

A Computational Model of the Development of Hemispheric Asymmetry of Face Processing

Panqu Wang (pawang@ucsd.edu)

Department of Electrical and Computer Engineering, University of California San Diego
9500 Gilman Dr 0407, La Jolla, CA 92093 USA

Garrison Cottrell (gary@ucsd.edu)

Department of Computer Science and Engineering, University of California San Diego
9500 Gilman Dr 0404, La Jolla, CA 92093 USA

Abstract

Extensive research effort has been invested in building neurocomputational models for face and object recognition. However, the relationship between the recognition model and the development of the visual system is rarely considered. Research on the development of contrast sensitivity shows that human infants can only perceive low spatial frequency information from visual stimuli, but their acuity improves gradually with age. Also, the right hemisphere (RH) develops earlier than the left hemisphere (LH), and is dominant in infants. Here we show that these constraints, coupled with a desire on the part of the infant to individuate its caretakers and family, leads naturally to the right hemisphere bias for face processing. We propose a developmental model for face and object recognition using a modular neural network based on Dailey and Cottrell (1999). This neural network represents the two hemispheres using two modules, with a competitive relationship between them mediated by a gating mechanism. The strong RH and low spatial frequency bias for face recognition emerges naturally in the model from the interaction of the slow development of acuity and the early dominance of the right hemisphere. Remarkably, this strong asymmetry does not appear to hold for the other object categories that we tried.

Keywords: face recognition; developmental model; contrast sensitivity; modular neural network.

Introduction

Computational models have been used extensively to instantiate hypotheses regarding face and object perception from a neurocomputational perspective (Dailey & Cottrell, 1999; O'Reilly & Munakata, 2000; Riesenhuber & Poggio, 1999). Due to their social importance and frequency, faces as a stimulus class have been studied most extensively. fMRI studies have shown that the fusiform face area (FFA) is activated in response to faces more than other stimuli, and this activation is found to be stronger in the right hemisphere (Kanwisher, McDermott, & Chun, 1997). This right hemisphere lateralization is supported by electrophysiological studies, which show a stronger face-specific wave 170 ms after the stimulus onset over the right hemisphere (Bentin, Allison, Puce, Perez, & McCarthy, 1996). In addition, according to neuropsychological data, the lesioning of the right FFA will result in prosopagnosia, a deficit in face recognition (Bouvier & Engel, 2006). To account for such hemispheric asymmetry, Ivry and Robertson (1998) proposed the Double Filtering by Frequency (dff) model which postulates differential frequency filtering on task-relevant frequency bands in each hemisphere, with a bias for low spatial frequency in the RH.

Hsiao, Shieh, and Cottrell (2008) implemented this DFF theory using a computational hemispheric model, and showed that the model could account for the left-side bias in face recognition.

All neurocomputational models so far investigate the effects of structural constraints, such as the visual field split, the interaction between the two hemispheres, and the localized nature of the FFA. Temporal constraints, such as the development of the brain through childhood, are rarely considered in these models. By including consideration of the developmental changes, we may generate new hypotheses concerning the lateralization of face processing.

We consider two main constraints. First, studies in human infants have shown that the right hemisphere develops its function earlier than the left hemisphere (Chiron et al., 1997). By measuring the regional cerebral blood flow (rCBF) changes at rest using single photon emission computed tomography (SPECT), Chiron et al. observed a right hemisphere predominance of the blood flow between 1 and 3 years of age, with the asymmetry shifting to the left hemisphere after 3 years.

The second constraint has to do with the development of visual acuity. Studies have shown that the contrast sensitivity function (CSF), a measure of the ability to distinguish between different levels of luminance in static images, changes radically over time. The CSF of a 3-month-old infant is over 1.0 log units lower than the adult (Peterzell, Werner, & Kaplan, 1995). By age 4, children's contrast sensitivity is still reduced by about 0.5 log units when compared to that of adults (Atkinson, Braddick, & Moar, 1995; Gwiazda, Bauer, Thorn, & Held, 1997). The studies regarding the time of contrast sensitivity maturity have obtained somewhat disparate results. They range from claims of maturity between 6 to 10 years of age (Bradley & Freeman, 1982; Mayer, 1977), to evidence of immaturity in 8-15 year old children (Arundale, 1978). More recent studies suggest that contrast sensitivity attains adult levels in all frequencies by age 7 (Ellemborg, Lewis, Liu, & Maurer, 1999), or 8 (Leat, Yadav, & Irving, 2009).

Combining these two constraints, we propose a neurocomputational developmental model for object and face recognition. The model is based on a previous study of the development of hemispheric asymmetry by Dailey and Cottrell

(1999). Dailey & Cottrell used a mixture of experts architecture (Jacobs, Jordan, & Barto, 1991), a neural network with two modules to represent the two hemispheres. The output of the modules is gated based on their contribution to the overall performance. Dailey & Cottrell also suggested that low acuity, and the need to individuate faces, would lead to right hemisphere dominance. However, in their model, they assumed the right hemisphere would receive low spatial frequencies, and the left hemisphere high spatial frequencies, which is an unrealistic assumption. Furthermore, their model did not change its visual acuity over time. Here, we model the changes in children's contrast sensitivity by low-pass filtering the data set, and gradually improving the fidelity of inputs over training iterations (i.e., as the model "ages"). To model the asymmetric developmental pattern of the brain, we give the two modules different learning rates over time. A detailed description of this methodology will be given in next section. In the result section, we will discuss the network's general performance on different objects, and the strong right hemisphere bias for face processing which emerges naturally from the interaction of the developmental trends.

Methods

We will first describe the structure of the modular neural network, and then discuss the modifications to create the development model.

The Model

Each input image goes through a two-step preprocessing stage: Gabor filtering and Principal Component Analysis (PCA). The biologically motivated 2-D Gabor filter (Daugman, 1985) is constructed by using a two-dimensional sinusoid localized by a Gaussian envelope. By tuning to particular spatial frequency and orientations, the Gabor filter magnitudes can be used to simulate the responses of complex cells in primary visual cortex (V1). Following Gabor filtering, PCA reduces the dimensionality of the information by simulating the information extraction mechanism beyond V1, up to the lateral occipital regions level. After these preprocessing steps, each image is represented by a vector of relatively low dimension to be fed into the modular neural network.

The mixture of experts architecture has been in development since 1990 (Dailey & Cottrell, 1999; Jacobs, Jordan, & Barto, 1991; Jacobs, Jordan, Nowlan, & Hinton, 1991). Our particular variant of the model is presented in Figure 1. The modular neural network has three components: two side-by-side hidden layers (the "modules"), an output layer and a gating layer. The hidden layers are used to learn features for a given task adaptively, and develop more sophisticated representations for the input stimuli; if the task is face recognition, we can consider the hidden layer as corresponding to FFA. There is one hidden layer for each of the hemispheres. There are sufficient hidden units in each network to perform the tasks. The output layer provides us with the labels of the input stimulus to perform the final object recognition task.

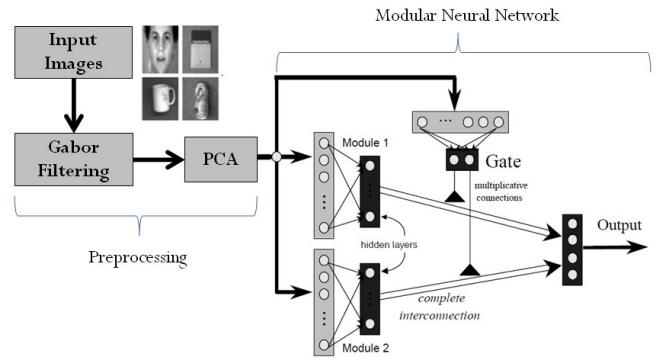


Figure 1: Architecture of Dailey and Cottrell's (1999) modular neural network model.

Because the labels to be discriminated have a strong influence on the hidden layer representation through the learning process, the discrimination level of the output layer will drive the representation developed by the hidden layer through error feedback. The gating layer has two nodes whose activity modulates the output of the hidden layers by multiplying the connection weights from the hidden layer to the output layer. We use a softmax function at the output of the gating network to ensure that the gating weights always sum to one. The network is trained using online backpropagation. During learning, the gating nodes act like a competitive selection mechanism and direct more information (error feedback) to the module that has more contribution to performance and better ability to process a given pattern. For example, if the task is better performed by module 1, the gating network will learn to weight the module 1 outputs more highly, and the hidden units of module 1 will also receive more training as a result. Thus there is a "rich get richer" effect.

Dailey and Cottrell (1999) trained this modular neural network to account for the development of the FFA. Based on the evidence that the right hemisphere has a relatively low spatial frequency preference and the left hemisphere has a relatively high spatial frequency bias (Sergent, 1982), they gave different spatial frequency information from each stimulus to each module by weighting the frequency component differently in the PCA step. When the model's goal was to individuate different faces while categorizing the other classes of stimuli, they observed a strong specialization of the low spatial frequency module to processing faces; no other tasks showed a similar specialization. Hence, they took these results as support for the hypothesis that the FFA developed as a natural consequence of the infant's low visual acuity, and the need to individuate faces.

Modeling Infant's Developmental Patterns

Although Dailey and Cottrell (1999) successfully modeled the right hemisphere lateralization of FFA for face recognition, their consideration of the developmental facts is inadequate.

quate. First, although the right hemisphere (RH) has a low spatial frequency (LSF) bias, both the left and right hemispheres should receive the same frequency information from the input stimuli, as psychophysical results show equal CSF's in the two visual fields. We hypothesize that the selectivity arises as a consequence of competition during the developmental period, rather than assuming it is already in force during development. Hence, instead of manually weighting the information provided by each spatial frequency differently in each module, in this work we give both modules the same images over time. Instead of manipulating the spatial frequencies so that they are different, we give the modules different learning rates, in accord with the data that the right hemisphere is dominant during the first three years (Chiron et al., 1997). We model this as a wave of plasticity in each hemisphere that is slightly out of phase, as in Figure 2. The right hemisphere will thus do more learning earlier than the left. This will drive the right hemisphere network to reach convergence more quickly, which in turn will make it win the competition and obtain a higher gating node value in the gating network. We used two Gaussians with different mean values and the same variance followed by a straight line to represent the evolution of the learning rate. At earlier epochs, module 1 has a higher learning rate than module 2. After a certain number of iterations, the learning rate of module 2 starts to increase and prevails over module 1. Finally, both learning rates drop to a small constant value until the end of training. This represents the maturity of both hemispheres, but they continue to learn into "adulthood."

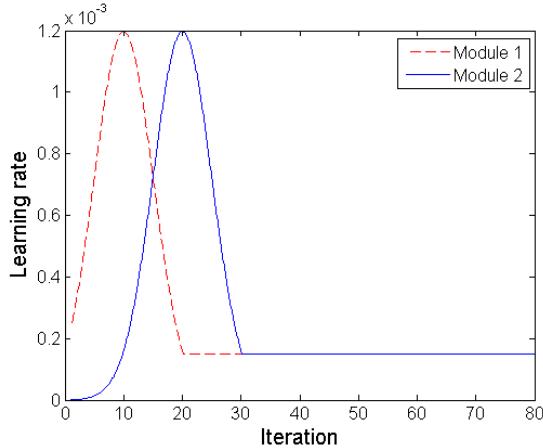


Figure 2: Learning rate example. Module 1 (dashed line) learns earlier than module 2 (solid line) and they converge to same value after 30 iterations.

Since the contrast sensitivity of infants is relatively low (Atkinson et al., 1995; Gwiadza et al., 1997; Peterzell et al., 1995), they can only obtain the low spatial frequency information of a given visual stimuli instead of receiving the full frequency details. As they grow older, their contrast sensitivity will reach adult levels, and then they will receive the full spatial frequency information from the input. To model

this change of children's contrast sensitivity over time, we low-pass filtered the dataset before sending them to the Gabor filter in the modular neural network model. We gradually improved the fidelity of the input as the training iteration increases. To be more specific, we used a 2-D circular averaging filter with decreasing radius r to filter the dataset. We set the largest radius be 6 pixels, and gradually decreased the radius by a same interval till it reached zero (unfiltered). Figure 3 shows the result of a image filtered by a disk filter with high to low radius.



Figure 3: Result of a sample image filtered by disk filter with radius decrease from 6 (leftmost) to 0 (rightmost).

Experiments and Results

Face/Object Stimuli and Preprocessing

Our studies used four category of objects: faces, books, cups, and handwritten digits. For the faces, books and cups, we collected the images from 12 different individuals for each category from the Cottrell and Metcalfe (1991) database. For digits 0-9, we utilized the MNIST handwritten database (LeCun, Bottou, Bengio, & Haffner, 1998) and sampled 20 images for each digit. Each image was transformed to grayscale and cropped to size of 64×64 . In the Gabor filtering step, we adopted the classical Gabor filter bank (Lades et al., 1993) with 5 different scales and 8 orientations ranging from 0 to $7\pi/8$. This step resulted in a 40-dimensional vector at each point in the image. After normalizing the response values across orientations and subsampling these vectors in a 8×8 grid, we computed the magnitude of the complex numbers and got a 2560-dimensional vector to represent the image. To extract a smaller number of features and maintain a segregation of responses from each scale of Gabor filter, we performed PCA separately on each spatial frequency component of the pattern vectors. Since we had 5 different scales, we extracted the subvectors corresponding to each scale from every pattern in the training set, computed the eigenvectors of the covariance matrix, and projected these subvectors onto these basis vectors. Here we used the Turk and Pentland trick (Turk & Pentland, 1991) to reduce the computational cost. We retained the eight most significant coefficients of each scale according to the eigenvalue, reassembled the pattern and finally obtained a 40-dimensional vector for each input image.

Network Training

Based on the task manipulations performed by Dailey & Cottrell, we trained the modular neural network to perform three tasks:

1. 4-Class basic-level classification on faces, books, cups and digits: 4 outputs in total.

2. Face expert: subordinate classification on 10 different faces, together with basic-level classification on books, cups and digits: 13 outputs in total.
3. Non-face expert: subordinate classification on 10 different digits (0-9), cups or books, together with basic-level classification on other 3 categories: 13 outputs in total.

For each task, we performed two experiments. As a control condition, we used the same learning rate λ for both modules. In the experimental condition, we assigned different learning rate over time to each module to model the asymmetric development pattern over 30 iterations, as in Figure 2:

$$\begin{aligned}\lambda_{\text{module1}} &= 0.015 \times \mathcal{G}(\text{iteration}, 10, 5) \\ \lambda_{\text{module2}} &= 0.015 \times \mathcal{G}(\text{iteration}, 20, 5)\end{aligned}$$

where $\mathcal{G}(\text{iteration}, 10, 5)$ means a Gaussian distribution with mean 10 and variance 5 evaluated at iteration . After 30 iterations, we set the constant learning rate $\lambda = 1.5 \times 10^{-4}$ for both modules. In all experiments, we considered module 1 to be the right hemisphere, which learns earlier, and module 2 to be the left hemisphere. To model the development of contrast sensitivity, we utilized 9 different filtered datasets from low to full spatial frequency (see Figure 3), changing the dataset every three epochs. While the mapping is arbitrary, we assume for now that three epochs correspond to a year, so that the contrast sensitivity matures around the "age" of 9, or 27 epochs. At that point, the dataset with full spatial frequency is used to train the model to convergence.

We used stochastic gradient descent (online learning) to train the neural network. We used a small momentum term, set to 0.01 in all experiments. The number of output nodes equals the number of categories to be classified. We performed a 10-fold cross validation on the training set to find the optimal settings for the number of nodes in hidden units and the stopping criteria based on the model's performance on the hold-out set. In the 4-way classification task, we used 48 images from each class to form the training set, and 12 images to form the test set. We determined that a mean squared error (MSE) of 0.001 was an adequate threshold to stop training: the value of the gating nodes was stable and the network reached good performance on the test set. For the face expert and non-face expert experiments, due to limitations on how many images we had of individual faces, we reduced each class to have 20 training images and 4 testing images, and the MSE threshold was set to 0.02 to avoid overtraining. Both hidden layers are determined to have 15 nodes as the model's performance is good and stable when we set it between 10 and 20. The connection weights between input layer, hidden layer and output layer were set randomly, while the weights between input to gating layer were all set to have the same value to give the softmax layer of the two gating nodes the same initial value of 0.5.

For each of the 3×2 experimental conditions, we ran the experiment 12 times and recorded the softmax output of the gating layer for each class after convergence. This indicated

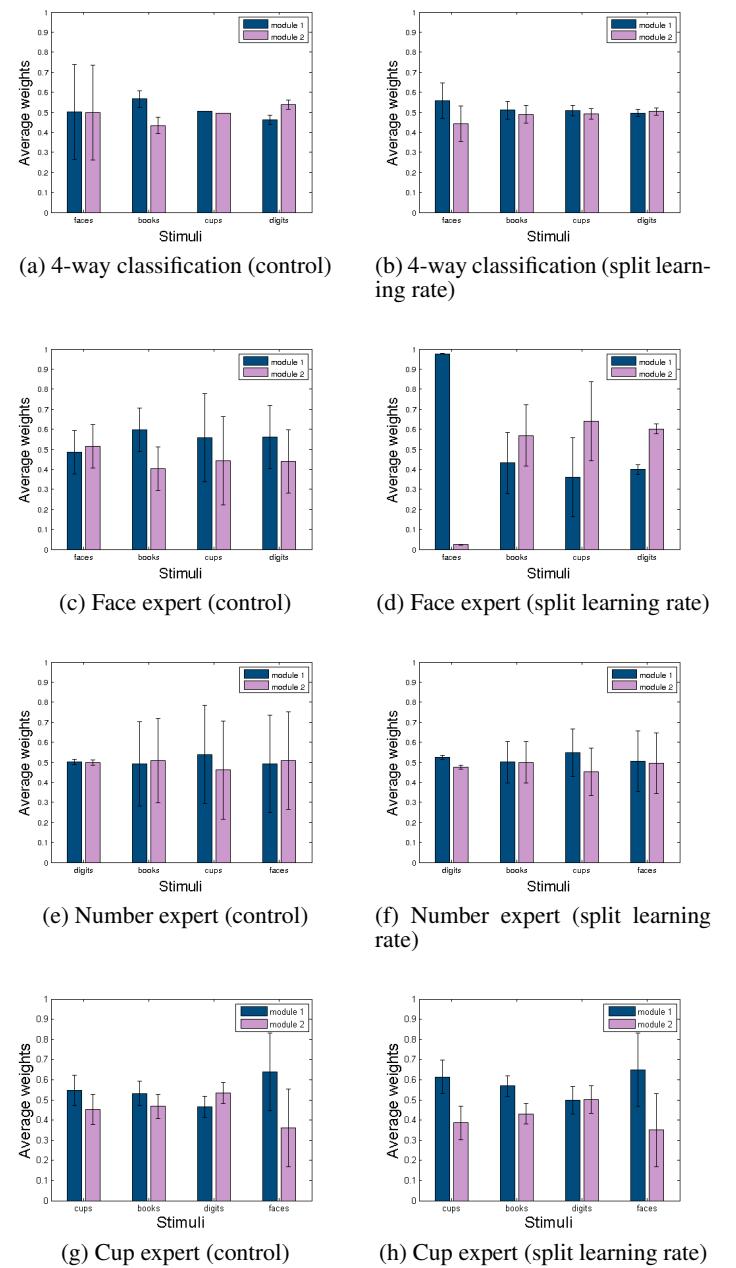


Figure 4: Average weights assigned to each module for each stimulus class. In control condition (left column), each module receives the same learning rate. We can see the average weights are almost symmetric in all tasks. In the experimental condition (right column), each module receives different learning rate. We can see a strong RH bias for face expert task, but no such bias exists in number and cup expert task. The error bar denotes the standard deviation.

the lateralization of each module for different classes of objects. Figure 4 displays the result for three of the tasks.

In the control condition, both modules receive the same learning rate, so it is expected that there is no preference for either module over any class of stimuli. Even if at certain

time one module wins for a given category, the average gating value for each module will be equal when averaged over many runs. As a result, no matter what the task is, no bias exists in the control task, so the average weights should be symmetric, as can be seen in the left column of Figure 4.

When we assigned different learning rate for each module, the task of recognizing different faces (face expert) shows a consistently strong bias for module 1 (right hemisphere). From Figure 4 (d), the average weight of RH reaches 0.96, 24 times higher than the LH weight value. However, this strong hemisphere lateralization does not occur on the task of differentiating the non-face objects, where the average weights for both hemispheres are close. We have also performed the equivalent experiment with a book expert, and this experiment only shows the weak right hemisphere bias (See Figure 5). Additionally, we have also performed the experiment using different learning rates and training epochs per dataset, and this RH lateralization for face recognition appeared very robust.

Brain Development, Contrast Sensitivity and RH Bias for Faces

The finding of a strong RH bias for face recognition directly raises two questions: at which point during the learning process (brain development) did the bias appear? What is the difference between face recognition and object recognition? To investigate the relationship between brain development and RH bias, we ran another face expert experiment, and kept track of the softmax value of the gating nodes for each class separately until they are stable. The result of gating node value vs. time is shown in Figure 5.

We observed some interesting phenomena. First, for all non-face objects, the value of the gating node for the RH increases rapidly before the 10th epoch. This is consistent with the fact that the learning rate for the RH is much higher than the LH (see Figure 2). In addition, a significant drop for RH value and an increase for LH value appear between 10-20th iteration for these same objects, when the learning rate switches to the LH. Thus these objects track the learning rate closely. However, the impact of the increased learning rate for the left hemisphere does not affect the allocation of face processing to the RH, as the gating node value for RH remains high and stable. Considering the hidden layer of the RH network to be analogous to the FFA, these findings are also consistent with the fact that the FFA is RH lateralized (Kanwisher et al., 1997). Combined with the fact that we mainly utilized the low spatial frequency (LSF) information to train the RH network before 10th iteration, and full spatial frequency thereafter, we've shown that in our model, face recognition has a strong LSF preference. The hypothesis then is that this strong LSF and RH lateralization of face recognition is a natural result of the interaction between children's brain development and contrast sensitivity maturity.

Finally, we investigated the relationship between the rate of contrast sensitivity development and RH bias for faces. By varying the step size of our circular filter to make smaller

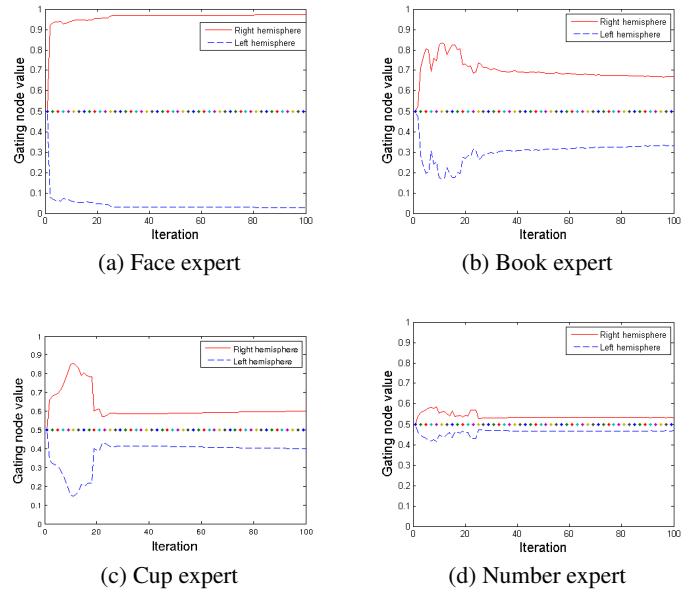


Figure 5: Gating node value vs. time for all objects in the experiments. We took the average value across all individuals in the same category to get the result for the corresponding expert tasks. The solid line represents the node value of RH and the dashed line represents the node value of LH. The dotted line set a middle threshold of 0.5. We set the maximum value at horizontal axis as 100 because the node value is stable afterwards.

or greater increments, we varied the rate of development. We built different numbers of filtered datasets ranges from 6 to 14 to test the effect on the degree of bias. All filtered datasets have the same range of filtering coefficients, and a larger number of dataset means a smaller interval step. We ran the training 12 times on each group of filtered datasets and recorded the value of RH and LH nodes when the network converges. We computed the mean value and standard deviation of both hemispheres. The results are shown in Table 1.

Table 1: Relationship between age of contrast sensitivity matures and the RH bias for face recognition

"Age"	RH Value	std(RH)	LH Value	std(LH)
6	0.7908	0.0908	0.2654	0.0968
8	0.8088	0.0269	0.1912	0.0269
10	0.8989	0.0241	0.1011	0.0333
12	0.8467	0.1042	0.1533	0.0709
14	0.8352	0.0533	0.1648	0.0525

The result shows the RH bias for face recognition is obvious regardless of rate of contrast sensitivity changes. However, we can observe a peak in the lateralization at 10 "years." This suggests that there could be an "optimal" age of CSF maturity with respect to lateralization.

Conclusion

In summary, we built a face and object recognition model, using a developmentally-inspired implementation. Our model suggests that the strong right hemisphere lateralization for face processing can arise from the interaction of two developmental facts: That the right hemisphere matures earlier than the left, and that visual acuity increases gradually over development. We should also note that, as in Dailey & Cottrell's 1999 model, there is a strong effect of task as well. Simply categorizing faces as faces does not lead to right hemisphere specialization. Hence the drive by the infant to individuate its parents, other caretakers, family and friends, leads to the right hemisphere specialization. In addition, as we observe a mild right hemisphere lateralization more or less for all expert experiments, we predict that both the task and the low spatial frequency nature of certain objects should contribute to the RH bias. Future work includes doing more analysis of the model to discover why the strong RH bias happens for faces, investigating the relationship between CS development and RH bias in more detail, and extending the model to other important classes of stimuli, such as word recognition, which shows a left hemisphere bias.

Acknowledgments

This work was supported in part by NSF grants SMA-1041755 and IIS-1219252 to G. W. Cottrell. We thank Gary's Unbelievable Research Unit (GURU) for comments on this work.

References

Arundale, K. (1978). An investigation into the variation of human contrast sensitivity with age and ocular pathology. *British Journal of Ophthalmology*, 62, 213-215.

Atkinson, J., Braddick, O., & Moar, K. (1995). Contrast sensitivity of the human infant for moving and static patterns. *Vision Research*, 35, 1045-1047.

Bentin, S., Allison, T., Puce, A., Perez, A., & McCarthy, G. (1996). Electrophysiological studies of face perception in humans. *Journal of Cognitive Neuroscience*, 8, 551-565.

Bouvier, S. E., & Engel, S. A. (2006). Behavioral deficits and cortical damage loci in cerebral achromatopsia. *Cerebral Cortex*, 16, 183-191.

Bradley, A., & Freeman, R. D. (1982). Contrast sensitivity in children. *Vision Research*, 22, 953-959.

Chiron, C., Jambaque, I., Nabbout, R., Lounes, R., Syrota, A., & Dulac, O. (1997). The right brain hemisphere is dominant in human infants. *Brain*, 120, 1057-1065.

Cottrell, G. W., & Metcalfe, J. (1991). Empath: face, gender and emotion recognition using holons. In R. P. Lippman, J. Moody, & D. S. Touretzky (Eds.), *Advances in neural information processing systems* (Vol. 3, pp. 564-571). San Mateo, CA: Morgan Kaufmann.

Dailey, M. N., & Cottrell, G. W. (1999). Organization of face and object recognition in modular neural network models. *Neural Networks*, 12, 1053-1073.

Daugman, J. G. (1985). Uncertainty relation for resolution in space, spatial frequency, and orientation optimized by two-dimensional visual cortex filters. *Journal of the Optical Society of America A*, 2, 1160-1169.

Ellemborg, D., Lewis, T. L., Liu, C. H., & Maurer, D. (1999). Development of spatial and temporal vision during childhood. *Vision Research*, 39, 2325-2333.

Gwiazda, J., Bauer, J., Thorn, F., & Held, R. (1997). Development of spatial contrast sensitivity from infancy to adulthood: psychophysical data. *Optometry and Vision Science*, 74, 785-789.

Hsiao, J. H., Shieh, D. X., & Cottrell, G. W. (2008). Convergence of the visual field split: Hemispheric modeling of face and object recognition. *Journal of Cognitive Neuroscience*, 22, 2298-2307.

Ivry, R. B., & Robertson, L. C. (1998). *The two sides of perception*. Cambridge, MA: MIT Press.

Jacobs, R. A., Jordan, M. I., & Barto, A. G. (1991). Task decomposition through competition in a modular connectionist architecture: The what and where vision tasks. *Cognitive Science*, 15, 219-250.

Jacobs, R. A., Jordan, M. I., Nowlan, S., & Hinton, G. E. (1991). Adaptive mixtures of local experts. *Neural Computation*, 3, 1-12.

Kanwisher, N., McDermott, J., & Chun, M. M. (1997). The fusiform face area: a module in human extrastriate cortex specialized for face perception. *Journal of Neuroscience*, 17, 4302-4311.

Lades, M., Vorbrüggen, J., Buhmann, J., Lange, J., Malsburg, C. von der, Witz, R. P., et al. (1993). Distortion invariant object recognition in the dynamic link architecture. *Computers, IEEE Transactions on*, 42(3), 300-311.

Leat, S. J., Yadav, N. K., & Irving, E. L. (2009). Development of visual acuity and contrast sensitivity in children. *Journal of Optometry*, 2, 19-26.

LeCun, Y., Bottou, L., Bengio, Y., & Haffner, P. (1998). Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11), 2278-2324.

Mayer, M. (1977). Development of anisotropy in late childhood. *Vision Research*, 17, 703-710.

O'Reilly, R., & Munakata, Y. (2000). *Computational explorations in cognitive neuroscience: Understanding the mind by simulating the brain*. Cambridge, MA: MIT Press.

Peterzell, D. H., Werner, J. S., & Kaplan, P. S. (1995). Individual differences in contrast sensitivity functions: longitudinal study of 4-, 6-, and 8-month-old human infants. *Vision Research*, 35, 961-979.

Riesenhuber, M., & Poggio, T. (1999). Hierarchical models of object recognition in cortex. *Nature Neuroscience*, 2, 1019-1025.

Sargent, J. (1982). The cerebral balance of power: confrontation or cooperation? *Journal of Experimental Psychology. Human Perception and Performance*, 8, 253-272.

Turk, M., & Pentland, A. (1991). Eigenfaces for recognition. *The Journal of Cognitive Neuroscience*, 3, 71-86.