	86.70
MFCC+HMM	89.90
MFCC+NN	86.10

## 7. Conclusion

This paper used a new method to classify the features that were extracted by the MFCC method for speech emotion recognition. The foundation of proposed model is inference the rules like minding decision. The proposed model is used the SLRP model of learning automata and the fuzzy method in Petri-net. FPN is a new model of classification which is introduced on emotional speech recognition. Also, the S-model learning automata used to find the membership functions of each vector [11]. In the first step of classification, learning automata gave a membership function to each of MFCC's feature. This model of learning works based on penalty and reward which is received from the environment. In the second step of classification, according to the selected operations, the amounts of input parameters are converted to a fuzzy module that called fuzzification. So, the inference engine is responsible for analyzing the rules like human's mind. In the final step, rules and achieved results are converted to numerical values called defuzzification like minding decision. Results obtained highlighted that the highest rate of speech emotion recognition among three emotions of anger, happy and sad situations in the proposed model of classification is for anger by 98.50%. In order to compare the recommended model with other methods, MFCC feature extraction is computed by the OpenSMILE toolkit for the proposed model of classification and the three layers of MLP classification on the Berlin database. It resulted in better values for the proposed model of classification. Then the proposed model of classification is compared with SVM, KNN, NN classifiers. Experimental results explained that the proposed model of classification achieved a higher rate of recognition than other models in the speech emotion recognition system. Therefore, it can be claimed that the combination of learning automata and the fuzzy method in Petri-net outperforms the other models.

## References

- [1] T.L. Pao, Y.T. Chen, and J.H. Yeh, Comparison of Classification Methods for Detecting Emotion from Mandarin Speech, *IEICE Transactions on Information and Systems*, vol. E91-D(4) (2008).
- [2] T. Nose, and T. Kobayashi, A Technique for Estimating Intensity of Emotional Expressions and Speaking Styles in Speech Based on Multiple-Regression HSMM, *IEICE Transactions on Information and Systems* (2010).
- [3] R. Cowie, E. Douglas-Cowie, N. Tsapatsoulis, G. Votsis, S. Kollias, W. Fellenz, and J.G. Taylor, Emotion recognition in human-computer interaction, *IEEE Signal Process*, vol. 18(1) (2001).
- [4] D. Erickson, Expressive speech: Production, perception and application to speech synthesis, *Acoustical Science and Technology*, vol. 26(4) (2005).
- [5] D. Neiberg, K. Elenius, and K. Laskowski, Emotion recognition in spontaneous speech using GMMs, *Proc. INTERSPEECH* (2006).
- [6] T.L. Nwe, S.W. Foo, and L.C.D. Silva, Speech emotion recognition using hidden Markov models, *Speech Communication*, vol. 41(4) (2003).
- [7] O. Pierre-Yves, The production and recognition of emotions in speech: features and algorithms, *International Journal of Human-Computer Studies*, vol. 59 (2003).
- [8] C.M. Lee, and S. Narayanan, Toward Detecting Emotions in Spoken Dialogs, *IEEE Transactions on Speech and Audio Processing*, vol. 13(2) (2005).
- [9] I. Albrecht, and M. Schroder, and J. Haber, and H.-P. Seidel, Mixed feelings: Expression of non-basic emotions in a muscle-based talking head, *Virtual Reality*, vol. 8(4) (2005).
- [10] D. Wu, and T.D. Parsons, and S. Narayanan, Acoustic Feature Analysis in Speech Emotion Primitives Estimation, *Proc. InterSpeech* (2010).
- [11] M. Schroder, and R. Cowie, and E.D.-cowie, M. Westerdijk, and S. Gielen, Acoustic Correlates of Emotion Dimensions in View of Speech Synthesis, *Proc. Eurospeech* (2001).
- [12] M. Grimm, and K. Kroschel, Emotion Estimation in Speech Using a 3D Emotion Space Concept, in *Robust Speech Recognition and Understanding*, M. Grimm and K. Kroschel (2007).
- [13] I. Kanluan, M. Grimm, and K. Kroschel, Audio-Visual Emotion Recognition Using an Emotion Space Concept, *Proc. EUSIPCO* (2008).
- [14] C.G. Looney, Fuzzy Petri-nets for rule-based decision making, *IEEE Transactions on Systems, Man and Cybernetics*, Vol. 18, No. 1 (1988).
- [15] L. Shiyon, Fuzzy Control, *Neuro Control And Intelligent Cybernetics*, Harbin Institute of Technology (1998).
- [16] T. Seehapoch, S. Wongathanavasu, Speech Emotion Recognition Using Support Vector Machines, *5th International Conference on Knowledge and Smart Technology* (2013).
- [17] J. Price, Design an automatic speech recognition system using Matlab, *University of Maryland Eastern Shore Princess Anne* (2004).
- [18] A. Kamarul, A. B. Ibrahim, Biomedical engineering laboratory student pack, *UTM Johor* (2005).
- [19] L. Muda, M. Begam, I. Elamvazuthi, Voice Recognition Algorithms using Mel Frequency Cepstral Coefficient and Dynamic Time Warping Techniques, *Journal of Computing* (2010).
- [19] M. Bojanic, V. Crojevic, V. Deliv, Application of Neural Network in Emotional Speech Recognition, *IEEE* 20-22 (2012).
- [20] M. Schroder, R. Cowie, E.D.-cowie, M. Westerdijk, S. Gielen, Acoustic Correlates of Emotion Dimensions in View of Speech Synthesis, *Proc. Eurospeech* (2001).
- [21] H. Beigy, M. R. Meybodi, A Mathematical Framework for Cellular Learning Automata, *Advances on Complex Systems*, Vol. 7, No. 3 (2004).
- [22] V. Raghunathan, C. Schurgers, S. Park, and M. B. Srivastava, Energy-Aware Wireless Microsensor Networks, *IEEE Signal Processing Magazine*, Vol 19 (2002).

- [23] K. S. Narendra, M. A. L. Thathachar, Learning automata: An introduction, Prentice Hall (1989).
- [24] M.A.L. Thatachar, P.S. Sastry, Varieties of Learning Automata: An Overview, IEEE Transaction on System, Man, and Cybernetics- part B: CYBERNETICS (2002).
- [25] J. Winter, Y. Xu, and W. C. Lee, Energy Efficient Processing of K Nearest Neighbor Queries in Locationaware Sensor Networks, the Second International Conference on Mobile and Ubiquitous Systems: Networking and Services (MobiQuitous'05), San Diego,CA (2005).
- [26] K.S. Narendra, M. A. L. Thathachar, Learning automata a survey, IEEE Transactions on Systems, Man and Cybernetics, vol. 4, no. 4 (1974).
- [27] N. Esau, L. Kleinjohann, B. Kleinjohann, An Adaptable fuzzy Model for Emotion Recognitio (2003).
- [28] F. Burkhardt, A.Paeschke, M. Rolfes, W. Sendlmeier, B. Weiss, A database of German emotional speech, Proceedings of the Interspeech (2005).
- [29] F. Eyden, M. Wollmer, and B. Schuller, Opensmile: the munich versatile and fast opensource audio feature extractor, 10 Proceedings of the international conference on Multimedia (2010).
- [30] M. M. Javidi, E. F. Roshan, Speech emotion recognition by using combinations of C5.0, Neural Network (NN), and Support Vector Machine (SVM) classification methods, Journal of mathematics and computer science (2013).
- [31] N. Moarefi, A. Yarmohamadi, The implementation of cellular automata with non-identical rule on serial base, Journal of mathematics and computer science (2012).