

An objective method for measuring face detection thresholds using the sweep steady-state visual evoked response

Justin M. Ales*

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



Faraz Farzin*

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



Bruno Rossion

Institute of Psychology and Institute of Neuroscience,
University of Louvain, Belgium



Anthony M. Norcia

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



We introduce a sensitive method for measuring face detection thresholds rapidly, objectively, and independently of low-level visual cues. The method is based on the swept parameter steady-state visual evoked potential (ssVEP), in which a stimulus is presented at a specific temporal frequency while parametrically varying (“sweeping”) the detectability of the stimulus. Here, the visibility of a face image was increased by progressive derandomization of the phase spectra of the image in a series of equally spaced steps. Alternations between face and fully randomized images at a constant rate (3/s) elicit a robust first harmonic response at 3 Hz specific to the structure of the face. High-density EEG was recorded from 10 human adult participants, who were asked to respond with a button-press as soon as they detected a face. The majority of participants produced an evoked response at the first harmonic (3 Hz) that emerged abruptly between 30% and 35% phase-coherence of the face, which was most prominent on right occipito-temporal sites. Thresholds for face detection were estimated reliably in single participants from 15 trials, or on each of the 15 individual face trials. The ssVEP-derived thresholds correlated with the concurrently measured perceptual face detection thresholds. This first application of the sweep VEP approach to high-level vision provides a sensitive and objective method that could be used to measure and compare visual perception thresholds for various object shapes and levels of categorization in different human populations, including infants and individuals with developmental delay.

Keywords: face detection, steady-state visual evoked potential, N170, object recognition

Citation: 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.

Introduction

The healthy adult human brain can detect visual patterns such as a face in a complex visual scene in a fraction of a second (e.g., Crouzet, Kirchner, & Thorpe, 2010; Fei-Fei, Iyer, Koch, & Perona, 2007; Fletcher-Watson et al., 2008; Lewis & Edmonds, 2003; Rousselet, Mace, & Fabre-Thorpe, 2003). Sensitivity to face patterns is even found at birth (Goren, Sarty, & Wu., 1975; Johnson, Dziurawiec, Ellis, & Morton, 1991), suggesting that newborns have an innate representation of a face template (although see Turati, Simion, Milani, & Umiltà, 2002).

In order to understand the mechanisms underlying face detection, or the categorization of a visual stimulus as a face, behavioral studies have investigated this process using various tasks and stimuli: detection of faces in complex visual scenes using manual responses (e.g., Lewis & Edmonds, 2003; Rousselet et al., 2003) or saccades (Cerf, Harel, Einhäuser, & Koch, 2008; Crouzet et al., 2010; Fletcher-Watson et al., 2008), categorization of normal faces versus faces presented under a variety of transformations such as inversion, feature masking, or jumbling (Cooper & Wojan, 2000; Lewis & Edmonds, 2003; Purcell & Stewart, 1986, 1988; Valentine & Bruce, 1986), visual-search paradigms with schematic faces or face photographs (Brown, Huey, &

Findlay, 1997; Garrido, Duchaine, & Nakayama, 2008; Hershler & Hochstein, 2005; Hershler, Golan, Bentin, & Hochstein, 2010; Lewis & Edmonds, 2003; Nothdurft, 1993; Van Rullen, 2006), detection of faces briefly presented with backward masking (Purcell & Stewart, 1986, 1988), or categorization of stimuli as faces based on their global configuration rather than on their local parts (e.g., two-tones Mooney figures or Arcimboldo's face-like paintings; McKeef & Tong, 2007; Mooney, 1957; Moore & Cavanah, 1998; Parkin & Williamson, 1987; Rossion, Dricot, Goebel, & Busigny, 2011).

The perception of a visual stimulus as a face has been associated with an increase in neural activation, relative to other object shapes and scrambled faces, in a set of high-level visual areas of the ventral processing stream, most prominently in the inferior occipital gyrus and middle fusiform gyrus, but also in the superior temporal sulcus and inferior temporal cortex (e.g., Haxby, Hoffman, & Gobbini, 2000; Kanwisher, McDermott, & Chun, 1997; Puce, Allison, Gore, & McCarthy, 1995; Sergent et al., 1992; Tsao, Moeller, & Freiwald, 2008; Weiner & Grill-Spector, 2010). Face perception has also been associated with an increase (relative to other visual stimuli) of the visual event-related potential (ERP) recorded on the occipito-temporal scalp at about 170 ms, the N170 (Bentin, Allison, Puce, Perez, & McCarthy, 1996; for early studies of face-sensitive ERPs, see Jeffreys [1989]; for reviews on the N170, see Rossion & Jacques [2008, 2011]; and for the analogous component recorded in MEG, M170, see e.g., Halgren, Raji, Marinkovic, Jousmäki, & Hari [2000]). Intracranial studies in epileptic patients have also reported large negative components at approximately the same latency on the ventral surface of the occipito-temporal cortex associated with the perception of a face (e.g., Allison, McCarthy, Nobre, Puce, & Belger, 1994; Barbeau et al., 2008).

Although these approaches have provided information regarding the stimulus characteristics, time-course, and neural basis underlying face processing in the healthy adult brain, they also have limitations that leave open the question of how a face is first detected. Behavioral detection thresholds reflect a complex chain of sensory and decision processes, and performance can be impacted by a number of extraneous factors, particularly in infants and children and in populations with cognitive impairments.

Traditional ERP measures based on the N170 face-sensitive response component typically involve the comparison between suitable face and control images (e.g., Rossion & Caharel, 2011; Rousselet, Husk, Bennett, & Sekuler, 2008a). However, subtraction of waveforms to isolate a face-specific response can be difficult to interpret due to differences in time (latency)

and space (topography) of the N170 elicited by a face versus a control image, as well as differences that are present in preceding ERP response components. The structure of face-selective components defined in this way can vary considerably across different populations and precise definition of the onset time, peak time, and amplitude can sometimes be challenging (see Kuefner, de Heering, Jacques, Palmero-Soler, & Rossion, 2010). Moreover, the low signal-to-noise ratio of the transient ERP method requires the recording and averaging of a substantial number of trials in order to obtain reliable transient ERP responses that differ between faces and control stimuli in a group of participants, let alone in a single observer. This limitation is particularly problematic when recording face perception responses from infants, children, or clinical populations (Kuefner et al., 2010).

What would be desirable is an objective method that not only tightly controls for the contribution of responses to extraneous low-level visual cues, but also provides adequate signal-to-noise ratio for defining face-sensitive response components in a small number of trials. Here, we used the steady-state visual evoked potential (ssVEP) method (Regan, 1966), in particular the sweep ssVEP (Regan, 1973), which has previously been used to isolate specific responses to simple visual stimuli. This method has provided a rapid and objective assessment of low-level visual function such as visual acuity and contrast sensitivity in infants and adults (e.g., Norcia & Tyler, 1985; Norcia, Tyler, & Hamer, 1990; Regan, 1977; Tyler, Apkarian, Levi, & Nakayama, 1979; for a recent review see Almoqbel, Leat, & Irving [2008]). To adapt the sweep ssVEP approach to the study of high-level vision, and face perception in particular, we used a phase-scrambling parameter to systematically vary face visibility. A comparison between responses evoked by phase-scrambled and intact images has been used in several recent ERP studies to isolate face-sensitive responses (e.g., Jacques & Rossion, 2004; Philiastides & Sajda, 2007; Rossion & Caharel, 2011; Rousselet, Husk, Bennett, & Sekuler, 2007; Rousselet et al., 2008a; Rousselet, Pernet, Bennett, & Sekuler, 2008b). In the present study, thresholds for the detectability of face-structure were measured using the sweep ssVEP method, in which the visibility of the face-structure was systematically increased (i.e., descrambled) while a face-specific response component was extracted using EEG spectrum analysis.

Materials and methods

Participants

Data are reported from 10 participants (six men; age range: 18–34 years; mean age: 25.8 years, SD: 6.1

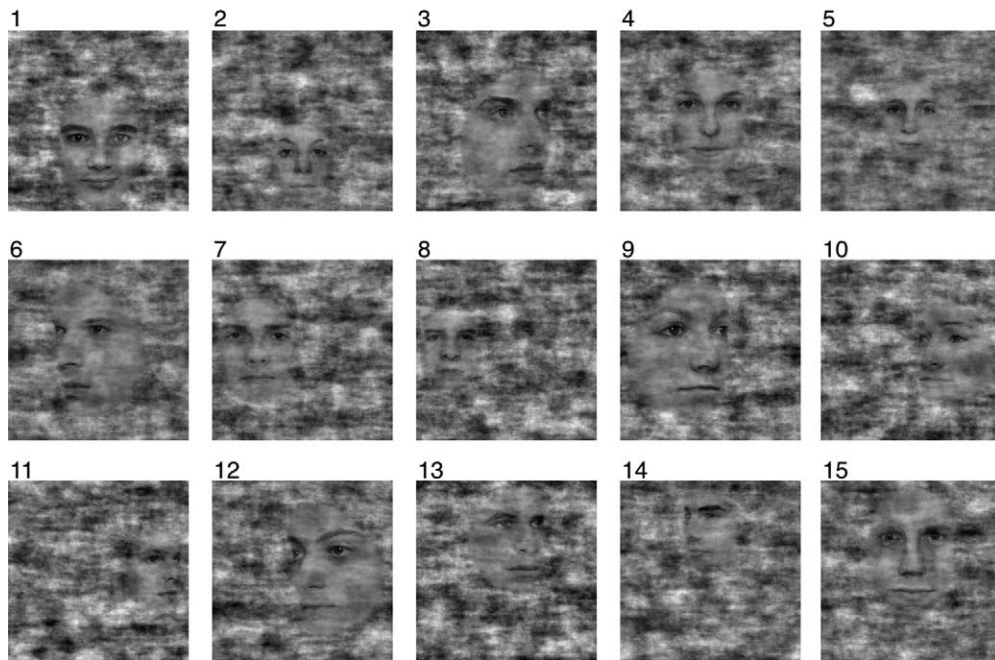


Figure 1. The full set (15) of 100% phase-coherent faces used in the study (with numbers corresponding to the data shown in the [Results](#) section). At the end of the 20-s stimulation sequence, a 100% phase-coherent face as displayed here alternated with a fully phase-scrambled version of the same stimulus.

years), each of whom had normal or corrected vision. Written informed consent in accordance with procedures approved by the Institutional Review Board of Stanford University was obtained from all participants prior to the start of the experiment.

Stimuli generation

Fifteen photographic face images were cropped to remove external features such as hair. The original stimuli varied in size (three levels), viewpoint (seven full-front, four left profile, four right profile) and spatial location on a uniform rectangular white background.

Previous studies have attempted to isolate evoked responses to faces from responses to low-level visual information such as luminance, contrast, and shape of the amplitude spectrum by comparing an entirely phase-scrambled face to an intact face (e.g., Nääsänen, 1999; Rossion & Caharel, 2011; Rousselet et al., 2008a, 2008b; Sadr & Sinha, 2004; Tanskanen et al., 2005). Our approach was different in that the image background remained fully scrambled throughout the entire sweep sequence. Also, face visibility was varied across steps (i.e., descrambled), which has been done previously in a few studies (e.g., Sadr & Sinha, 2004; Rousselet et al., 2008a, 2008b). As explained below, we varied face visibility by creating a graded sequence of

images with uniform degrees of scrambling and that maintained the same distribution of low-level image statistics, specifically equal power spectra and mean luminance. The 15 face images in their fully unscrambled state are shown in [Figure 1](#).

There were two distinct processes involved in the creation of the stimuli. The first was the creation of a set of face exemplars on noise backgrounds with identical power spectra from a set of unscrambled isolated face images, illustrated diagrammatically in [Figure 2](#). The second process involved the systematic degradation of these individual exemplars via phase scrambling.

To create the stimuli we first calculated the average power spectrum over the set of 15 isolated face exemplars. This power spectrum was then combined with the phase spectrum of each exemplar to create intermediate images with identical power spectra. Careful inspection of the face regions of [Figure 1](#) will reveal that the face regions contain noise. The face regions of these images are still 100% phase coherent with the face exemplars. The noise in the face regions is a result of balancing the power spectrum across the set of exemplars. The amount of noise added to the face regions as a result of changing the amplitude spectrum is shown in [Figure 2a](#) and [2c](#). If one replaces the white background of the top face in [Figure 2a](#) with a midgray background, then [Figure 2c](#) has a phase spectrum that is identical to that of the top image in [Figure 2a](#). Thus, the 100% coherent face stimulus is fully phase coherent

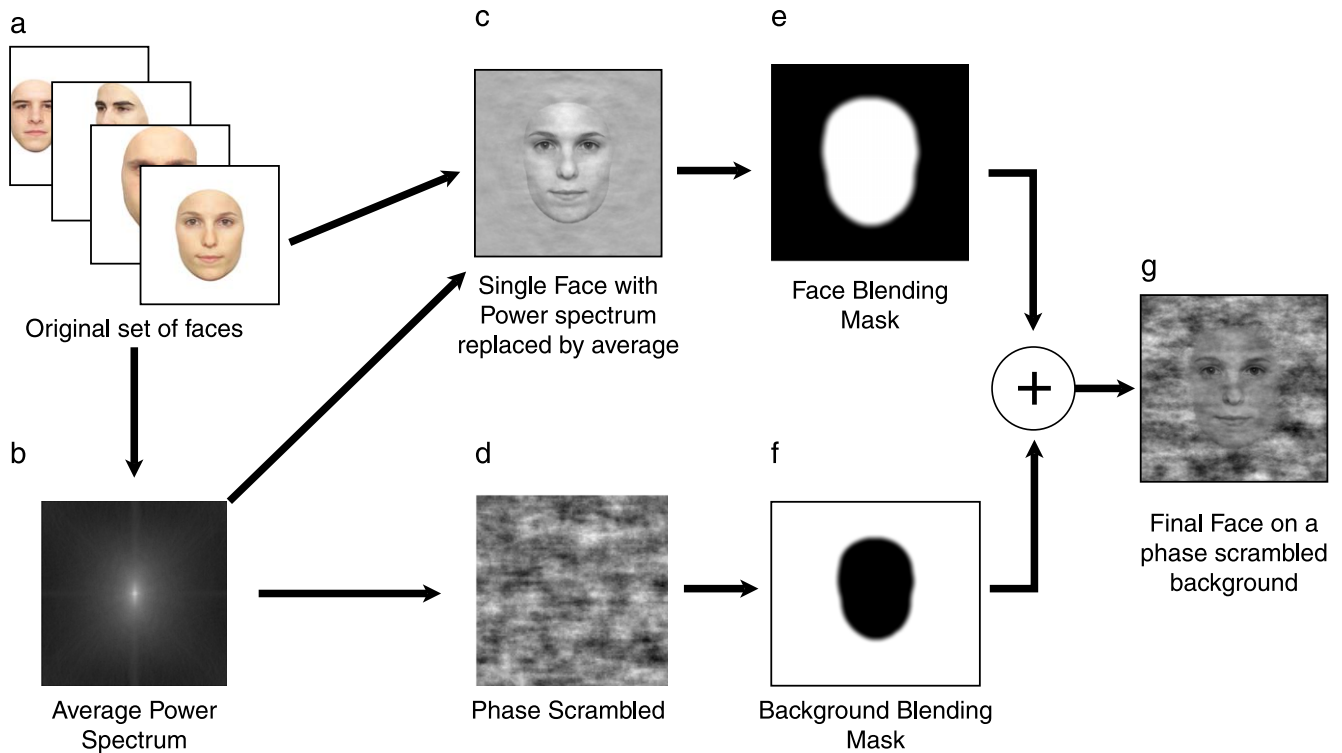


Figure 2. Flow-chart of stimulus generation. (a) Isolated, cropped faces of different sizes, poses, and spatial locations were derived from photographs. (b) The average power spectrum of the isolated faces was computed. (c) The power spectrum of each individual face exemplar was replaced with the power spectrum of the average, retaining the original phase spectrum of the exemplar. (d) A set of phase-randomized images was generated from the power spectrum of the average. (e) A smoothed blending mask was created for the face image (white indicates face visible, black not visible). (f) A complementary blending mask was generated for the background noise. (g) The face and background image were combined to create a face embedded in an equal power spectrum noise background.

in the face region, but is not 100% amplitude coherent. We wanted to embed each face in a random noise background of the same power spectrum as the faces in order to limit the introduction of a local contrast cue that would occur if isolated faces were scrambled. We thus created a set of background images from the average power spectrum image so that each had a uniform random phase distribution. The next step was to blend the isolated faces with the background images. The original isolated faces had an outline that created a visible discontinuity. To eliminate this discontinuity between the face region and the background region of the final images, we created complementary spatial blending masks that smoothly transitioned between regions. The blending masks were made such that they started within the face and ended by the face outline. Complementary masks for faces and the backgrounds were used to avoid an increase in contrast in the transition region. The complementary face and background images were then added to create the final equalized power spectrum faces.

The next step in creation of the stimuli was to generate a series of images that had progressively greater amounts of scrambling of the phase structure of

the face image. Interpolating between the unscrambled face and an image with uniform random phase, as done in previous studies (e.g., Rainer, Augath, Trinath, & Logothetis, 2001; Reinders, den Boer, & Büchel, 2005; Reinders et al., 2006), presents a problem. Phase is a circularly distributed quantity (Figure 3); therefore, progressive scrambling using simple linear interpolation introduces an artifact in the phase distribution (Dakin, 2002). Dakin (2002) introduced the weighted mean phase (WMP) procedure to solve the problem. WMP works by decomposing phase into individual sine and cosine components, interpolating these components, and transforming back to phase with the four-quadrant inverse tangent. While WMP avoids an overrepresentation of certain phases, it does not provide uniformly sized phase angle steps. Unequal phase angle steps is a limitation of previous EEG studies that have used this method to parametrically (de)scramble the phase of the stimulus (e.g., Rousselet et al., 2008b).

Another solution to the overrepresentation of phase was proposed by Sadr and Sinha (2004). In this solution, half of the Fourier coefficients in the power spectrum were assigned minimal-phase interpolation and the other half were assigned maximal-phase

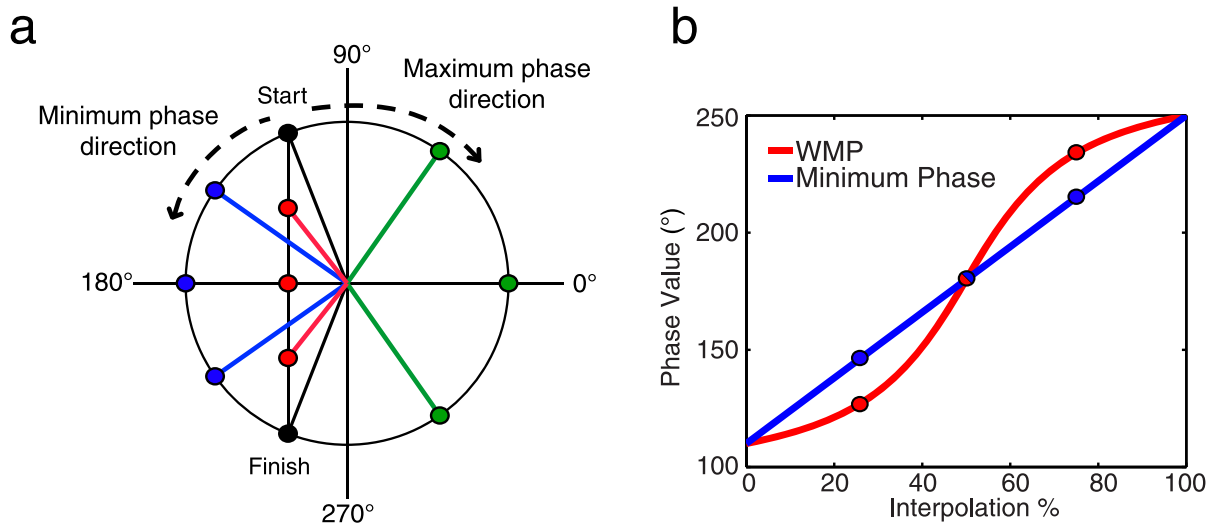


Figure 3. Graphical representation of phase circularity and phase scrambling algorithm used. (a) Start and finish phase values with three interpolation steps; red depicts steps created by weighted mean phase (WMP), green depicts steps created by maximum-phase method, and blue depicts steps created by minimum phase method (as used in the current study). (b) Comparison between step sizes created using WMP and the minimum-phase method (used here) of phase interpolation.

interpolation. This approach is nondeterministic and creates large transients in contrast for closely matched images, which is particularly problematic for EEG studies because these transients can generate spurious responses. The approach we took here was to linearly interpolate phase angle, but to choose the direction of interpolation that corresponded to the minimum distance between phases, irrespective of modulus boundaries. Using the minimum distance between phases preserves the uniformity of the phase distribution around the unit circle and provides equal sized steps.

The 20 steps that were swept for one face exemplar (one trial) are shown in Figure 4. For each face we interpolated between a starting image that had 100% randomized phases and the final unscrambled face exemplar. There were 20 equal steps in the interpolation. In order to destroy temporal correlations in luminance between successive scrambled images, the starting, fully random, image for the interpolation for each step in the sweep was chosen independently. The effects of the independent noise images can be seen by noting that on each step the noise background has been updated, and thus the noise masking of the face is different both because a new noise has been used and because the phase-coherence is different.

A total of 15 graded face image sequences were created for this study. These sequences contained faces that were highly variable in their visual appearance, size, and spatial location. The least scrambled image of each face exemplar is shown in Figure 1. Each sweep sequence included 20 steps, ranging from 0% to 100% interpolation of the original and random phase

spectra, with 5.26% change in coherence per step. A coherence level of 0 corresponded to a fully randomized phase spectrum of the original image and a coherence level of 100% corresponded to an unaltered phase spectrum.

Experimental design and procedure

The experiment consisted of the presentation of 45 20-s trials in which a face gradually emerged from a 0% coherence image on 1/3 of the trials. Each face-containing image was alternated with a 0% coherence image (face onset/offset presentation) at a rate of 3 Hz (Figure 5). An example trial for one face exemplar (face 9) is shown in Movie 1.

The 20 different steps of scrambling were presented for 1 s each using a newly computed random image for each step of the sweep. The sweep sequence was immediately preceded by a 1-s presentation of the first step of the sequence to allow the initial transient contrast appearance VEP to dissipate and the transition to the steady-state to begin. We used twice as many trials in which no face appeared in order to minimize participants' perceptual expectancies and guessing. Participants were instructed to press one response key (spacebar) as soon as they detected a face during the presentation of the sweep. They were asked to refrain from pressing a response key when no face was presented. Participants were also requested to maintain a constant level of confidence in their judgment across trials. They were informed that target faces were present in only a subset of the trials and that the faces

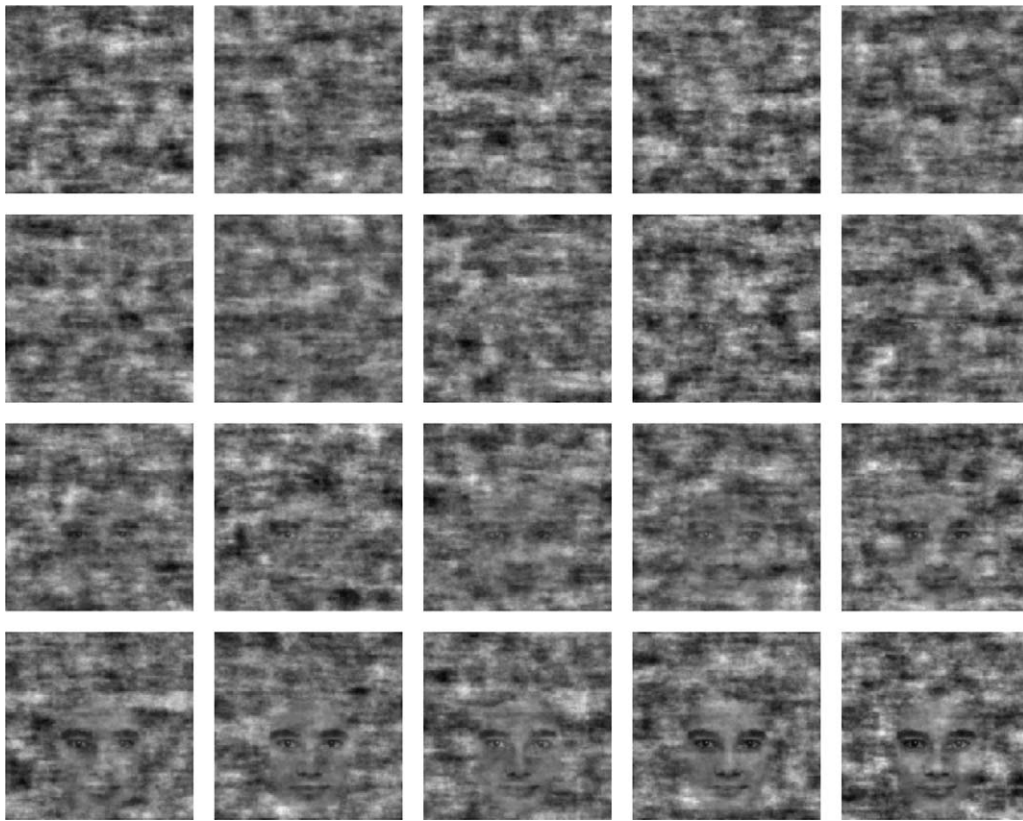


Figure 4. The 20 images of face 1 in decreasing order of scrambling. During the experiment, the first image of the sequence alternated with a fully scrambled stimulus for 1 s (three cycles) before the next image alternated with another fully phase-scrambled stimulus for 1 s, and so on.

could vary in size, appearance, and their spatial location within the image. Note that after the participant indicated their detection of a face, the presentation of the sweep continued until the last step.

Stimuli were presented as gray-scale images on a contrast linearized CRT at a resolution of 800×600 , a 72-Hz vertical refresh rate, and a mean luminance of 50.31 cd/m^2 . The images were always presented in the center of the screen and subtended a visual angle of approximately 15° .

ssVEP recording

The EEG data were collected using a 128-channel HydroCell Geodesic Sensor Net (Electrical Geodesics Inc., Eugene, OR), bandpass filtered from 0.1 to 200 Hz, and digitized at a rate of 432 Hz (Net Amps 300 TM, Electrical Geodesics, Inc.). Individual electrodes were adjusted until impedances were below $60 \text{ k}\Omega$ before starting the recording. Data were evaluated off-line with custom-made software (PowerDiva). 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 $30 \mu\text{V}$ were excluded on a sensor-by-sensor basis. Typically, these epochs included eye movements or blinks.

ssVEP threshold estimation

Individual VEP thresholds were estimated from an integrated first harmonic (1F; 3 Hz) response function. Voltages recorded from each step of the sweep were added together to form a cumulative response function that was guaranteed to be monotonically increasing. To estimate the EEG background noise, the same integration was performed at 2.5 and 3.5 Hz where there was no stimulus-related activity. We compared the cumulative sum of the signal to that of the noise, both normalized by the sum of the signal amplitude. This procedure reflects the percentage of the measured response that is signal. We then used an arbitrary threshold of 10% signal to determine the coherence level at which the integrated 1F response function diverged from the noise function. This coherence level was taken as the threshold of face detection.

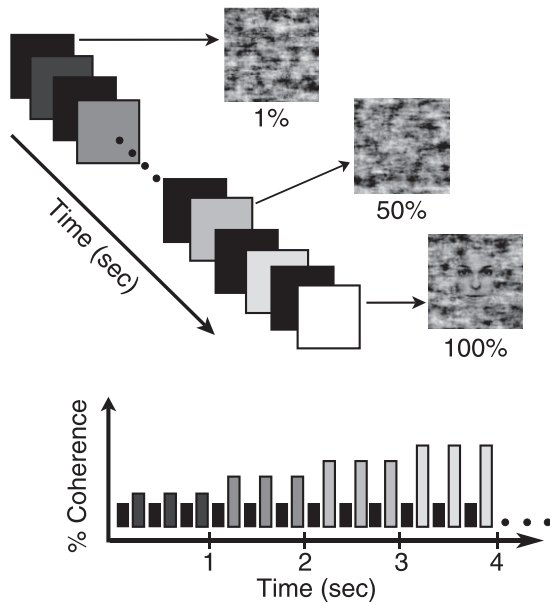


Figure 5. Schematic illustration of the face coherence sweep ssVEP paradigm. In this method, a phase-scrambled face alternates with a stimulus that evolves from a phase-scrambled face into a fully coherent face at 3 Hz over 20 s of stimulation. At the beginning of the sweep, the face-containing image has an almost entirely phase-randomized spectrum. Over the trial, the degree of phase-scrambling is decreased in a series of equal steps, three of which are illustrated. The black bars and black square icons indicate the fully randomized images. Gray bars and gray square icons indicate partially randomized images, with lighter colors representing lower levels of scrambling.

Results

First and second harmonics

Activity at the first harmonic (1F; 3 Hz) was only found for trials in which a face image was presented. Figure 6 (top panel) shows the topography of the group-averaged 1F response measured across all values of coherence (0%–100%). The response was distributed bilaterally with a maximum over the right hemisphere around channel 96 (P10). Activity at the first harmonic for the control trials that did not contain a face (Figure 6, top right) was not different from the experimental noise level. By contrast, the group averaged second harmonic (2F; 6 Hz) response was maximal over the occipital midline around channel 75 (OZ) and was comparable in magnitude between face and no-face trials (Figure 6, bottom left and right).

On the trials in which no face appeared in the sweep sequence, there were only 1.4% of the channels across all coherence steps that contained a signal significantly above noise level ($p < 0.00002$). Response phase was largely constant during the sweep (data not shown) so

collapsing across steps did not result in cancellation of responses that could have occurred if there were large phase differences over the different coherence values of the sweep. The large majority (92%) of these significant channels were located posteriorly, showing an effect only at the beginning of the stimulation (step 1 of the sequence). This activity may reflect a small residual of the transient VEP that is generated at the onset of the visual stimulation. At the end of the sweep there was no significant signal above noise on any of the channels.

Figure 7 shows the distribution of response components over the 0.5 to 15 Hz range is shown for face and no-face trials at three representative electrodes (two lateral and one mid-line electrode). The first and second harmonic components were found to be the largest, followed by the fourth harmonic (12 Hz). Odd harmonic responses (3 and 9 Hz) were present only for the face trials, especially over the right hemisphere where the first harmonic response was larger than the second harmonic response (Figure 7, top right panel). Even harmonic responses (6 and 12 Hz), but not odd harmonic responses, were present for the no-face trials (Figure 7, bottom right panel)

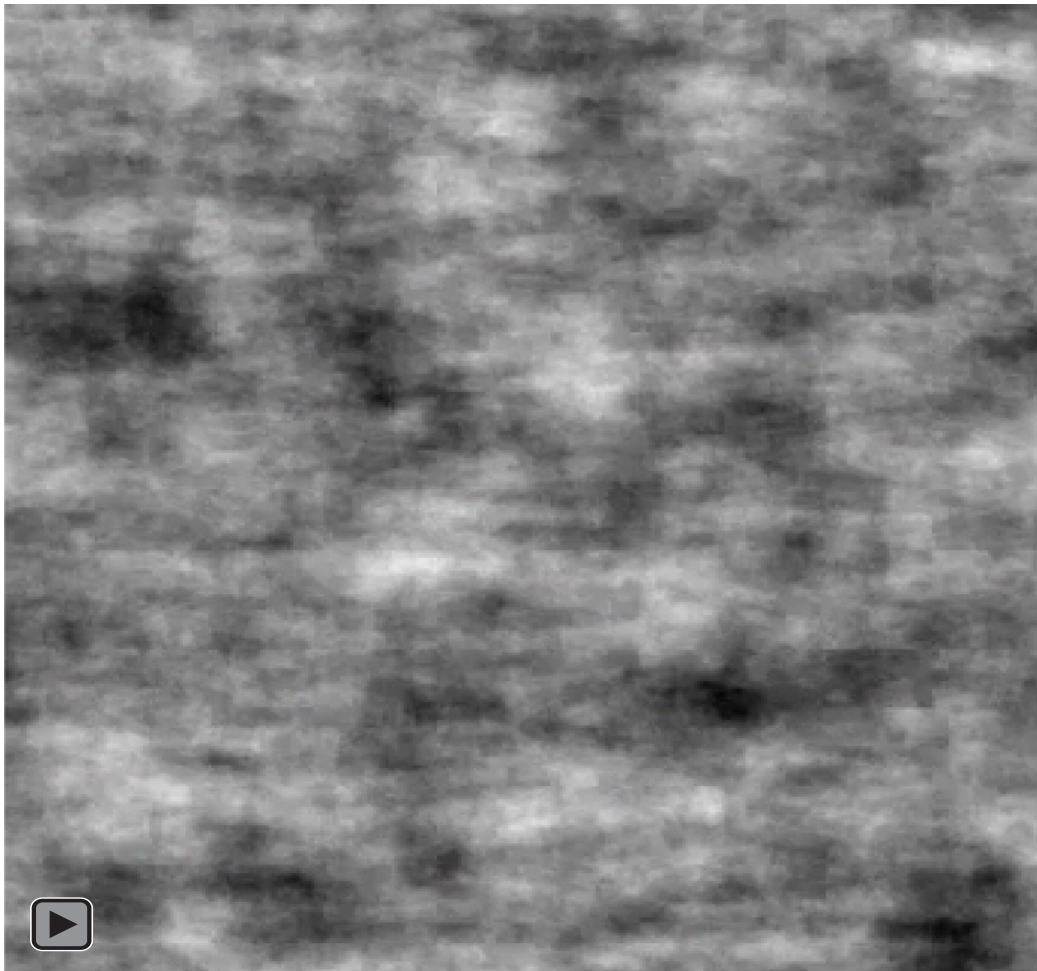
Figure 8 (left panel) plots the ratio of the first harmonic response relative to the sum of the first and second harmonic responses. This index ratio reflects the degree to which the total response is dominated by odd (face-specific) or even (not face-specific) activity. The index was plotted collapsed across all steps of the sweep. The selectivity index shows focal peaks bilaterally with maxima lying anteriorly to the maxima of the first harmonic itself. The values of the index are shown for channel 65, 75, and 96 in the right panel of Figure 8.

Sweep response functions

The 1F amplitude versus phase-coherence sweep response function averaged across all participants and all face exemplars is shown in Figure 9. We found that ssVEP amplitude at the first harmonic rose above the noise level abruptly rather than linearly, starting at about 30% phase-coherence (step 7). The response reached a plateau by about 40% coherence (step 15).

In contrast, for no-face trials, the first harmonic sweep response was not above the experimental noise level even at the end of the sequence, and did not rise above the noise level throughout the entire sweep. The first harmonic was thus specifically evoked by image sequences that alternated between face-containing and phase-randomized images.

The second harmonic sweep response function for face trials was nearly constant across all 20 steps of image coherence (Figure 10). This response is driven by the contrast changes that occur after each update of the image. These updates occur at 6 Hz. Comparable data



Movie 1. Example trial of the face coherence sweep ssVEP paradigm.

is shown for the no-face trials and the amplitudes were also constant across steps and of similar magnitude to those measured in the face trials (Figure 10).

Comparison of ssVEP and psychophysical face detection thresholds

The distribution of face detection behavioral response times is shown on the same axis as the group averaged 1F sweep response in Figure 9. The mean behavioral response for face detection occurred at around 45% coherence with the modal detection threshold that was slightly lower. Behavioral detection began at coherence levels where the evoked response first began to rise above the noise. The evoked response reached a plateau at face coherence values near the modal decision time.

The behavioral face detection thresholds varied substantially across face exemplars (range: 33%–73% coherence), likely a consequence of the variability in size, viewpoint, and spatial location of face presentation (Figure 11). Individual participants also showed a

range of detection thresholds (range: 41%–52% coherence) when pooled over face exemplars.

The inter-face and inter-participant variance was used to compare ssVEP with psychophysical face detection thresholds to test whether the electrophysiological and behavioral thresholds covaried. This analysis allowed us to determine whether the ssVEP thresholds tracked the variations in perceptual face detection. Figure 12 illustrates our procedure for determining ssVEP face detection thresholds. A standard method for determining the threshold for a swept parameter ssVEP measurement is to fit a line to the linear part of the response function and define threshold as the zero voltage intercept of this fit. This procedure works well when the response function is relatively linear with respect to the changing stimulus parameter. For the current stimulus, however, the response was closer to a step function. Because it was a step function, there were very few response steps that could be used for a regression-to-zero threshold estimation. Another method for determining the response threshold is to find the first step at which the response differs significantly from the noise. However, because this type of threshold measurement

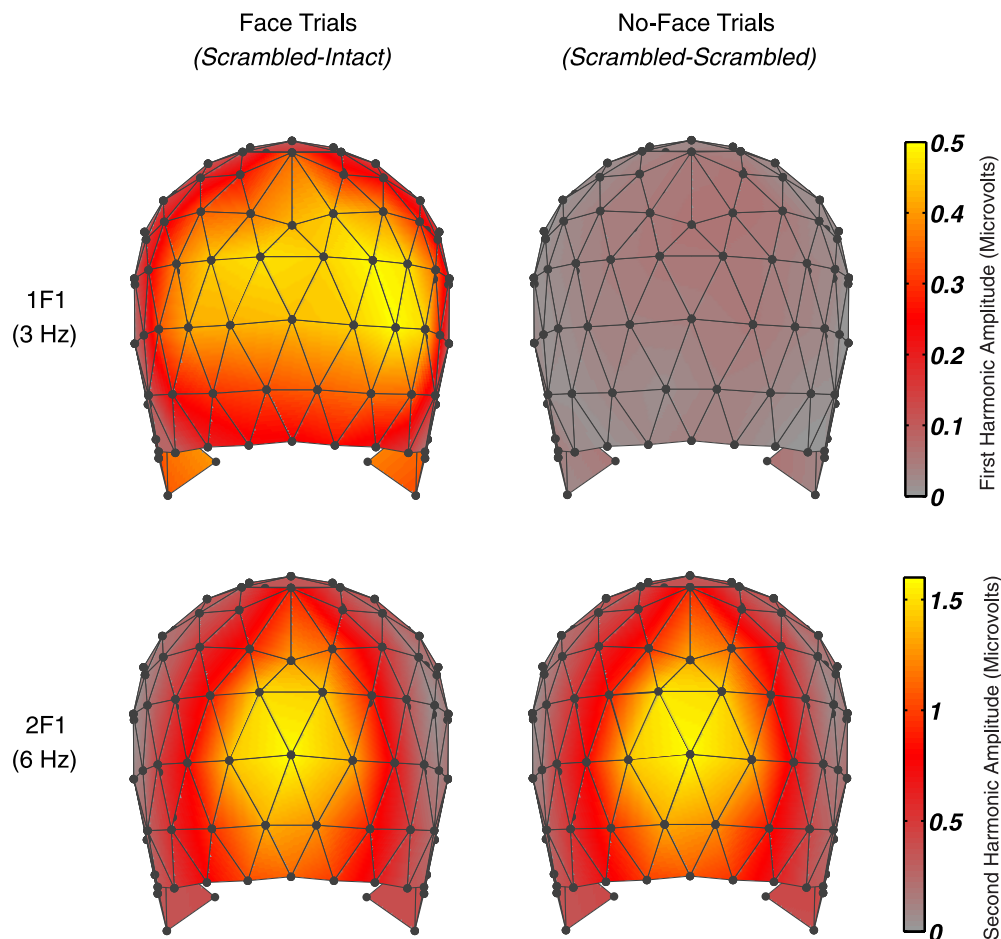


Figure 6. Scalp topography for first (top) and second (bottom) harmonic responses averaged across all sweep steps of face trials (left) and no-face trials (right). The first harmonic response was observed only for the face trials, and showed a broad distribution over the posterior scalp, maximal over right occipito-temporal electrodes. The nonspecific second harmonic response was distributed focally over the medial occipital electrodes, for both trial types.

relies on the lowest signal-to-noise ratio signals, it can be a highly variable estimation.

For the present threshold estimations, we thus adapted a method used to quantify ERP latencies. This method first determines the fractional area of a component, and defines the latency of the component as the time at which a certain fraction of the response has occurred (Luck, 2005). Adapting this method for use with sweep ssVEP data requires recognizing that amplitude is always positive (even for noise only measurements) and therefore one must also take into account the amount of noise that is contributing to the area measure. Figure 12a shows the response for a single face from a single trial averaged over the 10 participants. Figure 12b shows the results of taking the cumulative sum of the data in Figure 12a. We compared the cumulative sum of the signal to that of the noise, both normalized by the sum of the signal amplitude (therefore the final 1F value is 100% by

definition). Figure 12c shows the difference between the signal at 1F and the noise estimated from two adjacent EEG frequencies. This curve is indicative of the percentage of signal present at each coherence value. We then arbitrarily defined the ssVEP threshold as the coherence level at which 10% of the cumulative sum was signal. Figure 12d shows the same analysis as 12c, but for all face exemplars.

Figure 13 shows the correlation between ssVEP thresholds derived as described above and psychophysical thresholds for each face exemplar. The ssVEP and psychophysical face detection thresholds were significantly correlated ($R^2 = 0.93$; $p < 1e-8$). Figure 14 presents the same comparison across participants and here the correlation was also significant ($R^2 = 0.86$, $p < 0.001$). The slopes of the regression lines were both close to 1, indicating a 1:1 relationship between ssVEP and perceptual sensitivity.

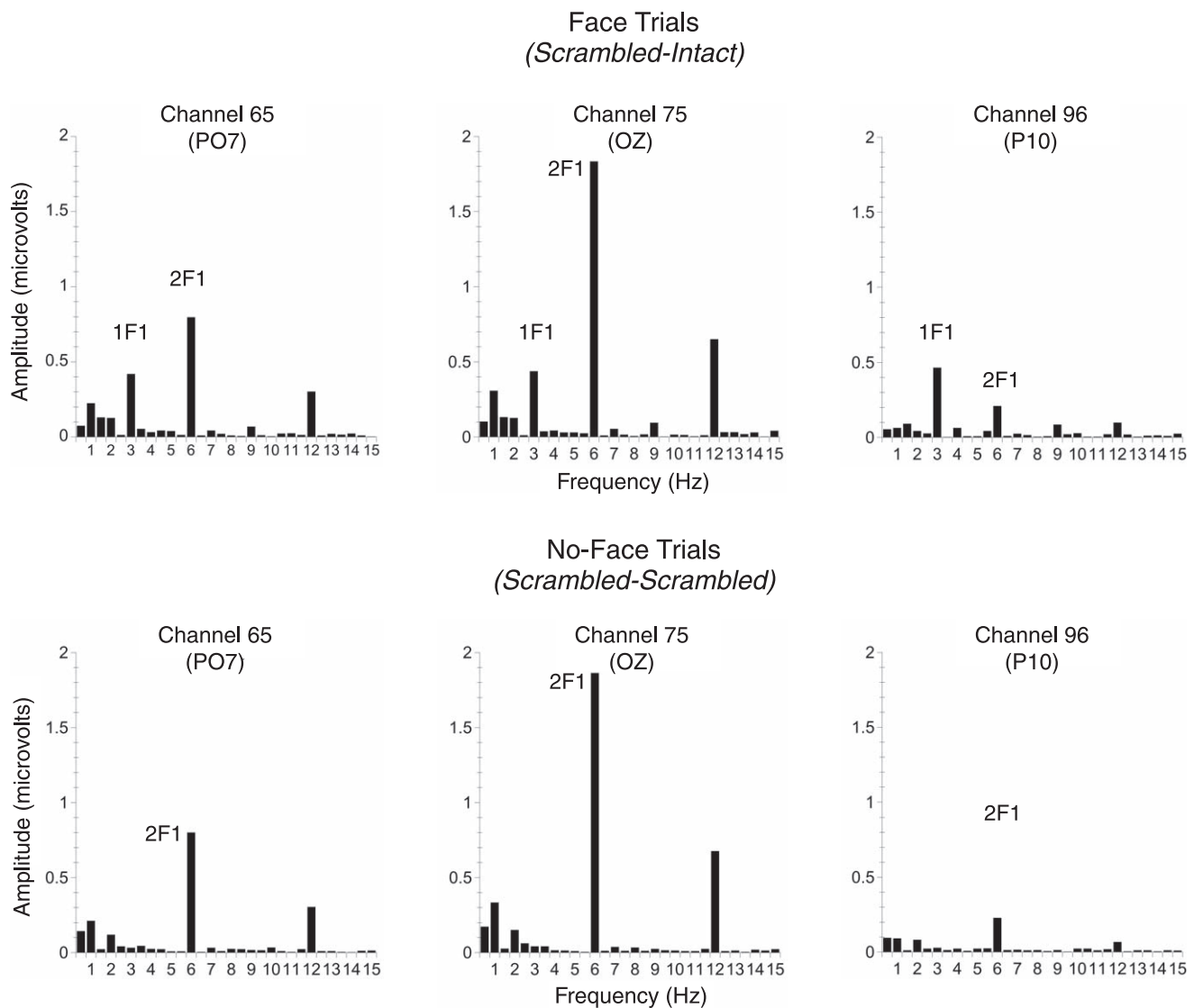


Figure 7. EEG spectra (0.5–15 Hz; frequency resolution of 0.5 Hz) at three occipital channels, averaged across all sweep steps of face trials (top) and no-face trials (bottom). For the face trials (top), the spectra show the distinct first harmonic response (3 Hz), which was particularly prominent on lateral occipital sites (PO7 on the left, P10 on the right). Over the right occipito-temporal site, the 1F response was the largest (note also the presence of the 3F response at 9 Hz). For the no-face trials (bottom), there was no distinct response at the first harmonic (3 Hz).

Discussion

We have developed a novel method based on the sweep ssVEP for obtaining an objective, sensitive, and behavior-free measure of face detection. Our stimuli were segmented faces of differing sizes, viewpoints, and spatial locations, which resulted in a group-level ssVEP face detection threshold at 30%–35% phase-coherence of the face image. Thresholds were reliably estimated from individual participants over 15 face trials, and for each of the 15 face trials when averaged over the 10 participants.

Behavioral measures of perceptual face detection reflect a complex chain of concurrent sensory and motor decision processes, and task performance can be impacted by a number of extraneous factors, such as response criterion, attention, motivation, and response selection. By contrast, our electrophysiological approach provides a sensitive neural measurement that isolates responses specific to the image structure of faces, but does not rely on a behavioral response. This is particularly important if one aims to obtain face detection thresholds from infants and children, individuals with cognitive impairments, or nonhuman populations.

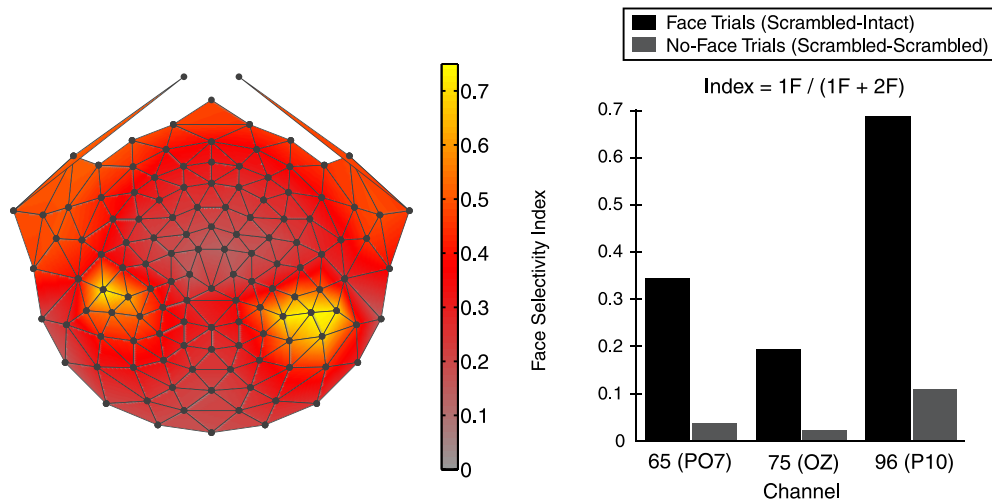


Figure 8. Two-dimensional scalp map showing the index of the first harmonic response relative to the sum of the two harmonic responses, for both trial types. Channel 96 (PO10) showed the most specific increase of the first harmonic response associated with face coherence.

Comparison to previous ERP studies of face detection

Previous studies of face processing have utilized transient ERPs, which provide important information about the time-course of face perception. When low-level visual cues were carefully controlled, for instance by means of phase-scrambling procedures similar to those used here, ERP results have shown that faces are detected at around 120–130 ms following face onset, with peak discrimination occurring at 160–170 ms on average (N170 face-sensitive component) (Jacques & Rossion, 2004; Rossion & Caharel, 2011; Rossion & Jacques, 2008; Rousselet et al., 2007, 2008a, 2008b). Transient ERPs, however, have several limitations that

are improved by the steady-state technique we present here.

A first limitation of transient ERP studies concerns the ambiguity in component selection. A flashed face stimulus elicits a sequence of evoked response components on the scalp that can be defined as visual potentials: C1(N170), P1, N1/N170, P2, N250, etc. These components vary in terms of their polarity, peak latency and amplitude, and topography. While these components provide a rich source of information about the time-course of a given process, for instance face detection, it is difficult to objectively associate a specific process to one of these components or to a defined time-window falling in between these components. This difficulty is largely based on the subjective criteria used

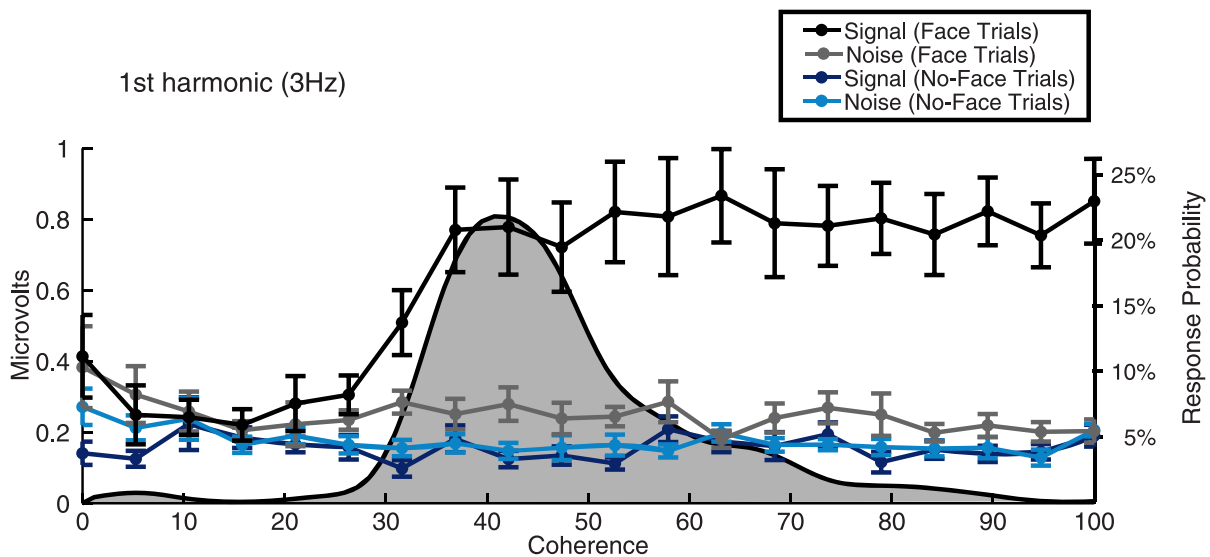


Figure 9. Amplitude of the first harmonic (3 Hz) as a function of coherence, as recorded on channel 96 (P10). Error bars represent 1 standard error of the mean across participants. The gray region shows the probability distribution of behavioral responses.

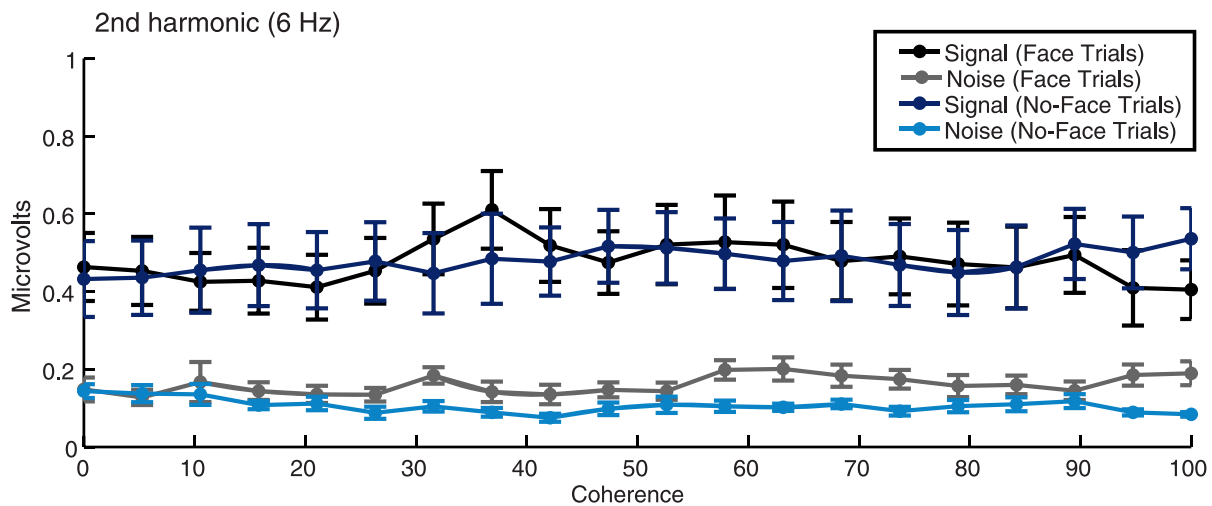


Figure 10. Amplitude of the second harmonic (6 Hz) as a function of coherence, as recorded on channel 96 (P10). Error bars represent 1 standard error of the mean across participants.

to identify these components. Moreover, components elicited by face stimulation can be particularly difficult to identify when they are measured in infants, children, or neurologically affected patients because there can be significant variability in the number, timing, and morphology of the components with development and clinical condition (Kuefner et al., 2010; Prieto et al., 2011).

Importantly, the limitation described above also applies to transient ERP studies that rely on time-point analyses rather than on defined ERP components (e.g., Rousselet et al., 2008b). Baseline or latency differences between two stimulus conditions can lead to spurious “face-specific” responses occurring at multiple time-points, and there is an inherent inefficiency in independently estimating the low-level feature response. In contrast, the sweep ssVEP approach allows for an unambiguous (i.e., objective) quantitative

analysis of the face-specific response: the first harmonic response (3 Hz here) is defined by the paradigm and selected by the experimenter and is demonstrably face specific (see Figure 1). This component can be measured from a single stimulus condition, rather than requiring a subtraction of separately measured test and control responses. By sweeping the level of phase-coherence of the face, a threshold can be objectively determined, thereby providing a direct measure of face detection.

Specificity of the first harmonic

In the ssVEP paradigm used here, the specificity of the first harmonic for face structure derives from image symmetry considerations and from careful stimulus

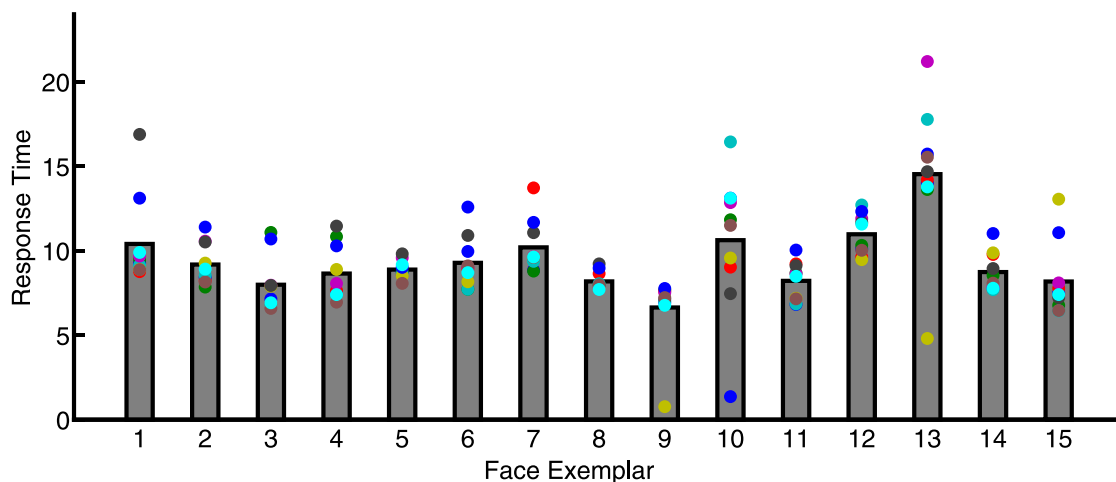


Figure 11. Average behavioral face detection response time for each face (10 s = half of the sequence, or 50% coherence). Dots represent individual participants’ response time for each face.

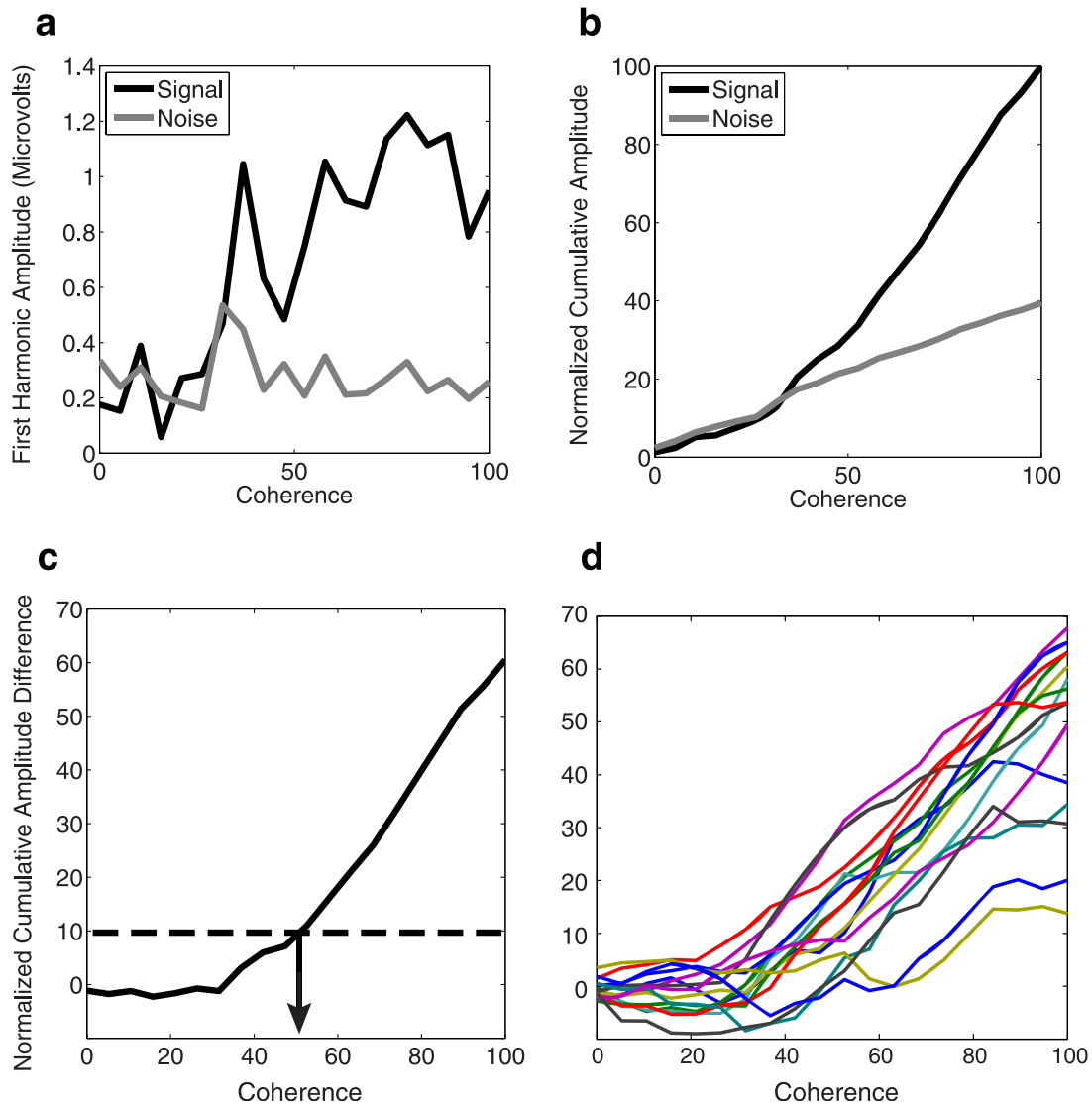


Figure 12. Method used to derive ssVEP threshold. (a) Amplitude of the first harmonic (3 Hz) as a function of coherence, as recorded on channel 96 (P10). These data are from a single presentation of face 4 averaged over 10 participants. Gray curve plots the noise level measured at nearby frequencies in the EEG. (b) Cumulative integral of the data from (a); both signal and noise are normalized by the sum of the signal amplitude. (c) Difference between signal and noise from (b) with ssVEP threshold criterion of 10% normalized signal shown as a dashed line. (d) Normalized cumulative amplitude difference for all 15 faces used in the study.

control. We alternated between two images that had equal power spectra and mean luminances. So, if the brain detects differences in the power spectrum or luminance of the two images, then transition responses from one image to the other should be identical because the underlying distribution of neural population activity should be the same at the level of resolution of the scalp-recorded VEP. If, on the other hand, there are populations of neurons that are sensitive to statistical regularities that are present in the face image and that are not captured by the power spectrum, then the populations that code face-containing images and the scrambled ones will not be the same. This nonequivalence of underlying neuronal responses opens the way to measuring nonequivalent evoked

responses to transitions between a face-containing and a scrambled image. These nonequivalent transition responses project onto the odd harmonics of the evoked response.

The crux of the ssVEP method is control over other factors that might lead to differential population responses from transitions between the different images, such as differences in mean luminance, or average contrast that could also lead to asymmetric, odd harmonic responses. Our stimulus set is sufficiently well controlled that we did not evoke an odd harmonic response at the beginning of the stimulation sequence, or at any step during the no-face sweep trials. Our phase scrambling method was carefully designed to create steps with equivalent changes in the stimulus.

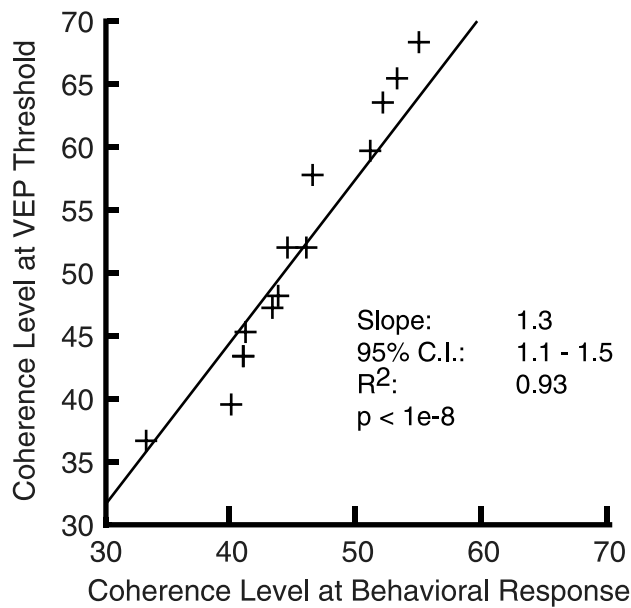


Figure 13. Correlation between ssVEP (channel 91) and psychophysical face detection thresholds for each face exemplar. Each data point represents the average of 10 participants. The best fitting two-parameter (slope and offset) line to the data is shown.

Thus, the odd harmonic we measured from 30%–35% phase coherence in the face-containing trials is specific to some level of structure in the face images that is higher order than the power spectrum. For this reason, the success of our approach depends to a greater extent than other approaches on a tight control of low-level visual features of the stimuli. As a result, the sweep ssVEP technique provides the advantage that lack of an adequate no-face control stimulus will be immediately visible from the shape of the response (i.e., the presence of a first harmonic response for “symmetrical” stimuli).

Signal-to-noise ratio advantage for ssVEP

Our sweep ssVEP approach to measure face detection overcomes yet a second limitation of transient ERP measures: their low signal-to-noise ratio (SNR), which requires the recording of a large number of independent trials. Here, because the visual system can be driven precisely by the periodic stimulation, all of the response, and thus all of the effect, is concentrated into a frequency band that occupies a very small fraction of the total EEG bandwidth. In contrast, biological noise is distributed throughout the EEG spectrum, so that the SNR in the bandwidth of interest can be very high (Regan, 1989). Moreover, the differential activity is present at an exactly known temporal frequency in the EEG, making it possible to

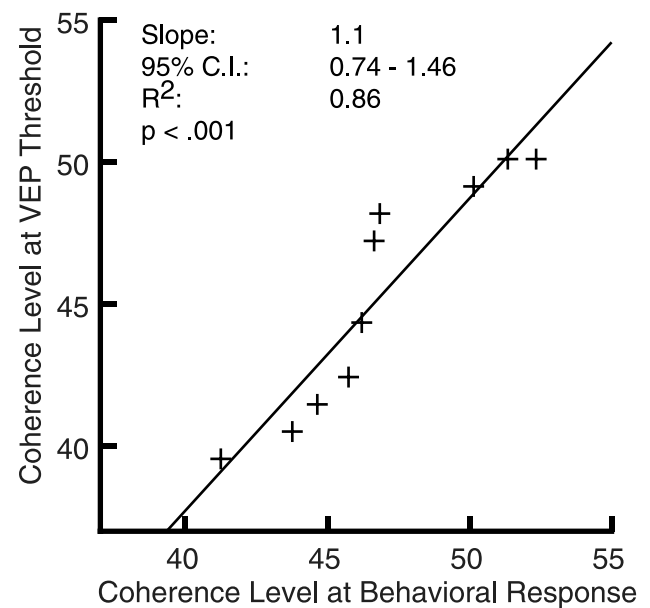


Figure 14. Correlation between ssVEP (channel 91) and psychophysical face detection thresholds for each participant. Each data point represents the average of 15 face exemplars. The best fitting two-parameter (slope and offset) line to the data is shown.

use a highly selective filter (spectrum analysis) to separate signal from noise.

Objective threshold estimation

A third advantage of the present sweep ssVEP approach to measure face detection is that it provides a threshold estimation by identifying the first image that leads to a first harmonic response, or by regression to zero amplitude, as has been done in the past for sweeps with low-level visual stimuli (Tyler et al., 1979). In contrast, despite the use of highly homogenous stimuli only (full-front faces with no variation in spatial location, viewpoint, and size), previous ERP studies that used parametric manipulations of face stimuli embedded in noise (Jemel et al., 2003; Rousselet et al., 2008b) were not designed to use the parametric variation as a means to estimate perceptual thresholds of face detection.

Future optimization of the approach

As observed in the grand-averaged first harmonic sweep response data, and for most participants, 30%–35% of phase coherence was sufficient to elicit a significant first harmonic response associated with face detection. Obviously, this amount of phase-coherence does not represent an absolute limit for the face

detection threshold but is only valid for the variable set of images used here. If we had used a more homogenous set of face stimuli, for instance a set of full-front faces presented centrally and of the same size, the face detection threshold might have been identified at a lower level of phase-coherence in the sweep sequence. However, under such highly predictable conditions, participants may have learned to anticipate the presence of a face from limited cues emerging constantly at the same location (e.g., one eye, the overall outline of the face). Here, variability was of interest as a means of creating unpredictability over which we could compare covariation of the electrophysiological and psychophysical thresholds observed for different face stimuli.

Despite this threshold variability, only a few (15) face trials were needed to estimate face detection thresholds reliably. This observation suggests that with a homogenous set of faces, the sweep ssVEP approach might be able to determine face detection thresholds from a smaller number of trials. Finally, sampling multiple frequency rates with the present paradigm could also be valuable in a future study, as it would provide an estimate of response latency from the phase values of the Fourier transform (Regan, 1989), while maintaining all of the advantages of the approach.

Face-specificity and generalization

Several factors motivated our decision to use faces as the image category for extension of the sweep ssVEP approach to high-level vision. Faces form a highly visually homogenous set of familiar stimuli, which are associated with large and well-defined neural responses. Faces are detected faster and more automatically than other stimuli (Crouzet et al., 2010; Fletcher-Watson et al., 2008; Hershler & Hochstein, 2005; Hershler et al., 2010; Kiani, Esteky, & Tanaka, 2005; although see Van Rullen, 2006), and computer scientists have devoted considerable efforts to building systems that automatically detect faces in images (e.g., Kemelmacher-Shlizerman, Basri, & Nadler, 2008; Viola & Jones, 2004; Yang, Kriegman, & Ahuja, 2002). However, the method developed here is not restricted to faces and could potentially be used to determine the thresholds for categorization of other classes of natural images. The sweep ssVEP could also be extended to the detection of faces or objects in nonsegmented images; that is, in complex visual scenes scrambled with a similar approach (e.g., Jiang et al., 2011).

Here, we cannot, and do not, claim that the 3-Hz first harmonic response obtained is specific to faces per se; rather, it reflects the detection of structure in the intact face stimuli that could be a specific feature of faces (e.g., eyes) or a feature that could have potentially

been obtained with other natural or with synthetic image classes. However, the observation of the largest and earliest first harmonic response over the right occipito-temporal cortex, at the same electrode sites where both the face-sensitive N170 component (Bentin et al., 1996; Rossion & Jacques, 2011) and the face-related ssVEP response (Rossion & Boremanse, 2011) have been found, is suggestive of responses from face-selective populations of neurons. Lastly, our data do not allow us to determine whether the face detection thresholds we have derived are determined solely by the physical attributes of the stimulus, or whether they depend on the task we have asked the observers to perform. These questions could be addressed in future studies using this method with appropriately designed stimuli and behavioral tasks.

Acknowledgments

This research was supported by National Institutes of Health grants EY06579 (AMN) and F32EY021389 (FF), Belgian National Fund for Scientific Research (BR), and ERC starting grant facessvep 284025 (BR). The authors wish to thank Corentin Jacques, Renaud Laguesse, and Ken Nakayama for providing stimuli used in the initial development of the face sweep VEP method, and Francesca Pei, who performed early recordings of face onset/offset responses based on modulation of the organization of image structure.

* These authors contributed equally.

Commercial relationships: none.

Corresponding author: Justin M. Ales.

Email: justin.ales@stanford.edu.

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

References

- Allison, T., McCarthy, G., Nobre, A., Puce, A., & Belger, A. (1994). Human extrastriate visual cortex and the perception of faces, words, numbers, and colors. *Cerebral Cortex*, 4, 544–554.
- Almoqbel, F., Leat, S. J., & Irving, E. (2008). The technique, validity and clinical use of the sweep VEP. *Ophthalmic & Physiological Optics*, 28(5), 393–403.
- Barbeau, E. J., Taylor, M. J., Regis, J., Marquis, P., Chauvel, P., & Liégeois-Chauvel, C. (2008). Spatio-temporal dynamics of face recognition. *Cerebral Cortex*, 18, 997–1009.

- Bentin, S., Allison, T., Puce, A., Perez, E., & McCarthy, G. (1996). Electrophysiological studies of face perception in humans. *Journal of Cognitive Neuroscience*, 8, 551–565.
- Brown, V., Huey, D., & Findlay, J. M. (1997). Face detection in peripheral vision: Do faces pop out? *Perception*, 26(12), 1555–1570.
- Cerf, M., Harel, J., Einhäuser, W., & Koch, C. (2008). Predicting human gaze using low-level saliency combined with face detection. In J. C. Platt, D. Koller, Y. Singer, & S. Roweis (Eds.), *Advances in neural information processing systems* (Vol. 20, pp. 241–248). Cambridge, MA: MIT Press.
- Cooper, E. E., & Wojan, T. J. (2000). Differences in the coding of spatial relations in face identification and basic-level object recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 26(2), 470–488.
- Crouzet, S. M., Kirchner, H., & Thorpe, S. J. (2010). Fast saccades toward faces: Face detection in just 100 ms. *Journal of Vision*, 10(4):16, 1–17, <http://www.journalofvision.org/content/10/4/16>, doi:10.1167/10.4.16. [PubMed] [Article]
- Dakin, S. C., Hess, R. F., Ledgeway, T., & Achtman, R. L. (2002). What causes non-monotonic tuning of fMRI response to noisy images? *Current Biology*, 12(14), R476–477.
- Fei-Fei, L., Iyer, A., Koch, C., & Perona, P. (2007). What do we perceive in a glance of a real-world scene? *Journal of Vision*, 7(1):10, 1–29, <http://www.journalofvision.org/content/7/1/10>, doi:10.1167/7.1.10. [PubMed] [Article]
- Fletcher-Watson, S., Findlay, J. M., Leekam, S. R., & Benson, V. (2008). Rapid detection of person information in a naturalistic scene. *Perception*, 37(4), 571–583.
- Garrido, L., Duchaine, B., & Nakayama, K. (2008). Face detection in normal and prosopagnosic individuals. *Journal of Neuropsychology*, 2(Pt 1), 119–140.
- Goren, C., Sarty, M., & Wu, R. (1975). Visual following and pattern discrimination of face-like stimuli by newborn infants. *Pediatrics*, 56, 544–549.
- Halgren, E., Raij, T., Marinkovic, K., Jousmäki, V., & Hari, R. (2000). Cognitive response profile of the human fusiform face area as determined by MEG. *Cerebral Cortex*, 10, 69–81.
- Haxby, J. V., Hoffman, E. A., & Gobbini, M. I. (2000). The distributed human neural system for face perception. *Trends in Cognitive Science*, 4, 223–233.
- Hershler, O., Golan, T., Bentin, S., & Hochstein, S. (2010). The wide window of face detection. *Journal of Vision*, 10(10):21, <http://www.journalofvision.org/content/10/10/21>, doi:10.1167/10.10.21. [PubMed] [Article]
- Hershler, O., & Hochstein, S. (2005). At first sight: A high-level pop-out effect for faces. *Vision Research*, 45, 1707–1724.
- Jacques, C., & Rossion, B. (2004). Concurrent processing reveals competition between visual representations of faces. *Neuroreport*, 15, 2417–2421.
- Jeffreys, D. A. (1989). A face-responsive potential recorded from the human scalp. *Experimental Brain Research*, 78, 193–202.
- Jemel, B., Schuller, A., Cheref-Khan, Y., Goffaux, V., Crommelinck, M., & Bruyer, R. (2003). Stepwise emergence of the face-sensitive N170 event-related potential component. *NeuroReport*, 16, 2035–2039.
- Jiang, F., Dricot, L., Weber, J., Righi, G., Tarr, M. J., Goebel, R., et al. (2011). Face categorization in visual scenes may start in a higher order area of the right fusiform gyrus: Evidence from dynamic visual stimulation in neuroimaging. *Journal of Neurophysiology*, 106, 2720–2736.
- Johnson, M. H., Dziurawiec, S., Ellis, H., & Morton, J. (1991). The tracking of face-like stimuli by newborn infants and its subsequent decline. *Cognition*, 40, 1–21.
- 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, 4302–4311.
- Kemelmacher-Shlizerman, L., Basri, R., & Nadler, B. (2008). 3D shape reconstruction of Mooney faces. Paper presented at the *IEEE Conference on Computer Vision and Pattern Recognition*. (pp. 1–8). Anchorage, AK.
- Kiani, R., Esteky, H., & Tanaka, K. (2005). Differences in onset latency of macaque inferotemporal neural responses to primate and non-primate faces. *Journal of Neurophysiology*, 94(2), 1587–1596.
- 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, doi:10.3389/neuro.09.067.2009.
- Lewis, M. B., & Edmonds, A. J. (2003). Face detection: Mapping human performance. *Perception*, 32, 903–920.
- Luck, S. J. (2005). *An introduction to the event-related potential technique*. Cambridge, MA: MIT Press.
- McKeeff, T. J., & Tong, F. (2007). The timing of

- perceptual decisions for ambiguous face stimuli in the human ventral visual cortex. *Cerebral Cortex*, *17*, 669–678.
- Mooney, C. M. (1957). Age in the development of closure ability in children. *Canadian Journal of Psychology*, *11*, 219–226.
- Moore, C., & Cavanagh, P. (1998). Recovery of 3D volume from 2-tone images of novel objects. *Cognition*, *67*, 45–71.
- Näasänen, R. (1999). Spatial frequency bandwidth used in the recognition of facial images. *Vision Research*, *39*, 3824–3833.
- Norcia, A. M., & Tyler, C. W. (1985). Spatial frequency sweep VEP: Visual acuity during the first year of life. *Vision Research*, *25*, 1399–1408.
- Norcia, A. M., Tyler, C. W., & Hamer, R. D. (1990). Development of contrast sensitivity in the human infant. *Vision Research*, *30*, 1475–1486.
- Nothdurft, H. C. (1993). Faces and facial expressions do not pop out. *Perception*, *22*(11), 1287–1298.
- Parkin, A. J., & Williamson, P. (1987). Cerebral lateralisation at different stages of facial processing. *Cortex*, *23*, 99–110.
- Philiastides, M. G., & Sajda, P. (2007). EEG-informed fMRI reveals spatiotemporal characteristics of perceptual decision making. *Journal of Neuroscience*, *27*, 13082–13091.
- Prieto, E. A., Caharel, S., Henson, R., & Rossion, B. (2011). Early (N170/M170) face-sensitivity despite right lateral occipital brain damage in acquired prosopagnosia. *Frontiers in Human Neuroscience*, *5*, 138. doi:10.3389/fnhum.2011.00138.
- Puce, A., Allison, T., Gore, J. C., & McCarthy, G. (1995). Face-sensitive regions in human extrastriate cortex by functional MRI. *Journal of Neurophysiology*, *74*, 1192–1199.
- Purcell, D. G., & Stewart, A. L. (1986). The face-detection effect. *Bulletin of the Psychonomic Society*, *24*, 118–120.
- Purcell, D. G., & Stewart, A. L. (1988). The face-detection effect: Configuration enhances detection. *Perception & Psychophysics*, *43*, 355–366.
- Rainer, G., Augath, M., Trinath, T., & Logothetis, N. K. 2001. Nonmonotonic noise tuning of BOLD fMRI signal to natural images in the visual cortex of the anesthetized monkey. *Current Biology*, *11*, 846–854.
- Regan, D. (1966). Some characteristics of average steady state and transient responses evoked by modulated light. *Electroencephalography and Clinical Neurophysiology*, *20*, 238–248.
- Regan, D. (1973). Rapid objective refraction using evoked brain potentials. *Investigative Ophthalmology*, *12*, 669–679.
- Regan, D. (1977). Steady-state evoked potentials. *Journal of the Optical Society of America*, *67*, 1475–1489.
- Regan, D. (1989). *Human brain electrophysiology: Evoked potentials and evoked magnetic fields in science and medicine*. New York: Elsevier.
- Reinders, A. A. T. S., Cläscher, J., de Jong, J. R., Willemsen, A. T. M., den Boer, J. A., & Büchel, C. (2006). Detecting fearful and neutral faces: BOLD latency differences in amygdala-hippocampal junction. *NeuroImage*, *33*, 805–814.
- Reinders, A. A. T. S., den Boer, J. A., & Büchel, C. (2005). The robustness of perception. *European Journal of Neuroscience*, *22*, 524–530.
- 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., & Caharel, S. (2011). ERP evidence for the speed of face categorization in the human brain: Disentangling the contribution of low-level visual cues from face perception. *Vision Research*, *51*, 1297–1311.
- Rossion, B., Dricot, L., Goebel, R., & Busigny, T. (2011). Holistic face categorization in higher-level cortical visual areas of the normal and prosopagnosic brain: Towards a non-hierarchical view of face perception. *Frontiers in Human Neuroscience*, *4*, 225. doi:10.3389/fnhum.2010.00225.
- 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*, 1959–1979.
- Rossion, B., & Jacques, C. (2011). The N170: Understanding the time-course of face perception in the human brain. In S. Luck & E. Kappenman (Eds.), *The Oxford handbook of ERP components* (pp. 115–142). New York: Oxford University Press.
- 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. (2008a). 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., Mace, M. J., & Fabre-Thorpe, M. (2003). Is it an animal? Is it a human face? Fast processing in upright and inverted natural scenes. *Journal of Vision*, 3(6):5, 440–455, <http://www.journalofvision.org/content/3/6/5>, doi:10.1167/3.6.5. [PubMed] [Article]
- Rousselet, G. A., Pernet, C. R., Bennett, P. J., & Sekuler, A. B. (2008b). Parametric study of EEG sensitivity to phase noise during face processing. *BMC Neuroscience*, 9, 98.
- Sadr, J., & Sinha, P. (2004). Object recognition and random image structure evolution. *Cognitive Science*, 28, 259–287.
- Sergent, J., Ohta, S., & MacDonald, B. (1992). Functional neuroanatomy of face and object processing: A positron emission tomography study. *Brain*, 115, 15–36.
- Tanskanen, T., Näsänen, R., Montez, T., Päällysaho, J., & Hari, R. (2005). Face recognition and cortical responses show similar sensitivity to noise spatial frequency. *Cerebral Cortex*, 15, 526–534.
- Tsao, D. Y., Moeller, S., & Freiwald, W. A. (2008). Comparing face patch systems in macaques and humans. *Proceedings of the National Academy of Science U S A*, 105, 19514–19519.
- Turati, C., Simion, F., Milani, I., & Umiltà, C. (2002). Newborns' preference for faces: What is crucial? *Developmental Psychology*, 38(6), 875–882.
- Tyler, C. W., Apkarian, P., Levi, D. M., & Nakayama, K. (1979). Rapid assessment of visual function: An electronic sweep technique for the pattern visual evoked potential. *Investigative Ophthalmology & Visual Science*, 18(7):703–713, <http://www.iovs.org/content/18/7/703>. [PubMed] [Article]
- Valentine, T., & Bruce, V. (1986). The effects of distinctiveness in recognising and classifying faces. *Perception*, 15(5), 525–535.
- Van Rullen, R. (2006). On second glance: Still no high-level pop-out effect for faces. *Vision Research*, 46, 3017–3027.
- Viola, P., & Jones, M. J. (2004). Robust real-time face detection. *International Journal of Computer Vision*, 57(2), 137–154.
- Weiner, K. S., & Grill-Spector, K. (2010). Sparsely-distributed organization of face and limb activations in human ventral temporal cortex. *NeuroImage*, 52, 1559–1573.
- Yang, M. H., Kriegman, D., & Ahuja, N. (2002). Detecting faces in images: A survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(1), 34–58.