

An applet for the Gabor similarity scaling of the differences between complex stimuli

Eshed Margalit¹ · Irving Biederman^{1,2} · Sarah B. Herald^{1,3} · Xiaomin Yue⁴ · Christoph von der Malsburg⁵

Published online: 24 August 2016
© The Psychonomic Society, Inc. 2016

Abstract It is widely accepted that after the first cortical visual area, V1, a series of stages achieves a representation of complex shapes, such as faces and objects, so that they can be understood and recognized. A major challenge for the study of complex shape perception has been the lack of a principled basis for scaling of the physical differences between stimuli so that their similarity can be specified, unconfounded by early-stage differences. Without the specification of such similarities, it is difficult to make sound inferences about the contributions of later stages to neural activity or psychophysical performance. A Web-based app is described that is based on the Malsburg Gabor-jet model (Lades et al., 1993), which allows easy specification of the V1 similarity of pairs of stimuli, no matter how intricate. The model predicts the psychophysical discriminability of metrically varying faces and complex blobs almost perfectly (Yue, Biederman, Mangini, von der Malsburg, & Amir, 2012), and serves as the input stage of a large family of contemporary neurocomputational models of vision.

Keywords 2-D shape and form · Similarity · Face perception

✉ Irving Biederman
bieder@usc.edu

¹ Neuroscience Program, University of Southern California, Los Angeles, CA, USA

² Department of Psychology, University of Southern California, Los Angeles, CA, USA

³ Department of Psychological and Brain Sciences, Dartmouth College, Hanover, NH, USA

⁴ Laboratory of Brain and Cognition, National Institute of Mental Health, Bethesda, MD, USA

⁵ Frankfurt Institute of Advanced Studies, Frankfurt, Germany

Consider the problem of determining whether people or monkeys are more sensitive to differences in nonaccidental properties (NAPs)—whether a contour is straight or curved, for example—than to differences in metric properties (MPs)—such as differences in degrees of curvature. If we assume that the sensitivity to differences in NAPs arises at a stage in the ventral pathway later than V1, how can the physical properties of the stimuli be selected in a principled manner, so that the comparisons are not confounded with differences in V1 activation? The same methodological problem arises if an investigator wishes to determine whether observers are more sensitive to differences in facial expression than to differences in identity (or sex, or orientation in depth, etc.). This problem arises not only in psychophysical scaling of stimuli, but also with studies designed to more directly reflect the underlying neural correlates, such as fMRI fast-adaptation designs and single-unit recordings. It can be argued that this problem of the scaling of shape similarity had been a major reason why, despite shape being the major input into visual cognition, the rigorous study of shape perception had clearly lagged the study of other perceptual attributes, such as color, motion, or stereo.

The value of an intuitive implementation of the Gabor-jet model

Despite the utility in employing such a scaling system, the Gabor-jet model is somewhat mathematically dense and cumbersome to explain to the uninitiated, thus diminishing its accessibility. Here, we introduce a Web-based applet designed to provide an engaging, graphically oriented guided tour of the model. The applet allows users to upload their own images, observe the transformations and computations made by the algorithm, customize the visualization of different processes, and retrieve a ranking of dissimilarity values for pairs of images. Such interactive experiences

can be valuable in fostering an understanding of otherwise challenging methodologies, rendering this tool accessible to a broad range of users. Since almost all contemporary neurocomputational models of vision assume a form of Gabor filtering as their input stage, an understanding of the Gabor-jet model also provides an introduction to the first stage of the larger family of computer vision approaches, including GIST (Oliva & Torralba, 2001), HMAX (Riesenhuber & Poggio, 1999), and recently popular convolution neural network (CNN) approaches (e.g., Krizhevsky, Sutskever, & Hinton, 2012). Of course, frivolous applications can be enjoyed. When Suri, the daughter of Tom Cruise and Katie Holmes, was a toddler, a (much too) lively debate raged as to which parent Suri most resembled. *People Magazine* requested that author I.B. weigh in with the model's choice (http://celebritybabies.people.com/2006/09/19/who_does_suri_r/).

An applet for the Gabor-jet model

The Gabor-jet model (Lades et al., 1993) is designed to capture the response properties of simple cells in V1 hypercolumns, whose receptive field spatial profiles can be described by two-dimensional Gabor functions (De Valois & De Valois, 1990; Jones & Palmer, 1987; Ringach, 2002). Gabor modeling of cell tuning in early visual cortex has also enjoyed great success in other computational models of visual processing (Kay, Naselaris, Prenger, & Gallant, 2008; Serre & Riesenhuber, 2004). By representing image inputs as feature vectors derived from convolution with Gabor filters, the Gabor-jet model can be used to compute a single value that represents the similarity of two images with respect to V1 cell filtering. These values have been shown to almost perfectly predict psychophysical similarity in discriminating metrically varying, complex visual stimuli such as faces and blobs (resembling teeth; Yue, Biederman, Mangini, von der Malsburg, & Amir, 2012). Under the assumption that V1 captures metric variation, sensitivity to the “qualitative” differences between complex stimuli, such as nonaccidental (i.e., viewpoint-invariant) properties (NAPs) versus metric (i.e., viewpoint-dependent) properties (MPs), or to differences in facial identity versus expression (which are presumably rendered explicit in later stages) can be more rigorously evaluated.

Prior to the implementation of the Gabor-jet model, a common scaling technique was to examine differences in pixel energy between pairs of stimuli. Of course, this method neglects the information present in orientation and scale. Yue et al. (2012) provided an example in which relatively slight differences in the orientations of two straight contours yielded pixel energy differences that were equivalent to the differences between a pair of contours, one straight and the other curved, which would have been much more readily discriminated. Perhaps the most general and well-documented effect in shape perception is that differences in NAPs of shape, such as straight versus curved, are much more readily discriminated than MPs, such as differences in degree of

curvature (e.g., Amir, Biederman, & Hayworth, 2012). However, this inference could not be made without a scaling that equated the NAP and MP differences according to early stage filtering. Otherwise, one could not know how much a particular difference in curvature could be equated to a NAP differences between straight and curved.

Gabor-like filters develop as part of the linear decomposition of natural images (Olshausen & Field, 1996) so the Gabor-like filtering characteristic of V1 simple cells is not unexpected. These basis sets emerge in the first layer of leading CNNs for image recognition (e.g., Krizhevsky et al., 2012) or are simply assumed as in the GIST model of Oliva and Torralba (2001), that adopts the use of multi-scale, multi-orientation Gabor filters to create a sparse description of image locations in much the same way that each jet in the Gabor-jet model is composed of a set of Gabor filters at different scales and orientations that share a common center in the image space. Similarly, the first layer of HMAX (Riesenhuber & Poggio, 1999) convolves image pixels with oriented Gabor filters before pooling responses (and then repeating those operations). So although the Gabor-jet model was developed almost a quarter of a century ago, its offering of an explicit measure of V1-based image similarity is still relevant given the widespread incorporation of Gabor filtering as the input stage in contemporary neurocomputational models of vision.

Our implementation of the Gabor-jet model follows that of Lades et al. (1993), in which each Gabor “jet” is modeled as a set of Gabor filters at five scales¹ and eight orientations, with the centers of their receptive fields tuned to the same point in the visual field (Fig. 1). We employ a 10×10 square grid—and therefore 100 jets—to mark the positions in the image space from which filter convolution values were extracted.

In an implementation of this model (version 2.0) that is specialized to scale the similarity of faces (Wiskott, Fellous, Krüger, & von der Malsburg, 1997), rather than a square grid, each jet is “trained” to center its receptive field over particular facial landmarks (or *fiducial points*), so that jet number 6, for example, automatically centers itself on the pupil of the right eye, and jet 17 on the tip of the nose (as is illustrated in Fig. 1, right panel). Version 2 can thus achieve translation invariance and, given appropriate training, some robustness over variations of orientation in depth and expression. Here we maintain the uniform 10×10 grid to simplify the model and preserve its ability to code objects as well as faces. In all, 40 filters (5 scales \times 8 orientations) are generated at each of the 100 locations specified by the grid, yielding a total of 4,000 Gabor filters. The input image is convolved with each filter, and the model stores both the magnitude and the phase of the filtered image. We construct a feature vector of 8,000 values for a given input image by concatenating the two values from each of the 4,000 filters. In our implementation, the Euclidean distance between any two output vectors—which has been

¹ Wilson, McFarlane, and Phillips (1983) showed that five scales could approximate human spatial frequency tuning.

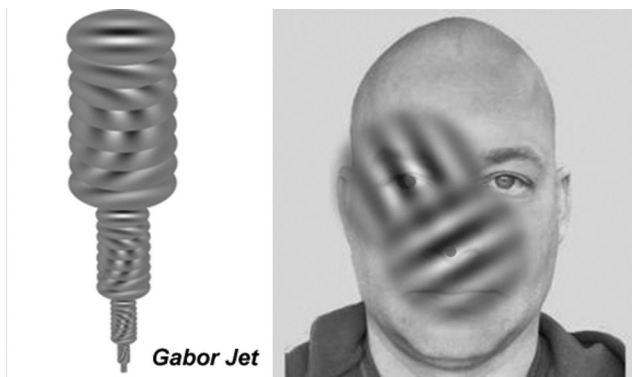


Fig. 1 (Left) Conceptualization of a “Gabor jet” that is composed of 40 filters whose receptive field centers are tuned to a common point in the visual field. (Right) Depiction of the filtering of a face by two Gabor filters of the same larger scale but different orientations and positions. In the current implementation, each of the 100 jets would be centered over a given node in a 10×10 grid. Modified from Fig. 2 of “A Neurocomputational Account of the Face Configural Effect,” by X. Xu, I. Biederman, and M. P. Shah, 2014, *Journal of Vision*, 14(8), article 9. Copyright 2014 by the Association for Research in Vision and Ophthalmology. Adapted with permission

shown to correlate most highly with human discrimination performance (Yue et al., 2012) and is more robust to outliers than other distance metrics such as correlation—is considered to be the perceptual distance between the two images that produced those arrays (Fig. 2). Thus, we are able to compute a single value representing the dissimilarity of two images

(according to an approximation of the V1 cell tuning). Scaling stimulus sets by this metric allows strong inferences to be made about both V1-level similarity effects and later-stage contributions to shape processing in the visual system.

The value of Gabor-jet scaling

It is one thing to compute a theoretical measure of V1 similarity, but it is quite another to document that such a measure actually predicts the psychophysical similarity of metrically varying complex stimuli. Yue et al. (2012) employed a match-to-sample task in which subjects viewed a briefly presented triangular display of either three complex blobs (resembling teeth) or three faces, as is illustrated in Fig. 3. One of the lower stimuli was an exact match to the sample (top stimulus). The advantage of match-to-sample tasks over same-different tasks is that in the latter, subjects must set an arbitrary criterion as to whether to respond “same” or “different” when the similarities are very high. In the match-to-sample task, the subject merely has to select the stimulus that is most similar to the sample, since it is an exact match. The correlations between the Gabor-jet computation of the dissimilarity of the foil and matching stimuli and response error rates were .985 for blobs and .957 for faces, which accounts for just about all of the predictable variance (given that there is some unreliability in subjects’ behavior).

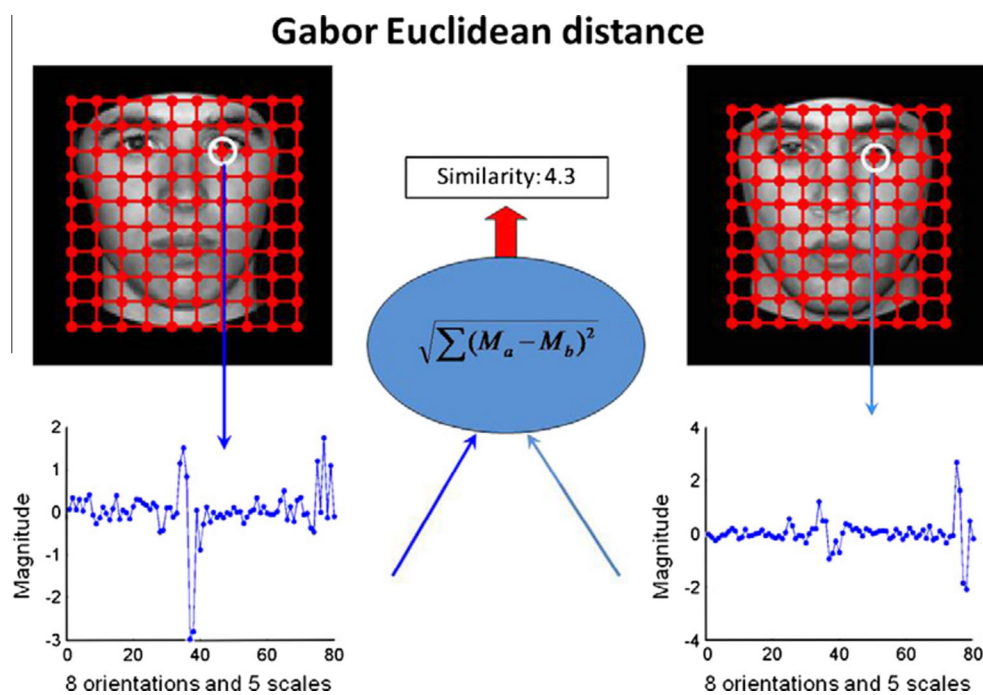


Fig. 2 Activation values of a sample filter from the 10×10 grid, taken from two frontal face images. The differences in activation between the corresponding filters of two stimuli are used to compute the dissimilarity of the images. From “Predicting the Psychophysical Similarity of Faces

and Non-Face Complex Shapes by Image-Based Measures,” by X. Yue, I. Biederman, M. C. Mangini, C. von der Malsburg, and O. Amir, 2012, *Vision Research*, 55, p. 43. Copyright 2012 by Elsevier. Reprinted with permission

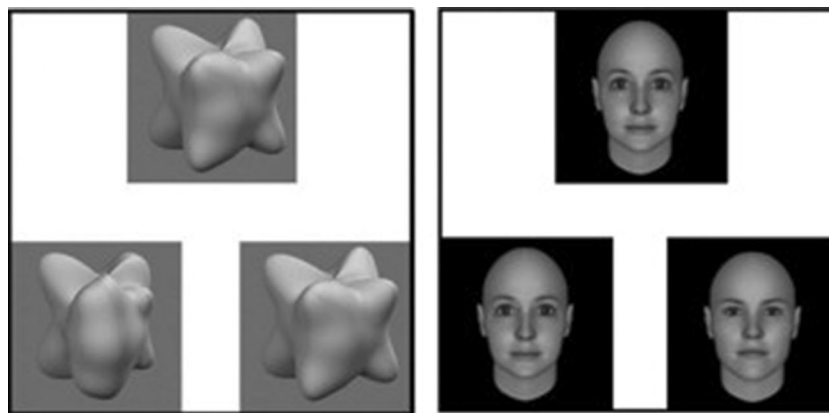


Fig. 3 Two sample trials from Yue et al.'s (2012) match-to-sample study. For both blobs (left) and artificial faces created in FaceGen (right; Singular Inversions, Vancouver, Canada), subjects were instructed to indicate which of the bottom two stimuli was an exact match of the sample above. Nearly all of the variance in error rates on this task was

predicted by the Gabor-jet model's computation of dissimilarity. The correct matching stimulus for the blobs is on the right, and for the faces is on the left. Modified from Fig. 1 of Yue et al. (2012), p.42. Copyright 2012 by Elsevier. Adapted with permission

This was the first time that the psychophysical similarity of complex stimuli had been predicted from a formal model. Most applications of the Gabor-jet model have been in the form of computer models of face recognition—for instance, those of Günther, Haufe, and Würtz (2012) and Jahanbin, Choi, Jahanbin, and Bovik (2008). Examples in which stimuli have been scaled according to Gabor-jet similarity can be found in Yue, Tjan, and Biederman (2006), for faces and blobs; Xu, Yue, Lescroart, Biederman, and Kim (2009) and Xu and Biederman (2010), for faces; Kim, Biederman, Lescroart, and Hayworth (2009), for the metric variation of line drawings of common objects and animals; Kim and Biederman (2012), for the relations between simple parts; Amir et al. (2012), for simple shapes; and Lescroart and Biederman (2012), for differences in part shapes and medial axis relations.

Platform

The Web app (<http://geon.usc.edu/GJW>) is presented as a webpage using HTML5 and CSS3, hosted on the geon.usc.edu server. Dynamic components of the page are written in JavaScript and utilize the popular jQuery library. External code is used, with acknowledgement from public libraries. Where possible, measures were taken to provide a complete experience across a variety of browsers and browsing devices, including smartphones and tablets.

Implementation details

Following a brief set of upload instructions, the user is prompted to upload three images. JPG and PNG file formats may be uploaded by using a file-selector button or the camera application

on most mobile devices. If necessary, images may be resized by the applet to 256×256 pixels. Although resizing may disrupt the aspect ratio if the uploaded image dimensions are nonsquare, this approach may be preferable to cropping the images (which deletes information altogether) or rescaling them such that the longest side is 256 pixels (which could pad the image space with uninformative values and introduce artificial boundaries).

If the user prefers, an option is present to use predefined sets of sample images, including colorized versions of artificially generated faces that have been used for in-house psychophysical studies (Fig. 4), textured blobs, or object stimuli from an fMRI study of lateral occipital complex function (Margalit et al., *in press*). In the artificial face default set, the difference between the first two default images is almost imperceptible, whereas Image 3 is readily perceived as being different from the first two. We would expect the Gabor dissimilarity values between pairs of these images to reflect these perceptual differences.

The application then grayscales the uploaded images by replacing each pixel of a color image by the scalar average of the red, green, and blue color layers (Fig. 5). A 10×10 grid of red dots is overlaid to demonstrate the points at which the Gabor jets will be centered.

Once an image is grayscaled, the model performs a two-dimensional fast Fourier transform (FFT) to convert the image into the frequency domain. Our two-dimensional FFT implementation first performs a one-dimensional FFT on the pixel columns of the image, then another one-dimensional FFT on the “pixel” rows of the output of the prior operation. For each uploaded image, 40 Gabor filters are generated (5 different scales \times 8 different orientations), according to the equation

$$G(x, y | W, \theta, \varphi, X, Y) = e^{-\frac{(x-X)^2 + (y-Y)^2}{2\sigma^2}} \sin[W(x\cos\theta - y\sin\theta) + \varphi],$$

where σ is the width of the Gaussian envelope, θ is the filter's orientation, W is its scale, φ is its phase shift, and X, Y describes

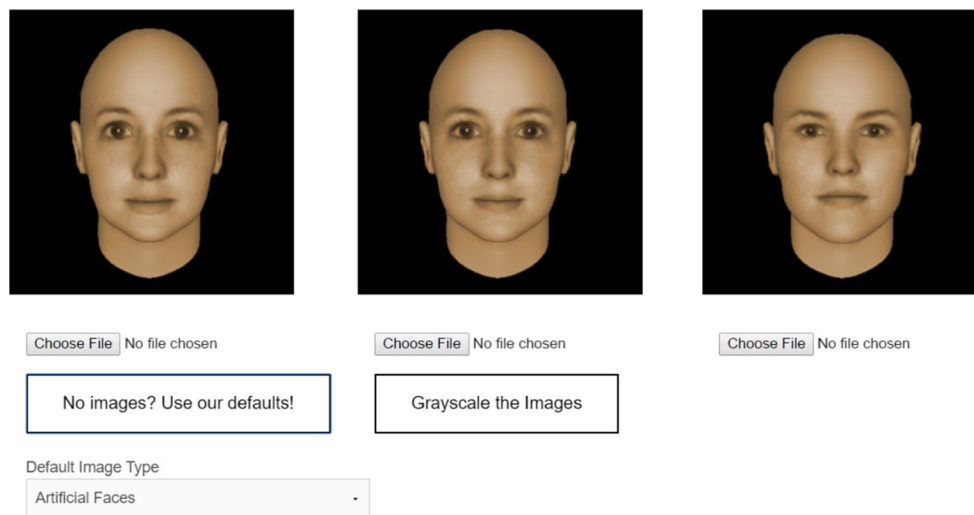


Fig. 4 The three default images used in the standard tutorial. These face images are colorized versions of stimuli used in Yue et al. (2012)

the center of the filter. In the JavaScript implementation used here, the equation is slightly modified to generate two kernels—identical except for a 90-deg phase shift—simultaneously at each location, scale, and orientation. Thus, 80 kernels are effectively used at each location. This phase shift allows sensitivity to the direction of contrast in the images, a key feature of the perception of certain complex objects such as faces. However, the visualizations generated by the applet represent complex cell responses, which are defined by a single filter at each scale, orientation, and position. Kernels are normalized to have a mean of zero and unit variance, making the model invariant to changes in global brightness (as long as changes in brightness do not saturate the image pixels, thus changing the edge structure of the image). While global brightness differences that do not saturate pixel values do not affect the model output, variations in global contrast will yield different filter values. That is, the same image at different contrast levels will alter the output of the filtering operation, since the Gabor filtering will be sensitive to differences in the structure of edges in the images. In the filtering stage, the input image matrix and the kernel (also represented as a matrix) are convolved by multiplying the kernel in the Fourier domain with the Fourier-transformed image and reverting the product into the spatial

domain via an inverse two-dimensional FFT (performed analogously to the forward 2-D FFT described above). In this domain, the real and imaginary filtered image values are segregated for the final dissimilarity computation. Magnitude is also explicitly calculated at this stage and stored for visualization later on. Phase values are computed and displayed in a table upon user request. The real and imaginary filter values are extracted at specific locations in the image—in this case, those positions are defined by a 10×10 grid of equally spaced points, although more advanced versions of this model utilize “fiducial” points that are automatically centered on facial landmarks, as we discussed earlier. Thus, 80 values are extracted from each of the 100 positions, yielding a matrix that, when collapsed, becomes an 8,000-value vector.

Visualization

For expository simplicity, we have chosen to illustrate the computations as complex cells that are invariant to the direction of contrast. However, given that the app will likely be used to scale faces—the perception of which is highly sensitive to the direction of the contrast (see, e.g., Biederman &

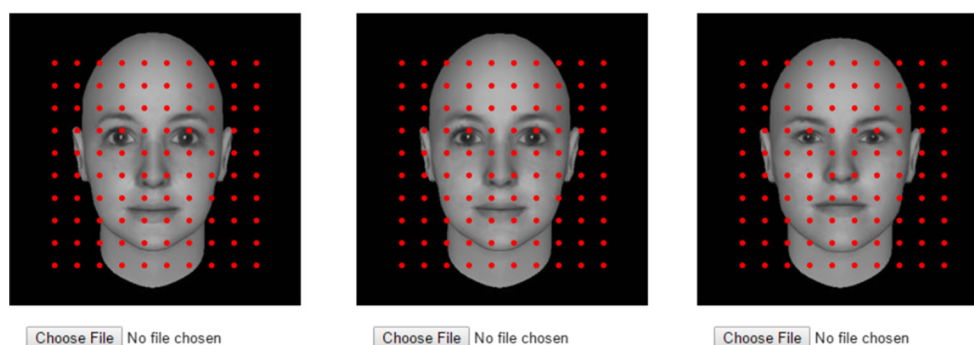


Fig. 5 Grayscale images of the stimuli from Fig. 4. The pixel intensity is taken as the scalar average of the red, green, and blue color layers of the original image. The grids of dots overlaid on the images indicate the image locations where each of the 100 jets will be centered

Kernel A

Row	Column	Orientation (°)	Scale (px)
8	5	112.5	24

Kernel B

Row	Column	Orientation (°)	Scale (px)
6	7	67.5	16

Kernel C

Row	Column	Orientation (°)	Scale (px)
5	5	90	32

Kernel D

Row	Column	Orientation (°)	Scale (px)
6	4	22.5	32

Kernel E

Row	Column	Orientation (°)	Scale (px)
4	7	90	16

Show Kernels

Fig. 6 The menu with which users specify which kernels will be visualized and have their magnitudes displayed in bar graphs. “Row” and “Column” specify the center of the kernel’s receptive field, “Orientation” dictates the angle at which the kernel is rotated, and

“Scale” dictates a kernel’s receptive field size. Although only values from these five kernels are plotted, all 4,000 kernels are used in the dissimilarity computation

Kalocsai, 1997)—the final computed similarities are based on simple cell tuning, which preserves the direction of contrast. The user is prompted to select the parameters (location, receptive field diameter in pixels, and orientation in degrees) of five kernels (Fig. 6). Parameter selection rows are color-coded to facilitate matching between the parameter selection stage and the following kernel visualization stage. The selected kernels are visualized and superimposed over the uploaded images (Fig. 7), and values extracted from convolutions using these five kernels are represented in a bar chart, thus representing

five out of the 4,000 combinations used by the applet (Fig. 8). This chart visualization allows the user to see where the images may differ with respect to the selected kernels. Although only values from these five kernels are plotted, all 8,000 kernels are used in the dissimilarity computation. Phase information is optionally available to the user in a table below the magnitude bar chart.

An aspect of the representation of faces, readily appreciated from a perusal of Fig. 4, is that we can often distinguish two similar faces without being able to articulate just what it is about

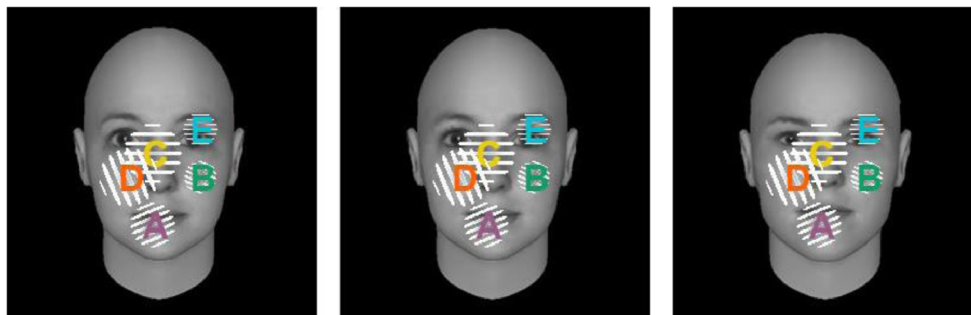


Fig. 7 The three default face images (from Fig. 4) with user-selected Gabor kernels (from Fig. 6) superimposed over the images. Colored labels facilitate matching between this visual representation and the parameter selection stage

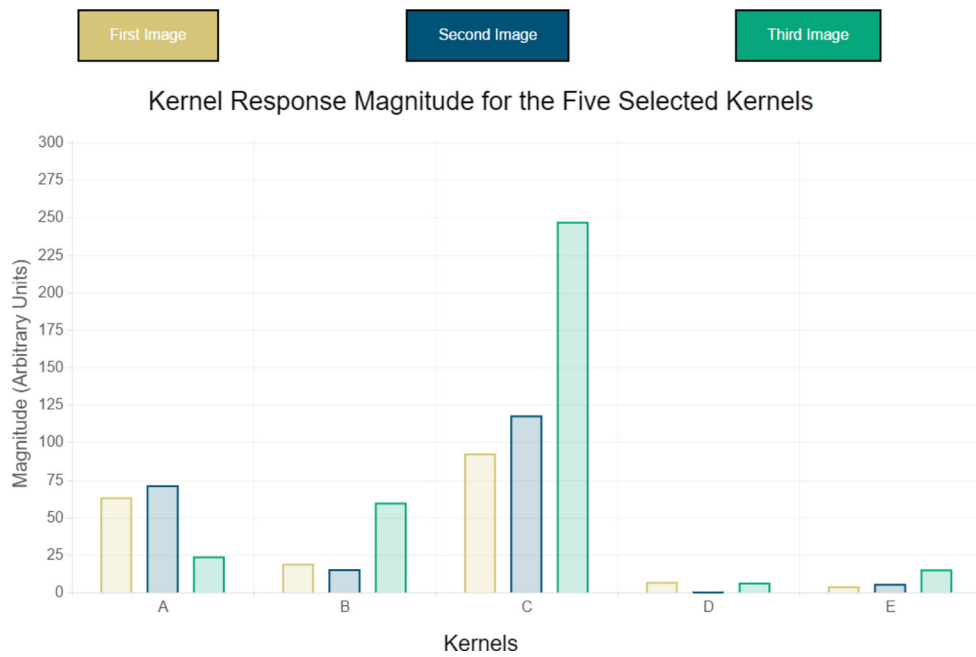


Fig. 8 Chart generated by the browser, showing convolution values from the sample kernels specified by the user during the “kernel selection” step. The kernel parameters from Fig. 6 were used in this visualization. For each of the three images, five values (each representing the response magnitude of a single kernel) are selected from the vector of 4,000 values and displayed in the bar graph. Note that the first two images, which

appear perceptually similar, share a common vector profile (i.e., they tend to share common output values for the default kernel display settings). The perceptual dissimilarity of the third image is also captured by this reduced vector comparison (its output values are markedly different from those of the first two images, especially for kernel C)

the faces that differs. That is, the differences between similar faces are *ineffable* (Biederman & Kalocsai, 1997). The ineffability of discriminability seems to be specific to faces and rarely

characterizes perceptible differences among objects (which are typically coded by their edges). A possible reason for this is that faces, but not objects, may retain aspects of the original spatial filtering (Yue et al., 2006)—activation of Gabor-like kernels, in the present case—and these kernels are not directly available to consciousness. The faces in Fig. 4 vary in the vertical distances between the eyes, nose, and mouth and in the heights of the cheekbones. The horizontally oriented Kernel C is directly affected by the variation in vertical distance, and we can see in Fig. 8 that it strongly signals that Face 3 differs from Faces 1 and 2. The output of the kernels can thus signal differences without an awareness of the particular kernels signaling that difference.

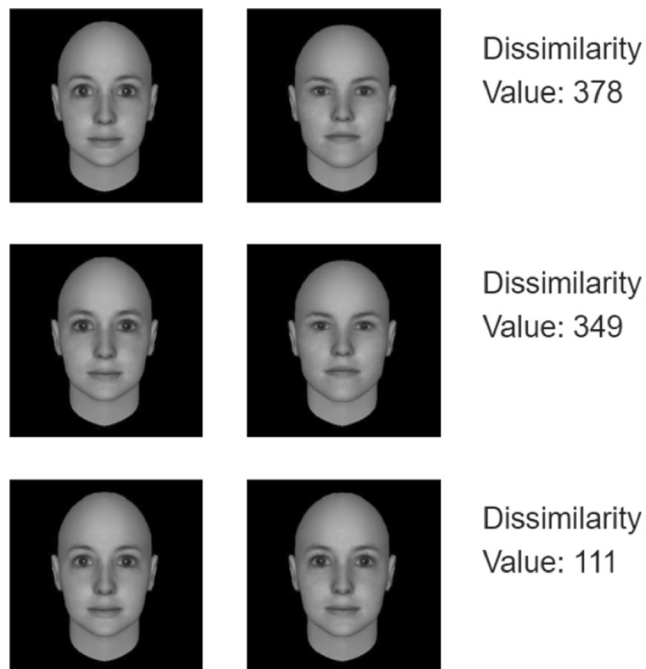


Fig. 9 Output of the application: image pairs ranked by dissimilarity. Consistent with subjective judgments, the top and middle pairs have high dissimilarity, whereas the dissimilarity rating for the bottom pair is markedly lower

The real and imaginary vectors for each image (which implicitly track magnitude and phase, respectively) are concatenated, yielding a single vector of 8,000 values for each image. The Euclidean distance between any two of these vectors is taken as a measure of the dissimilarity between the images. That is, the higher the Euclidean distance, the more dissimilar the Web app finds that pair of images. In this way, image pairs are ranked by similarity and presented to the user (Fig. 9).

Discussion

When learning about visual neuroscience, a typical student is provided a detailed explanation of low-level vision—beginning with retinal optics and extending to simple and complex

cells in V1—before jumping to topics such as object or face recognition. Often neglected, however, is an explicit account of the ways in which the functions of the early visual system give rise to representations computed in later stages of the visual pathway. Textbooks often reflect this chasm between understandings of low-level and high-level visual processing. To this end, educational tools that help demystify these processes may have significant didactic potential. The Gabor-jet model Web application, for example, illustrates the process by which V1 simple cell activation is used to differentiate similar metrically varying objects such as faces. The psychophysical similarity of faces, as well as of objects that vary metrically, such as the blobs in Fig. 3, can be predicted from the Gabor-jet model (Yue et al., 2012). However, only the neural coding of faces—but not blobs—retains aspects of the initial spatial coding, in that their release from adaptation in the fusiform face area depends on a change in the specific combinations of scale and spatial frequency values (Yue et al., 2006). Xu, Biederman, and Shah (2014) showed how face configural effects, which had previously defied neurocomputational explanation, could readily be derived from the action at a distance afforded by kernels with large, overlapping receptive fields. Furthermore, an understanding of V1-like convolution algorithms would provide the user with a strong foundation from which to understand more recent and intricate algorithms.

Above and beyond its didactic value, the Gabor-jet Web application has methodological utility. Researchers can take advantage of the interface of the simplified Web model, designed to be user-friendly, to test stimuli before employing the full MATLAB model available at http://geon.usc.edu/GWTgrid_simple.m. Unlike the Web app described here, the full-featured MATLAB code offers a wider range of parameters and lends itself more readily to batch processing, which is often necessary for stimulus scaling. Thus, the Web application can be considered an introduction to the Gabor-jet model that may encourage more frequent use of this valuable scaling system.

Author note This research was supported by NSF Grant No. BCS 0617699 and by the Dornsife Research Fund. Our 2-D FFT code utilizes a 1-D FFT implementation by Nayuki, which can be found at <https://www.nayuki.io/page/free-small-fft-in-multiple-languages>. Our code for resizing images, along with a fix for iPhone camera images, uses an image-rendering library by Shinichi Tomita, <https://github.com/stomita/ios-imagefile-megapixel>. Line charts were created with Chart.js, www.chartjs.org/.

References

- Amir, O., Biederman, I., & Hayworth, K. J. (2012). Sensitivity to nonaccidental properties across various shape dimensions. *Vision Research*, *62*, 35–43.
- Biederman, I., & Kalocsai, P. (1997). Neurocomputational bases of object and face recognition. *Philosophical Transactions of the Royal Society B*, *352*, 1203–1219. doi:10.1098/rstb.1997.0103
- De Valois, R. L., & De Valois, K. K. (1990). *Spatial vision*. New York, NY: Oxford University Press.
- Günther, M., Haufe, D., & Würtz, R. (2012). Face recognition with disparity corrected Gabor phase differences. In A. E. P. Villa, W. Duch, P. Erdi, F. Masulli, & G. Palm (Eds), *Artificial neural networks and machine learning—ICANN 2012, Part 1* (Lecture Notes in Computer Science Vol. 7552, pp. 411–418). Berlin, Germany: Springer.
- Jahanbin, S., Choi, H., Jahanbin, R., & Bovik, C. A. (2008). Automated facial feature detection and face recognition using Gabor features on range and portrait images. In *Proceedings of the 15th IEEE International Conference on Image Processing* (pp. 2768–2771). Piscataway, NJ: IEEE Press.
- Jones, J. P., & Palmer, L. A. (1987). The two-dimensional spatial structure of simple receptive fields in cat striate cortex. *Journal of Neurophysiology*, *58*, 1187–1211.
- Kay, K. N., Naselaris, T., Prenger, R. J., & Gallant, J. L. (2008). Identifying natural images from human brain activity. *Nature*, *452*, 352–355.
- Kim, J. G., & Biederman, I. (2012). Greater sensitivity to nonaccidental than metric changes in the relations between simple shapes in the lateral occipital cortex. *NeuroImage*, *63*, 1818–1826. doi:10.1016/j.neuroimage.2012.08.066
- Kim, J. G., Biederman, I., Lescroart, M. D., & Hayworth, K. J. (2009). Adaptation to objects in the lateral occipital complex (LOC): Shape or semantics? *Vision Research*, *49*, 2297–2305.
- Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). ImageNet classification with deep convolutional neural networks. In *Advances in Neural Information Processing Systems* (pp. 1097–1105). Cambridge, MA: MIT Press.
- Lades, M., Vorbruggen, J. C., Buhmann, J., Lange, J., von der Malsburg, C., Wurtz, R. P., & Konen, W. (1993). Distortion invariant object recognition in the dynamic link architecture. *IEEE Transactions on Computers*, *42*, 300–311. doi:10.1109/12.210173
- Lescroart, M. D., & Biederman, I. (2012). Cortical representation of medial axis structure. *Cerebral Cortex*, *23*, 629–637.
- Margalit, E., Shah, M. P., Tjan, B. S., Biederman, I., Keller, B., & Brenner, R. (in press). The lateral occipital complex shows no net response to object familiarity. *Journal of Vision*.
- Oliva, T., & Torralba, A. (2001). Modeling the shape of the scene: A holistic representation of the spatial envelope. *International Journal of Computer Vision*, *42*, 145–175.
- Olshausen, B. A., & Field, D. J. (1996). Emergence of simple-cell receptive field properties by learning a sparse code for natural images. *Nature*, *381*, 607–609.
- Riesenhuber, M., & Poggio, T. (1999). Hierarchical models of object recognition in cortex. *Nature Neuroscience*, *2*, 1019–1025. doi:10.1038/14819
- Ringach, D. L. (2002). Spatial structure and symmetry of simple-cell receptive fields in macaque primary visual cortex. *Journal of Neurophysiology*, *88*, 455–463.
- Serre, T., & Riesenhuber, M. (2004). *Realistic modeling of simple and complex cell tuning in the HMAX model, and implications for invariant object recognition in cortex* (Technical Report No. AI-MEMO-2004-017). Massachusetts Institute of Technology, Cambridge Computer Science and Artificial Intelligence Lab.
- Wilson, H. R., McFarlane, D. K., & Phillips, G. C. (1983). Spatial frequency tuning of orientation selective units estimated by oblique masking. *Vision Research*, *23*, 873–882.
- Wiskott, L., Fellous, J. M., Krüger, N., & von der Malsburg, C. (1997). Face recognition by elastic bunch graph matching. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *19*, 775–779.
- Xu, X., & Biederman, I. (2010). Loci of the release from fMRI adaptation for changes in facial expression, identity, and viewpoint. *Journal of Vision*, *10*(14), 36:1–13. doi:10.1167/10.14.36

- Xu, X., Biederman, I., & Shah, M. P. (2014). A neurocomputational account of the face configural effect. *Journal of Vision*, *14*(8), 9:1–9. doi:[10.1167/14.8.9](https://doi.org/10.1167/14.8.9)
- Xu, X., Yue, X., Lescroart, M. D., Biederman, I., & Kim, J. G. (2009). Adaptation in the fusiform face area (FFA): Image or person? *Vision Research*, *49*, 2800–2807.
- Yue, X., Biederman, I., Mangini, M. C., von der Malsburg, C., & Amir, O. (2012). Predicting the psychophysical similarity of faces and non-face complex shapes by image-based measures. *Vision Research*, *55*, 41–46. doi:[10.1016/j.visres.2011.12.012](https://doi.org/10.1016/j.visres.2011.12.012)
- Yue, X., Tjan, B. S., & Biederman, I. (2006). What makes faces special? *Vision Research*, *46*, 3802–3811.