NeuroImage 176 (2018) 465–476

Contents lists available at ScienceDirect

NeuroImage

journal homepage: www.elsevier.com/locate/neuroimage

Ultra-coarse, single-glance human face detection in a dynamic visual stream

Genevieve L. Quek [a,*,1,2], Joan Liu-Shuang [a,2], Valérie Goffaux [a,b], Bruno Rossion [a,c,d]

[a] Institute of Research in Psychology (IPSY) & Institute of Neuroscience (IoNS), University of Louvain, Belgium
[b] Department of Cognitive Neuroscience, Faculty of Psychology and Neuroscience, Maastricht University, The Netherlands
[c] Université de Lorraine, CNRS, CRAN, F-54000, Nancy, France
[d] Université de Lorraine, CHRU-Nancy, Service de Neurologie, F-5400, France

ARTICLE INFO

Keywords:
Face detection
Spatial frequency
EEG
Human vision

ABSTRACT

Effective human interaction depends on our ability to rapidly detect faces in dynamic visual environments. Here we asked how basic units of visual information (spatial frequencies, or SF) contribute to this fundamental brain function. Human observers viewed initially blurry, unrecognizable natural object images presented at a fast 12 Hz rate and parametrically increasing in SF content over the course of 1 minute. By inserting highly variable natural face images as every 8th stimulus, we captured an objective neural index of face detection in participants' electroencephalogram (EEG) at exactly 1.5 Hz. This face-selective signal emerged over the right occipito-temporal cortex at <5 cycles/image, suggesting that the brain can – at a single glance – discriminate vastly different faces from multiple unsegmented object categories using only very coarse visual information. Local features (e.g., eyes) are not yet discernable at this threshold, indicating that fast face detection critically relies on global facial configuration. Interestingly, the face-selective neural response continued to increase with additional higher SF content until saturation around >50 cycles/image, potentially supporting higher-level recognition functions (e.g., facial identity recognition).

Introduction

Vision in the real world requires the human brain to rapidly transform luminance input from the retina into meaningful and complex objects, often in dynamic visual environments. Nowhere is this capacity more striking than in the case of *face detection*, by which we mean the brain's ability to rapidly discriminate faces from many other potential object categories (i.e., to categorize a face *as a face*). There is enormous evolutionary pressure for efficient face detection in humans, since it acts as the mandatory gateway to higher-level perceptual face processing, such as identity and emotion recognition (Besson et al., 2017). That is, before you can know whether a face belongs to a friend or a stranger, whether its expression is angry or happy, the visual system must first categorize that stimulus *as a face*, rather than any other type of visual object. Whereas human face detection is largely automatic (Hershler and Hochstein, 2005) and can be achieved in just over 100 ms (i.e., at a single glance, Crouzet et al., 2010; Rousselet et al., 2003), artificial face detection systems still lag well behind human performance (Scheirer et al., 2014; Viola and Jones, 2004; Yang et al.; Zhang and Zhang, 2010). Indeed, despite tremendous progress in artificial face detection in recent years (in particular thanks to deep learning approaches, see Yang et al.), human observers still readily outstrip these systems, easily perceiving faces that are far away (i.e., small), embedded in complex backgrounds, or obscured by foreground objects (Lewis and Edmonds, 2003; Rossion et al., 2015).

Despite the importance of this fundamental brain function, surprisingly little is known about how basic units of visual information – spatial frequencies (SFs), i.e., the spatial scale of luminance variations in an image (Campbell et al., 1969; De Valois and De Valois, 1980) – support the categorization of a face as a face. This contrasts with the richer scientific literature concerning the role of spatial frequency in higher-level face processing functions (Ruiz-Soler and Beltran, 2006), such as identity (Goffaux et al., 2011; Gold et al., 1999; Näsänen, 1999; Ojanpää and Näsänen, 2003; Ramon et al., 2015), sex recognition (Khalid et al., 2013; Schyns and Oliva, 1999), and emotion recognition (Gao and Maurer, 2011; Vuilleumier et al., 2003). Here we used an original electrophysiological approach to *i)* determine the minimal amount of SF content capable of driving robust face detection in the human brain (i.e., the

* Corresponding author. Place Cardinal Mercier 10, 1348 Louvain-la-Neuve, Belgium.
E-mail address: g.quek@donders.ru.nl (G.L. Quek).
[1] Present Address: Donders Institute for Brain, Cognition and Behaviour, Radboud University, Montessorilaan 3, 6525 HR Nijmegen, The Netherlands.
[2] Indicates shared first authorship.

https://doi.org/10.1016/j.neuroimage.2018.04.034
Received 9 November 2017; Received in revised form 9 April 2018; Accepted 13 April 2018
Available online 18 April 2018

detection threshold), and *ii)* characterize how face detection signals (i.e., face-selective brain activity) accumulate from this minimal threshold onwards with the addition of finer scale information.

Since effective face detection in natural environments demands both speed and accuracy despite highly variable and complex visual inputs, we imposed these same constraints on the perceptual processes measured here. To this end, we employed a dynamic visual stimulation paradigm to objectively quantify neural discrimination between faces and a wide array of other object categories presented very rapidly (Jacques et al., 2016; Quek and Rossion, 2017; Retter and Rossion, 2016; Rossion et al., 2015). In this paradigm, observers view highly variable natural object images (e.g., plants, animals, buildings, artefacts, etc.) presented at a fast periodic rate of 12 Hz (SOA = 83.33 ms), with face images interleaved at regular intervals (1/8 images) (Fig. 1 and Movie 1) (Retter and Rossion, 2016; Rossion et al., 2015). To guard against face detection based on simple low-level visual features, faces/objects are not segmented from their natural backgrounds, and faces vary substantially in age, sex, expression, as well as position, size, and viewpoint (Fig. 1; Fig. S1). Together, the rapid presentation rate and high image variability emulate the demands placed on the visual system by effective object categorization in the real world (i.e., speed and high categorical diversity). In the context of this dynamic image stream, a neural response occurring exactly at 1.5 Hz (i.e., 12 Hz/8) as captured with scalp electroencephalography (EEG) provides an objective, experimentally pre-defined marker for face detection (i.e., a face-selective neural index of the perceptual categorization of faces amongst various object categories, Retter and Rossion, 2016). Importantly, where previous electrophysiological studies have often focused on individual, subjectively defined components related to face detection (e.g., the N170, Collin et al., 2012; Goffaux et al., 2003b; Halit et al., 2006), our approach here takes into consideration the *entire* face-selective response – a complex multi-component waveform spanning several hundred milliseconds (Retter and Rossion, 2016)[3] (see also Carlson et al., 2013; Cichy et al., 2014; Nemrodov et al., 2016).

Supplementary video related to this article can be found at https://doi.org/10.1016/j.neuroimage.2018.04.034.

In the present study, the SF content of this image sequence parametrically increased over the course of 1 min. Images initially appeared blurry and indistinguishable, and progressively sharpened every 4 s (from 0.5 to 128 cycles per image, or cpi) until they became clearly recognizable as faces and objects (see Movie 1). Importantly, this parametric filtering method does not merely decompose images into discrete SF bands (e.g., low vs. mid vs. high SF bands) (Collin et al., 2012; Costen et al., 1994; Gaspar et al., 2008; Goffaux et al., 2003b; Halit et al., 2006; Näsänen, 1999). Rather, the progressive increase in low-pass filter cut-off enables the quantification of cumulative integration of face information from low to high SFs. In this way, this approach aligns with the presumed coarse-to-fine accumulation of information in the visual system (Goffaux et al., 2011; Hegdé, 2008; Morrison and Schyns, 2001; Rossion, 2014; Sergent Elliset al, 1986); namely that coarse low SF information is extracted before finer high SF details.

We pursued two goals with this experimental design: First, we aimed to identify the minimal SF content capable of driving a significant face-selective response, so as to shed light on the nature of the visual information underlying early face detection (i.e., face vs. object discrimination). One possibility in this regard is that face detection is supported by a holistic/configural template, wherein the presence of distinct individual facial features is less relevant than the perception of a global integrated structure (Caharel et al., 2013; Goold and Meng, 2016; Purcell and Stewart, 1988). If this is indeed the case, then extremely coarse visual information, as in the early stages of our stimulation sequence, should be sufficient to drive a face-selective response. Several lines of research support this prediction. Firstly, categorizing a face as a face necessarily precedes higher-order face processing such as identity recognition, which itself can already be resolved at relatively low SFs (e.g., 8–12 cpi, Costen et al., 1994; Gaspar et al., 2008; Näsänen, 1999). Moreover, face detection efficiency correlates with increased contrast sensitivity at low SFs (e.g., 0.5 cycles per visual degree, or cpd, Owsley and Sloane, 1987), and rapid face detection in the visual periphery and early face-selective neural markers also appear to rely on such coarse information (Boucart et al., 2016; Goffaux et al., 2003a,b, 2011; Halit et al., 2006). Finally, despite having visual acuity limited to the very low SF range (Peterzell et al., 1995), even 4–6 month old infants generate clear neural face detection responses when viewing briefly presented natural faces in a similar paradigm (de Heering and Rossion, 2015). An alternative possibility is that face detection is critically contingent on the perception of certain face parts (e.g., eyes) (Gava et al., 2008; Itier et al., 2007; Paras and Webster, 2013; Rousselet et al., 2014). This would predict that the face-selective response we measure here should not emerge until spatial resolution is sufficiently high for these diagnostic features to be distinguishable.

Second, regardless of the exact minimal SF content needed for face detection, it remains unknown whether cumulatively increasing SFs above this threshold still contribute to the face vs. object categorization process. That is, if face detection can be achieved on the basis of very degraded visual information, what do finer scales add to this process? Existing work using parametric structure (phase) de-scrambling suggests that the visual system is capable of generating a *full* face detection response on the basis of only partial or degraded information (Ales et al., 2012; Liu-Shuang et al., 2015). If low SFs suffice to generate a full face detection response, despite high temporal constraints, then this response should saturate immediately after the detection threshold. Alternatively, in a challenging and dynamic visual environment, the visual system may continue to accumulate additional higher SF content to refine the face detection response. In that case, the point of saturation may be far away from the detection threshold, with a potential increasing response function between the two. Clarifying these outstanding issues by *i)* identifying the spatial frequency threshold for face detection, and *ii)* establishing whether and how the visual system integrates finer details, will not only shed light on the neural basis of rapid, automatic face detection in dynamic visual environments, but could also serve to inspire and constrain computational models of face detection (Scheirer et al., 2014).

Materials & methods

Participants

We tested 19 participants (11 females) in exchange for monetary compensation. We excluded the data of three individuals who either did not follow task instructions appropriately (1 participant) or whose EEG recording contained excessive noise/muscular artefact (2 participants). The final sample consisted of 16 participants (10 females, mean age = 22 ± 2), of whom 12 also took part in an additional experiment using full-spectrum images immediately prior to the main experiment (see section below on Face detection of full-spectrum faces). All reported to having normal/corrected-to-normal vision; none had a history of neurological or psychiatric disorder. We obtained written informed consent prior to testing in accordance with the guidelines of the Biomedical Ethical committee of the University of Louvain (Belgian Number: B403201111965).

Stimuli

Stimuli were greyscale images of 100 faces and 200 objects (e.g.,

[3] Note that whether the approach does indeed capture the full face-selective response depends critically on the face-presentation frequency employed, as this reflects the SOA between consecutive faces and therefore the degree to which they mask each other. We point the interested reader to our previous work for a full discussion of this issue (Retter and Rossion, 2016).

buildings, trees, animals, vehicles, etc.), sized to 256 × 256 pixels. All faces and objects were left embedded in their natural backgrounds (i.e., unsegmented), and varied greatly in orientation, lighting, size, and overall appearance (examples given in Fig. S1A). We verified that there was no structural regularity in either category by separately averaging all face and object images and determined that no recognizable object or face shapes could be distinguished (see Fig. S1B). Stimuli were displayed at a viewing distance of 80 cm on a 120 Hz BenQ LED computer screen and subtended 9.08° visual angle. In the face stimulus set, the average rectangular region delimited by the forehead, cheekbones, and chin of

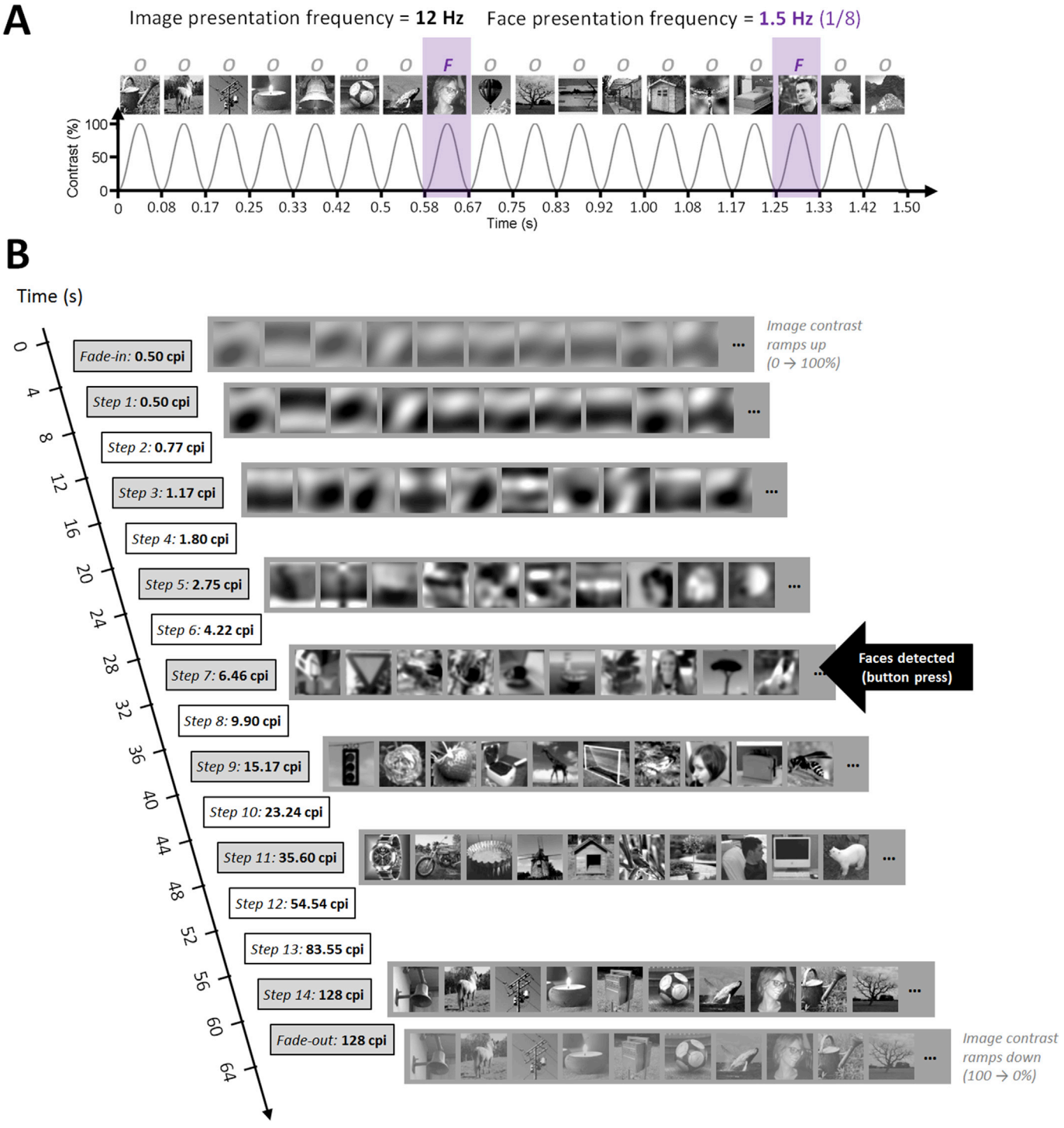


Fig. 1. The dynamic visual stimulation paradigm. **A)** Images were presented using sinusoidal contrast modulation at a fast periodic rate of 12 Hz (ISI = 83.33 ms), so that participants perceived a fast, continuous stream of images. Highly variable faces were interleaved with objects at fixed intervals throughout the sequence (1/8 images, i.e., 1.5 Hz). **B)** Schematic illustration of the coarse-to-fine stimulation with parametrically increasing SF content throughout the sequence (see also Movie 1), with example images for the shaded SF steps on the left. Observers were instructed to press the spacebar as soon as they perceived a face in the sequence, while brain activity was recorded with high-density EEG. Each sequence lasted 56 s, flanked at its start and end by an additional 4 s during which the maximum image contrast gradually ramped up and down, respectively (total duration = 64 s, see Movie 1 for an example Face sequence). A total of 768 images (96 faces and 672 objects) were shown during this time.

the face subtended 3.56° × 4.61° visual angle (median values; min = 2.09° × 2.52°, max = 6.74° × 9.04°), and thus occupied approximately 20% of the image surface (see Fig. S1C). Faces were positioned at a median eccentricity of 1.69° visual angle relative to the center of the image (min = 0.05°, max = 3.66°).

Full-spectrum face and object stimuli were low-pass spatial filtered at 14 logarithmic cut-offs (referred to here as SF Steps). We used logarithmic, rather than linear, cut-offs in order to maximize the sampling of low SFs since this is *i)* where most of image energy is contained (Fig. S1B), *ii)* where energy decays the fastest as a function of SF, and *iii)* where the visual system is most sensitive to contrast (Ginsburg et al., 1984). Fig. S1D presents these low-pass cut-off values with example images. Since faces only occupy a fraction of the total image, we also indicate the approximate SF cut-offs in terms of cycles per face width (cpf), corresponding to the median cut-off estimate across the 100 face exemplars. Following spatial filtering, we adjusted the luminance and contrast values of the resulting images to match the mean values of the original full-spectrum image set. Consequently, all 4200 (face and object) images were equivalent in luminance and contrast both within and across the 14 SF Steps.

Design & procedure

Our design was similar to previous face categorization experiments using fast periodic visual stimulation designs (Jacques et al., 2016; Retter and Rossion, 2016; Rossion et al., 2015), and was implemented using custom software written in Java. Each 56 s sequence consisted of 672 images shown through sinusoidal contrast modulation at a frequency of 12 images per second (12 Hz, SOA = 83.33 ms, Fig. 1A). Within a single sequence, the SF content of images parametrically increased over the course of 14 sequential SF Steps (4 s per SF Step). In this way, the sequence initially appeared blurry and gradually sharpened over the course of the 56 s (see Fig. 1B & Movie 1). The sequence was flanked by 4-s pre- and post-ludes, which were repetitions of the first and last SF Steps respectively. The depth of the contrast modulation ramped up during the pre-lude and down during the post-lude to minimize blinks or muscular artefacts elicited by the sudden appearance or disappearance of flickering stimuli. The total duration of a full stimulation sequence (sweeping across the 14 SF Steps and including the pre- and post-lude) was 64 s. A blue central fixation cross appeared between 2–5 s before sequence onset and remained superimposed on the periodically presented images until 2–5 s after sequence offset. Participants were asked to maintain fixation on the central cross throughout the entire sequence.

There were two sequence types: *Face* sequences contained randomly selected object images with face images interleaved at a fixed rate as every 8th image (i.e., 12 Hz/8 = 1.5 Hz). In this way, there were 6 face images contained within each 4 s SF step and 666.67 ms between consecutive faces. The participant's task was to press the spacebar only *once* per sequence, as soon as they thought that the sequence contained faces. We instructed them to respond as quickly and as accurately as possible. To ensure participants paid attention to this instruction, we also included *No Face* sequences as Catch Trials, in which only randomly selected object images appeared. Here the correct response was to withhold the button press response entirely. Note that although object exemplars did repeat more often than face exemplars did within each 64-s sequence, critically, no exemplar was ever repeated within any given SF Step. Additionally, previous work has shown that varying the ratio of face and object repetitions does not influence the properties of the face-selective response (Retter and Rossion, 2016).

In keeping with previous studies (e.g., Jacques et al., 2016), we expected that the stimulation in the critical *Face* sequences would elicit two responses in the EEG spectrum: one at the image presentation frequency (12 Hz), reflecting aspects of visual processing common to faces and objects (referred to here as the *common response*), and one at the face presentation frequency (1.5 Hz) reflecting face detection/categorization processes (*face-selective response*). This latter 1.5 Hz response can only arise if the neural response evoked by each briefly presented face consistently differs from that evoked by the many other object categories appearing in the sequence. Note that these stimulation frequencies were not chosen arbitrarily; a 12 Hz image presentation rate effectively limits perceptual processing to a single glance per image (i.e., no saccades possible), and 1.5 Hz face presentation frequency allows ample time (i.e., 667 ms) between face presentations to allow the full face response to unfold (~450 ms in duration, Retter and Rossion, 2016). These two frequencies are also highly dissociable in the frequency spectrum and in terms of their associated scalp topographies. It should be noted that although the appearance of faces in the sequence is strictly periodic, their temporal predictability does not contribute to or drive the face-selective response (Quek and Rossion, 2017).

EEG acquisition

Scalp EEG was acquired with a 128-channel BioSemi Active 2 system (BioSemi, Amsterdam, Netherlands), with electrodes including the standard 10–20 system locations and additional intermediate positions (http://www.biosemi.com/headcap.htm, relabeled to more conventional labels of the 10–20 system, see Fig. S1 in Rossion et al., 2015). Continuous EEG was sampled at 512 Hz and electrode offset was held below ±50 μV. Eye movements were monitored with four electrodes placed at the outer canthi of the eyes and above and below the right eye. During testing, digital triggers were sent via a parallel port from the stimulation PC to the EEG acquisition PC at the start and the end of each sequence, at the start of each SF Step, and at the onset of each 12 Hz stimulation cycle (i.e., at the minima of the sinewave, or 0% contrast). A trigger was also sent when the participant responded, allowing the experimenter to monitor performance on the behavioral task online. The experimenter manually initiated recordings only after the participant showed a stable EEG trace (i.e., free from muscular and ocular artefact) for at least 5 s.

EEG preprocessing

EEG data analysis was carried out in Letswave5 (http://nocions.webnode.com/letswave) running on MATLAB R2012b (MathWorks, MA, United States). We band-pass filtered the continuous EEG data between 0.1 Hz–100 Hz (4th order zero-phase Butterworth filter) and then downsampled it to 256 Hz for easier handling and storage. The data was then segmented into 68 s epochs, including an extra 2 s before and after each trial. For each participant, we used independent component analysis (ICA) (Jung et al., 2000) with a square mixing matrix to remove the component corresponding to eye blinks (identified by inspecting the topographical distribution and the waveform). Next, we identified artefact-ridden channels by visual inspection and replaced them using linear interpolation of the 3 neighboring channels (less than 5% of channels for each participant, see Picton et al., 2000). Finally, we re-referenced the cleaned data to the average of the 128 scalp channels and averaged each participant's trials for *Face* and *Catch Trial* (No Face) sequences separately. The preprocessed data epochs were then cropped again according to each SF Step (14 × 4 s epochs) and subjected to two analyses in the frequency-domain.

EEG frequency-domain analysis

Face detection at the supra-threshold level

Since behavioral data indicated that participants detected faces well before SF Step 10 (see behavioral results in Fig. 2), we considered the face-selective EEG response within these "supra-threshold" SF Steps (SF Steps 10–14) to be representative of a typical face-selective response (see Retter and Rossion, 2016 for a description of a "full" face-selective response). As such, we averaged across the top five SF Steps and extracted the amplitude spectra from a Fast Fourier transform (FFT; frequency resolution = 0.25 Hz). We then used this supra-threshold EEG

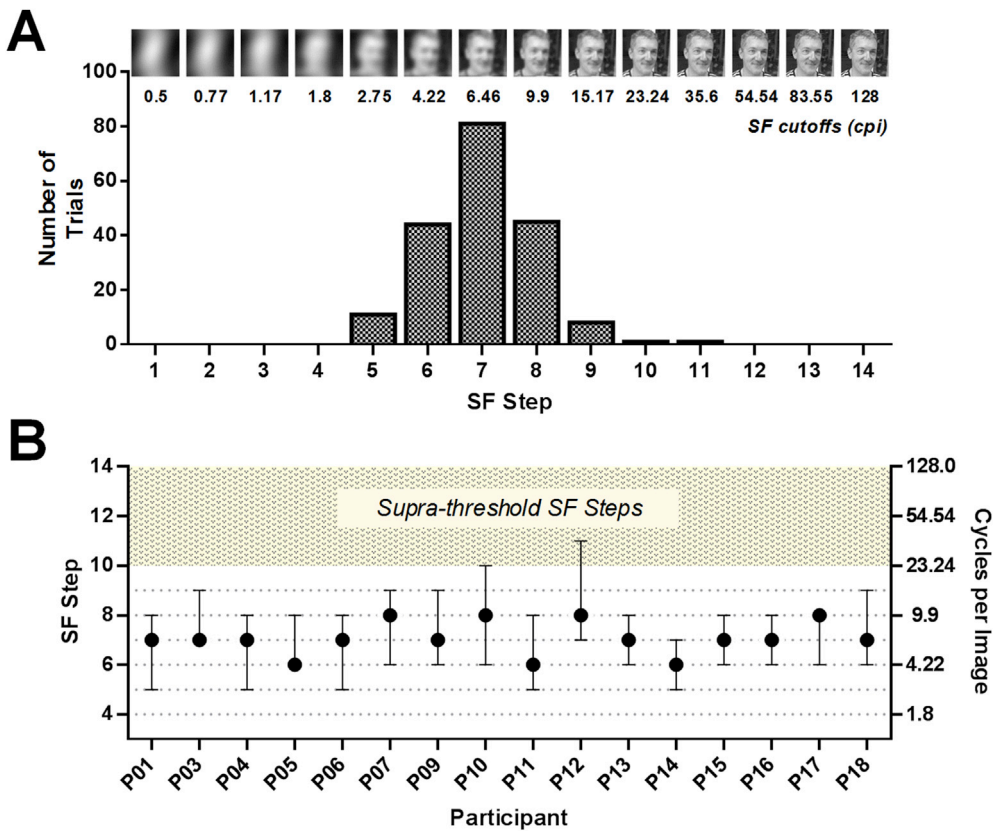


Fig. 2. Behavioral results. On each trial, participants had to press the spacebar as soon as they felt confident that the sequence contained face images. **A)** Frequency distribution of the behavioral responses across SF Steps. Example face images are given for each SF step. **B)** Individual data. Dots = the mode SF Step at which individuals indicated "faces present". Error bars are min-max range for each participant. Shaded region denotes supra-threshold SF range.

response to determine the relevant range of harmonics and regions-of-interest (ROIs) for the main analysis.

To determine the range of relevant harmonics to consider for the common and face-selective responses, we grand-averaged the amplitude spectra across all participants and all scalp channels, and transformed the resulting amplitude spectra into Z-scores: For each frequency bin, we subtracted the mean amplitude of the surrounding frequency bins (defined as three bins, i.e., 0.75 Hz, on either side of the bin of interest, excluding the immediately adjacent bins[4]) from the amplitude at the bin of interest, and divided the resulting value by the standard deviation of the surrounding frequency bins. Using a conservative threshold of $z > 2.57$ (i.e., $p < .005$, one-tailed, signal > noise), we identified four significant consecutive harmonics of the image presentation frequency (i.e., 12, 24, 36, & 48 Hz) and seven significant consecutive harmonics of the face presentation frequency (i.e., 1.5, 3, 4.5, 6, 7.5, 9, & 10.5 Hz) in the critical *Face* sequences. As expected, the *Catch trial* (No Face) sequences only yielded significant responses at harmonics of the image presentation frequency (i.e., 12, 24, 36, & 48 Hz). As a final step, we quantified the two responses of interest by summing baseline-corrected amplitudes (amplitude at each frequency harmonic minus mean amplitude of 6 surrounding frequency bins, see above) at the selected harmonics of each. Henceforth, mentions of the common response and the face-selective response will refer to these summed responses.

Since the scalp topography of the face-selective response typically varies across individuals (Quek and Rossion, 2017; Retter and Rossion, 2016; Rossion et al., 2015), we optimized our quantification of each participant's face-selective response by defining ROIs at the individual participant level. For each participant, the face-selective ROI was the average of the four channels with the strongest supra-threshold face--selective response. Although these channels were not necessarily contiguous, they were generally located within the same posterior lateral/medial occipital regions within the same hemisphere (except for 4/16 participants for whom the response was bilateral). We applied the same approach to determine the individual ROIs for the common response. Importantly, there was very little overlap between each participant's common response ROI and face-selective ROI; only two participants had one electrode which contributed to both ROIs.

Statistical analyses on these supra-threshold face-selective and common responses were carried out using planned pairwise *t*-tests with Bonferroni-adjusted alphas.

Face detection as a function of increasing SF content

To evaluate the effect of increasing SF content on face detection, we extracted the frequency spectra corresponding to each of the 14 SF Steps for each participant and computed the common response and face-selective response within their respective ROIs. To estimate the EEG *detection threshold* (i.e., the SF Step at which the face-selective response first emerged), we computed the 95% confidence interval (CI, calculated across individual participants) around the mean face-selective response amplitude at each SF Step, and identified the step at which this 95% CI did not include zero. Conversely, the EEG *saturation threshold* was defined as the SF Step for which the 95% CI around the mean face-selective response overlapped with the mean face-selective response elicited by full-spectrum face images (see below).

Considering that the face-selective response in such dynamic visual stimulation designs is predominantly right-lateralized (Retter and Rossion, 2016; Rossion et al., 2015), we also contrasted the impact of

[4] Note that previous studies (Jacques et al., 2016; Retter and Rossion, 2016) have used a larger range of surrounding bins made possible by the longer stimulation sequences and therefore higher frequency resolution (0.017 Hz) than we have here (0.25 Hz).

increasing SF content on the face-selective response across the left and right hemispheres. Note that we do not aim to establish the emergence of a lateralization effect by directly comparing responses between hemispheres. Rather, the goal here was to evaluate the amount of SF content required by each hemisphere for face detection, using the approach as described above to define *detection* and *saturation thresholds*. To this end, we created symmetrical left and right ROIs by averaging four occipito-temporal channels on each side of the head (Fig. 4C). Note that these bilateral ROIs were defined based on group-level data and were identical across participants.

Face detection of full-spectrum faces

For 12 out of 16 participants, the main experiment was preceded by four *Face* sequences that contained unfiltered, full-spectrum versions of faces and objects. We used this complementary dataset to compare the response obtained in our paradigm manipulating SF to the magnitude of the face-selective response for images containing all available SFs (i.e., establish a baseline). Stimuli were identical to those used in the main experiment (i.e., greyscale, 256 × 256 pixels), save that here the images were unfiltered. Since participants would easily be able to detect faces in these full-spectrum sequences, here the task was instead to monitor the central fixation cross and to respond when it changed color (Quek and Rossion, 2017; Retter and Rossion, 2016; Rossion et al., 2015). Analysis of these full-spectrum sequences was highly similar to the main experiment, save that the entire sequence was treated as a single 56 s epoch, rather than being segmented into smaller 4 s epochs. Naturally this resulted in a higher frequency resolution (0.016 Hz) in the frequency-domain. Since the concordance between the main and full-spectrum experiments was very high, both in terms of the range of significant harmonics and topography of responses, we quantified the common and face-selective responses using the same range of harmonics and individual ROIs as in the main experiment. Responses were averaged over the four trials for each participant and then grand-averaged across participants.

Results

Face detection at the supra-threshold level

Given that participants' behavioral responses necessarily reflect the summation of perceptual, decisional, and motor processes – and since in the present context they could be initiated based on the appearance of a single salient face in the sequence – behavioral response latencies (relabeled with the corresponding SF Step during which the response occurred) cannot accurately estimate the face detection threshold (i.e., minimal amount of SF content required by the brain to discriminate faces and objects). Rather, they identify the SF steps during which participants were confident of the presence of a face (i.e., a "supra-threshold" SF range). Participants most often indicated "faces present" at SF Step 7, corresponding to 6.46 cpi (interquartile range = 4.22–9.9 cpi, see Fig. 2). Since participants typically responded "faces present" prior to SF Step 10, we defined SF Steps 10–14 as supra-threshold steps. Note that false positive responses were very rare (less than 4% of the total number of catch trials across all participants) and accuracies were therefore not analyzed.

Taking the supra-threshold SF steps identified in behavioral analysis (SF Steps 10–14), we inspected the neural response at both the image presentation frequency (12 Hz) and the face presentation frequency (1.5 Hz). Where the former reflects visual processing common to faces *and* objects, the latter is an objective marker for face detection, in that it captures systematic deviations from the neural response evoked by both stimulus types (Retter and Rossion, 2016; Rossion et al., 2015). Fig. 3A shows the grand-averaged frequency spectrum for the average of all 128 channels for both sequence types. Applying our pre-defined significance threshold of $z > 2.57$ (i.e., $p < .005$, one-tailed, signal > noise), both *Face* and *Catch trial* (No Face) sequences elicited significant responses at the first four harmonics of the image presentation frequency (i.e., 12, 24, 36, and 48 Hz). More importantly, *Face* sequences further elicited significant peaks at the first 7 harmonics of the face presentation frequency (i.e., 1.5, 3, 4.5, 6, 7.5, 9, and 10.5 Hz), reflecting significant perceptual categorization of faces amongst objects (de Heering and Rossion, 2015; Retter and Rossion, 2016; Rossion et al., 2015).[5] At the group-level, the *common response* (defined as the sum of the first 4 harmonics of 12 Hz) emerged over medial occipital electrode sites, whereas the *face-selective response* (defined as the sum of first 7 harmonics of 1.5 Hz) emerged over lateral occipito-temporal electrodes. Here we observed a right hemispheric advantage (see Fig. 3B), a signature of face-selective processing (Rossion et al., 2015). The face-selective response was identifiable for all 16 observers tested. For each participant, we defined the four scalp channels with the largest common response and the four channels with the largest face-selective response as regions-of-interest (ROIs). Fig. 3C presents the group-averaged common and face-selective responses in the two ROIs. For the common response, we were specifically interested in whether the *Face* and *Catch trial* (No Face) sequences would elicit different response magnitudes, since this could suggest participants attended differently to the two sequence types. This was not the case in either ROI (common ROI: $t(15) = .349$, $p = .732$, $d = .09$; face-selective ROI: $t(15) = .201$, $p = .844$, $d = .05$; Bonferroni adjusted $\alpha = .025$). As expected from the design, *Face* sequences elicited a significantly larger face-selective response than did the *Catch trial* (No Face) sequences in both ROIs (common ROI: $t(15) = 7.07$, $p < .0001$, $d = 1.77$; face-selective ROI: $t(15) = 6.77$, $p < .0001$, $d = 1.69$; Bonferroni adjusted $\alpha = .025$).

Face detection as a function of increasing SF content

Having established the presence and scalp topography of a supra-threshold face-selective response, we next examined how this response emerged as a function of SF content. As shown in Fig. 4A, the face-selective response gradually increased with increasing SF information mainly over lateral (right) occipito-temporal regions. Averaging the face-selective response within each participant's unique face-selective ROI, the face detection threshold (defined as the first SF Step at which the 95% CI did not include zero) was located at SF Step 6, corresponding to 4.22 cpi (or 1.66 cycles per face width, cpf, see Fig. 4B). Importantly, due to the periodic nature of the face-selective response, this detection threshold does not reflect detection of a single salient face exemplar, but rather the point at which the brain is able to detect multiple variable faces within the image stream (thereby generating a periodic EEG response). As such, this estimate reflects face detection that has reached a certain level of stability in the context of highly variable stimulus images (see Fig. 6 for examples of faces at the detection threshold of 4.22 cpi).

Interestingly, the face-selective response profile we observed did not saturate immediately after onset, but rather continued to increase across sequential SF Steps before reaching plateau (see Fig. 4B). To determine the point of saturation, we contrasted the face-selective response at each SF Step with the face-selective response elicited by a full-spectrum version of the experiment (tested on 12/16 participants, run before the main experiment[6]). The full-spectrum experiment closely resembled the *Face* condition of the main experiment, save that here the four

[5] Harmonic responses of frequencies of interest arise since the periodic EEG response is not purely sinusoidal, but rather a complex, multi-component waveform (Norcia et al., 2015; Regan, 1989). For further discussion of the quantification of harmonic responses see (Retter and Rossion, 2016).

[6] We also ran the same analysis with the 12 participants who completed both the main and full-spectrum experiments. Detection (4.22 cpi) and saturation thresholds (35.36 cpi) identified in this secondary analysis were very similar to those identified in the primary analysis (including all 16 participants in the main experiment), leading to the same overall conclusions. We therefore report only the latter analyses.

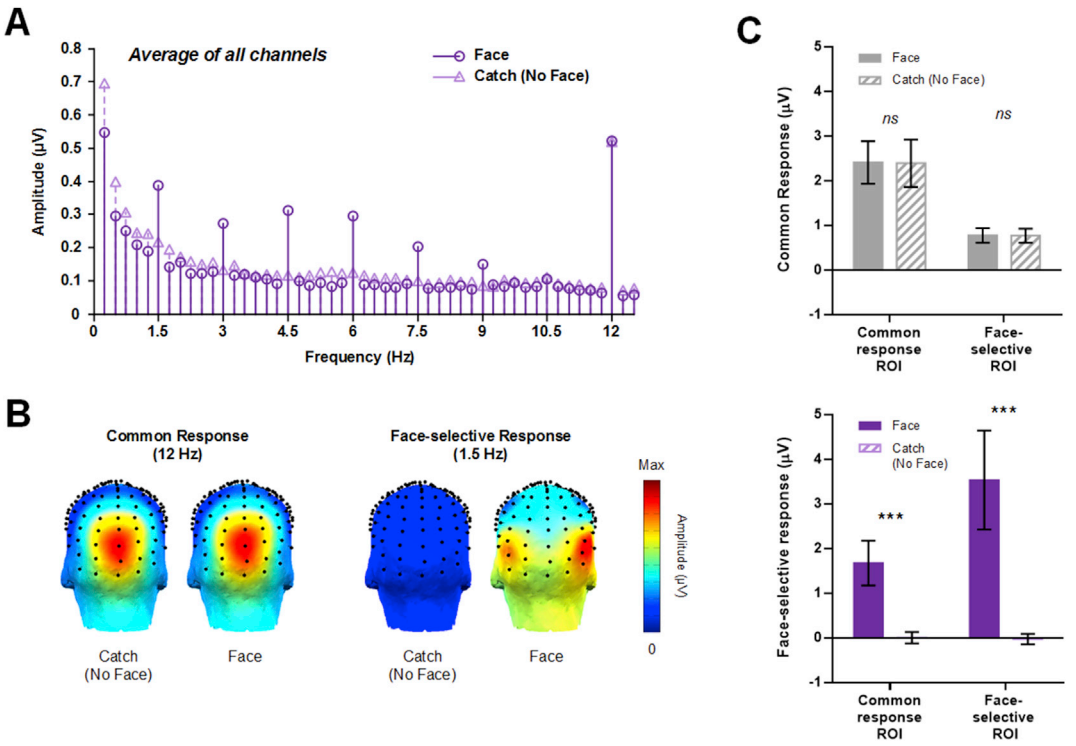


Fig. 3. Frequency-domain data for the supra-threshold SF Steps (Steps 10–14). **A)** Grand-averaged FFT amplitude spectra (averaged across all 128 channels). Both sequence types gave rise to a strong response at the image presentation frequency (12 Hz), however only *Face* sequences generated significant responses at the face-presentation frequency (1.5 Hz) and its harmonics. **B)** Group-level scalp topographies for the common and face-selective responses. **C)** Common and face-selective response magnitudes shown as a function of Sequence Type and ROI. Error bars are 95% confidence intervals (***$p < .0001$).

stimulation sequences contained only full-spectrum images (i.e., no spatial filtering, for similar examples see Retter and Rossion, 2016). Using the same 95% CI approach described above, we located the saturation threshold of the face-selective response at SF Step 12, or 54.54 cpi (21.41 cpf). Hence, while a minimum of 4.22 cpi was sufficient for consistent face detection response, the visual system continued to integrate relevant SF content until face images contained a higher level of detail with at least 54.54 cpi (see Fig. 6 for face images at the saturation threshold of 54.54 cpi). Importantly, the sigmoidal shape of this response function was not a consequence of averaging across step-like profiles with differing detection thresholds. Rather, all participants exhibited a similar gradually increasing face-selective response profiles (Fig. 5). In contrast, the face-selective response for *Catch trial* (No Face) sequences did not systematically depart from the zero baseline at any point in the stimulation sequence. Note that the common response, reflecting general visual processing of both faces and objects, also increased gradually as a function of SF content, although the profile was less steep and saturated earlier at 23.24 cpi (see Fig. S2).

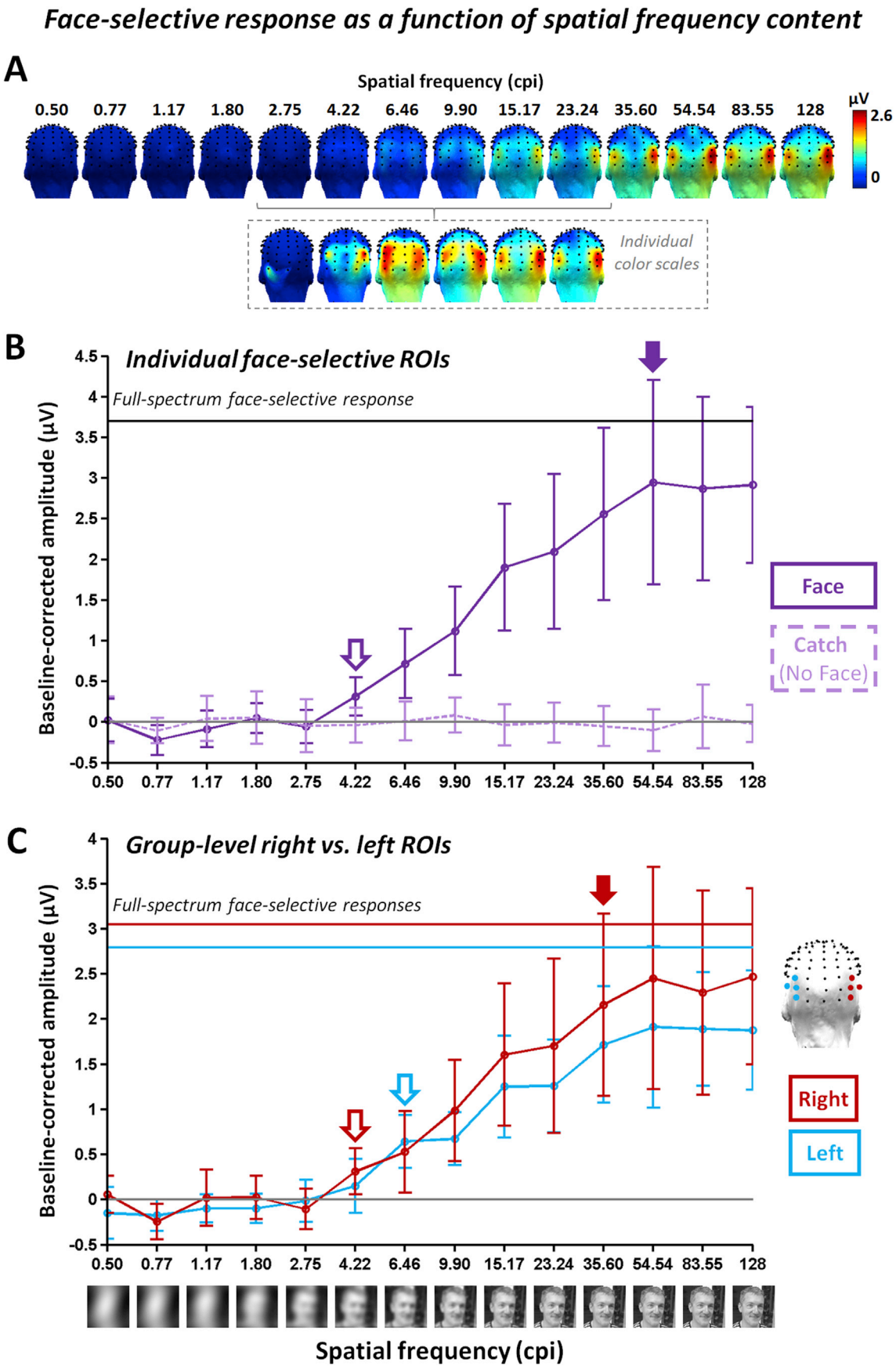
Given the right lateralization of the face-selective response at both the group-level (Fig. 4A) and in the majority of individual participants (Fig. 5), we further dissociated the right and left hemisphere response functions (note that here the bilateral ROIs were composed of the same channels for all participants). As is evident from the response topography in Fig. 4A and the response functions in Fig. 4C, the face-selective response emerged at an earlier SF Step over the right (4.22 cpi) relative to the left hemisphere (6.46 cpi). Moreover, the right hemisphere response saturated at 35.60 cpi while the left hemisphere response never quite reached saturation threshold. These results further underline the fundamental role of the right hemisphere in high-level face perception.

Discussion

Here we employed a paradigm that directly indexes face-selective neural responses to assess how SF content drives perceptual face categorization/face detection. Our data establish two important findings. First, even in a dynamic, highly variable visual environment, the human brain can detect briefly presented faces among other objects on the basis of very coarse visual information, i.e., minimal SF content of 4.22 cpi. At this threshold, individual local face parts are virtually invisible, such that faces are discriminated from objects based on only their distinctive global structure (Fig. 6). Second, the neural face detection response continued to increase well beyond the detection threshold, reaching saturation at a much finer spatial resolution of 54.54 cpi. Thus, while very low SFs are sufficient to achieve single glance face (vs. object) categorization, reaching a full, stable face-selective response requires the progressive extraction of additional higher SF content.

Face detection based on ultra-coarse visual information

That face detection emerged here on the basis of very degraded visual information, i.e., SF content <5 cpi (corresponding here to < 0.5 cpd; Fig. S1D), is consistent with previous studies highlighting the importance of low SFs for face detection (Goffaux et al., 2003a, 2011; Owsley and Sloane, 1987). This finding is also in agreement with fast face detection observed in 4–6 months old infants (de Heering and Rossion, 2015), for whom only coarse visual information is available at that stage of development (i.e., visual acuity threshold of ~2 cycles per degree, Peterzell et al., 1995). Importantly, this immature face-selective response in babies is strongly right-lateralized, a hallmark of specialised face processing (Bouvier and Engel, 2006; Jonas and Rossion, 2016; Sergent et al., 1992). Hemispheric lateralization is similarly evident in our data, in that face-selective responses emerged based on coarser information (4.22 cpi) over the right hemisphere than over the left hemisphere (6.46 cpi, see Fig. 4C). This observation thus provides rare and original support for the long-standing view that the right hemispheric specialisation for faces, which emerges early in development (Adibpour et al., 2018; de Heering



(caption on next page)

Fig. 4. The face-selective neural response as a function of increasing SF content. **A)** Top row: Topographical distribution of the face-selective response as a function of SF Step in the *Face* sequences, displayed on a common color scale. Bottom row: face-selective response topography at SF Steps near the detection threshold, displayed on individual color scales reflecting the maximum of each SF Step. **B)** The face-selective response, averaged within individually-defined face-selective ROIs, as a function of SF Step. Error bars represent 95% CIs. We defined the detection threshold (open arrow) as the first SF Step at which the 95% CI did not include zero. The saturation threshold (filled arrow) was the point at which the 95% CI included the face-selective response elicited by full-spectrum (unfiltered) images (solid black line). **C)** The face-selective response as a function of SF Step shown separately for the left and right hemispheres (channels composing the ROIs shown on the right, identical for all participants). Note that the detection threshold was earlier in the right hemisphere (red open arrow) than the left hemisphere (blue open arrow). Responses over the left hemisphere also do not appear to quite saturate.

and Rossion, 2015; de Schonen et al., 1989), is linked to low SF information sensitivity (Sergent and Weiskrantz, 1988).

What kind of information carried at this low spatial scale (i.e., <5 cpi) does the visual system rely on to detect faces amongst objects? From an image statistics perspective, natural face images contain more energy in the lower SF bands than do object images (see Fig. S1B, also Torralba and Oliva, 2003). However, while some studies have suggested that spectral amplitude differences alone drive fast face vs. object categorization, this has only been demonstrated for binary discriminations (e.g., face vs. car, Crouzet and Thorpe, 2011; Honey et al., 2008). In contrast, the face detection response recorded here reflects a selective neural response resulting from multiple comparisons to a wide range of object categories, both living and non-living (Fig. 1; Fig. S1A). Hence, this response most likely depends on phase information (i.e., structure) at this low spatial scale. Indeed, when such a large number of variable visual categories are presented in a full-spectrum dynamic paradigm as used here, phase-scrambling the images eliminates the occipito-temporal face detection responses (de Heering and Rossion, 2015; Gao et al., 2017; Gentile et al., 2017; Rossion et al., 2015). Assuming it is indeed structural information at this coarse scale that underpins face detection/categorization, an outstanding question is whether structural cues at a *local* (i.e., features) or *global* level are the most relevant for this visual discrimination. Given that local facial features (e.g., an eye, a nose, a mouth, etc.) are not clearly distinguishable in the images at the detection threshold (see Fig. 6) – and would be even less discernible at the rapid presentation rate we use – it would seem that at the point of emergence, face detection depends predominantly on the global face structure (Goffaux and Rossion, 2006; Sergent Elliset al, 1986). While our findings do not rule out the role of local face parts for detecting faces, they do suggest that such information might be less diagnostic and robust than global face structure in circumstances where the image quality or the visual input is impoverished, for example, in the visual periphery (Boucart et al., 2016) or at a distance. This conclusion has important implications for artificial face detection systems, which indeed are often based on specific feature detectors (i.e., eye, nose, mouth, hair) and known to struggle with degraded or very small face images (Yang et al.; Zhang and Zhang, 2010).

An interesting possibility regarding the global structural information contained at the coarse scale around the detection threshold, is that it captures the minimum global luminance variation necessary for perceiving internal face-like structure. In our design, <5 cpi corresponds to approximately 1.5 cpf (see Fig. S1D), which could be sufficient to evoke a global dark-light-dark pattern representing contrasting facial features (e.g., eye-nose-eye or eyes-nose-mouth). The view that ordinal contrast relationships between facial features are critical for face perception has been suggested previously (Dakin and Watt, 2009; Gilad et al., 2009). Indeed, not only does reversing the contrast relationships in face images hamper face detection (Lewis and Edmonds, 2003; Liu-Shuang et al., 2015), but this deleterious effect is particularly salient at lower SFs (Hayes et al., 1986). Although this is beyond the scope of this study, whether such internal contrast relationships do indeed drive face detection, and whether the face detection system is tuned particularly to contrast patterns in one cardinal direction (i.e., horizontal vs. vertical, Goffaux and Dakin, 2010; Goffaux et al., 2016), are testable hypotheses that warrant further investigation with the approach introduced here.

Since effective perceptual categorization in real world vision demands both speed and accuracy despite complex and varied visual input, a valid and relevant estimate of the minimum SF content capable of driving face detection in the human brain can only be obtained if face categorization processes are measured whilst subject to these same demands. By using a rapid image presentation rate (Retter and Rossion, 2016), we were able to both mimic the impressive speed of human face detection (Crouzet et al., 2010; Rousselet et al., 2003) and present a very large number of individual exemplars from many different categories, approximating the richness and categorical diversity of real world visual environments. In this context, using a large set of highly variable faces is critical, as it ensures that the periodically evoked 1.5 Hz response reflects consistent detection of faces despite broad variability in lighting, position, viewpoint, size, age, sex, etc. (i.e., an invariant response to multiple exemplars). As such, the electrophysiological face detection threshold reported here cannot have arisen based on the detection of a single, particularly noticeable face, but rather reflects the minimum SF content required for *consistent* perceptual discrimination of highly variable faces amongst many different object categories. This is in contrast with behavioral detection thresholds, which are vulnerable to influence from both salient exemplars and observer confidence thresholds related to task and instruction (e.g., "Respond as fast as possible when you see a face" vs. "Respond when you are sure you see a face").[7] Moreover, since faces and objects in our design remain embedded in their natural backgrounds (rather than isolated against artificial or homogenous backgrounds, Collin et al., 2012; Goffaux et al., 2003a), the face detection threshold estimated here also encompasses the complex process of figure-ground segregation required in natural vision (Regan, 2000).

Contribution of higher SFs to face detection

While extremely coarse SF content was sufficient to drive face detection, the face-selective response observed here continued to increase until images reached a much higher resolution (>50 cpi, or >6 cpd and >20 cpf; Fig. S1D), suggesting that the cortical visual system progressively integrates additional SF content. The saturation point we estimate here indicates the contribution of a higher range of SFs than has previously been reported for face detection (e.g., 22.63 cpi in Collin et al., 2012). This difference may relate to how we evaluated the contribution of SFs using a cumulative SF manipulation, instead of considering each SF band separately. However, we would argue that examining the role of higher SFs in the context of lower SFs, rather than in isolation (as is the case in bandpass filtering), is a more ecological approach for understanding how spatial frequency processing underpins perception of high-level stimuli. That is, while the visual system never encounters high-pass filtered stimuli in the real world, it regularly deals with low-resolution stimuli that gradually increase in SF content (e.g., a face seen in the distance will be blurry, but will become clearer/sharper as you approach it). Moreover, fine lines that define contours of facial features may have a stronger influence on perception when overlapped with low and middle SFs that define global structure, compared to when they are shown against a uniform background.

Since we manipulated SF content through cumulative low-pass

[7] Note that although a slower presentation rate or a more homogenous face stimulus set (e.g., segmented images with the same viewpoint/position/size) could perhaps be associated with an even lower detection threshold, the relevance of such a threshold to perceptual categorization of faces in more demanding, naturalistic settings would be unclear.

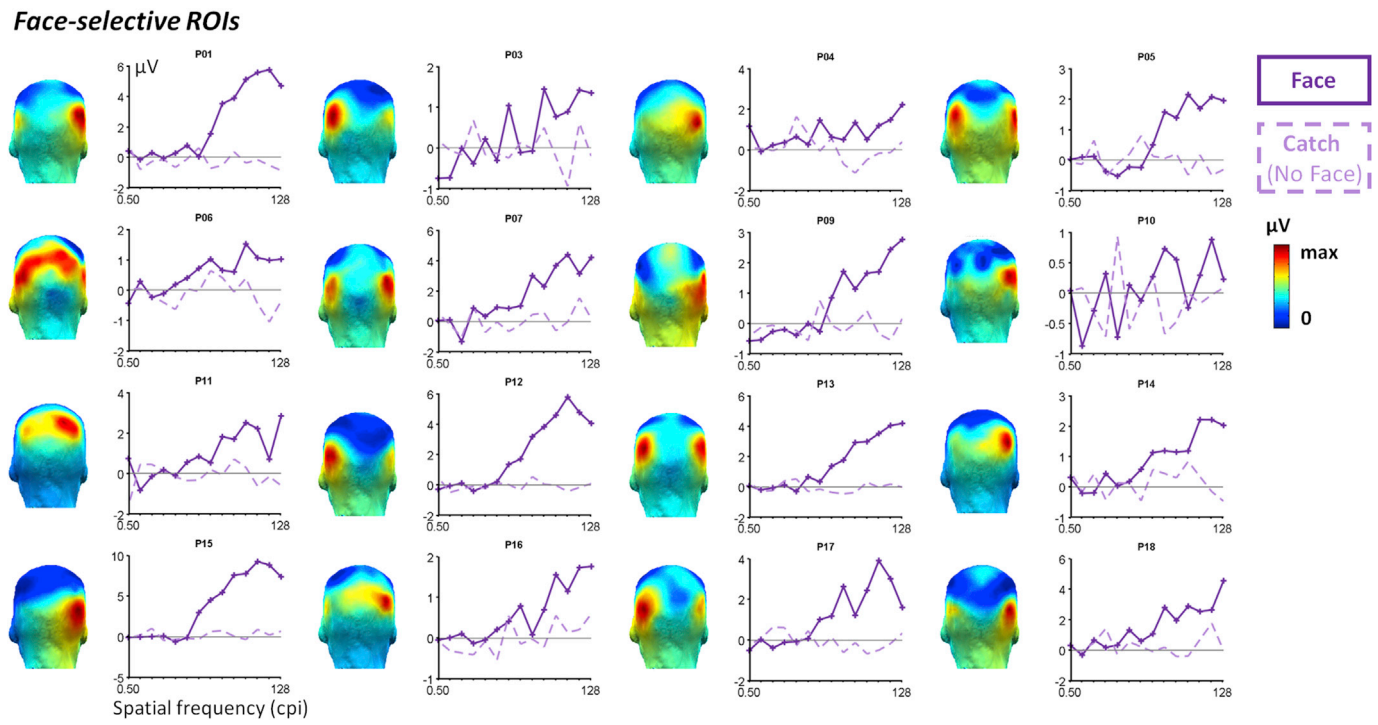


Fig. 5. The face-selective response function for each individual participant. Topographies on the left reflect each participant's face-selective response averaged across the supra-threshold SF Steps (used to define individual face-selective ROIs, see Methods). The line-plots show the face-selective response as a function of SF Step averaged within each participant's face-selective ROI. Solid/dashed lines are response profiles elicited by the *Face* and *Catch trial* (No Face) sequences respectively; the latter provides a reference for individual background EEG noise. Note that despite large variability in response topography and magnitude, participants showed comparable detection thresholds and a similar gradual increase of the face-selective response as a function of SF content.

filtering (rather than bandpass filtering), our approach also simulates the coarse-to-fine processing fundamental to high-level visual perception (Goffaux et al., 2011; Hegdé, 2008; Morrison and Schyns, 2001; Rossion, 2014; Sergent Elliset al, 1986). Within this framework, our results suggest that the human brain is capable of extracting higher SF content contained in the later part of the stimulation sequence within the very brief presentation duration of each image (83.33 ms). Again, this property of high SF processing might only be evident when lower SF face information is already present. Related to this, we cannot exclude the possibility that the extraction of even higher SF content (>50 cpi) requires a longer presentation duration than the current one (Goffaux et al., 2011). If so, running the current experiment at a slower temporal frequency (e.g., 6 Hz, as in de Heering and Rossion, 2015; Rossion et al., 2015) could potentially show both a continued increase of the face-selective response to filtered images until the last SF Step (128 cpi), as well as increased face-selective responses to full-spectrum images.

A relevant question here is why the visual system would continue to accumulate face-selective information within higher SF bands beyond the minimal SF content necessary for face detection. One possibility in this regard is that such evidence accumulation may progressively fine-tune the matching between degraded face inputs (especially those in more atypical positions/viewpoints) to internal face templates. In this context, these higher SFs may carry critical information for the extraction of facial emotion or identity (e.g., Ramon et al., 2015).

Caveats & limitations

It should be noted that we do not claim that the detection and saturation thresholds reported here are specific to faces as a category. It could be that these threshold values and the gradual increase until saturation might still be present if another category (e.g., houses or limbs, see Jacques et al., 2016) were presented periodically in the stimulation sequences. Indeed, it would be interesting to compare these characteristics of SF sensitivity across ecologically relevant categories. However, given that the response we measure here is, by design, *selective* only for faces, this possibility is beyond the scope of the present study.

One possible limitation of the design here concerns task specificity. The electrophysiological threshold was obtained while participants completed an explicit behavioral face detection task ("*Respond as soon as you think this sequence contains faces*"). Nevertheless, we would fully expect that the main conclusion – that faces can be detected at extremely coarse scales based on global structure – would still be valid even in the absence of such an explicit task. Indeed, in line with previous work, the results from our full-spectrum experiment show that a face-related task in itself is not necessary to obtain a robust measure of face categorization (Jacques et al., 2016; Quek and Rossion, 2017; Quek et al., 2018; Retter and Rossion, 2016; Rossion et al., 2015). More importantly, existing evidence suggests that explicitly attending to faces mainly engages non-essential, supplementary processes, but has little effect on the core processes involved in face detection (Quek et al., 2018). Indeed, selective attention to faces predominantly increases response amplitude over the left, rather than the right, hemisphere (Quek et al., 2018). Hence, while we would predict that the electrophysiological response profile and detection threshold would be largely similar in the context of an orthogonal behavioral task, we could potentially observe stronger right lateralization effects (i.e., larger offset in the detection threshold between right and left hemispheres).

Conclusions

The present findings undermine the view that face-selectivity in the human brain begins with identification of local features. Indeed, the face detection response emerged here based on coarse global facial structure alone, i.e., in the absence of distinguishable local facial features, and gradually increased with finer detail integration. These findings shed new light on both face detection and high-level perceptual categorization

Fig. 6. Twelve example faces shown at the detection (left) and response saturation thresholds (right). Note that at the point in the sequence at which face detection emerges, the global face configuration is evident where individual features are not. During the experiment, images were viewed from 80 cm at a size of 12.7 × 12.7 cm, such that each subtended 9.08° visual angle. See Fig. S3 in supplemental materials for examples of object images at the detection and saturation thresholds (cpi = cycles per image; cpf = cycles per median face width).

in general, and open up multiple avenues for future investigation. The ultra-coarse face detection threshold reported here agrees with recent evidence for fast (i.e., single-glance) face detection in human infants (de Heering and Rossion, 2015), which is presumably based on very low SF content. If so, then testing infants with a passive viewing task (as in de Heering and Rossion, 2015) should yield a similar detection threshold to that reported here, but a much lower saturation threshold that increases as sensitivity to higher SFs develops with age. Our approach is well-suited for investigating such developmental trajectories, as it can capture perceptual processing even in the absence of an explicit face-related task, enabling the same quantification of face detection processes from infancy through to adulthood. Moreover, since the approach enjoys a very high signal-to-noise ratio, it could also readily be applied to a more in-depth examination of inter-individual variability for face detection in adults. Finally, the response profile we observed here, in which coarse SFs are capable of driving the face detection response, while finer details progressively refine it, could serve to both inspire and constrain computational models of face detection (Scheirer et al., 2014).

Acknowledgements

The authors thank Andrea Conte for assistance with programming and Dr. Caroline Michel for assistance with collecting the stimuli. This work was supported in part by a European Research Council grant awarded to BR [grant no: facessvep 284025], in part by an FSR-FNRS postdoctoral grant awarded to JLS [grant no: FC 91608], and in part by a co-funded initiative by the University of Louvain and the Marie Curie Actions of the European Commission awarded to GLQ [grant no: F211800012]. VG is supported by the Belgian National Fund for Scientific Research (FSR-FNRS).

Appendix A. Supplementary data

Supplementary data related to this article can be found at https://doi.org/10.1016/j.neuroimage.2018.04.034.

References

Adibpour, P., Dubois, J., Dehaene-Lambertz, G., 2018. Right but not left hemispheric discrimination of faces in infancy. Nat. Hum. Behav. 2 (1), 67–79.

Ales, J.M., et al., 2012. An objective method for measuring face detection thresholds using the sweep steady-state visual evoked response. J. Vis. 12 (10), 18–18.

Besson, G., et al., 2017. From face processing to face recognition: comparing three different processing levels. Cognition 158, 33–43.

Boucart, M., et al., 2016. Finding faces, animals, and vehicles in far peripheral vision. J. Vis. 16 (2), 10.

Bouvier, S.E., Engel, S.A., 2006. Behavioral deficits and cortical damage loci in cerebral achromatopsia. Cereb. Cortex 16 (2), 183–191.

Caharel, S., et al., 2013. Early holistic face-like processing of Arcimboldo paintings in the right occipito-temporal cortex: evidence from the N170 ERP component. Int. J. Psychophysiol. 90 (2), 157–164.

Campbell, F., Cooper, G.F., Enroth-Cugell, C., 1969. The spatial selectivity of the visual cells of the cat. J. Physiol. 203 (1), 223.

Carlson, T., et al., 2013. Representational dynamics of object vision: the first 1000 ms. J. Vis. 13 (10), 1.

Cichy, R.M., Pantazis, D., Oliva, A., 2014. Resolving human object recognition in space and time. Nat. Neurosci. 17 (3), 455–462.

Collin, C.A., et al., 2012. Effects of band-pass spatial frequency filtering of face and object images on the amplitude of N170. Perception 41 (6), 717–732.

Costen, N.P., Parker, D.M., Craw, I., 1994. Spatial content and spatial quantisation effects in face recognition. Perception 23 (2), 129–146.

Crouzet, S.M., Thorpe, S.J., 2011. Low-level cues and ultra-fast face detection. Front. Psychol. 2, 342.

Crouzet, S.M., Kirchner, H., Thorpe, S.J., 2010. Fast saccades toward faces: face detection in just 100 ms. J. Vis. 10 (4), 16–16.

Dakin, S.C., Watt, R.J., 2009. Biological "bar codes" in human faces. J. Vis. 9 (4), 2.1-10.

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

de Schonen, S., Mathivet, E., Deruelle, C., 1989. Hemispheric specialization for face recognition in infancy. In: Neurobiology of Early Infant Behaviour. Springer, pp. 261–271.

De Valois, R.L., De Valois, K.K., 1980. Spatial vision. Annu. Rev. Psychol. 31, 309–341.

Gao, X., Maurer, D., 2011. A comparison of spatial frequency tuning for the recognition of facial identity and facial expressions in adults and children. Vis. Res. 51 (5), 508–519.

Gao, X., Gentile, F., Rossion, B., 2017. Fast periodic stimulation (FPS): a highly effective approach in fMRI brain mapping. bioRxiv.

Gaspar, C., Sekuler, A.B., Bennett, P.J., 2008. Spatial frequency tuning of upright and inverted face identification. Vis. Res. 48 (28), 2817–2826.

Gava, L., et al., 2008. Effect of partial occlusion on newborns' face preference and recognition. Dev. Sci. 11 (4), 563–574.

Gentile, F., Ales, J., Rossion, B., 2017. Being BOLD: The neural dynamics of face perception. Hum. Brain Mapp. 38 (1), 120–139.

Gilad, S., Meng, M., Sinha, P., 2009. Role of ordinal contrast relationships in face encoding. Proc. Natl. Acad. Sci. 106 (13), 5353–5358.

Ginsburg, A.P., et al., 1984. Large-sample norms for contrast sensitivity. Optom. Vis. Sci. 61 (2), 80–84.

Goffaux, V., Dakin, S.C., 2010. Horizontal information drives the behavioral signatures of face processing. Front. Psychol. 1, 143.

Goffaux, V., Rossion, B., 2006. Faces are "spatial" - holistic face perception is supported by low spatial frequencies. J. Exp. Psychol. Hum. Percept. Perform. 32 (4), 1023–1039.

Goffaux, V., et al., 2003a. ERP evidence for task modulations on face perceptual processing at different spatial scales. Cognit. Sci. 27 (2), 313–325.

Goffaux, V., Gauthier, I., Rossion, B., 2003b. Spatial scale contribution to early visual differences between face and object processing. Brain Res. Cogn. Brain Res. 16 (3), 416–424.
Goffaux, V., et al., 2011. From coarse to Fine? Spatial and temporal dynamics of cortical face processing. Cereb. Cortex 21 (2), 467–476.
Goffaux, V., et al., 2016. Horizontal tuning for faces originates in high-level Fusiform Face Area. Neuropsychologia 81, 1–11.
Gold, J., Bennett, P.J., Sekuler, A.B., 1999. Identification of band-pass filtered letters and faces by human and ideal observers. Vis. Res. 39 (21), 3537–3560.
Goold, J.E., Meng, M., 2016. Visual search of mooney faces. Front. Psychol. 7 (155).
Halit, H., et al., 2006. Is high-spatial frequency information used in the early stages of face detection? Brain Res. 1117 (1), 154–161.
Hayes, T., Morrone, M.C., Burr, D.C., 1986. Recognition of positive and negative bandpass-filtered images. Perception 15 (5), 595–602.
Hegdé, J., 2008. Time course of visual perception: coarse-to-fine processing and beyond. Prog. Neurobiol. 84 (4), 405–439.
Hershler, O., Hochstein, S., 2005. At first sight: a high-level pop out effect for faces. Vis. Res. 45 (13), 1707–1724.
Honey, C., Kirchner, H., VanRullen, R., 2008. Faces in the cloud: Fourier power spectrum biases ultrarapid face detection. J. Vis. 8 (12), 9.1-13.
Itier, R.J., et al., 2007. Early face processing specificity: it's in the eyes! J. Cogn. Neurosci. 19 (11), 1815–1826.
Jacques, C., Retter, T.L., Rossion, B., 2016. A single glance at a face generates larger and qualitatively different category-selective spatio-temporal signatures than other ecologically-relevant categories in the human brain. NeuroImage 137, 21–33.
Jonas, J., Rossion, B., 2016. Beyond the core face-processing network: intracerebral stimulation of a face-selective area in the right anterior fusiform gyrus elicits transient prosopagnosia. J. Vis. 16 (12), 385–385.
Jung, T.P., et al., 2000. Removal of eye activity artifacts from visual event-related potentials in normal and clinical subjects. Clin. Neurophysiol. 111 (10), 1745–1758.
Khalid, S., et al., 2013. Subcortical human face processing? Evidence from masked priming. J. Exp. Psychol. Hum. Percept. Perform. 39 (4), 989–1002.
Lewis, M.B., Edmonds, A.J., 2003. Face detection: mapping human performance. Perception 32 (8), 903–920.
Liu-Shuang, J., et al., 2015. The effect of contrast polarity reversal on face detection: evidence of perceptual asymmetry from sweep VEP. Vis. Res. 108, 8–19.
Morrison, D.J., Schyns, P.G., 2001. Usage of spatial scales for the categorization of faces, objects, and scenes. Psychon. Bull. Rev. 8 (3), 454–469.
Näsänen, R., 1999. Spatial frequency bandwidth used in the recognition of facial images. Vis. Res. 39 (23), 3824–3833.
Nemrodov, D., et al., 2016. The time course of individual face recognition: a pattern analysis of ERP signals. NeuroImage 132, 469–476.
Norcia, A.M., et al., 2015. The steady-state visual evoked potential in vision research: a review. J. Vis. 15 (6), 4–4.
Ojanpää, H., Näsänen, R., 2003. Utilisation of spatial frequency information in face search. Vis. Res. 43 (24), 2505–2515.
Owsley, C., Sloane, M.E., 1987. Contrast sensitivity, acuity, and the perception of 'real-world' targets. Br. J. Ophthalmol. 71 (10), 791–796.
Paras, C.L., Webster, M.A., 2013. Stimulus requirements for face perception: an analysis based on "Totem poles". Front. Psychol. 4, 18.
Peterzell, D.H., Werner, J.S., Kaplan, P.S., 1995. Individual differences in contrast sensitivity functions: longitudinal study of 4-, 6-and 8-month-old human infants. Vis. Res. 35 (7), 961–979.
Picton, T.W., et al., 2000. Guidelines for using human event-related potentials to study cognition: recording standards and publication criteria. Psychophysiology 37 (2), 127–152.
Purcell, D.G., Stewart, A.L., 1988. The face-detection effect: Configuration enhances detection. Percept. Psychophys. 43, 355–366.
Quek, G.L., Rossion, B., 2017. Category-selective human brain processes elicited in fast periodic visual stimulation streams are immune to temporal predictability. Neuropsychologia 104, 182–200.
Quek, G.L., Nemrodov, D., Rossion, B., Liu-Shuang, J., 2018. Selective attention to faces in a rapid visual stream: hemispheric differences in enhancement and suppression of category-selective neural activity. J. Cognit. Neurosci. 30 (3), 393–410.
Ramon, M., et al., 2015. Neural microgenesis of personally familiar face recognition. Proc. Natl. Acad. Sci. 112 (35), E4835–E4844.
Regan, D., 1989. Human Brain Electrophysiology: Evoked Potentials and Evoked Magnetic fields in Science and Medicine. Elsevier, Amsterdam.
Regan, D., 2000. Human Perception of Objects. Sinauer Associates, Sunderland, MA, pp. 295–342.
Retter, T.L., Rossion, B., 2016. Uncovering the neural magnitude and spatio-temporal dynamics of natural image categorization in a fast visual stream. Neuropsychologia 91, 9–28.
Rossion, B., 2014. Understanding face perception by means of human electrophysiology. Trends Cognit. Sci. 18 (6), 310–318.
Rossion, B., et al., 2015. Fast periodic presentation of natural images reveals a robust face-selective electrophysiological response in the human brain. J. Vis. 15 (1), 15.1.18.
Rousselet, G.A., Mace, M.J., Fabre-Thorpe, M., 2003. Is it an animal? Is it a human face? Fast processing in upright and inverted natural scenes. J. Vis. 3 (6), 440–455.
Rousselet, G.A., et al., 2014. Eye coding mechanisms in early human face event-related potentials. J. Vis. 14 (13), 7.
Ruiz-Soler, M., Beltran, F.S., 2006. Face perception: an integrative review of the role of spatial frequencies. Psychol. Res. 70 (4), 273–292.
Scheirer, W.J., et al., 2014. Perceptual annotation: measuring human vision to improve computer vision. IEEE Trans. Pattern Anal. Mach. Intell. 36 (8), 1679–1686.
Schyns, P.G., Oliva, A., 1999. Dr. Angry and Mr. Smile: when categorization flexibly modifies the perception of faces in rapid visual presentations. Cognition 69 (3), 243–265.
Sergent, J., 1988. Face perception and the right hemisphere. In: Weiskrantz, L. (Ed.), Thought without Language. Oxford University Press, pp. 108–131.
Sergent, J., 1986. Microgenesis of face perception. In: Ellis, H.D., et al. (Eds.), Aspects of Face Processing. Springer Netherlands, Dordrecht, pp. 17–33.
Sergent, J., Ohta, S., Macdonald, B., 1992. Functional neuroanatomy of face and object processing A positron emission tomography study. Brain 115 (1), 15–36.
Torralba, A., Oliva, A., 2003. Statistics of natural image categories. Network 14 (3), 391–412.
Viola, P., Jones, M.J., 2004. Robust real-time face detection. Int. J. Comput. Vis. 57 (2), 137–154.
Vuilleumier, P., et al., 2003. Distinct spatial frequency sensitivities for processing faces and emotional expressions. Nat. Neurosci. 6 (6), 624–631.
Yang, S., et al., Faceness-Net: Face Detection through Deep Facial Part Responses. https://arxiv.org/abs/1701.08393, (pre print).
Zhang, C., Zhang, Z., 2010. A Survey of Recent Advances in Face Detection. Microsoft Research, Microsoft Corporation, Washington, USA.