Published in final edited form as:
Neuron. 2012 May 10; 74(3): 567–581. doi:10.1016/j.neuron.2012.03.024.

What makes a cell face-selective: the importance of contrast

Shay Ohayon,
Division of Biology, Computation and Neural Systems, California Institute of Technology, 1200 East California Boulevard, Pasadena, CA 91125, USA

Winrich A Freiwald, and
Laboratory of Neural Systems, The Rockefeller University, 1230 York Avenue, New York, NY 10065, USA

Doris Y Tsao
Division of Biology, Computation and Neural Systems, California Institute of Technology, 1200 East California Boulevard, Pasadena, CA 91125, USA

Summary

Faces are robustly detected by computer vision algorithms that search for characteristic coarse contrast features. Here, we investigated whether face-selective cells in the primate brain exploit contrast features as well. We recorded from face-selective neurons in macaque inferotemporal cortex, while presenting a face-like collage of regions whose luminances were changed randomly. Modulating contrast combinations between regions induced activity changes ranging from no response to a response greater than that to a real face in 50% of cells. The critical stimulus factor determining response magnitude was contrast polarity, e.g., nose region brighter than left eye. Contrast polarity preferences were consistent across cells, suggesting a common computational strategy across the population, and matched features used by computer vision algorithms for face detection. Furthermore, most cells were tuned both for contrast polarity and for the geometry of facial features, suggesting cells encode information useful both for detection and recognition.

Keywords

face selectivity; macaque middle face patches; face detection; contrast features; fMRI

Introduction

Neurons in inferior temporal (IT) cortex of the macaque brain respond selectively to complex shapes (Desimone et al., 1984; Fujita et al., 1992; Logothetis and Sheinberg, 1996; Tanaka, 1996, 2003; Tanaka et al., 1991; Tsunoda et al., 2001). Previous studies have proposed that the key element in shape representation is contours and that this representation may be encoded by local curvature and orientation across the population of V4 cells which project to IT (Brincat and Connor, 2004; Pasupathy and Connor, 2002). However, contours are only one source of information available in the retinal image.

Another rich source of information about object shape is contrast. Humans can detect and recognize objects in extremely degraded images consisting of only a few pixels (Harmon

and Julesz, 1973; Heinrich and Bach, 2010; Sinha et al., 2006). Thus high frequency information and fine feature details may not be necessary for object detection. What types of features are available in the low frequency range? One possibility is features based on coarse-level contrast cues. Contrast features have been proposed as an intermediate feature representation in computer vision systems (Papageorgiou et al., 1998) and are ubiquitous in state-of-the art object recognition systems, in particular, for face detection (Lienhart and Jochen, 2002; Viola and Jones, 2001).

If contrast is an important component of object representation in IT cortex, one would expect cells to be strongly modulated by contrast manipulations such as global contrast reversal. Indeed, when Tanaka et al. (Ito et al., 1994; Tanaka et al., 1991) presented simple geometrical shapes such as stars or ellipses with different protrusions to IT cells and manipulated the contrast by global contrast reversal or outlining (removing contrast from filled regions and retaining only edges), many cells (> 95%) showed dramatic reductions in firing rate, suggesting that cells in IT carry information about contrast polarity (Fujita et al., 1992; Ito et al., 1994; Tanaka, 1996, 2003). While characterizing cell responses to contrast reversal reveals whether contrast is important, this approach does not address the more fundamental question of how contrast sensitivity might contribute to the form selectivity of a given neuron. Moreover, other studies report that IT cells do not change their firing rates with contrast reversal (Baylis and Driver, 2001; Rolls and Baylis, 1986), leading to the conclusion that a hallmark of object representation in IT cortex lies in its ability to generalize over global contrast reversal. Thus, the importance of contrast in shape encoding in IT has remained elusive.

Here, we ask whether contrast features serve as a fundamental building block for object selectivity in macaque IT cortex. This question has been difficult to answer in previous studies since cells were picked at random from IT cortex. The variance of cells' shape preferences in such random sampling was large and prohibited a systematic study involving local manipulations of parts and their contrasts. Here, we take advantage of the known shape selectivity in macaque face-selective regions. These regions have a high concentration of cells firing stronger to faces compared to other objects (Tsao et al., 2006). The known shape selectivity enabled us to focus on the individual parts constituting the face and to investigate the role of contrast by systematically manipulating contrast across parts while preserving effective contours.

If contrast plays a role in shape coding we would expect it to have an effect at early stages of face processing. By using functional magnetic resonance imaging (fMRI) with a face localizer stimulus, we targeted our recordings to the middle face patches. There are several indications that the middle face patches likely represent an early stage of face processing. First, cells in the middle face patches are still view-specific, unlike those in more anterior face-selective regions (Freiwald and Tsao, 2010). Second, some cells in the middle face patches still fire to object stimuli sharing rudimentary features with faces, such as apples and clocks (Tsao et al., 2006).

While face-selective cells have been shown to be tuned for fine structural details (Freiwald et al., 2009), their selectivity for coarse-level features has not been investigated. Many coarse-level contrast feature combinations are possible. However, only a few can be considered predictive of the presence of a specific object in an image. The predictive features can be found by an exhaustive search (Lienhart and Jochen, 2002; Viola and Jones, 2001), or by other considerations, such as consistency across presentations with different lighting conditions (i.e., invariance to illumination changes). Indeed, a simple computational model for face detection based on illumination-invariant contrast features was proposed by Sinha (Sinha, 2002). In Sinha's model, a face is detected in a given image if 11 conditions

are met. Each condition evaluates a local contrast feature (luminance difference across two regions of the face, e.g., nose and left eye) and tests whether contrast polarity is along the direction predicted from illumination invariance considerations. Here, we tested whether face-selective cells are tuned for contrast features useful for face detection. We measured responses to an artificial parameterized stimulus set, as well as to large sets of real face and non-face images with varying contrast characteristics to elucidate the role of contrast in object representation.

Results

Face-selective cells respond differently to different contrast combinations

We identified the locations of six face patches in the temporal lobes of three macaque monkeys with fMRI by presenting an independent face localizer stimulus set and contrasting responses to real faces with those to non-face objects (Moeller et al., 2008; Tsao et al., 2003; Tsao et al., 2008). We then targeted the middle face patches for electrophysiological recordings (Ohayon and Tsao, 2011), (see Experimental Procedures, Figure S1); We recorded 342 well-isolated single units (171 in Monkey H, 129 in Monkey R, and 42 in Monkey J) while presenting images in rapid succession (5 images / s). Images were flashed for 100 ms (ON period), and were followed by a gray screen for another 100 ms (OFF period). Monkeys passively viewed the screen and were rewarded with juice every 2–4 s during fixation.

We presented 16 real face images and 80 non-face object images to assess face selectivity (Tsao et al., 2006). We quantified how well each cell discriminated face images from non-face images using the d’ measure (see Experimental Procedures) and considered cells to be face selective if $d' > 0.5$. Under this criterion, 280/342 cells (137 in monkey H and 108 in monkey R and 35 in monkey J), were found to be face selective across the population (Figure 1, see Experimental Procedures). Similar results were obtained with other face selectivity metrics (Figures S2A-S2B).

Motivated by coarse contrast features that are ubiquitously used in state-of-art face detection systems (Figure 2A, Viola and Jones, 2001), we designed a simple 11 part stimulus (Figure 2B) to assess selectivity for luminance contrasts in the face. In brief, we decomposed the picture of an average face to 11 parts (Figure 2B), and assigned each part a unique intensity value ranging between dark and bright. By selecting different permutations of intensities we could generate different stimuli. We randomly selected 432 permutations to cover all possible pair-wise combinations of parts and intensities (see Experimental Procedures).

We first tested whether cells selective for real face images would respond to our artificial parameterized stimulus. Cells typically showed large variance of response magnitudes to the different parameterized stimuli. The example cell in Figure 2C fired vigorously for only a subset of the parameterized faces. The subset that was effective drove the cell to levels that were comparable to those to real faces, while other parameterized stimuli were less effective in driving the cell, leading to firing rates that were comparable to those to objects. A similar trend was observed across the population (Figure 2D). Parameterized face stimuli elicited responses ranging between nothing to strong firing (Figure 2D, right column). Thus different luminance combinations can either be effective or ineffective drivers for cells.

To test the extent to which a parameterized face could drive cells we computed the maximal response across all 432 parameterized face stimuli and compared it to the maximal response evoked by a real face (Figure S2C). In about half of the cells (145/280) the maximal evoked response by a parameterized face was stronger than the maximal evoked response by a real face. Furthermore, the minimal evoked response across the 432 parameterized face stimuli

was smaller than the maximal evoked response by objects. Thus, middle face patch neurons can be driven by highly simplified stimuli lacking many of the fine structural features of a real face such as texture and fine contours. On average, we found 60±76 parameterized stimuli per cell that elicited firing rates greater than the mean firing rate to real faces, indicating that the observed ratio of maximal responses was not due to a single stimulus. Thus some of the artificial stimuli seem to be good proxies for real faces.

Cells are tuned for contrast polarity features

Cells responded to the parameterized stimulus set with large variability. But are there any rules governing whether a given stimulus elicits a strong response or not? If so, what are these rules, and do they apply to all cells? We hypothesized that relative intensity, i.e., contrast, across parts and its polarity are the governing principles underlying the observed responses. To test this hypothesis, we analyzed firing rate as a function of the pair-wise contrast polarity among the 11 face parts. For each cell, we considered all 55 possible part pairs (pair table, Supplementary Table 1). For a given part pair (A–B), we compared the responses to stimuli with intensity of part A greater than part B with responses to the reversed contrast polarity, irrespective of the luminance values assumed by the remaining nine face parts. If contrast polarity plays a role in determining the observed variability, cells should show significant differences in firing rates for the condition A>B vs. the condition A<B.

We found that middle face patch neurons are indeed sensitive to the contrast between face parts and its polarity. This is illustrated by an example cell in Figure 3A (same cell that is shown in Figure 2C), whose firing rate was significantly modulated by 29 of 55 contrast pairs ($P < 10^{-5}$, Mann-Whitney U-test). Not only were these firing rate differences significant, they were also sizeable. For example, the example cell fired about twice as strongly when the intensity in the left eye region was lower than that of the nose region (30 Hz, vs. 15 Hz, Figure 3A), irrespective of all other 9 face parts.

The same pattern of results was observed across the population. Out of the 280 face-selective cells, 138 (62/135 in Monkey H, 57/108 in Monkey R and 19/35 in Monkey J) were significantly tuned for at least one contrast polarity pair ($P < 10^{-5}$, Mann Whitney U-test). Those cells sensitive to contrast polarity features were influenced by 8.13 ± 7.17 features (Figure S3). Different cells were tuned for different contrast polarity features. The tuning for contrast polarity features can be summarized in a tuning matrix indicating for each part-pair whether it was significant, and if so, which polarity evoked the stronger response. The tuning matrix of monkey R (Figure 3B) illustrates the diversity, but also consistency, of significant tuning in the population. Similar tuning matrices were observed for monkey H (Figure 3C) and monkey J (Figure 3D). Thus, about 50% of face-selective cells encode some aspect of contrast polarity across face parts.

Is there a common principle behind the observed tuning to contrast polarity? Computational models, as well as psychophysics observations (Sinha, 2002; Sinha et al., 2006; Viola and Jones, 2001) have suggested that if a certain feature is useful in predicting the presence of an object in an image, its contrast polarity should be consistent across different image presentations, and should generalize over different illumination conditions, and small changes in viewpoint. To test this, we conducted illumination invariance measurements for human and macaque faces (Figures S4A–S4D) and confirmed that a subset of contrast polarity features such as eye-forehead can predict the presence of a face in an image since polarity is consistent and eyes tend to be darker than the forehead in the majority of images tested. Thus, some contrast polarity features can serve as good indicators to the presence of a face under various light configurations.

To test whether middle face patch neurons coded contrast polarity consistently, we plotted the number of cells that significantly preferred A>B along the positive axis and the number of cells that significantly preferred A<B along the negative axis (Figure 4A). Notice that for a proposed part pair, each cell can either vote along the positive direction or along the negative direction (but not both), depending on which direction elicited the higher significant firing rate. The histogram of cells tuned for specific contrast pairs in Figure 4A demonstrates very strong consistency across the population for preferred polarity direction. For example, while 95 (42 in monkey H, 41 in monkey R, 11 in monkey J) cells preferred the left eye to be darker than the nose (pair index 11), just a single cell was found that preferred the opposite polarity. The same result was found across other pairs: if a contrast polarity direction was preferred by one cell, it was also preferred by almost all other cells that were selective for the contrast of the part combination. We quantified this by measuring the polarity consistency index (see Experimental Procedures). A consistency index of value one indicates that all cells agree on their contrast polarity preference while a consistency index of zero indicates that half of the cells preferred one polarity direction and the other half, the opposite polarity direction. Pooling data from all three monkeys, we found the consistency index to be 0.93±0.15 (discarding features for which less than two tuned cells were found). Furthermore, polarity histograms from each individual monkey show that preferred polarities were highly consistent across the three animals (Figure 4B). Thus face-selective cells are not encoding a random set of contrast polarities across face parts, but instead have a highly consistent preference for polarity depending on the part pair.

Do the preferred contrast polarities agree with predicted features that are useful for face detection? To test this we plotted the polarities proposed by the Sinha model (Sinha, 2002) as well as two other predictions from our illumination invariance measurements (Figure 4A). Overall, we found that many of the predicted contrast polarity features were represented across the population. Importantly, almost no cells were found to be tuned to a polarity opposite to the prediction (Figure S4E).

While cells were highly consistent in their contrast polarity preference for any given part pair, they varied widely as to which pairs they were selective for. Some contrast pairs were more prominently represented than others. The most common contrast pair was Nose > Left Eye, for which almost 70% of the cells were tuned, followed closely by Nose > Right Eye (Figure 4C). Although the most common features involved the eye region, many other regions were represented as well. A graphical representation of the tuning for several random cells is shown in Figure 4D. Green lines represent a significant part pair which does not include the eye region while yellow lines denote pairs including the eye region. Notice that for some of these cells, the significant feature included non-neighboring parts as well (e.g., top right corner, forehead–chin). Cells encoded on average 4.6 features involving eyes (out of a possible 19) and 3.3 features that did do not include the eye region (out of a possible 36). This suggests that cells are encoding a holistic representation which includes multiple face parts, but not necessarily the entire face.

Contrast polarity information arises from low spatial frequencies

The parts constituting the parameterized face stimulus consisted of large regions (Figure 2B), suggesting that selectivity for contrast polarity between these parts is based on low spatial frequency information. However, it is also possible that contrast information was extracted just from the borders between face parts, and could thus be based on high frequency information.

To test to what extent low and high frequency information contribute to the contrast selectivity we conducted two further experiments in which we presented two variants of the parameterized stimulus (Figures S5C and S5E). The first variant retained the contrast

relationships from the original experiment, but only along the contours (Figure S5C). The second variant was a heavily smoothed version of the original parameterized face. If high frequency information is critical, we would expect to see the same modulation for the first, but not the second variant.

We recorded from 18 additional face-selective units in monkey R and presented both the original parameterized face and the first variant. The cells showed similar patterns of tuning for the original parameterized face (Figure S5B), but almost no significant tuning was found for the first variant (Figure S5D). To further validate that high frequency information is not the critical factor, we recorded 34 additional face-selective units in monkey R while presenting the second, heavily smoothed variant of the parameterized face (Figure S5E). In this case, we found similar tuning for contrast polarity as for the original parameterized face stimulus (Figure S5F) .

To further evaluate the contribution of contours compared to contrast we generated a third parameterized face stimulus variant in which we varied the luminance level of all parts simultaneously, resulting in 11 different stimuli (Figure 5A). These stimuli lacked the contrast differences across parts, but maintained the same contours that were present in the normal parameterized face stimuli. The third variant stimuli were presented along with the main experiment stimuli (real faces, normal parameterized stimulus and non-face objects) to further characterize the same 280 face-selective units (from the analysis in Figures 2–4). To assess the contribution of contrast we considered a contrast relationship to be “correct” if its polarity agreed with the Sinha model (Sinha, 2002) (see Figure S4E for list of correct part pairs). We found that the stimuli that contained contours but no contrast relationships elicited a response that was comparable to stimuli with only a few correct features, but still significantly higher (almost three fold in magnitude) than the response to non-face objects (Figure 5B).

Thus both contours and correct contrast contribute to the overall firing rate of cells and sensitivity to contrast polarity features arises from low spatial frequency information across large regions of the face.

Contrast is necessary but not sufficient to elicit strong responses

Our results obtained from simplified face stimuli suggest that correct contrast is necessary to yield strong responses from face-selective cells. Do these results extend to real faces? And is correct contrast even sufficient, i.e., does correct contrast, when it occurs outside a face, trigger large responses, too? To investigate these issues we generated an image set using the CBCL library (Heisele et al., 2000) containing 207 real faces (registered and normalized in size) and 204 non-face images randomly sampled from natural images lacking faces. Face images lacked external contours such as the hair (Figure 6A).

To determine the number of correct contrast features in each of these images we manually outlined the parts on the average face (Figure 6A). Since all images were registered the template matched all faces. The same template was then overlaid on each of the non-face images and the number of correct contrast features was computed in a similar way (i.e., by averaging the intensity level in each region, see Figure 6A). In this way we could build an image set of faces and non-faces with varying numbers of correct contrast polarity features (Figure 6B). Although individual samples of 12 correct features in non-face images did not resemble a face, their average did (Figure 6B, last column).

We reasoned that if face-selective cells use a simple averaging scheme over fixed regions similar to proposed computational models (Lienhart and Jochen, 2002; Sinha, 2002; Viola

and Jones, 2001), they would respond strongly to non-face stimuli with correct contrast relationships.

We recorded the responses of 25 face-selective units in monkey H and 41 in Monkey R. The response of one cell as a function of the number of correct polarity features is presented in Figure 6C. When presented with pictures of faces, the cell increased its firing rate as the number of correct features increased. However, no significant change in firing rate was observed to non-face images, regardless of the number of correct polarity features (Figure 6D). We found similar behavior across the population (Figure 6E). The number of correct contrast polarity features was found to be a significant factor modulating firing rate for face images (one way ANOVA, $P < 0.0001$), but not for non-face images (one way ANOVA, $P > 0.8$). Thus contrast features, though necessary, are not sufficient to drive face-selective cells. The presence of higher spatial frequency structures can additionally modulate the responses of the cells, and interfere with the effects of coarse contrast structure.

Global contrast inversion

Our results so far demonstrate that contrast can serve as a critical factor in driving face-selective cells. From this finding one would predict that global contrast inversion of the entire image should elicit low firing rates. To test this prediction and directly relate our results to previous studies on effects of global contrast inversion in IT cortex (Baylis and Driver, 2001; Ito et al., 1994; Rolls and Baylis, 1986), we presented global contrast-inverted images of faces and their normal contrast counterparts and recorded from 20 additional face-selective cells from monkey H and monkey R (Figure 7A, black traces). The response to faces was indeed strongly reduced by global contrast inversion (Figure 7A, $P < 0.01$, t-test). Thus the prediction that global contrast inversion, by flipping all local feature polarities, would induce a low-firing rate for faces was verified. Surprisingly, responses to inverted contrast cropped objects were significantly larger compared to normal contrast cropped objects (Figure 7A, $P < 0.01$, t-test). One possible explanation is that face-selective cells receive inhibition from cells coding non-face objects, and the latter also exploit contrast-sensitive features in generating shape selectivity.

The role of external features in face detection

Behaviorally, it has been found that external features such as hair can boost performance in a face detection task (Torralba and Sinha, 2001). Up to now, all the experiments demonstrating the importance of contrast features for generating face-selective responses were performed using stimuli lacking external features (i.e., hair, ears, head outline). We next asked what the effect of global contrast inversion is for faces possessing external features. To our surprise, we found that the population average response to globally contrast-inverted faces possessing external features was almost as high as the average response to normal contrast faces ($P > 0.2$, t-test, Figure 7B). A significant increase in response latency was also observed ($P < 0.001$, t-test); the average latency (time to peak) for normal contrast faces was 106±29 ms, and 160±60 ms for contrast inverted faces. This result suggests that detection of external features provides an additional, contrast-independent mechanism for face detection, which can supplement contrast-sensitive mechanisms. In addition, we again noticed that images of globally contrast-inverted non-face objects elicited slightly higher responses compared to normal contrast objects ($P < 0.01$, t-test, Figure 7B); this was true for all object categories (hands, bodies, fruits, and gadgets), but not for scrambled patterns ($P > 0.05$, t-test, Figures S6A and S6B).

It seems plausible that the component which elicited the high firing rate in the inverted contrast uncropped faces was hair. To test this, we constructed artificial stimuli which were exactly like the original, but with black hair added (Figure 7C). This allowed us to directly

test the effect of adding hair on responses to stimuli with correct and incorrect contrast features (16 images per condition), and observe whether responses to hair can override responses to incorrect internal contrast features. We recorded 35 additional face-selective cells in monkey H; the average population response is shown in Figure 7C. When hair was added to incorrect contrast faces (magenta line), the response was delayed and almost as high as that to correct contrast faces without hair (Figure 7C, green line, $P > 0.3$, t-test, Figure 7D). This shows that a specific external feature, hair, can drive face-selective cells via a longer latency mechanism even when incorrect contrast is present in internal features.

Cell selectivity for the presence of a part depends on its luminance

Why do non-face images containing correct contrast relationships nevertheless elicit no response (Figures 6C-E)? What is the additional element present in a face that is lacking in these non-face images? One simple hypothesis is that the non-face images lack the correct contours, i.e., the presence of the correct face parts. A recent study examined in detail the coding of face parts in the middle face patches, and demonstrated that cells in this region are tuned for both the presence and geometry of different subsets of face parts (Freiwald et al., 2009). This conclusion was derived from two experiments exploiting cartoon faces: (1) Cells were presented with cartoon faces consisting of all possible combinations of seven basic parts (hair, bounding ellipse, irises, eyes, eyebrow, nose and mouth) and their sensitivity to the presence of specific parts was determined, (2) Cells were presented with cartoon faces in which the geometry of face parts was modulated along 19 different dimensions (e.g., iris size, intereye distance), and tuning was measured along each dimension.

To explore in detail the relationship between contrast tuning and selectivity for the presence of face parts within single face cells, we next repeated the experiments of Freiwald et al. in conjunction with our contrast tuning experiments. We hypothesized that tuning for the presence of a part depends not only on purely geometrical factors (i.e. the shape of the part), but also on part luminance or contrast relative to other parts. To test this hypothesis we presented three stimulus variants: (1) a parameterized face stimulus with correct contrast, (2) the same stimulus with fully inverted contrast, and (3) the original cartoon stimuli used in Freiwald et al., 2009 (“cartoon”) (Figure 8A); the first two stimuli were derived from the parameterized contrast stimulus introduced in Figure 2, but with eyebrows, irises and hair added to allow direct comparison to the third stimulus. For each variant we presented the decomposition of the face into seven basic parts (2^7 stimuli). Thus, we could directly compare the results of Freiwald et al. 2009 to our current results and test whether selectivity for the presence of specific face parts also depends on the contrast of those parts.

We recorded from 35 additional face-selective cells from monkey H. The responses of an example cell to the decomposition of all three stimuli (normal contrast, inverted contrast, cartoon) are shown in Figure 8A. We found that responses were similar between cartoon and normal contrast stimuli. Furthermore, we found that the inverted contrast decomposition elicited very different responses compared to the two normal contrast conditions. To determine whether the presence of a part played a significant role in modulating firing rate we performed seven way ANOVA with parts as the factors (similar to the analysis in Freiwald et al. 2009). Cells exhibited different tuning for parts for the three different stimulus variants (Figure 8B, 7-way ANOVA, $p<0.005$). To quantify the degree to which cells show similar tuning we counted the number of parts that were shared across two conditions. We found that cells were more likely to be tuned to the same part in the normal contrast and cartoon compared to inverted contrast and cartoon ($p < 0.001$, sign test). However, if a cell shows tuning for the presence of a part in the cartoon stimuli, this does not necessarily imply that it will also show preference for the same part in the artificial contrast stimuli (e.g., irises were found to be a significant factor for 16 cells in the correct contrast condition and 11 in the cartoon). More importantly, we found very different

preferences for presence of a part between the normal and inverted contrast conditions which cannot be explained by different shapes of the parts since they were exactly the same. For example, while irises were found to be a significant factor in 16 cells for the correct contrast condition, only one cell preferred irises in the incorrect contrast. Thus, preference for a specific part depends not only on the part shape (i.e. contour) but also on its luminance level relative to other parts.

Contrast features and geometrical features both modulate face cell tuning

The second major finding reported in Freiwald et al. 2009 was that cells are tuned to the metric shape of subsets of geometrical features, such as face aspect ratio, inter-eye distance, iris size, etc. Such features are thought to be useful for face recognition. Our present results suggest that face-selective cells use coarse-level contrast features to build a representation which might be useful for face detection. Are these two different types of features, contrast features and geometric features, encoded by different cells, or are the same cells modulated by both type of features?

To answer this, we repeated the Freiwald et al. experiment in which cartoon stimuli were simultaneously varied along 19 feature dimensions, and presented in addition our artificial face stimuli which varied in contrast (see Figure 2B). We recorded the responses of 35 face-selective cells (monkey J) and found similar ramp-shape tuning curves for subsets of geometrical feature dimensions as previously reported. The example cell in Figure 8C increased firing rate when aspect ratio dimension was modified, but not when the inter-eye distance changed (Figure S7A). To determine whether cells were significantly tuned for each one of the 19 geometrical feature dimensions we repeated the analysis described in Freiwald et al. 2009 and computed the heterogeneity index (Figure S7B, see Experimental Procedures).

Out of the 35 face-selective cells, 29 were modulated by at least one geometrical feature (Figures 8D and S7C), where the most common feature was aspect ratio (Figure S7D). Cells were also modulated by contrast polarity features (Figure 8D). Out of the 35 cells, 19 were modulated by at least one contrast polarity feature. Overall, 49% of the cells were modulated by both types of features (Figures 8E and S7E). Thus, tuning to low-spatial frequency coarse contrast features and to high spatial frequency geometrical features can co-occur in face-selective cells, suggesting that some cells encode information relevant for both detection and recognition.

Discussion

One of the most basic questions about face-selective cells in IT cortex is how they derive their striking selectivity for faces. Motivated by computational models for object detection that emphasize the importance of features derived from local contrast (Lienhart and Jochen, 2002; Sinha et al., 2006; Viola and Jones, 2001), this study focused on the question whether contrast features are essential for driving face-selective cells. Our main strategy was to probe cells with a parameterized stimulus set allowing manipulation of local luminance in each face part. The results suggest that detection of contrast features is a critical step used by the brain to generate face-selective responses. Four pieces of evidence support this claim. First, different combinations of contrasts could drive cells from no response to responses greater than that to a real face. Second, the polarity preference for individual features was remarkably consistent across the population in three monkeys. Third, the contrast feature preference followed with exquisite precision features that have been found to be predictive of the presence of a face in an image; these features are illumination invariant, agree with human psychophysics (Sinha et al., 2006), fMRI studies (George et al., 1999; Gilad et al., 2009), and are ubiquitously used in artificial real-time face detection (Lienhart and Jochen,

2002; Viola and Jones, 2001). Finally, the tuning to contrast features generalized from our artificial collage of parts to real face images.

Shape selectivity in IT has been proposed to arise from cells representing different feature combinations (Brincat and Connor, 2004; Fujita et al., 1992; Tanaka, 2003; Tsunoda et al., 2001). Elucidating exactly what features drive activity of randomly sampled cells in IT has been difficult due to the large space of shapes one needs to test (Kourtzi and Connor, 2011). Clever approaches such as parameterization of shape (Pasupathy and Connor, 2002) or genetic optimization (Yamane et al., 2008) are needed to make the problem tractable. Here, we took a different approach, focusing on a specific shape domain. The known shape selectivity of cells in face-selective regions in inferotemporal cortex allowed us to carefully test a specific computational model for generation of shape selectivity, the Sinha model (Sinha, 2002).

A plethora of computer vision systems have been developed to detect faces in images. We chose to test the Sinha model for the same reasons that Sinha proposed this scheme in the first place: it is motivated directly by psychophysical and physiological studies of the human visual system. Specifically, Sinha's model naturally accounts for (1) the robustness of human face detection to severe image blurring, (2) its sensitivity to contrast inversion, and (3) its holistic properties. The Sinha model provides a simple, concrete distillation of these three properties of human face detection. Thus it is an important model to test physiologically, and our study is the first to test its critical predictions.

Sinha's theory makes three straightforward predictions. First, at least a subset of face cells should respond to grossly simplified face stimuli. We found that 51% of face cells responded to a highly simplified 11-component stimulus and modulated their firing rate from no response to responses that were greater than that to a real face. Thus the first prediction of Sinha's theory was confirmed. Second, Sinha's theory predicts a subset of contrast polarity features to be useful for face detection. We found, first, that middle face patch cells selective for contrast across parts, were tuned for only a subset of contrasts. Second, all features predicted by Sinha were found to be important and were found with the correct polarity in all cases, and this was highly consistent across cells (Figures 4 and S4E). Thus, our results have a very strong form of consistency with Sinha's theory. A third prediction of Sinha's theory is that face representation is holistic: robust detection is a consequence of confirming the presence of multiple different contrast features. We found that the shapes of the detection templates used by many (though not all) cells indeed depended critically on multiple face parts and were thus holistic in Sinha's sense. Taken together, our results confirm the key aspects of the Sinha model and pose a tight set of restrictions on possible mechanisms for face detection used by the brain.

Despite these correspondences, our results also show that the brain does not implement an exact replica of the Sinha model. First, cells respond in a graded fashion as a function of the number of correct features, yet an all or none dependence is predicted by the model. Second, the simple Sinha model uses only 12 features to detect a face, while the population of middle face patch neurons encodes a larger number of features. Furthermore, these neurons do not respond to non-face images with 12 correct contrast features (Figure 6E), indicating additional mechanisms for detecting the presence of specific parts are in place.

Our results rule out alternative detection schemes. Models that use geometric, feature-based matching (Brunelli and Poggio, 1993) can be ruled out as incomplete, since not only the position of but also the contrast between features matters. The observation that some of our artificial face stimuli elicited responses stronger than that to a real face might also indicate that a fragment based approach (Ullman et al., 2002) is unlikely since that theory predicts

that the maximal observed response should be to a patch of a real face image and not to an artificial uniform luminance patch; in addition, the holistic nature of the contrast templates in the middle face patches (Figure 4D) suggests cells in this region are not coding fragments. However, our results do not rule out the possibility that alternative schemes might provide an accurate description for cells in earlier stages of the face processing system.

Surprisingly, we found the subjective category of “face” to be dissociated from the selectivity of middle face patch neurons. First, Figure 2 shows that a face-like collage of 11 luminance regions in which only the contrast between regions is modulated can drive a face cell from no response to a response greater than that to a real face. All of the stimuli used in this experiment, including the ineffective ones, would be easily recognizable as a face to any primate naïve to the goals of the experiment. Yet, despite the fast speed of stimulus update, face cells did not respond to “wrong contrast” states of the face. Second, in Figure 6 we show that real face images with incorrect contrast relationships elicited a much lower response than those with 12 correct relationships (indeed, on average, even faces with only 4 correct relationships yielded close to no response at all). Perceptually, all of the real face images are easily recognizable as faces. Thus, it seems that the human categorical concept of face is much less sensitive to contrast than the early detection mechanisms used by the face processing system.

Previous studies have found that global contrast inversion can either abolish responses in IT cells (Fujita et al., 1992; Ito et al., 1994; Tanaka, 1996; Tanaka et al., 1991), or have a small effect (Baylis and Driver, 2001; Rolls and Baylis, 1986). Our experiments shed some light on this apparent conflict, and suggest that at least for the case of faces, the response to global contrast inversion is highly dependent on the presence of external facial features. When external features are present, they can activate a contrast-independent mechanism for face detection. How internal and external features are integrated, however, remains unknown. One clue might be provided by the observation that middle face patch neurons respond to inverted contrast faces with external features with much longer latency. It is thus tempting to speculate that higher order face-selective regions are necessary for integrating internal and external facial features, yet, this remains to be validated in future experiments.

Our finding that cells are tuned to both contrast features and to geometrical features extends and complements the previous work by Freiwald et al. (Freiwald et al., 2009). The Freiwald et al. study probed cells with parameterized cartoon faces and revealed two important tuning characteristics of cells: they are tuned for the presence of different constellations of face parts and are further modulated by the geometric shape of features such as aspect ratio, inter-eye distance, etc. The cartoon stimuli used in that study contained significant contrast differences between parts (see Figure 8A), but the contrasts were held fixed, thus their contribution to face cell responses was left undetermined. The present study demonstrates the importance of having both correct contours and correct contrast to effectively drive face-selective cells. While contours alone can drive face-selective cells by a certain amount (Figure 5), correct contrast greatly increases the response, and under some circumstances may be necessary to elicit responses (Figures 6 and 8A).

The second main finding of the Freiwald et al. study was that cells are modulated by complex geometrical features encoded by high frequency information. The current study shows that cells are further modulated by coarse, low-level frequency contrast information. These two properties can in fact be represented in a single cell (Figure 8E), suggesting that cells may be encoding information that is useful both for detection of faces and recognition of individuals. Alternatively, such “dual” tuning characteristics could be a result of recognition processes occurring after detection processes, as predicted by computational models (Tsao and Livingstone, 2008); according to the latter view, cells with dual tuning

characteristics may nevertheless be contributing exclusively to recognition. Importantly, these two aspects of face cell tuning (tuning to coarse contrast features and tuning to high frequency geometrical contours) are not independent: images with correct contrast features but incorrect contours (Figure 6E), or correct contours but incorrect contrast features (Figure 8B), can both fail to elicit a significant response.

What mechanisms could provide the inputs for establishing the contrast sensitivity of face cells? Exploration of mechanisms for contour representation in area V4, a key area for mid-level object vision (Brincat and Connor, 2004; Pasupathy and Connor, 2002), suggests that cells in V4 are sensitive to contrast polarity (Pasupathy and Connor, 1999). These cells are plausible candidates to provide input to the contrast sensitive cells we observed. Direct recordings from the inputs to middle face patch cells, e.g., guided by in vivo tracer injections (Ichinohe et al., 2010) or antidromic identification (Hoffmann et al., 2009; Movshon and Newsome, 1996), will be necessary to elucidate the contour and contrast tuning properties of face cell inputs.

Faces are a privileged object class in the primate brain, impervious to masking (Loffler et al., 2005) and attracting gaze an order of magnitude more powerfully than other objects (Cerf et al., 2009). What is the chain of events that enables faces to capture the visual consciousness of a primate so powerfully? Our results shed new light on the nature of templates used by the brain to detect faces, revealing the importance of contrast features. An important question we have not addressed is how these detection templates are read out to drive behavior. We found that different cells encoded different contrast features, suggesting a population code is used to describe a single image. The diversity of contrast features coded by cells in the middle face patches suggests that pooling and readout may be a function of subsequent processing stages, i.e., the problem of face detection has not yet been entirely solved at this stage. Alternatively, cells with face detection capabilities matching perception may already exist in the middle face patches, but constitute a specialized subset which will require more refined targeting techniques to access. Behavioral evidence suggests that a powerful link should exist between face detection machinery and brain areas controlling attention, suggesting a possible approach for tracing the readout neurons.

Experimental Procedures

All procedures conformed to local and US National Institutes of Health guidelines, including the US National Institutes of Health Guide for Care and Use of Laboratory Animals.

Face Patch Localization

Two male rhesus macaques were trained to maintain fixation on a small spot for juice reward. Monkeys were scanned in a 3T TIM (Siemens) magnet while passively viewing images on a screen. MION contrast agent was injected to improve signal to noise ratio. Six face selective regions were identified in each hemisphere in both monkeys. Additional details are available in Tsao et al., 2006, Freiwald and Tsao, 2010 and Ohayon and Tsao 2011. We targeted middle face patches that are located on the lip of the superior temporal sulcus and in the fundus (Figure S1).

Visual Stimuli and Behavioral Task

Monkeys were head fixed and passively viewed the screen in a dark room. Stimuli were presented on a CRT monitor (DELL P1130). Screen size covered 21.6 x 28.8 visual degrees and stimulus size spanned 7 degrees. The fixation spot size was 0.25 degrees in diameter. Images were presented in random order using custom software. Eye position was monitored using an infrared eye tracking system (ISCAN). Juice reward was delivered every 2–4

seconds if fixation was properly maintained. We presented in rapid succession (5 images / s) a set of 16 real face images, 80 images of objects from non-face categories (fruits, bodies, gadgets, hands, scrambled images) and 432 images of a parameterized face. Each image was presented 3–5 times to obtain reliable firing rate statistics.

Parameterized face stimuli generation

The parameterized face stimuli were generated by manual segmentation of an average face. Each part was given a unique intensity level ranging between dark (0.91 cd/m^2) and bright (47 cd/m^2). We generated our stimuli using an iterative search algorithm that aimed to cover all possible pair-wise combinations of part intensities with the minimal number of permutations. That is, our data set contained at least one exemplar for every possible part-pair (55), and every possible intensity level (11x11). We used a greedy approach: starting with a single random permutation, we added the next permutation that contained the needed intensity values (if more than one was found, a random decision was made). In this way, we were able to reduce the number of possible combinations from 6655 (55x11x11) to 432. Each condition used for the analysis (intensity in Part A > intensity in Part B) aggregated on average 214 ± 8 stimuli. The stimulus set did not contain an intensity bias toward any of the parts. A 1-way ANOVA revealed that the mean intensity in each part did not significantly deviate from all other parts (P > 0.5).

Neural Recording

Tungsten electrodes (18–20 Mohm at 1 kHz, FHC) were back loaded into metal guide tubes. Guide tubes length was set to reach approximately 3–5 mm below the dura surface. The electrode was advanced slowly with a manual advancer (Narishige Scientific Instrument). Neural signals were amplified and extracellular action potentials were isolated using the box method in an online spike sorting system (Plexon). Spikes were sampled at 40 kHz. All spike data was re-sorted with off-line spike sorting clustering algorithms (Plexon). Only well-isolated units were considered for further analysis.

Data analysis

Data analysis was performed using custom scripts written in C and MATLAB (MathWorks).

A trial was considered to be the time interval from one stimulus onset to the next (200 ms). We discarded all trials in which the maximal deviation from the fixation spot was larger than 3 degrees. PSTHs were smoothed with a Gaussian kernel (σ = 15ms). Unless otherwise stated, stimulus response was computed by averaging the interval [50, 250] ms relative to stimulus onset and subtracting the preceding baseline activity, which was estimated in the interval [0, 50] ms.

We estimated cells ability to discriminate face images from non-face images using d'. d' was computed by $d' = \sqrt{2}Z^{-1}(AUC)$, where AUC is the area under the ROC curve and Z^{-};[1] is the normal inverse cumulative distribution function (AUC was ensured to be above 0.5 to capture units that were inhibited by faces as well). d' is more sensitive than our previously used face selectivity index (FSI) (Tsao et al., 2006), since it takes into account the response variance. Different measures of face selectivity yielded similar numbers of face-selective cells: 267/280 using the Area Under Curve (AUC) measure from signal detection theory (Figure S2A, AUC > 0.5, permutation test, p<0.05), and 298/342 using the Face Selective Index (FSI) measure (Figure S2B, FSI > 0.3). Similar results were obtained when cells were selected according to d', AUC, or FSI.

Unless otherwise stated, population average response was computed by normalizing each cell to the maximal response elicited by any of the probed stimuli.

Polarity consistency index

Given a contrast polarity feature across two parts (A,B), we counted how many cells fired significantly stronger ($P < 10^{-5}$, Mann-Whitney U-test) for the condition A>B vs. the condition A<B, and normalized the number to be between zero and one:

$Index = \left| \frac{\#(A>B) - \#(A<B)}{\#(A>B) + \#(A<B)} \right|$. An index of one corresponds to all cells preferring the same polarity direction and an index of zero corresponds to half of the population preferring A>B and the other half preferring A<B.

Determining geometrical feature significance

For each cell and feature dimension we computed time-resolved post-stimulus tuning profiles, (such as the ones shown in Figure 8C) over three feature update cycles (300 ms) and 11 feature values. Profiles were smoothed with a 1D Gaussian (5 ms) along the time axis. To determine significance we used an entropy-related measure called heterogeneity (Freiwald et al., 2009). Heterogeneity is derived from the Shanon-Weaver diversity index

$$-\sum_{i=1}^{k} p_i \log(p_i)$$

and is defined as $H = 1 - \frac{-\sum_{i=1}^{k} p_i \log(p_i)}{\log(k)}$, where k is the number of bins in the distribution (11 in our case) and p_i the relative number of entries in each bin. If all p_i values are identical, heterogeneity is 0, and if all values are zero except for one, heterogeneity is 1. Computed heterogeneity values were compared against a distribution of 5,016 surrogate heterogeneity values obtained from shift predictors. Shift predictors were generated by shifting the spike train relative to the stimulus sequence in multiples of the stimulus duration (100 ms). This procedure preserved firing rate modulations by feature updates, but destroyed any systematic relationship between feature values and spiking. From the surrogate heterogeneity distributions, we determined significance using Efron's percentile method; for an actual heterogeneity value to be considered significant, we required it to exceed 99.9% (5,011) of the surrogate values. A feature was considered significant if heterogeneity was above the surrogate value for continuous 15 ms. For additional information please refer to (Freiwald et al., 2009).

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

References

Baylis GC, Driver J. Shape-coding in IT cells generalizes over contrast and mirror reversal, but not figure-ground reversal. Nat Neurosci. 2001; 4:857–858. [PubMed: 11528408]

Brincat SL, Connor CE. Underlying principles of visual shape selectivity in posterior inferotemporal cortex. Nat Neurosci. 2004; 7:880–886. [PubMed: 15235606]

Brunelli R, Poggio T. Face Recognition: Features versus Template. IEEE Transactions on PAMI. 1993; 15:1042–1052.

Cerf M, Frady EP, Koch C. Faces and text attract gaze independent of the task: Experimental data and computer model. J Vis. 2009; 9:10, 11–15. [PubMed: 20053101]

Desimone R, Albright CG, Gross CG, Bruce C. Stimulus-Selective properties of inferior temporal neurons in the Macaque. J Neurosci. 1984; 4:2051–2062. [PubMed: 6470767]

Freiwald WA, Tsao DY. Functional compartmentalization and viewpoint generalization within the macaque face processing system. Science. 2010; 330:845–851. [PubMed: 21051642]

Freiwald WA, Tsao DY, Livingstone MS. A face feature space in the macaque temporal lobe. Nat Neurosci. 2009; 12:1187–1198. [PubMed: 19668199]

Fujita I, Tanaka K, Ito M, Cheng K. Columns for visual features of objects in monkey inferotemporal cortex. Nature. 1992; 360:343–346. [PubMed: 1448150]

George N, Dolan RJ, Fink GR, Baylis GC, Russell C, Driver J. Contrast polarity and face recognition in the human fusiform gyrus. Nat Neurosci. 1999; 2:574–580. [PubMed: 10448224]

Gilad S, Meng M, Sinha P. Role of ordinal contrast relationships in face encoding. Proc Natl Acad Sci. 2009; 106:5353–5358. [PubMed: 19276115]

Harmon LD, Julesz B. Masking in visual recognition: effects of two-dimensional filtered noise. Science. 1973; 180:1194–1197. [PubMed: 4707066]

Heinrich SP, Bach M. Less is more: subjective detailedness depends on stimulus size. J Vis. 2010; 10:2. [PubMed: 20884467]

Heisele, B.; Poggio, T.; Pontil, M. Face Detection in Still Gray Images. Center for Biological and Computational Learning, MIT; 2000.

Hoffmann KP, Bremmer F, Distler C. Visual response properties of neurons in cortical areas MT and MST projecting to the dorsolateral pontine nucleus or the nucleus of the optic tract in macaque monkeys. Eur J Neurosci. 2009; 29:411–423. [PubMed: 19200243]

Ichinohe, N.; Sato, T.; Tanifuji, M. In vivo connection imaging and its application to monkey temporal face system. Paper presented at: SFN; San Diego. 2010.

Ito M, Fujita I, Tamura H, Tanaka K. Processing of contrast polarity of visual images in inferotemporal cortex of the Macaque monkey. Cereb Cortex. 1994; 4:499–508. [PubMed: 7833651]

Kourtzi Z, Connor CE. Neural representations for object perception: structure, category, and adaptive coding. Annu Rev Neurosci. 2011; 34:45–67. [PubMed: 21438683]

Lienhart R, Jochen M. An extended set of Haar-like features for rapid object detection. ICIP. 2002:900–903.

Loffler G, Gordon GE, Wilkinson F, Goren D, Wilson HR. Configural masking of faces: evidence for high-level interactions in face perception. Vision Res. 2005; 45:2287–2297. [PubMed: 15924942]

Logothetis NK, Sheinberg DL. Visual object recognition. Annu Rev Neurosci. 1996; 19:577–621. [PubMed: 8833455]

Moeller S, Freiwald WA, Tsao DY. Patches with links: a unified system for processing faces in the Macaque temporal lobe. Science. 2008; 320:1355–1359. [PubMed: 18535247]

Movshon JA, Newsome WT. Visual response properties of striate cortical neurons projecting to area MT in macaque monkeys. J Neurosci. 1996; 16:7733–7741. [PubMed: 8922429]

Ohayon S, Tsao DY. MR-Guided Stereotactic Navigation. J Neurosci Methods. 2011; 204(2):389–397. [PubMed: 22192950]

Papageorgiou, CP.; Oren, M.; Poggio, T. General framework for object detection. ICCV Proc of the 6th Inter Conf on Computer Vision; 1998. p. 555-562.

Pasupathy A, Connor CE. Responses to contour features in macaque area V4. J Neurophysiol. 1999; 82(5):2490–2502. [PubMed: 10561421]

Pasupathy A, Connor CE. Population coding of shape in area V4. Nat Neurosci. 2002; 5:1332–1338. [PubMed: 12426571]

Rolls ET, Baylis GC. Size and contrast have only small effects on the responses to faces of neurons in the cortex of the superior temporal sulcus of the monkey. Exp Brain Res. 1986; 65:38–48. [PubMed: 3803509]

Sinha, P. Qualitative representations for recognition. Proc Annu Workshop on Biologically Motivated Computer Vision; 2002. p. 249-262.

Sinha, P.; Balas, B.; Ostrovsky, Y.; Russell, R. Face recognition by humans: nineteen results all computer vision researchers should know about. Proc of the IEEE; 2006. p. 94

Tanaka K. Inferotemporal cortex and object vision. Annu Rev Neurosci. 1996; 19:109–139. [PubMed: 8833438]

Tanaka K. Columns for complex visual object features in the inferotemporal cortex: clustering of cells with similar but slightly different stimulus selectivities. Cereb Cortex. 2003; 13:90–99. [PubMed: 12466220]

Tanaka K, Saito H, Fukada Y, Moriya M. Coding visual images of objects in the inferotemporal cortex of the Macaque monkey. J Neurophy. 1991; 66:170–189.

Torralba, A.; Sinha, P. Detecting faces in impoverished images (In AI Memo 2001–028, CBCL Memo 208). 2001.

Tsao DY, Freiwald WA, Knutsen TA, Mandeville JB, Tootell RBH. Faces and objects in macaque cerebral cortex. Nat Neurosci. 2003; 6:989–995. [PubMed: 12925854]

Tsao DY, Freiwald WA, Tootell RBH, Livingstone MS. A cortical region consisting entirely of face-selective cells. Science. 2006; 670:670–674. [PubMed: 16456083]

Tsao DY, Livingstone M. Mechanisms of face perception. Ann Rev Neurosci. 2008; 31:411–437. [PubMed: 18558862]

Tsao DY, Moeller S, Freiwald WA. Comparing face patch systems in macaques and humans. Proc Natl Acad Sci. 2008; 105:19514–19519. [PubMed: 19033466]

Tsunoda K, Yamane Y, Nishizaki M, Tanifuji M. Complex objects are represented in macaque inferotemporal cortex by the combination of feature columns. Nat Neurosci. 2001; 4:832–838. [PubMed: 11477430]

Ullman S, Vidal-Naquet M, Sali E. Visual features of intermediate complexity and their use in classification. Nat Neurosci. 2002; 5:1–6. [PubMed: 11753407]

Viola P, Jones M. Rapid object detection using a boosted cascade of simple features. CVPR. 2001:8–14.

Yamane Y, Carlson ET, Bowman KC, Wang Z, Connor CE. A neural code for three-dimensional object shape in macaque inferotemporal cortex. Nat Neurosci. 2008; 11:1352–1360. [PubMed: 18836443]

Highlights

- Cells are tuned for specific contrast relationships across face parts and the preference is for specific features that have been shown by past computational work to be useful for face detection under varying illumination.
- Although individual cells encode different sets of feature combinations, the population as a whole is consistent in the preferred contrast polarity of each feature.
- Correct contrast relationships in an image are necessary, but not sufficient, for a face-selective cell to fire.
- Face-selective cells in the middle face patch are modulated by both coarse contrast information useful for detection and by high frequency contour information useful for recognition.
- Contrast inversion induces low firing rate only when external features (e.g., hair) are absent. The presence of hair can override tuning for contrast features.

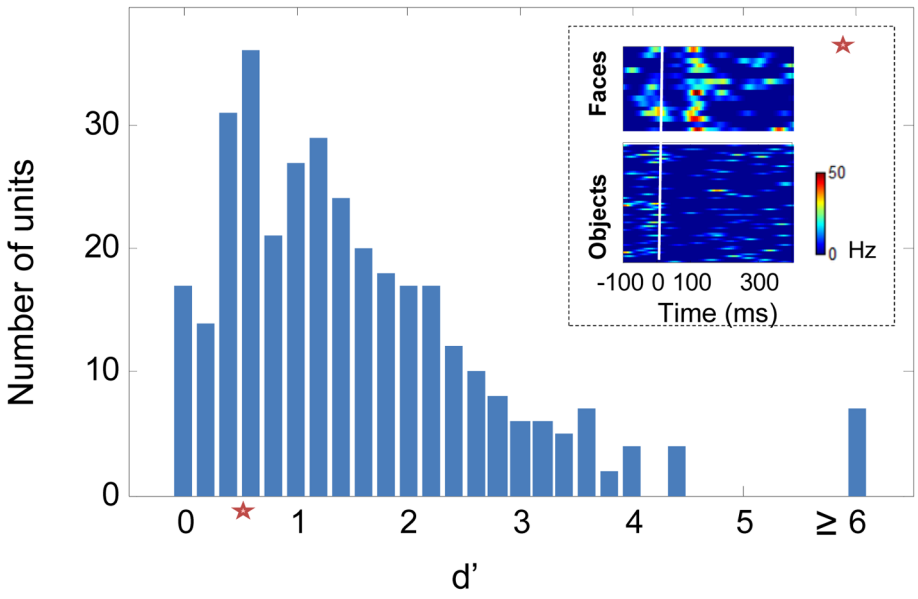

Figure 1.

Face discriminability histogram for 342 recorded cells from three monkeys. Discriminability between face and non-face images was quantified with the d' measure. 16 images of faces and 80 images of non-face objects were presented to the monkey in random order. The response for each image was estimated as the average firing rate between [50,250] ms relative to stimulus onset, minus baseline activity between [0,50] ms. Inset depicts responses of an example cell with d'=0.66 (denoted by a red star) to face and object images. Each line represents the PSTH for a given image. All cells with d' > 0.5 were considered to be face-selective.

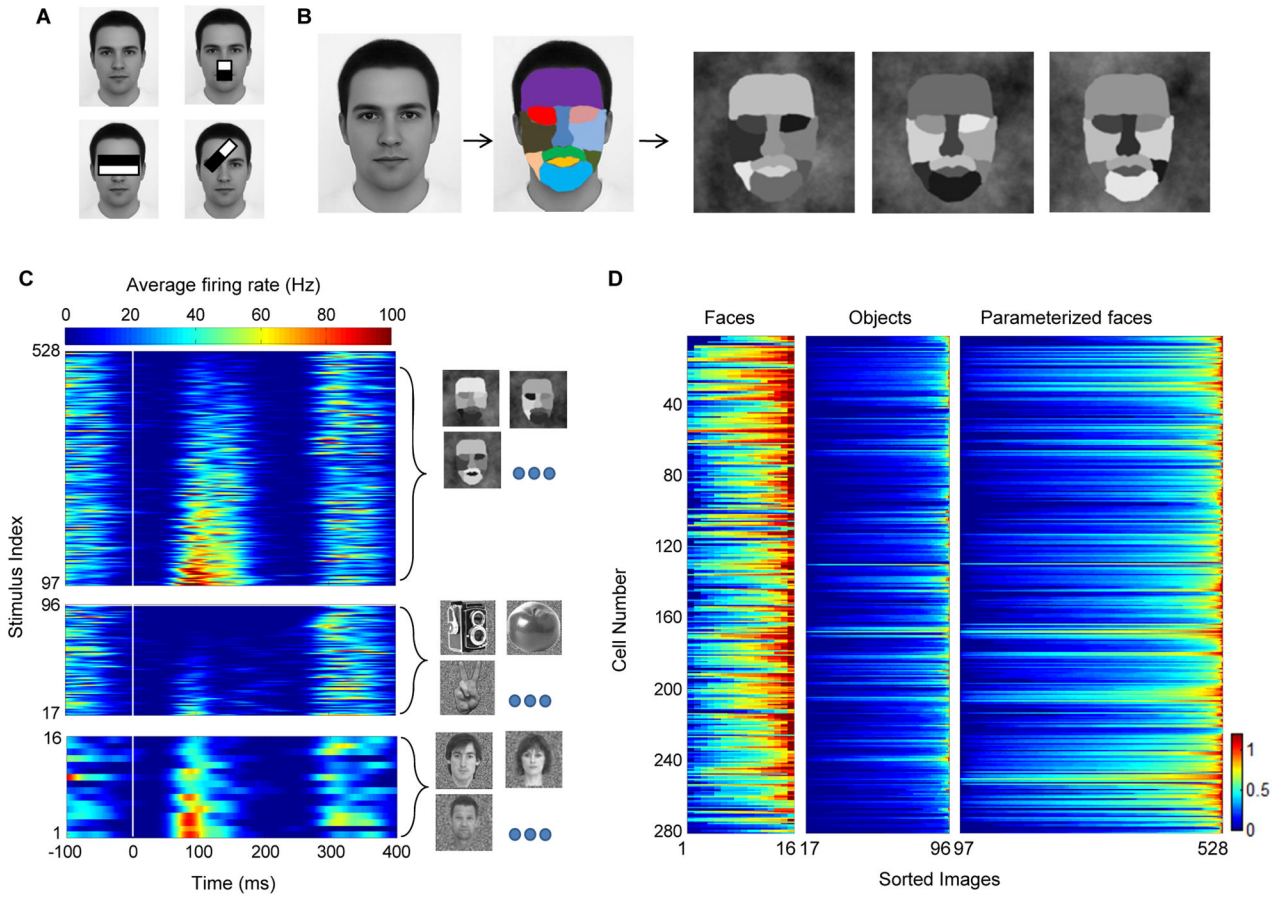

Figure 2.

Responses to artificial parameterized face stimuli. (a) Features proposed by computational models for face detection. Each contrast feature has two subparts. The value of a feature is evaluated by summing and subtracting the intensity levels in its sub-region components. (b) Construction of a parameterized face that was used to probe cells for effects of local contrast. An average face was segmented into eleven subparts. Each part was assigned a unique intensity level. Three different instances are shown. (c) PSTH of a single cell to the 432 artificial face stimuli, 80 object stimuli and 16 face stimuli (sorted by mean response magnitude). Images were presented at time zero (white vertical line), for 100 ms and were followed by a gray screen for an additional 100 ms. (d) Normalized average firing rate estimated between [50, 250] ms relative to stimulus onset for all recorded cells in three monkeys. Each row represents one cell. Each group (faces, objects, parameterized faces) was sorted such that the maximal firing rate is presented on the right for each cell (entries in each column do not correspond to the same stimulus).

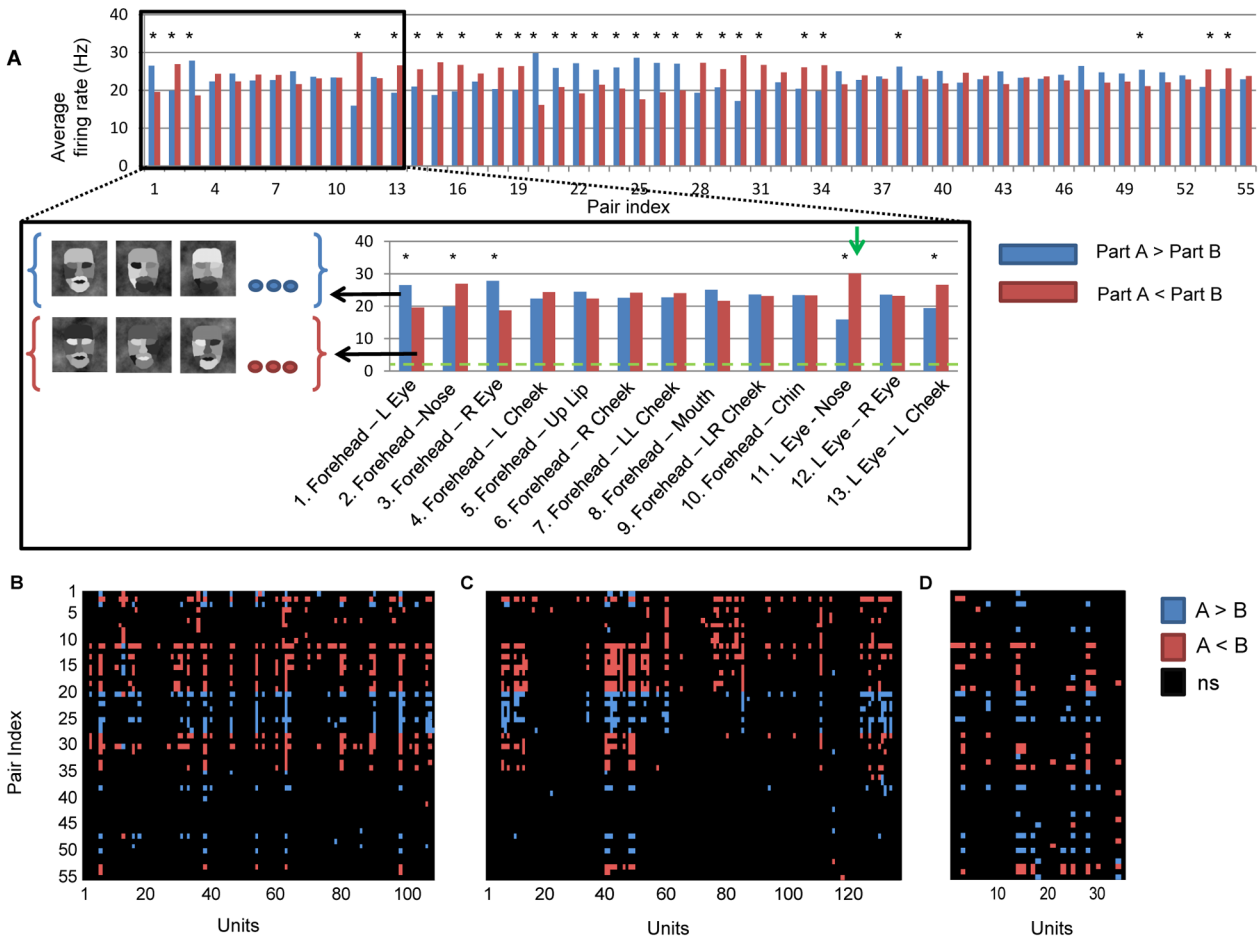

Figure 3.

Single cell tuning for contrast polarity features. (a) Parameterized face stimuli were grouped according to whether the intensity in part A was greater or smaller than the intensity in part B. Blue bars represent the firing rate in the condition A>B, and red bars represent the firing rate in the condition B<A. Average firing rate (baseline subtracted) of an example cell to the two polarity conditions across all part pairs is presented (* $P < 10^{-5}$, Mann-Whitney test). Inset shows the first 13 part pairs with several examples of stimuli used in the averaging of the pair (Forehead-Left eye). Green horizontal line represents baseline activity, and the green arrow represents the largest firing rate difference (15 Hz) (b,c,d) Tuning matrices for monkey (R,H, J) representing which part pair was found to be significant. Blue (red) pixels represent significant tuning for the A>B (A<B) condition.

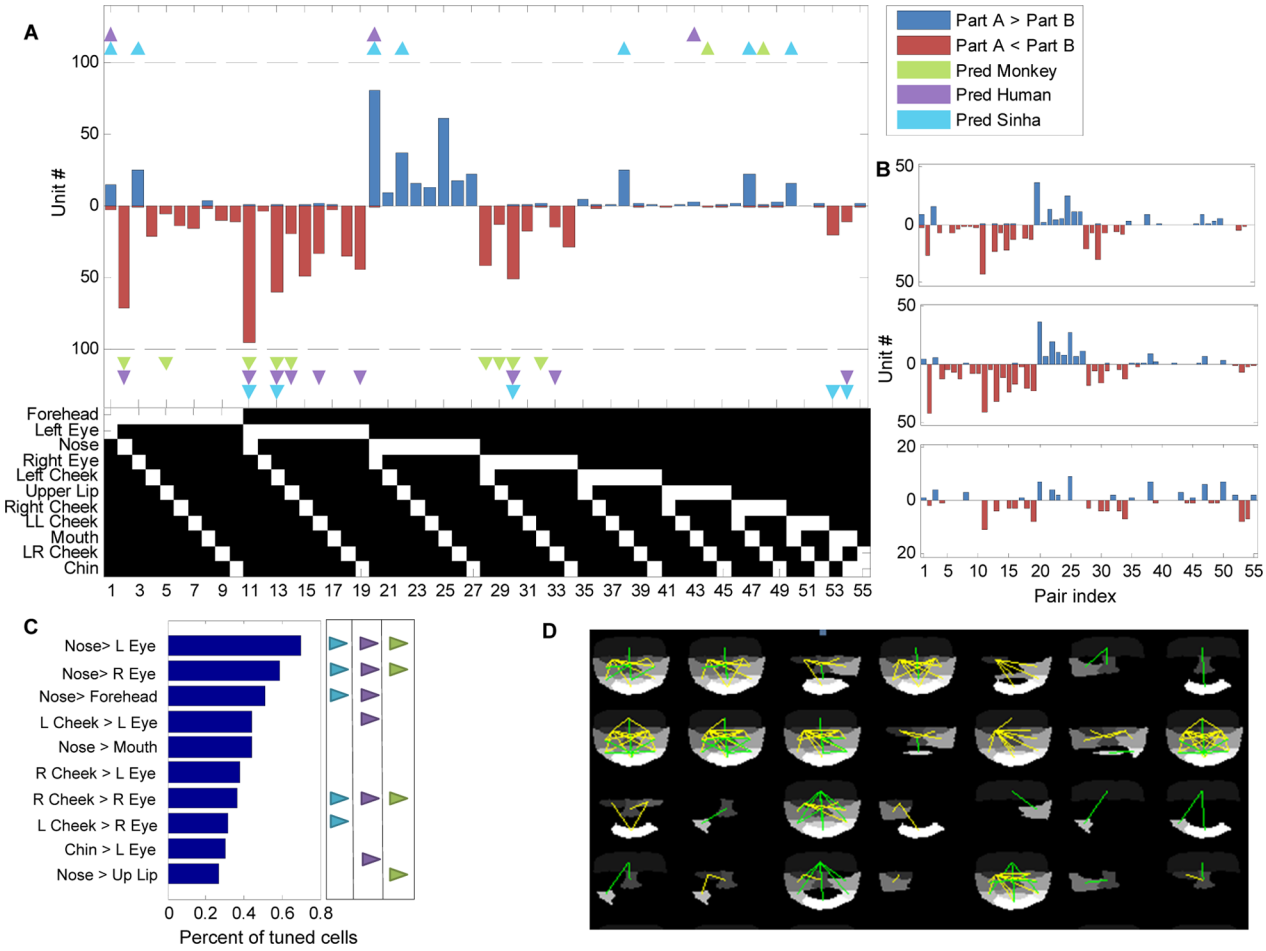

Figure 4.

Consistency in contrast polarity preference. (a) Significant contrast feature histogram (data pooled from all three monkeys). Blue (red) bars indicate the number of cells tuned for intensity in part A greater (less) than intensity in part B. Triangles indicate three different feature polarity direction predictions (see Supplementary Fig. 4, main text). The binary table below the histogram denotes the two parts that define each of the 55 pairs; the upper bit represents part A and the lower bit represents part B. (b) Significant contrast feature histogram for each of the three monkeys (R, H, J, from top to bottom). (c) Most common features and their preferred polarity across the population of cells that were tuned for at least one feature. Model predictions (and prediction's directionality) are represented by small triangles on the right (same convention as in (a)). (d) Graphical representation of feature tuning for a subset of random cells. Yellow lines represent features involving the eye region; green lines represent features which do not involve the eye region.

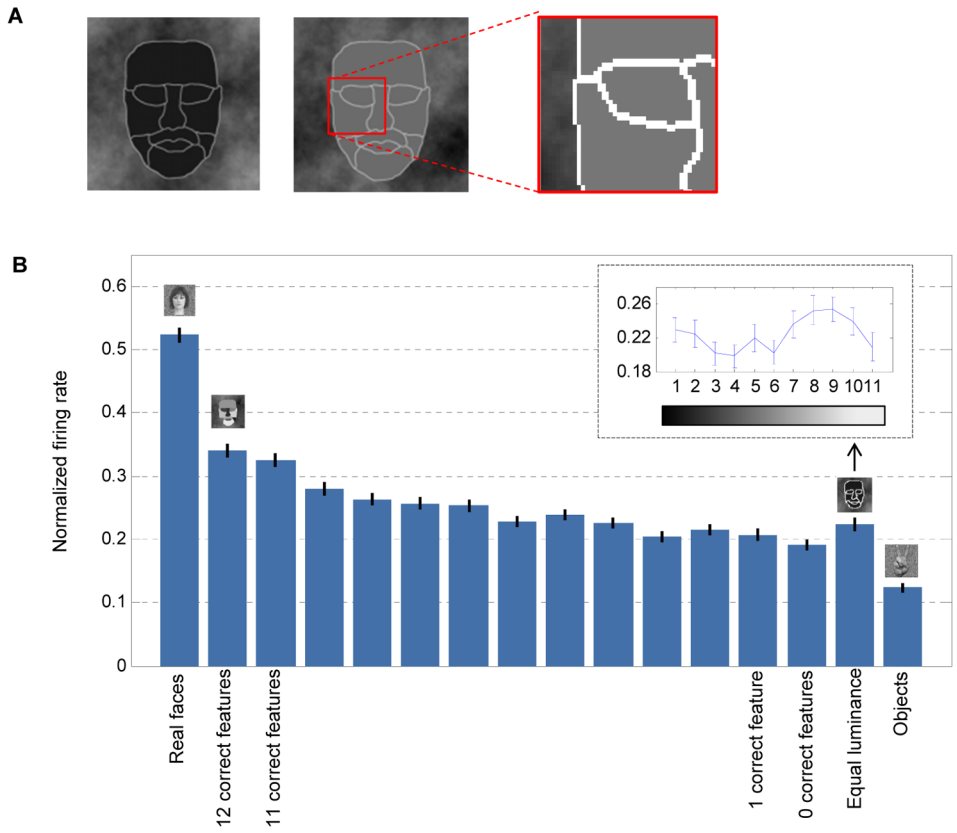

Figure 5.

Contribution of contours vs. contrast to firing rate of cells. (a) Two instances of an equal-luminance parameterized face. Each part has the same intensity level and all contours share the same (but brighter) intensity. (b) Normalized average firing rate (mean ± SEM) for 138 cells that were tuned for at least one contrast polarity feature (pooled across all monkeys). Firing rate of each cell was normalized to the stimulus which elicited the maximal response. Normal parameterized face stimuli are sorted by the number of their correct contrast polarity features. Correct contrast features were considered according to the Sinha model. Small inset shows firing rate variations for the equal luminance variant as a function of intensity level.

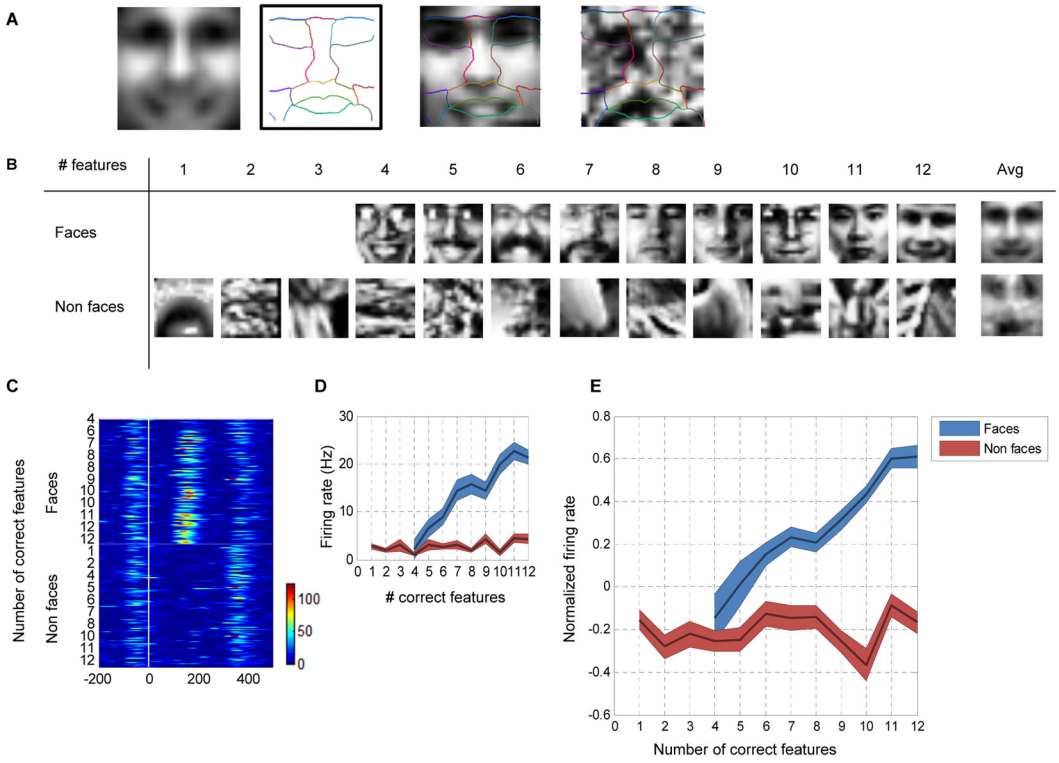

Figure 6.

Responses to real face and non-face images as a function of the number of correct contrast features. (a) Left to right: average face computed by averaging all face images in the data set; manual delineation of parts based on the average face; an instance of a face with the template overlaid; an instance of a non-face with the same template overlaid. (b) Examples of face and non-face images with indicated number of correct features (according to Sinha's model). Last column (Avg) shows the result of averaging all images containing 12 correct features. (c) Single cell PSTH to 207 face and 204 non-face images, sorted by the number of correct contrast features in each stimulus. (d) Average firing rate of the example cell shown in (c), as a function of the number of correct features for faces (blue curve) and non-faces (red). Firing rate was averaged on the interval [50,250] ms without baseline subtraction. Shaded area denotes standard error of the mean. (e) Population normalized firing rate (baseline subtracted) to face and non-face images as a function of the number of correct contrast features.

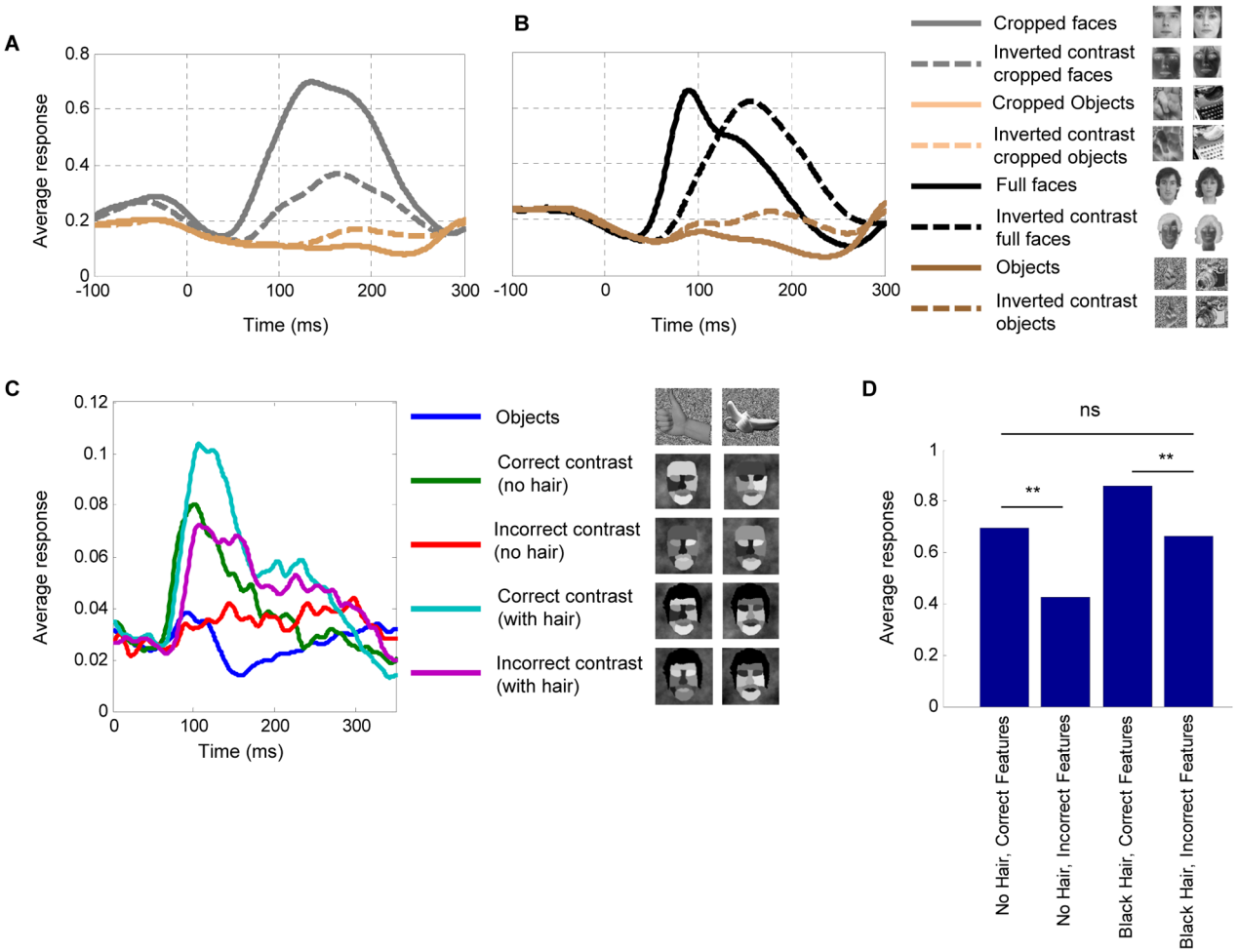

Figure 7.

Responses to global contrast inversion. (a) Average population response of 20 cells to normal and inverted contrast real faces and objects. (b) Average population response of 20 cells to normal and inverted contrast cropped faces and cropped objects. Two exemplars from each category are shown in the legend. (c) Average population response of 35 cells to the artificial stimuli controls testing the effect of hair on internal contrast features. Two exemplars from each category are shown in the legend. (d) Average firing rate across the four conditions (** t-test, $p<0.01$).

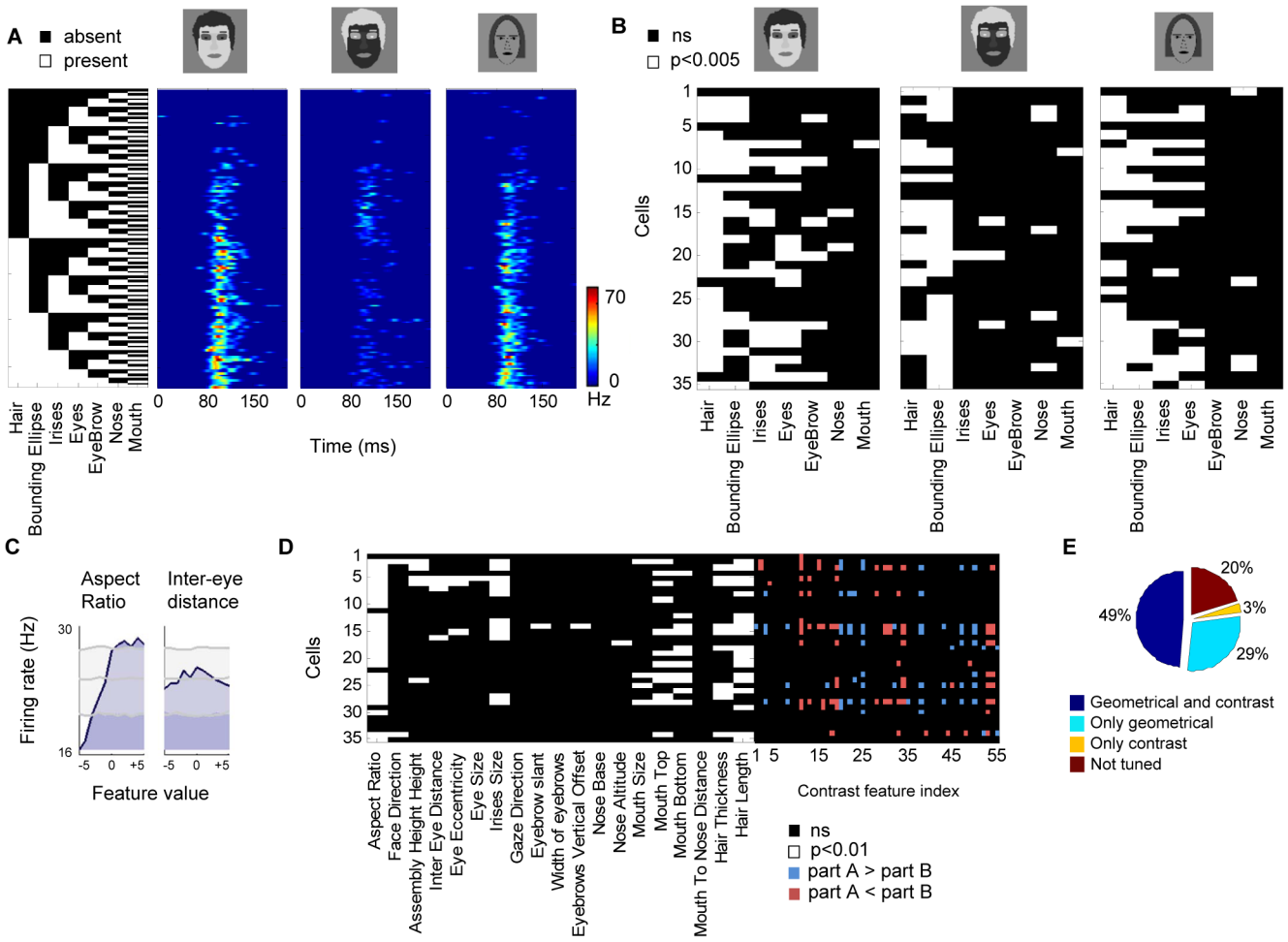

Figure 8.

Relationship between tuning to part contrast, part presence, and part geometry. (a) Responses of a single cell to a decomposition of a face stimulus with correct contrast (left), inverted contrast (middle) and cartoon (right). For each row, the parts present are indicated by the white squares in the black and white matrix. (b) Significant tuning of all cells to presence of parts across the three stimulus conditions (7-way ANOVA, $p<0.005$). Each row represents a single cell and its tuning to parts across the three different decompositions. The cell shown in (a) is represented in the last row. (c) Tuning for geometrical features. Tuning of an example cell to two feature dimensions (aspect ratio, inter-eye distance); the tuning curve (blue) is shown at a delay corresponding to maximal modulation. Maximal, minimal and mean values from the shift predictor are shown in gray. (d) Significant geometrical feature tuning across all 35 cells (each row represents tuning of a single cell). Right block, tuning of the same cells to contrast polarity features. (e) Percentage of cells tuned for geometrical and contrast features.