"Fast periodic presentation of natural images reveals a robust face-selective electrophysiological response in the human brain"

Rossion, Bruno ; Torfs, Katrien ; Jacques, Corentin ; Liu-Shuang, J.

ABSTRACT

We designed a fast periodic visual stimulation approach to identify an objective signature of face categorization incorporating both visual discrimination (from nonface objects) and generalization (across widely variable face exemplars). Scalp electroencephalographic (EEG) data were recorded in 12 human observers viewing natural images of objects at a rapid frequency of 5.88 images/s for 60 s. Natural images of faces were interleaved every five stimuli, i.e., at 1.18 Hz (5.88/5). Face categorization was indexed by a high signal-to-noise ratio response, specifically at an oddball face stimulation frequency of 1.18 Hz and its harmonics. This face-selective periodic EEG response was highly significant for every participant, even for a single 60-s sequence, and was generally localized over the right occipitotemporal cortex. The periodicity constraint and the large selection of stimuli ensured that this selective response to natural face images was free of low-level visual confounds, as confirmed by the absence of any oddball response for phase-scrambled stimuli. Without any subtraction procedure, time-domain analysis revealed a sequence of differential face-selective EEG components between 120 and 400 ms after oddball face image onset, progressing from medial occipital (P1-faces) to occipitotemporal (N1-faces) and anterior temporal (P2-faces) regions. Overall, this fast periodic visual stimulation approach provides a direct signature of natural face categorization and opens an avenue for efficiently measuring categorization responses of complex visual stimuli in the human brain.

CITE THIS VERSION

Fast periodic presentation of natural images reveals a robust face-selective electrophysiological response in the human brain

Bruno Rossion
Psychological Sciences Research Institute, Institute of Neuroscience, University of Louvain, Louvain-la-Neuve, Belgium

Katrien Torfs
Psychological Sciences Research Institute, Institute of Neuroscience, University of Louvain, Louvain-la-Neuve, Belgium

Corentin Jacques
Psychological Sciences Research Institute, Institute of Neuroscience, University of Louvain, Louvain-la-Neuve, Belgium

Joan Liu-Shuang
Psychological Sciences Research Institute, Institute of Neuroscience, University of Louvain, Louvain-la-Neuve, Belgium

We designed a fast periodic visual stimulation approach to identify an objective signature of face categorization incorporating both visual discrimination (from nonface objects) and generalization (across widely variable face exemplars). Scalp electroencephalographic (EEG) data were recorded in 12 human observers viewing natural images of objects at a rapid frequency of 5.88 images/s for 60 s. Natural images of faces were interleaved every five stimuli, i.e., at 1.18 Hz (5.88/5). Face categorization was indexed by a high signal-to-noise ratio response, specifically at an oddball face stimulation frequency of 1.18 Hz and its harmonics. This face-selective periodic EEG response was highly significant for every participant, even for a single 60-s sequence, and was generally localized over the right occipitotemporal cortex. The periodicity constraint and the large selection of stimuli ensured that this selective response to natural face images was free of low-level visual confounds, as confirmed by the absence of any oddball response for phase-scrambled stimuli. Without any subtraction procedure, time-domain analysis revealed a sequence of differential face-selective EEG components between 120 and 400 ms after oddball face image onset, progressing from medial occipital (P1-faces) to occipitotemporal (N1-faces) and anterior temporal (P2-faces) regions. Overall, this fast periodic visual stimulation approach provides a direct signature of natural face categorization and opens an avenue for efficiently measuring categorization responses of complex visual stimuli in the human brain.

Introduction

A fundamental function of the human brain is to organize sensory events into distinct classes. Categorization of sensory stimuli requires both discrimination—the ability to provide different responses to stimuli belonging to different categories—and generalization—the ability to provide the same response to different exemplars of the same category. In vision, this dual problem of categorization is well illustrated by arguably the most familiar stimuli in people’s visual environment: human faces. Discriminating faces from nonface stimuli is extremely challenging for the visual system because the natural world contains a wide range of stimuli that share visual properties with faces and that potentially look like faces, including biological categories such as fruits and vegetables or animals. Furthermore, there is a wide range of physical variation between faces encountered in the environment (e.g.,


doi: 10.1167/15.1.18
Received May 27, 2014; published January 16, 2015
ISSN 1534-7362 © 2015 ARVO

Downloaded From: http://jov.arvojournals.org/pdfaccess.ashx?url=/data/Journals/JOV/933689/ on 07/16/2015
differences in head orientation, relative size, gender, ethnic origin, emotional expression, age), making generalization across face exemplars very challenging. For this reason, although face detection is well established in the engineering literature (Hjelmås & Low, 2001; Yang, Kriegman, & Ahuja, 2002; Viola & Jones, 2004), automatic systems that can match human face detection performance remain elusive (Scheirer, Jones, 2004), automatic systems that can match human face detection performance remain elusive (Scheirer, Anthony, & Nakayama, 2014).

Despite the tremendous difficulty of face categorization, behavioral experiments have shown that humans categorize visual stimuli as faces accurately and rapidly (e.g., Lewis & Edmonds, 2003; Rousselet, Mace, & Fabre-Thorpe, 2003; Herschler & Hochstein, 2005; Fletcher-Watson, Findlay, Leekam, & Benson, 2008; Crouzet, Kirchner, & Thorpe, 2010; Crouzet & Thorpe, 2011; Scheirer et al., 2014). In these studies, face categorization is typically assessed by asking participants either to provide a behavioral response to faces only, among various distractors, or to provide one kind of behavioral response to exemplars of faces and another kind of response to exemplars of another visual category (e.g., cars, animals; see, e.g., Crouzet et al., 2010). However, these behavioral studies measure the explicit output of the system, which is a mixture of perceptual, attentional, decisional, and motor processes. Hence, they do not have direct access to the visual categorization process that generates this behavioral output.\(^1\)

Neuroimaging studies have identified a number of regions of the human brain that respond differently to the sudden onset of faces and objects, beyond differences in low-level visual cues, and without the need to behaviorally (i.e., explicitly) categorize face versus nonface stimuli (e.g., Sergent, Ohta, & MacDonald, 1992; Puce, Allison, Gore, & McCarthy, 1995; Kanwisher, McDermott, & Chun, 1997; Haxby, Hoffman, & Gobbini, 2000; Weiner & Grill-Spector, 2010; Rossion, Prieto, Boremanse, Kuefner, & Van Belle, 2012). Recording of electromagnetic signals on the human scalp or inside the human brain has also revealed consistent differences in responses to faces and nonface stimuli, with the advantage of providing a time stamp to these differences (e.g., Bentin, Allison, Puce, Perez, & McCarthy, 1996; Jeffreys, 1989, 1996; Moulson et al., 2011; Cauchoix et al., 2014; see Rossion & Jacques, 2011, and Rossion, 2014a, for reviews; for intracranial recordings, see, e.g., Allison, Puce, Spencer, & McCarthy, 1999; Engell & McCarthy, 2011). However, these approaches have several limitations that prevent them from providing direct signatures of face categorization, and visual categorization in general. First, with these approaches there is no direct identification of a face categorization response: The responses to control categories recorded at a different time must be subtracted out, or regressed out, from the responses evoked by faces. However, the subtracted components may in fact interact with the selective responses evoked by the different categories, so that this subtraction procedure is not transparent (i.e., the pure insertion problem; see Friston et al., 1996; D’Esposito, 2010). Second, these approaches do not provide evidence for generalization across face exemplars: A subset of the face stimuli generating a differential response with nonface stimuli could be sufficient to obtain a significant (faces–objects) discrimination response, even if generalization across the entire set of faces is relatively low. Third, with these approaches, the identification of a discrimination response—on face-selective response—is quite subjective, often requiring a post hoc search across space, time, and frequency bands to define the regions of interest. Finally, these approaches are time consuming: Their low signal-to-noise ratio (SNR) requires the estimation of responses from many stimuli presented at a slow rate, whether these responses are analyzed by averaging across trials, as in most studies, or analyzed trial by trial (e.g., Carlson, Tovar, Alink, & Kriegeskorte, 2013; Cauchoix et al., 2014).

The general goal of the present study is to overcome these limitations and provide an objective, direct (i.e., without subtraction), robust (i.e., high SNR) and automatic (i.e., without explicit instruction to categorize) signature of visual categorization of natural faces in the human brain, incorporating both generalization (across widely variable face exemplars) and discrimination (from nonface objects). In addition, we aimed to identify a face-selective response to natural images that is not accounted for by low-level visual cues, a real challenge for conventional approaches (e.g., Rousselet et al., 2008; Yue, Cassidy, Devaney, Holt, & Tootell, 2011; Carlson et al., 2013; Rice, Watson, Hartley, & Andrews, 2014; see Rossion, 2014a, for a discussion of this issue).

In order to provide a signature of face generalization, the approach relies on periodicity. A large number of widely variable unsegmented faces (Figure 1; full set of stimuli available at http://face-categorization-lab.webnode.com) are presented one by one to the visual system at a fixed rate. To isolate the common response of the system to each face stimulus (i.e., generalization), we extract a periodic neural response that is precisely related to the periodicity of the face input. This response can be defined thanks to three aspects of this original approach. First, the visual system’s precise temporal synchronization to periodic visual inputs is reflected by a brain response at the same frequency as the visual input. Second, we have the ability to capture the system’s response to each face with electroencephalography (EEG), the recording of electrical brain waves from the human scalp (Regan, 1989; Luck, 2012). More precisely, presenting a visual stimulus at a
constant rate elicits a periodic response—a steady-state visual evoked potential (Regan, 1966, 1989)—directly identifiable in the frequency domain of the EEG. Third, the use and analysis of a long stimulation sequence (about 1 min) provides a very high frequency resolution (e.g., a 60-s sequence gives a $1/60 = 0.0166$-Hz frequency resolution; Regan, 1989). In these conditions, the common response to the periodically presented faces can be captured in a tiny EEG frequency band that is minimally affected by broad-band EEG noise (Regan, 1989; Rossion & Boremanse, 2011; Rossion, 2014b).

To ensure that the periodic response to faces is selective to this category, i.e., to simultaneously provide a direct measure of visual discrimination between faces and nonfaces, a fixed number (≥2) of pictures of various unsegmented objects are inserted in between the periodically presented faces (i.e., four nonface stimuli in the current experiment; Figure 1). These nonface stimuli are selected at random, so that there is no object category other than faces that repeats periodically. However, all images are presented at a periodic rate, i.e., at a faster rate than the faces. This rate is the base stimulation frequency ($f$ Hz). With four object categories in between each face, the rate of face stimulation is $f/5$ Hz, or every fifth stimulus (oddball frequency). In these conditions, an EEG response at precisely $f/5$ Hz inherently reflects a differential response (i.e., face-selective) from the response evoked at every cycle ($f$) of stimulation.

Thus, considering these constraints altogether, a consistently different brain response at every $f/5$-Hz cycle in such a stimulation sequence provides a direct signature of both face discrimination and generalization. Here, generalization is mandatory: Categorizing only a subset of the face stimuli as faces would break the $f/5$ periodicity. This signature can be obtained directly (without the need to subtract responses obtained from different categories, the control being inserted in the paradigm) and identified objectively (exactly at the $f/5$-Hz frequency defined by the experimenter).

To be more specific, we designed a paradigm in which natural stimuli are presented for about 1 min at a fast rate of 5.88 Hz (5.88 c/s = six images/s), with faces introduced every fifth stimulus (i.e., 5.88/5 = 1.18 Hz; Figure 1; Movie 1). This type of periodic visual oddball paradigm has been designed in EEG to measure visual discrimination of simple elements such as orientation of gratings (Braddick, Wattam-Bell, & Atkinson, 1986;
Braddick, Birtles, Wattam-Bell, & Atkinson, 2005; Heinrich, Mell, & Bach, 2009) or, more recently, individual faces (Liu-Shuang, Norcia, & Rossion, 2014). However, in these studies the visual stimulus is identical at every cycle (or varies only according to a low-level property such as grating phase or size; see Braddick et al., 1986; Liu-Shuang et al., 2014). The originality of the present paradigm is in presenting entirely different stimuli—natural visual objects from various object categories—at every stimulation cycle (for fast periodic stimulation of different complex stimuli in a different context, i.e., attentional blink paradigms, see A. Keil, Ihssen, & Heim, 2006; Talsma, Doty, Strowd, & Woldorff, 2006).

Besides the potential of this approach to provide a direct face categorization response that incorporates discrimination and generalization, it has many other nonnegligible advantages. The response can be recorded without requiring a behavioral face categorization task (i.e., implicitly), since the processes involved in the orthogonal task do not elicit periodic responses and therefore do not contaminate the response. Moreover, the SNR of this fast periodic visual stimulation (FPVS; Rossion, 2014b) approach is extremely high, making it possible to rapidly obtain significant responses in individual participants (see, e.g., Regan, 1989; Srinivasan, Russell, Edelman, & Tononi, 1999; Rossion & Boremanse, 2011; Liu-Shuang et al., 2014). In addition, although the strength of the present approach is not in providing accurate timing information (i.e., when a given differential response occurs between faces and nonface object categories), the time delay between two face stimuli in our paradigm remains relatively long (i.e., stimulus onset asynchrony of 850 ms at 1.18 Hz). Hence, despite the rapid base stimulation rate, one can potentially extract complementary timing information by averaging the time segments following the onset of a specific visual stimulus (e.g., Appelbaum, Ales, & Norcia, 2012; Ales, Appelbaum, Cottereau, & Norcia, 2013), here a face. Since this response necessarily reflects a differentiation between the object and face stimuli, the time-domain waveforms obtained are differential waveforms rather than event-related potentials (ERPs), directly reflecting face-selective processes (Dzhelyova & Rossion, 2014; Liu-Shuang et al., 2014). Finally, thanks to periodicity, the FPVS oddball approach should remove the potential contribution of low-level visual differences across categories while preserving the naturality of stimuli. This is because low-level visual confounds such as different spatial-frequency amplitude spectra between faces and objects (VanRullen, 2006; M. S. Keil, 2008) can influence the periodic oddball response only if (a) they are systematically present in all or a large majority of stimuli and (b) they are stronger than the low-level differences between any two successive images of nonface objects. To test whether these low-level confounds are removed by the periodicity constraint, we created a condition containing the stimuli presented in the same oddball sequence described, but without any shape information (i.e., phase-scrambled stimuli). We hypothesized that, contrary to periodically presented natural face images, these phase-scrambled stimuli would not generate any face-selective oddball response.

Materials and methods

Participants

Twelve participants (three men, mean age = 22.8 ± 1.3, range: 21–26) were tested in the experiment. All participants gave written informed consent and received financial compensation for their participation in the study, which was approved by the Biomedical Ethical Committee of the University of Louvain and was in conformity with the 2013 WMA Declaration of Helsinki. They were all right-handed and reported normal or corrected-to-normal vision. None reported any history of psychiatric or neurological disorder. Participants were aware neither that in one of the conditions a face was presented at a rate of 1 out of 5
stimuli nor that the phase-scrambled stimuli were made from objects and faces.

**Stimuli**

We collected 200 photographic images of various nonface objects (animals, plants, built objects, and houses) and 50 images of faces from the Internet. All objects and faces were unsegmented, i.e., embedded in their original visual scene. The various objects and faces were centered, but they differed in terms of size, viewpoint, lighting conditions, and background (Figure 1; Movie 1). The entire set of stimuli could not be displayed as a figure or in supplementary material for copyright reasons, but they are available upon request from the authors or online at http://face-categorization-lab.webnode.com. The stimuli were converted to gray scale, resized to 200 × 200 pixels, and equalized in terms of pixel luminance and root-mean-square contrast in Matlab. Importantly, given that this normalization is performed on the whole image, the faces in these natural images still purposely differed substantially in local luminance, contrast, and power spectrum. Scrambled versions of the stimuli were made by replacing the phase of each image by random coefficients. Shown at a distance of 1 m and a resolution of 800 × 600 pixels, the stimuli subtended approximately 5.22° of visual angle.

**Procedure**

Figure 1A shows a schematic illustration of the experimental design. As in work by Liu-Shuang and colleagues (2014) and previous FPVS studies with faces (see Rossion, 2014b, for a review), the stimuli were presented through sinusoidal contrast modulation at a rate of 5.88 Hz using custom Matlab software (Sinstim). Each stimulation cycle lasted 170 ms (i.e., 1000 ms/5.88) and began with a uniform gray background from which an image appeared as its contrast increased following a sinusoidal function. Full contrast was reached at 85 ms and then decreased at the same rate. A rate of 5.88 Hz was used because this frequency leads to a large response over occipitotemporal regions (see Alonso-Prieto, Belle, Liu-Shuang, Norcia, & Rossion, 2013) and falls in an area of the EEG spectrum (theta band) where the noise level is low (i.e., above the EEG delta band but below the alpha band of 8–12 Hz). The periodic oddball sequence was composed of four objects (O) followed by a face (F), all randomly selected from their respective categories. Thus, the oddball faces were presented at a frequency of 5.88/5 = 1.18 Hz. As a result, EEG amplitude at the precise oddball face stimulation frequency and its
harmonics (2.35 Hz, 3.53 Hz, etc.) was used as an index of discrimination between faces and objects, and of generalization across faces. The 1/5 ratio was selected based on pilot tests and on a previous experiment (Liu-Shuang et al., 2014). This ratio is a good compromise between a face stimulation rate that is too high, preventing isolation of the responses to each face in the time domain, and one that is too low, leading to smaller face-selective responses.

The experiment consisted of three conditions: Periodic, Nonperiodic, and Scrambled Periodic. The Periodic condition is the periodic oddball sequence described earlier. In the Nonperiodic condition, the exact same stimuli were shown but in an entirely random order. This condition was included in the study in order to provide a dissociation between the rate responses—which should not differ between the Periodic and Nonperiodic conditions—and the oddball responses—which should be present only in the Periodic condition. Moreover, the Nonperiodic condition could be used as a comparison waveform in the time domain to identify the oddball face-selective components. In the Scrambled Periodic condition, phase-scrambled versions of objects and faces were presented in the periodic oddball sequence. Each condition was repeated four times in pseudorandom order for each participant. Stimuli were repeated randomly (but not consecutively) within each trial and across repetitions. At the start of each trial, a fixation cross was displayed against the blank background for 2–5 s (duration randomly jittered between trials) in order to stabilize participants’ fixation. The stimulation sequence lasted 60 s and was flanked by 2 s of fade-in and fade-out at the beginning and the end of the sequence, respectively. During the fade-in, the contrast modulation depth of the periodic stimulation progressively increased from 0% to 100% (full contrast), while the opposite manipulation was applied during the fade-out. This fading aimed at reducing blinks and abrupt eye movements due to the sudden appearance or disappearance of flickering stimuli. The total duration of the experiment was approximately 13 min (12 trials × 64 s), not including breaks. As mentioned in the Introduction, a long trial duration was used in order to obtain a high frequency resolution for the periodic response. Since this high frequency resolution allows for isolation of the majority of the response of interest (i.e., the signal) into a discrete frequency bin, while the EEG noise is distributed throughout many frequency bins in the spectrum, the signal-to-noise ratio of the response is very high (Regan, 1989; Rossion et al., 2012).

During EEG recording, participants were seated comfortably at a distance of 1 m from the computer screen and were instructed to fixate on a small black cross located in the center of the stimuli while continuously monitoring the flickering stimuli. Their task was to detect brief (500 ms) color changes (black to red) of this fixation cross. Color changes occurred at random times, 10 times within every trial. This task was orthogonal to the manipulation of interest in the study and was used to ensure that the participants maintained a constant level of attention throughout the experiment.

**EEG acquisition**

EEG was acquired using a 128-channel BioSemi Active 2 system (BioSemi, Amsterdam, The Netherlands), with electrodes including standard 10–20 system locations as well as additional intermediate positions (http://www.biosemi.com, relabeled according to more conventional labels; see Supplementary Figure S1). EEG was sampled at 512 Hz, and acquisition took place in a dimly lit and sound-attenuated room. Electrode offset was reduced between ±20 μV for each individual electrode by softly abrading the scalp underneath with a blunt plastic needle and injecting the electrode with saline gel. Eye movements were monitored using four electrodes placed at the outer canthi of the eyes and above and below the right orbit. During the experiment, triggers were sent via parallel port from the stimulation computer to the EEG recording computer at the start of each trial and at the minima of each 5.88-Hz stimulation cycle (gray background, 0% contrast). Recordings were manually initiated when participants showed an artifact-free EEG signal and included 10 s of resting-state EEG before stimulation.

**EEG analysis**

**Preprocessing**

All EEG analyses were carried out using Letswave 5 (http://nocions. webnode.com/letswave) running on Matlab 2012 (MathWorks, Natick, MA). EEG data were first band-pass filtered at a low cutoff of 0.1 Hz and a high cutoff of 100 Hz using a fast Fourier transform (FFT) band-pass filter with a Hanning window achieving full attenuation in a 2-Hz window. Subsequently, EEG was downsampled to 250 Hz to reduce file size and data processing time. EEG was then segmented in 68-s segments, keeping 2 s before and after each trial (~2–66 s). Noisy and artifact-ridden channels containing deflections larger than 100 μV in at least two trials were rebuilt using linear interpolation from immediately adjacent clean channels (no more than 5% of channels were interpolated per participant), and a common average reference computation was applied to all channels excluding ocular channels.
**Frequency-domain analysis**

Preprocessed data segments were further cropped down to an integer number of 1.18-Hz cycles beginning 2 s after onset of the trial (right at the end of the fade-in period, to avoid any contamination by the fade-in and initial transient responses) until approximately 60 s, before stimulus fade-out (68 cycles, 14,458 time bins in total = 58 s). Trials were averaged in the time domain, separately for each condition and each individual participant. Averaging was first done in the time domain to increase the signal-to-noise ratio by reducing EEG activity that was not phase-locked to the stimulation. Subsequently, a discrete Fourier transform (as implemented in Matlab with the algorithm of Frigo & Johnson, 1998) was applied to these averaged segments, and amplitude spectra were extracted for all channels. Thanks to the long time window, frequency analysis yielded spectra with a high frequency resolution of 0.0173 Hz, thus improving signal-to-noise ratio and allowing unambiguous identification of the response at the frequencies of interest (i.e., 5.88 Hz, 1.18 Hz, and their harmonics).

SNR was computed to take into account the variations of noise across the EEG spectrum. It was calculated as the ratio between the amplitude at each frequency and the average amplitude of the 20 surrounding frequency bins (10 on each side, excluding the immediately adjacent bin; e.g., Rossion et al., 2012). Additionally, z-scores were calculated in the same way (i.e., the difference between amplitude at the frequency of interest and mean amplitude of 20 surrounding bins divided by the standard deviation of the 20 surrounding bins). For the group analysis, individual SNR spectra were averaged within each condition. Group z-scores were calculated by averaging individual amplitude spectra, then computing z-scores on the resulting grand-averaged spectrum.

To analyze the responses at the base and oddball frequencies, we first determined a range of relevant harmonics for each frequency based on the group-level data. To do so, grand-averaged amplitudes were pooled across all channels and z-scores were calculated. The number of harmonic frequency responses was constrained by the highest harmonic response that was significant in at least one condition (z-score > 3.29, p < 0.001). Electrodes of interest were selected based on those that showed the largest response on group-averaged data across the range of harmonics defined earlier. For analyses focusing on face-related responses at individual harmonics, the fifth, 10th, and 15th harmonics of the 1.18-Hz oddball frequency (5.88, 11.76, and 17.64 Hz) were excluded, since these corresponded to the harmonics of the 5.88-Hz base frequency. Statistical comparisons between conditions was done using repeated-measures ANOVAs, and Greenhouse–Geisser corrections were applied to the degrees of freedom whenever the assumption of sphericity was violated.

**Time-domain analysis**

The periodic oddball responses were also examined in the time domain. Raw continuous EEG data were first band-pass filtered (0.1–100 Hz) and downsampled to 250 Hz to reduce file size. Each trial was then segmented with an extra 2 s before and after the stimulation sequence (−2–66 s). An independent component analysis (Jung et al., 2000) with a square mixing matrix was applied in order to remove blink artifacts. Only a single component was removed for each participant, chosen based on visual inspection of the waveform and its topography. Further artifact correction was carried out by interpolating noisy channels, and the data were rereferenced to the average of all channels. EEG data were then resegmented from 0 to 64 s. Data were then further low-pass filtered with a cutoff of 30 Hz (zero-phase Butterworth filter with a slope of 24 dB/octet). In a complementary analysis, a multinotch narrowband FFT filter (0.1-Hz width) was applied to selectively remove the contribution of the base stimulation frequency and its first five harmonics (5.88–29.39 Hz) from the time-domain waveforms. The filtered data were cropped into smaller epochs in two ways: (a) nonoverlapping epochs of 2551 ms (15 × 170-ms cycles), containing three FOOOO sequences, in order to illustrate the identical pattern of periodic components time-locked to each face; and (b) epochs of 1191 ms (7 × 170-ms cycles) overlapping by 340 ms, including the responses to a sequence of two object stimuli, one face stimulus, and four object stimuli (OOFOOOO), in order to obtain one epoch for each face stimulus presented. A baseline correction was applied to these smaller epochs by subtracting the mean response amplitude during the 340 ms (corresponding to the response to the two object images) preceding the face image from the waveform. The first and last few epochs in each trial were discarded, since these intervals contained the fade-in and fade-out periods, as were the initial event-related evoked potentials. The final 92 and 276 epochs, respectively, were averaged for each participant and then grand-averaged per condition for display of time-domain data.

**Statistical analysis**

We ran repeated-measures ANOVAs to compare conditions in the behavioral and EEG data. Mauchly's test for sphericity was used, and Greenhouse–Geisser correction for degrees of freedom was applied whenever sphericity was violated. For significant effects, post hoc pairwise t tests were conducted to examine differences between conditions.
13
12
11
9
8
7
6
4
3

Harmonic
1/5 = 1.18 Hz
2/5 = 2.35 Hz
3/5 = 3.53 Hz
4/5 = 4.70 Hz
6/5 = 7.05 Hz
7/5 = 8.23 Hz
8/5 = 9.41 Hz
9/5 = 10.58 Hz
11/5 = 12.93 Hz
12/5 = 14.11 Hz
13/5 = 15.28 Hz
14/5 = 16.46 Hz

Table 1. Group-level z-scores of responses at 1.18-Hz harmonics (z-scores based on the average of all channels). Notes: Numbers in italics indicate nonsignificant responses.

Results

Behavioral data

Accuracy rates at the fixation-cross task were near ceiling in all conditions (Periodic: 96% ± 4.2%; Nonperiodic: 97% ± 3.5%; Scrambled Periodic: 97% ± 2.7%). and response times were rapid in all conditions (Periodic: 451.8 ± 46.8 ms; Nonperiodic: 474.71 ± 50.37 ms; Scrambled Periodic: 435.02 ± 42.36 ms). There were no significant differences between conditions in accuracy rates, $F(2, 22) = 1.06, p = 0.36$. There were significant differences in correct response time, $F(2, 22) = 8.65, p < 0.002, \eta = 0.44$, due to faster correct response times in the Scrambled Periodic condition compared to the Nonperiodic condition, $t(11) = -4.81, p < 0.002$. Differences between the Periodic condition and the Nonperiodic, $t(11) = 1.90, p = 0.07$, and Scrambled Periodic conditions, $t(11) = 2.01, p = 0.08$, did not reach significance.

EEG data

Although the main interest of the study is the response at the oddball (i.e., face) stimulation rate, we first report the large response at the base rate (5.88 Hz) across the three conditions.

Base stimulation frequency (5.88 Hz)

There were large responses at the base stimulation frequency and at its harmonics in all the conditions. They remained significant until the seventh harmonic (41.15 Hz) in at least one condition (see Supplementary Table S1). There was a marked topographical dissociation between conditions at the fundamental frequency (5.88 Hz). As can be seen in Figure 2, in the conditions containing intact original images (Periodic and Nonperiodic), the responses at 5.88 Hz were lateralized over the right occipitotemporal region, peaking around electrode PO8. However, in the Scrambled Periodic condition, which did not contain recognizable shapes, the 5.88-Hz response was distributed over the medial occipital region, peaking on electrode Oz.2 To statistically test this pattern, a 3 × 2 repeated-measures ANOVA was carried out on the 5.88-Hz response amplitude, with Condition (Periodic, Nonperiodic, Scrambled Periodic) and Channel (Oz, PO8) as within-subject factors. There were no main effects of Condition, $F(2, 22) = 2.14, p = 0.14, \eta^2 = 0.16$, or Channel, $F(1, 11) = 0.003, p = 0.96, \eta^2 = 0$, but a significant Condition × Channel interaction, $F(2, 14.84) = 29.84, p < 0.0001, \eta^2 = 0.73$. This interaction was due to a larger response at Oz for the Scrambled Periodic condition as compared to the other two conditions with natural images—$t(11) = 6.51, p < 0.001$; $t(11) = 5.20, p < 0.001$—without any difference between the latter two, $t(11) = -0.76, p = 0.23$. At PO8, the response was smaller for the Scrambled Periodic condition than for the two conditions with natural images—$t(11) = -2.64, p < 0.01$; $t(11) = -3.34, p < 0.005$—without any difference between the latter two conditions, $t(11) = 1.15, p = 0.14$.

Oddball discrimination frequency (1.18 Hz)

At the group level and pooling across all channels, we observed robust responses at the oddball frequency and at its harmonics in the Periodic condition with intact images only (Figure 2; Table 1). These responses were significant until the 14th harmonic (16.46 Hz; all $z$-scores > 3.29, $p < 0.001$; $z$-score range: 3.81–16.37). There were no significant oddball responses in this frequency range (1.18–16.46 Hz; Figure 3) in the Nonperiodic and Scrambled Periodic conditions (highest $z$-score = 2.10, second highest = 1.46). Using this range of harmonics ($f/5$–$14f/5$), we found significant oddball responses in each individual participant, even though the number of significant harmonics varied across individuals (range: 4–15; Figure 3). These oddball responses were present not only when averaging all the trials ($N = 4$) but also when considering only one (the first) 60-s trial (Figure 3).

The largest oddball response was observed over the right occipitotemporal region, specifically over channel P10 (at nine out of 14 harmonics) and adjacent channels (PO8, PO10; Figure 4). This was true not only for the group data but also for individual participants.
Figure 3. Significant oddball frequency harmonics in the Periodic condition in individual participants and group-averaged data pooled across all channels, as assessed via z-scores (i.e., population mean and variance estimated from 20 neighboring frequency bins, see Materials and methods). Significance thresholds are color-coded according to the legend on the right. (A) Results based on the average of four trials (4 \times 60 \text{s}). (B) Results based on the first 60-s trial. Significant oddball responses are present even for a single trial, for each individual participant. Participants S04, S09, and S12 are the male participants in the sample.

Figure 4. Signal-to-noise ratio (SNR) spectrum (grand average) of the right occipitotemporal channel P10 in the Periodic condition, depicting the range of oddball frequency harmonics at the group level. The topographical map of each harmonic is shown below, with the color scale adapted for each harmonic between 1 and the value displayed at the top of each map (see color bar on the right). Overall, there is a clear occipitotemporal distribution of the response, with a right lateralization for most of the harmonics.
(Figure 5), with the exception of three individuals who showed the largest responses on homologous channels on the left hemisphere (P7, P9, TP7) and one individual whose largest response was over the medial parietal region (Pz). The magnitude of oddball responses decreased as harmonic frequency increased, with the first three harmonics being consistently the largest across all participants.

Time-domain analysis

Since the oddball response reflects a differentiation between the object and face stimuli, the time-domain analysis provides differential time-domain waveforms and components, which directly reflect a (face-)selective process (see Dzhelyova & Rossion, 2014; Liu-Shuang et al., 2014). These EEG waveforms, time-locked to the periodic face stimuli, reveal several distinctive components underlying the oddball response (Figures 6 and 7): a positive component peaking at approximately 160 ms centered on medial and lateral occipital channels (“P1-faces”), followed by a large negative component situated on bilateral occipitotemporal regions and peaking on PO8 at a 220-ms latency (“N1-faces”) and finally by a large positive component with a more anterior and ventral distribution that peaks on P10 at around 410 ms (“P2-faces”). The spatiotemporal differences between these components suggest that they reflect different aspects of the face categorization process.

Discussion

We identified a signature of visual categorization of natural face images in the human brain which incorporates both visual discrimination (from various kinds of objects) and generalization (across widely variable face exemplars), thanks to a fast periodic visual stimulation (FPVS) approach in EEG. This response is obtained without requiring an explicit categorization of the faces, and without participants even noticing the periodicity of the face stimuli in the sequence (i.e., they were all asked and were unable to tell whether there was a difference between the Periodic and Nonperiodic sequences). This response is identified objectively in the EEG spectrum because it occurs exactly at the frequency of stimulation defined by the experimenter, and at exact multiples of this frequency. It is particularly robust, emerging significantly above noise following only a single 60-s stimulation sequence and in every individual tested in the experiment.

A high-level face-selective response to natural images

Importantly, the face-selective periodic response is not accounted for, even partially, by low-level visual cues: There is no face categorization response whatsoever when the exact same stimuli are phase-scrambled. This is important because, on average, natural pictures of faces and nonface objects differ in terms of low-level visual cues, in particular the slope of the power spectrum, which is steeper at low spatial frequencies for built objects (e.g., VanRullen, 2006; M. S. Keil, 2008). This difference partly accounts for rapid behavioral face categorization performance (Cerf, Harel, Einha, & Koch, 2008; Honey, Kirchner, & VanRullen, 2008; Crouzet & Thorpe, 2011) and differences between faces and nonface stimuli in neuroimaging (Andrews, Clarke, Pell, & Hartley, 2010; Yue et al., 2011; Rossion et al., 2012) and standard ERP experiments (Rousselet et al., 2008; Rossion & Caharel, 2011). Here, in order to contribute even minimally to the periodic face categorization response, low-level visual cues would have to vary
systematically (i.e., periodically) in the same direction at the frequency of interest (i.e., at $f/5$). For instance, every face in the set should have a higher contrast or more power in low spatial frequencies than all of the nonface objects used in the sequence. Given the large variability in the stimulus set used here, this is unlikely. Thus, insofar as one uses a large set of stimuli that are highly variable in terms of low-level visual properties (e.g., power spectrum, local luminance and contrast), the constraint of periodicity offers an elegant way of controlling for low-level visual differences between categories while preserving the naturality of the stimuli. This is yet another strength of the approach used here, because the strict equalization of low-level visual properties between face and object stimuli usually degrades the stimulus quality (i.e., methodological control by elimination; e.g., Rousselet et al., 2008; see Rossion, 2014a, for a discussion of this issue).
Nevertheless, in this fast periodic oddball study, there is at least one low-level visual cue that was not present in the stimuli and may be relevant for face detection: color. Grayscale rather than full-color images of faces and objects were used for the sake of simplicity, in particular for phase-scrambling the images. However, color is a highly diagnostic cue for face detection in artificial systems (e.g., de Dios, 2007) and there is behavioral evidence that colored faces are detected faster in colored scenes than grayscale faces in grayscale scenes (Lewis & Edmonds, 2003; Bindemann & Burton, 2009). In the large set of images used here, there are nonface objects of similar color to faces, and the faces also vary naturally in color. However, under natural conditions, faces are never, for example, green or blue, and skin colors tend to differ in their intensity rather than their chromaticity (Graf et al., 1995, 1996; Bindemann & Burton, 2009). Hence, future investigations with this approach could use color stimuli and test whether the presence of diagnostic color enhances the electrophysiological face-selective response.

A complex spatiotemporal gradient of face-selective responses in the human brain

Our observations indicate that a complex nonlinear brain response subtends the high-level categorization of faces among natural and built objects. That is, we recorded a multiform response in the time domain of several hundreds of milliseconds in duration, which incorporates more than one deflection over time (Figure 7). Hence, this response is projected onto many harmonics in the frequency domain of the EEG spectrum by the Fourier decomposition. Interestingly, this response appears to vary in spatial location over time: It is initially widespread over the medial and lateral occipital cortices, then it becomes more focused over the (right) lateral occipital cortex and progresses more anteriorly and ventrally over the (right) occipitotemporal cortex (Figure 7).

Fully characterizing and understanding this complex response in space and time is not the main goal of this study, because the approach introduced here has other strengths than providing precise information about the timing of face-selective processes. Moreover, the localization of brain processes from scalp EEG (or magnetoencephalography [MEG], for that matter) is seriously limited. Nevertheless, to our knowledge, such a complex spatiotemporal face-selective response has never been described before. Rather, standard ERP studies of face and object categorization—i.e., studies that compare the change of EEG activity to transient presentations of faces and nonface stimuli—have consistently identified a single ERP component that arises during the 130–200-ms time window and shows a larger response to faces: the N170 (Bentin et al., 1996; see Rossion & Jacques, 2011, for a review) and its positive counterpart on the vertex, the vertex positive potential (Jeffreys, 1996; Joyce & Rossion, 2005). Inconsistent differences between faces and objects on earlier components such as the P1 in EEG (Eimer, 1998; Itier & Taylor, 2004; Rossion & Caharel, 2011) or the M1 in MEG (e.g., Halgren, Raji, Marinkovic, Josmäki, & Hari, 2000; Liu, Harris, & Kanwisher, 2002; Okazaki, Abrahamyan, Stevens, & Ioannides, 2008) have been attributed to low-level visual cues such as amplitude spectrum or color (Halgren et al., 2000; Rossion & Caharel, 2011). Post-N170 components have also been observed, such as the N250, a component which is triggered essentially by familiar faces (e.g., Tanaka, Curran, Porterfield, & Collins, 2006; Kaufmann, Schweinberger, & Burton, 2009; Gosling & Eimer, 2011) and which is enhanced by the repetition of familiar faces more than unfamiliar faces (N250r; Schweinberger, Pickering, Jentzsch, Burton, & Kaufmann, 2002; Schweinberger, Huddy, & Burton, 2004). However, this N250r response is dependent on immediate exemplar repetition or long-term face representations, and its specificity to faces is not well established (although see Schweinberger et al., 2004; Nasr & Esteky, 2009). Moreover, in transient ERP studies, later components such as the N250r often overlap with eye movements or decisional or motor processes, making it difficult to isolate face-selective responses at these latencies (Rossion, 2014a). Here, not only did such decisional and motor processes occur very rarely during a stimulation sequence (i.e., when participants detected a change of color of the fixation cross), but their contribution is inevitably eliminated by the analysis, which reveals only the electrophysiological correlate processes that are time-locked to the periodic face presentation. Hence, we were able to disclose multiple novel face-selective responses on the human scalp that we tentatively labeled P1-faces, N1-faces, and P2-faces.

In terms of scalp topography, the overall face categorization response identified here both in the frequency domain and in the time domain bears some striking resemblance to the N170 (Bentin et al., 1996; Rossion & Jacques, 2011) and to face-related EEG periodic responses as reported in recent studies (Rossion & Boremanse, 2011; Ales, Farzin, Rossion, & Norcia, 2012; Rossion, 2014b). This is particularly the case for the second deflection, of negative polarity over occipitotemporal sites (N1-faces). However, the relationship between the original face categorization response identified here and the N170 obtained following transient visual stimulation is unclear, for several reasons. First, the face categorization response recorded here is an inherent contrast response between faces and other visual stimuli (i.e., a face-selective...
response), not an ERP component to the sudden onset of a face from a uniform background, as in standard ERP studies. The common periodic response to faces and objects does not appear at 1.18 Hz or its harmonics in the frequency domain, or as a deviation from the 5.88-Hz oscillation time-locked to the face stimuli in the time domain. Second, the face stimuli are not presented abruptly but are gradually revealed through sinusoidal contrast modulation. Hence the exact onset of the face categorization response recorded here is probably underestimated. However, considering that a portion of the face categorization response recorded here is sinusoidal contrast modulation. Hence the exact onset of the face categorization response could be shifted forward by 40–60 ms. This estimated onset latency would be around 100–120 ms, with the negative component peaking around 160–180 ms. This earlier latency would be compatible with the onset of the face-sensitive N170, and with the timing of behavioral face categorization in general if low-level visual cues are controlled for (Crouzet & Thorpe, 2011).

The observation of a prolonged face-selective response made of three differential components brings two outstanding issues: First, why does face categorization trigger such a relatively long response, composed of multiple components interlocked in time? Second, where does this response—or these responses—come from? These two issues are somewhat related: Since there are many brain regions responding preferentially to faces and distributed from the lateral occipital cortex to the temporal pole, in particular in the right hemisphere (e.g., Sergent et al., 1992; Allison, McCarthy, Nobre, Puce, & Belger, 1994; Haxby et al., 2000; Weiner & Grill-Spector, 2010; Rossion et al., 2012), it is not surprising to observe a multiform response on the scalp that appears to progress from posterior to anterior ventral regions, with a similar right hemispheric dominance.

One possible account of the spatiotemporal signature of this response is that the perception of the stimulus as a face is based on the earlier part of the response—before 200 ms after stimulus onset (Rossion, 2014a)—while the prolongation of the response in the form of a wide positive component peaking at about 400 ms reflects a face-specific periodic increase in attention (e.g., Hajcak, MacNamara, Foti, Feri, & Keil, 2013) triggered by the high saliency of face stimuli (Hershler & Hochstein, 2005; Crouzet et al., 2010). Alternatively, the early face categorization response could be based on face-selective viewpoint-dependent representations, with face-selective viewpoint-invariant representations emerging gradually as information spreads more anteriorly to temporal regions (Pourtois, Schwartz, Seghier, Lazeyras, & Vuilleumier, 2005; Axelrod & Yovel, 2012; see also Booth & Rolls, 1998; Freiwald & Tsao, 2010; Eifuku et al., 2011). Yet another possibility is that the later component reflects the memory encoding of visual representations in face-selective anterior regions of the temporal lobe (Sergent et al., 1992; Nakamura et al., 1994; Rajimehr, Young, & Tootell, 2009; Gainotti & Marra, 2011; Avidan et al., 2013). This latter hypothesis is compatible with the presence of a face-specific late potential (AP350) over the ventral anterior temporal lobe reported with intracranial recordings that follows an earlier face-specific N170/N200 component originating from the occipitotemporal cortex (Allison et al., 1994; Allison et al., 1999; Rosburg et al., 2010). This AP350 component is sensitive to face repetition, suggesting that it is related to the encoding of individual faces in memory (Allison et al., 1999). Future scalp EEG studies could test these hypotheses with the current fast periodic oddball paradigm by manipulating the task (e.g., asking participants to pay attention to the faces, or to encode as many faces as possible in memory for subsequent recognition) or the viewpoint variation of face stimuli and assessing which subcomponents of the complex face-selective response are affected. Nevertheless, a more direct way to address this question and understand the neural basis of this response would be to record intracerebral local field potentials in humans during this paradigm in order to disentangle the functional contribution of the different regions of the ventral occipitotemporal cortex to visual face categorization.

Conclusions and perspectives on visual (face) categorization: The power of a fast periodic visual input

In the present study, we took advantage of the human brain’s precise synchronization to periodic input in order to rapidly and objectively define a high-level spatiotemporal signature of face-selective processes that generalizes across a wide variety of natural face exemplars. Our observations indicate that the selective face categorization process is a prolonged response composed of at least three components—labeled P1-faces, N1-faces, and P2-faces—overlapping in space and time, these components being largest over the right-hemisphere occipitotemporal regions. Nevertheless, the strengths of the approach to indexing face categorization are best appreciated when considering the frequency-domain representation. Thanks to the high frequency resolution provided by the long stimulation sequence and the temporal precision offered by EEG, a high-SNR face-selective categorization response is obtained exactly at the frequency determined by the experimenter, providing objectivity to the approach.

The approach presented here is only constrained by the strict periodicity of the visual input. This could be a
possible weakness, because periodicity means predict-
ability and a response that could potentially be
contaminated by the observer’s expectations (e.g.,
Summerfield, Egner, Mangels, & Hirsch, 2006). How-
ever, the fast rate of the input and the large range of
variability in the stimuli limit the potential impact of
top-down expectation. As mentioned earlier, partici-
pants tested with this paradigm notice the presence of
faces in the stimulation sequence but are unable to tell
that faces are presented periodically, let alone to report
their periodicity. Moreover, for the visual system, the
fast periodic input is certainly not less natural than a
slow transient (i.e., abrupt) stimulation, as typically
used in standard EEG studies. In fact, a fast periodic
input may well fit with a natural periodic sampling rate
of information in the visual system (VanRullen &
Koch, 2003; VanRullen, Zoefel, & Ilhan, 2014), even
though the exact rate of this optimal sampling may
depend on the kind of visual information that has to be
processed (Alonso-Prieto et al., 2013).

Finally, and although this was not a goal of the
present study, the face-selective response could potential-
ly be quantified at the individual level as the sum of
the EEG amplitude spread over the different oddball
frequency harmonics (e.g., Appelbaum, Wade, Vildav-
ski, Pettet, & Norcia, 2006) and correlated with
behavioral measures of face detection obtained inde-
pendently. The approach could also be implemented as
a sweep visual evoked potential paradigm (Regan,
1973). A sequence would start with fully phase-
scrambled stimuli that parametrically decrease in the
scrambling of the image power spectra. Category-
selective neural face detection thresholds could then be
measured (Ales et al., 2012). This approach could
therefore be used in future studies to understand the
nature and neural basis of face-selective processes in
human adults, but also as a diagnostic tool for
characterizing face categorization during develop-
ment—in infants and young children who cannot
provide overt behavioral responses of face categoriza-
tion—and in clinical populations experiencing difficul-
ties with face and object categorization (i.e., visual
agnosia, acquired and congenital prosopagnosia). More
generally, the fast periodic oddball visual
stimulation approach could be relatively easily ex-
tended beyond face perception to understand visual
and semantic categorization in the human brain.

Keywords: face perception, eeg, ssvep, natural images,
periodicity

Acknowledgments

This work was supported by an ERC grant (face-
ssvep 284025). The authors thank Talia Retter for
collecting the stimuli and also for providing useful
comments on a previous version of this paper, together
with Greg Appelbaum and an anonymous reviewer. BR
and JLS are supported by the Belgian National
Foundation for Scientific Research (FNRS). CJ is
supported by the Belgian Federal Science Policy Office
(BELSPO).

Commercial relationships: none.
Corresponding author: Bruno Rossion.
Email: bruno.rossion@uclouvain.be.
Address: Psychological Sciences Research Institute,
University of Louvain, Louvain-la-Neuve, Belgium.

Footnotes

1 For a continuous behavioral response measure
indexing face processing before the end of the process,
see Quek and Finkbeiner (2013).
2 At the second harmonic (11.76 Hz), the scalp
topographies were essentially identical in all the
conditions, spreading over all posterior sites with a
focus on medial occipital channels (i.e., Oz). Beyond
the second harmonic, the spatial distribution of
activation was also focused on medial occipital
channels, similarly across conditions, and these
responses were not analyzed further.

References

Ales, J. M., Appelbaum, L. G., Cottereau, B., &
Norcia, A. M. (2013). The time course of shape
discrimination in the human brain. NeuroImage,

Ales, J. M., Farzin, F., Rossion, B., & Norcia, A. M.
(2012). An objective method for measuring face
detection thresholds using the sweep steady-state
18, 1–18, http://www.journalofvision.org/content/
12/10/18, doi:10.1167/12.10.18. [PubMed] [Article]

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.

Allison, T., Puce, A., Spencer, D. D., & McCarthy, G.
(1999). Electrophysiological studies of human face
perception. I: Potentials generated in occipitotem-
poral cortex by face and non-face stimuli. Cerebral
Cortex, 9, 415–430.

Alonso-Prieto, E., Belle, G., Liu-Shuang, J., Norcia, A.
M., & Rossion, B. (2013). The 6 Hz fundamental


on Acoustics, Speech and Signal Processing, 3, 1381–1384.


temporal areas related to face and scene processing: A PET study. Brain, 123, 1903–1912.


