International Journal of Innovative
Computing, Information and Control
Volume 8, Number 5(B), May 2012
ICIC International ©2012 ISSN 1349-4198
pp. 3451–3466

A BIO-INSPIRED COMPUTATIONAL NEURAL MODEL FOR ILLUSTRATION FACE AND CAR EXPERTISE EFFECT ON THE GATEWAY TO THE RIGHT FFA

Reza Ebrahimpour[1,2], Atena Sajedin[2,3], Seyed Zeinolabedin Moussavi[1]
and AbbasAli Gholamhoseini Fariz hendi[4]

[1]Brain and Intelligent Systems Research Lab
Department of Electrical and Computer Engineering
Shahid Rajaee Teacher Training University
Tehran, P.O.Box: 14115-143, Iran

[2]School of Cognitive Sciences
Institute for Research in Fundamental Sciences (IPM)
Niavaran, P.O.Box: 19395-5746, Tehran, Iran
Ebrahimpour@ipm.ir

[3]Islamic Azad University, South Tehran Branch
No. 209, North Iranshahr Street, Tehran, Iran

[4]Faculty of Educational Science and Psychology
Alzahra University
Vanak, P.O.Box:199389-1176, Tehran, Iran

Received December 2010; revised April 2011

Abstract. *The human visual system consists of a hierarchy of multiple cortical areas and it has been reported that a cortical region in the fusiform gyrus called the Fusiform Face Area, FFA, responds much more strongly to faces than to any other class of stimulus. Recent studies have also revealed that objects of visual expertise activate the FFA more strongly than non-expertise stimuli, and it was argued that the right FFA is involved in expertise-specific rather than face-specific visual processing. According to these evidences, we propose a new biologically plausible computational model to illustrate face and car expertise effect on the gateway to the right FFA. In addition, there has been reported a difference in the onset latency of macaque inferotemporal neural responses. This latter case is also considered in the proposed model, where faces are recognized in the first layer and in the second layer, a discrimination task between cars and other objects is carried out.*

Keywords: Biologically plausible computational model, Visual expertise, Gateway to the right FFA, Face perception

1. **Introduction.** The human visual system consists of a hierarchy of multiple cortical areas performing feedforward neural computation on the incoming visual signals. At the early steps, the visual cortical areas V1 and V2 perform edge and line detection. In a higher stage of processing, area V4 represents partially complex shapes with information about the structural description of the represented features [1]. In the final stage of visual processing lies the inferotemporal cortex, IT, which is thought to execute visual object recognition. Near 25% of cells in IT have been shown to respond selectively to face images, making IT the ultimate cortical machinery for performing face/nonface recognition tasks [2-4].

Consistent with prior patient research on prosopagnosia [6], in functional Magnetic Resonance Imaging, fMRI, studies, a cortical region in the fusiform gyrus called the fusiform face area, FFA, has been shown to respond much more strongly to faces than to any other class of stimulus [6,7]. These findings have led to the hypothesis that the FFA contains specialized mechanisms for face processing. There are, however, two different hypotheses for the FFA mechanisms: many researchers agree that the mechanisms involved in face processing are "special", but few agree on what exactly these mechanisms are specialized in [8-10]. According to the Face Specificity Hypothesis, cognitive and neural machinery exists that is selectively involved in the perception of faces *per se*. According to the Expertise Hypothesis, however, those mechanisms that appear to be selectively involved in face perception are in fact engaged more generally in the identification of any class of visual stimuli that share the same basic configuration and for which the subject has gained substantial visual expertise [11]. Neuropsychologists have found enhanced activity in face-selective extrastriate areas with the development of expertise for novel objects. It could be demonstrated that approximately 164 ms after presentation, objects of expertise can be differentiated neurologically from objects of lesser-known categories [12]. Car and bird expertise also recruit face-selective areas. At the spatial resolution of fMRI, the peaks of activity for faces, birds or cars in experts, could not be dissociated. Thus, the hypothesis that different expertise domains recruit the same underlying processes seems plausible. For example, activity for Greebles in the fusiform 'face area', FFA, increased with expertise [13]. In other words, the FFA may not be specific for faces *per se*, but rather only for the operations we perform typically on faces and by default when perceiving faces [10,14].

A necessary condition for becoming an expert is extensive training. The neurons needed for the fine discrimination processes have to be tuned for fast and accurate identification of individuals [15]. Changes in the timing and morphology of the ERP (Event-Related Potential) responses lead to the conclusion that cortical specificity for face-processing increases through development [16,17]. From an evolutionary standpoint, it makes sense that humans should have developed an area of the brain that is specific for the processing of faces or highly similar objects, to inaugurate specialized mechanisms, which are superior to other tasks. The ability to recognize particular faces is crucial to many aspects of social interaction, for example, to distinguish between family and strangers or friends and enemies. Xu investigated the role of the FFA in visual expertise for car and bird experts [18], reporting that the same neural mechanisms may be involved in the processing of both face and non-face expertise visual stimulus, inconsistency with the Expertise Hypothesis. The deduction is that faces and objects of expertise would selectively activate independent populations of neurons, interspersed within the same fMRI voxel. In other words, it can be thought of as a gateway, allowing only objects of visual expertise into the FFA. Thus, to facilitate the modeling of such independent populations, we need a basic-level categorization mechanism to discriminate between expertise visual stimuli faces and cars, and other objects.

The eigenfaces is well known method for face recognition. The idea of using eigenfaces was motivated by a technique developed by Sirovich and Kirby (1987) [19] for efficiently representing human faces using principal component analysis (PCA). Starting with an ensemble of different face images, they calculated the principal components of the distribution of faces, expressed in terms of eigenvectors of the covariance matrix of the distribution. Each individual face in the face set can then be approximated by a linear combination of the eigenvectors with largest eigenvalues, more commonly referred to as eigenfaces, using appropriate weights.

Turk and Pentland [20] developed the near real-time eigenfaces systems for face recognition using eigenfaces and Euclidean distance. Their method exploits the distinct nature of the weights of eigenfaces in individual face representation. Since the face reconstruction by its principal components is an approximation, a residual error is defined in the algorithm as a preliminary measure of "faceness". This residual error which they termed "distance-from-face-space" gives a good indication of face existence through the observation of global minima in a distance map.

There are two main strategies in combining classifiers: fusion and selection. In classifier fusion, it is supposed that each ensemble member is trained on the whole feature space, whereas in classifier selection, each member is assigned to learn a part of the feature space. This way, in the former strategy, the final decision is made considering the decisions of all members, while in the latter strategy, the final decision is made by aggregating the decisions of one or a few of experts [21]. Combining classifiers based on the fusion of outputs of a set of different classifiers have been developed in the field of face recognition as a method of improving the recognition performance [22]. Classifier selection has not attracted as much attention as classifier fusion. However, classifier selection is probably the better of the two strategies, if trained well. One of the most popular methods of classifier selection is the mixture of experts, ME, originally proposed by Jacobs et al. [23]. The ME models the conditional probability density of the target output by mixing the outputs from a set of local experts, each of which separately derives a conditional probability density of the target output. The outputs of expert networks are combined by a gating network which is trained to select the expert(s) that is performing the best at solving the problem [24-26].

For the first time, Ebrahimpour et al. in [27] under the classifier selection category, proposed a model, in an ME framework, for view-independent face recognition. In their model, a global eigenspace is used as the representation layer for both gating and expert networks. In order to sustain the specialization of experts, [28] employs a single-view representation scheme for the ME experts. In this work, each expert has its own eigenspace computed from the faces of a single view. According to the input layer eigenspace, each expert obtains specialization in recognizing faces of a single view. Also Ebrahimpour et al. in [29,30] propose to achieve view-independent face recognition by combining the output of experts, specialized in recognizing faces in specific view(s), in an ME framework. [27] uses single view eigenspaces to direct each expert towards a specific view, which is extended to overlapping eigenspaces in [28] to direct the experts towards a range of views. The authors try to improve the face recognition accuracy by applying TDL to ME. In their proposed method, TDL tries to direct each expert towards a single view [27] and in another work, towards a range of views [29]. Specialization of exerts in a range of views results in improved recognition performance on previously intermediate unseen views.

In this paper, we present a new and biologically plausible computational model for the Gateway to the Right FFA. According to neuro-imaging studies, a special gating mechanism exists that discriminates between faces/objects of visual expertise and other objects and allows only the face/object of visual expertise into the FFA. In agreement with recent findings on the FFA gateway mechanisms, we propose a model with a sequential architecture, which consists of constructing and training a neural network model which is mainly composed of two networks. Each network is trained on faces or cars exclusively, and in other words, each expert network has gained expertise over a specific domain: face or car. Our simulation consists of three major processing layers: Perceptual Layer, face layer and car layer. Face and car layers are composed of two sub layers: object representation and categorization. Each face and car layer is dealing with a part of input space. The face layer categorizes faces from nonfaces and then nonfaces are fed into the layer which categorizes cars from other objects.

The rest of this paper is organized as follows: Section 2 describes our proposed computational model of the gateway to FFA and its three processing layers; Section 3 applies the model to real-world data and investigating the neurophysiological results; Section 4 discusses sequential and parallel processing structures and finally Section 5 concludes and summarizes the paper.

2. **Computational Modeling of the Gateway to the Right FFA.** We propose a novel model based on biological pathways as shown in Figure 1. The first layer of the model is a set of neurons whose response properties are similar to those of complex cells in the visual cortex. A common way to model these cells in other visual recognition networks is to use the so-called "Gabor filters". Normally, these units are defined by forty Gabor different filters are derived, considering five different scales and four different orientations (0, 45, 90 and 135 degrees). Each of these filters is convolute with the input image, resulting in forty filtered copies of the input image. To encompass all the features produced by the different Gabor, the resulting Gabor wavelet features are concatenated to derive an augmented Gabor feature vector. A Gabor filter is defined as:

$$G(u, v) = \exp\left(-\frac{x^2}{2\sigma^2} - \frac{\gamma^2 y^2}{2\sigma^2}\right) \times \cos\left(2\pi \frac{x}{\lambda}\right) \tag{1}$$

where $x = u\cos\theta + v\sin\theta$, $y = -u\sin\theta + v\cos\theta$. By varying σ, λ, and the wave θ, the properties of the Gabor filter can be adjusted. We term this level the "perceptual" layer (See Figure 2 for examples of the filters at four particular orientations and five scales). The next two layers are face and car layers. From a psychological viewpoint, the face layer is first constructed and then the car layer is formed. Neuropsychologists believe that it takes a long time to become an expert. For instance, in the field of chess it takes about ten years to reach the world-class level [31]. The same is commonly said about face expertise, whether one is a child or an adult [32]. Only after a development phase of not less than 10-14 years with an ongoing cortical restructuring [33], the full expertise is ready to work properly [34]. Distinguished improvement in recognition tasks is observed between the ages of 2 and 10 years [35]. Therefore, it is justified to speak of a downright face recognition ontogeny.

1- Perceptual Layer:
Applying Gabor Filters to the stimulus
2- Face Layer:
2-1- Projecting Perceptual Layer Output into face space.
2-2- Face/ nonface Categorization.
2-3- If face then go to end; else
{
3- Car Layer:
3-1- Projecting Perceptual Layer Output into car space.
3-2- Car/ other object Categorization.
}
End.

FIGURE 1. The implementation steps to model expertise effect on the gateway to the right FFA. The proposed model is composed of three layers: perceptual layer, face and car layers.

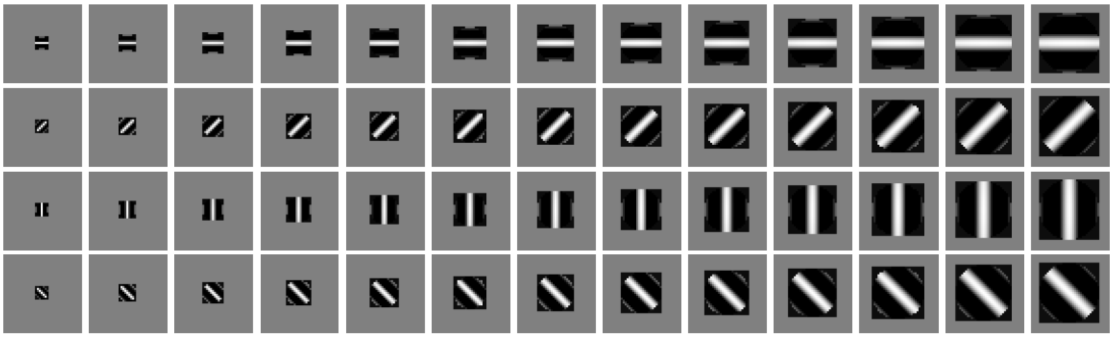

FIGURE 2. Gabor filters with circular receptive field at four different orientations and five scales

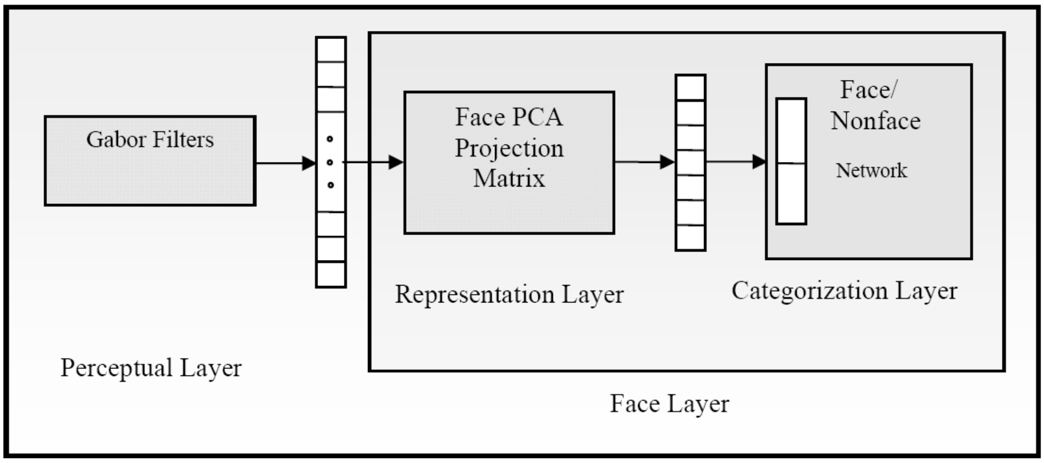

FIGURE 3. The face layer is composed of two sub-layers: representation and categorization layers. In the representation layer, PCA mapping is carried out and in the categorization layer a MLP or ME networks performs the categorization task.

It is believed that a part of this time is devoted to the formation of a gateway which discriminates between faces and nonfaces. We relate this module, the face/nonface gateway, to the face layer in our model. In other words, during the period of becoming a face critic, an experienced neuron population is generated (Figure 3). This brain development can be interpreted as a training of specialized brain areas or can be seen as a selection mechanism based on an unspecific 'resonance' between the sensorial input and the brain [36]. Due to our suggestion, in either case, this brain development corresponds to the PCA and classifier module of the model.

After obtaining face expertise as a default, in the process of forming the second layer, a cortical restructuring is carried out. It should be mentioned that the restructuring process of the second experienced neuron population takes not less than a decade. During the period of becoming a car critic, a new PCA and classifier module is generated, acting as the car gateway, which categorizes cars from other objects, excluding faces. This recent module corresponds to the car layer in the proposed model.

There is also a difference in the onset latency of inferotemporal neural responses [2] which is considered in the model. Where faces are labeled as "face" in the face layer and then in the case of nonface stimuli, the output of the perceptual layer is allowed to enter the car layer, to separate cars and other objects. Thus, faces are categorized faster than cars and other objects.

Face and car layers are feedforward network consisting of two sub-layers common to most object recognition models [3]: the first sub-layer is the "representation" layer, and the second is the "categorization" layer, Figure 3 illustrates the details of the face Layer; the car-expert layer differs in Car PCA projection matrix, "eigencars", and car/other objects network as representation and categorization layers, respectively. To avoid a high dimensional and redundant input space we employed the representation layer, which is common in both face and car layer. In the favor of an easier decision making we completely separate the input classes by extracting orthogonal dimensions from the data using PCA technique. The resulting low-dimensional object-level representation is specific to the face or car processing, as is in the population of the so-called "experienced cell" in inferior temporal cortex [11,14]. On the other hand, there are biological evidences indicating the existence of IT neurons which completely separate the input classes by extracting orthogonal dimensions from the data [37].

In the second sub layer of the face network, categorization layer, the outputs of representation layer are then categorized into "faces and nonfaces" by two network strictures a multilayer perceptron network and ME with MLP experts. For the second sub layer of the car-network, the same process is repeated, with the distinction that the discrimination task is performed between cars and other objects, excluding faces.

3. **Experimental Results.** Our experiments are divided into two parts. The first experiment investigates model implementation. Our second experiment is designed to examine the accordantly of the proposed model with the agreed neuroscientists findings in connection with face expertise efficiency in the FFA.

3.1. **Experiment I.** The experiment carried out in this work can classified into the following stages. First, eigenfaces are generated from the face image set after the Gabor filtering. Second, the face/nonface classifier is designed and trained. Third, the face/nonface classification performance is evaluated. Fourth, eigencars are generated from the car image set after the Gabor filtering. Fifth, the car/other object classifier is designed and trained. In the last step, the car/other object classification performance is evaluated.

TABLE 1. Filter specifications used to create the filters. Filter size indicates the size of the filter used for the Gabor filters.

	Size	**9 × 9**	**15 × 15**	**21 × 21**	**27 × 27**	**33 × 33**
Parameters	σ	**3.6**	**6.3**	**9.2**	**12.3**	**15.8**
	λ	**4.6**	**7.9**	**11.5**	**15.4**	**19.7**
	θ	$\mathbf{0, \pi/4, \pi/2, 3\pi/4}$				

3.1.1. *Image dataset.* We applied a set of 2085 face images to train the face layer collected from Olivetti [38], UMIST [39], Harvard [40], Yale [41], and BioID [42] databases resized to 40×40 pixels. Meanwhile we collected 2085 nonface samples from images containing landscapes, trees, buildings, cars, rocks, flowers etc. (Figure 4). After applying Gabor filters with properties as shown in Table 1, the PCA matrix has built from exclusively face images, it was used to project face and nonface samples from a 32000 dimension image space to an optimal case 50 dimension space, which was revealed from different experiments [43,44].

Figure 4. Sample images from face and nonface categories used to train the face network

Figure 5. Sample images from car and other object categories used to train the car layer

To train the car layer, we collected a set of 1450 side views of cars mainly from UIUC Image Database [45] and complementary from the various available web pages cropped and resized to 40×40 pixels. We used 1500 samples of other objects containing landscapes, trees, buildings, rocks, flowers etc. (Figure 5). After applying Gabor filter, the PCA matrix, built exclusively from car images, was used to project cars and other objects from a 32000 dimension image space to a 50 dimension eigencar space.

Moreover, the images from face, car and nonface categories used in training set has been established over a wide variety of conditions, including gray-scale images, variations in pose, lighting conditions, image cluttered scenes and varied camera-subject distances (see Figures 4 and 5).

3.1.2. *Training phase.* In the categorization layer, the classification task is handled using two neural networks structures: First a three-layer Multi-Layered Perceptrons (MLPs) network and secondly Mixture of Experts (ME). MLPs network is the most popular neural network for regression and classification. The weights of the MLPs classifier are learned using the error back-propagation algorithm [46-48]. In the basic form of ME [21], the expert and gating networks are linear classifiers, however, for more complex classification tasks, the expert and gating networks should be more complicated. For instance, [49] proposed a face/nonface recognition model, in which they use MLPs in forming the gating and expert networks to improve the face detection accuracy (details in appendix A). During our experiments, for the face layer with MLPs network the optimal structure was found to be 50 : 10 : 1 and for the car layer with MLPs network the optimal structure turned out to be 50 : 15 : 1 (see Table 2). Furthermore experiments for face layer

TABLE 2. The details of the training parameters of the MLP networks for face and car layer and its performance

	Hidden neurons number	η	*Output threshold value*	*Percentage of face layer (%)*	*Percentage of car layer (%)*
Face layer	**12**	**0.05**	**0.603**	**92.87**	–
	14	**0.05**	**0.561**	**96.29**	–
	16	**0.05**	**0.653**	**94.33**	–
	18	**0.05**	**0.439**	**95.20**	–
	20	**0.05**	**0.461**	**93.71**	–
Car layer	**12**	**0.05**	**0.589**	–	**93.13**
	14	**0.05**	**0.518**	–	**94.91**
	16	**0.05**	**0.618**	–	**95.36**
	18	**0.05**	**0.67**	–	**95.17**
	20	**0.05**	**0.613**	–	**94.42**

by combining structure, ME with MLPs experts, to find optimal topology and parameters of ME structure repeated more than 20 times. The details of the training parameters are shown in Table 3. It should be mentioned that the training parameters and output node threshold values in all experiments for the MLP and ME structures, were selected in a manner that the error on the validation set obtained minimum.

At the face categorization layer, outputs 1 and 0 were considered as faces and non-faces, respectively. Choosing an appropriate threshold value can noticeably improve the recognition rate of the model. A reliable approach to find a proper threshold value is to use "ROC curves". ROC curves reveal recognition and false rates at the same time, and hence can be utilized to define the threshold value of maximum recognition and minimum false rates. Using this approach the best threshold value, regarding the face layer output node by MLP network, was found to be 0.551 which yields a recognition rate of 96.29%. The same process was repeated for the car categorization layer, and the best threshold value was 0.618 with which a recognition rate of 95.36% is attained (see Table 2).

Whereas, in conformity with Table 3, for face layer the ME network with two experts and gating net work topologies 50 : 7 : 1 and 50 : 5 : 2, respectively and $\eta_e = 0.045$, $\eta_g = 0.015$, the best threshold value was found to be 0.469 which yields a recognition rate of 100% . Also for car layer the ME network with two experts and gating net work topologies 50 : 9 : 1 and 50 : 4 : 2, respectively and $\eta_e = 0.045$, $\eta_g = 0.005$, the best threshold value was found to be 0.49 which obtains a recognition rate of 100%.

For finding optimum training neural network parameters, the training data is often divided into training and validation sets. All of 4170 collected face and nonface images are divided in to 3000 samples for training and 1170 samples for validation. Also the 2950 collected car and non car images are divided to 1500 samples for training and 1450 samples for validation.

3.1.3. *Results.* To evaluate the model performance we used a test set of 650 face images, 500 car images and 60 images of other objects that were not present in training phase. In conformity with Figure 1, the test set images were first delivered to the perceptual layer where Gabor filtering was carried out. The outputs of the perceptual layer were projected into the "face space" through PCA mapping, and in the categorization layer faces, were distinguished from nonfaces. Those images labeled as "nonface" were then guided to the car layer; this time projected into the "car space" through PCA mapping

Table 3. The details of the training parameters of the ME with MLPs networks for face and car layer and its performance

	Gating network hidden neurons number	*Expert networks hidden neurons number*	η_e	η_g	*Output threshold value*	*Percentage of face layer (%)*	*Percentage of car layer (%)*
Face layer	**4**	**5**	**0.045**	**0.015**	**0.460**	**98.13**	–
	4	**4**	**0.040**	**0.015**	**0.463**	**97.98**	–
	5	**5**	**0.040**	**0.020**	**0.465**	**99.63**	–
	5	**4**	**0.050**	**0.010**	**0.467**	**99.22**	–
	5	**3**	**0.045**	**0.015**	**0.469**	**100**	–
Car layer	**3**	**6**	**0.040**	**0.015**	**0.471**	–	**96.47**
	4	**4**	**0.045**	**0.005**	**0.490**	–	**99.81**
	5	**5**	**0.040**	**0.015**	**0.521**	–	**98.15**
	6	**6**	**0.045**	**0.010**	**0.535**	–	**99.02**
	6	**8**	**0.040**	**0.005**	**0.540**	–	**97.62**

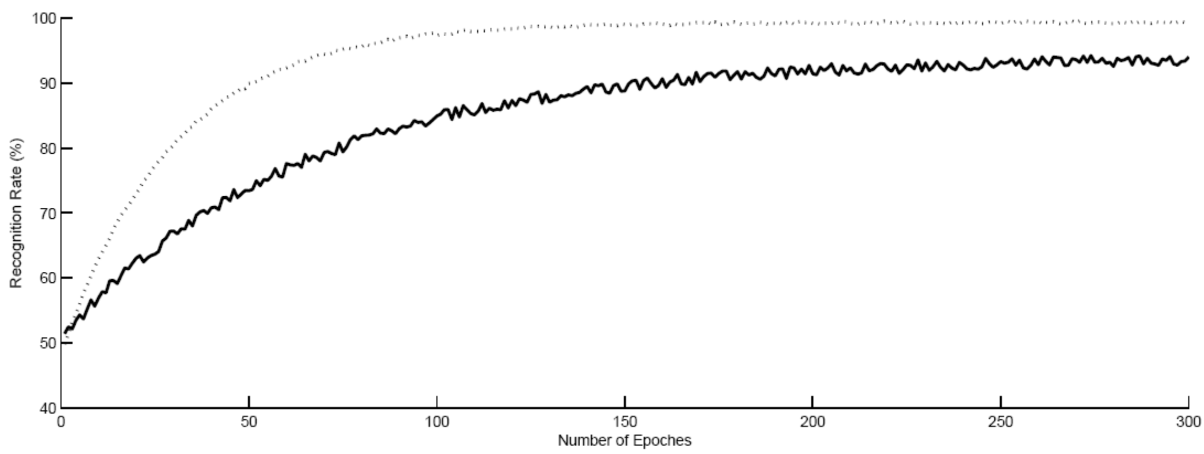

Figure 6. Recognition performance curves. The model based on mixture of experts (solid line) exhibits much better performance in comparison with The model based on MLP (dotted line).

and in the categorization layer, cars were separated from other objects. The resulting recognition rate of the model by MLPs for categorization layers on the test set was 93.87% and by using the ME structure for the face and car categorization layers performance of recognition rate on the same test set was improved to 99.33%.

It should be mentioned, the results are attained for the best structure designed in training phase on test data set. Figure 6 illustrates recognition rate of the model versus increasing the number of epochs on test data set. Consider in this figure: both speed of convergence and learning stability are more increased in the model based on ME than the same on MLP. Meanwhile, in comparison from the viewpoint of the computational load and the number of free parameters, the model based on ME had a suitable condition relative to the same on MLP. For example, the number of weights in the model based on MLP is about1500 while in the model based on ME are approximately 1100.

3.2. **Experiment II.** Several neurophysiological evidences have shown the modality of FFA activities for expertise or non expertise stimuli. This experiment is designed to examine the accordantly of the proposed model with the agreed neurophysiological findings in the human visual system.

3.2.1. *Method and material.* Faces are one of the most relevant stimulus classes in everyday life. Although faces in principle are a very homogenous visual category, adult observers are able to detect subtle differences between facial parts and their spatial relationship. Humans are able to recognize familiar faces with an accuracy of over 90%, even after fifty years. These evolutionary very adaptive abilities seem to be remarkably disrupted if faces are turned upside-down. We thus compared face/nonface recognition with upright and inverted faces. Consider the results in Figure 8: Recognizing is difficult when faces are inverted.

Next investigation which recognizing faces that are misaligned along the horizontal or vertical midline. The parts in fractured faces were displaced spatially both in the horizontal and vertical direction. As well, many parts were displaced from their original location. We wished to determine whether our model recognition would be impaired if displacements involved whole units, such as the top or bottom of the face, and if it mattered whether the displacement was horizontal or vertical (see Figure 7).

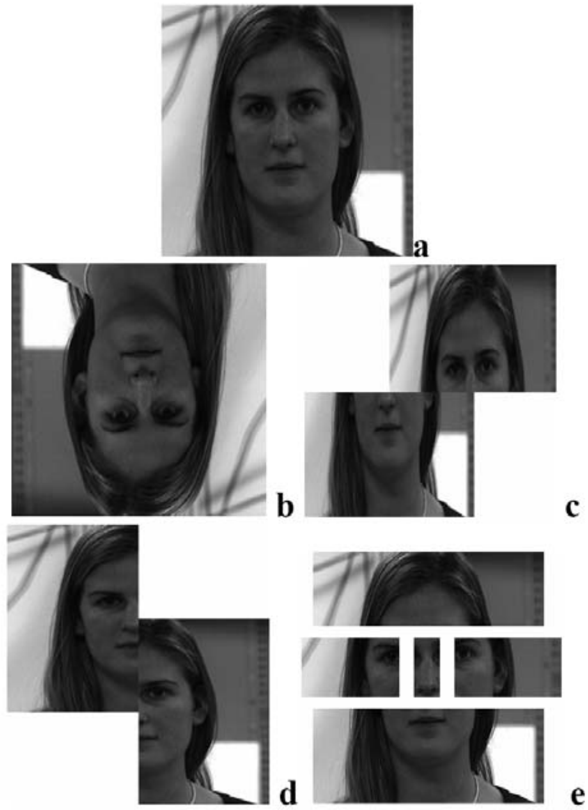

FIGURE 7. Examples of a) upright face, b) inverted face, c) misaligned along the horizontal midline, d) misaligned along the vertical midline, and e) fractured face that were used in experiment II

Consequently, our model had to identify faces that were misaligned along the vertical or horizontal midline. Our performance would provide information about the spatial relations that are crucial for face recognition consistently with neurophysiological results [50,51]. Then we altered the spatial relations among components of features without inversion. We accomplished this by cutting photos of faces of test set of face images into four or five parts and spreading them apart while retaining a semblance of the first-order relations among them. That is, the arrangement, from top to bottom, of forehead, eyes, nose, mouth and chin was preserved although the spatial distance between them was altered (see Figure 7).

3.2.2. *Results.* The proposed model was impaired at recognizing inverted faces and also those faces that were misaligned along the horizontal but not the vertical in accordance with neurophysiological findings. The drop in performance for the model from the intact condition was comparable to that observed for fractured faces, even though in this experiment most of the components retained their relation to one another. Altering the spatial relations among components had as deleterious an effect on recognition as did inversion and misalignment (see Figure 8).

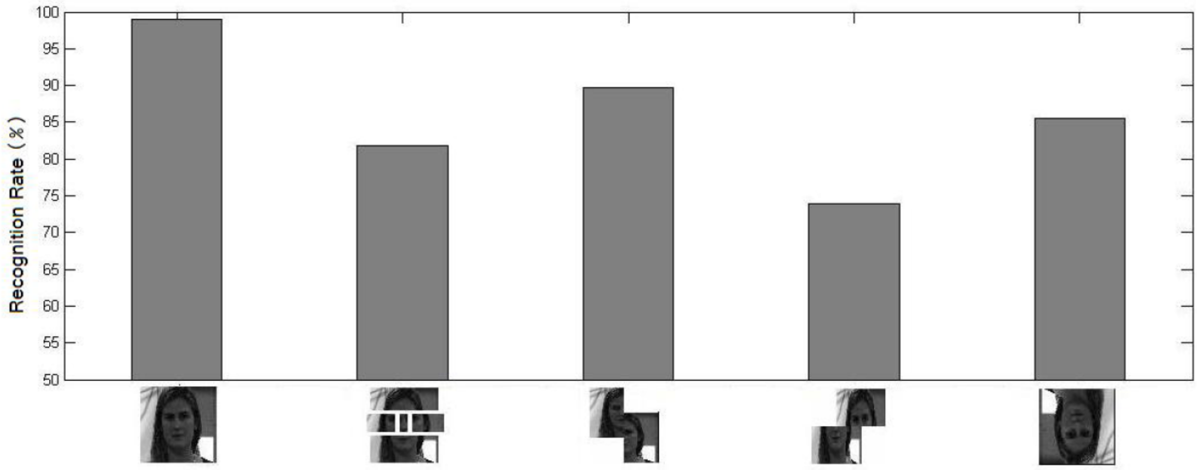

FIGURE 8. Computational modeling results for upright face, inverted face, misaligned along the horizontal midline, misaligned along the vertical midline, fractured face recognition (each result is the average of ten times run)

4. **The Proposed Sequential Method vs. Parallel Method.** We now proceed to outline the strength and weakness of our sequential model in comparison with an alternative parallel model. We first describe the sequential model preferences:

1. The sequential structure of the proposed model is strongly supported by biological evidences. As mentioned before, a recent study on macaque inferotemporal cortex has reported a difference in the onset latency of neural responses to primate and non-primate faces [2]. This latency can be modeled in a sequential structure, where faces are recognized in the first layer, and then in the second layer, with latency, another visually experienced object is processed.
2. From a computational viewpoint, the sequential model is more laborsaving. To clarify this, consider a face image at the input: the sequential model projects the input image into the face space in its first layer, if it is correctly recognized; the task is finished without going through car categorization. However, in a parallel structure both face and car processes are carried out and then a decision is made about the entity of the input image. In this latter case, the parallel processing of the input image increases the computational complexity.

On the other side, a parallel model has some advantages:

1. The decision made by a parallel system about the entity of an input image is more reliable than a sequential system, as the decision is made through a comparison between the outputs of, say, two networks. Whereas in a sequential system, the output of the first layer is compared with a fix threshold and the decision is made. Now if the system makes a wrong decision at this step, which might be caused by an inappropriate threshold, the output will certainly be mistaken. Note that this kind

of problem in a sequential system can be avoided to a large extent by choosing an appropriate threshold. To maximize the recognition performance of the model, we chose an appropriate threshold for each layer with the aid of ROC curves.

2. A parallel model is faster than a sequential model in responding to a nonface stimulus. In a sequential model, after labeling an input image as a nonface, it is guided to the car layer to see whether the input image is a car or not. But in a parallel model, both face and car processes are executed at the same time, so it is faster to find that the nonface input image is a car or not. Although the omission of the time latency makes the parallel model faster, but at the same time makes it more biologically implausible [2]. As our purpose was to model expertise effect on the gateway to the Right FFA, we chose to base our model on a sequential structure.

5. **Conclusion.** In this paper we proposed a bio-inspired computational model to explicate how faces and cars are allowed to enter the FFA. In accordance with neuropsychological evidences, after a development phase of not less than a decade, with an ongoing cortical restructuring, the full expertise is ready to work properly. After a period of at least 10 years, second expertise comes to existence and another cortical restructuring is carried out. As face expertise is a default cortical restructuring, in our model we first arranged face layer. Inside the face layer, a face PCA and classifier module was utilized to handle face/nonface discrimination task. In the next layer of our model, car experience is considered to be the second expertise. The car layer is composed of a car PCA and classifier module to discriminate between cars and other objects. The PCA and classifier modules, in critic layers, represent the cortical restructurings caused by expertise. Another noticeable point of the proposed model is the sequential processing structure supported by biological evidences reporting a difference in the onset latency of macaque inferotemporal neural responses.

REFERENCES

[1] A. Pasupathy and C. E. Connor, Population coding of shape in area V4, *Nature Neuroscience*, vol.5, no.12, pp.1252-1254, 2002.

[2] R. Kiani, H. Esteky and K. Tanaka, Differences in onset latency of macaque inferotemporal neural responses to primate and non-primate faces, *Journal of Neurophysiology*, vol.94, no.2, pp.1587-1596, 2005.

[3] M. Riesenhuber and T. Poggio, Models of object recognition, *Nature Neuroscience*, vol.3, pp.1199-1204, 2005.

[4] K. Tanaka, Inferotemporal cortex and object vision, *Annual Review Neuroscience*, vol.19, pp.109-139, 1996.

[5] T. Shimada, T. Fukami and Y. Saito, Brain activation in response to face images with backgrounds, *International Journal of Innovative Computing, Information and Control*, vol.5, no.3, pp.589-600, 2009.

[6] J. Sergent and J. L. Signoret, Varieties of functional deficits in prosopagnosia, *Cerebral Cortex*, vol.2, pp.375-388, 1992.

[7] N. G. Kanwisher, J. McDermott and M. M. Chun, The fusiform face area: A module in human extra striate cortex specialized for face perception, *Journal of Neuroscience*, vol.17, pp.4302-4311, 1997.

[8] N. Kanwisher, Domain specificity in face perception, *Nature Neuroscience*, vol.3, pp.759-763, 2002.

[9] B. Rossion, T. Curran and I. Gauthier, A defense of the subordinate-level expertise account for the N170 component, *Cognition*, vol.85, pp.189-196, 2002.

[10] M. J. Tarr and I. Gauthier, FFA: A flexible fusiform area for subordinate-level visual processing automatized by expertise, *Nature Neuroscience*, vol.3, pp.764-769, 2000.

[11] I. Gauthier, P. Skudlarski, J. C. Gore and A. W. Anderson, Expertise for cars and birds recruits brain areas involved in face recognition, *Nature Neuroscience*, vol.3, pp.191-197, 2000.

[12] J. W. Tanaka and T. Curran, A neural basis for expert object recognition, *Psychological Science*, vol.12, no.1, pp.43-47, 2001.

[13] I. Gauthier and M. J. Tarr, Unraveling mechanisms for expert object recognition: Bridging brain activity and behavior, *Journal of Experimental Psychology: Human Perception and Psychophysics*, vol.28, no.2, pp.431-446, 2002.
[14] R. Williams, M. Gugin and I. Gauthier, Perceptual expertise with objects predicts another hallmark of face perception, *Journal of Vision*, vol.10, pp.1-12, 2010.
[15] R. Williams, I. Gauthier and M. J. Tarr, Feature learning during the acquisition of perceptual expertise commentary on Schyns, Goldstone & Thibault: The development of features in object concepts, *Behavioral and Brain Sciences*, vol.21, no.1, pp.40-41, 1998.
[16] C. Bruce, R. Desimone and C. Gross, Visual properties of neurons in a polysensory area in the superior temporal sulcus of the macaque, *Journal of Neurophysiology*, vol.46, pp.369-384, 1981.
[17] L. Mi, X. Luo and F. Ren, An ERP research on Chinese Japanese learners' processing of Japanese Kanji and sentences, *International Journal of Innovative Computing, Information and Control*, vol.6, no.3, pp.1257-1270, 2010.
[18] Y. Xu, Revisiting the role of the fusiform and occipital face areas in visual expertise, *Cereb Cortex*, vol.15, pp.1234-1242, 2005.
[19] L. Sirovich and M. Kirby, Low-dimensional procedure for the characterization of human faces, *Journal of the Optical Society of America*, vol.4, pp.519-524, 1987.
[20] M. Turk and A. Pentland, Eigenfaces for recognition, *Journal of Cognitive Neuroscience*, vol.3, no.1, pp.71-86, 1991.
[21] L. Kuncheva, *Combining Pattern Classifiers: Methods and Algorithms*, Wiley, New York, 2004.
[22] J. Kittler, M. Hatef, R. P. W. Duin and J. Matas, On combining classifiers, *IEEE Trans. Pattern Anal. Mach. Intell.*, vol.20, no.3, pp.226-239, 1998.
[23] R. A. Jacobs, M. I. Jordan, S. E. Nowlan and G. E. Hinton, Adaptive mixture of experts, *Neural Computation*, vol.3, pp.79-87, 1991.
[24] K. Chen, L. Xu and H. Chi, Improved learning algorithms for mixture of experts in multiclass classification, *Neural Networks*, vol.12, no.9, pp.1229-1252, 1999.
[25] X. Hong and C. J. Harris, A mixture of experts network structure construction algorithm for modeling and control, *Applied Intelligence*, vol.16, no.1, pp.59-69, 2002.
[26] I. Guler and E. D. Ubeyli, A modified mixture of experts network structure for ECG beats classification with diverse features, *Engineering Applications of Artificial Intelligence*, vol.18, pp.845-856, 2005.
[27] R. Ebrahimpour, E. Kabir and M. R. Yousefi, Teacher-directed learning with mixture of experts for view- independent face recognition, *Lecture Notes in Computer Science*, vol.4362, pp.601-611, 2007.
[28] R. Ebrahimpour, E. Kabir and M. Yousefi, Improving mixture of experts for view-independent face recognition using teacher-directed learning, *Machine Vision and Applications*, vol.22, no.2, pp.421-432, 2011.
[29] R. Ebrahimpour, E. Kabir and M. Yousefi, Teacher-directed learning in view-independent face recognition with mixture of experts using single-view eigenspaces, *Journal of the Franklin Institute*, vol.345, pp.87-101, 2008.
[30] R. Ebrahimpour, E. kabir and M. Yousefi, View-independent face recognition with mixture of experts, *Neurocomputing*, vol.71, pp.1103-1107, 2008.
[31] W. G. Chase and H. A. Simon, Perception in chess, *Cognitive Psychology*, vol.4, pp.55-81, 1973.
[32] S. Carey and R. Diamond, Are faces perceived as configurations more by adults than by children? *Visual Cognition*, vol.1, no.2/3, pp.253-274, 1994.
[33] J. C. Bartlett and J. Searcy, Inversion and configuration of faces, *Cognitive Psychology*, vol.25, pp.281-316, 1993.
[34] R. Le Grand, C. J. Mondloch, D. Maurer and H. P. Brent, Early visual experience and face processing, *Nature*, vol.410, pp.890, 2001.
[35] R. Diamond and S. Carey, Developmental changes in the representation of faces, *Journal of Experimental Child Psychology*, vol.23, pp.1-22, 1977.
[36] R. Pallbo, Mind as evolution and evolution as such, in *The Evolution of Complexity*, F. Heylighen, J. Bollen and A. Riegler (eds.), Kluwer Academic, Dordrecht, 1999.
[37] G. Kayaert, I. Biederman and R. Vogels, Representation of regular and irregular shapes in macaque inferotemporal cortex, *Cerebral Cortex*, vol.15, pp.1308-1321, 2005.
[38] F. S. Samaria, *Face Recognition Using Hidden Markov Models*, Ph.D. Thesis, University of Cambridge, 1994.

[39] D. B. Graham and N. M. Allinson, Characterizing virtual eigensignatures for general purpose face recognition, in *Face Recognition: From Theory to Applications*, H. Wechsler, P. J. Phillips, V. Bruce, F. Fogelman-Soulie and T. S. Huang (eds.), Springer-Verlag, Berlin, 1998.

[40] M. Lades, J. C. Vorruggen, J. Buhmann, J. Lange, C. von der Malsurg, R. P. Wurtz and W. Konen, Distortion invariant object recognition in the dynamic link architecture, *IEEE Trans. Comput.*, vol.42, no.3, pp.300-311, 1993.

[41] P. Belhumeur, J. Hespanha and D. Kriegman, Eigenfaces vs. fisherfaces: Recognition using class specific linear projection, *IEEE Trans. Pattern Anal. Mach. Intell.*, vol.19, pp.711-720, 1997.

[42] *The Bio ID Face Database*, http://www.bioid.com/research/index.htm, BioID-Technology Research, 2001.

[43] R. Ebrahimpour, S. R. Ehteram and E. Kabir, Face recognition by multiple classifiers, a divide-and-conquer approach, *Lecture Notes in Artificial Intelligence*, vol.3683, pp.225-232, 2005.

[44] C. Chang and H. Hsu, Application of principle component analysis to a radial-basis function committee machine for face recognition, *International Journal of Innovative Computing, Information and Control*, vol.5, no.11, pp.4145-4154, 2009.

[45] S. Agarwal and D. Roth, Learning a sparse representation for object detection, *Proc. of European Conference on Computer Vision*, 2002.

[46] D. E. Rumelhart, G. E. Hinton and R. J. Williams, Learning representations by back-propagation error, *Nature*, vol.323, no.9, pp.533-536, 1982.

[47] M. R. Nasiri and R. Ebrahimpour, In-service video quality measurements in optical fiber links based on neural network, *Neural Network World*, vol.16, no.5, pp.457-468, 2007.

[48] H. Wu, C. Hsu, T. Lee and F. Fang, Improved SVM and ANN in incipient fault diagnosis of power transformer using clonal selection algorithms, *International Journal of Innovative Computing, Information and Control*, vol.5, no.7, pp.1959-1974, 2009.

[49] R. Ebrahimpour, E. Kabir, H. Esteky and M. R. Yousefi, A mixture of multilayer perceptron experts network for modeling face/nonface recognition in cortical face processing regions, *Intelligent Automation and Soft Computing*, vol.14, no.2, pp.145-156, 2008.

[50] M. Moscovitch, G. Winocur and M. Behrmann, What is special about face recognition? Nineteen experiments on a person with visual object agnosia and dyslexia but normal face recognition, *Journal of Cognitive Neuroscience*, vol.9, pp.555-604, 1997.

[51] K. M. Curby and I. Gauthier, The temporal advantage for individuating objects of expertise: Perceptual expertise is an early riser, *Journal of Vision*, vol.12, pp.1-13, 2009.

[52] M. N. Dailey and G. W. Cottrell, Organization of face and object recognition in modular neural networks, *Neural Networks*, vol.12, pp.1053-1073, 1999.

Appendix A.

Mixture of Experts. In a revised version of mixture of experts model, to improve the performance of the expert networks, we use MLPs instead of linear networks or experts in Figure 9. The application of MLPs in the structure of expert networks calls for a revision in the learning algorithm. In order to match the gating and expert networks, the learning algorithm is corrected by using an estimation of the posterior probability of the generation of the desired output by each expert. Using this new learning method, the MLP expert networks' weights are updated on the basis of those estimations and this procedure is repeated for the training data set. It should be mentioned that we do not use the notation of [36] to formulize the learning rules of the modified ME, but we follow the one which is described in Appendix A of [52], since its clear explanation of learning rules makes its extension easier for our purpose.

Each expert is a feedforward network and all experts receive the same input and have the same number of outputs. The gating network is also feedforward, and typically receives the same input as the expert networks. The experts compete to learn the training patterns and the gating network mediates the competition. The gating network is simultaneously trained to combine the experts' outputs. Each expert is an MLP network with one hidden layer that computes an output O_i as a function of the input stimuli vector, x, and a set of weights of hidden and output layers and a sigmoid activation function. We assume that

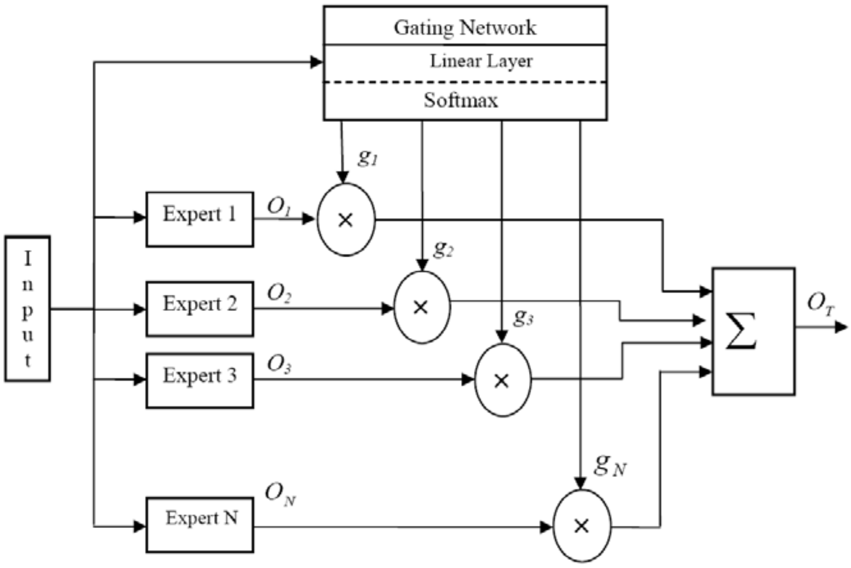

FIGURE 9. The mixture of experts is composed of expert networks and a gating network

each expert specializes in a different area of the input space. The gating network assigns a weight g_i to each of the experts' outputs, O_i. The gating network determines the g_i as a function of the input vector x and a set of parameters such as weights of its hidden and output layers and a sigmoid activation function. The g_i can be interpreted as estimates of the prior probability that expert i can generate the desired output y. The gating network is composed of two layers: the first layer is an MLP network, and the second layer is a softmax nonlinear operator. Thus the gating network computes O_g, which is the output of the MLP layer of the gating network, then applies the softmax function to get:

$$g_i = \frac{\exp(O_{g_i})}{\sum\limits_{j=1}^{N} \exp(O_{g_i})}, \quad i = 1, \dots, N \tag{A1}$$

where N is the number of expert networks. So the g_i are nonnegative and sum to 1. The final mixed output of the entire network is:

$$O_T = \sum_{i=1}^{N} g_i \cdot O_i, \quad i = 1, \dots, N \tag{A2}$$

The weights of MLPs are learned using the error back propagation, BP, algorithm. For each expert i and the gating network, the weights are updated according to the following equations:

$$\Delta w_y = \eta_e h_i (y - O_i)(O_i(1 - O_i))O_{hi}^T \tag{A3}$$

$$\Delta w_h = \eta_e h_i w_y^T (y - O_i)(O_i(1 - O_i))O_{hi}(1 - Oh_i)x_i \tag{A4}$$

$$\Delta w_{yg} = \eta_g (h - g)(O_g(1 - O_g))O_{hg}^T \tag{A5}$$

$$\Delta w_{hg} = \eta_g w_{yg}^T (h - g)(O_g(1 - O_g))O_{hg}(1 - O_{hg})x_i \tag{A6}$$

where η_g and η_e are learning rates for the expert and the gating networks, respectively. w_h and w_y are the weights of input to hidden and hidden to output layer, respectively, for experts and w_{hg} and w_{yg} are the weights of input to hidden and hidden to output layer, respectively, for the gating network. O_{hi}^T and O_{hg}^T are the transpose of O_{hi} and O_{hg}, the outputs of the hidden layer of expert and gating networks, respectively. h_i is an estimate of the posterior probability that expert i can generate the desired output y:

$$h_i = \frac{g_i \exp\left(-\frac{1}{2}(y - O_i)^T (y - O_i)\right)}{\sum_j g_j \left(-\frac{1}{2}(y - O_j)^T (y - O_j)\right)} \tag{A7}$$

As pointed out by [52], in the network's learning process, "the expert networks "compete" for each input pattern, while the gate network rewards the winner of each competition with stronger error feedback signals. Thus, over time, the gate partitions the input space in response to the expert's performance".