

Spatial summation of face information

Christopher W. Tyler

The Smith-Kettlewell Eye Research Institute,
San Francisco, CA, USA



Chien-Chung Chen

Department of Psychology, National Taiwan University,
Taipei, Taiwan



Do all parts of the face contribute equally to face detection or are some parts more detectable than others? To evaluate this issue, we studied detection of the presence of normalized frontal-face images within aperture windows of varying extent. We performed a face summation study using two-alternative forced-choice psychophysics. The face stimuli were scaled to equal eye-to-chin distance, centered on the bridge of the nose, and windowed by fourth-power Gaussian envelopes of various sizes. The faces were intermixed with control stimuli consisting inverted faces to test for configuration effects, split-half inverted faces to perturb the symmetry, and phase-scrambled versions of the faces with equal Fourier energy. Face detectability improved rapidly at first, then at a progressively shallower rate for larger window sizes, in a similar fashion for the three face-based stimulus types. The spectrally equated noise stimuli were less detectable than the face stimuli for all except the smallest apertures. The results were fit with a model incorporating global face-specific and local nonspecific spatial integration mechanisms. Detection of the noise images was consistent with local detection mechanisms accessed through a wide-field attention mechanism. The data for face detection implied detection mechanisms that integrated linearly up to some small size, integrated more slowly up to an intermediate size, and failed to gain any improvement for information beyond some larger size. This performance supports the concept of a specialized face configuration mechanism operating at detection threshold, similar in extent among the observers.

Keywords: detection threshold, face perception, attentional window, facial feature

Introduction

The perceptual processing of face images has been the subject of scientific study since the time of Charles Darwin and Francis Galton, in the mid-19th century. A key question that arises is whether there is specialized processing for the face configuration relative to other generic stimuli. Among the vast array of techniques, few support the idea of such specialized processing. For example, careful studies of the visual search for face images show no popout effect for faces (Brown, Huey, & Findlay, 1997; Nothdurft, 1993; Wolfe, 1998), particularly when the spatial-frequency content of the images is equated (Vanrullen, 2006). Interestingly, however, a face superiority effect has been identified for detection studies of faces against blank fields (Gorea & Julesz, 1990; Lewis & Edmonds, 2003; Purcell & Stewart, 1986, 1988; Rolls, Tovee, Purcell, Stewart, & Azzopardi, 1994; van Santen & Jonides, 1978). There appears to be something special about face visibility in the vicinity of detection threshold that reveals the special status of its processing.

Three levels of specialization can be envisaged: a global image-based face template, purely local feature templates, or a specialization for feature configuration. In the last case, the features are recognized by local (oriented) templates but their configuration (in terms of angles and

distances among the features) plays an important role in the detection and recognition processes. One could envisage that the relationships among eyes, nose, and mouth would be acceptable over some range of variation, a range large enough to render fixed templates ineffective. Indeed, the evidence suggests that visual processing is specialized for global relationships among the local features, but that these relationships are carried by configurational encoding rather than a generic face template. For example, Webster and MacLin (1999) report a profound figural aftereffect for adaptation to distorted faces, suggesting the operation of a generic face template against which the distortion was determined. However, the perceived distortion generalizes across orientation (Webster & MacLin, 1999) and size (Rhodes, Jeffery, Watson, Jaquet, Winkler, & Clifford, 2004; Zhao & Chubb, 2001), implying that the face configuration is encoded at a more abstract level than is implied by a simple face template.

The issue that remains to be determined, however, is the mechanism of the face superiority effect near detection threshold. What specialized processing for face information contributes to the detection of face images in comparison with other comparable images? One difficulty in this endeavor is to specify the other comparable images in a meaningful fashion. For example, if one compared the detection for faces and for random-dot noise, differences in detectability could be attributed to the differences in spatial

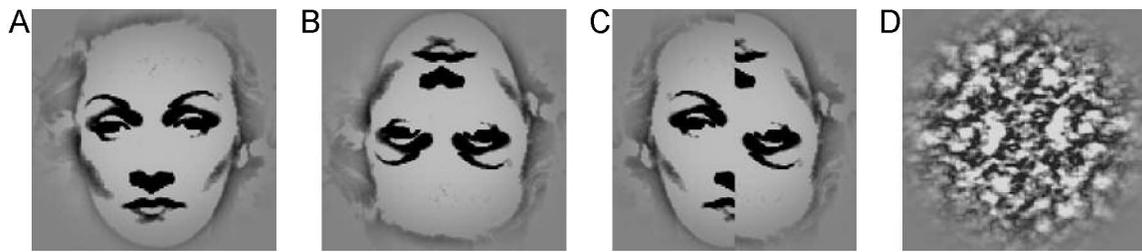


Figure 1. Sample stimuli. (A) High-pass filtered face image. (B) Fully inverted face image. (C) Split-half inverted face image. (D) Phase-randomized noise with matched spatial frequency spectrum and centric symmetry.

frequency spectrum between the two types. The comparison stimuli will therefore be equated for the spectral characteristics of the face stimuli (and the visual extent within which they are presented, see Figure 1D).

A second property of faces is that they are bilaterally symmetric (to first approximation). This symmetry imposes a pairwise constraint on the points in the image. We therefore wished to provide a corresponding constraint in the noise stimulus. However, the bilateral symmetry may be an important component of the face detectability. We therefore imposed a pairwise, but different, constraint on the noise—centric symmetry, where each point matches the point diametrically across the center of the pattern (equivalent to split-half inversion of the pattern). We also included two further control conditions; one in which the faces were fully inverted (Figure 1B), to determine whether the face inversion effect (Carey & Diamond, 1977; Thompson, 1980; Yin, 1969) operates at detection threshold, and a split-half inversion of the face images (Figure 1C) to match the centric symmetry of the noise patterns in the nonscrambled face images.

To evaluate the spatial processing properties at threshold, we varied the aperture in which the face stimuli were

visible (Figures 1 and 2). For the noise stimuli, the contrast energy formalism validated in previous studies (Watson, 2000) implies that detection sensitivity should increase uniformly with some fractional power of the area of the aperture. The presence of specialized face detection mechanisms would be revealed by a more rapid improvement in detectability with area for the face stimuli than for the noise stimuli. In particular, if there is linear summation of spatial information up to some extent, asymptoting to zero summation beyond that, it implies the presence of a single filter, template, or feature configuration module mediating the detection performance (Tyler & Chen, 2000a). This is the hypothetical form of the data that would reveal specialized processing for face information. Other forms representing more complex models of face processing are evaluated in Results and analysis.

The inverted faces (Figure 1B) were included to test the specificity of a face template governing threshold performance because they have the same local features arranged in an unfamiliar configuration. If performance is the same for the inverted configuration, it implies the existence of an orientation-independent template; a loss in detectability with inversion implies specificity for

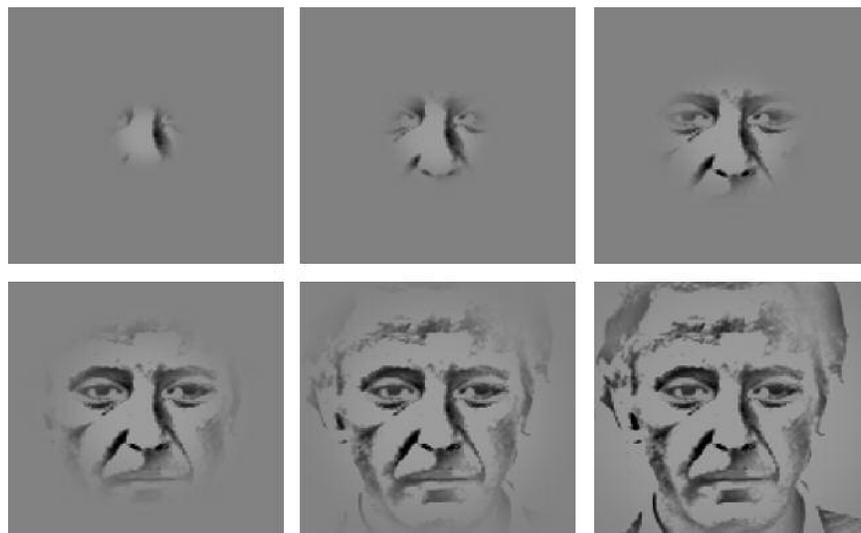


Figure 2. Depiction of the six sizes of the fourth-power Gaussian aperture in which the faces were shown.

the upright face configuration governing detection performance. The split-half face transformation (Figure 1C) converts the approximate bilateral symmetry of the face images into centric symmetry without affecting most of the local feature structures and radically perturbs the relative feature configuration. If there is an upright face template, it might be expected to use the split-half configuration at half the efficiency of the full upright face.

Methods

Stimuli

We used four types of images as stimuli (Figure 1). The face images comprised four male and four female faces. These face stimuli were centered at the bridge of the nose, scaled to the separation between the eyes and the mouth and processed by a high-pass filter with a cut-off frequency of four cycles per image. The average eye separation (distance between the centers of the two pupils) was 0.7° . The inverted face images were simply upside-down versions of the face images. The split-face images had their left sides the same as the face images whereas the right sides were inverted version of the same faces. The noise control stimuli were scrambled patterns derived from the same power spectrum as the face images whereas the phase spectrum was generated from a uniformly distributed random number generator with a range from zero to 2π radians. This procedure is equivalent to generating random-dot images and filtering them to match the amplitude spectrum of each face.

All images were scaled to the same eye separation and normalized to equalize the contrast energy across the set. Each image was then presented in a smooth-edged circular aperture whose width formed the independent variable of the experiment. The aperture was centered at the nose. The aperture was defined by a two-dimensional fourth-power Gaussian function $\exp(-(x^4 + y^4) / \sigma^4)$ where σ is the scale parameter that controls the size of the image. We used six σ values ranging from 0.2° to 1.2° in half octave steps (Figure 2). At σ values of 0.8° and beyond, all the facial features were visible.

The application of the fourth-power Gaussian envelope did not produce an imbalance of contrast energy among different stimulus types. We computed the root mean square error (RMSE) contrast for stimuli at all envelope sizes. We found no systematic effect of windowing on RMSE contrast. The difference between the faces and the noises was within 1 dB (0.003–0.847 dB), which was about the precision of our dynamic threshold-seeking method. The measured threshold difference between the faces and the noises, as reported below, were well beyond 5 dB for different conditions and observers.

Apparatus and procedure

The stimuli were presented on a CRT monitor controlled by a Macintosh computer. The monitor input–output intensity function was measured with a LightMouse photometer (Tyler & McBride, 1997). A bit-stealing technique (Tyler, Chan, Liu, McBride, & Kontsevich, 1992) was used to compute the linear look-up table to an accuracy of better than 0.1%. The mean luminance of the monitor was 30 cd/m^2 . At the viewing distance of 114 cm, each pixel on the monitor corresponded to $1'$ in visual angle.

We used a temporal two-alternative forced-choice paradigm and the Ψ dynamic threshold-seeking method (Kontsevich & Tyler, 1999) to measure the detection thresholds relative to a uniform blank field at 75% proportional correct level. The limits of both test and null fields were delineated by corner markers. The temporal waveform of the stimulus intervals was a square pulse with a duration of 300 ms and an 800-ms blank period between the two temporal intervals. The reported contrast threshold values were averages of four threshold measurements.

Four observers participated in the experiment. Among them, CC was one of the authors and all others were naive to the purpose of the study but experienced in psychophysical experiments. Observers AV, CC, and HB had corrected-to-normal and observer AM had normal visual acuity (20/20).

Results and analysis

The results of the detection task for the noise stimulus and the three face-based stimuli are shown for the four data sets for each of the four observers in Figure 3. The null hypothesis is that detection performance would be the same for all stimulus types, based on the matched contrast-energy content of the stimuli. The main experimental question was, do observers perform any differently on the detection task when the information is structured into a face configuration? The answer seems to be yes, in general. The data for face detection showed rapid improvement from the small sizes in all observers, with approximately linear integration up to some aperture size. Moreover, the sensitivity for the original faces was greater than that for the matching noise at large aperture sizes but showed significant evidence of a face inversion effect (Yin, 1969) in only one of the four observers. The form of the detection functions implies the existence of specialized mechanisms underlying face processing and provides information about their organization. To interpret this information, we need to make some basic assumptions about a plausible model structure of the underlying mechanisms. These are the minimum assumptions that can account for the form of the data within a physiologically plausible mechanism structure.

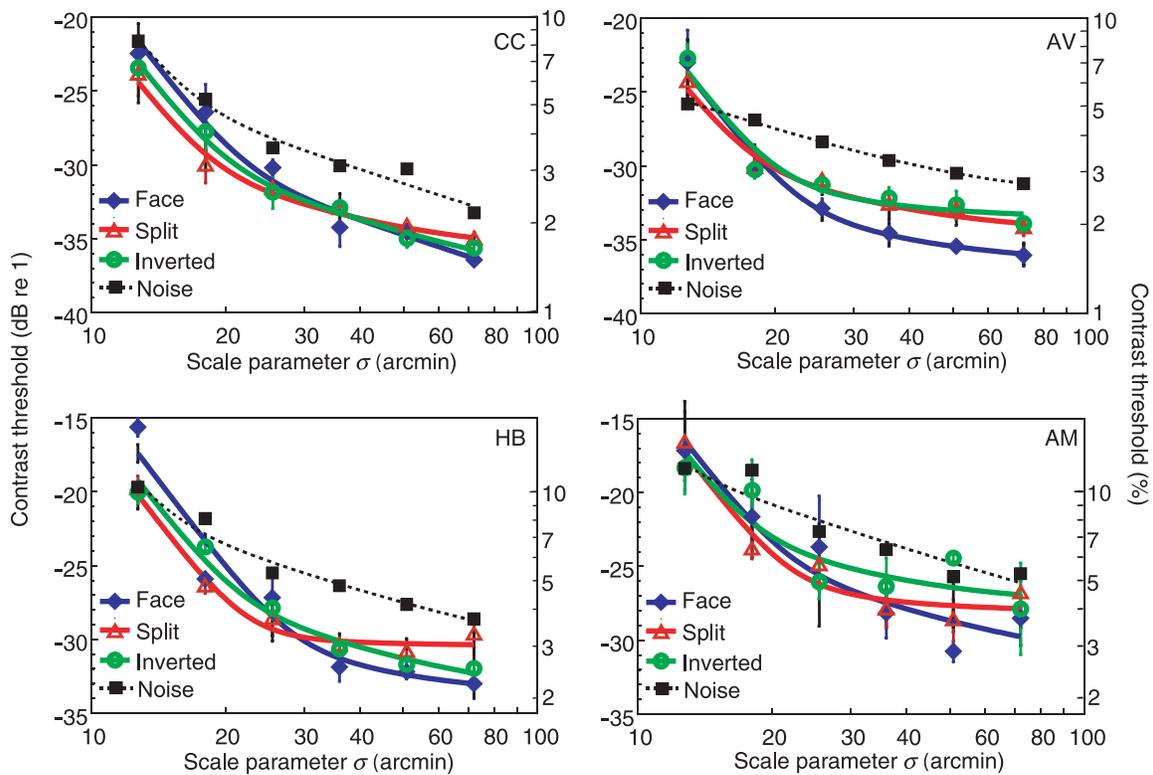


Figure 3. Spatial summation curves for four observers. The abscissa is the length of the stimuli defined by the scale parameter(s) of the fourth-power Gaussian envelope. The ordinate on the left gives the threshold in decibels whereas that on the right is in percentage contrast. The smooth curves are fits of the model defined in the text. The error bars are one standard error of the means.

Model structure

To set the scene, we can summarize the model used to fit the four data sets for each observer as involving a limited set of matching face filters for the three face-based stimulus types, together with a local filter model for the randomized stimuli that, by virtue of its randomization, do not permit any form of matching filter to operate. To convey the local filter outputs to a unitary decision site, they are combined by an attentional pooling mechanism with a limited aperture (Graham, Robson, & Nachmias, 1978; Tyler & Chen, 2000a). This model is schematized in Figure 4 with the following constraints derived from the general form of the data from Figure 3. The summation functions involve three segments implied by matching face filters (centered continuous circles in Figure 4). First, if a stimulus is small enough, it lies within the spatial extent of the smallest relevant detector (center purple circle in Figure 4). Increasing stimulus size up to this extent would simply increase the areal overlap between the detector and the stimulus and generate a proportional increase in the response of the detector. If we assume such detector collection information linearly within its receptive field, the threshold determined by such detector should decrease according to Ricco’s law (e.g., Barlow, 1958; Baumgardt, 1959), as a direct reciprocal function of stimulus area (i.e., a slope of -1 in double-log coordinates). When the

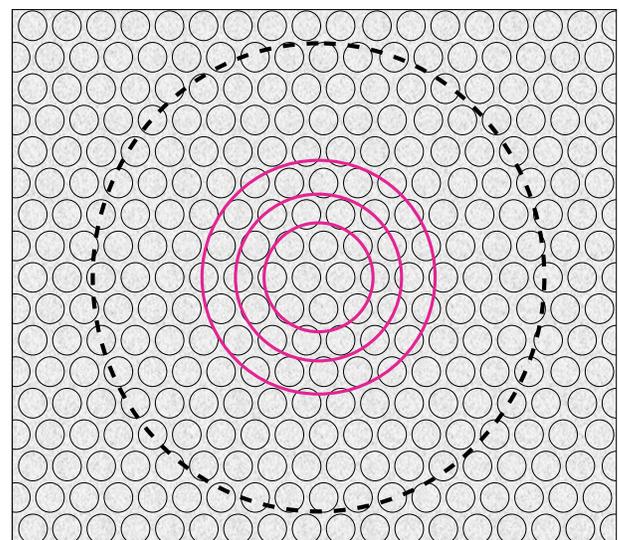


Figure 4. Schematic model of mechanisms involved in face and noise detection. Small circles: local RF filters. Purple circles: a limited range of sizes of matched filters for face detection. Dashed circle: aperture for attentional pooling. Background stippling depicts local noise limiting performance.

stimulus size is large enough, it exceeds the range of one detector and makes no further response increment in that detector. However, there may be larger summing fields beyond the smallest (larger purple circles in Figure 4), providing further summation capability. Under the assumptions that these further mechanisms are limited by local Gaussian noise (texture fill in Figure 4) and are combined by attention to the select one with the strongest response to the current stimulus, the threshold will show a further decrease with a shallower slope of $-1/2$ (Green & Swets, 1966; Tyler & Chen, 2000a). Finally, when the stimulus extends even beyond the largest summation field, the threshold would be constant regardless further size increases. Thus, the parameters in this model of face summation mechanisms are the sensitivity of each mechanism type and the smallest and largest sizes of the matched filter mechanisms. (The latter two parameters are coincident if there is only one matched filter mechanism.)

The second component of the model is the detection of the randomized images, in which the location and form of the features are unpredictable and do not allow the operation of any detection templates. As a result, the only type of summation available is attentional (or ‘probability’) summation model that assumes that detectability is mediated by the maximum response of the array of local filters (Tyler & Chen, 2000a). Linear summation up to the size of these filters (small circles in Figure 4) generates a summation slope of -1 in the range of small stimulus sizes, progressing to a slope of $-1/4$ over the range of operation of the attentional summation, limited by the size of the attentional window (dashed circle in Figure 4). Thus, the model of noise summation has three parameters, the size and sensitivity of the local filters and the size of the attentional window.

Consider first the data for summation of randomized information (dashed curves, Figure 3). Sensitivity continuously improves with aperture size with a slope that approximates $-1/4$ on double-log coordinates. However, across the four observers, the noise summation slopes were not uniform but tended to be steeper for smaller sizes and shallower at larger sizes. This behavior implies that the size of the local filters (Figure 4) had a measurable impact on the lower end of the curve, heading toward the predicted asymptotic slope of -1 for small values. The data thus conform well with the form predicted for randomized patterns.

The local filter model, however, has only one size of local filter operative for noise detection (smallest circles in Figure 4). Moreover, given an extended array of local filters, the shallower slope for some observers at large aperture sizes implies that the information is being truncated by the limited attention window (dashed circle in Figure 4). Hence, the noise summation curve can be fitted with a combination of three line segments with slopes -1 , $-1/4$, and 0 , respectively, combined through a fourth-power attentional summation operator (Tyler & Chen, 2000a) that provides for smooth transitions between the segments of the curve. That is, the threshold θ_n in

linear contrast as a function of stimulus area A was given by

$$\theta_n = ((s_1 \cdot A^{-1})^4 + (s_2 \cdot A^{-1/4})^4 + s_3^4)^{1/4}, \quad (1)$$

where s_1 , s_2 , and s_3 were the relative weights of the three summation regimes. The area A was computed as σ^2 , where σ is the scale constant of the stimulus window.

For the face stimuli, the improved sensitivity over the noise stimuli requires a model structure with the additional component of a set of matched filters for the face-specific information. Because the three face-based stimuli are different, the filter set for each stimulus is allowed to vary for best fit to each data set for each observer. Thus, the thresholds θ_{fi} for the face-based stimuli are fit with an equation of the form

$$\theta_{fi} = ((s_4 \cdot A^{-1})^4 + (s_5 \cdot A^{-1/2})^4 + s_6^4)^{1/4}. \quad (2)$$

(Note that because the face information has a characteristic spatial arrangement, it is not necessarily compatible with the noise distribution, so the local mechanism was omitted from the model in the fitting of the face-based stimuli.)

The model fits are shown as the continuous curves in Figure 3. The parameters for the s_{1-3} are all determined by the fit of Equation 1 to the data sets for the noise alone. They are translated into the model properties of the size σ_L of the local filters and the size of the attentional window σ_W the following relations:

$$\sigma_L = [(s_1/s_2)^{4/3}]^{1/2}, \quad (3)$$

$$\theta_W = [(s_3/s_2)^4]^{1/2}. \quad (4)$$

In the case of the fit of Equation 2, the logic is slightly different. The first intersection, between the slopes of -1 and $-1/2$ (or slopes of -1 and 0 if this pairing provided a better fit to the data), represents the size of the *smallest* matched filter for the particular type of face stimulus in each data set, whereas the second intersection, between the slopes of $-1/2$ and 0 (or -1 and 0 if this pairing provided a better fit), represents the size of the *largest* matched filter for the data set, beyond which there is no improvement in detection sensitivity. They are translated from Equation 2 into the properties of the size σ_{F1} of the smallest filter and the size σ_{F2} of the largest filter by following relations:

$$\sigma_{F1} = [\min((s_4/s_5)^2, s_4/s_6)]^{1/2}, \quad (5)$$

$$\sigma_{F2} = [\max((s_5/s_6)^2, s_4/s_6)]^{1/2}. \quad (6)$$

	σ_{F1}	σ_{F2}	RMSE
Face	0.37 ± 0.09	0.62 ± 0.16	0.98
Invert	0.28 ± 0.05	0.56 ± 0.14	0.72
Split	0.31 ± 0.05	0.43 ± 0.04	0.88
	σ_L	σ_W	RMSE
Noise	0.20 ± 0.10	2.84 ± 1.96	0.62

Table 1. Characteristic sizes for the model fits (in degrees).

Model fit

The RMSE for the joint fit of the respective model components was remarkably low, at 0.81 dB, or between 0.53 and 1.21 dB for the four observers, and between 0.62 and 0.98 dB across conditions. These RMSEs are close to the mean standard error of measurement, which ranged from 0.61 to 1.40 dB across observers, implying that the fitted functions account for essentially all the systematic variation in the data. The parameter values of the fits, in terms of the parameters of Equations 3–6, are specified in Table 1 in terms of visual angle. (For reference, note that the all facial features were available only when the envelope scale parameter was greater than 0.8°.) The error terms represent the standard errors for the model parameters across the group of observers. These show that neither the σ_{F1} nor the σ_{F2} parameters differ significantly for any of the face types. However, both parameters for the noise stimuli differ significantly from all the face stimuli. The data therefore support a dual-process model of one summation structure for the noise stimuli (Equation 1) and a second set of summation sizes for all the face stimuli (Equation 2 with one set of parameters), although it does not tell us that the specific filters are the same for the three face types.

The fits of this dual-process model structure may be compared with (a) that of a unitary multiple-mechanism summation model fitted to all four stimulus types, and (b) that of separate single-mechanism summation models fitted for each stimulus type. These models are representative of the two closest mechanism structures competing with the full model. The unitary multiple-mechanism summation

model suggests that detectability is controlled by a uniform model regardless of stimulus type; that is, the parameters σ_{F1} and σ_{F2} are the same for all stimulus types. The full model (schematized in Figure 4) gave a significantly better fit to the data sets for the four observers than the unitary multiple-mechanism model, $F(9,88) = 6.56, p < .0001$.

In the single-mechanism summation models, there are specialized “template” mechanisms for each stimulus type consisting of only one summation field. That is, $\sigma_{F1} = \sigma_{F2}$, or s_5 in Equation 2, is zero. Again, the full model gave a significantly better fit to the data sets for the four observers than the model of separate single mechanisms for each stimulus type, $F(4,83) = 7.33, p < .0001$. Thus, there is good statistical support for the dual-process model structure of Figure 4 over its two closest competitors.

As shown in Table 1, the difference in averaged fit parameters σ_{F1} and σ_{F2} , between faces and inverted faces was small compared with standard deviation across observers. However, it does not necessary mean that summation curves for faces and inverted faces are the same. Instead, the above analysis may be insensitive to small differences in the fits for individual observers because they are masked by larger differences among the observers. We therefore moved to a within-subject chi-square analysis of the goodness of fits for each function to the condition for which it was optimized in comparison to the fits for the other two face types. This analysis may be applied to data set for each observer and cumulated across observers. The results are provided in Table 2. They show that the fit increases significantly for the inverted-face condition individually for three of the four observers and for the split-face condition for all four individual observers. The overall chi-square is significantly different for both conditions. These results for the more sensitive chi-square analysis imply that the filter sizes defined by parameters σ_{F1} and σ_{F2} are, in fact, significantly larger for upright face processing than for processing of the inverted and split-face stimuli.

Discussion

There have been numerous studies with reverse correlation and “bubbles” techniques that identify particular features as

Observer	Upright	Inverted	Split	Difference: inverted and upright	Difference: split and upright
AM	7.93	18.07	16.57	10.14*	8.64*
AV	5.69	35.68	27.11	29.99*	21.42*
CC	0.92	5.32	26.78	4.40	25.86*
HB	4.11	17.61	41.43	13.50*	37.32*
Overall chi-square difference				58.03*	93.24*
p value				7.32×10^{-05}	2.018×10^{-10}

Table 2. Chi-square values for the fit of the upright model function to the nonfitted data sets. *Significant difference at $\alpha = .01$.

relevant to face identification and discrimination (Gold, Murray, Bennett, & Sekuler, 2000; Gosselin & Schyns, 2001; Kontsevich & Tyler, 2004; Mangini & Biederman, 2004; Schyns, Bonnar, & Gosselin, 2002; Sekuler, Gaspar, Gold, & Bennett, 2004; Smith, Cottrell, Gosselin, & Schyns, 2005). Although these methods have identified the key features involved in face processing, they have not revealed the structure of the long-range summation fields underlying such feature processing. By utilizing the theory of threshold summation behavior (Tyler & Chen, 2000a), this study evaluates the long-range of information processing involved in face perception near detection threshold. The data indicate that there is specialized processing of human face information relative to that for Fourier-equated noise at detection threshold. All observers show a strong summation of upright face information up to an intermediate size, continuing with weaker summation up to a larger size and asymptoting to negligible summation beyond that size. The limiting size for full summation was 0.37° , compatible with the existence of a single configuration template for the central portion of the face (essentially the eyes and nose in our configuration). The mean size of the second corner in the summation function was 0.62° (about the range of the internal features of the face), implying the existence of additional face configuration mechanisms out to this extent, beyond which there was no further improvement in sensitivity. Thus, all observers showed no further advantage for the information around the edge of the face—the hair, ears, and jaw line. Although these facial characteristics are well known to be significant for facial discrimination (Ullmann, 1996), they seem to play no incremental role in threshold detectability, at least under our test conditions. [It should be stated that the specialization for face stimuli does not imply the presence or absence of such processing for other classes of object stimuli such as shapes (Weisstein & Harris, 1974), words (Reicher, 1969), and greebles (Gauthier & Tarr, 1997).]

In contrast to the model structure for upright face processing, the detection function for noise alone had markedly different properties, showing full summation up to only 0.2° and fourth-root summation for an extended region beyond that point, terminating with a variable asymptote averaging nearly 3° in diameter (see Figure 4). The interpretation of these fits is that the smallest mechanism available for the unpredictable noise (matching the spectral characteristics of the face stimuli) was about 0.2° in extent (17% of the full face width), and that attentional pooling was available to integrate information from such mechanisms up to the limit of the attentional window, whose extent varied from 0.94° to 5.47° across observers. The noise control is important because it sets the limiting sensitivity for local filter properties, which must be exceeded in the face-based conditions to constitute evidence for specialized face-processing mechanisms.

The other two face-based stimulus types cast further light on the mechanisms available for the human detection of face stimuli. These two manipulations of the face images resulted in summation functions that were rather

similar to those for the upright face, on average. However, a chi-square analysis on the individual data show the both the small and large size of the summation range for the filters modeling face detection are larger than those for the two control face types. This result appears to be an example of the face superiority effect measured by various psychophysical techniques (Farah, Tanaka & Drain, 1995; Tanaka & Farah, 1993; Valentine, 1988; Yin, 1969). The ratio of thresholds obtained between the upright faces and the other face types is between 3 and 6 dB, comparable with the size of the superiority effect obtained with other contrast tasks (e.g., Tanaka & Sengco, 1997).

The curtailment of the summation for the split-face condition may have been the result of splitting the mouth feature into two unfamiliar halves, one half above and the other below the eye features. Thus, although the summation was similar regardless of the location (or orientation) of the features within the summation zone, small differences in the estimates of underlying mechanism sizes were evident.

It is also of interest that the detection thresholds for faces deviate markedly from the prediction of a contrast energy model (Watson & Ahumada, 2005), which assumes that the observers are able to sum the local contrast information over the entire aperture of the fovea (which is much larger than a typical receptive field). In its generalized form (general contrast energy, GCE), this model assumes that there is an accelerating transducer for contrast preceding the summation, with a power exponent of 2.6. Even if we allow the power component to be a free parameter, application of this prediction substantially degrades the fit to the data, increasing the RMSE by more than twofold (0.81–1.68 dB). Even accounting for the one extra parameter per summation curve of our summation model, this degradation is significant, $F(16,63) = 12.79$, $p < .0001$. This result may not be totally surprising. The GCE model was developed to account for detection thresholds for a large bank of visual stimuli provided by the Modelfest group (Carney, Tyler, Watson, Makous, Beutter, Chen, et al., 2000). That stimulus set was dominated by Gabor patches. Actually, even in that data set, the GCE model had more difficulty with complex stimuli such as random noise and a natural scene. The poor fit of the GCE model for both the complex stimuli in the Modelfest stimulus set and the complex face stimuli of this study suggests that it may not be suitable to account for the detectability of complex stimuli in general.

In general, the comparisons both with the reduced model structures in the analysis and with the GCE model in the previous paragraph illustrate that many plausible models of the spatial summation of face information are excluded by the present data. The excellent quality of the fits, with RMSEs close to the standard error of measurement, suggests that the full model fit of Equation 1 to the noise data and Equation 2 to the three face-stimulus data has captured almost all of the systematic variation in the data (despite some individual variation in the relative sensitivities to the four stimulus types). The threshold summation approach thus provides a valuable adjunct to other suprathreshold

analyses of the mechanisms underlying the processing of facial information.

Conclusions

We performed a summation experiment with face, inverted face, split-face, and phase-scrambled versions of the faces with equal Fourier energy using two-alternative forced-choice psychophysics. Face detectability improved rapidly at first, then at a progressively shallower rate for larger window sizes, in a similar fashion for the three face-based stimulus types. The noise stimuli were less detectable than the face stimuli for all except the smallest apertures. The results were fit with a model incorporating global face-specific and local nonspecific spatial integration mechanisms. Detection of the noise images was consistent with local detection mechanisms accessed through a wide-field attention mechanism. The data for face detection implied detection mechanisms that integrated linearly up to some small size, integrated more slowly up to an intermediate size, and failed to gain any improvement for information beyond some larger size. This performance supports the concept of a specialized face configuration mechanism at detection threshold, similar in extent among the observers. The functions for the Fourier-equated noise stimuli not only indicate the presence of a separate spatial summation process dominated by an attentional combination rule, but one that has lower sensitivity than the face detection mechanisms and has a limited aperture for the attentional combination operation. This noise-specific summation process (Equation 1) corresponds to the ‘Standard Model’ of Watson & Ahumada (2005) and its reduced sensitivity indicates the limitations of that model in relation to the detection of more real-world stimuli such as faces.

Acknowledgments

Supported by NIH/NEI EY13025 to CWT and NSC 94-2752-H-002-007-PAE to CCC. A preliminary version of this study was published as Tyler and Chen (2000b).

Commercial relationships: none.

Corresponding author: C.W. Tyler.

Email: cwt@ski.org.

Address: The Smith-Kettlewell Eye Research Institute, 2318 Fillmore St., San Francisco, CA, USA.

References

- Barlow, H. B. (1958). Temporal and spatial summation in human vision at different background intensities. *The Journal of Physiology*, *141*, 337–350. [PubMed] [Article]
- Baumgardt, E. (1959). Visual spatial and temporal summation. *Nature*, *184*, 1951–1952. [PubMed]
- Brown, V., Huey, D., & Findlay, J. M. (1997). Face detection in peripheral vision: Do faces pop out? *Perception*, *26*, 1555–1570. [PubMed]
- Carey, S., & Diamond, R. (1977). From piecemeal to configurational representation of faces. *Science*, *195*, 312–314. [PubMed]
- Carney, T., Tyler, C. W., Watson, A. B., Makous, W., Beutter, B., Chen, C. C., et al. (2000). Modelfest: First year results and plans for future years. *SPIE Proceedings*, *3959*, 140–151. [Link]
- Farah, M. J., Tanaka, J. W., & Drain, H. M. (1995). What causes the face inversion effect? *Journal of Experimental Psychology: Human Perception Performance*, *21*, 628–634. [PubMed]
- Gauthier, I., & Tarr, M. J. (1997). Becoming a “Greeble” expert: Exploring mechanisms for face recognition. *Vision Research*, *37*, 1673–1682. [PubMed]
- Gold, J. M., Murray, R. F., Bennett, P. J., & Sekuler, A. B. (2000). Deriving behavioural receptive fields for visually completed contours. *Current Biology*, *10*, 663–666. [PubMed] [Article]
- Gorea, A., & Julesz, B. (1990). Context superiority in a detection task with line-element stimuli: A low-level effect. *Perception*, *19*, 5–16. [PubMed]
- Gosselin, F., & Schyns, P. G. (2001). Bubbles: A technique to reveal the use of information in recognition tasks. *Vision Research*, *41*, 2261–2271. [PubMed]
- Graham, N., Robson, J. G., & Nachmias, J. (1978). Grating summation in fovea and periphery. *Vision Research*, *18*, 815–825. [PubMed]
- Green, D. M., & Swets, J. A. (1966). *Signal detection theory and psychophysics*. New York: Wiley.
- Kontsevich, L. L., & Tyler, C. W. (1999). Bayesian adaptive estimation of psychometric slope and threshold. *Vision Research*, *39*, 2729–2737. [PubMed]
- Kontsevich, L. L., & Tyler, C. W. (2004). What makes Mona Lisa smile? *Vision Research*, *44*, 1493–1498. [PubMed]
- Lewis, M. B., & Edmonds, A. J. (2003). Face detection: Mapping human performance. *Perception*, *32*, 903–920. [PubMed]
- Mangini, M. C., & Biederman, I. (2004). Making the ineffable explicit: Estimating the information employed for face classification. *Cognitive Science*, *28*, 209–226.
- Nothdurft, H. C. (1993). Faces and facial expressions do not pop out. *Perception*, *22*, 1287–1298. [PubMed]

- Purcell, D. G., & Stewart, A. L. (1986). The face-detection effect. *Bulletin of the Psychonomic Society*, *24*, 118–120.
- Purcell, D. G., & Stewart, A. L. (1988). The face-detection effect: Configuration enhances detection. *Perception & Psychophysics*, *43*, 355–366. [PubMed]
- Reicher, G. M. (1969). Perceptual recognition as a function of meaningfulness of stimulus material. *Journal of experimental psychology*, *81*, 275–280. [PubMed]
- Rhodes, G., Jeffery, L., Watson, T. L., Jaquet, E., Winkler, C., & Clifford, C. W. (2004). Orientation-contingent face aftereffects and implications for face-coding mechanisms. *Current Biology*, *14*, 2119–2123. [PubMed] [Article]
- Rolls, E. T., Tovee, M. J., Purcell, D. G., Stewart, A. L., & Azzopardi, P. (1994). The responses of neurons in the temporal cortex of primates, and face identification and detection. *Experimental Brain Research*, *101*, 473–484. [PubMed]
- Schyns, P. G., Bonnar, L., & Gosselin, F. (2002). Show me the features! Understanding recognition from the use of visual information. *Psychological Science*, *13*, 402–409. [PubMed]
- Sekuler, A. B., Gaspar, C. M., Gold, J. M., & Bennett, P. J. (2004). Inversion leads to quantitative, not qualitative, changes in face processing. *Current Biology*, *14*, 391–396. [PubMed] [Article]
- Smith, M. L., Cottrell, G. W., Gosselin, F., & Schyns, P. G. (2005). Transmitting and decoding facial expressions. *Psychological Science*, *16*, 184–189. [PubMed]
- Tanaka, J. W., & Farah, M. (1993). Parts and wholes in face recognition. *Quarterly Journal of Experimental Psychology A, Human Experimental Psychology*, *46*, 225–245. [PubMed]
- Tanaka, J. W., & Sengco, J. A. (1997). Features and their configuration in face recognition. *Memory & Cognition*, *25*, 583–592. [PubMed]
- Thompson, P. (1980). Margaret Thatcher: A new illusion. *Perception*, *9*, 483–484. [PubMed]
- Tyler, C. W., Chan, H., Liu, L., McBride, B., & Kontsevich, L. L. (1992). Bit stealing: How to get 1786 or more gray levels from an 8-bit color monitor. In B. Rogowitz (Ed.), *Human vision, visual processing & digital display III* (pp. 351–364). Bellingham, WA: SPIE.
- Tyler, C. W., & Chen, C. C. (2000a). Signal detection theory in the 2AFC paradigm: Attention, channel uncertainty and probability summation. *Vision Research*, *40*, 3121–3144. [PubMed]
- Tyler, C. W., & Chen, C. C. (2000b). Spatial summation of face information. *SPIE Proceedings*, *3959*, 451–457.
- Tyler, C. W., & McBride, B. (1997). The Morphonome image psychophysics software and a calibrator for Macintosh systems. *Spatial Vision*, *10*, 479–484. [PubMed]
- Ullman, S. (1996). *High-level vision*. Cambridge, MA: MIT Press.
- Valentine, T. (1988). Upside-down faces: A review of the effect of inversion upon face recognition. *British Journal of Psychology*, *79*, 471–491. [PubMed]
- Vanrullen, R. (2006). On second glance: Still no high-level pop-out effect for faces. *Vision Research*, *46*, 3017–3027. [PubMed]
- van Santen, J. P. H., & Jonides, J. (1978). A replication of the face-superiority effect. *Bulletin of the Psychonomic Society*, *12*, 378–380.
- Watson, A. B. (2000). Visual detection of spatial contrast patterns: Evaluation of five simple models. *Optic Express*, *6*, 12–33. [PubMed] [Article]
- Watson, A. B., & Ahumada, A. J. Jr. (2005). A standard model for foveal detection of spatial contrast. *Journal of Vision*, *5*(9), 717–740, <http://journalofvision.org/5/9/6/>, doi:10.1167/5.9.6. [PubMed] [Article]
- Webster, M. A., & MacLin, O. H. (1999). Figural aftereffects in the perception of faces. *Psychonomic Bulletin & Review*, *6*, 647–653. [PubMed]
- Weisstein, N., & Harris, C. S. (1974). Visual detection of line segments: An object-superiority effect. *Science*, *186*, 752–755. [PubMed]
- Wolfe, J. M. (1998). Visual search. In H. Pashler (Ed.), *Attention* (pp. 13–73). London, UK: University College London Press.
- Yin, R. K. (1969). Looking at upside-down faces. *Journal of Experimental Psychology*, *81*, 141–145.
- Zhao, L., & Chubb, C. (2001). The size-tuning of the face-distortion after-effect. *Vision Research*, *41*, 2979–2994. [PubMed]