The human brain rapidly detects faces in the visual environment. We recently presented a sweep visual evoked potential approach to objectively define face detection thresholds as well as suprathreshold response functions (Ales, Farzin, Rossion, & Norcia, 2012). Here we determined these parameters are affected by orientation (upright vs. inverted) and contrast polarity (positive vs. negative), two manipulations that disproportionately disrupt the perception of faces relative to other object categories. Face stimuli parametrically increased in visibility through phase-descrambling while alternating with scrambled images at a fixed presentation rate of 3 Hz (6 images/s). The power spectrum and mean luminance of all stimuli were equalized. As a face gradually emerged during a stimulation sequence, EEG responses at 3 Hz appeared at ≈35% phase coherence over right occipito-temporal channels, replicating previous observations. With inversion and contrast-reversal, the 3-Hz amplitude decreased by ≈20%–50% and the face detection threshold increased by ≈30%–60% coherence. Furthermore, while the 3-Hz response emerged abruptly and saturated quickly for normal faces, suggesting a categorical neural response, the response profile for inverted and negative polarity faces was shallower and more linear, indicating gradual and continuously increasing activation of the underlying neural population. These findings demonstrate that inversion and contrast-reversal increase the threshold and modulate the suprathreshold response function of face detection.
nals, reflecting an objective (i.e., at a predefined frequency) threshold for face detection. Moreover, despite the progressive increase of shape information, this 3-Hz increase was abrupt, reaching the maximal response amplitude within one level of additional phase-coherence, suggesting a strongly saturated response function for face detection. These findings are unique in the sense that they provide a sensitive (i.e., high signal-to-noise ratio, SNR) and objective (i.e., exactly at an experimentally defined frequency) threshold and response function for face detection (Ales et al., 2012).

The present study builds upon these observations to investigate how face detection is influenced by orientation and contrast polarity reversal. Since plane inversion (Yin, 1969) and contrast polarity reversal (i.e., black pixels become white and vice versa; Galper, 1970) are two well-documented manipulations that disproportionately impair perception for faces compared to objects (for inversion see Farah, Wilson, Drain, & Tanaka, 1998; Rossion, 2008, for reviews; for contrast-reversal see Nederhouser, Yue, Mangini, & Biederman, 2007; Vuong, Peissig, Harrison, & Tarr, 2005), a first aim is to use these manipulations to test whether the face detection threshold and response function reflect high-level visual processes specific to facial structure (face vs. scrambled), rather than more general (midlevel) structure detection mechanisms (structured vs. scrambled). It is in this sense that we will use the term face specific in the text that follows. For this purpose, orientation and contrast inverted faces are better comparisons than non-face object categories because they share the same low-level characteristics and structure as upright faces (i.e., rounded shape and internal features).

It should be noted that these behavioral effects of inversion and contrast-reversal have been most often reported in the context of individual face discrimination (i.e., the differentiation of facial identities, e.g., Freire, Lee, & Symons, 2000; Russell, Biederman, Nederhouser, & Sinha, 2007) while the effect of inversion and contrast-reversal on face detection (i.e., the perception of a face as a face) is less well described. Studies that have addressed this issue generally show that inversion or contrast-reversal slows down and reduces face detection accuracy (Garrido, Duchaine, & Nakayama, 2008; Lewis & Edmonds, 2003, 2005; Parkin & Williamson, 1987; Purcell & Stewart, 1986, 1988; Rousselet et al., 2003; VanRullen, 2006) relative to upright faces.

At the neural level, both inversion (e.g., Kanwisher, Tong, & Nakayama, 1998; Mazzard, Schiltz, & Rossion, 2006; Nasr & Tootell, 2012) and contrast-reversal (George et al., 1999; Nasr & Tootell, 2012; Rossion, Dricot, Goebel, & Busigny, 2011; Yue et al., 2013) reduce the amplitude of face-selective fMRI responses and increase the latency of the N170 component (Bentin et al., 1996; Jeffreys, 1993) and of SSVEPs elicited by faces (Liu-Shuang, Norcia, & Rossion, 2014; Rossion, Alonso Prieto, Boremanse, Kuefner, & Van Belle, 2012). However, the interpretation of a reduced response for highly suprathreshold inverted and negative polarity faces is ambiguous: Either the threshold for face detection is elevated for inverted/contrast reversed faces, or the detection threshold is the same for upright and inverted/contrast reversed faces but activation only rises to a lower level as one goes further above threshold. The second aim of this sweep VEP study is to disentangle these two possibilities. By presenting faces with systematically increasing levels of structural integrity, we are able to determine at what level of structure typical and inverted/contrast reversed faces deviate, rather than just quantify differences at the endpoint where face structure is fully visible. Here we use the coherence of the phase spectrum of the images to vary structural integrity. Using this method, we compare detection thresholds, which reflect the amount of information necessary to perceive a face, as well as suprathreshold response functions, which are indicative of how the incoming visual information activates the underlying neural population (e.g., gradually vs. abruptly).

To pursue the two goals of this study, we used sequences of gradually revealed faces, similarly to Ales et al. (2012), which varied here in orientation and contrast polarity. We then compared the effect of these manipulations on several aspects of the resulting SSVEP response: overall amplitude, amplitude across face visibility, shape of the response function, and response threshold. Based on previous studies, we formulated the following predictions. If our sweep...
paradigm even partially captures face-specific processing, we should observe a reduced overall amplitude of the 3-Hz response following inversion and contrast-reversal. Importantly, the 6-Hz response, reflecting low-level visual processing, should remain constant across conditions and face visibility levels if there are no general differences in arousal or attention across conditions. Furthermore, if inversion and contrast-reversal affect the amount of face structure information (% phase coherence) necessary to generate a 3-Hz response, we should observe differences between conditions in the estimated face detection thresholds. Alternatively, if rate of activation is the key difference across conditions, we could observe slower supra-threshold response increase for inverted or polarity reversed faces. A combination of both response elevation and changes in the shape of the face-related response function is also possible.

Methods

Participants

We tested 17 participants (eight males, mean age 23.35 ± 5.73), 15 received monetary compensation and two received course credit. Two participants were excluded due to below chance performance on the behavioral task, one participant was excluded due to amblyopia, and another was excluded due to excessive noise and artifacts in their signal. A total of 13 participants were kept for analyses (seven males, mean age 24.1 ± 6.32). None of the participants reported any history of psychiatric or neurological disorders, and had normal or corrected-to-normal vision. All gave written informed consent according to the guidelines of the Institutional Review Board of Stanford University.

Stimuli

Stimuli were composed of the same set that was used in the original face detection sweep study (Ales et al., 2012) as well as modified versions of this stimulus set. Detailed information on the creation of the original stimulus set can be found in aforementioned paper and will be briefly summarized here. A set of 15 grayscale faces varying in gender, viewpoint, and size were equalized in terms of their power spectra. Each face was then smoothly embedded at different locations on a background composed of the phase-scrambled average power spectrum (Figure 1A). Hence, all images had the exact same power spectrum. Face sequences
were generated by parametrically increasing the phase coherence of the embedded face from 0% to 100% in 5.26% steps. The resulting sequences showed a face gradually emerging from a scrambled background in 20 steps (see Figures 1C & D for an abridged version of the sequence with 10 steps). Complementary 100% scrambled sequences composed of phase-scrambled backgrounds with no faces (20 images) were also created. In both face and scrambled sequences the phase scrambling at each step was newly randomized. For the current experiment, we created three additional stimulus sets based on the original set with (a) inverted faces, (b) negative contrast polarity faces, and (c) inverted negative contrast polarity faces (Figure 1B). Inverted faces were created by rotating the images 180° around the x axis and negative polarity faces were created by inverting the contrast polarity around the mean luminance value of the images (127). The image dimensions were 512 × 512 pixels and subtended 11.44° of visual angle at a distance of 110 cm. Stimuli were shown on an 800 × 600 pixel CRT screen with a 72-Hz refresh rate.

Procedure

We presented five types of sweep sequences in the experiment: upright positive polarity faces, inverted positive polarity faces, upright negative polarity faces, inverted negative polarity faces, and scrambled images (Figure 1B). With the exception of the scrambled condition, each sweep sequence consisted of images from the scrambled sequence alternating with images from the face sequence at a rate of 3 Hz (6 images/s; Figure 1C). At each coherence step, a face and a scrambled image were repeated three times. In the scrambled condition, only scrambled images were shown (two different scrambled images per second). Sequences lasted 20 s, with one extra second of stimulation at the start and the end of the sequence (repetitions of the first and last steps, respectively) to exclude event-related potentials generated by the abrupt onset and offset of images. The 15 faces were shown once in each condition (15 × 5 = 75 sequences). The 75 sweep sequences were presented in pseudo-random order and the total duration of the experiment was 28 min, with short breaks every 10 min. Participants were instructed to fixate on a central fixation cross and to respond with a gamepad as rapidly and accurately as possible whenever they detected a face appearing in the sequence. They were informed that faces could appear at any location and with different viewpoints but that not all sequences contained faces. The experimenter verified that the instructions were well understood before starting the experiment.

EEG acquisition

EEG was acquired with a 128-channel Hydrocell Geodesic Sensor Net (Electrical Geodesics Inc., Eugene, OR, USA) at a sampling rate of 432 Hz. Electrode impedances were reduced below 60 kΩ prior to recording and readjusted during breaks. The data was band-pass filtered offline (0.3–50 Hz) and preprocessed with a custom software (PowerDiva). First, artifact rejection was applied by replacing channels containing more than 15% of data points exceeding a threshold of 30 μV with the average of their six neighboring channels. EEG data was then rereferenced to the average of all channels. Further artifact rejection consisted of excluding 1 s epochs on a channel-by-channel basis if they contained more than 10% of samples exceeding 30 μV. Entire 22-s trials were rejected if they contained blinks (> 7 channels exceeding 60 μV). For the SSVEP analysis, the first and last seconds were excluded.

Behavioral data analysis

Button press response times were converted into units of phase coherence by subtracting one second (= one step) and multiplying them by 5.26% (since at the first step, the coherence is at 0%). These estimated detection thresholds were not corrected for delays in the participants’ decision or motor response generation. Accuracy and detection thresholds were recorded and averaged both per participant and per face exemplar. Effects of conditions were assessed via repeated measures analyses of variance (ANOVA) with Greenhouse-Geisser correction was used to adjust degrees of freedom whenever the assumption of sphericity was violated. The behavioral detection thresholds were then correlated with their respective neural detection thresholds to investigate the relationship between these two measures.

SSVEP analysis

The SSVEPs were analyzed in two ways. First, a discrete Fourier Transform was applied to average data across the entire trial to extract the amplitude of the overall SSVEP response in each condition for individual participants. These values were grand averaged in the frequency domain to compute group-level frequency spectra. This analysis was used to assess the overall presence of 3-Hz responses reflecting face detection and of 6-Hz responses reflecting low-level visual processes. To isolate channels that were specifically involved in face processing, we computed an index of the 3-Hz response (3 Hz/3 Hz + 6 Hz). Based
on the spatial distribution of face-related responses, we then selected the channels composing regions of interest (ROIs) for further analyses.

Second, 3-Hz and 6-Hz response profile as a function of stimulus coherence were computed by a recursive least squares filter (RLS; Tang & Norcia, 1995). Compared to computing a discrete Fourier transform analysis on each step separately, the RLS method adaptively estimates the response phase from the recorded data, with a resulting improvement of SNR of 3 dB. To test for differences between conditions across each coherence level, we ran a permutation test against zero difference (null hypothesis) with all possible pairwise comparisons between conditions throughout all coherence levels. For each pairwise comparison, 213 total within-subject permutations of condition labels were carried out at every coherence level and a distribution of mean differences was created. Differences between conditions in the actual data were considered as significant if they were situated among the 5% extreme values in the distribution (one-tailed) and if the difference persisted over at least three consecutive coherence levels.

To determine the onset of the 3-Hz response relative to background EEG activity, i.e., the neural face detection threshold, we used the difference between integrated signal and noise functions (see Ales et al., 2012; see Figure 2). The signal was the cumulatively summed 3-Hz amplitude across the sweep sequence. Noise was the mean amplitude at two frequencies adjacent to 3 Hz (2.5 Hz and 3.5 Hz) averaged across the entire sequence and cumulatively summed 20 times. The cumulative amplitudes were normalized to the maximum value of the signal, thus reducing the range to 0%–100%. The normalized signal and noise functions were subtracted and a hinge function was fitted to the difference using a nonlinear least-square curve fitting algorithm (Levenberg-Marquardt) as implemented by the curve_fit method in the scipy Python toolbox. The hinge function was defined as a continuous piecewise linear function made of two parts that transition at a certain value of $x$, the first one being constant and the second one being linearly increasing. The onset of the 3-Hz response was defined as the inflection point of the hinge function. In case of a negative signal-noise difference, the threshold was set to 100% coherence (no face detection). These modified thresholds can be seen on Figure 9B (gray data points). In other cases when the algorithm’s estimate was unrealistic (i.e., < 10% coherence or negative coherence value), thresholds were manually set according to visual inspection (coherence at which the longest increasing segment of the difference function began rising above zero; see Supplementary Table). This second adjustment was applied to one participant in the left occipito-temporal ROI and seven face exemplars, four in the left occipito-temporal ROI (one for upright positive polarity faces and three for inverted negative polarity faces) and three in the right occipito-temporal ROI (one for upright positive polarity faces and two for inverted positive polarity faces). Overall, 92% of original threshold estimates were kept. This analysis was applied both across participants (averaged across 15 trials) and across face exemplars (averaged across 13 participants).

## Results

### Behavior

Mean behavioral detection thresholds derived from the button press task and expressed as percent coherence are shown on Figure 3. Mean accuracy rates are reported in Table 1. Both participant and face exemplar data show a decline in performance according

---

**Figure 2.** Schematic illustration of the estimation procedure of the 3-Hz response threshold. The amplitude was extracted at every coherence step at 3 Hz, representing the signal of interest (face detection), and at two adjacent frequencies (2.5 Hz and 3.5 Hz), representing the background EEG noise (far left). Both signal and noise were cumulatively summed and then normalized relative to the maximal value of the signal (middle). The difference between the signal and noise function (far right) can be used to estimate the stimulus value needed for response onset by fitting a hinge function to the data. The inflection point of the hinge function is then considered as the threshold of face detection.
The same trend was visible for the face exemplar analyses. A 2 × 2 repeated measures ANOVA with Orientation (upright vs. inverted) and Polarity (positive vs. negative) confirmed additive effects of these manipulations on response times for both participant, Orientation: $F(1, 12) = 25.6, p < 0.001, \eta^2 = 0.68$; Polarity: $F(1, 12) = 130.38, p < 0.001, \eta^2 = 0.92$, and face exemplar analyses, Orientation: $F(1, 14) = 16.34, p < 0.001, \eta^2 = 0.54$; Polarity: $F(1, 14) = 100.35, p < 0.001, \eta^2 = 0.88$. There were no significant Orientation × Polarity interactions, participant analysis: $F(1, 12) = 0.42, p = 0.53, \eta^2 = 0.03$; face exemplar data: $F(1, 14) = 0.83, p = 0.38, \eta^2 = 0.06$. For accuracy, there was a significant main effect of Polarity for participant analysis, $F(1, 12) = 5.66, p < 0.035, \eta^2 = 0.32$, reflecting decreased detection rates for negative polarity faces. The same trend was visible for the face-exemplar analysis, $F(1, 14) = 4.38, p = 0.055, \eta^2 = 0.24$. There were no effects of Orientation, participant analysis: $F(1, 12) = 1.68, p = 0.22, \eta^2 = 0.12$; face-exemplar analysis: $F(1, 14) = 3.03, p = 0.1, \eta^2 = 0.12$, nor any Orientation × Polarity interactions, participant analysis: $F(1, 12) = 0.42, p = 0.53, \eta^2 = 0.03$; face-exemplar analysis: $F(1, 14) = 1.37, p = 0.26, \eta^2 = 0.1$, for accuracy.

### Overall sweep VEP responses

There were consistent 3-Hz responses in all face conditions but not in the scrambled condition, whereas 6-Hz responses, corresponding to general low-level responses to the general visual stimulation, were present in all conditions (Figure 4). The representative frequency spectra at three channels (58/P7, 75/Oz, 96/P8) with peak SSVEP responses are shown on Figure 5. The topographical distribution of the 3-Hz and 6-Hz responses were clearly distinct (Figure 4). Activity at 3 Hz was spread bilaterally along the occipital and occipito-temporal regions and decreased as the faces were inverted and contrast reversed. By contrast, the 6-Hz responses were focally located on medial occipital electrodes and did not differ between conditions. In order to better isolate the topography of face-related responses, we calculated a 3-Hz index (3Hz/(3 Hz + 6 Hz)). The resulting topographical maps of this index highlighted the implication of bilateral occipito-temporal regions in face detection (Figure 4, bottom). Based on these scalp topographies, we created three regions of interest (ROIs) among posterior occipito-temporal channels in order to compare the magnitude of 3-Hz and 6-Hz responses: right occipito-temporal (channels 90/PO8, 91/P6, 95/PPO10h, 96/P8), left occipito-temporal (channels 58/P7, 59/P5, 64/PPO9h, 65/PO7), and medial occipital (channels 70/Oz, 71/PPO3h, 75/O1, 76/PPO4h, 83/O2). These channels consistently showed the largest relevant responses across participants and conditions (see Figure 6 for the position of these channels and Supplementary Figure 1 for individual participant data). Statistical testing was then carried out separately for 3-Hz and 6-Hz responses at their respective ROIs.

A repeated-measures ANOVA on the 6-Hz response amplitude in the occipital ROI with Condition (face conditions and scrambled) as within-subject factor showed a significant effect of Condition, $F(4, 48) = 4.31, p < 0.005, \eta^2 = 0.26$. This was due to higher amplitudes in the scrambled compared to the face conditions. However, within face conditions, the

<table>
<thead>
<tr>
<th>Condition</th>
<th>Participants</th>
<th>Face exemplars</th>
</tr>
</thead>
<tbody>
<tr>
<td>Up + Pol</td>
<td>0.96 (0.02)</td>
<td>0.96 (0.01)</td>
</tr>
<tr>
<td>Inv + Pol</td>
<td>0.95 (0.02)</td>
<td>0.95 (0.02)</td>
</tr>
<tr>
<td>Up − Pol</td>
<td>0.92 (0.02)</td>
<td>0.92 (0.03)</td>
</tr>
<tr>
<td>Inv − Pol</td>
<td>0.89 (0.03)</td>
<td>0.89 (0.03)</td>
</tr>
<tr>
<td>Scrambled</td>
<td>0.85 (0.06)</td>
<td>0.84 (0.03)</td>
</tr>
</tbody>
</table>

Table 1. Response accuracy rates (Mean ± SEM). Notes: Up = upright; Inv = inverted; +Pol = positive polarity; −Pol = negative polarity.
ANOVA did not reveal any significant differences, $F(3, 36) = 1.07, p = 0.37, \eta^2 = 0.08$. Hence, the general processing of low-level visual features did not differ across face conditions and cannot account for the differences of face-related 3-Hz responses.

A 2 × 5 repeated measures ANOVA on the 3-Hz response amplitude with ROI (left vs. right occipito-temporal) and Condition (upright positive polarity face, inverted positive polarity face, upright negative polarity face, inverted negative polarity face, and scrambled) as within-subject factors showed a significant main effect of Condition, $F(2.13, 25.51) = 36.65, p < 0.0001, \eta^2 = 0.77$, as well as a significant ROI × Condition interaction, $F(4, 48) = 2.92, p < 0.031, \eta^2 = 0.20$. The 3-Hz response was larger in all face conditions compared to the scrambled condition. The ROI × Condition interaction was due to the larger amplitude over the right than the left hemisphere for upright faces, as previously found (Ales et al., 2012) as well as smaller amplitudes in the right relative to the left ROI with face inversion and polarity reversal. Based on these results, we ran a more detailed 2 × 2 × 2 repeated-measures ANOVA on face conditions only with ROI (left vs. right occipito-temporal), Orientation (upright vs. inverted), and Polarity (positive vs. negative) as within-subject factors. There were significant main effects of Orientation, $F(1, 12) = 15.56, p < 0.002, \eta^2 = 0.57$, Polarity, $F(1, 12) = 28.97, p < 0.001, \eta^2 = 0.71$, as well as a significant ROI × Polarity interaction, $F(1, 12) = 5.36, p < 0.036, \eta^2 = 0.32$. The 3-Hz response amplitudes were larger for upright compared to inverted faces and for positive compared to negative polarity faces. The lack of Orientation × Polarity interaction, $F(1, 12) = 2.68, p = 0.13, \eta^2 = 0.18$, suggests that the effects were mainly additive. The ROI × Polarity interaction was due to stronger right lateralization (i.e., larger response in the right than the left ROI) for positive polarity faces compared to negative polarity faces (pairwise t-test on right-left difference: $t(12) = 2.36, p < 0.018$; see Figure 6). This difference was present in 10/13 participants. There was no main effect of ROI, $F(1, 12) = 0.14, p = 0.72, \eta^2 = 0.01$, nor any ROI × Orientation interaction, $F(1, 12) = 1.20, p = 0.30, \eta^2 = 0.09$.

In summary, EEG data show that face detection performance and its related neural response decrease when face orientation and contrast polarity are manipulated. The general pattern of results suggests that the detrimental effects are amplified additively. It is noteworthy that the effect of contrast polarity inversion was larger than the effect of orientation inversion in both behavioral and frequency analyses.
(i.e., partial $\eta^2$ of the main ANOVA effects). This observation suggests that contrast polarity plays a more important role than picture-plane orientation in face detection. To determine to which extent the reduced neural response reflects overall weaker activity and/or increased response threshold, we next examined the evolution of the 3-Hz response across the sweep sequence. Additionally, we investigated the correlation between neural and behavioral response latencies. Given the topography of the face-related response, we focused our analyses on the bilateral occipito-temporal regions.

Figure 5. Frequency spectra of channels 58/P7, 75/Oz, and 96/P8 showing the largest 3-Hz and 6-Hz responses across conditions. Note that although a 3-Hz response is present on all channels in face conditions, the 6-Hz response is largely reduced on bilateral occipito-temporal channels (58 and 96) compared to the medial occipital channel (75). In the scrambled condition, only a 6-Hz response is present.

Figure 6. (A) Mean amplitude difference ($\pm$ SEM) between right and left occipito-temporal ROIs across face conditions. (B) Amplitude difference between right and left occipito-temporal ROIs averaged across contrast polarities in individual participants. A larger right-lateralized response in the positive compared to the negative polarity condition was present for 10/13 participants.
Sweep VEP responses as a function of stimulus visibility

As can be seen in Figure 7, the amplitude of the 3-Hz response increases as a function of face visibility while the 6-Hz response function remains stable across face coherence levels. This distinction further supports the notion that the 3-Hz response represents coherence-dependent face detection processes while the 6-Hz response reflects the constant low-level visual input that is equalized across conditions. Accordingly, the 3-Hz response differs between conditions on several aspects while the 6-Hz response does not. Although all face conditions elicited a rising 3-Hz response, this response varies in terms of absolute amplitude, slope, and onset.

In line with mean response amplitudes, visual inspection reveals that upright positive polarity faces elicit the largest 3-Hz response in both regions, followed by inverted positive polarity faces. Negative polarity faces show the lowest response, which did not vary between orientations. This order was present mainly at low and intermediate levels of face coherence (≈30%–80%). At higher coherence levels, amplitude values tended to converge. Furthermore, this pattern was more salient and was also more consistent across participants and face exemplars (i.e., smaller SEM) in the right compared to the left occipito-temporal region. We ran pairwise permutations test in order to determine significant deviations between response functions. Results are summarized on Figure 8. In the right occipito-temporal region, upright positive polarity faces showed significantly larger amplitudes than inverted positive polarity faces for a large part of the sequence and negative polarity faces from around 32% until approximately the end of the sequence. Inverted positive polarity faces also diverged from negative polarity faces for a large part of the sequence. In the left occipito-temporal region however, differences between conditions were less robust, with smaller amplitude differences and fewer consecutive significant differences. Consistently across both hemispheres, responses for negative polarity faces overlapped throughout nearly all coherence levels, regardless of orientation.
Aside from the magnitude of the 3-Hz response, its profile also varied across conditions. For upright positive polarity faces, the 3-Hz amplitude behaved in a step-like manner, with an abrupt onset followed by a constant level of activation. This pattern was also present in the other face conditions but with the initial response increase being more gradual and linear before reaching a plateau. For instance, in the right occipito-temporal region, upright positive polarity faces elicited a sharp 3-Hz increase around $\approx 30\%$ coherence before saturating around $\approx 40\%$ coherence. In comparison, response to upright negative polarity faces emerged a little later around $\approx 35\%$ coherence and continued to increase until approximately $\approx 80\%$ coherence. Once again, differences in terms of the slope of the response function were more evident in the right compared to the left hemisphere. A potential cause for the shallower slopes for inverted and negative polarity faces may be a larger variation in the onset of the 3-Hz response, such that step-like increases are smoothed when averaged (see Supplementary Figure 2).

To investigate this possibility, we estimated the onsets of the face-related response using cumulatively summed data (see Methods). Table 2 shows the mean goodness-of-fit values of the hinge function used to estimate the detection thresholds. Additionally, we also tested whether neural face detection was correlated with behavioral face detection, either by considering the data in a participant-wise (i.e., variation between participants) or face exemplar-wise manner (i.e., variation between stimuli). The mean response threshold in the bilateral occipito-temporal ROIs for both participant and face exemplar analyses are shown on Figure 9A (the differential cumulative sum function that were used are depicted on Supplementary Figure 3 and individual participant and face exemplar response functions can be seen on Supplementary Figure 4). In accordance with the visual observation of sweep

<table>
<thead>
<tr>
<th></th>
<th>Up +Pol</th>
<th>Inv +Pol</th>
<th>Up −Pol</th>
<th>Inv −Pol</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>R²</strong></td>
<td>0.98 (0.01)</td>
<td>0.95 (0.02)</td>
<td>0.94 (0.02)</td>
<td>0.87 (0.03)</td>
</tr>
<tr>
<td><strong>RMSE</strong></td>
<td>0.095 (0.012)</td>
<td>0.084 (0.007)</td>
<td>0.104 (0.010)</td>
<td>0.111 (0.009)</td>
</tr>
<tr>
<td>Participants</td>
<td>0.93 (0.03)</td>
<td>0.90 (0.03)</td>
<td>0.84 (0.06)</td>
<td>0.83 (0.05)</td>
</tr>
<tr>
<td>Face exemplars</td>
<td>0.088 (0.008)</td>
<td>0.096 (0.009)</td>
<td>0.114 (0.008)</td>
<td>0.110 (0.008)</td>
</tr>
</tbody>
</table>

Table 2. Goodness-of-fit measures for the hinge function used to estimate neural face detection thresholds (Mean ± SEM). Note. Up = upright; Inv = inverted; +Pol = positive polarity; −Pol = negative polarity; RMSE = root mean squared error (smaller values indicate better fit).
response profiles, the quantified response thresholds increased with inversion and contrast-reversal in both right and left occipito-temporal ROIs.

To verify whether the hinge function was equally well fitted across conditions, we ran $2 \times 2 \times 2$ repeated measures ANOVAs on the RMSE values with ROI (left vs. right occipito-temporal), Polarity (positive vs. negative), and Orientation (upright vs. inverted) as within-subject factors. RMSE values were used because $R^2$ was near ceiling in some conditions. For participant analyses, there was a significant main effect of ROI, $F(1, 12) = 9.46, p < 0.01, \eta^2 = 0.44$, due to better fits in the right than the left occipito-temporal ROI. There was a trend for an effect of Polarity, $F(1, 12) = 4.29, p = 0.06, \eta^2 = 0.26$, suggesting better fits for positive compared to negative polarity faces. No other main effect, Orientation: $F(1, 12) = 0.07, p = 0.8, \eta^2 = 0.01$, nor interactions were significant, Orientation $\times$ Polarity: $p = 0.22$; Orientation $\times$ ROI: $p = 0.53$; Polarity $\times$ ROI: $p = 0.79$; Orientation $\times$ Polarity $\times$ ROI: $p = 0.18$. For face exemplars, there was only a main effect of Polarity indicating better fits for positive relative to negative polarity faces, $F(1, 14) = 16.03, p < 0.001, \eta^2 = 0.53$. All other effects were nonsignificant, Orientation: $p = 0.80$; ROI: $p = 0.99$; Orientation $\times$ Polarity: $p = 0.32$; Orientation $\times$ ROI: $p = 0.1$; Polarity $\times$ ROI: $p = 0.80$; Orientation $\times$ Polarity $\times$ ROI: $p = 0.80$. These results suggest that thresholds for negative polarity faces need to be interpreted with caution.

For the participant analyses, a $2 \times 2 \times 2$ repeated-measures ANOVA with ROI (left vs. right occipito-temporal), Orientation (upright vs. inverted), and Polarity (positive vs. negative) as within-subject factors showed main effects of Orientation, $F(1, 12) = 14.1, p < 0.003, \eta^2 = 0.54$, and Polarity, $F(1, 12) = 20.43, p < 0.001, \eta^2 = 0.63$, but not ROI, $F(1, 12) = 1.64, p = 0.23, \eta^2 = 0.12$. Neural face detection thresholds followed the same pattern as behavioral thresholds with the lowest thresholds for upright positive polarity faces, followed by inverted positive polarity faces, upright negative polarity faces, and, finally, inverted negative polarity faces. The effects were additive as there was no Orientation $\times$ Polarity interaction, $F(1, 12) = 1.25, p = 0.29, \eta^2 = 0.09$. There were some trends towards interactions between ROI $\times$ Polarity, $F(1, 12) = 4.58, p = 0.053, \eta^2 = 0.28$, and between ROI $\times$ Orientation $\times$ Polarity, $F(1, 12) = 3.91, p = 0.071, \eta^2 = 0.25$. These near-significant
interactions were due to increased thresholds for inverted negative polarity faces in the left compared to the right occipito-temporal region. In the other conditions, thresholds were equivalent across hemispheres. The ROI × Orientation interaction was nonsignificant, \( F(1, 12) = 2.15, p = 0.17, \eta^2 = 0.15 \).

For face exemplar analyses, the same ANOVA also showed main effects of Orientation, \( F(1, 14) = 6.84, p < 0.02, \eta^2 = 0.33 \), and Polarity, \( F(1, 14) = 9.08, p < 0.009, \eta^2 = 0.39 \), as well as trends towards interactions between Orientation × Polarity, \( F(1, 14) = 4.18, p = 0.06, \eta^2 = 0.23 \), and ROI × Polarity, \( F(1, 14) = 4, p = 0.065, \eta^2 = 0.22 \). There was no main effect of ROI, \( F(1, 14) = 1.78, p = 0.20, \eta^2 = 0.11 \), nor any ROI × Orientation, \( F(1, 14) = 2.23, p = 0.16, \eta^2 = 0.14 \), or Orientation × Polarity × ROI interactions. \( F(1, 14) = 2.34, p = 0.15, \eta^2 = 0.14 \). The main effects reflected the same pattern as the participant analysis. However, here the interactions suggested no threshold differences between upright and inverted polarity faces in the left occipito-temporal region and increased thresholds to inverted negative polarity faces in the right but not the left hemisphere. Overall, the patterns were the most consistent in the right hemisphere across subject-wise and face exemplar analyses. Statistical comparison did indeed not reveal effects of Analysis-type on the threshold patterns in the right occipito-temporal region, \( F(1, 12) = 1.26, p = 0.28, \eta^2 = 0.1 \).

Given that the 3-Hz response onsets paralleled the response time data, we calculated the correlations between the neural and behavioral face detection thresholds (see Figure 9B). For participant analyses, there were low to modest correlations in the right occipito-temporal ROI especially, with only the correlation for inverted negative polarity faces reaching significance. The same analysis taking into account the individual participant’s peak ROI for the 3-Hz response did not improve correlations. Instead, these decreased and were not significant (upright positive polarity = 0.33; inverted positive polarity = 0.02; upright negative polarity = 0.41; inverted negative polarity = 0.16). Filtering data according to correct detection responses did not change the pattern of correlations either.

Correlations for face exemplar data were higher and present in both hemispheres, with significant neural–behavioral correlations for both upright face conditions. Selecting the best ROI for each trial did improve correlations for all conditions except for inverted negative polarity faces (upright positive polarity = 0.75, \( p < 0.001 \); inverted positive polarity = 0.57, \( p < 0.025 \); upright negative polarity = 0.72, \( p < 0.002 \); inverted negative polarity = 0.1, \( p = 0.73 \)). Using only correct responses did not change the overall pattern but did reveal a significant correlation for inverted negative polarity faces (\( r = 0.54, p < 0.038 \)).

### Discussion

**Summary of results**

We investigated the effects of orientation and contrast polarity inversion on face detection using a sweep VEP paradigm, which provides objectively defined and high signal-to-noise ratio responses in a limited number of trials (Ales et al., 2012; Norcia & Tyler, 1985; Regan, 1973; see Almqvist, Leat, & Irving, 2008, for a review).

Our first aim was to test whether the sweep face detection response defined by Ales et al. (2012) reflected high-level visual processes, at least partially specific to faces. This was indeed the case as a larger response was recorded to upright than inverted and contrast-reversed faces. Moreover, there were no differences between face conditions at the 6-Hz presentation frequency, which reflects general low-level processes. In this context, it is important to note that the stimuli used here were fully equalized in terms of their power spectrum and mean luminance. This is not the case in most studies that have used Caucasian faces and manipulated contrast polarity: mean luminance is not controlled so that negative polarity faces are darker than their positive counterpart and thus a comparison between positive and negative polarity faces is biased (Bruce & Langton, 1994; Farroni et al., 2005; Galper, 1970; Itier & Taylor, 2002; Nasr & Tootell, 2012; Nederhouser et al., 2007; Ohayon, Freiwald, & Tsao, 2012; Otsuka, Hill, Kanazawa, Yamaguchi, & Spehar, 2012; Russell, Sinha, Biederman, & Nederhouser, 2006; Taubert & Alais, 2011; Vuong et al., 2005; Yue, Nasr, Devaney, Holt, & Tootell, 2013; see George et al., 1999; Liu, Collin, Burton, & Chaudhuri, 1999; Tomalski & Johnson, 2012, for controlled stimuli). Furthermore, since SSVEP responses are sensitive to attention (e.g., Morgan, Hansen, & Hillyard, 1996), especially over medial occipital sites where the 6-Hz response is prominent (Müller et al., 2006; Pei, Pettet, & Norcia, 2002), the lack of differences between conditions suggest that there were no general differences in allocation of attentional resources across conditions.

Our second aim was to examine how the face detection process was affected by these two image manipulations (picture-plane inversion and contrast-reversal). Behaviorally, more face structure was necessary for participants to detect inverted and contrast-reversed faces (15%–20% additional coherence) compared to upright normal faces. Several factors appear to account for the behavioral effects. First, face detection responses on bilateral occipito-temporal regions decreased in amplitude for inverted and
contrast-reversed faces. The response profile as a function of face image coherence indicates that this effect was largely sustained throughout the stimulation sequence. Second, the estimated face detection thresholds increased with inversion and contrast-reversal. Third, the response function was shallower for inverted and contrast-reversed faces than typical faces. Our results show that impaired face detection can be attributed to the combination of elevated detection threshold and inefficient activation of neural populations.

Upright positive polarity face detection

The detection of upright positive polarity faces varying in position, viewpoint, gender, and size was associated with an abrupt increase at the face presentation frequency (3 Hz) over the right occipito-temporal region around channel 96/P8. The onset of the 3-Hz response, occurred on average at a coherence level of ~35% and quickly saturated at suprathreshold levels (~40%). Behaviorally, performance was near ceiling and the average detection threshold was around 42%. Our findings thus replicate the results from the original face detection sweep study (Ales et al., 2012). Interestingly, the two sets of data were comparable despite a large difference in the proportion of face trials in each experiment (1/3 face trials–2/3 scrambled trials in Ales et al., 2012 vs. 4/5 face trials–1/5 scrambled trials in the present experiment). The role of expectation in generating this precise response profile and threshold therefore appears to be limited.

The step-like response profile is consistent with previous ERP studies that parametrically varied the phase coherence of face images and measured the N170 face-sensitive component (Philiastides & Sajda, 2006; Rousselet et al., 2008). Responses also rose around similar coherence levels (30% in Rousselet et al., 2008; 35% in Philiastides & Sajda, 2006, though in this case the range was restricted to 30%–45% phase-coherence), despite using a different approach of phase scrambling (weighted mean phase, Dakin et al., 2002). Response functions like these indicate that detection of normal faces is categorical in nature: face versus nonface.

Several mechanisms are thought to underlie efficient face detection in natural scenes. Some authors have suggested that the typical amplitude spectrum present in human faces provides enough information for face detection (Crouzet & Thorpe, 2011; Honey et al., 2008). However, since this property is equalized across all images as in our study, it cannot drive a neural face-detection response (see also Rousselet et al., 2008). Midlevel properties such as local shape and edge information also play a part in face detection, but again these factors are not sufficient (Honey, Kirchner, & VanRullen, 2008; Rousselet et al., 2008). While the eyes are important for fast face detection (Burton & Bindemann, 2009; Lewis & Edmonds, 2003), the absence of clearly defined facial features does not necessarily hinder face perception (e.g., Caharel et al., 2013; George et al., 2005). Face detection probably relies on a combination of visual properties summarized within a face template with which the visual input is compared (Perrett, Oram, & Ashbridge, 1998; Rossion & Boremanse, 2008; Rousselet et al., 2003). Such a template is likely to be based both on biological constraints (Morton & Johnson, 1991; Turati, Simion, Milani, & Umiltà, 2002) and visual experience, so that only upright positive polarity faces can match this template.

Effects of orientation and contrast polarity on face detection

Inverted and negative polarity faces are unusual for the visual system and because of this the brain cannot rely on a pre-existing internal template. The lack of such a template may then lead to the need for more structure (phase coherence) in the stimulus in order for a differential response to be generated. We found a significant threshold increase of the 3-Hz response for inverted (~9% additional coherence on average) and negative polarity faces (14%–20% additional coherence) relative to normal upright faces. Moreover, even after detection was achieved, the responses still differed from normal upright faces in terms of their overall amplitude and response profile.

The decrease of response amplitude at suprathreshold coherence values was especially salient on right occipito-temporal channels and lasted throughout the entire sequence for negative polarity faces. Response amplitudes for inverted faces eventually converged with those of upright faces at higher coherence levels. Previous studies have consistently reported reduced neural activity to inverted (e.g., Haxby et al., 1999; Kanwisher et al., 1998; Mazard et al., 2006; Nasr & Tootell, 2012) and negative polarity faces (e.g., George et al., 1999; Nasr & Tootell, 2012; Yue et al., 2013) in face-selective regions. In transient ERP studies by contrast, these manipulations generally evoke a paradoxical increase in the N170 amplitude (e.g., Eimer, 2000; Itier & Taylor, 2002; Rossion et al., 1999; Sadeh & Yovel, 2010). However, a reduction of the N170 amplitude for inverted faces can be observed when face stimuli are degraded by masking the features with noise (Linkenkaer-Hansen et al., 1998; Schneider, DeLong, & Busey, 2007), transformed into two-tone images (George et al., 2005) or when the features are schematic/unrealistic (Caharel et al., 2013;
Sagiv & Bentin, 2001; but see Tomalski & Johnson, 2012). In these cases, the decrease of N170 amplitude with inversion is interpreted as a failure to perceive the stimulus as a face (Rossion & Jacques, 2011). Interestingly, Schneider et al. (2007) parametrically varied the visibility of phase-scrambled faces and reported an increase of the N170 amplitude with increasing visibility. Critically, amplitudes for inverted faces were smaller relative to upright faces throughout visibility levels. This kind of reasoning can be applied to both our inverted and negative polarity conditions in which facial features were masked by phase scrambling.

Besides the quantitative differences, the profile of the response amplitude relative to face structure was also different for inverted and negative polarity faces. While the response function was step-like for upright positive polarity faces, it had a shallower slope for inverted and negative polarity faces and levelled off at a higher coherence level (~75% vs. 40% for normal faces), suggesting that the face-related information is integrated differently in these conditions (Perrett et al., 1998). If we consider the hypothesis of an internal upright positive polarity face template which optimally activates the neural population to saturation point, these modified faces would only partially match the template and therefore gradually activate neural populations and reach saturation level only when there is enough structure to fully match the template.

Origins of orientation and contrast polarity inversion effects

For picture-plane inversion, it is proposed that faces are processed by suboptimal analytical (feature-based) mechanisms, as opposed to the more efficient holistic processing (integrated perception of multiple facial features) underlying the perception of upright positive polarity faces (Rossion, 2008; Sergent, 1984; Tanaka & Farah, 1993). The effect of contrast polarity inversion seems to stem from a different cue given the relatively intact holistic processing with negative polarity faces (Taubert & Alais, 2011). Some suggest that shape-from-shading cues are distorted with contrast inversion (Kemp, Pike, White, & Musselman, 1996; Lewis & Johnston, 1997; Liu, Collin, & Chaudhuri, 2000) while others underline the role of texture and pigmentation cues (Bruce & Langton, 1994; Nederhouser et al., 2007; Russell et al., 2006). A third alternative postulates that a more global property underlies this effect, namely, the pattern of light and dark regions (Dakin & Watt, 2009; Gilad, Meng, & Sinha, 2008). In a face, the eyes are always darker than the forehead and cheeks, and it is this stable contrast relationship that drives efficient face processing.

Comparing effects of orientation and contrast polarity inversion

Directly comparing the effects of picture-plane and contrast polarity inversion reveals larger effects of negating contrast polarity. This pattern is present in other studies as well both behaviourally (Bruce & Langton, 1994; Gaspar, Bennett, & Sekuler, 2008; Lewis & Edmonds, 2003) and electrophysiologically (Itier & Taylor, 2002; Rossion et al., 2012), even though contrast-reversed faces were not equated to normal faces in terms of luminance in these studies. An upside-down face may be taken as an extreme case of the range of face orientations normally experienced. However, contrast polarity inversion never occurs in the natural world and would not benefit from the existing template processing. In another sweep VEP study (Liu-Shuang, Ales, Rossion, & Norcia, 2015) with a similar design as the current experiment, the results show that the perception of a positive polarity face dominates that of its negative counterpart when they are directly competing, suggesting that the visual system exclusively uses the information contained in the positive polarity face. These findings suggest a more fundamental role of the contrast pattern in activating face-selective neural populations for face detection. This hypothesis is also supported by infant research showing that infants’ preference for faces is inherently dependent on whether the real or schematic face stimulus contains the correct contrast information (Farroni et al., 2005; Otsuka et al., 2012). Additionally, face-selective cells in the monkey infero-temporal cortex show sensitivity to this parameter as well (Ohayon et al., 2012).

The combined picture-plane and contrast polarity inversion caused the most substantial decrement in behavioral performance and neural response magnitudes. With the exception of the amplitude analysis across coherence levels, the data generally point towards an additive effect of both manipulations. This additivity is also in line with previous evidence (Bruce & Langton, 1994; Gaspar et al., 2008; Itier et al., 2006; Johnston et al., 1992; Lewis & Edmonds, 2003; Rossion et al., 2012).

Hemispheric lateralization

The face-related SSVEP response at 3 Hz was distributed along the posterior occipito-temporal channels. However, a more detailed examination taking into account its amplitude relative to the non-face-
related 6-Hz response revealed activation clusters over bilateral occipito-temporal regions. These scalp topographies correspond to the typical locations of the face-sensitive N170 component (e.g., Bentin et al., 1996; for review see Rossion & Jacques, 2011) and SSVEP responses to sequences of different faces presented below 10 Hz (Alonso-Prieto et al., 2013; Liu-Shuang et al., 2014; Rossion & Boremanse, 2011). However, there were hemispheric differences. Responses were right lateralized for positive polarity faces but bilateral for negative polarity faces, and differences between conditions in terms of response profile and thresholds were more salient in the right hemisphere. Given the dominance of the right hemisphere in face perception and recognition in humans (e.g., Bentin et al., 1996; Hécaen & Angelergues, 1962; Hilger & Koenig, 1991; Kanwisher et al., 1997; Rossion, 2014; Sergent et al., 1992), our data reflect the involvement of specialized face detection processes and underline the magnitude of the disturbance caused by contrast polarity inversion (i.e., not recruiting a dedicated right occipito-temporal face processing network).

Correlation with behavior

At the behavioral level, our face detection thresholds conform with data from previous studies where a roughly linear increase in response times with inversion and contrast-reversal have been found (e.g., Lewis & Edmonds, 2003). Compared to our previous study (Ales et al., 2012), the correlations between the neural thresholds and behavioral measures were rather weak when considering the variability across participants. In a similar vein, the N170 face-sensitive component was not related to behavioral reaction times in a face detection task with phase-scrambled images (Philiaстides, Ratcliff, & Sajda, 2006; Philiastides & Sajda, 2006). These findings suggest that neural signatures of face detection are not strongly related to overt behavioral detection responses, which is understandable when considering the added sources of variability that are present in the behavioral response. Decisional strategies and criteria, as well as delays in motor response preparation and output across individual participants, contribute heavily to interindividual variance in behavioral responses. Correlations were much higher when considering data across individual face stimuli rather than participants. Thus, some faces were consistently detected faster or slower irrespective of individual variations in response speed. In our stimulus set, face n°3 was the most easily detected and face n°13 was the least well detected in all conditions, similar to the original face detection sweep study (Ales et al., 2012).

Conclusions

In conclusion, our sweep VEP study reveals that orientation and contrast polarity inversion additively impair the face detection process at several levels. First, these manipulations increase behavioral and neural response thresholds. Second, they decrease the overall neural response amplitude over bilateral occipito-temporal regions. Third, contrast inversion also reduces the right hemisphere lateralization that is characteristic of high-level face processing. Overall, this manipulation produces greater effects behaviorally and neurally compared to picture-plane inversion, indicating that contrast polarity may be fundamental for face detection. Finally, with orientation and contrast inversion the suprathreshold response profile as a function of face visibility changes from a step-like function towards a more progressive increase, suggesting that the underlying neural populations require additional face structure information to reach saturation level.

Keywords: face detection, sweep SSVEP, parametric image variation, face inversion, contrast polarity reversal

Acknowledgments

This research was funded by a grant from the European Research Council (facesvtep 284025) awarded to B. R. and a travel grant from the Belgian National Scientific Research Fund (A2/5-ROI/JA-1240) awarded to J. L.-S. The authors would like to thank Faraz Farzin and Claudia Arrighi for their help for data acquisition and participant recruitment, Vladimir Vildavski and Corentin Jacques for helpful comments regarding data analysis, and Bastien Gossissen for providing the function fitting scripts.

Commercial relationships: none.
Corresponding author: Joan Liu-Shuang.
Email: joan.liu@uclouvain.be.
Address: Psychological Sciences Research Institute and Institute of Neuroscience, University of Louvain, Belgium.

References


