

# Piecing it together: Infants' neural responses to face and object structure

**Faraz Farzin**

Department of Psychology, Stanford University, Stanford, CA, USA



**Chuan Hou**

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



**Anthony M. Norcia**

Department of Psychology, Stanford University, Stanford, CA, USA



Integration of local elements into a coherent global form is a fundamental aspect of visual object recognition. How the different hierarchically organized stages of visual analysis develop in order to support object representation in infants remains unknown. The aim of this study was to investigate structural encoding of natural images in 4- to 6-month-old infants and adults. We used the steady-state visual evoked potential (ssVEP) technique to measure cortical responses specific to the global structure present in object and face images, and assessed whether differential responses were present for these image categories. This study is the first to apply the ssVEP method to high-level vision in infants. Infants and adults responded to the structural relations present in both image categories, and topographies of the responses differed based on image category. However, while adult responses to face and object structure were localized over occipitotemporal scalp areas, only infant face responses were distributed over temporal regions. Therefore, both infants and adults show object category specificity in their neural responses. The topography of the infant response distributions indicates that between 4 and 6 months of age, structure encoding of faces occurs at a higher level of processing than that of objects.

**Keywords:** object perception, face processing, lateral occipital cortex, image structure, steady-state evoked potentials

**Citation:** Farzin, F., Hou, C., & Norcia, A. M. (2012). Piecing it together: Infants' neural responses to face and object structure. *Journal of Vision*, 12(13):6, 1–14, <http://www.journalofvision.org/content/12/13/6>, doi:10.1167/12.13.6.

## Introduction

A critical function of the human visual system is to spatially and temporally integrate local image elements into a global form percept in order to accurately and meaningfully represent the identity of an object. Formation of this visual representation occurs effortlessly within a fraction of a second and allows us to recognize objects across widely varying viewing conditions including differences in size, location, orientation, lighting, and color.

Findings from decades of neuropsychological, neurophysiological, and neuroimaging research have shed light on the anatomical and functional organization underlying visual object processing in adult human and nonhuman primates. From these studies we have learned that visual object processing occurs along the ventral pathway of the brain, beginning from primary visual cortex through extrastriate areas such as V2 and V4, where local analysis of individual features and conjunctions of features takes place, and finally extending anteriorly into temporal cortex, where object

categorization and ultimately semantic identification occurs (Grill-Spector et al., 1998; Haxby, Hoffman, & Gobbini, 2000; Kanwisher, 2010; Malach et al., 1995; Puce, Allison, Asgari, Gore, & McCarthy, 1996).

Further, a collection of regions in lateral and ventral occipitotemporal cortex has been found to selectively process the complex structure of a shape, and less so the simple features comprising a shape (Kourtzi & Kanwisher, 2000, 2001; Lerner, Hendler, Ben-Bashat, Harel, & Malach, 2001). Form-specific activity has been most clearly demonstrated by reduced responses to scrambled versions of shapes that contain the same simple elements as the intact shape, such as lines and edges, but that lack the global form. Beyond cortical specificity for encoding object structure, functional magnetic resonance imaging (fMRI) studies have shown distinct response patterns for single, basic-level object categories. For example, brain regions within the lateral occipital complex (LOC) show category-specific activity for faces (fusiform face area, FFA; Kanwisher, McDermott, & Chun, 1997; McCarthy, Puce, Belger, & Allison, 1999), places and scenes (parahippocampal place area, PPA; Aguirre, Zarahn, & D'Esposito, 1998;

Epstein & Kanwisher, 1998), body parts (extrastriate body area; Downing, Bray, Rogers, & Childs, 2004), and tools (Beauchamp, Lee, Haxby, & Martin, 2002; Martin, Wiggs, Ungerleider, & Haxby, 1996). In these studies, category-selective activity was found by contrasting patterns of the BOLD signal induced from images belonging to one category with that of another category, or patterns of responses to intact versus within-category scrambled images or noise. Depending on the contrasted categories, different levels of functional specialization have been ascribed to different areas.

Visual event-related potential (ERP) studies, which measure transient changes in the brain's electrical activity that are time-locked to the presentation of an image, have also been used to examine category-specific responses. Studies using ERP have consistently found that faces elicit a negative potential around 150–200 ms (N170) post-stimulus onset that is maximal over occipitotemporal scalp regions (Bentin, Allison, Puce, Perez, & McCarthy, 1996; Rossion & Jacques, 2008). The “N170” is characterized by a larger amplitude and shorter latency in response to the presentation of faces versus nonface objects (Bentin et al., 1996; Botzel, Schulze, & Stodieck, 1995; Carmel & Bentin, 2002; Eimer, 2000b; George, Evans, Fiori, Davidoff, & Renault, 1996; Itier & Taylor, 2004; Jeffreys, 1989; Rousselet, Husk, Bennett, & Sekuler, 2008). In addition, a face-inversion effect, comprised of a delayed and more negative N170 response to inverted faces but not to inverted objects, has been considered a marker for face-specific processing (Bentin et al., 1996; Eimer, 2000a; Jacques, d'Arripe, & Rossion, 2007; Rossion et al., 2000). In these studies, the defining signature of category-specificity has typically been differences in amplitude or timing of an ERP component between two stimulus categories, presumably reflecting differences in perceptual processing of structural information. However, it has been argued that the N170 may not constitute category-specific neural processing per se, as there are often differences in basic low-level properties within and between categories that contribute to the response (Johnson & Olshausen, 2003; Pernet, Schyns, & Demonet, 2007; Rousselet, Gaspar, Wiczorek, & Pernet, 2011; Rousselet, Husk, Bennett, & Sekuler, 2007; Rousselet et al., 2008; Rousselet, Pernet, Caldara, & Schyns, 2011; VanRullen & Thorpe, 2001). Nevertheless, collectively, multiple lines of evidence support the conclusion that despite similar computational processing of individual local visual features, there exist qualitatively and quantitatively selective neural mechanisms for extracting higher-level structural information across image categories.

Due to the limited applicability of brain imaging techniques to infants, less is known about the emergence of category-selective visual processing in

the developing brain. Specifically, how do different cortical areas process specific object structures early in development? As in adults, much of the work examining neural systems underlying object recognition has focused on face processing, and the nature of infants' face-specific responses has been most extensively characterized using ERPs. ERP studies with infants have revealed two distinct components, the N290 and the P400, that become increasingly sensitive to upright human faces between 3 and 12 months of age. These components have therefore been considered precursors to the child (found at 4 years; Kuefner, de Heering, Jacques, Palmero-Soler, & Rossion, 2010; Taylor, McCarthy, Saliba, & Degiovanni, 1999) and adult face-sensitive N170 component described above (de Haan, Johnson, & Halit, 2003; de Haan, Pascalis, & Johnson, 2002; Halit, de Haan, & Johnson, 2003; Scott & Monesson, 2009; Scott, Pascalis, & Nelson, 2007). However, the evidence for these face-specific components originates from comparison between upright human faces and other face-related stimuli such as inverted faces, monkey faces, familiar or unfamiliar faces, or scrambled faces. While these face-related control stimuli were matched on some low-level properties, these studies did not determine category-specific processing per se because the comparison between test and control images was within the same category.

Less than a handful of studies have used images from different basic-level categories to investigate category-specific cortical processing in infants (de Haan & Nelson, 1999; McCleery, Akshoomoff, Dobkins, & Carver, 2009). In the first and one of the only studies to assess differences in ERP components in response to faces versus objects, 6-month-olds were shown familiar and unfamiliar faces or objects (de Haan & Nelson, 1999). This study found that the P400 component peaked earlier over the midline occipital electrode (Oz) for faces than for objects, independent of familiarity. de Haan and Nelson (1999) also showed that the topography of the differential response was unlike that of the adult N170 face response found over posterior temporal regions because it was medial rather than lateral. Limitations of the study that should be considered when interpreting the findings are that it was not a within-infant design, did not include an adult comparison group, some of the toys contained faces, and there were confounded low-level differences between stimuli categories in structure, spatial frequency, and color. McCleery et al. (2009) used a within-subject design with 10-month-old infants and found no significant difference in N290 latency, but did find larger N290 amplitude for faces compared with objects. In line with the de Haan and Nelson (1999) finding, 10-month-olds demonstrated significantly faster latency and lower amplitude of the P400 component for faces

than objects. In a study with 3- to 4-year-old children, Dawson et al. (2002) reported that the latency of the N290 was faster for faces than objects (not within-subject design). Collectively, these studies provide some information regarding category-specific neural responses in the first year of life, but due to limitations in design and analysis, they leave open the question of how and when category-selective neural responses emerge in development.

Typically, ERP studies present an image following a blank field, evoking activity that is a mixture of responses to the onset of local visual properties such as luminance and contrast as well as responses to higher-level organization, or structure, of the image. The challenge with this technique is to isolate the part of the response specific to the structural information that forms the object representation from other components of the response that are nonspecific. ERP analysis, therefore, conventionally relies on identifying peaks of interest, or components, within a large (for example, 200–500 ms) time-window of the transient response. While these components can provide information about the overall time-course of a process, the criteria for individual component selection can be problematic for developmental studies in which the polarity, number, latency, amplitude, and topography of peaks in the waveform can each vary with age as a result of physiological maturation processes such as increased myelination and synaptic density (DeBoer, Scott, & Nelson, 2004).

Here we introduce a novel steady-state visual evoked potential (ssVEP) paradigm as an objective measure of structure encoding of faces and objects in infants and adults. ssVEPs are cortical responses that are generated at frequencies that are exact integer multiples of the stimulus presentation frequency, and which can be recorded from the scalp (Regan, 1989). Stimulus-specific responses can be isolated using EEG spectrum analysis of the time series and thereby analyzed separately from other frequencies in the EEG recording. By recording ssVEP responses to alternating pairs of images, one scrambled and one intact, which are equated for low-level attributes (identical pixel content) but which differ in structural organization, we can isolate the response to the global structure from the response to the local elements of the images. Neural responses occurring at the intact image presentation frequency, or the first harmonic (1F), arise from mechanisms that are sensitive to changes in structure that occur at the onset or offset of only the intact images. Responses at the second harmonic (2F), the image alternation frequency, capture activity that is common after each image presentation such as sensitivity to local contrast changes that are not specific to the image structure. However, due to possible nonlinearities, the 2F may also contain some struc-

ture-related responses. Because the 2F potentially contains a mixture of structure-specific and non-specific responses, it is not directly interpretable with respect to global structure processing, and consequently, the 1F is considered the signature of structure-specific processing. ssVEPs have previously been used with infants and adults to investigate low- and mid-level visual processing, including contrast, acuity, orientation, texture, and motion sensitivity (Ales & Norcia, 2009; Baker, Norcia, & Candy, 2011; Braddick, Birtles, Wattam-Bell, & Atkinson, 2005; Hamer & Norcia, 1994; Heinrich & Bach, 2003; Norcia, Tyler, & Hamer, 1990; Pei, Pettet, & Norcia, 2007; Shirai et al., 2009; Skoczenski & Norcia, 1999), and more recently with adults to investigate high-level visual processing (Kaspar, Hassler, Martens, Trujillo-Barreto, & Gruber, 2010; Keil et al., 2003; Rossion & Boremanse, 2011). The current study is the first to extend the ssVEP method to examine infants' neural responses to high-level natural visual stimuli such as faces and objects.

## Methods

### Participants

Seventeen healthy, full-term infants between the age of 4 and 6 months (eight males) completed both object and face stimulus conditions (mean age = 5 months, 6 days  $\pm$  10 days). We chose to conduct a within-subjects design with the infants in order to minimize between-category error variance and to be able to directly compare face versus object responses. An additional eight infants were tested, but were excluded because they did not provide data for both conditions ( $N = 3$ ), or because of excess recording artifacts ( $N = 5$ ). Adults were tested using a between-subjects design; 16 adults were tested using object images and 10 adults were tested using face images. Informed consent was obtained from the adult participants or from the parent/guardian of the infant participant under a protocol that was approved by the Institutional Review Board of the California Pacific Medical Center. Infants were recruited from the San Francisco area through letters sent to parents and adults were recruited through local advertisement.

### Stimuli

Visual stimuli were presented through an in-house software package run by a Power Macintosh G4 computer (Mitsubishi Electric, Tokyo, Japan) on a contrast linearized CRT monitor (Apple Inc., Cupertino, CA) with a resolution of  $800 \times 600$  and a vertical



refresh rate of 72 Hz. Stimuli consisted of 15 grayscale photographs of objects and 15 grayscale photographs of female faces, and their corresponding scrambled images (Figure 1). Object images were a subset of those originally used by Kourtzi and Kanwisher (2000). Face images were chosen from the Karolinska Directed Emotional Faces set (KDEF, Lundqvist, Flykt, & Öhman, 1998), and were frontal views of females exhibiting a happy expression, with external features cropped (KDEF identities AF01, AF02, AF05, AF06, AF07, AF08, AF09, AF11, AF13, AF14, AF17, AF19, AF20, AF22, AF25). Faces and objects subtended approximately  $14^\circ \times 14^\circ$  of visual angle and were positioned in the center of the screen.

Scrambled versions of images were created by dividing the intact images into a  $20 \times 20$  pixel grid ( $0.8^\circ$ ) and randomly rearranging the positions of each of the resulting squares (black gridlines present in both

intact and scrambled images). This grid-scrambling method controls for contour fragments, edge contrast and luminance distribution between scrambled and intact images, and has been used in a number of previous studies of object processing (Fang, Murray, Kersten, & He, 2005; Kim, Biederman, Lescroart, & Hayworth, 2009; Kourtzi, Tolias, Altmann, Augath, & Logothetis, 2003; Lerner, Hendler, & Malach, 2002). Because these object images are routinely used for functionally defining LOC in both human and nonhuman primates using fMRI (Appelbaum, Wade, Vildavski, Pettet, & Norcia, 2006; Denys et al., 2004; Kourtzi & Kanwisher, 2000), we wanted to extend their use to the study of functional brain development in infants by relating our data to this literature. Grid scrambling was also used for the face images to make them comparable on the parameters described above.

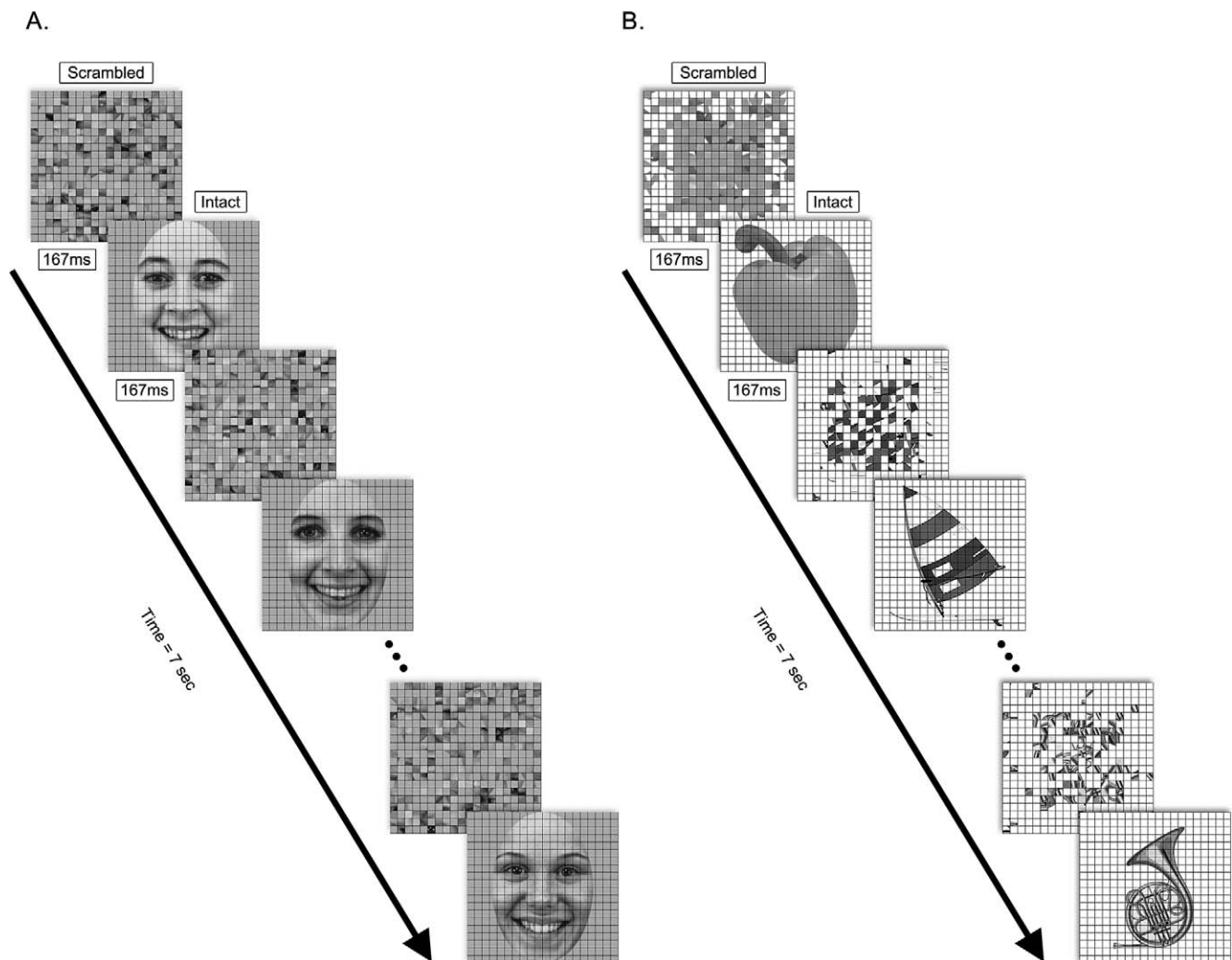


Figure 1. Examples of stimulus frames used in the (a) face (KDEF identities AF14, AF13, AF19) and (b) object conditions of the experiment. Trials consisted of alternating scrambled and intact images at 3 Hz for 7 seconds. Decomposition of response waveforms into spectral harmonics isolated responses to the global structure of the intact image (1F; 3 Hz) from responses at the image update rate (2F; 6 Hz). Statistical analyses were conducted on the scalar amplitude of these harmonics.

The grid scrambling procedure, while controlling for some features across intact and scrambled images, does not result in identical Fourier amplitude spectra (see [Supplementary Material](#)). Results from Fast Fourier Transform analysis confirmed that the grid-scrambling process altered the spatial frequency spectrum of the images. Specifically, at very low frequencies, intact images had higher amplitudes than scrambled images, and this was more pronounced for the face category. Our analysis also showed that object images had higher root-mean-squared (RMS) contrast (Mean = 0.784,  $SD = 0.08$ ) than faces (Mean = 0.456,  $SD = 0.02$ ) because the objects were against a white background rather than a gray background. All stimuli were well above infants' contrast and spatial frequency thresholds.

### Steady-state recording procedure

ssVEPs were recorded using a whole-head 128-channel Geodesic Sensor Net (Electrical Geodesics Inc., Eugene, OR), bandpass filtered from 0.1 Hz to 50.0 Hz, and digitized at a rate of at 432 Hz (Net Amps 200 TM, Electrical Geodesics, Inc., Eugene, OR). Individual electrodes were adjusted until impedances were below 60 k $\Omega$  before starting the recording.

Within each image category, trials consisted of a scrambled and corresponding intact image alternating every 166.67 ms, for 7 seconds ([Figure 1](#)). The frequency of intact image presentation was therefore 3 Hz, referred to as the fundamental frequency. Infants passively viewed the images while seated on their parent or caregiver's lap approximately 60 cm from the computer monitor. To control infant fixation and accommodation, a small toy on a string was suspended at the center of the screen. An experimenter observed the reflection of the monitor in the infant's pupil and trials were paused (by button press) if the infant was judged to be looking away or fussy. Data were not included 1 second before interruption and recording resumed 1 second after the experimenter indicated that the infant had regained fixation. Infants provided four to nine trials per category. Infant testing was done in two visits, counterbalanced for image category. For adults, 10 trials were presented in two blocks of five trials, in random order. Adults were instructed to fixate on a cross at the center of the monitor and to refrain from blinking during trials.

### Data processing and analysis

Artifact rejection was performed offline. Epochs were extracted from the continuously recorded EEG relative to the start of the trial, and were digitally filtered with a 0.8 to 20 Hz bandpass filter. Artifact

rejection was done according to a sample-by-sample thresholding procedure to remove noisy electrodes and replace them with the average of the six nearest neighboring electrodes. The EEG was then re-referenced to the common average of all the remaining electrodes. Epochs with more than 20% of the data samples exceeding 300  $\mu V$  were discarded on a sensor-by-sensor basis. Participants with fewer than three artifact-free trials in each condition were excluded from analyses.

Individual participants' time averages for each stimulus condition were computed over an approximately 2-second epoch that contained an exact integer number of cycles of the first harmonic. The time averages were converted to amplitude spectra at a frequency resolution of 0.5 Hz via a Discrete Fourier Transform. Incoherent averages of the amplitude at the first (1F; 3 Hz) and second (2F; 6 Hz) harmonic of the stimulus frequency were then evaluated (in both groups higher harmonics showed signals that were not reliably different from adjacent nonharmonic frequencies).

Based on recent infant ERP studies of face and object processing (McCleery et al., 2009; Scott & Monesson, 2009) and the scalp topographies of averaged data, three groups of electrodes were included in the analyses. Electrodes over the medial occipital region: 70, 75, 83, 74, 82 (corresponding to O1 [electrode 70], Oz [electrode 75], O2 [electrode 83] in the International 10-20 system), and the left: 58, 59, 64, 65, 69 (corresponding to P9 [electrode 58], P7 [electrode 59], PO7 [electrode 65]) and right: 91, 96, 90, 95, 89 (corresponding to P8 [electrode 91], P10 [electrode 96], and P08 [electrode 90] in the International 10-20 system) lateral occipital regions were grouped for statistical analysis for both infant and adult participants ([Figure 2](#)).

## Results

### Adult responses to image structure

Cortical responses to differences in the global organization between scrambled and intact images recorded at the first harmonic of the stimulus frequency result from neural mechanisms that are involved in structural encoding of intact images. Responses that are identical after transitions between scrambled and intact images and intact and scrambled images, such as those from local contrast mechanisms, are reflected in the second harmonic of the response. Our primary question concerns the presence of a differential structure-specific response to faces versus objects, operationalized as a significant response at the first

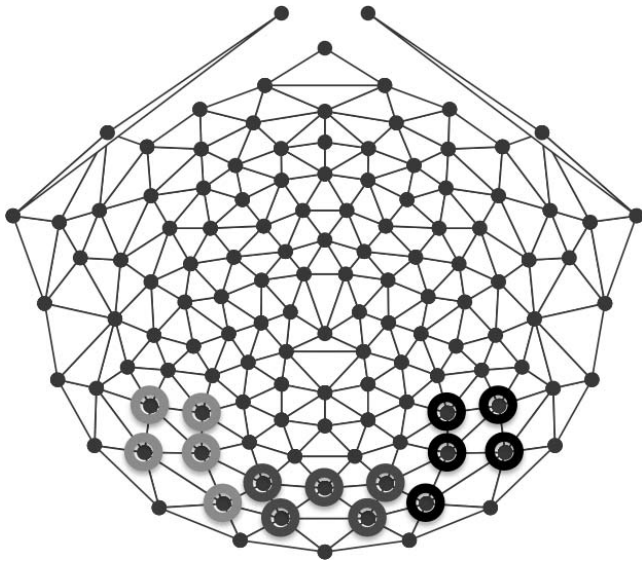


Figure 2. Electrode groups used for analysis. Responses were calculated over left (light gray), middle (gray), and right (black) occipital scalp regions.

harmonic, and the scalp topography corresponding to this selectivity.

Scalp distributions of 1F and 2F responses in adults illustrated both quantitative and qualitative differences between faces and objects (Figure 3). The 1F response for faces was distributed bilaterally over occipitotemporal electrodes, extending further lateral than the 1F response to objects. Both categories showed equivalent 1F responses over frontal regions. The topography of the 2F response for faces was also distributed bilaterally over occipitotemporal electrodes, but peaked on electrode 70 (corresponding to O1), while the 2F response for objects was distributed purely over medial-occipital electrodes and was centered more anteriorly than the 2F face response. Differences in the topography of first and second harmonic response components reflect differences in the underlying sources of these components. To highlight the scalp distribution that was unique to global structure processing of each image category, we subtracted the 2F from the 1F responses. This difference calculation revealed a clear response over occipitotemporal sites in the right hemisphere for faces and a bilateral occipitotemporal response for objects (Figure 3b).

We tested for significant 1F responses by comparing amplitudes at 3 Hz with noise levels measured from the mean amplitude at 2.5 and 3.5 Hz. Both faces and objects elicited 1F responses significantly above the noise levels, and this was true for all groups of electrodes (two-tailed, paired samples *t*-tests, all  $ps < 0.001$ ). Significant 1F responses confirmed that the adult brain is sensitive to the global structure present in intact face and object images. Faces and objects also evoked significant 2F responses relative to noise levels

for each group of electrodes ( $ps < 0.003$ ). The presence of a first harmonic component reflects the encoding of the structural differences between intact and scrambled images.

Regarding category specificity of adult responses, we found that faces elicited stronger responses than objects both at the 1F and at the 2F (Figure 4). The fact that the second harmonic for faces is larger than that for objects may reflect an overall larger response to this image category. It may also denote a combination of structure-related and local feature responses that were present in the 2F for faces. These results were quantified by entering amplitude as the dependent measure in to a repeated measures, mixed-model analysis of variance (ANOVA) with two within-subject variables (electrode group: left OT, medial occipital, right OT and harmonic: 1F, 2F) and one between-subject variable (image category: faces, objects). The analysis produced significant main effects of harmonic,  $F(1, 24) = 5.13$ ,  $p = 0.033$ ,  $\eta_p^2 = 0.176$ , and image category,  $F(1, 24) = 27.68$ ,  $p = 0.0001$ ,  $\eta_p^2 = 0.536$ . The main effect of harmonic was generated by stronger responses to global structure as measured by the first harmonic (1F Mean = 1.28 microvolts,  $SEM = 0.11$ ) than responses to local image elements as measured by the second harmonic (2F Mean = 1.02 microvolts,  $SEM = 0.108$ ). The main effect of image category reflects over twofold stronger responses to faces (Mean = 1.64 microvolts,  $SEM = 0.148$ ) than objects (Mean = 0.653 microvolts,  $SEM = 0.117$ ).

## Infant responses to image structure

Topographic maps of infant object and face responses are shown in Figure 5. Infant 1F responses to face structure were distributed over medial and lateral visual areas, much like those of adults. Infant 1F responses to objects, however, were restricted to the midline while adults produced topographies with bilateral occipitotemporal activity. Infant 2F responses showed nearly identical scalp distributions as their within-image category 1F responses, with faces eliciting a significant 2F response over medial and lateral regions and objects doing so only over medial occipital electrodes. While only structure encoding processes that can recognize the difference between scrambled and intact images will yield a signal at 1F, these neural processes may also contribute to the 2F response. As in adults, we subtracted the 2F from the 1F responses to examine the topography of neural responses produced only when recognizing differences in global structure. As shown in Figure 5b, this difference calculation in infants generated a bilateral response over occipitotemporal response for faces and strictly over the midline for objects.



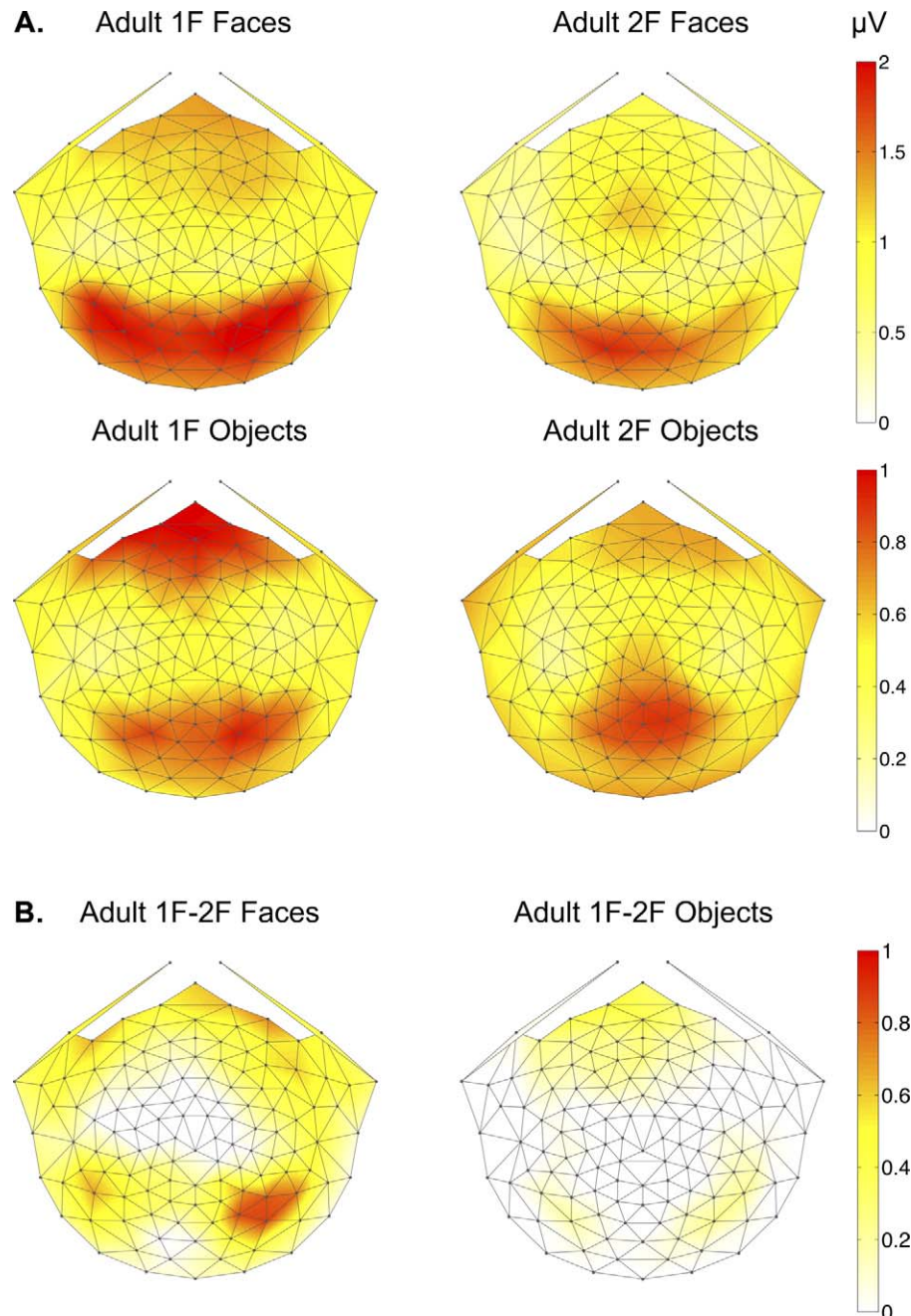


Figure 3. 2D scalp topographic maps of adult (a) 1F and 2F responses to faces and objects (b) 1F-2F difference responses to faces and objects.

Two-tailed, paired samples *t*-tests, revealed that infants produced significant 1F responses relative to the neighboring frequency noise level for faces at each electrode group (left OT:  $p = 0.038$ , medial occipital:  $p = 0.001$ , right OT:  $p = 0.002$ ). For objects, however, infant 1F responses were significantly different from the noise only over medial occipital electrodes ( $p = 0.002$ ). These results indicate that infant structure-specific responses were substantially more similar to those of adults for faces than for objects.

A repeated measures ANOVA with electrode group, harmonic, and image category as within-subject factors yielded significant main effects of electrode group,  $F(2, 32) = 14.81$ ,  $p = 0.0001$ ,  $\eta_p^2 = 0.481$ , and harmonic,  $F(1, 16) = 45.90$ ,  $p = 0.0001$ ,  $\eta_p^2 = 0.742$ . This was qualified by a nearly significant interaction effect between electrode group and harmonic,  $F(2, 32) = 3.85$ ,  $p = 0.067$ ,  $\eta_p^2 = 0.194$ , reflecting infants' stronger responses to image structure (1F) over medial occipital electrodes (Mean = 3.75;  $SEM = 0.408$ ) compared with left (Mean = 2.42;  $SEM = 0.312$ ) or right OT electrodes (Mean =

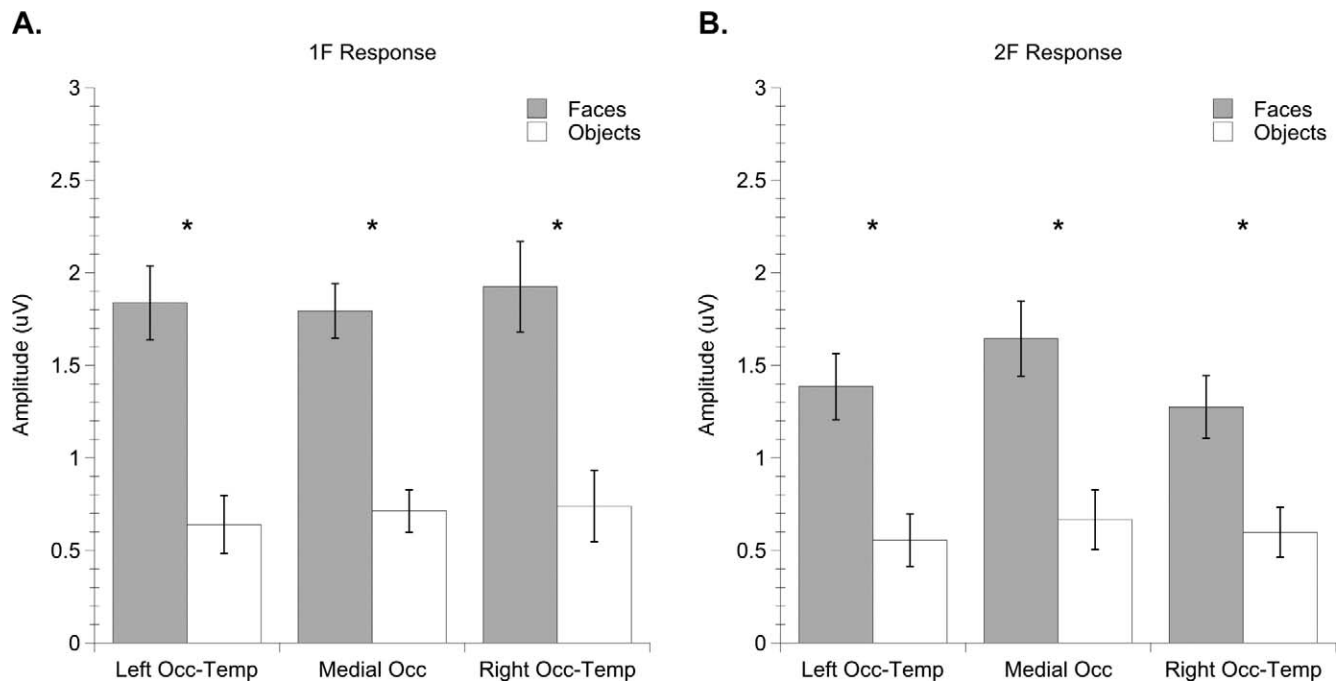


Figure 4. Mean amplitude of adult (a) 1F and (b) 2F responses to faces and objects. Error bars indicate standard errors of the mean. Asterisks indicate significant difference between conditions.

2.53;  $SEM = 0.214$ ). There was a trend for an interaction effect between electrode group and category,  $F(2, 32) = 4.08$ ,  $p = 0.061$ ,  $\eta_p^2 = 0.203$ , and a trend for a three-way interaction between electrode group, harmonic, and category,  $F(2, 32) = 3.39$ ,  $p = 0.084$ ,  $\eta_p^2 = 0.175$ . Notably, the direction of this trend was toward greater responses for faces (Mean = 2.93;  $SEM = 0.38$ ) than objects (Mean = 2.13;  $SEM = 0.17$ ) at right occipitotemporal electrodes. Post-hoc paired samples  $t$ -tests confirmed that 1F responses over right occipitotemporal electrodes were nearly reliably distinguishable between faces and objects,  $t(16) = 1.97$ ,  $p = 0.066$ . Figure 6 plots infant 1F and 2F amplitudes to faces and objects by electrode region.

Together, these results indicate that infants, similar to adults, were sensitive to the differences in global structural between scrambled and intact images, and that infants showed signs of selective processing of face, relative to object, image structure based on larger and more lateral responses to faces than objects.

## Discussion

The current report presents the first application of the ssVEP technique to the study of high-level visual processing in infants. The technique was used to investigate the topographic organization of responses to global structure in face and object images. We also determined whether a differential response to image

structure was present for faces versus objects, which would provide evidence for the emergence of category-specific object representations in infancy.

The electrophysiological method used here yields a sensitive and objective neural measurement that separates the response to image structure from the low-level feature response, without the challenges associated with component criteria and selection that are present in conventional transient ERP studies. This feature of the ssVEP technique facilitates between-group comparison of responses that reflect functionally equivalent processes. In addition to the unambiguity in the quantitative analysis of the structure-specific response measured at the fundamental presentation frequency, the ssVEP method also has the advantage of yielding high signal-to-noise ratios of the responses, making it possible to record a greater number of stimulus conditions in a given period of time, which is valuable for research with infants and allowed us to conduct a within-subject design in the infants.

Infants and adults produced evoked responses that were specific to differences in global structure between scrambled and intact image transitions, for both faces and objects. The scalp topography of the 1F responses in infants and adults also differed by image category. For faces, infant and adult responses were distributed bilaterally over occipitotemporal scalp areas, and the adult 1F-2F difference-response matched the ssVEP face response reported in Rossion and Boremanse (2011). For objects, infant 1F responses were restricted to the occipital midline in contrast to adults' more



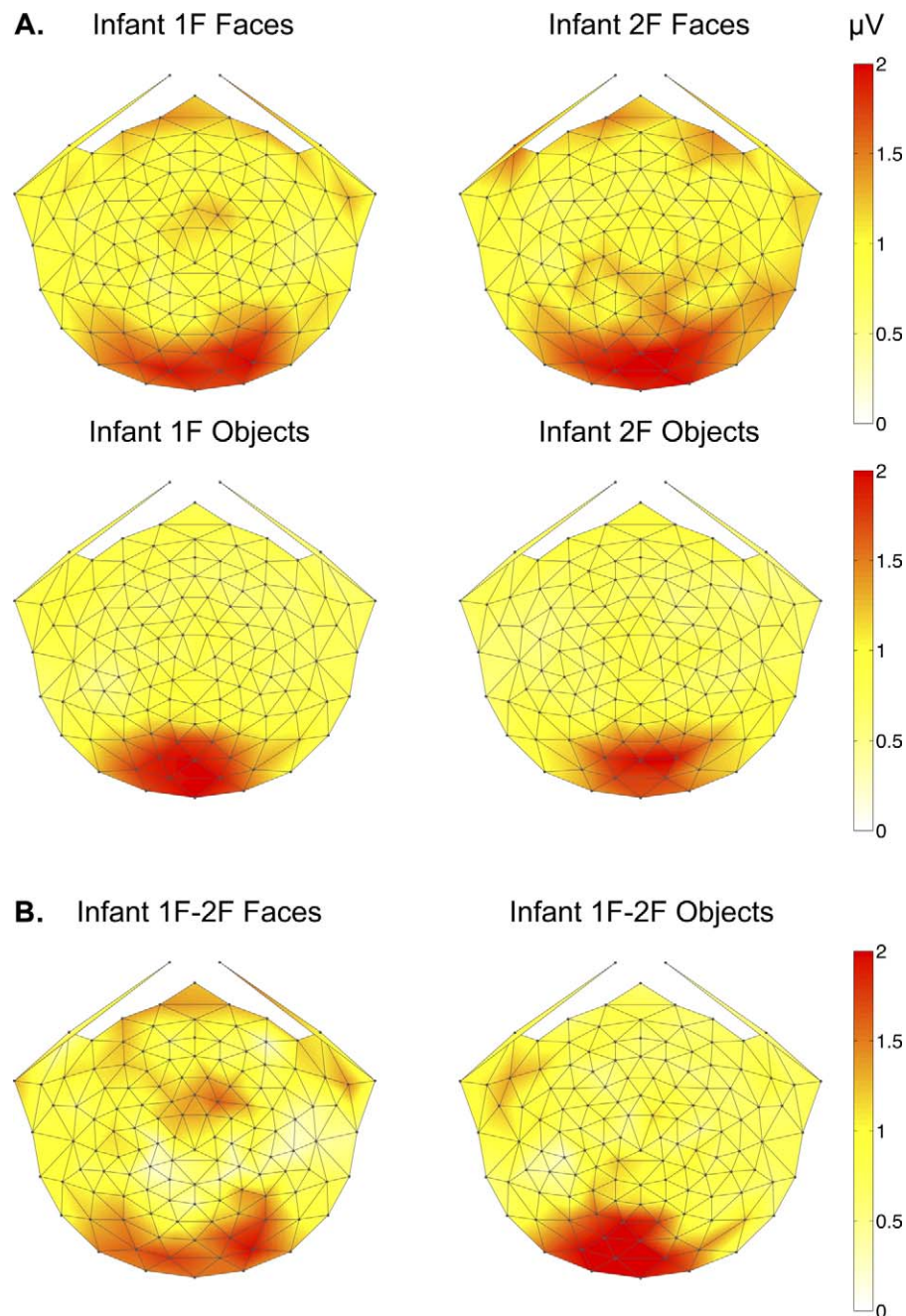


Figure 5. 2D scalp topographic maps of infant (a) 1F and 2F responses to faces and objects and (b) 1F-2F difference responses to faces and objects.

lateral distribution. The object images used here are commonly used in fMRI studies to define the location of cortex comprising the LOC; contrasting BOLD responses to intact versus grid-scrambled objects selectively activates LOC with little activation in low-level areas (Appelbaum et al., 2006; Denys et al., 2004; Kourtzi & Kanwisher, 2001). Consistent with fMRI-defined object-selective regions in the LOC, our results show that adult neural responses to differences in structure between scrambled and intact object images were topographically distributed over lateral occipito-

temporal regions of the scalp. Infant evoked responses did not extend over lateral occipitotemporal regions, suggesting that portions of cortex considered to be object-selective in adults are immature at 6 months of age. On the other hand, the cortical responses underlying structure encoding of faces appeared similar between infants and adults. While these findings do not necessarily constitute evidence of structure processing at a level required for *perceptual* image categorization, they indicate the presence of a mid- to high-level neural substrate for structure-specific responses to faces and

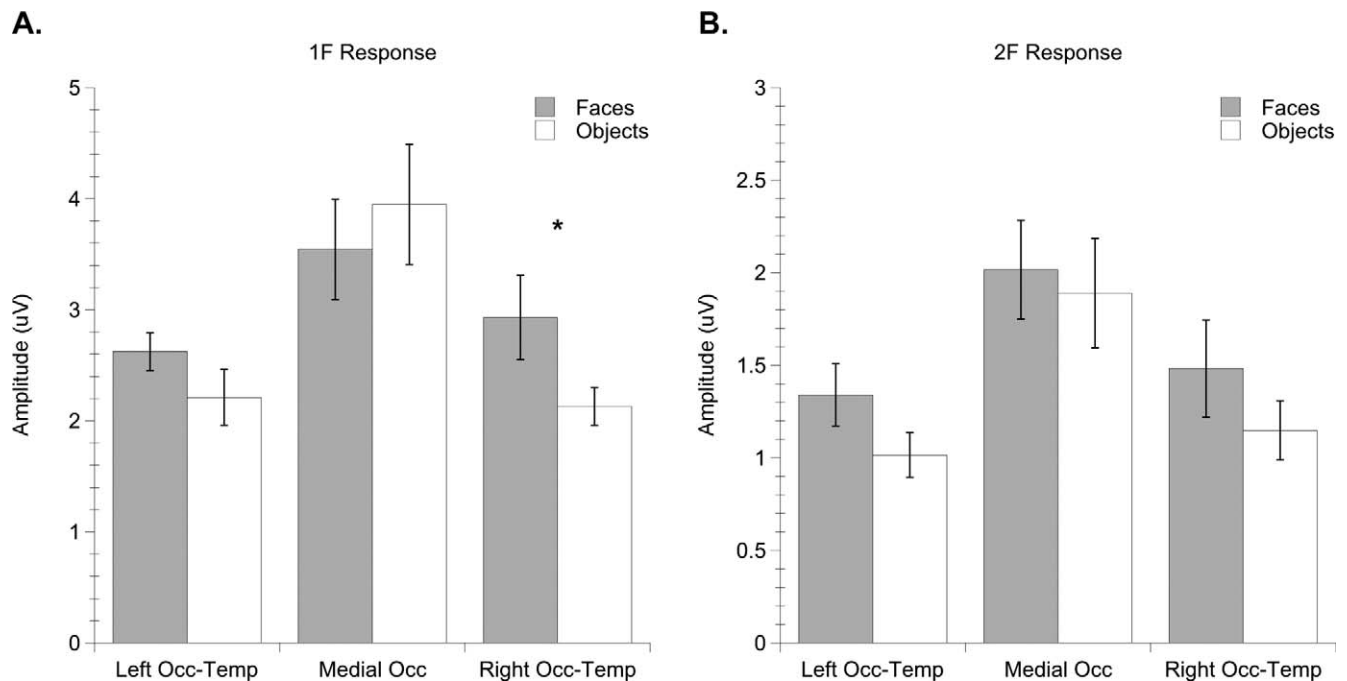


Figure 6. Mean amplitude of infant (a) 1F and (b) 2F responses to faces and objects. Error bars indicate standard errors of the mean. Asterisk indicates significant difference between conditions ( $p < 0.06$ ).

objects. Further, although infants showed high-level structure-specific responses to faces, the present data, alone, do not resolve whether visual areas responsible for face processing are necessarily fully adult-like by 6 months. A suggestion that brain areas involved in face processing might still be undergoing structural and functional changes is the more restricted topography of the infant 1F response compared with that of adults. Also, infants did not exhibit a twofold stronger response to faces than objects, as was found in adults. A more direct source analysis with an accurate conductivity model is needed to determine if the underlying source distribution matches the scalp distribution.

What factors could underlie the topographically restricted processing of object versus face image structure that we observed in infants? The structure of an object or a face can be extracted using a variety of different visual cues processed at different stages of the visual processing hierarchy. Therefore, the ability to encode natural image structure could be limited at one of many stages of hierarchy in the infant visual system. At the earliest stage of visual processing, infants are limited by reduced contrast sensitivity and visual acuity (Banks & Salapatek, 1983). In our experiment it is unlikely that basic visibility of the local visual information differentially implicated structure processing of face versus object images. First, infants produced significant responses to faces and objects at the 2F frequency, indicating that they were equally responsive

to the local elements present in both scrambled and intact face and object images.

Second, the contrast level of the images was well above threshold levels for infants in this age range. The contrast of the face images was lower than the objects, but this should have resulted in reduced rather than enhanced evoked responses for faces relative to objects, contrary to what we observed. Also, we expect this contrast difference had minimal effect on the overall findings because the RMS contrast was identical between the scrambled and intact image pairs and first and second harmonic responses were calculated within image category.

Early visual areas may also have responded to differences in the spatial frequency spectrum between scrambled and intact images, which were different for faces and objects. While the grid-scrambling method used here preserves the contour fragments, edge contrasts, and luminance distributions of the intact images, by rearranging sections of the image, scrambling alters the spatial frequency content. Intact images had more energy at low spatial frequencies than their scrambled counterparts, and this difference was more substantial for faces (see [Supplementary Material](#)). Therefore, it is plausible that changes in spatial frequency could have contributed to the first harmonic response and this contribution could have differed between face and object conditions. However, if sensitivity to spatial frequency rather than structure was driving the first harmonic responses, these responses likely would have been generated from a lower rather

than higher level of the visual hierarchy. The topographic distribution of the first harmonic response to faces, in both adults and infants, extended beyond medial occipital scalp regions, arguing against this explanation. Additional evidence against the interpretation that the 1F represents sensitivity to spatial frequency differences comes from data we have collected using a phase-scrambling method that maintains the power spectrum and mean luminance of an intact face and scrambled image pair (Ales, Farzin, Rossion, & Norcia, 2012). Phase-scrambling allows us to consider how much of the 1F response may have been generated by the spectral difference. If the 1F responses obtained in the present study were generated solely by the differences in spatial frequency, we would expect phase-scrambling to eliminate the 1F response because it eliminates the difference in spatial frequency content. Instead, we found that after controlling for spatial frequency, the 1F adult response distribution, at least for faces, is largely the same as that presented here. The phase-scrambled stimuli were separately used with infants and adults and resulted in response distributions similar to those presented here (unpublished data). These data indicate that even when the power spectra are maintained, the responses to faces are distributed laterally. Spatial frequency differences, if they do contribute to the first harmonic responses, are most likely to impact responses measured over the occipital pole.

Based on the qualitative differences between infant and adult evoked responses to object structure, together with the similarities in responses to face structure, we propose that infants are able to extract mid-level structure present in at least some categories of natural images. In support of this view, the scrambled versus intact images in both categories differed on a number of mid-level structure attributes that undergo analysis during the processing of complex global form. One such attribute includes the degree of organization of local features or elements that need to be integrated in order to correctly encode structure. Specifically, statistical pair-wise correlations in spatial frequency and orientation between neighboring spatial regions, which are prevalent in natural images but are disrupted by scrambling (Geisler, Perry, Super, & Gallogly, 2001). The statistics of these local correlations enable the visual system to group together local elements and ultimately allow us to interpret images accurately and efficiently.

Previous ssVEP studies have examined the emergence and development of mid-level structure processing using textures (Norcia et al., 2005; Palomares, Pettet, Vildavski, Hou, & Norcia, 2010; Pei, Pettet, & Norcia, 2007; Wattam-Bell et al., 2010). By manipulating the level of orientation correlations within a texture, these studies have demonstrated that infants show neural sensitivity to the organization of elements

in orientation-defined textures and contours. The restricted topography of responses to the object images in the present study resembled that previously reported for texture-defined form responses of 5-month-olds (Wattam-Bell et al., 2010). This similarity in response distribution suggests that the object images were being processed only up to a mid-level of analysis akin to that involved in processing textures, but not beyond. The more anteriorly extended topography of the infant face-structure response, on the other hand, suggests that a higher level of image organization could have been extracted for faces. This higher level of processing may be possible due to differences in the mid-level statistics of object and face images, or to the precocious development of a face “template” in higher-level cortex.

Another possible distinction between the face and object images presented here is within-category inter-stimulus variance. That is, the face stimuli were less variable in structure than the objects. All faces share the same basic parts and have a similar spatial configuration of features in terms of two eyes above a centered nose and above a mouth, while objects may have more variability in the configuration of their features. Given this, infants could have been more mature in their processing of global face structure because the statistical correlations across face exemplars were more consistent. To address this, exemplars from a single object category (i.e., houses or cars) could be used in a future study. Finally, the amount of experience with exemplars from our two categories is likely to be different, which could have led to the relative response difference.

From our data we conclude that 4- to 6-month-old infants could encode at least some of the structural cues that differed between intact versus scrambled images, and it appears that the infant visual system carries some specificity in mid- or high-level integrative stages that allows a distinction to be made between face and object images. While we have demonstrated that infant visual cortex shows mid- to high-level category-specificity, functional organization of object and face processing pathways is not yet fully mature. Future studies that link neural responses to perceptual behaviors are warranted.

## Acknowledgments

This research was supported by National Institutes of Health grant F32EY021389 (FF).

Commercial relationships: none.

Corresponding author: Faraz Farzin.

Email: [ffarzin@stanford.edu](mailto:ffarzin@stanford.edu).

Address: Department of Psychology, Stanford University, Stanford, CA, USA.



## References

- Aguirre, G. K., Zarahn, E., & D'Esposito, M. (1998). Neural components of topographical representation. *Proceedings of the National Academy of Sciences of the United States of America*, *95*(3), 839–846.
- Ales, J. M., & Norcia, A. M. (2009). Assessing direction-specific adaptation using the steady-state visual evoked potential: Results from EEG source imaging. *Journal of Vision*, *9*(7):8, 1–13, <http://www.journalofvision.org/content/9/7/8>, doi:10.1167/9.7.8. [PubMed] [Article]
- Ales, J. M., Farzin, F., Rossion, B., & Norcia, A. M. (2012). An objective method for measuring face detection thresholds using the sweep steady-state visual evoked response. *Journal of Vision*, *12*(10):18, 1–18, <http://www.journalofvision.org/content/12/10/18>, doi:10.1167/12.10.18. [PubMed] [Article]
- Appelbaum, L. G., Wade, A. R., Vildavski, V. Y., Pettet, M. W., & Norcia, A. M. (2006). Cue-invariant networks for figure and background processing in human visual cortex. *Journal of Neuroscience*, *26*(45), 11 695–11 708.
- Baker, T. J., Norcia, A. M., & Candy, T. R. (2011). Orientation tuning in the visual cortex of 3-month-old human infants. *Vision Research*, *51*(5):470–478.
- Banks, M. S., & Salapatek, P. (1983). Infant visual perception. In P. H. Mussen (Ed.), *Handbook of child development* (pp. 435–571). New York: Wiley.
- Beauchamp, M. S., Lee, K. E., Haxby, J. V., & Martin, A. (2002). Parallel visual motion processing streams for manipulable objects and human movements. *Neuron*, *34*(1), 149–159.
- Bentin, S., Allison, T., Puce, A., Perez, E., & McCarthy, G. (1996). Electrophysiological studies of face perception in humans. *Journal of Cognitive Neuroscience*, *8*(6), 551–565.
- Botzel, K., Schulze, S., & Stodieck, S. R. (1995). Scalp topography and analysis of intracranial sources of face-evoked potentials. *Experimental Brain Research*, *104*(1), 135–143.
- Braddick, O., Birtles, D., Wattam-Bell, J., & Atkinson, J. (2005). Motion- and orientation-specific cortical responses in infancy. *Vision Research*, *45*(25–26), 3169–3179.
- Carmel, D., & Bentin, S. (2002). Domain specificity versus expertise: Factors influencing distinct processing of faces. *Cognition*, *83*(1), 1–29.
- Dawson, G., Carver, L., Meltzoff, A. N., Panagiotides, H., McPartland, J., & Webb, S. J. (2002). Neural correlates of face and object recognition in young children with autism spectrum disorder, developmental delay, and typical development. *Child Development*, *73*(3), 700–717.
- de Haan, M., Johnson, M. H., & Halit, H. (2003). Development of face-sensitive event-related potentials during infancy: A review. *International Journal of Psychophysiology*, *51*(1), 45–58.
- de Haan, M., & Nelson, C. A. (1999). Brain activity differentiates face and object processing in 6-month-old infants. *Developmental Psychology*, *35*(4), 1113–1121.
- de Haan, M., Pascalis, O., & Johnson, M. H. (2002). Specialization of neural mechanisms underlying face recognition in human infants. *Journal of Cognitive Neuroscience*, *14*(2), 199–209.
- DeBoer, T., Scott, L., & Nelson, C. A. (2004). Event-related potentials in developmental populations. In T. Handy (Ed.), *Event-related potentials: A methods handbook*. (pp. 263–297). Cambridge, MA: MIT Press.
- Denys, K., Vanduffel, W., Fize, D., Neliseen, K., Peuskens, H., Van Essen, D., et al. (2004). The processing of visual shape in the cerebral cortex of human and nonhuman primates: A functional magnetic resonance imaging study. *Journal of Neuroscience*, *24*(10), 2551–2565.
- Downing, P. E., Bray, D., Rogers, J., & Childs, C. (2004). Bodies capture attention when nothing is expected. *Cognition*, *93*(1), B27–38.
- Eimer, M. (2000a). Effects of face inversion on the structural encoding and recognition of faces. Evidence from event-related brain potentials. *Brain Research Cognitive Brain Research*, *10*(1–2), 145–158.
- Eimer, M. (2000b). Event-related brain potentials distinguish processing stages involved in face perception and recognition. *Clinical Neurophysiology*, *111*(4), 694–705.
- Epstein, R., & Kanwisher, N. (1998). A cortical representation of the local visual environment. *Nature*, *392*(6676), 598–601.
- Fang, F., Murray, S. O., Kersten, D., & He, S. (2005). Orientation-tuned fMRI adaptation in human visual cortex. *Journal of Neurophysiology*, *94*(6), 4188–4195.
- Geisler, W. S., Perry, J. S., Super, B. J., & Gallogly, D. P. (2001). Edge co-occurrence in natural images predicts contour grouping performance. *Vision Research*, *41*(6), 711–724.
- George, N., Evans, J., Fiori, N., Davidoff, J., & Renault, B. (1996). Brain events related to normal

- and moderately scrambled faces. *Brain Research Cognitive Brain Research*, 4(2), 65–76.
- Grill-Spector, K., Kushnir, T., Hendler, T., Edelman, S., Itzchak, Y., & Malach, R. (1998). A sequence of object-processing stages revealed by fMRI in the human occipital lobe. *Human Brain Mapping*, 6(4), 316–328.
- Halit, H., de Haan, M., & Johnson, M. H. (2003). Cortical specialisation for face processing: Face-sensitive event-related potential components in 3- and 12-month-old infants. *Neuroimage*, 19(3), 1180–1193.
- Hamer, R. D., & Norcia, A. M. (1994). The development of motion sensitivity during the first year of life. *Vision Research*, 34(18), 2387–2402.
- Haxby, J. V., Hoffman, E. A., & Gobbini, M. I. (2000). The distributed human neural system for face perception. *Trends in Cognitive Science*, 4(6), 223–233.
- Heinrich, S. P., & Bach, M. (2003). Adaptation characteristics of steady-state motion visual evoked potentials. *Clinical Neurophysiology*, 114, 1359–1366.
- Itier, R. J., & Taylor, M. J. (2004). Source analysis of the N170 to faces and objects. *Neuroreport*, 15(8), 1261–1265.
- Jacques, C., d'Arripe, O., & Rossion, B. (2007). The time course of the inversion effect during individual face discrimination. *Journal of Vision*, 7(8):3, 1–9, <http://www.journalofvision.org/content/7/8/3>, doi:10.1167/7.8.3. [PubMed] [Article]
- Jeffreys, D. A. (1989). A face-responsive potential recorded from the human scalp. *Experimental Brain Research*, 78(1), 193–202.
- Johnson, J. S., & Olshausen, B. A. (2003). Timecourse of neural signatures of object recognition. *Journal of Vision*, 3(7):4, 499–512, <http://www.journalofvision.org/content/3/7/4>, doi:10.1167/3.7.4. [PubMed] [Article]
- Kanwisher, N. (2010). Functional specificity in the human brain: a window into the functional architecture of the mind. *Proceedings of the National Academy of Sciences of the United States of America*, 107(25), 11 163–11 170.
- Kanwisher, N., McDermott, J., & Chun, M. M. (1997). The fusiform face area: A module in human extrastriate cortex specialized for face perception. *Journal of Neuroscience*, 17(11), 4302–4311.
- Kaspar, K., Hassler, U., Martens, U., Trujillo-Barreto, N., & Gruber, T. (2010). Steady-state visually evoked potential correlates of object recognition. *Brain Research*, 1343, 112–121.
- Keil, A., Gruber, T., Müller, M. M., Moratti, S., Stolarova, M., Bradley, M. M., et al. (2003). Early modulation of visual perception by emotional arousal: Evidence from steady-state visual evoked brain potentials. *Cognitive, Affective, & Behavioral Neuroscience*, 3, 195–206.
- Kim, J. G., Biederman, I., Lescroart, M. D., & Hayworth, K. J. (2009). Adaptation to objects in the lateral occipital complex (LOC): Shape or semantics? *Vision Research*, 49(18), 2297–2305.
- Kourtzi, Z., & Kanwisher, N. (2000). Cortical regions involved in perceiving object shape. *Journal of Neuroscience*, 20(9), 3310–3318.
- Kourtzi, Z., & Kanwisher, N. (2001). Representation of perceived object shape by the human lateral occipital complex. *Science*, 293(5534), 1506–1509.
- Kourtzi, Z., Tolias, A. S., Altmann, C. F., Augath, M., & Logothetis, N. K. (2003). Integration of local features into global shapes: Monkey and human fMRI studies. *Neuron*, 37(2), 333–346.
- Kuefner, D., de Heering, A., Jacques, C., Palmero-Soler, E., & Rossion, B. (2010). Early visually evoked electrophysiological responses over the human brain (P1, N170) show stable patterns of face-sensitivity from 4 years to adulthood. *Frontiers in Human Neuroscience*, 3, 67.
- Lerner, Y., Hendler, T., Ben-Bashat, D., Harel, M., & Malach, R. (2001). A hierarchical axis of object processing stages in the human visual cortex. *Cerebral Cortex*, 11(4), 287–297.
- Lerner, Y., Hendler, T., & Malach, R. (2002). Object-completion effects in the human lateral occipital complex. *Cerebral Cortex*, 12(2), 163–177.
- Lundqvist, D., Flykt, A., & Öhman, A. (1998). The Karolinska directed emotional faces set (KDEF) [CD-ROM from Department of Clinical Neuroscience, Psychology section, Karolinska Institutet]. ISBN 91-630-7164-9.
- Malach, R., Reppas, J. B., Benson, R. R., Kwong, K. K., Jiang, H., Kennedy, W. A., et al. (1995). Object-related activity revealed by functional magnetic resonance imaging in human occipital cortex. *Proceedings of the National Academy of Sciences of the United States of America*, 92(18), 8135–8139.
- Martin, A., Wiggs, C. L., Ungerleider, L. G., & Haxby, J. V. (1996). Neural correlates of category-specific knowledge. *Nature*, 379(6566), 649–652.
- McCarthy, G., Puce, A., Belger, A., & Allison, T. (1999). Electrophysiological studies of human face perception. II: Response properties of face-specific potentials generated in occipitotemporal cortex. *Cerebral Cortex*, 9(5), 431–444.

- McCleery, J. P., Akshoomoff, N., Dobkins, K. R., & Carver, L. J. (2009). Atypical face versus object processing and hemispheric asymmetries in 10-month-old infants at risk for autism. *Biological Psychiatry*, *66*(10), 950–957.
- Norcia, A. M., Pei, F., Bonneh, Y., Hou, C., Sampath, V., & Pettet, M. W. (2005). Development of sensitivity to texture and contour information in the human infant. *Journal of Cognitive Neuroscience*, *17*(4), 569–579.
- Norcia, A. M., Tyler, C. W., & Hamer, R. D. (1990). Development of contrast sensitivity in the human infant. *Vision Research*, *30*(10), 475–486.
- Palomares, M., Pettet, M., Vildavski, V., Hou, C., & Norcia, A. (2010). Connecting the dots: How local structure affects global integration in infants. *Journal of Cognitive Neuroscience*, *22*(7), 1557–1569.
- Pei, F., Pettet, M. W., & Norcia, A. M. (2007). Sensitivity and configuration-specificity of orientation-defined texture processing in infants and adults. *Vision Research*, *47*(3), 338–348.
- Pernet, C., Schyns, P. G., & Demonet, J. F. (2007). Specific, selective or preferential: Comments on category specificity in neuroimaging. *Neuroimage*, *35*(3), 991–997.
- Puce, A., Allison, T., Asgari, M., Gore, J. C., & McCarthy, G. (1996). Differential sensitivity of human visual cortex to faces, letterstrings, and textures: A functional magnetic resonance imaging study. *Journal of Neuroscience*, *16*(16), 5205–5215.
- Regan, D. (1989). *Human brain electrophysiology. Evoked potentials and evoked magnetic fields in science and medicine*. New York: Elsevier.
- Rossion, B., & Boremanse, A. (2011). Robust sensitivity to facial identity in the right human occipitotemporal cortex as revealed by steady-state visual-evoked potentials. *Journal of Vision*, *11*(2):16, 1–21. <http://www.journalofvision.org/content/11/2/16>, doi:10.1167/11.2.16. [PubMed] [Article]
- Rossion, B., Gauthier, I., Tarr, M. J., Despland, P., Bruyer, R., Linotte, S., et al. (2000). The N170 occipitotemporal component is delayed and enhanced to inverted faces but not to inverted objects: An electrophysiological account of face-specific processes in the human brain. *Neuroreport*, *11*(1), 69–74.
- Rossion, B., & Jacques, C. (2008). Does physical interstimulus variance account for early electrophysiological face sensitive responses in the human brain? Ten lessons on the N170. *Neuroimage*, *39*(4), 1959–1979.
- Rousselet, G. A., Gaspar, C. M., Wieczorek, K. P., & Pernet, C. R. (2011). Modeling single-trial ERP reveals modulation of bottom-up face visual processing by top-down task constraints (in some subjects). *Frontiers in Psychology*, *2*, 137.
- Rousselet, G. A., Husk, J. S., Bennett, P. J., & Sekuler, A. B. (2007). Single-trial EEG dynamics of object and face visual processing. *Neuroimage*, *36*(3), 843–862.
- Rousselet, G. A., Husk, J. S., Bennett, P. J., & Sekuler, A. B. (2008). Time course and robustness of ERP object and face differences. *Journal of Vision*, *8*(12):3, 1–18. <http://www.journalofvision.org/content/8/12/3>, doi:10.1167/8.12.3. [PubMed] [Article]
- Rousselet, G. A., Pernet, C. R., Caldara, R., & Schyns, P. G. (2011). Visual object categorization in the brain: What can we really learn from ERP peaks? *Frontiers in Human Neuroscience*, *5*, 156.
- Scott, L. S., & Monesson, A. (2009). The origin of biases in face perception. *Psychological Science*, *20*(6), 676–680.
- Scott, L. S., Pascalis, O., & Nelson, C. A. (2007). A domain-general theory of the development of perceptual discrimination. *Current Directions in Psychological Science*, *16*(4), 197–201.
- Shirai, N., Birtles, D., Wattam-Bell, J., Yamaguchi, M. K., Kanazawa, S., Atkinson, J., et al. (2009). Asymmetrical cortical processing of radial expansion/contraction in infants and adults. *Developmental Science*, *12*(6), 946–955.
- Skoczenski, A. M., & Norcia, A. M. (1999). Development of VEP Vernier acuity and grating acuity in human infants. *Investigative Ophthalmology & Visual Science*, *40*(10), 2411–2417, <http://www.iovs.org/content/40/10/2411>. [PubMed] [Article]
- Taylor, M. J., McCarthy, G., Saliba, E., & Degiovanni, E. (1999). ERP evidence of developmental changes in processing of faces. *Clinical Neurophysiology*, *110*(5), 910–915.
- VanRullen, R., & Thorpe, S. J. (2001). The time course of visual processing: From early perception to decision-making. *Journal of Cognitive Neuroscience*, *13*(4), 454–461.
- Wattam-Bell, J., Birtles, D., Nyström, von Hofsten, C., Rosander, K., Anker, S., et al. (2010). Reorganization of global form and motion processing during human visual development. *Current Biology*, *20*(5), 411–415.