Connectivity Asymmetry Can Explain Visual Hemispheric Asymmetries in Local/Global, Face, and Spatial Frequency Processing

Ben Cipollini (bcipolli@cogsci.ucsd.edu)
Department of Cognitive Science, University of California San Diego
9500 Gilman Dr 0515, La Jolla, CA 92093 USA

Janet Hsiao (jhsiao@hku.hk)
Department of Psychology, University of Hong Kong
604 Knowles Building, Pokfulam Road, Hong Kong SAR

Garrison Cottrell (gary@eng.ucsd.edu)
Department of Computer Science and Engineering, University of California San Diego
9500 Gilman Dr 0404, La Jolla, CA 92093 USA

Abstract

Left-right asymmetries have been noted in tasks requiring the classification of many types of visual stimuli, including Navon figures, spatial frequency gratings, and faces. The Double Filtering by Frequency (DFF) model (Ivry & Robertson, 1998), which postulates asymmetric frequency filtering on task-relevant frequency bands, has been used to implement computational models accounting for the human behavior in each type of study above, but does not provide a fully mechanistic account, nor does it have direct correlates in the brain. The Differential Encoding (DE) model (Hsiao, Shahbazi, & Cottrell, 2008), which postulates that a known anatomical asymmetry in patch connectivity drives visual processing asymmetries, to date has been used to account for only one dataset. Here, we implement the DE model more closely to the parameters of the published patch asymmetry, show that the DE is doing spatial frequency filtering, then show that under these conditions the DE generalizes both within the original task, but also to three of the four datasets mentioned above. Examination of the failure to model the fourth dataset suggest a novel solution and a reinterpretation of the data.

Keywords: local/global processing, left-side bias, hemispheric asymmetry, visual perception, Differential Encoding, Double Filtering by Frequency, computational model

Introduction

A large literature of experimental psychology and cognitive imaging studies has established the existence of a wide range of left-right asymmetries in the classification of many visual stimuli. A typical paradigm consists of briefly presenting a stimulus to the left or right of fixation, then requiring subjects to perform a classification task, such as whether a target stimulus was present or not. Because information from the left visual field (LVF) is initially directed exclusively to the right cerebral hemisphere (RH), and right visual field (RVF) to the left cerebral hemisphere (LH), comparisons of task performance between LVF/RH and RVF/LH can indicate asymmetries in hemispheric processing. Visual stimuli which have shown such asymmetries in these types of tasks include Navon figures (Sergent, 1982) consisting of a large, "global"-level figure composed of smaller, "local"-level figures (see Figure ~~3a)~~ **3b, not 3a**, spatial frequency gratings (Christman, Kitterle, & Hellige, 1991; Kitterle, Hellige, & Christman, 1992), and faces (Young & Bion, 1981; Brady, Campbell, & Flaherty, 2005).

Ivry and Robertson (1998) developed the Double Filtering by Frequency (DFF) theory to account ~~for a large literature of local/global processing~~ **these data** asymmetries. The computational model they implemented based on the DFF theory aimed to account for three particular experiments from the literature, thought to express core features of the ~~literature~~:

- Sergent (1982): showed the basic hemisphere $\times$ level interaction of the local/global literature, with targets presented at the smaller, "local" level responded to faster when presented to the RVF/LH, and targets presented at the larger, "global" level responded to faster when presented to the LVF/RH.
- Kitterle et al. (1992): showed that reaction times in two different classification tasks, using the same stimuli but requiring use of information at different spatial frequencies, interacted with the visual field/hemisphere of presentation. The task requiring high spatial frequency (HSF) information was responded to faster when presented to the RVF/LH, and the task requiring low spatial frequency (LSF) information was responded to faster when presented to the LVF/RH.
- Christman et al. (1991): showed that discrimination between two stimuli which differed by a single spatial frequency interacted with the visual field/hemisphere of presentation, based on the *relative* frequency of the discriminative spatial frequency compared to the rest of the spatial frequencies contained in the stimuli. When the discriminative frequency was higher than the frequency content of the rest of the stimulus, responses were faster for presentation to the RVF/LH; when the discriminative frequency was lower than the frequency content of the rest of the stimulus, responses were faster for presentation to the LVF/RH.

The DFF model replicated the core features of each of the above studies. Later, Hsiao, Shieh, and Cottrell (2008) showed that the DFF theory could also account for the so-called 'left-side bias'–the tendency for people to associate

face identity with the right-side of a person's face (appearing in the LVF of the viewer) (Brady et al., 2005).

The DFF model, while it accounts for all these data, requires the modeler to input the frequency range of interest for the particular task being modeled, rather than discovering the task-relevant frequency range through training. In addition, no neurophysiological evidence has been found for spatial frequency filtering in cortex.

Hsiao, Shahbazi, and Cottrell (2008) took a very different approach to the problem of explaining visual processing asymmetries. Rather than starting with a theory of the algorithms behind the asymmetries (Marr's "algorithmic" level), they created a model of an anatomical asymmetry (Marr's "implementation" level) and asked whether it could account for the asymmetries in classification of visual stimuli observed in behavioral studies. They used an asymmetry in inter-patch connectivity (Galuske, Schlote, Bratzke, & Singer, 2000), one of the few known network-level asymmetries, and the only ~~which~~ has been suggested to be related to local/global processing asymmetry (Galuske et al., 2000; Hutsler & Galuske, 2003). This asymmetry was found in BA22, an auditory association area; previous researchers had suggested that local/global processing differences would occur beyond early primary sensory areas (Sergent, 1982; Ivry & Robertson, 1998), lending support for its use as the basis of their model.

one

that

Hsiao et al used this connectivity pattern as inspiration to implement a simple "autoencoder" neural network (see Methods section for details on this model) and tested to see whether this anatomical asymmetry could account for a subset of the data modeled by Ivry and Robertson. Using precisely the same small, 1D inputs that Ivry and Robertson created for modeling a reduced version of Sergent's study, Hsiao et al's "Differential Encoding" (DE) model showed a hemisphere × level interaction that matched Sergent's human data more closely than that of Ivry and Robertson's DFF model. Hsiao et al then constructed realistic 2D bitmaps of Sergent's Navon stimuli, trained 2D versions of the DE models on these realistic stimuli, and again showed a hemisphere × level interaction that was a better quantitative match to the original human data than the results published by Ivry and Robertson using their DFF model.

et al ->

et al.

it

1-dimensional (1D)

The DE model was able to directly address issues with the DFF model by implementing an anatomical hemispheric asymmetry. ~~The use of a secondary network for classification does not manipulate the inputs with hand selected weights, allowing task relevant information to be selected by the learning algorithm, rather than manually selected by the modeler.~~

However, the DE did not address most of the data accounted for by the DFF, including two local/global studies (Kitterle et al., 1992; Christman et al., 1991), one face-processing study (Young & Bion, 1981), and the relationship between local/global processing and spatial frequency processing. In addition, the DE model used parameters for number of connections and distribution shape that were different from the parameters reported in the literature.

Here, we present a version of the DE model which is closer to the original anatomy. We show that this model accounts for three of the four studies described above. We also show that this model is doing frequency filtering in a qualitatively similar way as described in the literature.

Methods

The "Differential Encoding" (DE) model is based on an asymmetry in "patch" connectivity found in BA 22 (Galuske et al., 2000). "Patches" are found in many cortical areas across species, including sensory areas such as V1, V4, A1, and BA22 in monkeys and barrel cortex in rodents. Patches are thought to be a level of organization akin to a macro-column, consisting of thousands of selectively interconnected neurons within a cortical area. These patches selectively interconnect to a small subset of other local patches through horizontal connections through the grey matter. These patches are named because an injection of dye into the cortical surface will label cortex at the injection site, as well as rather discrete "patches" of surrounding cortex (see Figure 1) (Amir, Harel, & Malach, 1993; Levitt & Lund, 2002).

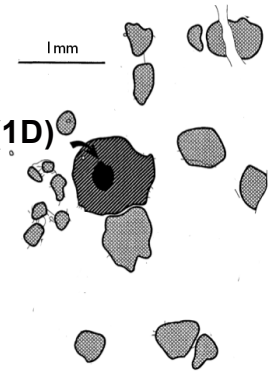

Figure 1: Drawing of "patches" in V4. Dark arrow indicates site of dye injection. Reproduced (without permission) fro Amir et al. (1993).

The function of these inter-patch connections is not known. Briefly, we propose here that horizontal connections lead to interconnected patches, biasing each other to process features shared across the group. We therefore implement a feed-forward model, where the hidden units discover the correlated features shared across interconnected patches for a particular input stimulus.

The "Differential Encoding" (DE) model includes two autoencoder neural networks with differences in connectivity, one for each hemisphere. Unlike most autoencoders, the hidden units of these models connect to a small subset of the

input and output banks (see Figure 2). Each hidden unit has a position in the input (and output) arrays, and a fixed number of connections to the input (and output) arrays are sampled from a Gaussian distribution centered at that hidden unit's position in the input (and output). The LH and RH autoencoders have the same number of hidden units and sample the same number of connections to the input (and output) for each hidden unit. The only difference between the networks, then, is the width (σ) of the Gaussian distribution. In accordance with the findings of Galuske et al. (2000), the left hemisphere network had a wider distribution than the right hemisphere network (see Figure 2).

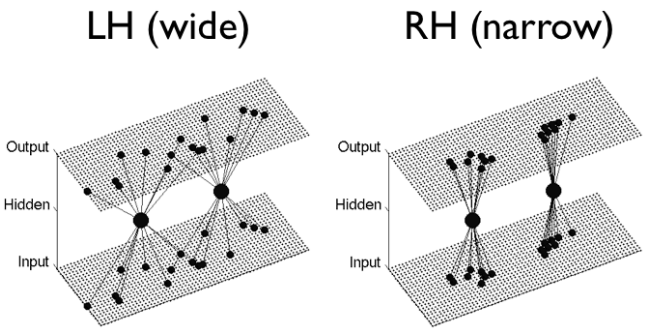

Figure 2: Representation of two hidden units for LH (left) and RH (right) autoencoder networks, along with their connections. The connections are randomly sampled from a Gaussian distribution centered on each hidden unit's position in the input array. The Gaussian distribution used for the LH is wider than that used for the RH. Not pictured is the classification networks which operate on the hidden unit encodings extracted from the autoencoder networks after training.

The number of hidden units was varied from extremely small (13) to extremely large (800) values. Results did not differ qualitatively when the number of connections per hidden unit varied to allow for the same number of overall weights to be used to learn the images. If too few weights were used, the networks could not learn the training set well enough for a meaningful analysis. The number of connections per hidden unit were fixed to be close to values reported in the literature, and the number of hidden units were chosen to allow equal spacing across the input image with enough total parameters to learn the images.

Each LH and RH network is constructed by randomly sampling connections for each hidden node. Gaussian distributions were used such that inter-patch distance values were similar to those reported in (Galuske et al., 2000).

Greyscale images are constructed for each task stimulus. The autoencoders are trained via of backpropagation of error (Rumelhart, Hinton, & Williams, 1986) to reproduce these greyscale images from the input to the output. Once the autoencoders reach a pre-determined performance level, training stops.[1] Each stimulus image is then presented to the trained autoencoder network, and the activation of the hidden units is recorded. These encodings, which will differ only due to the differences in connectivity structure between LH and RH networks, are then used as inputs to a separate feed-forward neural network which is trained to classify these encoded stimuli according to the behavioral task for the experiment.

For a single experiment, multiple "instances" of each LH and RH network are constructed and trained; each "instance" differs only in the random samples of its connections. The number of instances is determined in order to match the total number of trials used in the corresponding behavioral experiment, to match statistical power of results. Performance is evaluated on each model individually, then performance for all instances of each hemisphere are averaged. Average model error for each model hemisphere is compared to average reaction time in the corresponding human experiment, with both conceived as measures of difficulty or uncertainty in processing.

In order to examine how the different connectivity structures affect spatial frequency encoding, each stimulus image is presented to a trained autoencoder. Each output image produced is examined for spatial frequency content, and a 2D spectrogram across all images in the stimulus set is constructed. Each 2D spectrogram is translated to 1D spatial frequencies. Each spectrogram is compared to the spectrogram of the original image. The difference in spectrograms is then compared across hemispheres, showing for each frequency which hemisphere has encoded information closer to the original image than the other.

Experiments and Results

Sergent (1982) simulations

16 binary images (31x13 pixels) of Navon stimuli (letters [H, T, F, L] each appeared at local and global levels in all possible combinations) (see Figure 3b for example stimuli) were presented to 68 LH ($\sigma = 6.0$) and RH ($\sigma = 3.0$) autoencoder models to match the total number of trials in Sergent's human data. Each autoencoder network had 360 hidden units, with each hidden unit connecting to 8 input and output units. Each autoencoder network was trained to 0.005 average error per output unit, then hidden unit encodings were extracted. A perceptron classifier with 360 input units and one output unit was trained to classify each of the 16 Navon stimuli as containing a target or not.

As in Hsiao, Shahbazi, and Cottrell (2008), the network showed a significant hemisphere $\times$ level interaction (Two-factor, within-subject repeated measures ANOVA; $F(1,67)=8.62$; $p< 0.01$). We ran this same modeling experiment and analysis for all 6 combinations of target and distracter letters, to see if results would generalize. This only required training new classification neural networks, as the

[1] Similar to Ivry and Robertson (1998), autoencoder training is stopped before convergence to avoid overtraining. All simulations were re-run using weight decay to avoid over-training, and qualitatively similar results were found.

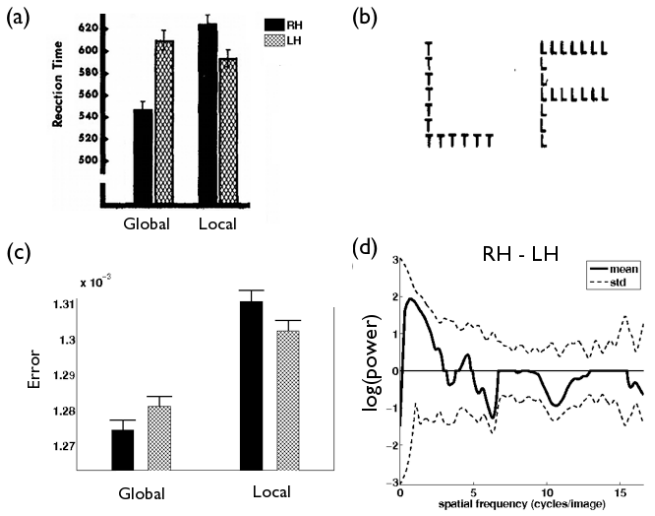

Figure 3: Original and model results for Sergent task

(a) Original hemisphere × level interaction; reproduced (without permission) from Sergent (1982)
(b) Sample Navon stimuli; reproduced (without permission) from Sergent (1982)
(c) DE model hemisphere × level interaction
(d) DE model spatial frequency analysis of output images, showing a RH advantage (above zero on Y-axis) for LSF (towards left side of X-axis) and a LH advantage (below zero on Y-axis) for HSF (towards right side of X-axis)

stimuli in each experiment remained the same. Each of these experiments showed a statistically significant hemisphere × level interaction.

Comparing the 1D spectrograms created from the output images of the autoencoder neural networks, we saw a clear tendency for the RH network to be closer to the spectrogram of the original image for LSFs, and the LH network to be closer to the spectrogram of the original image for HSF (see Figure 3d). These trends matched the large literature reporting better performance on LSF for LVF/RH and better performance on HSF for RVF/LH.

Kitterle et al (1992) simulations

40 greyscale images (31x13 pixels), each consisting of a low or high frequency grating, the grating either being a *Sine* or square wave, and shown at one of 10 phases (see Figure 4b for example stimuli), were presented to 40 LH ($\sigma = 6.0$) and RH ($\sigma = 3.0$) autoencoder models. Each autoencoder network had 360 hidden units, with each hidden unit connecting to 8 input and output units. Each autoencoder network was trained to 0.005 average error per output unit, then hidden unit encodings were extracted. Two identical neural networks with 360 input units, ten hidden units, and one output unit were trained to classify each of the 40 stimuli. One classification network was trained to discriminate between wave type (sin vs square) and to ignore the frequency of the waves; this task required use of HSF information. The other classification network was trained to discriminate between the two frequencies of the waves and to ignore the wave type; this task required use of LSF information. Both networks were trained with the same training parameters.

The first classification network showed a significant effect of hemisphere ($F(1,39)=4.06$; $p< 0.05$), with the LH network showing better performance. The second classification network showed a significant effect of hemisphere ($F(1,39)=4.53$; $p< 0.04$), with the RH network showing better performance. No statistical test was necessary to confirm the hemisphere × task interaction. Note that the main effect of task-type was not preserved; our models found discriminating the wave type to be an easier task than discriminating between two frequencies.

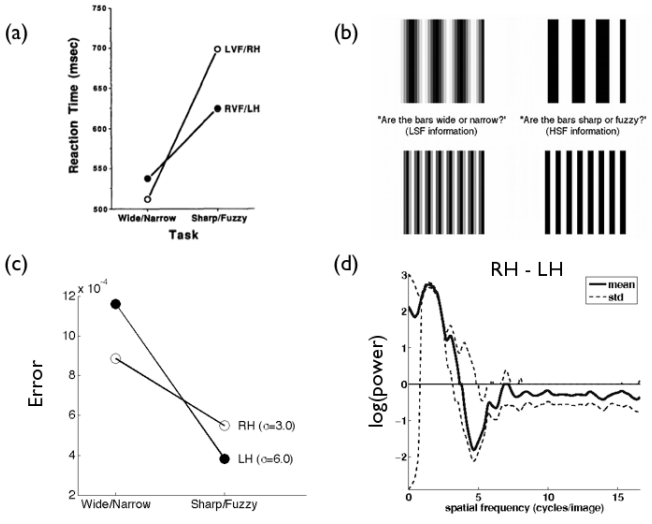

Figure 4: Original and model results for Kitterle task

(a) Original hemisphere × task interaction; reproduced (without permission) from Kitterle et al. (1992)
(b) Sample stimuli used in the modeling study
(c) DE model hemisphere × task interaction
(d) DE model spatial frequency analysis of output images, showing a RH advantage (above zero on Y-axis) for LSF (towards left side of X-axis) and a LH advantage (below zero on Y-axis) for HSF (towards right side of X-axis)

Face Processing simulations

Young and Bion (1981): Face Recognition Young and Bion (1981) (and other) studies have shown a RH advantage for face recognition. We set out to replicate this general finding.

The same face stimuli used in Hsiao, Shieh, and Cottrell (2008) were used to construct greyscale images in this study. 240 greyscale face images (34x25 pixels). The dataset contained 30 individuals with 8 expressions each; 4 expressions

used in training, and 4 different expressions were used in the data collection/testing phase. These face stimuli were more complex, and so required more parameters to train to a looser error criterion. The face stimuli were presented to 40 LH ($\sigma = 8.0$) and RH ($\sigma = 3.0$) autoencoder models. Each autoencoder network had 360 hidden units, with each hidden unit connecting to 12 input and output units; ; the wider LH gaussian was selected to space out the greater number of connections per hidden unit. Each autoencoder network was trained to 0.01 average error per output unit, then hidden unit encodings were extracted. A neural network with 360 input units, 25 hidden units, and 30 output unit was used to classify each of the 120 test images as one of the 30 individuals.

A significant effect of hemisphere on face identification accuracy was found (t-test; $p < 1 * 10^{-11}$). These effects were consistent across the training and test sets.

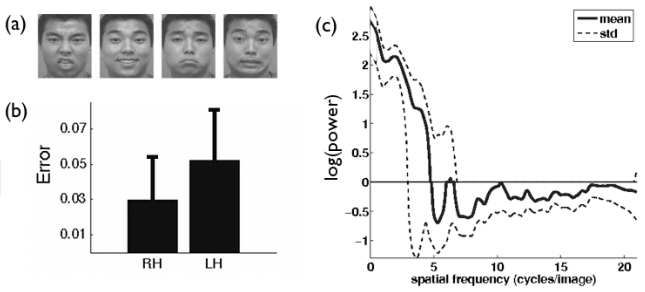

Figure 5: Stimuli and spectrogram for left-side bias task

(a) Sample stimuli for one individual across four expressions (from the CAFE dataset)
(b) DE model classification of individual face recognition identity (Young & Bion task); error-bars represent standard error of the mean
(c) DE model spatial frequency analysis of output images, showing a RH advantage (above zero on Y-axis) for LSF (towards left side of X-axis) and a LH advantage (below zero on Y-axis) for HSF (towards right side of X-axis)

Brady et al. (2005): Left-Side Bias The same face stimuli used in Hsiao, Shieh, and Cottrell (2008) were used to construct greyscale face images. 240 greyscale face images (34x25 pixels; 30 individuals; 4 expressions used in training, 4 different expressions used in testing) were used to create left and right chimeric faces: faces with one side duplicated across the midline to the other. The same network parameters were used as above for training.

For each set of chimeric faces, a significant effect for hemisphere was found, with a RH advantage for face recognition in each case (left chimeric: $F(1,39)=7.58$, $p < 0.01$; right chimeric: $F(1,39)=8.83$; $p < 0.01$), again replicating (twice) RH advantage for face identification. Comparing across left and right chimeric faces, both hemispheres showed a significant preference for left chimeric images, replicating the left-side bias effect.

Christman et al. (1991) simulations

Two sets of 16 greyscale images (31x13 pixels) were constructed, each consisting of a two types of stimuli. The first type stimulus consisted of two frequency gratings at different relative phases to each other. The second stimulus type consisted of the first set of stimuli, with a third frequency grating superimposed upon them. In one stimulus set, the third frequency grating was at a higher spatial frequency than the other two frequency gratings; in the second stimulus set, it was at a lower spatial frequency than the other two. Importantly, the third frequency grating was of exactly the same spatial frequency in both stimulus sets (see Figure 6b for example stimuli). There were 4 phase variations for each stimulus type in each stimulus set.

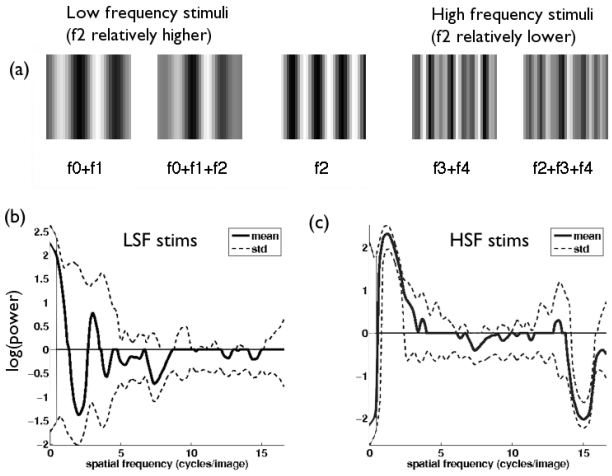

Figure 6: Original results, sample model stimuli, and model spectrograms

(a) Sample stimuli created to train DE models
(b) DE model spatial frequency analysis of output images, showing much flatter spectrogram differences than all other stimuli, with weak RH advantage (above zero on Y-axis) for LSF (towards left side of X-axis) and a weak LH advantage (below zero on Y-axis) for HSF (towards right side of X-axis)

Each set of greyscale images was presented to 64 LH ($\sigma = 6.0$) and RH ($\sigma = 3.0$) autoencoder models. Each autoencoder network had 360 hidden units, with each hidden unit connecting to 8 input and output units. Each autoencoder network was trained to 0.005 average error per output unit, then hidden unit encodings were extracted. For each set of greyscale images, a neural networks with 360 input units, ten hidden units, and one output unit were trained to classify each of the 16 stimuli.

For both stimulus sets (LSF and HSF), there was a significant effect of stimulus class, with the 3-component stimu-

lus being harder to classify than the 2-component stimulus. However, there was no hemisphere $\times$ stimulus class interaction. Looking at the spectrogram differences between the two stimulus classes (Figure 6c and 6d), and comparing them to the previous spectrogram differences, it is clear that there is much less encoding asymmetry between the two model hemispheres for these stimuli as compared to all other stimuli used in model experiments within this paper.

We tested a few possible explanations for this. We tried many different combinations of spatial frequency gratings; this varied which model hemisphere showed better performance, but no reliable hemisphere $\times$ stimulus class interaction. We tried larger images, to expand the range of spatial frequencies that could be encoded, but again no consistency was found. Lastly, we tried training the autoencoder on separate dataset, then extracting hidden unit encodings on the task-relevant stimuli. Again, this did not show any consistent interaction.

Further work is warranted to better characterize whether the DE model can account for this critical dataset. We have developed (elsewhere) a developmental model of this asymmetry which suggests that this dataset may be modeled by engagement of more than one cortical area showing asymmetry. This pattern is seen in neuroimaging results reported by (Hopf et al., 2006), for example. Variations in average inter-patch distance based on cortical area Amir et al. (1993) suggest that different areas may have different frequency preferences. This would suggest that "relative frequency" processing may in fact be simply selecting different absolute frequency filters based on task demands. We are currently investigating whether this is might provide an alternate explanation to these data, and the idea of relative frequency encoding in general.

Conclusions

Here, we showed that an asymmetry in local connectivity can account for local/global behavioral data, face processing data, and matches spatial frequency asymmetries reported in the literature. This model provides a biologically grounded implementation for these phenomena, and the analyses here showing consistent frequency filtering differences in the model hemispheres are consistent with the current algorithmic explanation for visual processing asymmetries via frequency filtering. Unlike the DFF model, however, these frequency filtering differences are found at a post-sensory encoding stage. Further work must be done to investigate whether our failure to model the results of Christman et al. (1991) is due to practical modeling concerns, or suggests a fundamentally different approach to modeling local/global processing asymmetry.

Acknowledgments

This work was partly funded by a Center for Academic Research and Training in Anthropogeny (CARTA) fellowship.

References

Amir, Y., Harel, M., & Malach, R. (1993). Cortical hierarchy reflected in the organization of intrinsic connections in macaque monkey visual cortex. *The Journal of Comparative Neurology*, *334*(1), 19–46.

Brady, N., Campbell, M., & Flaherty, M. (2005). Perceptual asymmetries are preserved in memory for highly familiar faces of self and friend. *Brain and Cognition*, *58*(3), 334–342.

Christman, S., Kitterle, F. L., & Hellige, J. (1991). Hemispheric asymmetry in the processing of absolute versus relative spatial frequency. *Brain and Cognition*, *16*(1), 62–73. (PMID: 1854470)

Galuske, R. A., Schlote, W., Bratzke, H., & Singer, W. (2000). Interhemispheric asymmetries of the modular structure in human temporal cortex. *Science (New York, N.Y.)*, *289*(5486), 1946–1949. (PMID: 10988077)

Hopf, J., Luck, S. J., Boelmans, K., Schoenfeld, M. A., Boehler, C. N., Rieger, J., et al. (2006). The neural site of attention matches the spatial scale of perception. *The Journal of Neuroscience*, *26*(13), 3532–3540. Available from `http://www.jneurosci.org/content/26/13/3532`

Hsiao, J., Shahbazi, R., & Cottrell, G. (2008). Hemispheric asymmetry in visual perception arises from differential encoding beyond the sensory level. *Proceedings of the 30th Annual Meeting of the Cognitive Science Society.*.

Hsiao, J., Shieh, D., & Cottrell, G. (2008). Convergence of the visual field split: Hemispheric modeling of face and object recognition. *Journal of Cognitive Neuroscience*, *20*(12), 2298–2307.

Hutsler, J., & Galuske, R. A. W. (2003). Hemispheric asymmetries in cerebral cortical networks. *Trends in Neurosciences*, *26*(8), 429–35. (PMID: 12900174)

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

Kitterle, F. L., Hellige, J. B., & Christman, S. (1992). Visual hemispheric asymmetries depend on which spatial frequencies are task relevant. *Brain and Cognition*, *20*(2), 308–314. (PMID: 1449760)

Levitt, J., & Lund, J. (2002). Intrinsic connections in mammalian cerebral cortex. In A. Schuez & R. Miller (Eds.), *Cortical areas: unity and diversity.* CRC Press.

Rogers, L. J., & Andrew, R. (2002). *Comparative vertebrate lateralization* (1st ed.). Cambridge University Press.

Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1986). Learning representations by back-propagating errors. *Nature*, *323*(6088), 533–536.

Sergent, J. (1982). The cerebral balance of power: confrontation or cooperation? *Journal of Experimental Psychology. Human Perception and Performance*, *8*(2), 253–72. (PMID: 6461721)

Young, A. W., & Bion, P. J. (1981). Accuracy of naming laterally presented known faces by children and adults. *Cortex; a Journal Devoted to the Study of the Nervous System and Behavior*, *17*(1), 97–106.