Uncovering the neural magnitude and spatio-temporal dynamics of natural image categorization in a fast visual stream

Talia L. Retter[1,2] & Bruno Rossion[1]

1 Psychological Sciences Research Institute, Institute of Neuroscience, University of Louvain, Belgium

2 Department of Psychology, Center for Integrative Neuroscience, University of Nevada, Reno, USA

Corresponding author:

Talia Retter

1664 N Virginia St.

Mail Stop 296

Reno, Nevada 89557

+1 (775) 303-3472

tlretter@nevada.unr.edu

Brief running title: Image categorization magnitude and dynamics

Keywords: electroencephalography; face; humans; frequency tagging; visual evoked potentials

Abstract

Perceptual categorization occurs rapidly under natural viewing conditions. Yet, the neural spatio-temporal dynamics of category-selective processes to single-glanced, natural (i.e., unsegmented) images in a rapidly changing presentation stream remain unknown. We presented human observers with natural images of objects at a fast periodic rate of 12.5 Hz, i.e., every 80 ms. Images of faces were inserted every 3, 5, 7, 9, or 11 stimuli, defining stimulus-onset-asynchronies (SOAs) between 240-880 ms, i.e., presentation frequencies (*F*s) between 4.17-1.14 Hz. Robust face-selective responses were objectively identified and quantified at *F* and its harmonics (2*F*, 3*F*, etc.) for every condition in the electroencephalogram (EEG). The summed-harmonic face-selective response was significantly reduced by 25% at the lowest face SOA, i.e. 240 ms between two faces, but remained stable from 400 ms SOA onward. This high-level, right lateralized face-selective response emerged at about 100 ms post-stimulus onset and progressed spatially throughout four successive time-windows (i.e., P1-face, N1-face, P2-face, P3-face) from posterior to anterior occipito-temporal electrode sites. The total duration of a category-selective response to a briefly presented face stimulus in a rapid sequence of objects, was estimated to be 420 ms. Uncovering the neural spatio-temporal dynamics of category-selectivity in a rapid stream of natural images goes well beyond previous evidence obtained from spatially and temporally isolated stimuli, opening an avenue for understanding human vision and its relationship to categorization behavior.

Introduction

Perceptual categorization, the process by which sensory events are differentiated and classified in subgroups, is critical in enabling human interaction with the world. Human faces, which carry ecologically important social information, constitute the most salient class of visual images for understanding perceptual categorization. Indeed, faces can be differentiated from other objects with astounding accuracy and speed (Hershler & Hochstein, 2005; Crouzet et al., 2010; Hershler et al., 2010; Crouzet & Thorpe, 2011; Scheirer et al., 2014). Furthermore, the perception of segmented images of faces is known to elicit a large, widely distributed and partly specific neural response in the human ventral occipito-temporal (VOT) cortex, with a right hemisphere advantage (Sergent et al., 1992; Allison et al., 1994; 1999; Puce et al., 1995; Kanwisher et al., 1997; Weiner & Grill-Spector, 2010; Rossion et al., 2012; Zhen et al., 2015).

Scalp electroencephalography (EEG), or more rarely magnetoencephalography (MEG), defines the speed and temporal dynamics of face-selective responses in the millisecond range at a system-level of organization. Most significantly, an early response peaking at about 170 ms following stimulus onset (i.e., the N170/VPP complex) differs in amplitude in response to faces compared to other object categories (Jeffreys, 1989; Jeffreys & Tukmachi, 1992; Bötzel et al., 1995; Bentin et al., 1996; Eimer, 2000; Halgren et al., 2000; Rossion et al., 2000; Itier & Taylor, 2004; Rousselet et al., 2008; Ganis et al., 2012; for reviews, Rossion & Jacques, 2011; Rossion, 2014a). Contrary to earlier, potentially spurious differences between faces and objects, this N170 selectivity to faces is not accounted for by Fourier amplitude information, which carries global low-level statistical properties of images (Rossion & Caharel, 2011; see also Tanskanen et al., 2005; Rousselet et al., 2008).

However, crucially, despite natural viewing conditions providing us with continually changing streams of information in complex scenes, categorization of faces, and perceptual categorization in general, have almost exclusively been investigated at the behavioral and neural level with images presented in spatial and temporal isolation.

Spatial isolation refers to face and non-face object stimuli being segmented from their natural backgrounds. In the rare use of natural images (Itier & Taylor, 2004; Rousselet et al., 2004; 2007; Hershler & Hochstein, 2005; Hershler et al., 2010; Crouzet et al., 2010; Cauchoix et al., 2014), controlling for low-level statistical properties differing between faces and objects (e.g., Torralba & Oliva, 2003; VanRullen, 2006; Keil, 2008) is particularly challenging, and their contribution to behavioral and neural face-selective responses is difficult to exclude (Itier & Taylor, 2004; VanRullen, 2006; Rousselet et al., 2007; Cerf et al., 2008; Honey et al., 2008; Crouzet & Thorpe, 2011; Cauchoix et al., 2014; but see Hershler & Hochstein, 2006).

Temporal isolation of the stimuli of interest is the norm in behavioral and neural studies of perceptual categorization and refers to the stimuli being presented as unique events separated by a long and often variable stimulus onset asynchrony (SOA). Alternatively, a train of stimuli with a brief SOA is sometimes used in neuroimaging (i.e., a block design), but the response to the individual stimuli are lumped into a global brain response. Moreover, in EEG/MEG studies, unmasked faces and nonface objects are typically presented for long stimulus duration (e.g., Bötzel et al., 1995: 3-4 seconds; Rossion et al., 2000: 500ms; Crouzet et al., 2010: 400 ms; Ganis et al., 2012: 800ms; Carlson et al., 2013: 533ms; Cauchoix et al., 2014: 300-

600ms; Cichy et al., 2014: 500 ms) and SOAs of 1 to 2 seconds at least. Thus, object/face categorization may appear as a prolonged process (Cichy et al., 2014; Mur & Kriegekorte, 2014) merely because of this long and uninterrupted stimulus duration: in reality, while a single glance suffice for categorization of faces (Crouzet et al., 2010), the duration of category-selective processes from this brief encounter with the stimulus, in context, remains completely unknown.

An alternative stimulus presentation mode has been offered by rapid serial visual presentation (RSVP), which has been used with natural stimuli presented in such rapid succession that they are backward- and forward-masked and may be visible for only a single glimpse; this technique has been employed to investigate the contributions of memory and attention processes to behavioral image recognition over time (Potter & Levy, 1969; Potter, 2012; Potter et al., 2014). However, to derive behavioral performance (i.e., detection) from RSVP, a limited number of stimuli are presented in each sequence (i.e., fewer than 20 in the previously cited studies), and the rapidity of within-category stimulus presentation (e.g., as fast as 13 ms per stimulus in Potter et al., 2014) limits the availability of temporal information in response to a stimulus category at a neural system-level.

Here, we provide the first comprehensive report of the magnitude, onset, and duration, or more generally the temporal dynamics, of the differential neural response between natural images of faces and other object categories viewed at a single glance within a rapid visual presentation stream. The approach that we use is termed *Fast Periodic Visual Stimulation* (FPVS; Rossion, 2014b), in which stimuli are presented in a fast periodic stimulation stream (here, at 12.5 Hz, i.e., one stimulus every 80 ms) while EEG is recorded. Similarly to RSVP, natural and highly variable images are forward- and backward-masked and are visible only long enough to be seen in a single fixation. However, since a neural response to a selected image category (i.e., faces) is investigated here rather than an explicit behavioral response, we are able to present long stimulation sequences (2 minute sequences, each containing about 1,500 images, i.e., 200 s/12.5 Hz) and to periodically embed face images within the sequence at a lower rate, for instance every five items (i.e., 400 ms). Thus, we build on a recently introduced FPVS-EEG paradigm to measure high-level perceptual categorization in the human adult (Rossion et al., 2015; Jacques et al., in press; Jonas et al., in press) and infant (de Heering & Rossion, 2015) brain.

Since face stimuli are presented periodically in this paradigm as a proportion of rapidly presented images from various non-face object categories, two distinct response types emerge in the EEG recording: 1) general visual responses synchronized with the base presentation rate of object stimuli (here 12.5 Hz) and 2) face-selective responses, representing differential responses to faces in contrast to non-face objects, present at the slower face stimulation rate (**Figure 1**). In these conditions, note that a response to faces would not emerge were it identical to the response to non-face objects: thus, the response at the face stimulation rate inherently represents the *differential* response to faces, eliminating the need for post-hoc subtraction across conditions (Rossion et al., 2015; Jacques et al., in press; Jonas et al., in press; see also Liu-Shuang et al., 2014). Moreover, the response to faces in this paradigm reflects not only discrimination (since faces are contrasted to numerous object categories, about 250 variable object stimuli in total), but also generalization (i.e. invariance) across face exemplars (about 50 different face stimuli are used, varying in background, identity, expression, size, viewpoint and lighting conditions), and thus truly reflects face categorization (see **Figure 1**).

Additionally, the embedded periodic presentation of a natural image category within the periodic base stimulation stream allows the contribution of low-level image features to the face-selective response to be restricted with minimal artificial stimulus standardization: putative amplitude spectrum differences that may vary across face and non-face images on average, but which do not vary consistently within the face stimulus set, are not present periodically and so are not captured at the face presentation rate. The variance within the face stimulus set is put in competition with the variability of a large number of natural stimuli in the non-face stimulus set: changes of local contrast, luminance and spatial frequency that occur at every stimulation cycle project to the 12.5 Hz base stimulus presentation rate. Finally, low-level visual cues which might vary systematically (i.e., periodically) at the slower face stimulation rate are reduced by the variability within the natural face stimulus set. Thus, electrophysiological activity at the face-stimulation rate reflects high-level face-selective responses that are absent when the amplitude spectrum is preserved, i.e., for periodically presented phase-scrambled face stimuli *vs.* phase-scrambled non-face object stimuli (Rossion et al., 2015; de Heering & Rossion, 2015).

An important advantage of a FPVS-EEG categorization paradigm is that it enables the review of the EEG data in both the frequency and time domains, each providing its unique advantages. The periodicity of the stimulus presentation can be exploited in the EEG frequency domain, which captures periodic responses exactly at the frequency (or frequencies) of stimulation[1]. Such periodic responses, typically referred to as "Steady-State Visual Evoked Potentials" (SSVEPs, Regan, 1966, 1989; Norcia et al., 2015), are known for their objective localization and extremely high

[1] Thanks to periodicity, parallel streams may be frequency-tagged at different presentation rates, for instance in the right and left visual fields, yet another difference from the RSVP approach.

signal-to-noise ratio (SNR) in the frequency domain, and will be utilized here for face-selective response quantification. Moreover, the spatio-temporal dynamics of the face-selective response may be observed in the time domain: given the relatively low stimulation frequencies of face stimuli afforded by their spaced placement within the relatively fast presentation stream, FPVS-EEG is able to provide a rich description of information flow in response to faces in the time-domain (e.g., Dzhelyova & Rossion, 2014; Rossion et al., 2015; Jacques et al., in press).

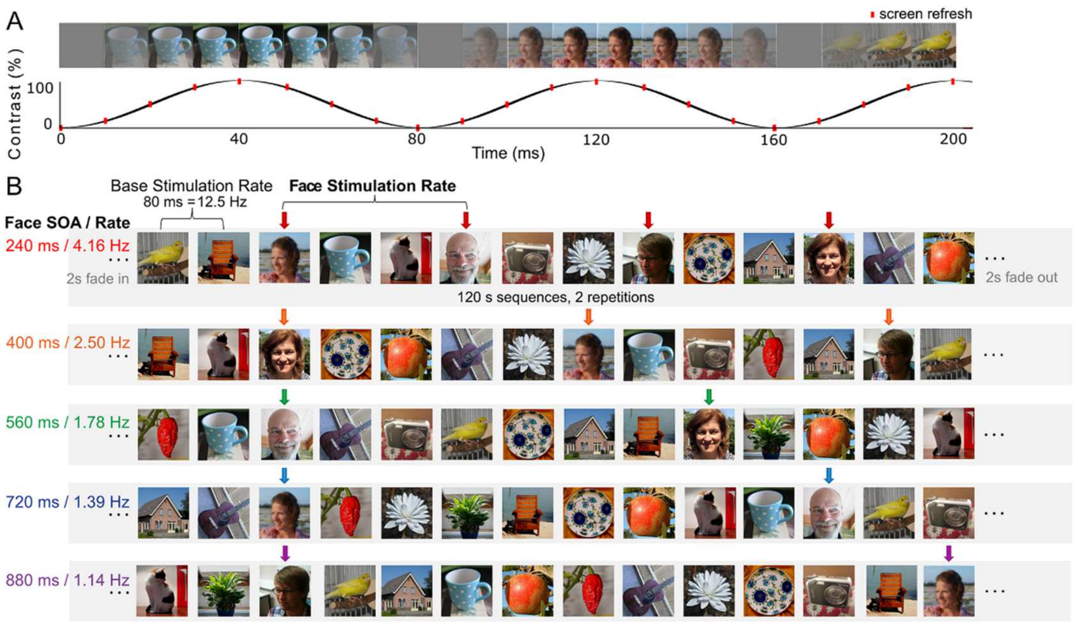

Figure 1. A) *Stimuli are presented at a fast base stimulation frequency (i.e., 80 ms per image = 12.5 Hz) synchronized with the refresh rate of the monitor (i.e., 10 ms per frame = 100 Hz): each image is displayed for eight successive frames. The presentation of each image occurs smoothly through a sinusoidal modulation of luminance contrast, progressing with every frame in steps of 0-15-50-85-100-85-50-15 percent contrast.* ***B)*** *For each condition, face stimuli appear throughout two 120 s sequences as different proportions of base object stimuli, specifically, as every 1/3, 1/5, 1/7, 1/9, or 1/11 stimuli. This defines the five different face stimulus-onset-asynchronies (SOAs)/stimulation rates used in the experiment. For example, in the top*

row, faces appear every 1/3 images, so the face stimulation rate = 12.5 Hz/3 = 4.16 Hz, and the face SOA = 1/4.16 Hz = 240 ms.

In summary, the specific goals of the present study were to exploit the FPVS-EEG paradigm to determine for the first time the magnitude of comprehensive face-categorization responses in a rapid visual stream of non-face objects, as well as to define their exact onset, duration and spatio-temporal pattern. These goals were achieved by 1) modifying the base stimulus presentation rate (i.e., 12.5 Hz here *vs.* 5.88 Hz previously) to severely constrain stimulus duration and segregate in time and space (i.e., scalp topography) face-selective responses from this fast base rate response (Alonso-Prieto et al., 2013); 2) quantifying multi-harmonic face-selective responses at a group and individual level across five manipulations of temporal distance between face stimuli, i.e., face SOAs, in the rapid visual stimulation stream.; and 3) comparing a typically used sinusoidal contrast modulation stimulation mode to an abrupt (i.e., squarewave) stimulation (Experiment 2) in order to determine the exact onset, propagation and temporal dynamics of face-selective responses in a rapid stimulation stream.

Materials and Methods

Experiment 1: Temporal distance between faces

Participants

Sixteen healthy participants (age range 19-25 years, 8 female), from whom no data was rejected, were tested individually in a single EEG recording session for Experiment 1. All participants reported normal or corrected to normal vision and all were right-handed according to an adapted Edinburgh Handedness Inventory measurement (Oldfield, 1971). Participants were recruited from a university campus and received monetary compensation for their time. Signed informed consent was given by all participants before the start of the experiment, which was approved by the Biomedical Ethical Committee of the University of Louvain and the 2013 WMA Declaration of Helsinki.

Stimuli

Stimuli were 294 color images of natural, unsegmented faces (46 images) and objects (248 images), cropped to a square, sized to 200 by 200 pixels, and equalized for mean pixel luminance (at 112/255; in order to standardize the modulation of luminance contrast described in the Procedure). These stimuli are from the same set as used by Rossion and colleagues (2015), although they were presented in greyscale in the previous study (examples of stimuli are available in **Figure 1** and here: http://face-categorization-lab.webnode.com/resources/natural-face-stimuli/). While mean pixel contrast (across-pixel standard deviation) was not equalized across stimuli, this property did not differ on average across face (59.6) and object (56.8) categories ($t(292) = 1.33$, $p > .05$, Cohen's $d = .22$). Images were taken with varying viewpoints, backgrounds, and lighting conditions. Object images consisted of 15 sets of diverse categories of objects (between 5-48 images in category set), including: cats (9), dogs (5), horses (5), birds (24), fruit (29), vegetables (20), flowers (15), house plants (15), telephones (13), chairs (15), cameras (6), dishes (15), guitars (15), lamps (14), and houses (48). Displayed on a monitor with an 800 by 600 pixel resolution from a distance of 1 m, the stimuli subtended approximately 5.2 degrees of visual angle.

Procedure

The experiment was run in a quiet, low-lit room. During testing, the participant was seated in front of a table, on which rested a keyboard and the cathode ray tube (CRT) testing monitor. A curtain isolated the participant from the experimenter; participant behavior was monitored with a camera. Participants viewed two trial repetitions of five conditions, for a total testing time of about 20 minutes.

Each trial consisted of: 1) 2-5 s of a fixation cross on a grey background; 2) 2 s of gradual stimulus fade-in; 3) 120 s testing sequence; 4) 2 s of gradual fade-out; 5) 2 s of a blank gray screen. Fade-in and fade-out were included to reduce abrupt eye movements or blinks due to abrupt stimulation onset or offset, respectively. Within each testing sequence, stimuli were presented at a constant rate of 12.5 images per second (12.5 Hz = 80 ms per image), in synchrony with screen refreshes (at 100 Hz), by means of sinusoidal modulation of contrast from 0-100%, using MATLAB R2009a (MathWorks, USA) with PsychToolbox (**Figure 1A**). A sinusoidal stimulation mode has been typically used in FPVS studies with face stimuli (e.g. Rossion & Boremanse, 2011; Liu-Shuang et al., 2014; Rossion et al., 2015), since the visual stimulation is smoother than with a squarewave (i.e., abrupt onset), and can be described with a single parameter (SOA). Moreover, importantly, with such a sinusoidal contrast stimulation mode, the visual stimulation is virtually continuous, with only one frame (10 ms) by cycle where the contrast is at 0% and a subjective perception of continuous visual stimulation (**Movies 1 & 2**).


Movie1.avi *(http://gofile.me/24HTb/RAZqKhst6)*

Movie 1. *The 240 ms face SOA condition (10 s stimulation extract).*


Movie2.avi *(http://gofile.me/24HTb/k7Ypt1VBV)*

Movie 2. *The 880 ms face SOA condition (10 s stimulation extract). Compare to* ***Movie 1*** *for the range of face SOAs used in the experiment.*

Stimuli from the object categories were used as base stimuli, and face stimuli were interleaved at a fixed temporal distance, i.e., as a fixed proportion of stimuli, within the base sequence. This follows the recent paradigm of Rossion et al. (2015); for visual item-embedded periodic paradigms, see Braddick et al. (1986) as well as Heinrich et al. (2009), and for the first use of such paradigms with high-level visual stimuli, see Liu-Shuang et al. (2014). Here, the temporal distance between face stimuli, i.e., face signal-onset-asynchronies (SOAs), defined five experimental conditions: faces were presented as every 1/3 base stimuli (i.e., at a SOA/rate of 240 ms/4.16 Hz), 1/5 stimuli (400 ms/2.50 Hz), 1/7 stimuli (560 ms/1.78 Hz), 1/9 stimuli (720 ms/1.39 Hz), or 1/11 stimuli (880 ms/1.14 Hz) (**Figure 1B**).

EEG Acquisition

The EEG was recorded using a BioSemi ActiveTwo system with 128 Ag-AgCl Active-electrodes, arranged in the default BioSemi configuration, which centers around nine standard 10/20 locations on the primary axes (BioSemi B.V., Amsterdam, Netherlands; for exact position coordinates, see http://www.biosemi.com/headcap.htm). Electrode labels were changed to closely match a more conventional 10/20 system (for exact relabeling, see Rossion et al., 2015, **Figure S2**). The magnitude of the offset of all electrodes, referenced to the common mode sense (CMS), was held below 50 mV. Vertical and horizontal electrooculogram (EOG) was recorded using four additional flat-type Active-electrodes: two electrodes above and below the participant's right eye and two lateral to the external canthi. The EEG and EOG were digitized at a sampling rate of 512 Hz.

Analysis

Analysis of the recorded EEG was performed using Letswave 5, an open source toolbox (http://nocions.webnode.com/letswave), running over MATLAB R2012b (MathWorks, USA),

Preprocessing

A fourth-order zero-phase Butterworth band-pass filter, with cutoff values of 0.1-120 Hz, was applied to the continuously recorded individual participant data. A Fast Fourier Transform (FFT) multi-notch filter with a width of 0.5 Hz was used to remove electrical noise at three harmonics of 50 Hz. Segmentation was performed, including two seconds before and after each trial. To remove a single component accounting for blink artifacts, independent component analysis (ICA) was applied on the data of two participants who blinked more than 0.2 times/s (mean = 0.10, SD = 0.116) on average during the sequences. Channels which were artifact-prone across multiple trials (less than 1% of channels on average) were interpolated. Finally, all EEG channels were referenced to a common average.

Frequency Domain Analysis

Each trial was re-segmented to contain an integer number of face presentation cycles: the trial was segmented from sequence onset to between 119.5 and 120 s for each condition (this ensures that upon frequency-domain transformation, the frequency resolution will be an integer multiple of the face stimulation frequency, i.e., that a discrete frequency bin will be present at the target face stimulation frequency). Trials were averaged within each condition. A FFT transform was computed to represent the data of each channel as a normalized amplitude spectrum (μV) in the frequency domain (ranging from 0 to 256 Hz). The frequency resolution of such a spectrum is the inverse of the sequence duration, and thus was approximately equal to 0.008 Hz (1/120 s). For group-level display, the grand-averaged amplitude spectra were computed for each channel.

In addition to a response at the face stimulation frequency (*F*), additional face stimulation harmonic frequency responses (i.e., 2*F*, 3*F*, etc.) were expected (e.g., see Rossion et al., 2015). In order to determine the number of significant harmonic responses to include in each condition, all EEG channels of the grand-averaged amplitude spectra were pooled; Z-scores were calculated on this amplitude spectrum for each discrete frequency bin (x) according to the formula Z = (x-baseline mean)/(baseline standard deviation). The baseline was defined as the twenty frequency bins surrounding each target bin excluding the immediately adjacent bins and the local maximum and minimum amplitude bins, i.e., a range of approximately 0.2 Hz around each target bin (Rossion et al., 2012). Continuous face-stimulation frequency harmonics with a Z-score of greater than 2.32 ($p<.01$; 1-tailed, i.e., signal>noise) were included, excluding harmonics which coincided with the base stimulation frequency. Significant base frequency harmonics were defined according to homologous criteria.

Quantification of the face-categorization response was performed across conditions varying in face stimulation frequency (*F* = 4.16 Hz to *F* = 1.14 Hz, for the 240 ms and 880 ms face SOAs, respectively). Since baseline noise levels vary across the frequency spectrum of human EEG data (being generally higher at lower frequencies and locally higher in certain bands, e.g., in the alpha band), a local baseline-subtraction was applied (e.g., as in Mouraux et al., 2011; Dzhelyova & Rossion, 2014; Jacques et al., in press), using the same baseline definition as for the

Z-score calculations. Additionally, harmonic frequency responses also occurred at varying frequencies across conditions: these harmonic responses were found to be distributed within a common frequency range, such that conditions with higher *F*s contained fewer harmonic responses (see Results). Thus, in order to permit a comparison of conditions, significant harmonic response baseline-subtracted amplitudes were summed within each condition (see Heinrich, 2009; Appelbaum et al., 2006; Dzhelyova & Rossion, 2014); this method was also validated by comparison with visualized response amplitudes in the time-domain. For presentation at the group-level, at each channel baseline-subtracted amplitude spectra were grand-averaged.

We performed two variations of this quantification for each condition. First, we averaged across all channels for each condition, essentially reducing the dataset to a single channel. Second, for the main focus of the results reported, we performed a region-of-interest (ROI) analysis for each condition by averaging three channels with the maximal summed-harmonic response across conditions: for the face-categorization responses, we averaged channels PO10, PO8, and P10, defining the *right occipito-temporal region;* a homologous *left occipito-temporal region* was also included, comprised of the average of channels PO9, PO7, and P9; and for the base stimulation responses, we averaged channels POO6, Oz, and O2, defining the *central occipito-parietal region*. Statistical comparisons of conditions were performed separately for face-categorization and base stimulation frequency responses, using repeated measures ANOVAs, with factors of *Region* (right and left occipito-temporal and central occipito-parietal) and *Condition* (240, 400, 560, 720, and 880 ms face SOAs). In the case that Mauchly's test of sphericity was significant, a Greenhouse-Geisser correction was applied to the degrees of freedom.

For analysis at an individual level, performed for the 720 ms/1.39 Hz condition, a significant face-categorization response was identified by summing the raw amplitude of the harmonic frequency responses significant at the group level (including about 0.25 Hz of surrounding baseline noise), pooling all 128 channels, and computing Z-Scores on the resulting spectrum (again with a significance threshold of Z = 2.32, p<.01). The lateralization of the face-categorization response was defined using the magnitudes of the right and left occipito-temporal ROIs (R and L, respectively) as follows: 100*(R-L)/(R+L).

Time Domain Analysis

Data were more conservatively filtered with a fourth-order, zero-phase Butterworth low-pass filter, with a cutoff value of 30 Hz, as typically used in time-domain analyses of ERPs (e.g., Jacques et al., 2007). Each trial was re-segmented so that the frequency resolution would be a multiple of the base stimulation frequency, corresponding to approximately 120 s, before an FFT multi-notch filter with a width of 0.05 Hz was applied to remove the first three harmonics of the base stimulation frequency (Rossion et al., 2015; Jacques et al., in press). Both base-frequency filtered and unfiltered data were included in the subsequent processing steps.

The data were segmented by each face stimulus presentation into overlapping 2 s epochs, including a baseline of two base cycles (160 ms) before and 1,000 ms after each face stimulus presentation; this led to 134-490 epochs per condition, with conditions containing more frequent face stimulus presentations producing more epochs. A complementary analysis over the same number of epochs per condition does not make any difference to the conclusions (**Figure S1**). These epochs were averaged by condition within participants. Since conditions with lower face SOAs did not allow time without face-selective responses, no baseline correction was applied in the comparison of conditions. However, the 720 ms face SOA condition was used as a representative example in separate analyses to examine the time course of the face-selective response, and in this case a baseline subtraction was applied from the time of two base stimulation cycles, i.e., 2/12.49 Hz = 0.16 s, to 0 s before the face stimulus presentation. Finally, individual data for each channel were grand-averaged by condition.

To statistically determine when a deflection in the time domain was significantly different from zero across participants, two-tailed t-tests were run on each bin (512 bins were sampled/s) from 0 to 800 ms after oddball face onset, with a significance threshold of $p<.01$; to correct for multiple comparisons, a nonparametric, percentile cluster permutation test (Maris & Oostenveld, 2007) was applied with 10,000 permutations.

Experiment 2: Sinusoidal Contrast Modulation vs. On/Off Presentation

All materials and methods were identical to that of Experiment 1 expect as stated below.

Participants

Sixteen participants (age range 19-25 years, 10 female) were tested, five of whom also participated in the first experiment.

Procedure

There were two experimental conditions. In both, base stimuli were presented at a rate of 12.5 Hz and face stimuli appeared as every 1/9 base stimuli, i.e., at a SOA/rate of 720 ms/1.39 Hz. However, in the *sinewave* condition, the base stimulation occurred through sinusoidal contrast modulation, exactly as in Experiment 1 (see **Figure 1A**). In the *squarewave* condition, the base stimulation occurred through on/off, i.e., squarewave, contrast modulation with a 50% duty cycle. Thus, for each 80 ms stimulus presentation cycle, the image was displayed at 100% contrast for the first 40 ms (i.e., four frames of 10 ms each) and at 0% contrast for the next 40 ms (four frames). In both stimulation modes, the presentation appears perceptually continuous.

EEG Analysis

Frequency Domain Analysis

ICA was again applied on two participants blinking more than 0.2 times/s on average (mean = 0.19, SD = 0.093). The same ROIs were used, giving the repeated measures ANOVA the same two levels of *Region* (right occipito-temporal and central occipito-parietal), although there were also only two levels of *Condition* (sinewave and squarewave).

Time Domain Analysis

A baseline-correction was applied here as in the 720 ms/1.39 Hz condition of Experiment 1.

Results

Experiment 1: Varying temporal distance between faces

1. Frequency Domain

1.1 The Face-Categorization Response is Distributed Across Frequency-Characterized Harmonics

A significant response maximal over the right occipito-temporal ROI at the fundamental face stimulation frequency (*F*) for each condition was revealed in the grand-averaged frequency-domain amplitude spectrum (**Figure 2A**). Additional harmonic frequency face-categorization responses (i.e., 2*F*, 3*F*, etc.) also emerged clearly: between 4 and 14 significant harmonic responses were identified for each condition (including the fundamental, i.e., first, harmonic; **Table 1**).[2]

Z-Scores	**Harmonic number**														
Face SOA (ms)	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
240	**114.9**	**64.6**	***119.0***	**15.0**	**9.9**	***36.6***	-0.5	**2.5**	***31.0***	1.2	-0.4	0.0	0.4	-0.2	0.3
400	**37.0**	**60.6**	**63.4**	**15.5**	***93.3***	**11.8**	**5.2**	**4.5**	**3.1**	***36.6***	1.9	**4.6**	1.6	-0.9	***36.3***
560	**16.4**	**33.2**	**99.9**	**54.3**	**20.6**	**7.1**	***104.9***	**10.5**	**7.7**	**7.2**	**3.2**	-0.2	-0.9	***35.3***	1.3
720	**12.6**	**30.8**	**37.1**	**46.5**	**36.6**	**18.3**	**8.7**	**6.5**	***126.5***	**7.5**	**2.9**	**2.5**	1.9	1.1	1.8
880	**7.3**	**16.9**	**22.8**	**29.5**	**25.2**	**22.9**	**14.6**	**10.6**	**14.5**	**5.6**	***117.6***	**5.2**	**5.0**	**9.9**	**3.0**

Table 1. *Z-scores by condition for face stimulation harmonic responses, calculated from the average of all channels in the grand-averaged amplitude spectrum. Significant responses are shown in bold (Z-score > 2.32; p<.01, 1-tailed). Base frequency harmonic responses, shown in italics, were excluded from the selection of significant harmonic responses. Although the table displays harmonics by number, note that the frequency of each harmonic varies by condition.*

---

[2] Note that the face-categorization harmonic responses are not considered at frequencies overlapping with the base stimulation response; however, there is some evidence that face-selective responses do not combine with base stimulation responses: e.g., the response to base stimulation does not vary across conditions containing periodic vs. non-periodic embedded face images (Rossion et al., 2015; Jonas et al., in press).

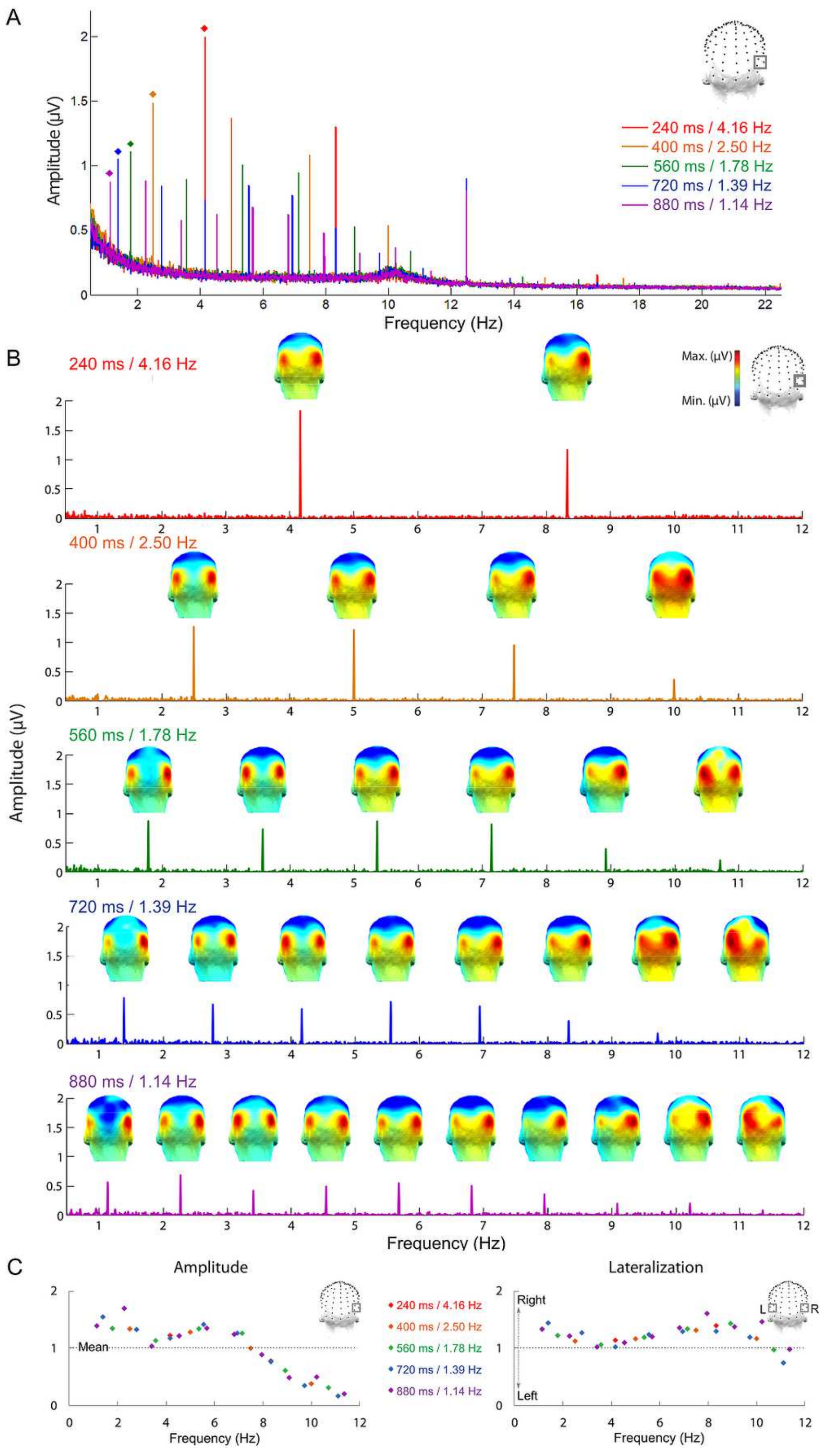

Figure 2. *Face-categorization responses are revealed in the frequency domain over the right occipito-temporal region-of-interest, as indicated on the topographical head maps (EEG channels PO10, P08, and P10).* ***A)*** *In the raw amplitude spectrum, a significant face-categorization response is revealed at the face stimulation rate (F),*

denoted by a diamond, for each face SOA/rate condition. Additional harmonic frequency responses (i.e., 2F, 3F, etc.) are distributed within a common frequency range depicted here, i.e., 0.5 to 22.5 Hz, in each condition. ***B)*** *Baseline-subtracted amplitudes; where a response to stimulation is not present, the amplitude is expected to be equal to 0 μV. Conditions are graphed separately here, so that topographic head plots may be included above each significant face-categorization harmonic response in this range. The scale of each head plot ranges from 0 μV to its maximum corrected amplitude (reported exactly in* ***Table S1B****).* ***C)*** *These baseline-subtracted face-categorization harmonic responses are plotted to emphasize trends in amplitude (left) and lateralization (right) as a function of frequency. Amplitude is normalized across conditions for display: the amplitude of each harmonic response is divided by the mean harmonic response amplitude for that condition. Lateralization across hemispheres is described as the right divided by left occipito-temporal region (i.e., ([PO10, PO8, P10]/[PO9, PO7, P9])); values above one represent a right-hemisphere advantage.*

These harmonic face-categorization responses were restrained within a common frequency range, with the mean maximal significant harmonic response frequency across conditions occurring at 19.3 Hz (SD = 1.11 Hz). Correspondingly, conditions with higher *F*s have fewer harmonic responses: e.g., the 240 ms/4.16 Hz condition had only 4 significant harmonic responses, while the 880 ms/1.14 Hz condition had 14 significant harmonic responses. Interestingly, the amplitude of the face-selective response appeared to be *distributed* among these harmonic frequency responses in each condition (**Figure 2B; Table 2**[3]). That is, there is an inverse

3

relationship between the number of harmonic responses and their amplitudes: the largest harmonic response amplitudes are present in the conditions with the lowest numbers of harmonic responses.

		Harmonic number														
A. Amp.	**Face SOA (ms)**	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
	240	1.99	1.30	*0.84*	0.16	0.09	*0.16*	0.05	0.05	*0.09*	0.03	0.03	*0.00*	0.03	0.02	*0.03*
	400	1.48	1.36	1.08	0.54	*0.86*	0.14	0.13	0.07	0.09	*0.17*	0.05	0.04	0.04	0.04	*0.09*
	560	1.11	0.89	1.01	0.94	0.53	0.34	*0.84*	0.14	0.12	0.09	0.07	0.05	0.06	*0.17*	0.05
	720	1.05	0.84	0.73	0.85	0.77	0.52	0.33	0.21	*0.90*	0.13	0.10	0.10	0.07	0.06	0.06
	880	0.87	0.88	0.58	0.62	0.68	0.63	0.48	0.32	0.36	0.19	*0.81*	0.11	0.10	0.09	0.07
B. Sub.																
	240	1.83	1.17	*0.75*	0.09	0.03	*0.11*	0.01	0.01	*0.06*	0.00	0.00	*0.00*	0.00	-0.01	*0.00*
	400	1.27	1.21	0.95	0.36	*0.77*	0.06	0.06	0.01	0.04	*0.12*	0.01	0.00	0.00	0.00	*0.05*
	560	0.87	0.73	0.87	0.81	0.39	0.20	*0.75*	0.07	0.05	0.03	0.02	0.00	0.02	*0.13*	0.01
	720	0.78	0.67	0.59	0.71	0.64	0.39	0.18	0.08	*0.81*	0.06	0.03	0.03	0.01	0.00	0.01
	880	0.56	0.68	0.42	0.49	0.54	0.50	0.36	0.20	0.20	0.08	*0.72*	0.04	0.03	0.03	0.01
C. SNR																
	240	12.35	9.52	*9.53*	2.23	1.64	*4.13*	1.25	1.35	*3.17*	1.15	1.00	*1.37*	1.05	0.79	*1.21*
	400	7.54	8.88	8.17	3.36	*9.31*	1.78	1.90	1.24	1.87	*4.14*	1.34	1.15	1.14	1.17	*2.87*
	560	4.66	5.57	7.21	7.11	4.06	2.61	*9.30*	1.83	1.62	1.40	1.25	1.14	1.24	*4.32*	1.39
	720	3.78	4.94	5.29	6.47	5.81	4.17	2.54	1.66	*10.2*	1.83	1.44	1.55	1.19	1.09	1.20
	880	2.79	4.51	3.52	4.57	4.87	4.96	3.92	2.52	2.28	1.76	*9.19*	1.56	1.54	1.44	1.11

Table 2. *Values for the amplitude (**A**), baseline-subtracted amplitude (**B**), and SNR (**C**), calculated from the right occipito-temporal ROI (channels PO10, PO8, P10), for*

Although this baseline correction is commonly applied to reduce the influence of activity unrelated to the stimulation at each harmonic frequency, it is not a perfect correction because the signal is not expected to simply sum linearly with the noise and the interaction of signal and noise varies depending on their relative magnitudes (Strasburger, 1987; Norcia et al., 1989). In previous studies, another method of baseline correction, signal-to-noise ratio (SNR), has been used (e.g., Liu-Shuang et al., 2014; Rossion et al., 2015). SNR uses a division of the baseline noise rather than a subtraction, emphasizing the clarity of the signal. Nevertheless, the pattern of activity is similar across these computations (**Table 2**).

the first fifteen harmonic face-categorization frequencies. Base harmonic frequencies are shown in italics.

Additionally, harmonic face-categorization responses appear to be characterized in terms of amplitude and hemispheric lateralization by the frequency at which they occur (**Figure 2C**). That is, across conditions, harmonic response amplitude generally decreases as frequency increases, although there is a local increase around 6 Hz. Concerning harmonic response lateralization, right hemisphere dominance appears strongest at the lowest frequencies, i.e., around 1 Hz, as well as broadly around 8 Hz (from about 6-10 Hz). Thus, while *F* determines the number of significant harmonic responses, which is inversely related to the amplitude of each harmonic response, the characteristics of each harmonic response depend on the frequency at which it falls (e.g., around 6 Hz or from 0-18 Hz) rather than by its number (e.g., *F*, 2*F*, 3*F*, etc.).

In comparison, for the base stimulation frequency (12.5 Hz), which did not differ across conditions, three significant base harmonic responses (i.e., up to 37.5 Hz) were identified for every condition.

1.2 Magnitude of the Face-Categorization Response

Harmonic face-categorization frequency responses were combined in order to quantify the full face-categorization response. Given the distribution of amplitude across significant harmonic responses revealed by the varying face SOAs/rates across conditions, we summed baseline-subtracted harmonic response amplitudes within each condition. Since, across conditions, each harmonic response is characterized by the frequency at which it falls (e.g., similar harmonic responses occur around 6 Hz) rather than by its number (e.g., similar responses do not occur at 2*F*) in the frequency domain, we considered all significant harmonic responses in each condition, occurring within a similar frequency range, rather than considering the same number of harmonic responses in each condition. Thus, the magnitude of the full face-categorization response was defined as *the sum of the baseline-subtracted amplitudes of all significant face-categorization harmonic responses in each condition.*

The magnitude of the full face-categorization response was remarkably stable across conditions, despite the variance in harmonic responses (**Figure 3**). Across four conditions, i.e., face SOAs/rates between 400 ms/2.50 Hz to 880 ms/1.14 Hz, the face-categorization response amplitude was about 4 μV. However, in the 240 ms/4.16 Hz condition, the magnitude of the face-categorization response was reduced by about 25%. Strikingly, in every condition, the largest face-selective response magnitude occurred in the right occipto-temporal region at channel P010, followed by adjacent channels P10 and P08. The channels giving the next largest response were at the three homologous locations in the left occipito-temporal region (PO9, PO7, and P9). The summed base stimulation harmonic response, which is highly similar across all conditions, peaked at about 2.3 μV over the medial occipito-parietal region (channels Oz, O2, and POO6), in contrast.

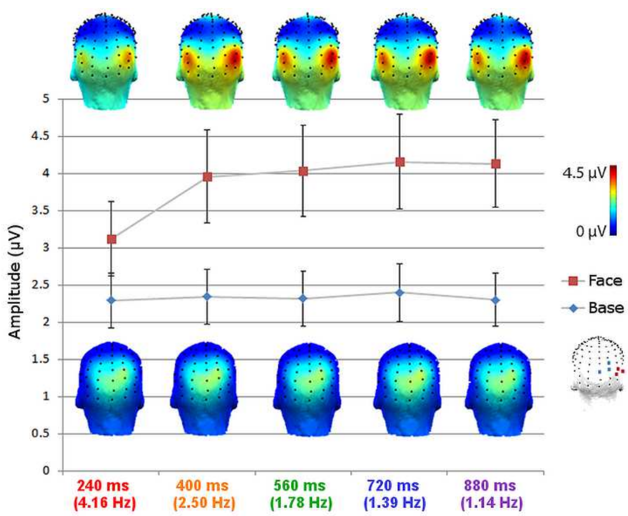

Figure 3. *Comparison of conditions for both face-categorization and base stimulation responses. The face-selective response is plotted as an average of the three maximum right occipito-temporal channels indicated on the far right topographical while head map (PO10, PO8, and P10; red squares); the base stimulation response as an average of the three maximum medial occipito-parietal channels (Oz, O2, and POO6; blue diamonds). A significantly lower amplitude face-categorization response is found only for the 240 ms face SOA condition. For a complementary comparison of conditions in the time domain, i.e., an overlay of responses revealing the lowest amplitude deflections for the 240 ms face SOA condition, see* ***Figure S2****.*

We chose the 720 ms/1.39 Hz condition as an example to further explore the face-categorization response magnitude (see Section *2.2.1*). Note that this condition was replicated in Experiment 2, so that a combination of participants from the two experiments yields a total of 27 unique participants. A significant face-categorization response was found in 15/16 participants from Experiment 1, with Z-scores ranging from 3.6 to 37 (M = 18) (**Figure 6**); considering the two experiments together, 26/27 participants had a significant face-selective response (Z-score M = 18, range 3.6 to 49). For those 26 participants, the amplitude of the right occipito-temporal ROI was significantly larger than that of the homologous left ($t(25) = 2.66$, $p = .01$, Cohen's d = .51). Note that while just 15% (SD = 4.2%) of the face-categorization response magnitude is concentrated in the right and left occipito-temporal ROIs, constituting less than 5% of the channels, these occipito-temporal ROIs predict extremely well the variance in all other channels (i.e., the average of all the remaining 122 channels) across participants ($r^2 = .84$, $p<.01$), suggesting that virtually all of the response of interest is captured by these ROIs.

Considering these 26 participants at the individual level, the magnitude of the face categorization response over the right occipito-temporal ROI ranged from 1.52 to 10.7 μV (M = 4.52 μV, SD = 2.45 μV). Over the homologous left occipito-temporal ROI, magnitudes ranged from 1.16 to 7.62 μV (M = 3.41 μV, SD = 1.88 μV). The hemispheric lateralization of the face-categorization response had a coefficient ranging from -28 to 68 (M = 13.1, SD = 24.4), with positive numbers indicating right dominance (20 participants) and negative numbers indicating left dominance (6 participants). Over the average of all 128 channels, magnitudes ranged from 0.53 to 2.82 μV (M = 1.30 μV, SD = 0.63 μV).

1.2.1 Statistical Comparison of Face-Categorization Responses

There was a main effect of *Region* ($F(2,14) = 11.9$, $p = .001$, $\eta_p^2 = 0.63$), reflecting the increased activity in the right and left occipito-temporal regions relative to the medial occipito-parietal region. Subsequent ANOVAs contrasting only two regions at a time showed a main effect of *Region* for ROT vs. MO: $F(1,15) = 24.7$, $p < .001$, $\eta_p^2 = .62$; and LOT vs. MO: $F(1,15) = 6.05$, $p = .03$, $\eta_p^2 = .29$; but not for LOT vs. ROT: $F(1,15) = 3.72$, $p = .07$, $\eta_p^2 = .20$). The main effect of *Condition* was also significant ($F(2.56,7.68) = 4.62$, $p = .02$, $\eta_p^2 = 0.61$); these factors were not qualified by a significant interaction ($F(8,8) = 2.46$, $p = .11$, $\eta_p^2 = 0.71$). The effect of condition was produced by a weaker response in the 240 ms (1/3) condition: when this condition was removed from the ANOVA, there were no significant differences between conditions (*Region*: $F(2,14) = 12.9$, $p = .001$, $\eta_p^2 = 0.65$; *Condition*: $F(3,13) = 0.21$, $p = .89$, $\eta_p^2 = 0.05$; interaction: $F(6,10) = 1.46$, $p = .28$, $\eta_p^2 = 0.47$). The analysis performed on all the electrodes averaged (in which *Region* was thus not a factor) also showed a main effect of *Condition* at the level of significance ($F(4,12) = 3.24$, $p =$

.05, η_p^2 = 0.52); there was no significant difference between conditions when removing the 240 ms (1/3) condition (F(3,13) = 2.68, p = .09, η_p^2 = 0.38).

1.2.2 Statistical Comparison of Base Stimulation Responses

Only a main effect of *Region* was present (F(1.30,9.11) = 18.2, p <.001, η_p^2 = 0.72), reflecting the higher activity in the medial occipito-parietal than right and left occipito-temporal regions. Following ANOVAs contrasting only two regions at a time showed a main effect of *Region* for MO vs. ROT: F (1,15) = 23.6, p < .001, η_p^2 = .61; and MO vs. LOT: F (1,15) = 37.1, p < .001, η_p^2 = .71; but not for LOT vs. ROT: F (1,15) = 0.589, p = .46, η_p^2 = .04). There was no main effect of *Condition* (F(4,12) = 2.17, p = .13, η_p^2 = 0.42), and no significant interaction between these factors (F(8,8) = 0.88, p = .57, η_p^2 = 0.47).

2. Time Domain

Responses in the time domain and frequency domains reflect much of the same information, however, these analysis modes have complementary strengths: response magnitude is more objectively and accurately determined in the frequency domain, while in the time domain, there is critical information for characterizing the temporal dynamics of the face categorization process. In classical SSVEPs, rapid presentation of stimuli creates a quasi-sinusoidal response in the time domain (Keitel et al., 2010; Norcia et al., 2015), for which only relative timing information in terms of phase is available in the frequency domain, as will be exemplified here for the base stimulation frequency of 12.5 Hz. However, the present design generates a differential response to face stimuli with the possibility of a long enough face SOA to identify a baseline and complex response components, with a determinable onset latency, in the time domain.

2.1. Face-Categorization Responses Vary Across Conditions as in the Frequency Domain

The face-categorization response is stable across face SOA conditions between 400 to 880 ms; however, for the 240 ms face SOA condition, there is not enough time in between face stimulus presentations, such that the deflections in the waveform are overlapping (**Figure 4**, right**).** This overlap produces interference which decreases the relative amplitude of maximal deflections for this condition (see also **Figure S2**), as well as affects response timing, e.g., the negative peak around 220 ms is delayed by approximately 12 ms. In the 400 ms SOA condition, there is also not a return to baseline, although the waveform is very similar to that of the other conditions with longer SOAs, which do contain distinct baseline periods. Similarly to the results in the frequency domain, in all conditions the right occipito-temporal channel PO10 displays the largest magnitude, about -3.5 µV, at the largest negative deflection peaking at about 210 ms.

In contrast to the complex face-categorization responses, responses to the base stimulation appeared as a simple, sinusoidal response at that frequency, similarly in terms of amplitude and phase across conditions (**Figure 4**, left). Corresponding with the results in the frequency domain, the magnitude of this response, estimated between 0 and 100 ms after stimulus onset, was largest at the occipito-parietal channel POO6, with peak amplitudes of approximately -2 to 2 µV, for all conditions except 240 ms (maximum at the right occipito-temporal channel PO8; this time window is influenced by the previous face-categorization response).

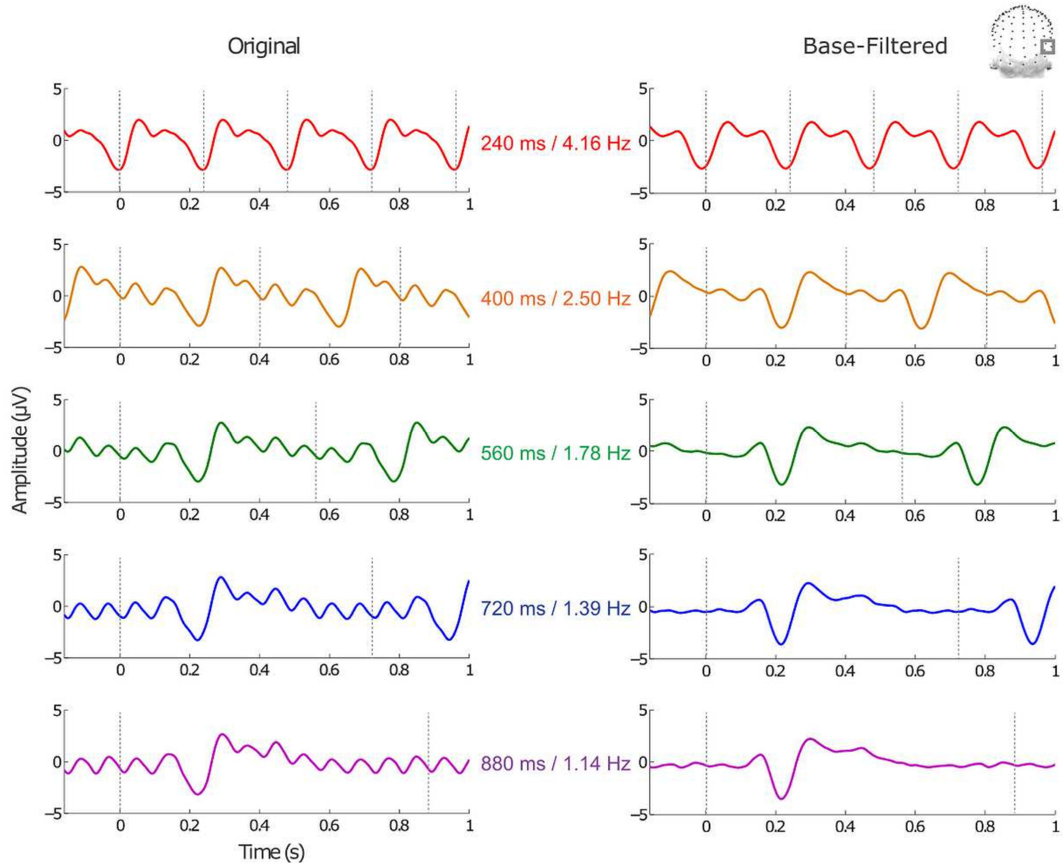

Figure 4. *Time-domain data is plotted from 160 ms before face stimulus onset (0 s) to 1 s after. Note that additional face presentations occur within this time window, according to the face SOA: each face presentation is indicated by a dotted gray line. The data is plotted for the grand-averaged right occipito-temporal region-of-interest, as indicated on the topographical head maps (EEG channels PO10, P08, and P10).* ***A)*** *In addition to the face-categorization response, these 'original' waveforms show responses to the base stimulation, i.e., a sinusoidal response with 80 ms per cycle.* ***B)*** *The face-categorization response is isolated by notch filtering to remove the base stimulation response at 12.5 Hz its harmonics. This response is stable across conditions, with the exception of the 240 ms face SOA, in which the face stimuli are presented too rapidly for a complete response to develop and a stable return to baseline; only three deflections are evident, with lower amplitude.*

2.2 Time Course of the Face-Categorization Response

To further explore the face-categorization response, the 720 ms/1.39 Hz condition was subjected to further analyses. This condition was selected as an example because it gives a large response magnitude in the frequency domain and provides a long enough interval between the presentations of face stimuli for a baseline of two pure base stimulation cycles, i.e., 160 ms, before face stimulus onset. This condition is not exceptional: the 880 ms condition, or even the 560 ms condition, could also have been used.

2.2.1. Response Components

Significant deflections between 123 and 545 ms, i.e., over a range of 422 ms, are evident over the right occipito-temporal ROI (**Figure 5A & B**). Within this time, no fewer than four occipito-temporal face-selective deflections, or "components" are clearly present: 1) a positive component with a peak latency of approximately 150 ms (termed "P1-face" in Rossion et al., 2015); 2) a large negative component with a peak latency of approximately 215 ms ("N1-face"); 3) a large positive component with a peak latency of approximately 290 ms ("P2-face"); and 4) a previously undescribed positive component with a peak latency of approximately 445 ms ("P3-face"). These differential (i.e., face-selective) components display distinctive spatial distributions (**Figure 5C**): the first component is spread over ventral medial and right occipital channels (peak at PO10), the second is bilateral and slightly more dorsal, peaking over the right occipito-temporal channel PO8, the third is lateralized over temporal channels, peaking at P10, and the fourth also peaks at channel P10 but appears particularly right-lateralized.

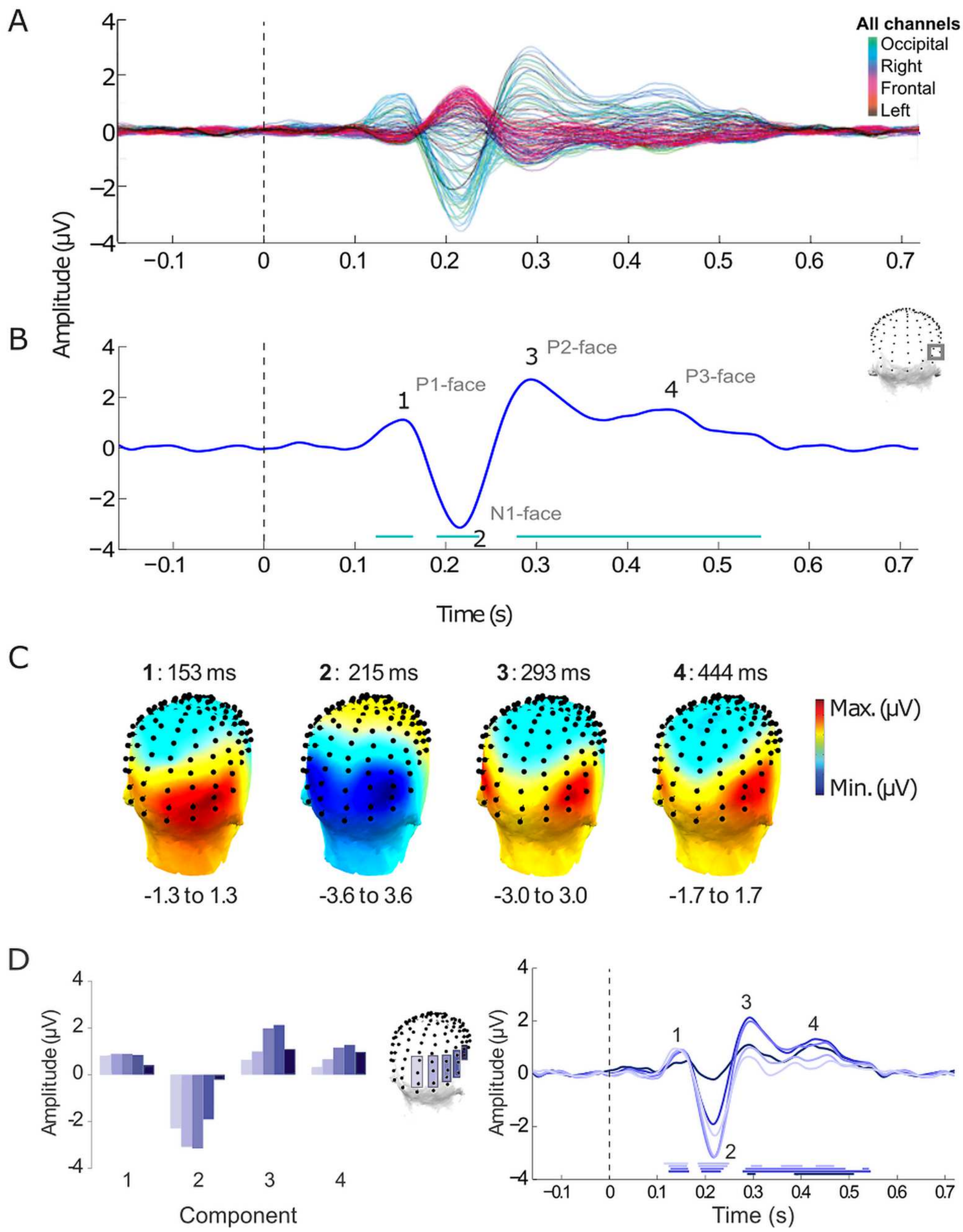

Figure 5. *Time course of face-categorization responses for the grand-averaged, base-filtered 720 ms face SOA condition.* ***A)*** *Responses for all 128 EEG channels, displaying four distinct deflections from about 100 ms, with respective peaks of approximately 150, 210, 290, and 450 ms, and residual activation continuing up to about 550 ms.* ***B)*** *The average of the right occipito-temporal ROI s is plotted here, as depicted on the topographical head map. The bars underneath the waveform indicate the time when the amplitude is significantly different from zero (*$p < .01$*). The four deflections are also evident here, as well as their precise peaks in time: P1-face (153*

ms), N1-face (215 ms), P2-face (293 ms), and P3-face (444 ms). ***C)*** *The scalp topographies of these four components, showing differing spatial patterns. The number and exact time at which the deflection peaks, and for which the topography is plotted, is indicated above each plot; beneath, the scale is indicated.* ***D)*** *A view of the posterior to anterior evolution of the face-categorization response over time. The relative amplitude are shown for five regions across a lateral line from the central occipital to right temporal-parietal areas, as depicted on the topographical head plot in the center; lighter shading corresponds to more posterior regions (and the order of these five regions is the same in the graph as represented on the head plot). From anterior to posterior, these five regions contain the respective electrodes: TP8h and TP8; P6, P8 and P10; PP06, P08 and PO10; P00, Oz and PO12; POOz, Oz and OIz.* ***Left:*** *The amplitude of these five regions is shown for the peak times of the four components identified in Panel B. Earlier components have relatively larger posterior than anterior response amplitudes.* ***Right****: Over the time course of the face-categorization response, more anterior regions also have relatively greater impact on later components. The most posterior region, e.g., only produces a significant response for the first two components, while the most anterior region has a significant response only for the last two components.*

2.2.2 Spatial Progression

In examining these topographies, a spatial progression is evident over time: the face-selective response, originating in medial and right occipital areas, evolves to be more lateralized, shifting from occipital regions towards a more dorsal and anterior location (**Figure 5D**). Indeed, from P1-face to P3-face, the response is nearly reversed in its posterior-anterior weight distribution. Quantitatively, for P1-face 44% of the response magnitude is the two most posterior regions and 33% in the most anterior two regions; while for P3-face only 23% of the response is the in two most posterior regions of P3-face and 51% in the most anterior regions. Visualization of the spatio-temporal progression of the response over the full 0 to 550 ms time window, in precise, 5 ms steps, is also available (**Movie 3**).

Movie3.mp4 (http://gofile.me/24HTb/1pPKFqI56)

Movie 3. *Visualization of the grand-averaged, base-filtered face-categorization response for the 720 ms condition. The topographical head plot viewing angle is centered over the right occipito-temporal region. Activity is plotted from 0 ms (stimulus onset) to 550 ms, with frames sampled at 5 ms. The scale is held constant at -3 to 3* μV.

In summary, the magnitude of face-selective responses is shown to be measured with high stability when summing baseline-subtracted harmonic response amplitudes across different face SOAs, with the exception of the overly rapid 240 ms SOA showing interference (i.e., decreased magnitude) from temporally overlapping face-selective responses. Generally, these face-categorization responses are complex, showing at least four significant deflections in the time domain and being spread over multiple harmonic frequencies.

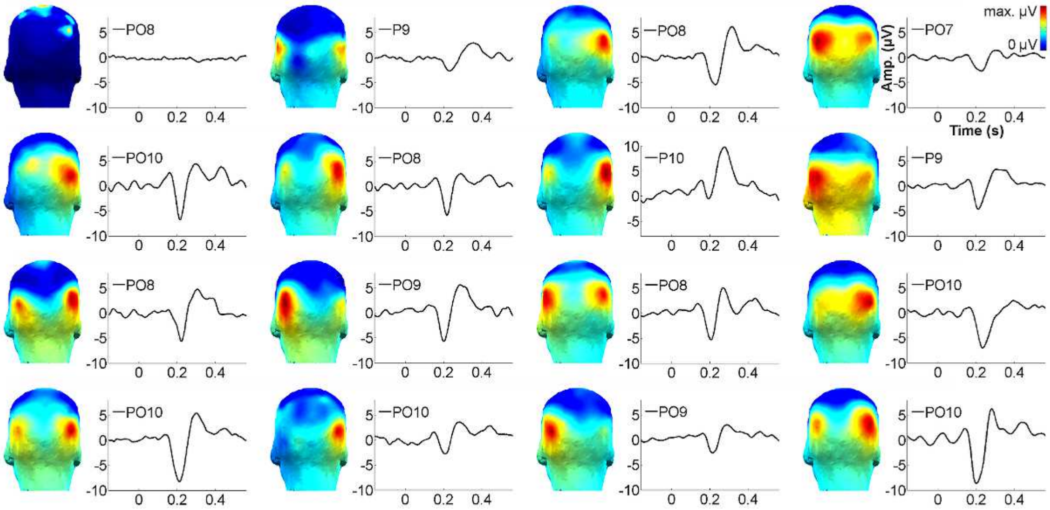

***Figure 6**. Individual participant data for the 16 participants of Experiment 1, from the four-minute recording of the 720 ms (1.39 Hz) condition. Each topographical head plot and the waveform to its right depicts the data of a single participant. The head plots show the sum of significant baseline-subtracted harmonics of 1.39 Hz and are scaled from 0 µV (dark blue) to the voltage of the maximal channel (red) separately for each participant. The waveforms show the same response in the time domain, with an x-axis of time (s) and a y-axis of amplitude (µV); a sinusoidal face presentation begins at 0 ms. The channel displayed is the maximal channel from the right or homologous left occipito-temporal ROI for each participant. Only one participant (top left corner) out of 16 did not show any face-selective response.*

Experiment 2: Sinusoidal vs. Squarewave Contrast Modulation Presentation

One goal of this study is to determine the exact onset of face-categorization responses in a fast visual stream, i.e., to briefly presented forward- and backward masked natural face images. In the main experiment, stimuli were presented progressively with sinusoidal contrast modulation (**Figure 1A**). As a consequence, 0 ms does not indicate the real time of stimulus onset: here, for the first frame (0-10 ms) the stimulus is displayed at 0% contrast (the next frames are at 38% then 71% contrast). Some delay is thus expected before the stimulus becomes detectable to the visual system, preventing precise inferences about onset information to be drawn from the data.

In order to measure the exact temporal delay elicited by sinusoidal contrast modulation of stimulus presentation, and so determine the exact onset of differential face-selective responses generated here, the second experiment with a new group of participants was performed with the 720 ms/1.38 Hz face presentation at an 80 ms/12.5 Hz base stimulation rate: in one condition, stimuli are presented with sinewave contrast modulation as before, and in a second condition, stimuli are presented with squarewave contrast modulation, i.e., in an on/off pattern, beginning with 100% contrast at 0 ms.

1. *Sinewave and Squarewave Stimulation Modes Generate Similar Responses in the Frequency Domain*

The harmonic distribution of face-categorization responses was similar across sinewave and squarewave presentation modes (**Figure 7A**). The number of statistically significant face-categorization harmonic responses was the same between the two conditions, i.e., there were 11 significant harmonic responses, up to 16.7 Hz (the same as for the 720 ms condition in Experiment 1; **Table 3**). Their summed magnitude was also similar, being approximately 4.7 μV on average, as well as their topographic distribution (**Figure 7B**). The face-categorization response amplitude peaked over right occipito-temporal regions at channel P10 in both conditions, followed by channels PO10 (the maximal channel in Experiment 1), PO12, PO8, and P8 in both conditions. The homologous channels over the left-hemisphere constituted the channels with the next largest response: i.e., in decreasing order for both conditions: P9, PO9, PO11, PO7, and P7.

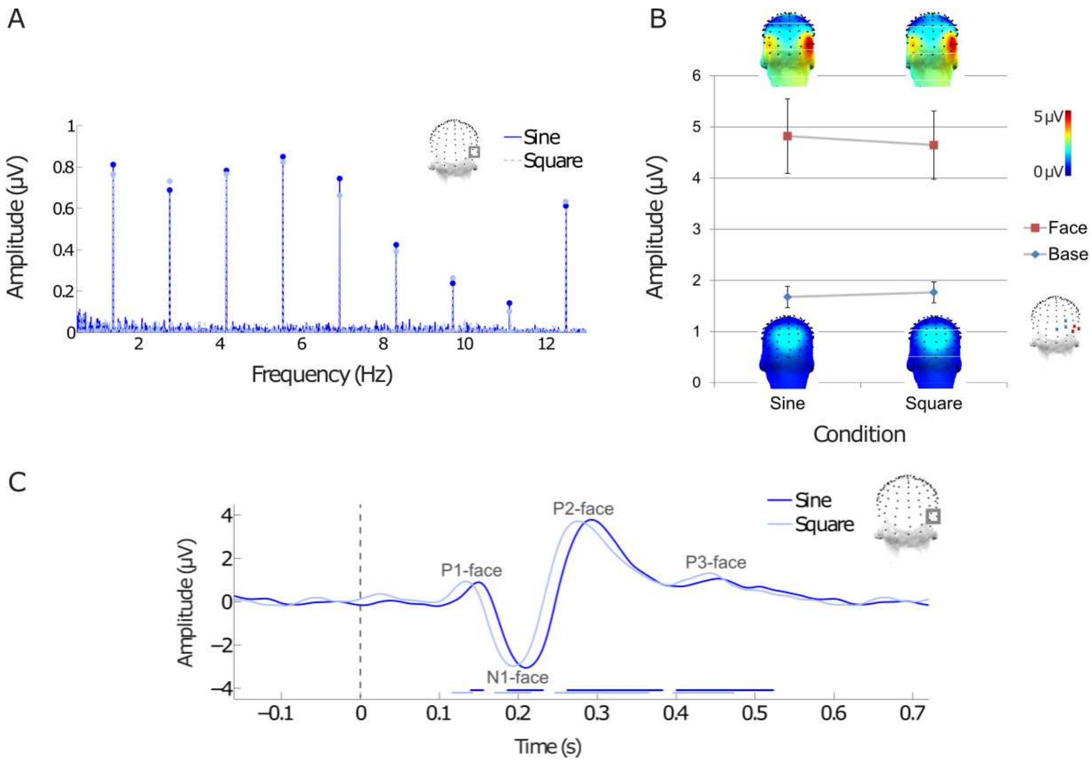

Figure 7. A) *The grand-averaged, baseline-subtracted, frequency-domain amplitude spectra for both sinewave ("sine") and squarewave ("square") conditions of Experiment 2. Data is graphed from the average of right occipito-temporal ROI (channels PO10, PO8, P10), as indicated on the topographic head map. In both conditions, the fundamental face-categorization response occurs at 1.39 Hz, and significant harmonic frequency responses continue until 16.7 Hz (indicated by markers at peak maxima). One base stimulation response is present at 12.5 Hz in the displayed range of the frequency spectrum here; significant harmonic base responses continued up to 37.5 Hz in both conditions.* ***B)*** *The conditions are compared for both face-categorization and base stimulation responses after summing their significant harmonics. The face-categorization response is quantified with the right occipito-*

temporal ROI (red squares on the far-right topographic head map); the base stimulation response is quantified with the medial occipito-parietal ROI (channels Oz, O2, and POO6; blue diamonds). No significant differences are found between the two conditions for either face-categorization or base stimulation response magnitudes. ***C***) *The grand-averaged response in the time-domain over the right occipito-temporal ROI to a face stimulus presentation at 0 ms in Experiment 2. In both conditions, the four face-categorization response components are evident, i.e., P1-face, N1-face, P2-face, and P3-face. Presenting stimuli with sinusoidal contrast modulation, i.e., gradually increasing stimulus contrast from 0 to 100% and back over a sinusoidal function ("sine", dark blue), is shown to delay all components of the face-categorization response by approximately 20 ms, i.e., a quarter of the sinusoidal presentation cycle duration, or two screen refreshes, until stimuli are presented at or above 50% contrast. This delay is relative to when stimuli are presented immediately at 100% contrast at 0 ms ("square", light blue) to the same participants. The bars underneath the waveforms indicate when the amplitude is significantly different from zero (p<.01).*

		Harmonic Number														
A. Z-Score	**Condition**	1	2	3	4	5	6	7	8	*9*	10	11	12	13	14	15
	Sine	**17.0**	**20.0**	**36.7**	**55.9**	**34.1**	**16.4**	**8.5**	**9.8**	***122.3***	**8.2**	**5.8**	**5.7**	1.4	**3.2**	1.9
	Square	**22.4**	**29.7**	**49.8**	**35.7**	**30.0**	**25.3**	**16.5**	**7.6**	***104.3***	**5.7**	**6.8**	**4.3**	0.3	1.0	1.3
B. Amp.																
	Sine	1.09	0.87	0.93	0.98	0.88	0.57	0.39	0.24	*0.70*	0.14	0.10	0.11	0.07	0.08	0.06
	Square	1.04	0.93	0.92	0.97	0.80	0.53	0.41	0.21	*0.72*	0.12	0.11	0.09	0.08	0.08	0.05
C. Sub.																
	Sine	0.81	0.69	0.78	0.85	0.74	0.42	0.25	0.14	*0.61*	0.07	0.03	0.05	0.02	0.02	0.01
	Square	0.77	0.74	0.77	0.82	0.66	0.39	0.27	0.10	*0.64*	0.05	0.04	0.03	0.02	0.02	0.00
D. SNR																
	Sine	4.24	4.87	6.16	7.25	6.46	4.15	3.09	2.27	*8.23*	1.89	1.50	1.67	1.35	1.40	1.23
	Square	4.18	4.99	6.00	6.47	5.77	4.00	3.14	1.94	*8.83*	1.76	1.64	1.52	1.30	1.43	1.08

***Table 3. A**) Statistical values by condition for the face-categorization harmonic frequency responses, shown for the first fifteen harmonic frequencies. Z-Scores were calculated from the average of all channels in the grand-averaged amplitude spectrum. Significant responses are shown in bold (Z-score > 2.32; p<.01, 1-tailed). The base frequency harmonic response in this range, shown in italics, was excluded from the selection of significant face-selective harmonic responses. Below, values for the amplitude (**B**), baseline-subtracted amplitude (**C**), and SNR (**D**), are calculated from the average of three right occipito-temporal chanels (PO10, PO8, P10. The base harmonic frequency is again shown in italics.*

The response to base stimulation also did not differ across conditions, despite the different presentation modes. There were three significant harmonic responses for the base stimulation frequency in both conditions (i.e., up to 37.5 Hz, as in Experiment 1). Their summed magnitude was also similar across conditions, equaling approximately 1.7 μV (**Figure 7B**). The response to the base stimulation peaked at the central-occipital channel Oz in both conditions (as in Experiment 1), followed by medial occipito-parietal channels POO6, POOz, POO5, O1, and O2.

1.1 Statistical Comparison of Face-Categorization Responses

There was only a main effect of *Region* ($F(2,14) = 16.8$, $p < .001$, $\eta_p^2 = 0.71$), reflecting the greater activity in the right then left occipito-temporal regions relative to the medial occipito-parietal region: following ANOVAs to contrast only two regions at a time showed a main effect of *Region* for ROT vs. MO: F (1,15) = 33.2, p < .001, $\eta_p^2 = .619$; and LOT vs. MO: F (1,15) = 6.51, p = .022, $\eta_p^2 = .20$; and also for ROT vs. LOT: F (1,15) = 5.39, p = .035, $\eta_p^2 = .26$). *Condition* (sinewave *vs* squarewave) did not produce a significant main effect ($F(1,15) = 0.62$, $p = .45$, $\eta_p^2 = 0.04$). The interaction between these factors did not reach significance ($F(2,14) = 1.86$, $p = .19$, $\eta_p^2 = 0.21$). When the average of all 128 channels was used to examine *Condition*, there was also no main effect ($F(1,15) = 0.02$, $p = .90$, $\eta_p^2 = 0.001$)

1.2 Statistical Comparison of Base Stimulation Responses

As above, only a main effect of *Region* was present ($F(2,14) = 30.5$, $p < .001$, $\eta_p^2 = 0.81$), reflecting the greater activity in the medial occipito-parietal region than right and left occipito-temporal regions. Subsequent ANOVAs to contrast only two regions at a time showed a main effect of *Region* for MO vs. ROT: F (1,15) = 35.3, p < .001, $\eta_p^2 = .70$; and MO vs. LOT: F (1,15) = 59.9, p < .001, $\eta_p^2 = .80$; and not for ROT vs. LOT: F (1,15) = 0.00, p = .99, $\eta_p^2 = .00$). There was neither a main effect of *Condition* ($F(1,15) = 0.126$, $p = .16$, $\eta_p^2 = 0.13$) nor a significant interaction between these factors ($F(2,14) = 1.02$, $p = .39$, $\eta_p^2 = 0.13$).

2. *Onset Delay Produced from Sinewave Stimulation in the Time Domain*

The four face-categorization response components are again identified for both stimulation modes (**Figure 7C**). In the sinewave condition, components *P1-face, N1-face, P2-face*, and *P3-face* peak at, respectively, 150 ms, 209 ms, 293 ms, and 453 ms. In the squarewave condition, each component peaks at, respectively, 133 ms, 193 ms, 276 ms, and 442 ms.

The onset of the response to sinewave relative to squarewave stimulation corresponds approximately to a delay of two screen refreshes, i.e., 20 ms: at the electrode showing the maximal face-categorization response (P10), a temporal delay of exactly 20 ms is produced by sinusoidal contrast modulation on the first component (*P1-face*). Over the right occipito-temporal region, the significant onset of each of the face-categorization response components is delayed by approximately 22 ms (SD = 8 ms), appearing to remain stable across components. In summary, while neither the face-selective response magnitude nor the harmonic distribution is affected by the sinusoidal stimulation mode, a delay of about a quarter of a stimulation cycle, i.e., 20 ms here, is produced in the time domain. This delay also projects to the onset time of all subsequent components.

Discussion

In the following sections we will discuss: 1) multi-harmonic frequency-domain response quantification; 2) the magnitude of the face-selective response; 3) the 100-ms onset of the face-selective response; 4) the 420-ms duration of the face-selective response; 5) the spatio-temporal dynamics of the face-selective response; and 6) cyclical and acyclical electrophysiological responses; we will finish with the Summary and Perspectives.

1. Objective quantification of a multi-harmonic selective response in the frequency domain

Previous studies comparing the neural response to faces and other object categories at a large system-level of organization, i.e., with EEG or MEG, have not been able to quantify face-selectivity. In event-related potential studies, this has been due to a difficulty of objectivity in identification and measurement of a response expressed in the time-domain (e.g., the N170 face effect, see Rossion & Jacques, 2008), as well as the pitfalls of a necessary subtraction procedure between time-domain waveforms recorded separately for faces and non-face objects. These limitations may be circumvented by frequency-domain quantification with the present paradigm, without the need for subtraction across conditions (Rossion et al., 2015). However, in frequency-domain analyses, quantification of a comprehensive response has posed difficulties because a standard method for combining multi-harmonic responses has not yet been introduced. Indeed, it is debated whether, and if so, how, multi-harmonic responses are relevant for evaluating a common response (see Norcia et al., 2015, Appendix 2): it has even been stated that "*there is no simple rule that would tell us how to combine the amplitude values at different harmonics into one single number that could be used as a measure of neural activity*" (Heinrich, 2010, p9). Thus, frequency-domain response quantification is traditionally performed at the level of individual harmonic responses (e.g., Morgan et al., 1996; Muller et al., 2006; Alonso-Prieto et al., 2013; Painter et al., 2014; see Norcia et al., 2015 for a large-scale review).

Here, we provide the first empirical evidence, to our knowledge, that the magnitude of a comprehensive selective response can be quantified through frequency-domain analysis by a summation of stimulation-specific baseline-corrected harmonic response amplitudes. This is principally evidenced by 1) the reliability of our results across stimulation rates: reliable (i.e., not statistically different) response magnitudes were found across four experimental conditions with differing face SOAs (400 to 880 ms) despite differing fundamental stimulation frequencies. And 2) the validity of our frequency-domain quantification in matching the apparent magnitudes of responses in the time domain: the 240 ms face SOA condition produced the statistically lowest magnitude in the frequency-domain quantification and in the time domain. Although we did not attempt to subjectively quantify the responses in the time domain across components, it is obvious that all the response components are also of the lowest magnitude for the 240 ms SOA condition (compare Fig. 3 to Supplemental Fig. 2). Thus, for the first time, a comprehensive response magnitude, computed over multiple face-selective harmonic responses, is shown to be independent of the specific face stimulation frequencies used. As an aside, it may be noted that in terms of methodological approach, a summation of harmonic response magnitudes has been found to perform better than other approaches (e.g., the maximum or median SNR across harmonics or T_{circ} statistics) in determining the significance of a multi-harmonic frequency-domain response (Heinrich, 2009).

Importantly, traditional frequency-domain analysis approaches would not have produced the same conclusions, given the differing distributions of harmonic response across conditions, i.e., a face-selective periodic response elicited every 880 ms (1.14 Hz) spreads over a large number of relatively low amplitude harmonics as compared to a face-selective response evoked every 400 ms (2.50 Hz) which spreads over fewer, but higher amplitude, harmonics.. For example, considering only the fundamental harmonic response amplitude or signal-to-noise ratio, or even an average over significant harmonic responses, would have determined that the 240 ms face SOA would have given the highest response magnitude, in clear contradiction to the results in the time domain.

Going a step further, we reasoned that if harmonic responses higher than the fundamental are indeed relevant to a common functional process, then we should be able to characterize the nature of their distribution in the frequency spectrum across conditions (note, however, that individual harmonic responses are not thought to be independent of one another): this is because different components of the response may be maximally captured in different (non-independent) frequency ranges. Across conditions, we found that face-selective harmonic responses were distributed within a common frequency range, despite the variance of the fundamental harmonic frequency (**Figure 2A**). Moreover, the harmonic responses follow similar trends over frequency in terms of their amplitude and topographic lateralization across conditions (**Figure 2B&C**). The maximal amplitude of the face-selective response, peaking at about 6 Hz, matches the frequency producing the maximal response for facial identity discrimination using FPVS in both EEG (Alonso-Prieto et al., 2013) and fMRI (Gentile & Rossion, 2014). The trends in right hemispheric lateralization, also peaking around 6-10 Hz and below 2 Hz, have not previously been reported to our knowledge, and should be further explored in future studies. However, we may already conclude that the frequency-characterization of harmonic responses implies that when considering multi-harmonic responses, the frequency at which a harmonic response falls is more significant than its number in the harmonic sequence.

Both the outcome of the quantification procedure and the frequency-distribution suggest that harmonic responses beyond the fundamental frequency continue to reflect relevant information regarding a category-selective response. Thus, these findings have important implications for the treatment of higher harmonic frequencies in future studies (see Heinrich, 2010; in the application to brain-computer interfaces, see, e.g., Müller-Putz et al., 2005).

<u>2. A robust face-selective response focused over the (right) occipito-temporal cortex</u>

We are able to report the magnitude of a comprehensive face-selective response for the first time with multi-harmonic frequency domain quantification: it is approximately 4 μV, with a range of about 8 μV across individual brains giving a significant response, as measured over three maximally-responding right hemisphere occipito-temporal channels on the scalp. This magnitude was found reliably, i.e., without significant statistical difference, across four separate conditions with differing face stimulus SOAs, in Experiment 1 (SE: 0.02 μV). Only in the condition where faces occurred every 240 ms, i.e., at the most rapid face presentation rate, was this response reduced by about 25%. This may potentially be explained by interference from overlapping face-selective responses as evident in the time domain. In addition, Experiment 2 revealed a face-selective response with an equal distribution of harmonic response amplitudes and equal total magnitude, whether the stimulus presentation was abrupt (squarewave) or sinusoidal. It is unlikely that habituation plays an important role in this reduction, since faces are always separated by object images here and the other conditions with relatively short face SOAs (e.g., 400 ms) are not significantly reduced relative to longer face SOAs. Additionally, habituation effects at the category level do not require short SOAs, usually also affect response latency and do not appear to be specific to faces (e.g., Kovacs et al., 2006; Nemrodov & Itier, 2012; Feuerriegel et al., 2015). Across all conditions in which the face-selective response could fully be expressed, the response peaked on the same right occipito-temporal channels, even though it spread more largely over all occipital and temporal channels, with variations over time. Considering one of these conditions as an example (i.e., 720 ms face SOA/1.39 Hz), at the group level across experiments, this response was about 3.5 times larger over the right occipito-temporal region of interest as compared to the average response recorded over the whole scalp. The response was 1.3 times larger over the right than the homologous left occipito-temporal channels at the group level, a statistically significant difference, and 77% of participants had a stronger response over the right OT ROI.

This large response reflects the wide distribution of face-selective neural responses across occipital and temporal cortices (Sergent et al., 1992; Allison et al., 1999; Tsao et al., 2008; Weiner & Grill-Spector, 2010; Rossion et al., 2012; Duchaine & Yovel, 2015; Zhen et al., 2015; Jonas et al., in press) and is in line with the well-established right hemispheric dominance of face perception in human adults (e.g., Hecaen & Angelergues, 1962; Hillger & Koenig, 1991; Sergent et al., 1992; Bentin et al., 1996; Rossion et al., 2012; Jonas et al., in press) and young (4-6 months) infants (de Heering & Rossion, 2015). However, this lateralization essentially reflects a group effect: despite testing only right-handed individuals here (see Bukowski et al., 2013), 23% of the individual participants presented with a larger face-selective response over the left hemisphere electrode sites. Note that this observation does not mean that the left hemisphere is dominant in these individuals, as a face-selective response recorded on the scalp depends on many factors that are not directly related to the underlying neural activity (e.g., orientation of the cortical sources with respect to the surface).

The 4 μV face-selective response obtained here cannot be compared to previous EEG/MEG studies, not only because of a lack of objective quantification in previous studies, but also due to the original paradigm used here: this response follows a very brief duration (i.e., about 50 ms in Experiment 1; 40 and about 50 ms in Experiment 2), allowing only a single glance, and the face is both forward- and backward-masked within the dynamic sequence of nonface objects presented. Yet, this face-selective response is rather large in absolute terms, corresponding to a high signal-to-noise ratio (SNR=4.4) across harmonics from the 27 participants across experiments viewing the 720 ms face SOA condition with sinusoidally modulated contrast presentation. It is also largely significant in all but one individual participant despite only a few minutes of stimulation. The reduction of the response by about 25% in the 240 ms SOA condition (4.16 Hz) in Experiment 1 also shows that the bulk of the response (i.e., 75%) is contained in the first 240 ms of a face-selective response (i.e., about 360 ms in total following the 120 ms response delay following sinusoidal face stimulus presentation onset).

3. High-level face-selectivity emerges at 100 ms post-stimulus onset

We found an early significant onset of face-selectivity for natural images at about 100 ms in our study, taking into account stimulus onset delays introduced by sinusoidal stimulation in Experiment 1. This early difference is highly significant at the group level, and observed in a few minutes of recording (for a given condition). Importantly, this latency value is found for natural images, which vary in low-level image statistics carried out in the amplitude spectrum (Torralba & Oliva, 2003; VanRullen, 2006). Yet, there are several reasons that this value most likely derives from a high-level face-selective response.

Firstly, the effects found here may not be explained by low-level differences in image amplitude spectrum: in previous studies with this paradigm, it has been shown that phase-scrambled versions of these stimuli do not produce significant periodic face-selective responses (Rossion et al., 2015; de Heering & Rossion, 2015). Secondly, a high-level response is indicated by the scalp topography of this early response, which is lateralized to the occipito-temporal cortex with a right hemispheric dominance (e.g., in contrast to the early response observed in Cauchoix et al., 2014 for instance, localized over medial occipital sites). Perhaps most importantly, the 100 ms onset of face-selectivity in the present study truly reflects face categorization, reflecting a time-locked periodic (i.e., generalized) deflection to the onset of variable face stimuli, differentiated from the responses to variable images from many object categories (**Figure 4**). Thus, without equating images for amplitude spectrum, which would degrade image quality in a widely variable set, the contributions of low-level differences between face and non-face object stimuli are controlled for here by the great amount of variation across natural images both between and within categories.

How does this 100 ms onset latency compare to previous evidence? Human observers can release a button when a face is present in a centrally presented natural scene as fast as about 250–300 ms following stimulus onset (Rousselet et al., 2003). Eye-tracking studies have shown that the fastest saccades to faces in contrast to objects in grayscale natural images occur at about 100-110 ms, with average face detection reaction times at approximately 150 ms (Crouzet et al., 2010). However, these early effects appear to be driven by amplitude spectra differences between faces and nonface objects (Honey et al., 2008; Crouzet & Thorpe, 2011). Most importantly, these extremely rapid latencies are obtained in two-alternative forced choice detection tasks, in which a face stimulus is compared to a single non-face category at a time.

The exact onset of category-selective responses, and specifically of face-selective responses, in the human brain has remained controversial (Bieniek et al., 2015; Rossion & Jacques, 2008). Previous EEG/MEG studies have reported conflicting results. Some studies, using mainly segmented stimuli, reported differences between faces and other object categories starting before 100 ms and peaking on the early ERP component to the sudden onset of images, the P1 (Eimer, 1998; Itier & Taylor, 2004) or M1 in MEG (Halgren et al., 2000; Liu et al., 2002; Okazaki et al., 2008). However, again, these early effects may be accounted for by differences in low-level visual cues such as surface properties or amplitude spectrum (Halgren et al., 2000; Rossion & Caharel, 2011). More recent MEG studies rely on a multivariate pattern (MVP) classification approach to distinguish and classify, over time, multiple categories of (segmented) stimuli based on their pattern of response across all recording channels on the scalp (Carlson et al., 2013; Cichy et al., 2014; Izik et al., 2014; van de Nieuwenhuijzen et al., 2013). These studies report early differences (i.e., before 100 ms) between categories such as faces, bodies, or inanimate objects, with peak decodability at about 120-130 ms. Yet, again, the use of a small number of naturalistic images by category and the decoding analysis performed across the exact same images suggest that this early categorization is based on low-level feature differences[4] (see also Cauchoix et al., 2014 for pre-100 ms differences evoked by target natural images of human body/faces *vs.* animal pictures accounted for by low-level visual cues).

As noted in the introduction, in typical EEG/MEG studies, i.e., studies in which stimuli are temporally and spatially isolated, robust differences between faces and other categories that cannot be explained by low-level cues emerge at the onset (120-130 ms) and peak of a N170 component over occipito-temporal sites (e.g., Bötzel et al., 1995; Bentin et al., 1996; Eimer, 2000; Halgren et al., 2000; Rossion et al., 2000; Rousselet et al., 2008; Rossion & Caharel, 2011; Ganis et al., 2012) although the exact onset is ambiguous and rarely measured by means of point-by-point statistical comparisons of waveforms (Rousselet et al., 2008; Rossion & Caharel, 2011). This latency also agrees with evidence for a clear EEG difference emerging at about 150 ms between target and non-target visual categories (i.e., animals or vehicles) independently of low-level visual cues (Thorpe et al., 1996; VanRullen & Thorpe, 2001).

Here, we set this onset time of face-selectivity on the human scalp to a slightly earlier latency, with a deflection starting at 100 ms for an abrupt onset of face stimuli. We argue that the particularly early 100 ms onset time of face-selectivity observed at a global level of organization (i.e., on the scalp) is due to the high sensitivity of the present approach, which isolates the specific activity evoked by natural faces by

---

4

This interpretation is supported by the particularly early (about 50 ms; Carlson et al., 2013; Cichy et al., 2014; Izik et al., 2014) decoding for specific exemplar images, i.e., when neural signals reach the primary visual cortex. See also Rice et al., 2014 for decoding of categories based on low-level visual differences in the fMRI signal recorded in the human ventral visual pathway.

inserting them in a rapid stream of other object categories. In previous studies, stimuli are presented in temporal isolation, so that each stimulus onset generates a large response in early visual areas (e.g., a standard P1 component), this response blurring genuine category-selective differences. Here, by using a dynamic visual stream, all of the process of interest concentrates on the *differential* response to faces relative to other objects, while the common brain response(s) with other objects, a mixture of low- and high-level visual processes, project to the base rate and can be selectively filtered out to isolate face-selective responses (**Figure 4**). Additionally, the early response found here is obtained implicitly, i.e., without an explicit categorization of faces and attentional resources focused on this category, as in many behavioral and electrophysiological studies reporting early differences (e.g., Cauchoix et al., 2014; Rousselet et al., 2007; see also Thorpe et al., 1996; VanRullen & Thorpe, 2001; Fabre-Thorpe, 2011 for a review).

Interestingly, a 100 ms onset of face-selectivity on the scalp when isolating the contribution of high-level categorization processes agrees remarkably with latency values of category-selective responses (to faces) observed with human intracerebral recordings performed directly in high-level visual areas (e.g., Davidesco et al., 2013; Jacques et al., 2016; Liu et al., 2009; Tang et al., 2014). It is also relatively consistent with the timing of onset latency of face-selective neurons in the monkey infero-temporal cortex, which is around 100 ms (e.g., Baylis et al., 1987; Perrett et al., 1982; Keysers et al., 2001; e.g., Tsao et al., 2006; Taubert et al., 2015 for recordings in fMRI-defined face-selective areas), although with a substantial amount of variance across studies (see Mormann et al., 2008, Table 1). This is interesting because (macaque) monkeys are much faster than humans at simple visual categorization tasks (Delorme et al., 2000) and the slightly longer latencies of face-selective responses found for humans in previous studies have often been accounted for by differences in brain size (Thorpe 2001). In fact, the present data suggests that the human brain may be as fast as the monkey brain to categorize faces, and that reaction time differences observed in behavioral tasks may be due to further (e.g., decisional) processes differing in speed between the two species rather than visual categorization time differences.

4. A prolonged (420 ms duration) face-selective response to briefly masked face appearances

The duration of face-selective responses to natural images seen at a single glance was examined here with FPVS-EEG by implementing five temporal distances between natural face stimuli (from a range of 240 to 880 ms, i.e., 4.17 to 1.14 Hz) within the rapid 12.5 Hz stream of non-face objects. We hypothesized that below a minimal onset asynchrony of face stimuli, face-selective responses interfering in time would decrease the response magnitude. In the time domain, a decreased response magnitude is evident across components for the 240 ms (1/3) condition. Overlapping responses are visible which produce interference only in this condition (**Figure 4**). Correspondingly, the quantification results revealed a significantly lower response magnitude only for the temporal distance of 240 ms (1/3 face proportion), as compared to all other conditions (**Figure 3**). Thus, analysis in the frequency domain (i.e., objectively at harmonics of a predetermined face presentation frequency) provides evidence that the duration of a face-selective response exceeds 240 ms.

On one hand, a weaker response for the shortest temporal distance between face presentations may seem counterintuitive: a more frequent presentation of faces might be expected to produce a larger response by decreasing the impact of the noise; however, in this design, this difference in the number of repetitions seems trivial in light of the high SNR across conditions: an analysis of the same number of repetitions in all conditions produced comparable results (**Figure S1**).

In the example 720 ms (1/9) condition, the duration over which a significant face-selective response may be recorded spans 420 ms in Experiment 1 (i.e., from about 120 ms to 540 ms poststimulus onset; 415 ms in Experiment 2, **Figure 7C**). Thus, the 400 ms SOA condition still provides enough time in between two faces to evoke a face-selective response that is not significantly lower than in the other three conditions with longer SOAs. This furthermore shows that the human brain is able to continue its face-selective response to the previous stimulus presentation in the time beyond when the next stimulus is presented and before the subsequent response is evoked (notice that the P3-face occurs after the next face image onset but before the P1-face in the 400 ms condition; **Figure 4**).

A face-selective response prolonged over hundreds of miliseconds is in line with results obtained in the few studies that reported post 200 ms face-selective responses (e.g., Schweinberger et al., 2004; Nasr & Esteky, 2009) and recent emphasis on the long duration of category-selective processes as identified with MVPA (Cichy et al., 2014; Mur & Kriegekorte, 2014; van de Nieuwenhuijzen et al., 2013). However, in these studies, stimuli were presented for durations and SOAs of hundreds of milliseconds. Here, stimuli were displayed briefly, at a single glance, i.e., appearing at or above 50% contrast for only 50 ms. Thus, given the quasi-continuous stimulus presentation mode but brief stimulus duration, the prolonged response found here is not confounded with eye movement exploration of the face; given the lack of face-related task (i.e., implicit processing), contributions of attentional, decisional and motor processes are also reduced. However, future studies will be needed to investigate the precise role of presentation time on face-selective responses.

5. Spatio-temporal dynamics of the face-selective response

An important practical/methodological implication of our findings of a prolonged face-selective response following a brief encounter with a face is that, in order to record the largest face-selective neural responses, face stimuli should not be presented at a faster rate than 2.5 Hz (400 ms SOA).

At the theoretical level, we identified 4 successive face-selective deflections, tentatively labeled *P1-face, N1-face, P2-face and P3-face*. It is necessary to emphasize that these deflections should not be equated to typical ERPs obtained to temporally isolated face stimuli such as the P1, N170, etc.: there are differential (i.e., face-selective) responses, previously undescribed in studies of temporally isolated face (and object) stimuli.) The first three, which form the bulk of the response, were already described in the previous study (Rossion et al., 2015), albeit with a lower signal-to-noise ratio due to the interference with the slower base rate used in that study (i.e., 5.88 Hz). Here, face-selective responses are statistically evaluated in space and time for the first time. Moreover, we were able for the first time, despite the low spatial resolution of surface recordings in EEG, to demonstrate a progression from posterior to anterior ventral regions of face-specific processes over 400 ms at the group-level, in line with a coarsely hierarchical processing of face-specific information in the ventral visual stream.

Although it is premature to relate these deflections to different face(-specific) processes, a possible account of the spatio-temporal signature of this response is that the perception of the stimulus as a face is based essentially on the earlier part of the response, within 200 ms post-stimulus onset (Rossion, 2014a), which would be sufficient to trigger a correct behavioral decision of face categorization in such a rapid visual stream. Although such decisions may be performed in 150 ms (saccadic RTs, Crouzet et al., 2010; Crouzet & Thorpe, 2011)) or 200-300 ms (manual RTs, Rousselet et al., 2003), they might be a bit slower when faces are temporally embedded in many highly variable distractors rather than a binary forced choice.

The late and more anteriorly weighted part of the face-selective response, i.e. a wide positive deflection with two peaks (P2-face, P3-face) persisting beyond 500 ms following stimulus onset, is not simply a reflection of periodic image presentation, which would lead to a generic “oddball” detection for instance. Indeed, these responses are not present to other periodically embedded categories (i.e., body parts or houses) in these fast visual streams (Jacques et al., in press). Rather, they might reflect one or more automatic face-selective processes. For example, it may reflect a specific increase in attention (e.g. Hajcak et al., 2013) triggered by the high saliency of face stimuli (Hershler & Hochstein, 2005; Crouzet et al., 2010), viewpoint-invariant representations which emerge gradually as information spreads more anteriorly to temporal regions (Pourtois et al., 2005; Axelrod & Yovel, 2012; see also Booth & Rolls, 1998; Freiwald & Tsao, 2010; Eifuku et al., 2011), or the memory encoding of visual representations in anterior regions of the temporal lobe (Sergent et al., 1992; Nakamura et al., 2000; Rajimehr et al., 2009; Gainotti & Marra, 2011; Avidan et al., 2013) compatible with the presence of a face-specific late potential (“AP350”) over the ventral anterior temporal lobe reported with human intracranial recordings (Allison et al., 1994; 1999; Rosburg et al., 2010).

Beyond speculation, now that these face-selective responses have been well characterized, they open a new world of possibilities to understand face categorization, and perceptual categorization in general, providing much more room in time and space, to disentangle the contribution of task and stimulus factors impacting face-categorization than in standard EEG studies focusing on a single component (i.e., mainly the N170; Bentin et al., 1996; Rossion, 2014a).

6. Cyclical and acyclical electrophysiological responses

Being able to characterize components of the response across time, i.e., examining the onset, duration, and spatiotemporal profile of face-selective responses, brings to light the correspondence between the time and frequency domains, which generally belong to two different traditions of research (i.e., “transient” *vs.* “steady-state” potentials; Regan, 1982; 2009). The face-selective responses produced here for longer temporal distances (i.e., 560 ms and above) demonstrate a reasonably flat baseline and clear component deviation peaks. This complexity is mirrored in the frequency domain, with a response spread across a range of higher harmonic frequencies, up to about 19 Hz (i.e., containing up to 14 significant face-selective harmonic responses in the 880 ms SOA condition). Importantly, the detailed responses found here are not typical ERPs, occurring to the abrupt and transient presentation of visual stimuli; nor do they resemble cyclical “steady-state” visual evoked potentials (SSVEPs) as have been traditionally reported (for a recent review, see Norcia et al., 2015). The consistent (“steady”) shape of the SSVEP response has been a key feature which has been used to define “the SSVEP”, e.g., as having a perfect oscillation (i.e., a sinusoidal waveform, e.g., Müller et al., 1997), or as not having a return to a baseline (Heinrich, 2010), or as having discrete frequency components remaining constant in amplitude and phase over an infinitely long time period (Regan, 1966; 1989; 2009), or as a response that has clear peaks in the frequency-domain representation (Vialatte et al., 2010), or even a response to a sinusoid stimulation as opposed to a squarewave stimulation (Victor & Zemon, 1985). A SSVEP has also been defined simply as the response to periodic stimulation

(Norcia et al., 2015). These distinctions may not be meaningful in practice, since they relate neither to functional processes nor to methodological implications: "SSVEPs" may simply reflect overlapping and interfering responses to the onset of periodic stimuli that are temporally too close to each other.

Besides referring to the approach (i.e., FPVS) rather than the kind of assumed responses, we propose that the term "SSVEP" be redefined as a "cyclical" electrophysiological response, as opposed to the "acyclical" responses typically labeled as ERPs. Following time-domain averaging, the former response occurs without a flat baseline while the latter has a return to baseline, more likely occurring as a complex, multi-harmonic response. This distinction is artificial and independent of the exact mechanism generating these responses (i.e., transient increases in amplitude or phase-resetting of ongoing oscillations, see e.g., Rousselet et al., 2007; Mouraux & Iannetti, 2008) but methodologically relevant, affecting how to extract temporal information: if cyclical, timing information can be taken from phase and interpreted only relative to other responses in the frequency domain; if acyclical, an additional aspect, peak latency, can be interpreted independently in the time domain.

Summary and perspectives

In measuring differential responses evoked by briefly presented natural images of faces inserted periodically in streams of natural object images, this study provides the first comprehensive report of the magnitude (at the group and individual levels), onset, duration, and spatio-temporal dynamics of category-selective responses in a rapid and continuously changing visual stream of stimulation.

Validating a multi-harmonic frequency domain response quantification, we report the magnitude of a comprehensive face-selective response for the first time: it is about 4 μV, with a range of about 8 μV across individual brains, as measured over three maximally-responding channels. This response is about 3.5 times larger over the right occipito-temporal cortex as compared to an average whole scalp response, and 1.3 times larger over the right than homologous left occipito-temporal channels, with 77% of individuals having a larger response over the right hemisphere. Importantly, despite short recording session (two trials of 120 s), significant face-selective responses can be identified in nearly all individual brains in this paradigm (i.e., 26/27 here).

The high-level face-selective response emerges at about 100 ms following face presentation. It lasts for 420 ms, i.e., until about 520 ms post-stimulus onset, despite each face appearing s briefly (50-80 ms) and being forward- and backward-masked. The bulk of the response (i.e., 75%) is contained in the first 240 ms of a face-selective response. Given that squarewave and sinewave stimulation modes provide identical responses, despite the stimulus being visible for a slightly longer duration in the sinewave stimulation mode, it seems that a face appearing for 40 ms in a perceptually continuous presentation stream generates a full face-selective response. However, in natural viewing conditions, changes in fixation typically occur at a slower rate, such that future studies are needed to investigate whether there is an advantage for longer image presentation durations.

Four successive face-selective components (P1-face, N1-face, P2-face, P3-face) are present, which largely overlap on the scalp but present a posterior (lateral occipital) to anterior (temporal) gradient shift. These differential components emerge as complex deflections (“differential ERPs”) from cyclical electrophysiological responses to the rapid periodic inputs, suggesting that the distinction between “ERPs” and “SSVEPs”, although methodologically relevant, appears largely artificial. Compared to transient stimulation modes in EEG, which have only identified the

N170 as a reliable index of face-selectivity, the identification of four successive face-selective components opens an avenue for understanding the nature of human face categorization, its development and neural basis.

Finally, the present quantification and spatio-temporal characterization of the selective response to single-glanced natural images of faces also provides a reference frame for future investigation of perceptual categorization (i.e., discrimination and generalization). Here, faces were used as the stimulus of interest to measure category-selectivity due to their social and biological importance for humans, as well as their well-studied category-selective neural responses with multiple modalities (e.g., EEG, MEG, intracerebral recordings, fMRI, and single neurons), however the paradigm may be extended to other image categories (Jacques et al., in press). Thus, stimulation with perceptually continuous natural images in FPVS-EEG opens a world of opportunities to dissociate the contribution of various task and stimulus factors on perceptual (face) categorization.

Acknowledgments

This work was supported by the European Research Council (ERC; grant number facessvep 284025 to BR) an “Action de Recherche Concertee” grant (ARC; 13/18-053) and the Belgian National Foundation for Scientific Research (FNRS; grant number FC7159 to TR). The authors have no conflict of interests to report.

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.

Allison T, Puce A, Spencer DD, McCarthy G. 1999. Electrophysiological studies of human face perception I: Potentials generated in occipitotemporal cortex by face and non-face stimuli. Cerebral Cortex. 9:415-430.

Alonso-Prieto EA, Van Belle G, Liu-Shuang J, Norcia AM, Rossion B. 2013. The 6 Hz fundamental frequency rate for individual face discrimination in the right occipito-temporal cortex. Neuropsychologia. 51:2863-2975.

Appelbaum LG, Wade AR, Vildavski VY, Pettet MW, Norcia AM. 2006. Cue-invariant networks for figure and background processing in human visual cortex. Journal of Neuroscience. 26:11695–11708.

Avidan G, Tanzer M, Hadj-Bouziane F, Liu N, Ungerleider LG, Behrmann M. 2013. Selective dissociation between core and extended regions of the face Processing network in congenital prosopagnosia. Cerebral Cortex. 246:1565-1578.

Axelrod V. and Yovel G. 2012. Hierarchical Processing of Face Viewpoint in Human Visual Cortex. Journal of Neuroscience. 32:2442-2452.

Baylis GC, Rolls ET, Leonard CM. 1987. Functional subdivisions of the temporal lobe neocortex. Journal of Neuroscience. 7:330–342.

Bentin S, McCarthy G, Perez E, Puce A, Allison T. 1996. Electrophysiological studies of face perception in humans. Journal of Cognitive Neuroscience. 8:551–565.

Bieniek MM, Bennett PJ, Sekuler AB and Rousselet GA. 2015. A robust and representative lower bound on object processing speed in humans. European Journal of Neuroscience. 1-11.

Booth MC, Rolls ET. 1998. View-invariant representations of familiar objects by neurons in the inferior temporal visual cortex. Cerebral Cortex. 86:510–523.

Bötzel K, Schulze S, Stodieck SRG. 1995. Scalp topography and analysis of intracranial sources of face-evoked potentials. Experimental Brain Research. 104:135–143.

Braddick OJ, Wattam-Bell J, Atkinson J. Orientation-specific cortical responses develop in early infancy. Nature. 320: 617-619.

Bukowski H, Dricot L, Hanseeuw B, Rossion B. 2013. Cerebral lateralization of face-sensitive areas in left-handers: only the FFA does not get in right. Cortex. 49 :2853-2859.

Carlson T, Tovar DA, Alink A, Kriegeskorte N. 2013. Representational dynamics of object vision: the first 1000 ms. Journal of Vision. 1310:1.1-1.19.

Cauchoix M, Barragan-Jason G, Serre T, Barbeau E.J. 2014. The neural dynamics of face detection in the wild revealed by MVPA. The Journal of Neuroscience. 343:846-854.

Cerf M, Harel J, Einhäuser W, Koch C. 2008. Predicting human gaze using low-level saliency combined with face detection. Platt, J.C., Koller, D., Singer, Y. and Roweis, S. Eds.. Advances in neural information processing systems, MIT Press, Cambridge, MA.

Cichy RM, Pantazis D, Oliva A. 2014. Resolving human object recognition in space and time. Nature Neuroscience. 17:455:462.

Crouzet SM, Kirchner H, Thorpe SJ. 2010. Fast saccades toward faces: face detection in just 100 ms. Journal of Vision. 10:16.1–16.17.

Crouzet SM, Thorpe, SJ. 2011. Low level cues and ultra-fast face detection. Frontiers in Psychology. 2:342.

Davidesco I, Harel M, Ramot M, Kramer U, Kipervasser S, Andelman F, Neufeld MY, Goelman G, Fried I, Malach R. 2013. Spatial and object-based attention modulates broadband high-frequency responses across the human visual cortical hierarchy. The Journal of Neuroscience. 33:1228–1240.

Delorme A, Richard G, Fabre-Thorpe M. 2000. Ultra-rapid categorisation of natural images does not rely on colour: a study in monkeys and humans. Vision Research. 40:2187–2200.

de Heering A, Rossion B. 2015. Rapid categorization of natural face images in the infant right hemisphere. eLife. 4:e06564.

Dzhelyova M, Rossion B. 2014. Supra-additive contribution of shape and surface information to individual face discrimination as revealed by fast periodic visual stimulation. Journal of Vision. 1414:15.1–15.14.

Duchaine B, Yovel G. 2015. A Revised Neural Framework for Face Processing. Annual Review of Vision Science. 1:393–416.

Eifuku S, De Souza, WC, Nakata R, Ono T, Tamura R. 2011. Neural representations of personally familiar and unfamiliar faces in the anterior inferior temporal cortex of monkeys. PLoS One. 64:e18913.

Eimer M. 1998. Does the face-specific N170 component reflect the activity of a specialized eye processor? NeuroReport. 9:2945–2948.

Eimer M. 2000. Event-related brain potentials distinguish processing stages involved in face perception and recognition. Clinical Neurophysiology. 111:694-705.

Fabre-Thorpe M. 2011. The characteristics and limits of rapid visual categorization. Frontiers in Psychology. 2:243.

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

Gainotti G, Marra C. 2011. Differential contribution of right and left temporo-occipital and anterior temporal lesions to face recognition disorders. Frontiers in Human Neuroscience. 5:55.1-55.11.

Ganis G, Smith D, Schendan HE. 2012. The N170, not the P1, indexes the earliest time for categorical perception of faces, regardless of interstimulus variance. NeuroImage. 623:1563-1574.

Gentile F, Rossion B. 2014. Temporal frequency tuning of cortical face-sensitive areas for individual face perception. NeuroImage. 90:256-265.

Hajcak G, MacNamara A, Foti D, Ferri J, Keil A. 2013. The dynamic allocation of attention to emotion: Simultaneous and independent evidence from the late positive potential and steady state visual evoked potentials. Biological Psychology. 9:447–455.

Halgren E, Raij T, Marinkovic K, Jousmaki V, Hari R. 2000. Cognitive response profile of the human fusiform face area as determined by MEG. Cerebral Cortex. 10:69–81.

Hecaen H, Angelergues R. 1962. Agnosia for faces prosopagnosia. Archives of Neurology, 7:92–100.

Heinrich SP. 2009. Permutation-based significance tests for multiharmonic steady-state evoked potentials. IEEE Transactions on Biomedical Engineering. 562:534–537.

Heinrich SP. 2010. Some thought son the interpretation of steady-state evoked potentials. Documenta Ophthalmology. 120:205–214.

Hershler O, Hochstein S. 2005. At first sight: A high-level pop-out effect for faces. Vision Research. 45:1707–1724.

Hershler O, Hochstein S. 2006. With a careful look: Still no low-level confound to face pop-out. Vision Research. 46:3028–3035.

Hershler O, Golan T, Bentin S, Hochstein, S. 2010. The wide window of face detection. Journal of Vision. 10(10):21.

Hillger LA, Koenig O.1991. Separable mechanisms in face processing: evidence from hemispheric specialization. Journal of Cognitive Neuroscience. 31:42-58.

Honey C, Kirchner H, VanRullen R. 2008. Faces in the cloud: Fourier power spectrum biases ultrarapid face detection. Journal of Vision. 812:9.1–9.13.

Isik L, Meyers EM, Leibo JZ, Poggio T. 2014. The dynamics of invariant object recognition in the human visual system. Journal of Neurophysiology. 111:91–102.

Itier RJ, Taylor MJ. 2004. N170 or N1? Spatiotemporal differences between object and face processing using ERPs. Cerebral Cortex. 14:132–142.

Jacques C, Rossion B. 2007. Early electrophysiological responses to multiple face orientations correlate with individual discrimination performance in humans. NeuroImage. 36:863-876.

Jacques C, Witthoft N, Weiner KS, Foster BL, Rangarajan V, Hermes D, Miller KJ, Parvizi J, Grill-Spector K. 2016. Corresponding ECoG and fMRI category-selective signals in human ventral temporal cortex. Neuropsychologia. 83: 14-28.

Jacques C, Retter TL, Rossion B. In press. A single glance at natural face images generates larger and qualitatively different category-selective spatio-temporal signatures than other ecologically-relevant categories in the human brain. NeuroImage. doi:10.1016/j.neuroimage.2016.04.045.

Jeffreys DA. 1989. A face-responsive potential recorded from the human scalp. Exp. Brain Res. 78:193–202.

Jeffreys DA, Tukmachi ESA. 1992. The vertex-positive scalp potential evoked by faces and by objects. Experimental Brain Research. 91:340-350.

Kanwisher N, McDermott J, Chun MM. 1997. The fusiform face area: a module in human extrastriate cortex specialized for face perception. The Journal of Neuroscience. 17:4302-4311.

Keil MS. 2008. Does face image statistics predict a preferred spatial frequency for human face processing. Proceedings of the Royal Society B: Biological Sciences. 275:2095–2100.

Keitel C, Andersen SK. and Mülller MM. 2010. Competitive effects on steady-state visual evoked potentials with frequencies in- and outside the α band. Experimental Brain Research. 2054:489-95.

Keysers C, Xiao DK, Földiák P, Perrett DI. 2001. The speed of sight. Journal of Cognitive Neuroscience. 13:90–101.

Liu J, Harris A, Kanwisher N. 2002. Stages of processing in face perception: an MEG study. Nature Neuroscience. 5:910–916.

Liu H, Agam Y, Madsen JR, Kreiman G. 2009. Timing, timing, timing: fast decoding of object information from intracranial field potentials in human visual cortex. Neuron. 62:281–290.

Liu-Shuang J, Norcia AM, Rossion B. 2014. An objective index of individual face discrimination in the right occipito-temporal cortex by means of fast periodic visual stimulation. Neuropsychologia. 52:57-72.

Maris E, Oostenveld R. 2007. Nonparametric statistical testing of EEG- and MEG-data. Journal of Neuroscience Methods. 164:177-190.

Morgan ST, Hansen JC, Hillyard SA. 1996. Selective attention to stimulus location modulates the steady state visual evoked potential. Proceedings of the National Academy of Sciences USA. 94:4770-4774.

Mormann F, Kornblith S, Quiroga RQ, Kraskov A, Cerf M, Fried, I, Koch C. 2008. Latency and selectivity of single neurons indicate hierarchical processing in the human medial temporal lobe. Journal of Neuroscience. 2836 :8865-8872.

Mouraux A, Iannetti GD. 2008. Across-trial averaging of event-related EEG responses and beyond. Magnetic Resonance Imaging. 267:1041-1054.

Mouraux A, Iannetti GD, Colon E, Nozaradan S, Legrain V, Plaghki L. 2011. Nociceptive steady-state evoked potentials elicited by rapid periodic thermal stimulation of cutaneous nociceptors. The Journal of Neuroscience. 3116:6079-6087.

Müller MM, Teder W, Hillyard SA. 1997. Magnetoencephalographic recording of steady-state visual evoked cortical activity. Brain Topography. 93:163-168.

Müller-Putz GR., Scherer R., Brauneis C, Pfurtscheller G. 2005. Steady-state visual evoked potential SSVEP-based communication: impact of harmonic frequency components. J. Neural Engineering. 2:1–8.

Mur M, Kriegeskorte N. 2014. What's there, distinctly, when and where? Nature Neuroscience. 173:332-333.

Nakamura K, Kawashima R, Sato N, Nakamura A, Sugiura M, Kato T, Zilles K. 2000. Functional delineation of the human occipito-temporal areas related to face and scene processing: A PET study. Brain. 123:1903–1912.

Nasr S, Esteky H. 2009. A study of N250 event-related brain potential during face and non-face detection tasks. Journal of Vision. 95:5.1–5.14.

Norcia AM, Tyler CW, Hamer RD, Wesemann W. 1989. Measurement of spatial contrast sensitivity with the swept contrast VEP. Vision Research. 29:627–637.

Norcia AM, Appelbaum LG, Ales AM, Cottereau B, Rossion B. 2015. The steady-state visual evoked potential in vision research: a review. Journal of Vision. 156:4.1–4.46.

Okazaki Y, Abrahamyan A, Stevens CJ, Ioannides AA. 2008. The timing of face selectivity and attentional modulation in visual processing. Neuroscience. 152:1130–1144.

Painter DR., Dux PE., Travis SL, Mattingley JB. 2014. Neural responses to target features outside a search array are enhanced during conjunction but not unique-feature search. Journal of Neuroscience. 34(9):3390-3401.

Perrett DI, Rolls ET, Caan W. 1982. Visual neurons responsive to faces in the monkey temporal cortex. Experimental Brain Research. 47:329–342.

Potter MC, Levy EI. 1969. Recognition memory for a rapid sequence of pictures. Journal of Experimental Psychology 81:10–15.

Potter MC. 2012. Recognition and memory for briefly presented scenes. Frontiers in Psychology, **3**:32.1-32.9.

Potter MC, Wyble B, Hagmann CE, McCourt E.S. 2014. Detecting meaning in RSVP at 13 ms per picture. Attention Perception and Psychophysics. 76:270-279.

Pourtois G, Schwartz S, Seghier ML, Lazeyras F, Vuilleumier P. 2005. Portraits or people? Distinct representations of face identity in the human visual cortex. Journal of Cognitive Neuroscience. 17:1043–1057.

Puce A, Allison T, Gore JC, McCarthy G. 1995. Face sensitive regions in human extrastriate cortex studied by functional MRI. Journal of Neurophysiology. 74:1192-1199.

Rajimehr R, Young JC, Tootell RB. 2009. An anterior temporal face patch in human cortex, predicted by macaque maps. Proceedings of the National Academy of Sciences of the United States of America. 106:1995-2000.

Regan D. 1982. Comparison of transient and steady-state methods. Annals of the New York Academy of Sciences. 388:45–71.

Regan D. 2009. Some early uses of evoked brain responses in investigations of human visual function. Vision Research. 49:882–897.

Rice GE, Watson DM, Hartley T, Andrews TJ. 2014. Low-level image properties of visual objects predict patterns of neural response across category-selective regions of the ventral visual pathway. Journal of Neuroscience. 34:8837–8844.

Rosburg T, Ludowig E, Dümpelmann M, Alba-Ferrara L, Urbach H, Elger CE. 2010. The effect of face inversion on intracranial and scalp recordings of event-related potentials. Psychophysiology. 47:147–57.

Rossion B, Gauthier I, Tarr MJ, Despland P-A, Linotte S, Bruyer R, Crommelinck M. 2000. The N170 occipito-temporal component is enhanced and delayed to inverted faces but not to inverted objects: an electrophysiological account of face-specific processes in the human brain. Neuroreport. 11:1-6.

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, Boremanse A. 2011. Robust sensitivity to facial identity in the right human occipito-temporal cortex as revealed by steady-state visual-evoked potentials. Journal of Vision. 112:16.1–16.21

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, Jacques C. 2011. The N170 : understanding the time-course of face perception in the human brain. The Oxford Handbook of ERP Components, Edited by S. Luck and E. Kappenman. Oxford University Press.

Rossion B, Alonso-Prieto E, Boremanse A, Kuefner D, Van Belle G. 2012. A steady-state visual evoked potential approach to individual face perception: effect of inversion, contrast-reversal and temporal dynamics. NeuroImage. 63:1585-1600.

Rossion B, Hanseeuw B, Dricot L. 2012. Defining face perception areas in the human brain: a large-scale factorial fMRI face localizer analysis. Brain and Cognition. 79:138-157.

Rossion B. 2014a. Understanding face perception by means of human electrophysiology. Trends in Cognitive Sciences. 18:310-318.

Rossion B. 2014b. Understanding individual face discrimination by means of fast periodic visual stimulation. Experimental Brain Research. 232: 1599-1621.

Rossion B, Torfs K, Jacques C, Liu-Shuang J. 2015. Fast periodic presentation of natural face images reveals a robust face-selective electrophysiological response in the human brain. Journal of Vision. 15:18.1-18.18.

Rousselet GA, Macé M J-M, Fabre-Thorpe M. 2003. Is it an animal? Is t a human face? Fast processing in upright and inverted natural scenes. Journal of Vision. 36:5-5.

Rousselet GA, Mace MJ, Fabre-Thorpe M. 2004. Animal and human faces in natural scenes: how specific to human faces is the N170 ERP component? Journal of Vision. 41:13–21.

Rousselet GA, Macé M J-M, Thorpe SJ, Fabre-Thorpe M. 2007. Limits of event-related potential differences in tracking object processing speed. Journal of Cognitive Neuroscience. 198:1241-1258.

Rousselet GA, Husik JS, Bennett PJ, Sekuler AB. 2008. Time course and robustness of ERP object and face differences. Journal of Vision. 812:3-3.

Scheirer WJ, de Rezende Rocha A, Sapkota A, Boult TE. 2014. Perceptual annotation: Measuring human vision to improve computer vision. IEEE Transactions on Pattern Analysis and Machine Intelligence. 36:1679–1686.

Schweinberger SR, Huddy V, Burton AM. 2004. N250r: A face-selective brain response to stimulus repetitions. NeuroReport. 15:1501–1505.

Sergent J, Ohta S, MacDonald B. 1992. Functional neuroanatomy of face and object processing. Brain. 115 :15–36.

Strasburger H. 1987. The analysis of steady-state visual evoked potentials revisited. Clinical Vision Science. 13:245-256.

Tang H, Buia C, Madhavan R, Crone NE, Madsen JR, Anderson WS, Kreiman G. 2014. Spatiotemporal dynamics underlying object completion in human ventral visual cortex. Neuron. 83:736–748.

Tanskanen T, Näsänen R, Montez T, Paallysaho J, Hari R. 2005. Face recognition and cortical responses show similar sensitivity to noise spatial frequency. Cerebral Cortex. 15:526–534.

Taubert J, Van Belle G, Vanduffel W, Rossion B, Vogels R. 2015. The effect of face inversion for neurons inside and outside fMRI-defined face-selective cortical regions. Journal of Neurophysiology. 1135:1644-1655.

Thorpe S, Fize D, Marlot C. 1996. Speed of processing in the human visual system. Nature. 381:520–522.

Thorpe SJ, Fabre-Thorpe M. 2001. Neuroscience. Seeking categories in the brain. Science. 291:260–263.

Torralba A, Oliva, A. 2003. Statistics of natural image categories. Network: Computational Neural Systems. 14:391-412.

Tsao DY, Freiwald WA, Tootell RB, Livingstone MS. 2006. A cortical region consisting entirely of face-selective cells. Science. 3115761:670–674.

Tsao DY, Moeller S, Freiwald WA. 2008. Comparing face patch systems in macaques and humans. Proceedings of the National Academy of Sciences. 105:19514–19519.

VanRullen R, Thorpe SJ. 2001. Is it a bird? Is it a plane? Ultra-rapid visual categorization of natural and artifactual objects. Perception. 306:655-668.

VanRullen R. 2006. On second glance: still no high-level pop-out effect for faces. Vision Research. 46:3017-3027.

Van de Nieuwenhuijzen ME, Backus AR., Bahramisharif A, Doeller CF, Jensen O, van Gerven MAJ. 2013. MEG-based decoding of the spatiotemporal dynamics of visual category perception. NeuroImage. 83:1063–1073.

Vialatte FB, Maurice M, Dauwels J, Cichocki A. 2009. Steady state visual evoked potentials in the delta range 0.5-5 Hz. International Conference on Neural Information Processing. 5506:399-406.

Vialatte FB, Maurice M, Duwels J, Cichocki A. 2010. Steady-state visually evoked potentials: focus on essential paradigms and future perspectives. Progress in Neurobiology. 904:418-438.

Victor JD, Zemon V. 1985. The human visual evoked potential: Analysis of components due to elementary and complex aspects of form. Vision Research. 25:1829–1842.

Weiner KS, Grill-Spector K. 2010. Sparsely-distributed organization of face and limb activations in human ventral temporal cortex. NeuroImage. 52:1559-1573.

Zhen Z, Yang Z, Huang L, Kong X, Wang X, Dang X. et al. 2015. Quantifying interindividual variability and asymmetry of face-selective regions: A probabilistic functional atlas. NeuroImage. 113:13-25.