# 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 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 implemented to account for asymmetric processing of each stimulus type above, but does not provide a fully mechanistic explanation, nor does it have direct neural correlates. The Differential Encoding (DE) model (Hsiao, Shahbazi, & Cottrell, 2008), which postulates that a known asymmetry in patch connectivity drives visual processing asymmetries, has previously been used to account for only one stimulus type. Here, we refine the DE model to match the published patch asymmetry more precisely and show that the DE model generalizes to three of the four datasets mentioned above. Examination of the failure to match all datasets suggest a possible reinterpretation of the original dataset itself.

**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 only 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 3b), 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 these asymmetries. The computational model they implemented aimed to account for three particular experiments from the literature, thought to express core features of the data:

- Sergent (1982): showed the basic hemisphere $\times$ level interaction of the local/global literature, with responses to targets presented at the smaller, "local" level showing faster reaction times when presented to the RVF/LH, and responses to targets presented at the larger, "global" level showing faster reaction times 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. Responses to the task requiring high spatial frequency (HSF) information showed faster reaction times when presented to the RVF/LH, and responses to the task requiring low spatial frequency (LSF) information showed faster reaction times 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 a homunculus 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 one 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, and not in primary auditory cortex. This matches current theory in the local/global literature, where it has been suggested that local/global processing differences occur beyond early primary sensory areas (Sergent, 1982; Ivry & Robertson, 1998), motivating the use of this asymmetry to model tasks involving local/global processing.

Hsiao et al. used this connectivity pattern as inspiration to implement a simple "autoencoder" neural network with differential connectivity (see Figure 2) to encode the stimuli, and a simple perceptron to perform the task using the autoencoders' learned representations (see Methods section for details on this model). They tested it to see whether this anatomical asymmetry could account for a subset of the data modeled by Ivry and Robertson. Using precisely the same small, 1-dimensional 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 $\times$ 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 $\times$ 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.

The DE model was able to address issues with the DFF model by implementing an anatomical asymmetry. The use of a secondary network for classification allows the learning algorithm to select task-relevant information, rather than manually selecting frequency bands as does the DFF model.

However, the DE model 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 very different from the parameters reported in the literature.

Here we improve the model's fidelity to neural data. We show that this model accounts for three of the four studies described above. We also show that this model filters frequencies in a manner consistent with the human 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 (monkeys, humans, cats, rodents) and across sensory and association areas. 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).

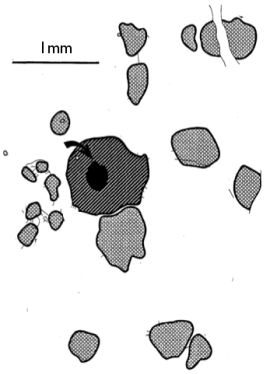


Figure 1: Drawing of "patches" in V4. Dark arrow indicates site of dye injection. Reproduced (without permission) from Amir et al. (1993).

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 interconnected group. We therefore implement a feed-forward model, where the hidden units discover the correlated features shared across interconnected input "patches" across a set of input stimuli.

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 ($\sigma$) 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).

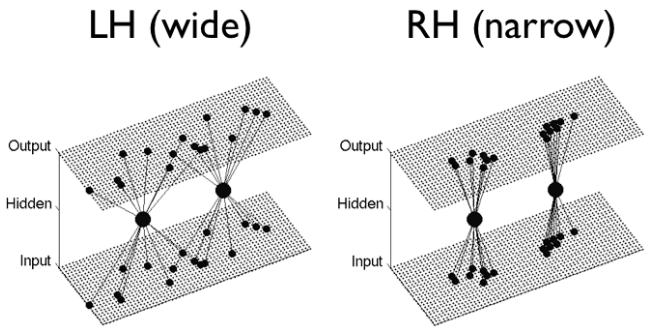


Figure 2: Representation of two hidden units for LH (left) and RH (right) autoencoder networks, along with their connections. The connections are randomly sampled from a Gaussian distribution centered on each hidden unit's position in the input array. The Gaussian distribution used for the LH is wider than that used for the RH. Not pictured are 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. After these initial explorations, the final simulations were run with the number of connections per hidden unit fixed to be within the range reported by Amir et al. (1993) in visual cortex (between 8 and 15), 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 calculated from data reported in (Galuske et al., 2000). On average, patches had 1.75 times the width of a single patch between them in the RH, and 2.05 in the LH. Here, we considered the size of a patch to be a pixel, and chose sigmas such that when inter-patch distance was measured after randomly sampling connections, on average there were 1.75 pixels between connections in the RH, and 2.05 in the LH.

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. Each stimulus image is then presented to the trained autoencoder network, and the activation of the hidden units is recorded. These encodings, which 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 encodings 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 by matching the total number of trials used both in the modeling experiment and in the corresponding behavioral experiment, such that the statistical power of each experiments are equated. Performance is evaluated on each model individually, then performance over all "instances" of a hemisphere are averaged. Average model error for each model hemisphere is compared to average reaction time of the corresponding visual field in the human experiment, with both measures 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 power spectrum across all images in the stimulus set is constructed. Each 2D power spectrum is translated to a 1D power spectrum by measuring the linear distance from the pixel carrying the f0 (intensity) power to each pixel in the 2D power spectrum, then sorting all linear distances and averaging over any pixels with the same linear distance. Each 1D power spectrum is then compared to the power spectrum of the original image. The difference in 1D power spectrums 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

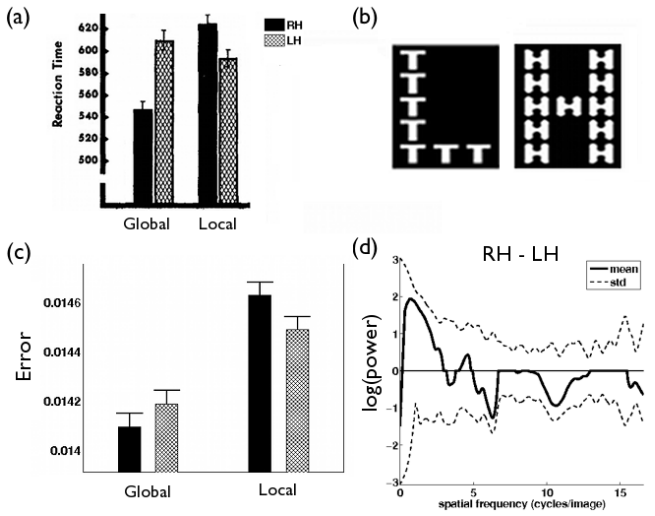


Figure 3: Original and model results for Sergent task
(a) Original hemisphere × level interaction; reproduced (without permission) from Sergent (1982)
(b) Sample Navon stimulus used in our modeling experiment
(c) DE model hemisphere × level interaction
(d) DE model spatial frequency analysis of output 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)

required training new classification neural networks, as the stimuli in each experiment remained the same. Each of these experiments showed a statistically significant hemisphere × level interaction.

Comparing the 1D power spectrums created from the output images of the autoencoder neural networks, we saw a clear tendency for the RH network to be closer to the power spectrum of the original image for LSFs, and the LH network to be closer to the power spectrum 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 (sine 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. Across the two tasks, the networks showed a significant hemisphere × task interaction (Two-factor, within-subject repeated measures ANOVA; $F(1,39)=9.89$; $p< 0.002$). 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; Kitterle et al. (1992) reported the opposite result for their human participants.

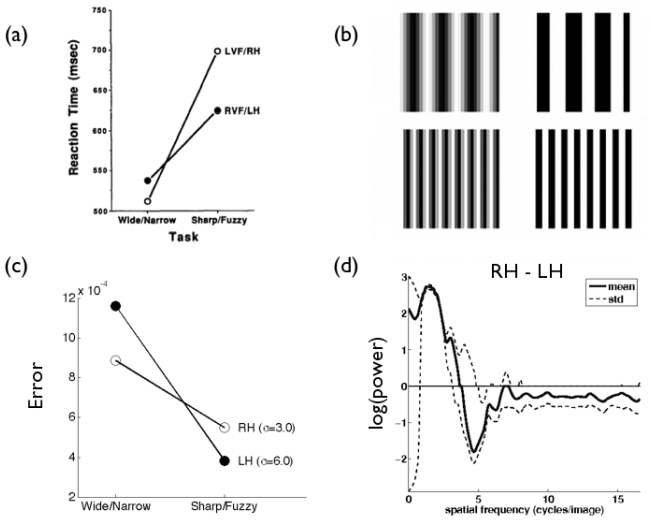


Figure 4: Original and model results for Kitterle task
(a) Original hemisphere × task interaction; reproduced (without permission) from Kitterle et al. (1992)
(b) Sample stimuli used in the modeling study
(c) DE model hemisphere × task interaction
(d) DE model spatial frequency analysis of output images, 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), which used the DFF model to show a left-side bias, were used to construct greyscale images in this study. 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 lesser 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 (ANOVA; F(1,39) =10.33, $p < 0.002$). These effects were consistent across the training and test sets.

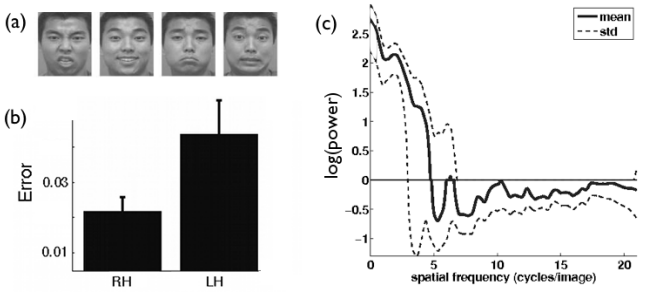


Figure 5: Stimuli and power spectrum for left-side bias task
(a) Sample stimuli for one individual across four expressions (from the CAFE dataset)
(b) DE model classification of individual face recognition identity (Young & Bion task); error-bars represent standard error of the mean
(c) DE model spatial 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), which used the DFF model to show a left-side bias, 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 the face recognition simulation 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 a 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.

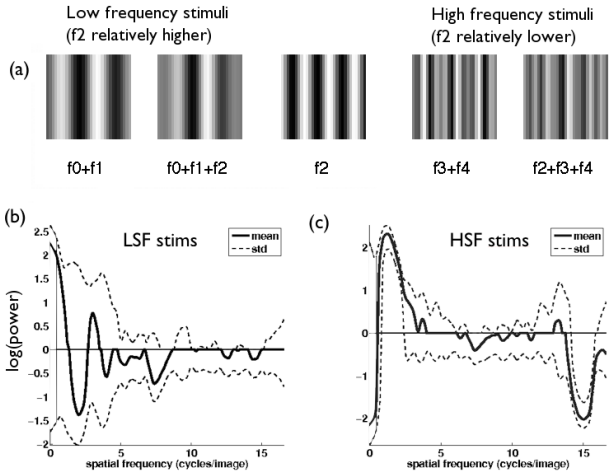


Figure 6: Original results, sample model stimuli, and model power spectrum
(a) Sample stimuli created to train DE models
(b) DE model spatial frequency analysis of output images, showing much flatter power spectrum differences than all other stimuli, with weak RH advantage (above zero on Y-axis) for LSF (towards left side of X-axis)
(c) 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 network 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 stimulus being harder to classify than the 2-component stimulus. However, there was no hemisphere $\times$ stimulus class interaction. Looking at the power spectrum differences between

the two stimulus classes (Figure 6b and 6c), and comparing them to the previous power spectrum 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 created (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, where a leftward asymmetry of an earlier visual processing area is engaged by a task at the "local" level, and a rightward asymmetry of a later visual processing area is engaged by a task at the "global" level. 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, implemented in different cortical areas, 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, as well as by NSF grant SMA 1041755 to the Temporal Dynamics of Learning Center, an NSF Science of Learning Center

## 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.

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.

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.

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.

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.

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

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.