Current Biology, Vol. 14, 391–396, March 9, 2004, ©2004 Elsevier Science Ltd. All rights reserved. DOI 10.1016/j.cub.2004.02.028

Inversion Leads to Quantitative, Not Qualitative, Changes in Face Processing

**Allison B. Sekuler,[1,]* Carl M. Gaspar,[1]
Jason M. Gold,[2] and Patrick J. Bennett[1]**
[1]Department of Psychology
McMaster University
Hamilton, Ontario L8S 4K1
Canada
[2]Department of Psychology
Indiana University
Bloomington, Indiana 47405

Summary

Humans are remarkably adept at recognizing objects across a wide range of views. A notable exception to this general rule is that turning a face upside down makes it particularly difficult to recognize [1–3]. This striking effect has prompted speculation that inversion qualitatively changes the way faces are processed. Researchers commonly assume that configural cues strongly influence the recognition of upright, but not inverted, faces [3–5]. Indeed, the assumption is so well accepted that the inversion effect itself has been taken as a hallmark of qualitative processing differences [6]. Here, we took a novel approach to understand the inversion effect. We used response classification [7–10] to obtain a direct view of the perceptual strategies underlying face discrimination and to determine whether orientation effects can be explained by differential contributions of nonlinear processes. Inversion significantly impaired performance in our face discrimination task. However, surprisingly, observers utilized similar, local regions of faces for discrimination in both upright and inverted face conditions, and the relative contributions of nonlinear mechanisms to performance were similar across orientations. Our results suggest that upright and inverted face processing differ quantitatively, not qualitatively; information is extracted more efficiently from upright faces, perhaps as a by-product of orientation-dependent expertise.

Results and Discussion

Despite its wide acceptance, the configural/featural distinction in face recognition has not led to the development of precise models for this process, in large part because of a lack of agreement on what comprises configural processing and local features [11–12]. Here, rather than starting with assumptions about the kinds of configural cues and/or local features that may be important for face recognition, our approach assumed that an ensemble of processes encodes various aspects of faces and that the outputs of these processes can be used in a flexible manner to solve different tasks (e.g., identification, gender discrimination, and recognition of emotional expression). Furthermore, we assumed these processes could be represented as a set of quasilinear and nonlinear filters. The response classification technique [7–8] enabled us to estimate the influences of these filters on observers' responses in face discrimination. In response classification, external noise is added to stimuli that the observer must classify (e.g., face A or face B). Then, from the trial-by-trial variation in the observer's responses, one can determine how noise in different parts of the stimulus image biases the observer toward a specific response. Imagine that the observer's task is to discriminate face A from face B when each face is embedded in white Gaussian external noise that varies across trials. In some trials, the observer's classifications will be correct. However, in other trials the noise may make one face look more like the other and thereby lead to incorrect classifications. One simple analysis averages all the noise fields within a given stimulus response class: N_{AA}, N_{AB}, N_{BA}, and N_{BB}. For example, N_{AB} is the average of all noise fields in trials in which stimulus A was presented and the observer responded "B." One can then calculate the classification image as follows [9–10]:

$$C = (N_{AA} + N_{BA}) - (N_{BB} + N_{AB}) \qquad (1)$$

The classification image C described by Equation 1 is an estimate of the linear calculation that affects performance [9–10], but response classification can also be generalized to estimate the effects of nonlinear mechanisms [13].

The classification image can be thought of as a behavioral receptive field [14] that reveals the stimulus regions in which noise consistently affects responses. Regions that appear to be gray in the classification image are locations where noise has no consistent effect on responses; regions that appear to be black or white are locations where noise leads to consistent response biases. Thus, response classification can be thought of as showing the parts of the stimulus an observer uses to make a decision. In this sense, response classification has a similar goal as the Bubbles technique [15], in which one estimates influential stimulus regions by tracking responses made to distinct areas of a stimulus (see [16–18] for a discussion of when response classification and Bubbles may be appropriate techniques).

Previously, response classification was used for investigating tasks ranging from stimulus detection [13, 19] and disparity discrimination [20] to perceptual organization [14], attention [21], and learning [22]. By applying this technique to the discrimination of upright and upside-down faces, we can obtain a direct view of how the previously invisible processing strategies differ as a function of orientation.

Classification Images

In each trial, observers viewed a single face embedded in noise. This was followed by a response window displaying noise-free, high-contrast versions of two faces (see Figure 1). The observer's task was to select the

*Correspondence: sekuler@mcmaster.ca

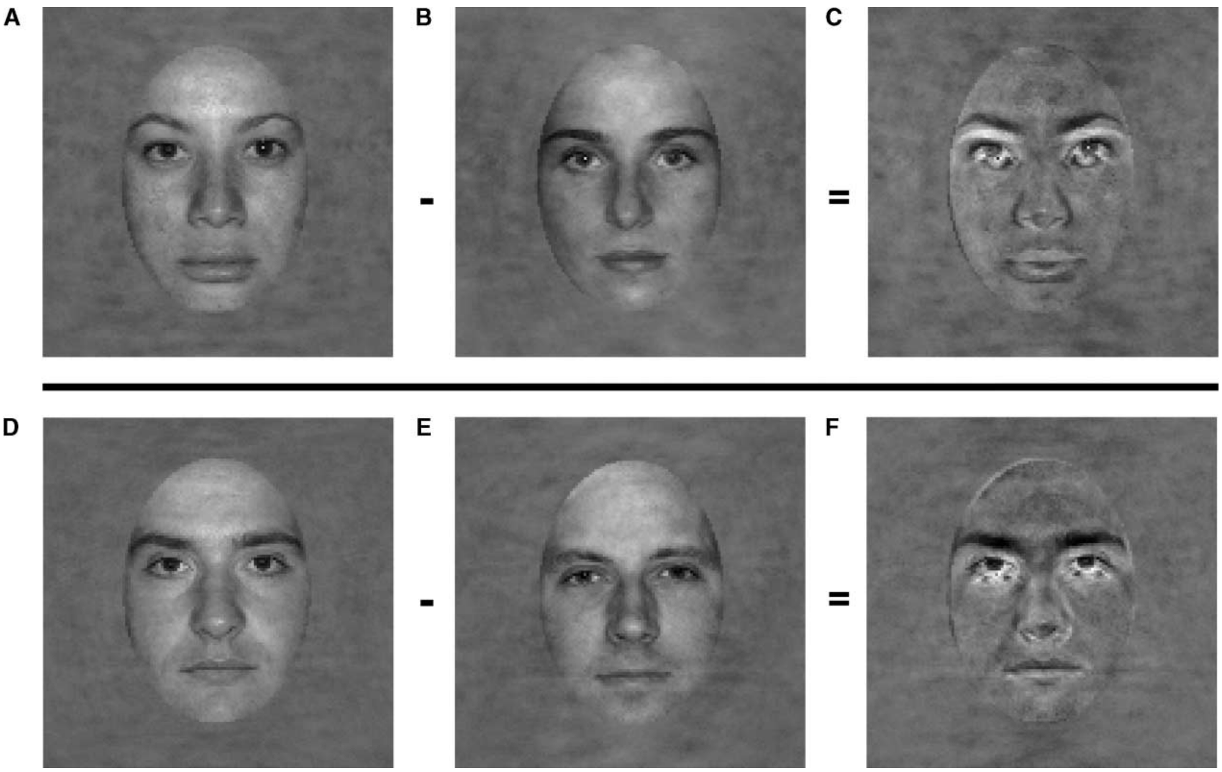

Figure 1. Face Stimuli and Ideal Observer Templates

Stimuli were one pair of female faces (A and B) and one pair of male faces (D and E). The most discriminative regions within each face pair are shown in the ideal observer classification images (C and F).

face presented in each trial. The combination of two factors (upright vs. inverted and male vs. female faces) led to four conditions, which were tested in different blocks. Contrast detection thresholds were determined for 71% correct performance, and noise fields from each trial were stored to derive classification images (see Experimental Procedures for details). Inverted face discrimination required 53% more contrast than upright face discrimination; root mean square (RMS) contrast thresholds for upright and inverted faces were 0.0172 and 0.0264, respectively [$t(8) = 3.20$, $p < 0.05$]. Thus, despite the fact that only two faces were shown in each condition, observers exhibited significant inversion effects similar to those expected from experiments with multiple faces. This result indicates that it is unlikely that observers were treating the stimuli like random patches of dark and light in an image-matching task because one would not expect, a priori, to find an inversion effect when simply matching one meaningless texture to another.

Visual inspection of the classification images (Figure 2) suggests that only highly localized regions of the stimulus, primarily around the eyes and eyebrows, have consistently affected observers' responses. Although it may seem surprising that such a restricted region would influence face perception, this result is consistent with recent results from studies regarding face perception [23–25]. The region around the eyes and eyebrows is also the region in which our face stimuli differ most from one another and therefore contains most of the available discrimination information. Thus, if observers can encode information from only a portion of the face, then the most efficient strategy for our stimulus set and task would be to focus on the eyes and eyebrows.

To quantify the analyses, we determined both the number of pixels that correlated significantly with an observer's response in each classification image and the normalized cross correlation of the observer's classification image with the classification image for an ideal discriminator. For our task, the ideal classification image was simply the pixel-by-pixel difference between the two faces being discriminated. The number of significant pixels indicates how much of the stimulus observers used in each condition, whereas the correlation analysis indicates how efficiently observers used that information. As seen in Figure 3, the number of significant pixels did not vary consistently with stimulus orientation [$t(8) = 1.61$, not significant] and those pixels generally were confined to the eye/brow region (blue rectangles, as described in Figure 2). Across all observers and face gender, an average of 75.4% and 76.8% of the significant pixels fell within this region for upright and inverted faces, respectively. However, the correlation between real and ideal classification images was always higher for upright stimuli; correlations were on average 1.72 times greater for upright faces than for inverted faces [$t(8) = 8.29$, $p < 0.0001$]. Thus, although observers generally used about the same number of highly localized

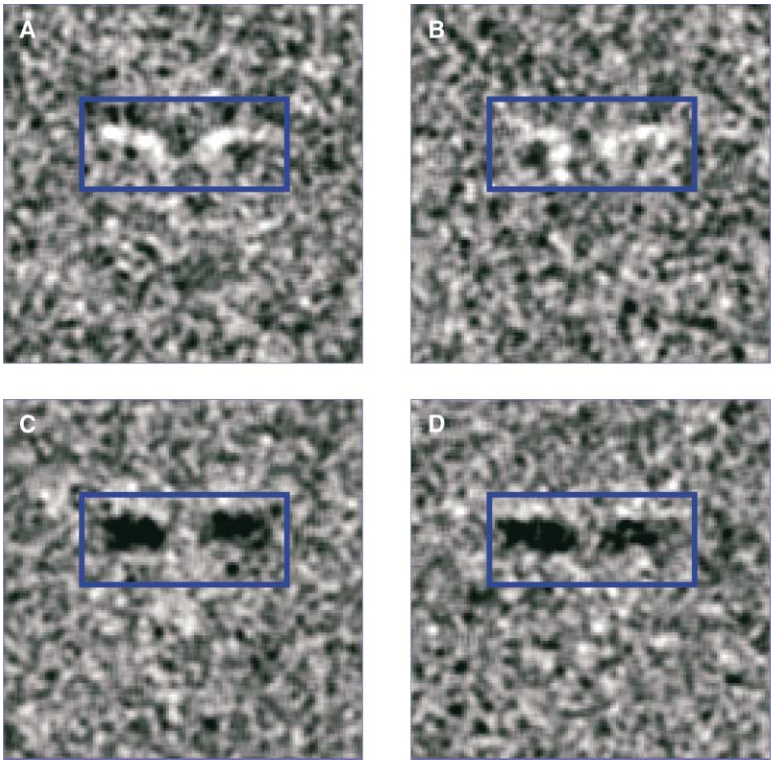

Figure 2. Smoothed Classification Images
Sample classification images for female faces (ALG [A and B]) and male faces (KAT [C and D]). The left column shows results for upright faces, and the right column shows results for inverted faces (rotated 180°). For computation of smoothed classification images, raw classification images were convolved with a 5×5 convolution kernel (the outer product of $[1, 2, 3, 2, 1]^T$). Blue rectangles show the 71×29 pixel region surrounding the eyes and eyebrows in the original face images (12.6% of the total pixels).

pixels in each case, the information contained in those pixels was used much more efficiently for upright faces than for inverted faces.

Nonlinear Contributions

Our interpretation of classification images, as defined by Equation 1, is limited by the fact that the technique reveals only the linear association between each pixel's contrast and the observer's responses and is not sensitive to nonlinearities. However, the contribution of nonlinearities can be estimated indirectly from classification images [26]. If, in each trial, the decision of an observer is affected only by linear mechanisms, then the observer's absolute efficiency (i.e., measured performance relative to the best possible performance of an ideal discriminator) can be predicted from the obtained classification image. If absolute efficiency is higher than the predicted value, then nonlinear mechanisms whose influence is not captured by the classification image must have contributed to the observer's behavior. Figure 4 shows the relationship between predicted and obtained values of absolute efficiency for upright and inverted conditions. The results point to two important conclusions. First, there is a strong relationship between predicted and obtained absolute efficiencies; observed efficiency was correlated with predicted efficiency [$r^2 = 0.76$], and the addition of face orientation as an additional, binary predictor variable did not significantly improve the fit of the regression model. Thus, variation in the structure within classification images is strongly correlated with differences in thresholds obtained with upright and inverted faces. Second, observed efficiency was slightly but consistently greater than predicted efficiency, thereby suggesting that nonlinear processes contributed to performance in our task. Interestingly, the fact that deviations from the predicted efficiency were similar for upright and inverted faces suggests that the contributions to performance that can be attributed to such nonlinear processes were similar in the two conditions (although it remains possible that different nonlinear processes are involved for faces viewed at different orientations).

Conclusions

Like previous researchers, we found that observers were more sensitive to differences between upright faces than between inverted faces. Such a result is typically interpreted to mean that upright faces are processed configurally, whereas the processing of upside-down faces is based primarily on features. However, our results do not support the conclusion that qualitatively distinct modes of processing are used in the two conditions. In our experiments, performance differences between upright and inverted faces were correlated with the structure contained in linear classification images, and observers used highly localized regions near the eyes to discriminate faces at both orientations. Although nonlinear mechanisms contributed to performance, the magnitude of this contribution was small and similar for both upright and inverted faces. Overall, our results suggest that the primary difference between processing upright and inverted faces is quantitative rather than qualitative. Discriminative regions are processed more efficiently in upright faces than in upside-down faces.

This view differs from the standard view of configural

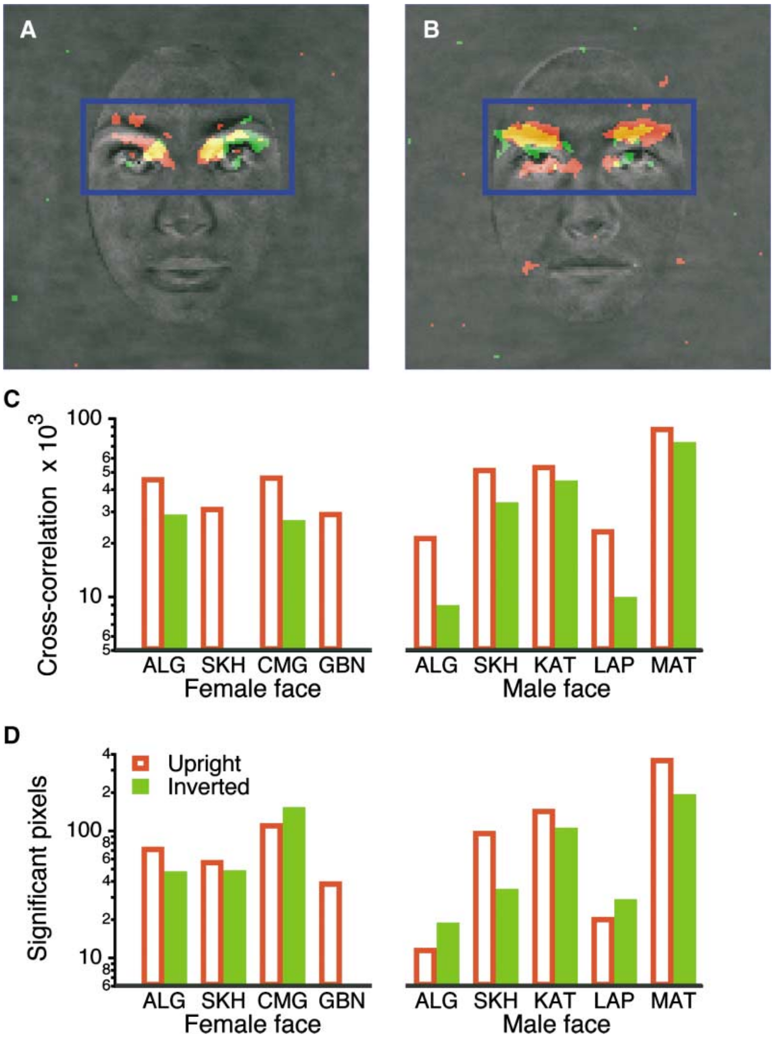

Figure 3. Significant Pixels and Cross-Correlation Results

Statistically significant pixels from classification images derived across all trials for all observers for (A) female and (B) male faces superimposed on a low-contrast version of the ideal observer classification image. Red pixels indicate significance ($p < 0.001$) for upright faces, green for inverted, and yellow for both. (C) Correlations between human and ideal classification images differed significantly across orientation. Open red bars show results for upright faces, filled green bars for inverted. Missing bars represent values at (SKH) or below (GBN, -0.002) the abscissa. (D) There were no consistent differences in the number of significant pixels across orientation. Symbols are as in (C). The missing bar represents a value at the abscissa.

versus part-based processing, but it is consistent with results from some recent behavioral and neuroimaging studies. For example, inversion effects have been found for the isolated eye/brow region of the face [24, 27]. These results suggest that the inversion effect does not require global processing across the entire face. Observers can also recognize upright faces in the presence of large geometric distortions, suggesting that face processing strategies are not based on the relative positions of discrete facial features in any simple way [28]. In addition, studies with fMRI show preferential activation in the so-called fusiform face area (FFA) for both upright and inverted faces compared to other objects such as houses [29]. All of these results support the idea that upright and inverted faces engage similar neural mechanisms. Although it is possible that different populations of neurons lead to these responses for upright and inverted faces, it is equally possible that both types of stimuli lead to activation of the same expertise-related mechanisms [30–31]. In this latter context, the advantage for processing upright faces may simply be a by-product of relative expertise levels. For example, previous researchers have shown that perceptual learning can be quite specific (e.g., to stimulus orientation [32]), and past research has also shown that perceptual learning increases processing efficiency [22, 33]. Throughout our lives, we certainly have more experience recognizing upright faces than upside-down faces, and this increased experience may lead to increased processing efficiency.

It remains an empirical question whether our approach can account for other results typically explained in terms of configural processing for upright faces (e.g., the Thatcher Illusion [34], the Composite Effect [35]). For example, the failure to recognize grotesque expressions in inverted, Thatcherized faces may be due to decreased ability to extract relevant emotional information from inverted faces (similar to the decreased ability to extract relevant identity information from inverted faces). However, it also remains possible that nonlinearities play a bigger role in such tasks and, more generally, that nonlinear processes are more important at suprathreshold signal-to-noise ratios.

Experimental Procedures

Data were collected in two laboratories via slightly different methods (versions A and B). However, the results were qualitatively and quan-

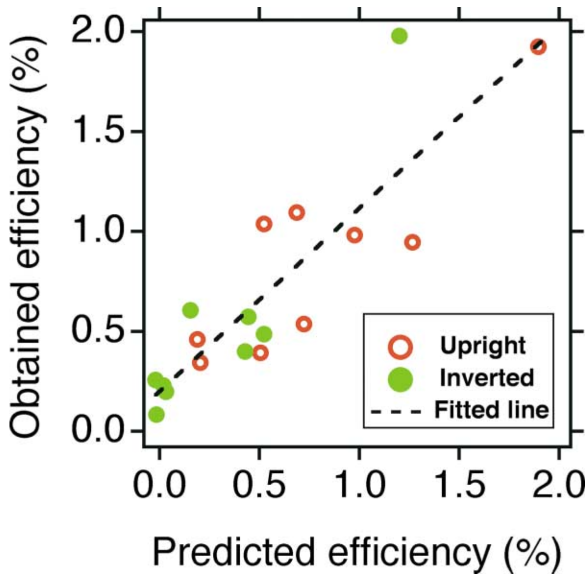

Figure 4. Relationship between Predicted and Obtained Absolute Efficiencies

Each point represents results from a single observer's classification image for upright (open red) and inverted (filled green) face discrimination. Observed efficiency was strongly correlated with predicted efficiency, and the relationship was constant across orientations.

titatively similar and so were combined. Both versions used Matlab with the Psychophysics Toolbox extensions [36, 37] to control stimulus presentation and response collection.

Stimuli

Each face was 99 pixels high and ranged from 62 to 90 pixels wide. Average luminance was 13.3 cd/m^2. Screen resolution was always 640 $\times$ 480 pixels, but pixel size was approximately 2.5 and 4.2 minarc for versions A and B, respectively. At a viewing distance of 100 cm, the height of each face subtended 4.1° or 6.9°, and the width ranged from 2.6° to 3.7° or 4.3° to 6.2° for versions A and B, respectively. Each face was centered within a 128 $\times$ 128 pixel array (5.3° $\times$ 5.3° or 8.9° $\times$ 8.9° for versions A and B, respectively). Additional details about stimulus generation can be found elsewhere [23]. For derivation of classification images, static, white Gaussian noise was added to each face (noise root-mean-square contrast: 0.14%). A new noise field covering the entire stimulus was generated in each trial. Half of the blocks contained two female faces, and half of the blocks contained two male faces; within each block, the face orientation was held constant, and two different conditions were tested for each face gender: upright and inverted.

Procedure

Three females (average age: 21 yr) participated in version A; one female and three males (average age: 24 yr) participated in version B. All observers except C.M.G. were naïve in regards to the purpose of the experiment and were unfamiliar with the face stimuli at the beginning of the experiment. All observers had normal or corrected to normal vision. In each trial, one of two faces was randomly selected as the target. Each trial consisted of an initial 1 s fixation on a small spot, followed by a face plus noise stimulus displayed for 500 ms. Finally, a selection window containing noise free, high-contrast versions of two faces was displayed. The subject indicated which face had served as the stimulus by clicking on that face with a computer mouse. The selection window remained visible until this response, and response accuracy was indicated by auditory feedback.

Version A

Orientation (upright and inverted) and gender (male and female) of face stimuli were crossed in a within-subjects design. All participants completed 10,000 trials in each condition, and each classification image was based on all 10,000 trials. There were 4–6 sessions per condition (500–2500 trials per session). Throughout the session, stimulus contrast was adjusted with a staircase procedure so that response accuracy remained at approximately 71% correct. Results from one observer's male conditions are not included in the analyses because a computer error led to the presentation of incorrect contrast values in those conditions.

Version B

Two observers were tested with female faces and two with male faces. Each observer was tested with both upright and inverted faces. All observers received several thousand practice trials in the face discrimination tasks before the main experiment. Each testing session consisted of 2150 trials. During the first 150 trials, stimulus contrast was adjusted with a staircase procedure so that the contrast needed to produce 71% correct responses could be estimated. Stimulus contrast remained at this value for the session's remaining 2000 trials, of which results were used to calculate the classification image. Each classification image was based on a total of 10,000 trials.

Acknowledgments

This research was supported by grants from the Natural Sciences and Engineering Research Council of Canada, the Canada Research Chair Council, the Canada Foundation for Innovation, and the Ontario Innovation Trust. We thank Richard Murray, Robert Sekuler, Denis Pelli, and an anonymous reviewer for their many helpful suggestions and comments.

Received: November 13, 2003
Revised: January 20, 2004
Accepted: January 20, 2004
Published: March 9, 2004

References

1. Yin, R.K. (1969). Looking at upside-down faces. J. Exp. Psychol. *81*, 141–145.
2. Diamond, R., and Carey, S. (1986). Why faces are and are not special: an effect of expertise. J. Exp. Psychol. Gen. *115*, 107–117.
3. Valentine, T. (1988). Upside-down faces: a review of the effect of inversion upon face recognition. Br. J. Psychol. *79*, 471–491.
4. Tanaka, J.W., and Farah, M.J. (1993). Parts and wholes in face recognition. Q. J. Exp. Psychol. A *46*, 225–245.
5. Murray, J.E., Yong, E., and Rhodes, G. (2000). Revisiting the perception of upside-down faces. Psychol. Sci. *11*, 492–496.
6. Reed, C., Stone, V.E., Bozova, S., and Tanaka, J. (2003). The body-inversion effect. Psychol. Sci. *14*, 302–308.
7. Ahumada, A., Jr., and Lovell, J. (1971). Stimulus features in signal detection. J. Acoust. Soc. Am. *49*, 1751–1756.
8. Beard, B.L., and Ahumada, A.J. (1988). A technique to extract relevant image features for visual tasks. In SPIE Proceedings: vol. 3299. Human Vision and Electronic Imaging III, B.E. Rogowitz, and T. N. Pappas T.N., eds. (Bellingham, WA: SPIE), pp. 79–85.
9. Murray, R.F., Bennett, P.J., and Sekuler, A.B. (2002). Optimal methods for calculating classification images: weighted sums. J. Vis. *2*, 79–104, http://journalofvision.org/2/1/6/, doi:10.1167/2.1.6.
10. Ahumada, A.J. (2002). Classification image weights and internal noise level estimation. J. Vis. *2*, 121–131, http://journalofvision.org/2/1/8/, doi:10.1167/2.1.8.
11. Farah, M.J., Wilson, K.D., Drain, M., and Tanaka, J.N. (1998). What is "special" about face perception? Psychol. Rev. *105*, 482–498.
12. Maurer, D., Grand, R.L., and Mondloch, C.J. (2002). The many faces of configural processing. Trends Cogn. Sci. *6*, 255–260.
13. Neri, P., and Heeger, D.J. (2002). Spatiotemporal mechanisms for detecting and identifying image features in human vision. Nat. Neurosci. *5*, 812–816.

14. Gold, J.M., Murray, R.F., Bennett, P.J., and Sekuler, A.B. (2000). Deriving behavioural receptive fields for visually completed contours. Curr. Biol. *10*, 663–666.
15. Gosselin, F., and Schyns, P.G. (2001). Bubbles: a technique to reveal the use of information in recognition tasks. Vision Res. *41*, 2261–2271.
16. Murray, R.F., and Gold, J.M. (2004). Troubles with bubbles. Vision Res. *44*, 461–470.
17. Gosselin, F., and Schyns, P.G. (2004). No troubles with bubbles: a reply to Murray and Gold. Vision Res. *44*, 471–477.
18. Murray, R.F., and Gold, J.M. (2004). Reply to Gosselin and Schyns. Vision Res. *44*, 479–482.
19. Solomon, J.A. (2002). Noise reveals visual mechanisms of detection and discrimination. J. Vis. *2*, 105–120, http://journalofvision.org/2/1/7, doi:10.1167/2.1.7.
20. Neri, P., Parker, A.J., and Blakemore, C. (1999). Probing the human stereoscopic system with reverse correlation. Nature *401*, 695–698.
21. Eckstein, M.P., Shimozaki, S.S., and Abbey, C.K. (2002). The footprints of visual attention in the Posner cueing paradigm revealed by classification images. J. Vis. *2*, 25–45, http://journalofvision.org/2/1/3/, doi:10.1167/2.1.3.
22. Gold, J., Sekuler, A.B., and Bennett, P.J. Characterizing perceptual learning with external noise. Cog. Sci., in press.
23. Gold, J., Bennett, P.J., and Sekuler, A.B. (1999). Identification of band-pass filtered letters and faces by human and ideal observers. Vision Res. *39*, 3537–3560.
24. Leder, H., Candrian, G., Huber, O., and Bruce, V. (2001). Configural features in the context of upright and inverted faces. Perception *30*, 73–83.
25. Sadr, J., Jarudi, I., and Sinha, P. (2003). The role of eyebrows in face recognition. Perception *32*, 285–293.
26. Murray, R.M. (2002). Perceptual organization and the efficiency of shape discrimination. Ph.D. Thesis, University of Toronto, Toronto, Canada.
27. Rakover, S.S., and Teucher, B. (1997). Facial inversion effects: parts and whole relationship. Percept. Psychophys. *59*, 752–761.
28. Hole, G.J., George, P.A., Eaves, K., and Rasek, A. (2002). Effects of geometric distortions on face-recognition performance. Perception *31*, 1221–1240.
29. Kanwisher, N., Tong, F., and Nakayama, K. (1998). The effect of face inversion on the human fusiform face area. Cognition *68*, B1–11.
30. Tarr, M.J., and Gauthier, I. (2000). FFA: a flexible fusiform area for subordinate-level visual processing automatized by expertise. Nat. Neurosci. *3*, 764–769.
31. Gauthier, I., Curran, T., Curby, K.M., and Collins, D. (2003). Perceptual interference supports a non-modular account of face processing. Nat. Neurosci. *6*, 428–432.
32. Gauthier, I., Williams, P., Tarr, M.J., and Tanaka, J. (1998). Training ‘greeble’ experts: a framework for studying expert object recognition processes. Vision Res. *38*, 2401–2428.
33. Gold, J., Bennett, P.J., and Sekuler, A.B. (1999). Signal but not noise changes with perceptual learning. Nature *402*, 176–178.
34. Thompson, P. (1980). Margaret Thatcher: a new illusion. Perception *9*, 483–484.
35. Young, A.W., Hellawell, D., and Hay, D.C. (1987). Configurational information in face perception. Perception *16*, 747–759.
36. Pelli, D.G. (1997). The VideoToolbox software for visual psychophysics: transforming numbers into movies. Spat. Vis. *10*, 437–442.
37. Brainard, D.H. (1997). The Psychophysics Toolbox. Spat. Vis. *10*, 443–446.