Biologically Inspired Cognitive Architectures 19 (2017) 32–38

Contents lists available at ScienceDirect

Biologically Inspired Cognitive Architectures

journal homepage: www.elsevier.com/locate/bica

Research article

Speech emotion recognition based on a modified brain emotional learning model

Sara Motamed [a], Saeed Setayeshi [b], Azam Rabiee [c,*]

[a] *Department of Computer Engineering, Science and Research Branch, Islamic Azad University, Tehran, Iran*
[b] *Department of Medical Radiation, Amirkabir University of Technology, Tehran, Iran*
[c] *Department of Computer Science, Dolatabad Branch, Islamic Azad University, Isfahan, Iran*

A R T I C L E I N F O

Article history:
Received 4 August 2016
Revised 28 December 2016
Accepted 29 December 2016

Keywords:
Emotional speech
Brain emotional learning (BEL)
Adaptive neuro fuzzy inference system (ANFIS)
Multi-layer perceptron (MLP)

A B S T R A C T

This paper introduces an optimized model of brain emotional learning (BEL) that merges the Adaptive Neuro-Fuzzy Inference System (ANFIS) and Multilayer Perceptron (MLP) for speech emotion recognition. This model is developed based on the limbic system of the mammalian brain in order to present a desirable learning model for speech emotion recognition in dynamic situations like the brain's emotional networks. The proposed model has four main parts including thalamus, sensory cortex, orbitofrontal cortex and amygdala. As this model is inspired from human brain, ANFIS has been used in this study to develop a model in subsystems of amygdala and orbitofrontal cortex to make rules. To classify the rate of speech emotion signals, the ANFIS outputs are given to MLP network. Experimental results show that the proposed algorithm has a powerful capability to identify a human's speech emotion.

Introduction

Emotion recognition refers to the ability to detect human feelings and conditions. It is also one of the most effective methods to obtain information from humans in order to improve the interaction between humans and machines (R'azuri et al., 2015). Many researchers have investigated different computational models of emotional learning based on the mammalian brain and proved that the limbic system is responsible for human emotion. In this context, a famous computational model was presented by Morén and Balkenius (2000). This model which is called brain emotional learning (BEL), shows the function of the limbic system considering that the brain's limbic system is an important part of the emotional process. BEL is controlled by reinforcement signals (reward) and reinforcement learning. Lucas showed an optimized model of the BEL in various control applications (Morén & Balkenius, 2000). Lotfi (2013), Lotfi & Akbarzadeh, (2013a, 2013b) introduced an adaptive version of the BEL model in different subjects.

Despite BEL simple structure in modelling amygdala and orbitofrontal cortex systems. In this study, the original BEL model has indicated undesirable performance for speech emotion recognition. Thus, emotional learning models with different structures and functions using the amygdala-orbitofrontal cortex subsystem are suggested in this research.

In another study, Parsapoor et al. (2012) worked on the brain emotional learning based on fuzzy inference system (BELFIS) model developed based on BEL with changes in the subsystem of the amygdala and orbitofrontal cortex. The BELFIS model employed the adaptive neuro fuzzy inference system (ANFIS) model in the subsystem of amygdala and orbitofrontal cortex. While, in our study, BELFIS did not perform as well as BEL in speech emotion recognition, we worked to develop a model to increase the rate of speech emotion recognition.

Therefore, our aim in this study is to improve the original BEL model, especially for speech emotion recognition. The main advantage of the proposed model is that it is a biologically inspired method, therefore, it is expected to be usable in cognitive classification models.

In the proposed model, the main changes are made in the learning parameters and the subsystems of the amygdala and the orbitofrontal cortex. As the ANFIS has good ability and performance in identification, prediction and control of problems and has been applied in many different systems (Güler & Übeyli, 2004; Kamaruddin & Wahab, 2008; Silarbi, Abderrahmane, & Phonetical, 2014), therefore, we have used ANFIS model to make the fuzzy inference rules. On the other hand, as the MLP is found to have an adaptive well-defined training algorithm and is easy to implement (Han, Yu, & Tashev, 2014; Khanchandani & Hussain, 2009; Shaw, Vardhan, & Saxena, 2016), it has been

* Corresponding author.
E-mail addresses: samotamed@yahoo.com (S. Motamed), setayesh@aut.ac.ir (S. Setayeshi), azrabiee@gmail.com (A. Rabiee).

http://dx.doi.org/10.1016/j.bica.2016.12.002
2212-683X/

employed in making decision based on the amygdala and orbitofrontal cortex outputs, instead of using a linear function. In other words, for improving the rate of the speech emotion classification, we utilize the MLP model to fuse the amygdala and orbitofrontal cortex outputs.

We employ the ANFIS and MLP (multi-layer perceptron) networks in the subsystems of the amygdala and the orbitofrontal cortex to find nonlinear boundaries separating the emotional states. The rest of the paper is organized as follows. Section 'Inspired from the human auditory pathway' discusses about the human auditory pathway. The modified BEL model and experiments are explained in Sections 'Modified BEL-based speech emotion recognition' and 'Experiments', respectively. The conclusion is given in the last section.

Inspired from the human auditory pathway

According to many studies, the limbic system is responsible for emotional responses to auditory stimuli (LeDoux, 1998). The auditory pathway has many different levels, which leads to an activity, such as music perception, speaker identification, speech perception, and speech production (Watts, 2012). The aim of this paper is to find a computational model of speech emotion recognition based on the human auditory cortex. Fig. 1 shows a block diagram of the human auditory pathway for speech perception which is extensively inspired by the human auditory pathway discussed by Watts (2012).

As shown in Fig. 1, speech enters to the left and right cochleas. Then, the cochleas create a spectro-temporal representation of the sound. Later, the 2D representation projects from the cochlear nerve into the lower brainstem. The lower brain stem and colliculus are responsible for preprocessing, scaling, and normalization. From there, signals enter the inferior and superior colliculus and thalamus. The thalamus is responsible for controlling attention and causes signals go from there to the limbic system and to the primary auditory cortex. The limbic system is responsible for processing emotional stimuli when Morén argues that it has four main parts including sensory cortex, thalamus, orbitofrontal cortex, amygdala (Morén, 2002; Morén & Balkenius, 2000). Finally, the processed signals in limbic system and the primary auditory cortex are projected to the specialized pathways for speech recognition, production, speaker identification, and music perception (Watts, 2012).

Modified BEL-based speech emotion recognition

Review of the BEL model

One of the most important models developed based on the limbic system is the BEL model, which is inspired by the mammalian brain. This model consists of four main subsystems including the amygdala, orbitofrontal cortex, thalamus, and sensory cortex (Lotfi, 2013; Lotfi & Akbarzadeh, 2013a, 2013b; Morén & Balkenius, 2000), and each of them has an important role in making decisions. For example, the thalamus is responsible for preprocessing, noise reduction and filtering, and giving the preprocessed data to sensory cortex and amygdala. The sensory cortex is responsible for normalizing and analyzing the received signals. The orbitofrontal cortex tries to prevent inappropriate responses from reaching the amygdala (Grossberg & Seidman, 2006; Hall & Guyton, 2010). The amygdala is responsible for making decisions, so it has an essential role in emotional learning. The subsystem of the amygdala and orbitofrontal cortex also receive a reinforcing signal (primary reward) that has been left unspecified in the Morén and Balkenius model (2000).

The overall block diagram of the BEL model is shown in Fig. 2. In this figure, $\boldsymbol{s}$ is the stimuli inputs and there is one $\boldsymbol{a}$ node for each stimulus $\boldsymbol{s}$ in the network model of amygdala. The input of the sensory cortex is $\boldsymbol{a}_{th}$, which is the maximum of the input feature vector (Eq. (1)). It means that in thalamus, the maximum value of the input vector is sent to the sensory cortex (Morén & Balkenius, 2000).

$$\boldsymbol{a}_{th} = \max(\boldsymbol{s}_i) \tag{1}$$

The sensory cortex analyzes and normalizes $\boldsymbol{s}$ by Eq. (2), and distributes it between the amygdala and the orbitofrontal cortex.

$$\boldsymbol{s_i} = \frac{\boldsymbol{s}_i - \boldsymbol{s}_{min}}{\boldsymbol{s}_{max} - \boldsymbol{s}_{min}} \tag{2}$$

In the amygdala, there is a plastic connection weight $\boldsymbol{v}_i$ to $\boldsymbol{s}_i$. The output of each node is obtained by multiplying any input with the weight $\boldsymbol{v}$ (Morén & Balkenius, 2000).

$$\boldsymbol{a} = \boldsymbol{s} \cdot \boldsymbol{v} \tag{3}$$

The output of the amygdala, $\boldsymbol{e}_a$, is computed by the simple summation of its elements (Morén & Balkenius, 2000):

$$\boldsymbol{e}_a = \sum_i \boldsymbol{a}(i) \tag{4}$$

The r is a reinforcement signal and γ is used to adjust the learning speed (Morén & Balkenius, 2000):

$$\Delta\boldsymbol{v} = \mathrm{diag}(\gamma \cdot \max(r - \boldsymbol{e}_a, 0) \cdot \boldsymbol{s}) \tag{5}$$

The max operation causes monotonic learning between input stimulus and information formed in learning weights. The outputs of the amygdala go to the inputs of the orbitofrontal cortex. The task of the orbitofrontal cortex is to react to critical situations. $\boldsymbol{o}$ is the output of the orbitofrontal cortex. According to Eq. (6), $\boldsymbol{o}$ nodes behave analogously, with a connection weight $\boldsymbol{w}$ employed to the input signal to create an output (Morén & Balkenius, 2000).

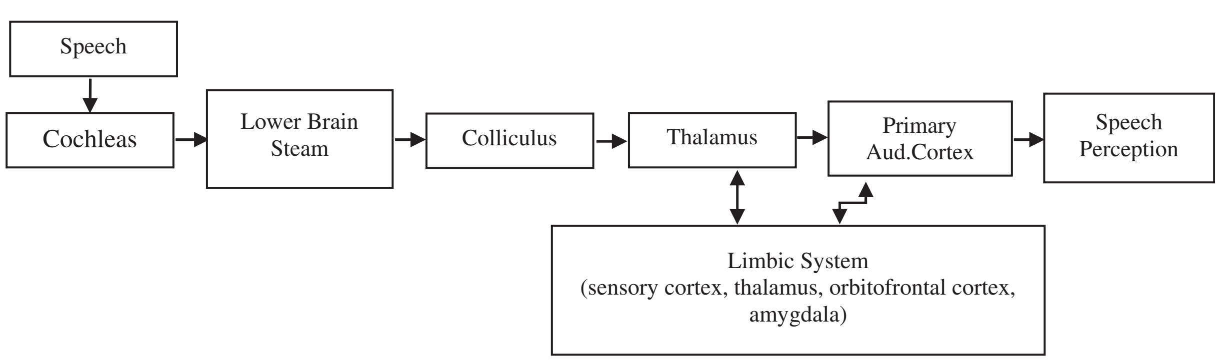

Fig. 1. Recommended model is inspired by human auditory pathway, starting from speech signal as input and ending to speech perception.

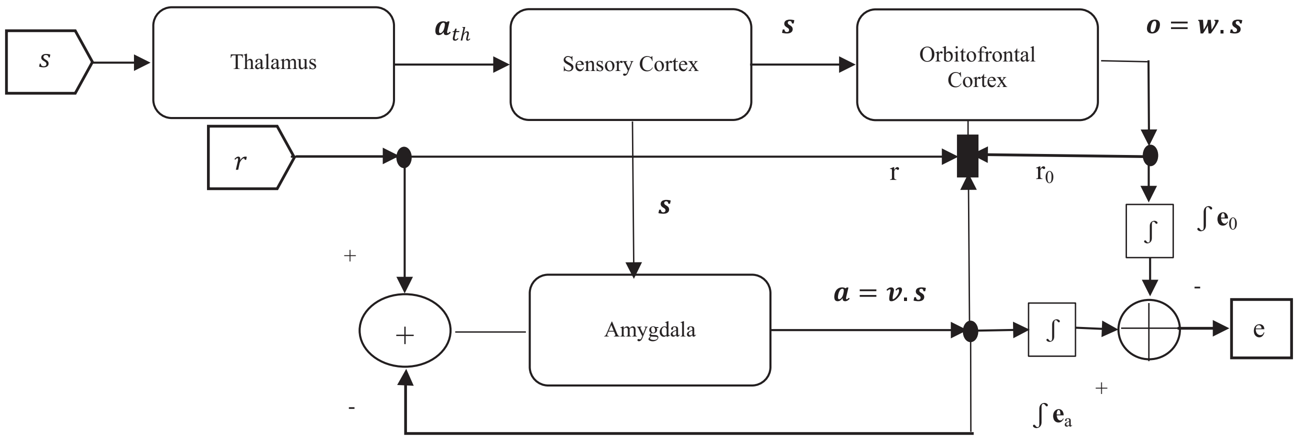

Fig. 2. The BEL model inspired by limbic system has four main subsystems including thalamus, sensory cortex, orbitofrontal cortex and amygdala (Morén & Balkenius, 2000).

$$\boldsymbol{o} = \boldsymbol{w} \cdot \boldsymbol{s} \tag{6}$$

The total output of the orbitofrontal cortex is the sum of all output nodes:

$$\boldsymbol{e}_o = \sum_j \boldsymbol{o}(j) \tag{7}$$

β is another learning rate. r_0 is the internal reinforcement signal of the orbitofrontal cortex cortex. The connection weights ($\boldsymbol{w}$) are updated by Eq. (8) (Morén & Balkenius, 2000):

$$\Delta \boldsymbol{w} = \mathrm{diag}(\beta \cdot r_0 \cdot \boldsymbol{s}) \tag{8}$$

r_0 is the internal reinforcement signal of the orbitofrontal cortex and is calculated by Eq. (9) (Morén & Balkenius, 2000):

$$r_0 = \begin{cases} \max(\boldsymbol{e}_a - r, 0) - \boldsymbol{e}_o & \text{if } r \neq o \\ \max(\boldsymbol{e}_a - \boldsymbol{e}_o, 0) & \text{Otherwise} \end{cases} \tag{9}$$

In the previous papers of the BEL model, r is computed differently (Lucas, 2010; Morén & Balkenius, 2000). According to Eq. (9), r_0 is internal reward and r is the input binary reinforce signal. In the training phase, if $r \neq 0$, then the max $(\boldsymbol{e}_a - r, 0) - \boldsymbol{e}_o$ is used to calculate the internal reward; otherwise it is calculated for other instances by max $(\boldsymbol{e}_a - \boldsymbol{e}_o, 0)$. The output of the BEL model is computed by Eq. (10) (Morén & Balkenius, 2000):

$$e = \boldsymbol{e}_a - \boldsymbol{e}_o \tag{10}$$

Proposed model

Fig. 3 presents the proposed model, inspired by the limbic system, which is the main part of the human auditory pathway for emotion recognition. The input of this model is a feature vector extracted from the emotional speech signal. This figure shows all connections in each part of the limbic system.

According to Fig. 3, the limbic system, in the proposed model, has four main regions: the thalamus, sensory cortex, orbitofrontal cortex, and amygdala. These regions have an important role in emotional learning. As shown in this figure, the input of the system is the feature vector (s') extracted from the speech signal. The feature vector (s') is sent to the thalamus. The thalamus is connected to the amygdala and the sensory cortex. According to the figure, the thalamus produces two different outputs. The first output of the thalamus is the same feature vector. The second output of the thalamus, a_{th}, is the maximum value of the feature vector (Eq. (11)).

$$\boldsymbol{a}_{th} = [\max(\boldsymbol{s}_i)] \tag{11}$$

The sensory cortex analyzes and provides $\boldsymbol{s}$ and then distributes it between the amygdala and the orbitofrontal cortex. (Lotfi & Akbarzadeh, 2013a, 2013b; Morén & Balkenius, 2000; Parsapoor & Bilstrup, 2012). $\boldsymbol{a}_{th}$ and $\boldsymbol{s}$ are two inputs of the amygdala. According to Fig. 3, the output of the amygdala, $\boldsymbol{e}_a$, is calculated using Eq. (12).

$$\boldsymbol{e}_a = F_{ANFIS}([\boldsymbol{s}, \boldsymbol{a}_{th}]) \tag{12}$$

The function of F_{ANFIS} is explained in Section 'Adaptive neural-fuzzy model of amygdala and orbitofrontal cortex'. There is a dual connection between the amygdala and orbitofrontal cortex. In Fig. 3, the input of orbitofrontal cortex is $\boldsymbol{s}$. The output of this part, ($\boldsymbol{e}_o$), is calculated by Eq. (13).

$$\boldsymbol{e}_o = F_{ANFIS}([\boldsymbol{s}]) \tag{13}$$

According to the BEL model, the reinforcement signals (r and r_0) are sent to the amygdala and the orbitofrontal cortex, respectively. Finally, the MLP network is used on $\boldsymbol{e}_a$, $\boldsymbol{e}_o$ in Eq. (14). It means that e_a and e_o are inputs of the MLP network. Finally, e is the output of the proposed model (Eq. (14)).

$$e = F_{MLP}([\boldsymbol{e}_a, \boldsymbol{e}_o]) \tag{14}$$

The system output is the emotional states, including happiness, anger and sadness.

Feature extraction

Many useful features can be extracted from the speech signals such as energy, MFCC (Mel frequency cepstral coefficients) and LPC (linear predictive coding). This set of features has important information for discriminating different types of emotions (Adell Mercado, Bonafonte Cávez, & Escudero Mancebo, 2005) (Wu, Parsons, & Narayanan, 2010). In this work, we have selected the MFCC to extract as the emotional features (Ayadi, Kamel, & Karray, 2011; Lotfi, 2013). Fig. 4 shows the MFCC features that are extracted and then used as the input of proposed model.

Adaptive Neural-fuzzy model of amygdala and orbitofrontal cortex

The ANFIS architecture contains both in artificial neural networks and fuzzy logic. So this method combines if-then rules (Sugeno type) according to the fuzzy method and the learning algorithm of the neural network (Boyacioglu & Avci, 2010). Fig. 5 shows a simple ANFIS method with a two-dimensional input vector with five layers (Boyacioglu & Avci, 2010).

According to Fig. 5, x and y are two inputs and f is an output of the ANFIS model, which is associated with the following rules:

$$\text{If x is A1 and y is B1, then } f_1 = p_1 x + q_1 y + r_1$$

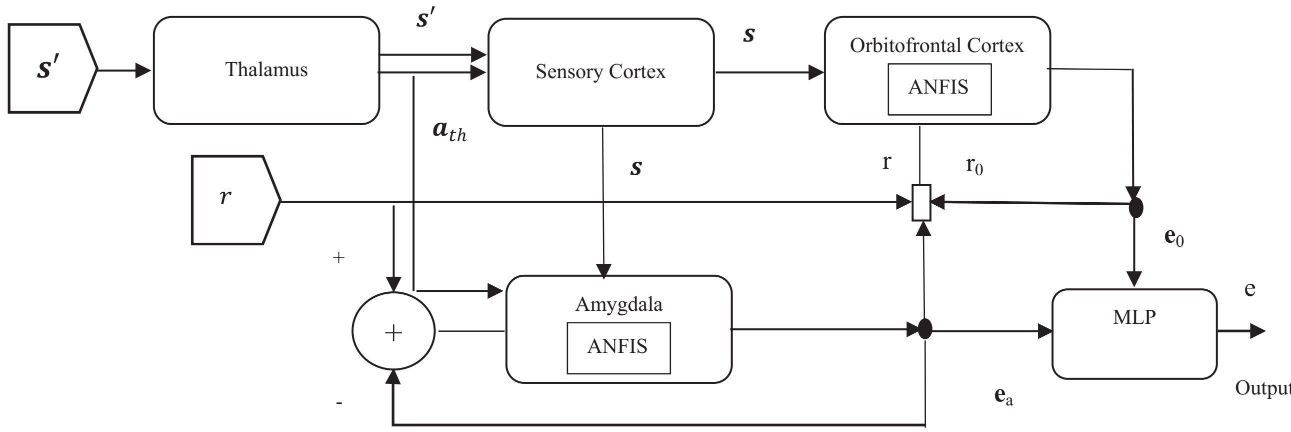

Fig. 3. The proposed BEL based model. The ANFIS model is used in the subsystems of the orbitofrontal cortex and amygdala. In addition, MLP is used to fuse the outputs of orbitofrontal cortex and amygdala.

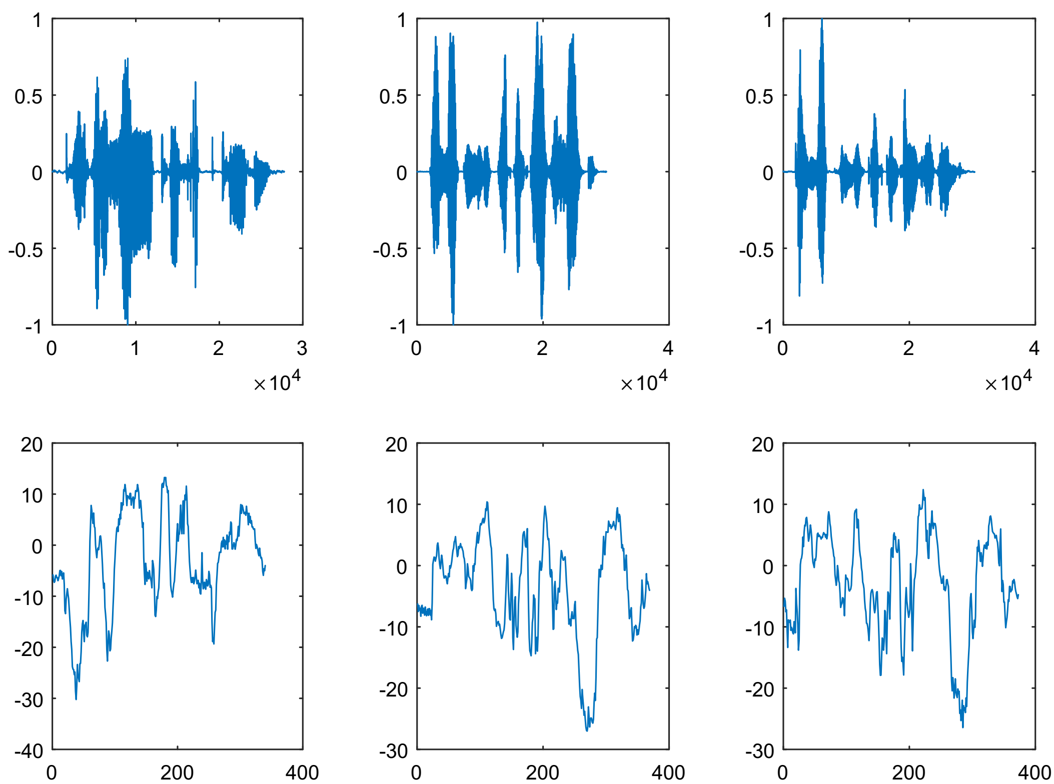

Fig. 4. The original signals and MFCC features of three different emotions: anger, happiness, and sadness.

If x is A2 and y is B2, then $f_2 = p_2x + q_2y + r_2$.

Ai, *Bi* and *fi* are fuzzy sets, and *pi*, *qi*, and *ri* are the output of the system that are obtained during the learning process.

Layer 1 (fuzzification): In the first layer, the membership grades are generated by Eqs. (15) and (16) (Boyacioglu & Avci, 2010).

$$O_{1,i} = \mu_{Ai}(x), \quad \text{for } i = 1,2 \tag{15}$$

$$O_{1,i} = \mu_{Bi-2}(y), \quad \text{for } i = 3,4 \tag{16}$$

Layer 2 (production): Each node calculates the firing strength or weights according to Eq. (17) (Boyacioglu & Avci, 2010):

$$O_{2,i} = w_i = \mu_{Ai}(x).\mu_{Bi}(y), \quad \text{for } i = 1,2 \tag{17}$$

Layer 3 (normalization): The nodes calculate the relative weights of the rules by Eq. (18). The result is a normalized firing strength (Boyacioglu & Avci, 2010).

$$O_{3,i} = \overline{W} = \frac{w_i}{w_1 + w_2} \tag{18}$$

Layer 4 (defuzzification): The nodes compute a parameter function of layer 3 output using Eq. (19) (Boyacioglu & Avci, 2010):

$$O_{4i} = \overline{w_i}f_i = \overline{w_i}(p_ix + q_iy + r_i), \quad i = 1,2 \tag{19}$$

Layer 5 (output): The summation of all incoming signals is computed by Eq. (20) (Boyacioglu & Avci, 2010):

$$O_{5i} = \overline{w_i}f_i = \frac{\sum_i w_if_i}{\sum_i w_i} \tag{20}$$

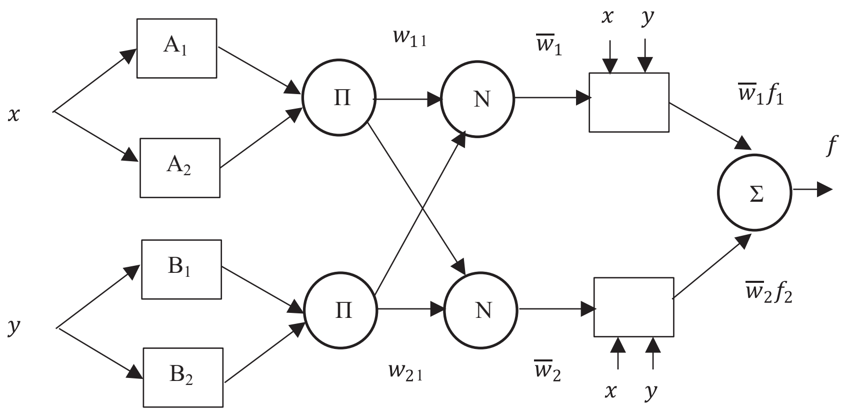

Fig. 5. Structure of the ANFIS network with two inputs of x and y.

The ANFIS method is executed on feature vectors in subsystems of the amygdala and orbitofrontal cortex.

Multi-layer perceptron (MLP)

A multi-layer perceptron (MLP) is a feedforward neural network model, and it consists of three or more layers consisting of input, output, and one or more hidden layers. In the MLP, the nodes are fully connected between layers without connections between units in the same layer. This model is a supervised learning for multilayer nodes and is performed through backpropagation that is a generalization of the least mean squares algorithm in the linear perceptron (Khanchandani & Hussain, 2009; Rosenblatt, 1962; Schuller, Rigoll, & Lang, 2004).

Experiments

Dataset

The "Berlin Dataset of Emotional Speech" was used to train and test the algorithm in this paper (Eyben, Wöllmer, & Schuller, 2010; Harimi, Shahzadi, Ahmadyfard, & Yaghmaie, 2014). The Berlin Dataset consists of 535 speech samples, consisting of German utterances related to emotions such as anger, disgust, fear, happiness, sadness, surprise, and neutrality, as performed by five male and five female voice actors. Each one of the ten professional actors expresses ten words and five sentences covering each of the emotional categories. The corpus was evaluated by 25 judges who classified each emotion with a score rate of 80%.

This Dataset was chosen for the following reasons: (i) the quality of its recording is very good and (ii) it is a public and popular Dataset of emotion recognition that is recommended in the literature (Burkhardt, Paeschke, Rolfes, Sendlmeier, & Weiss, 2005). This paper has focused on only three main emotions from the Berlin Dataset including anger, happiness, and sadness in order to achieve a higher and more accurate rate of recognition.

Competitor approaches

All models of recognition have three main steps: preprocessing, feature extraction, and classification. We have used three emotional states (happiness, anger, and sadness), on Berlin dataset, and our using simulation environment is MATLAB 2014. For each emotion we have used 62 emotional speech samples randomly divided into 70% to train and 30% for testing dataset. Then, we have extracted 12 MFCC features from all the input samples. All feature vectors in each different run feed to the proposed modified BEL. For updating the orbitofrontal and amygdala weights, we find out that 0.059 is the best value for training rate γ and β.

The suggested model is compared with machine learning models which are already studied in different human recognition problem researches including the MLP (Muda, Begam, & Elamvazuthi, 2010), ANFIS (Taleb, 2012) (Agrawal & Agrawal, 2013), support vector machine (SVM) (Schuller et al., 2004) and k-nearest neighbor (KNN) (Winter, Xu, & Lee, 2005). As well the proposed model as a cognitive approach is compared with the BEL, BELFIS, and BEL based on learning automata (BELBLA) models as they are inspired by mammalian limbic system (Morén & Balkenius, 2000). In the BELBLA, the variable structure stochastic automata (VSSA) model is used for learning parameters (γ, β) in the amygdala and orbitofrontal cortex in the BEL model (Esau, Kleinjohann, & Kleinjohann, 2003; Moarefi & Armohamadi, 2012; Narendra & Thathachar, 1974a, 1974b; Raghunathan, Schurgers, Park, & Srivastava, 2002; Thathachar & Sastry, 2002; Winter et al., 2005). Those models were applied to predict problems in the dynamic environment in emotional speech recognition.

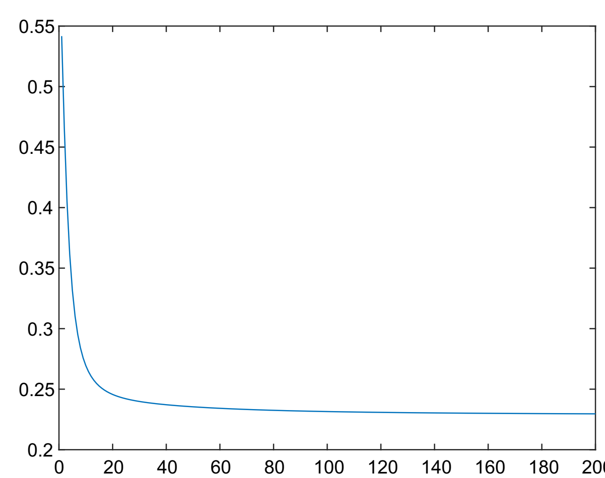

Fig. 6. Training the feed forward backpropagation of the proposed model.

Table 1
Speech emotion recognition using the proposed model in different epochs.

Recognition based the proposed model	All emotions (Happiness, Anger, Sadness)	Happiness	Anger	Sadness
Epoch = 100	69.20%	70.00%	52.50%	85.00%
Epoch = 150	70.80%	72.50%	62.50%	77.50%
Epoch = 200	72.50%	67.50%	52.50%	97.50%

Table 2
Confusion matrix of speech emotion recognition employing the proposed model.

Proposed model	Anger	Happiness	Sadness	Rate
Anger	**27**	11	1	**69.20%**
Happy	10	**21**	0	**67.70%**
Sad	3	8	**39**	**78.00%**
Precision	**67.50%**	**52.50%**	**97.50%**	Total = **72.50%**

In the confusion matrix, the main diagonal shows the correct rate of recognition.

Table 3
Speech emotion recognition using different algorithms.

Name of algorithms	All emotions (Anger, Happiness, Sadness)	Anger	Happiness	Sadness
ANFIS	51.60%	45.00%	30.00%	80.00%
MLP	53.00%	59.10%	31.80%	68.20%
BEL	53.40%	38.20%	30.00%	90.90%
BELFIS	55.00%	40.00%	37.50%	87.50%
BELBLA	68.70%	63.60%	48.50%	94.00%
KNN (k = 3)	71.30%	69.00%	55.00%	90.00%
SVM	72.33%	69.70%	56.40%	90.90%
Proposed Algorithm	72.50%	67.50%	52.50%	97.50%

Experimental results

In this study, ANFIS is used with the effective hybrid learning rule. To learn the node outputs that go forward until level 4 and the consequent parameters, the least mean squares method is used. In the backward pass, the error signals propagate backwards and premise parameters are updated by gradient decent (GD), and in MLP part, back propagation is used. This model is a supervised learning for multilayer nets.

In the recognition problems, accuracy and mean square error (MSE) are used by the feed forward back propagation. The performance measures are used to evaluate the model. In the training phase the network error at various iterations can be seen in Fig. 6.

Table 1 shows the recognition rates of the proposed model, in different epochs 100, 150 and 200 on three emotions (happiness, anger and sadness).

The training was repeated 10 times, and the average of the accuracy in the test set was recorded. Table 1 shows the rate of recognition in 100, 150, and 200 epochs obtained from the proposed model. According to Table 1 the highest rate of recognition across all emotions belongs to the proposed model by 72.5% in 200 epochs. In the proposed model of recognition, sadness has a higher recognition rate than anger and happiness.

Table 2 shows the confusion matrix of the proposed model by 200 epochs on three emotions (anger, happiness, and sadness) of the Berlin Dataset. The recognition accuracies have been evaluated by the 10-fold cross validation technique of this Dataset.

The comparison between the accuracy rate of the proposed model and different models is presented in Table 3.

To compare the results between cognitive classification models and different machine learning classifications, we applied the 12 MFCC feature extraction for different algorithms on three speech emotion (Anger, Happiness, and Sadness) signals. The training was repeated 10 times. Table 3 shows the accuracy average in the test set for different algorithms such as SVM, KNN, MLP, and ANFIS as machine learning models and the BEL, BELFIS and BELBLA as cognitive classification models. According to Table 3 the highest rate of speech emotion recognition belongs to the proposed model by 72.5% in 200 epochs. However, it could be as the result of first selecting the reward and penalty dynamically in the proposed model instead of considering them statically when reinforcement signals play an essential role in updating the amygdala and orbitofrontal weights. The second reason could be the fusion of the amygdala and orbitofrontal outputs by MLP instead of using a linear function. Then the second highest rate is obtained by the SVM. Across all models of classifications presented in Table 3, the highest rate of recognition goes to sadness emotion.

Conclusion

This paper introduces an applicable model of speech emotion recognition analyzing human emotion in such a way that improves the communication between humans and machines. Since the human brain is the main organ of the human central nervous system, the computational models inspired by humans' brain are believed to work properly. But the original BEL model, inspired by mammalian limbic system, didn't obtain good results in classifying speech emotion in this study. Therefore, in this study, different structures in amygdala and orbitofrontal cortex are examined to obtain a model with desirable performance in speech emotion recognition. The proposed model is found an applicable method to solve cognitive classification and dynamic problems. Comparing the experimental results revealed that the proposed method analyzed and recognized emotional characteristic parameters more accurately than other methods including SVM, KNN, MLP, ANFIS, BEL, BELFIS, and BELBLA. In future work, the development of the BEL model to recognize speech emotion based on the relationship between human's brain and mind would be of interest.

References

Adell Mercado, J., Bonafonte Cávez, A., & Escudero Mancebo, D. (2005). Analysis of prosodic features: Towards modelling of emotional and pragmatic attributes of speech. *SEPLN*, 277–284.

Agrawal, R. S., & Agrawal, U. N. (2013). A review on emotion recognition using hybrid classifier. In *Special issue for national conference on recent advances in technology and management for integrated growth 2013 (RATMIG 2013)*.

Ayadi, M. E., Kamel, M. S., & Karray, F. (2011). Survey on speech emotion recognition: Features, classification schemes and databases. *Pattern Recognition*, 572–587.

Boyacioglu, M. A., & Avci, D. (2010). An adaptive network-based fuzzy inference system (ANFIS) for the prediction of stock market return: The case of the Istanbul stock exchange. *Expert System's Istanbul Prediction, 37*(12), 7908–7912.

Burkhardt, F., Paeschke, A., Rolfes, M., Sendlmeier, W. F., & Weiss, B. (2005). A database of German emotional speech. *Interspeech, 5*, 1517–1520.

Esau, N., Kleinjohann, L., & Kleinjohann, B. (2003). *An adaptable fuzzy model for emotion recognition* (pp. 191–201).

Eyben, F., Wöllmer, M., & Schuller, B. (2010). Opensmile: The Munich versatile and fast open source audio feature extractor. *Proceedings of the international conference on multimedia, 10*, 1459–1462.

Grossberg, S., & Seidman, D. (2006). Neural dynamics of autistic behaviors: Cognitive, emotional, and timing substrates. *Psychological Review, 113*, 483–525.

Güler, I., & Übeyli, E. D. (2004). Application of adaptive neuro-fuzzy inference system for detection of electrocardiographic changes in patients with partial epilepsy using feature extraction. *Expert Systems with Applications, 27*, 323–330.
Hall, J. E., & Guyton, A. C. (2010). *Guyton and Hall textbook of medical physiology: Enhanced E-book*. Elsevier Health Sciences.
Han, K., Yu, D., & Tashev, I. (2014). *Speech emotion recognition using deep neural network and extreme learning machine* (pp. 223–227). Singapore, Malaysia: Inerspeech.
Harimi, A., Shahzadi, A., Ahmadyfard, A. R., & Yaghmaie, K. (2014). Classification of emotional speech spectral pattern features. *Journal of AI and Data Mining, 2*(1), 53–61.
Kamaruddin, N., & Wahab, A. (2008). Speech emotion verification system (sevs) based on mfcc for real time application. In *4th International conference on intelligent environments* (pp. 1–7).
Khanchandani, K. B., & Hussain, M. (2009). Emotion recognition using multilayer perceptron and feedforward neural network. *Journal of Scientific and Industrial Research, 68*, 367–371.
LeDoux, J. (1998). The emotional brain. *Simon & Schuster, 5*, 582–589.
Lotfi, E. (2013). Mathematical modeling of emotional brain for classification problems. *Proceedings of IAM, 2*(1), 60–71.
Lotfi, E., & Akbarzadeh, M. R. (2013a). Adaptive brain emotional decayed learning for online prediction of geomagnetic activity indices. *Neurocomputing, 126*, 188–196.
Lotfi, E., & Akbarzadeh, M. R. (2013b). Brain emotional learning based pattern recognizer. *Cybernetics and Systems, 44*(5), 402–421.
Lucas, C. (2010). Introducing BELBIC: Brain emotional learning based intelligent controller. *Integrated Systems, Design and Technology, 3*, 203–214.
Moarefi, N., & Armohamadi, A. (2012). The implementation of cellular automata with non-identical rule on serial base. *Journal of Mathematics and Computer Science, Poland, 4*(2), 264–269.
Morén, J. (2002). *Emotion and learning – A computational model of the amygdala* Ph.D. Dissertation. Lund, Sweden: Dept. of Cognitive Science, Lund University.
Morén, J., & Balkenius, C. (2000). A computational model of emotional learning in the amygdala. From animals to animats 6. In *Proceedings of the 6th international conference on the simulation of adaptive behaviour* (pp. 115–124). Cambridge, MA, USA: MIT Press.
Muda, L., Begam, M., & Elamvazuthi (2010). Voice recognition algorithms using mel frequency cepstral coefficient (MFCC) and dynamic time warping (DTW Techniques). *Journal of Computing, 2*(3), 138–143.
Narendra, K. S., & Thathachar, M. A. L. (1974a). *Learning automata. An Introduction*. Prentice Hall.
Narendra, K. S., & Thathachar, M. A. L. (1974b). Learning automata – A survey. *IEEE Transactions on Systems, Man and Cybernetics, 4*(4), 323–334.
Parsapoor, M., & Bilstrup, U. (2012). Brain emotional learning based fuzzy inference system (BELFIS) for solar activity forecasting. In *IEEE 24th international conference on tools with artificial intelligence* (Vol. 1, pp. 532–539).
Raghunathan, V., Schurgers, C., Park, S., & Srivastava, M. B. (2002). Energy-aware wireless micro sensor networks. *IEEE Signal Processing Magazine*.
R'azuri, J. G., Sundgren, D., Rahmani, R., Larsson, A., Cardenas, A. M., & Bonet, I. (2015). Speech emotion recognition in emotional feedback for human–robot interaction. *International Journal of Advanced Research in Artificial Intelligence, 4*, 20–27.
Rosenblatt, F. (1962). *Principles of neurodynamics: Perceptrons and the theory of brain mechanisms* (p. 616). Washington, DC: Spartan Books.
Schuller, B., Rigoll, G., & Lang, M. (2004). *Speech emotion recognition combining acoustic features and linguistic information in a hybrid support vector machine-belief network architecture* (vol. 1. IEEE.
Shaw, A., Vardhan, R. K., & Saxena, S. (2016). Emotion recognition and classification in speech using artificial neural networks. *International Journal of Computer Applications, 145*(8), 5–9.
Silarbi, S., Abderrahmane, B., & Phonetical, A. (2014). *Classification by adaptive network based fuzzy inference system and subtractive clustering*. Computer science & information technology (CS & IT) (pp. 187–196).
Taleb, A. (2012). Speech recognition by fuzzy-neuro ANFIS network and genetic algorithms. *International conference on intelligent computational systems (ICICS)*, 41–44.
Thathachar, M. A. L., & Sastry, P. S. (2002). Varieties of learning automata. *An Overview in IEEE Transaction on System, 32*(6), 711–722.
Watts, L. (2012). Reverse engineering the human auditory pathway. *IEEE WCCI*, 47–59.
Winter, J., Xu, Y., & Lee, W. C. (2005). Energy efficient processing of k- nearest neighbor queries in location aware sensor networks. In *The second annual international conference on mobile and ubiquitous systems*: *Networking and services, San Diego, CA* (pp. 281–292).
Wu, D., Parsons, T. D., & Narayanan, S. S. (2010). Acoustic feature analysis in speech emotion primitives estimation. *Interspeech*, 785–788.