Behavioral/Cognitive

The Neural Dynamics of Face Detection in the Wild Revealed by MVPA

Maxime Cauchoix,[1]* Gladys Barragan-Jason,[1]* Thomas Serre,[2]† and Emmanuel J. Barbeau[1]†

[1]Centre de Recherche Cerveau et Cognition, Université de Toulouse, CNRS-UMR 5549, Toulouse 31000, France and [2]Cognitive Linguistic and Psychological Sciences Department, Institute for Brain Sciences, Brown University, Providence, Rhode Island 02912

Previous magnetoencephalography/electroencephalography (M/EEG) studies have suggested that face processing is extremely rapid, indeed faster than any other object category. Most studies, however, have been performed using centered, cropped stimuli presented on a blank background resulting in artificially low interstimulus variability. In contrast, the aim of the present study was to assess the underlying temporal dynamics of face detection presented in complex natural scenes.

We recorded EEG activity while participants performed a rapid go/no-go categorization task in which they had to detect the presence of a human face. Subjects performed at ceiling (94.8% accuracy), and traditional event-related potential analyses revealed only modest modulations of the two main components classically associated with face processing (P100 and N170). A multivariate pattern analysis conducted across all EEG channels revealed that face category could, however, be readout very early, under 100 ms poststimulus onset. Decoding was linked to reaction time as early as 125 ms. Decoding accuracy did not increase monotonically; we report an increase during an initial 95–140 ms period followed by a plateau ~140–185 ms–perhaps reflecting a transitory stabilization of the face information available–and a strong increase afterward. Further analyses conducted on individual images confirmed these phases, further suggesting that decoding accuracy may be initially driven by low-level stimulus properties. Such latencies appear to be surprisingly short given the complexity of the natural scenes and the large intraclass variability of the face stimuli used, suggesting that the visual system is highly optimized for the processing of natural scenes.

Introduction

How much time do we need to detect a face in a natural environment? It is likely that very little time would be needed considering how crucial faces may have been to our ancestors, since they signaled either a danger or an opportunity. Human observers can robustly recognize faces presented in complex natural scenes very rapidly (~260–290 ms poststimulus onset) even when there are large changes in appearance (Fabre-Thorpe, 2011). Human participants can initiate a saccade to a face as early as ~100–110 ms poststimulus onset; this is faster than toward any other object category (Crouzet et al., 2010). Such results highlight the formidable robustness and efficacy of the primate visual system to detect faces in natural scenes. However, our understanding of the precise timing and corresponding neural dynamics underlying this process remains relatively coarse.

Magnetoencephalography/electroencephalography (M/EEG) studies have sometimes reported an event related potential (ERP) differential (face vs no-face) signal over occipitotemporal sites during the P1 component ~80–120 ms poststimulus onset (Halgren et al., 2000; Eimer and Holmes, 2002; Liu et al., 2002; Itier and Taylor, 2004; Thierry et al., 2007a; Dering et al., 2009, 2011; Rossion and Caharel, 2011). However, a subsequent N170 component (~140–200 ms) has been identified as a more reliable correlate of face perception (Jeffreys, 1989; Bentin et al., 1996; Rossion et al., 1999). One important caveat, however, is that these studies have relied on the use of isolated and cropped faces with limited interstimulus variability (Dering et al., 2011). Whether highly variable faces presented in complex natural scenes would elicit or not elicit a distinct pattern of neural activity at such early latencies remains to be investigated (Rousselet et al., 2004, 2005, 2007a; Dering et al., 2011).

The aim of the present study was thus to assess the timing and characterize the neural stages underlying face processing using complex, cluttered, and natural scenes.

Participants performed a go/no-go task during which they had to detect human faces. We first focused our analyses on a classical (ERP) univariate analysis on the P1 and N170 components. Only modest modulations were found on classic electrodes associated with face processing. Given that our scene stimuli (compared with cropped stimuli) could potentially modify the classic topography of neural activity evoked by faces (Rousselet

Received July 16, 2013; revised Nov. 18, 2013; accepted Nov. 24, 2013.

Author contributions: G.B.-J. and E.J.B. designed research; G.B.-J. performed research; M.C. analyzed data; M.C., G.B.-J., T.S., and E.J.B. wrote the paper.

The data analysis component of this work was supported in part by the National Science Foundation early career award (IIS-1252951), Office of Naval Research (N000141110743), and the Robert J. and Nancy D. Carney Fund for Scientific Innovation. Additional support was provided by the Brown Institute for Brain Sciences, the Center for Vision Research, and the Center for Computation and Visualization.

*M.C. and G.B.-J. contributed equally to this work.

†T.S. and E.J.B. contributed equally to this work.

The authors declare no competing financial interests.

Correspondence should be addressed to Maxime Cauchoix, Centre de Recherché Cerveau and Cognition, CNRS CERCO UMR 5549, Pavillon Baudot, CHU Purpan, BP 25202, 31052 Toulouse Cedex, France. E-mail: mcauchoixxx@gmail.com.

DOI:10.1523/JNEUROSCI.3030-13.2014

 0270-6474/14/340846-09$15.00/0

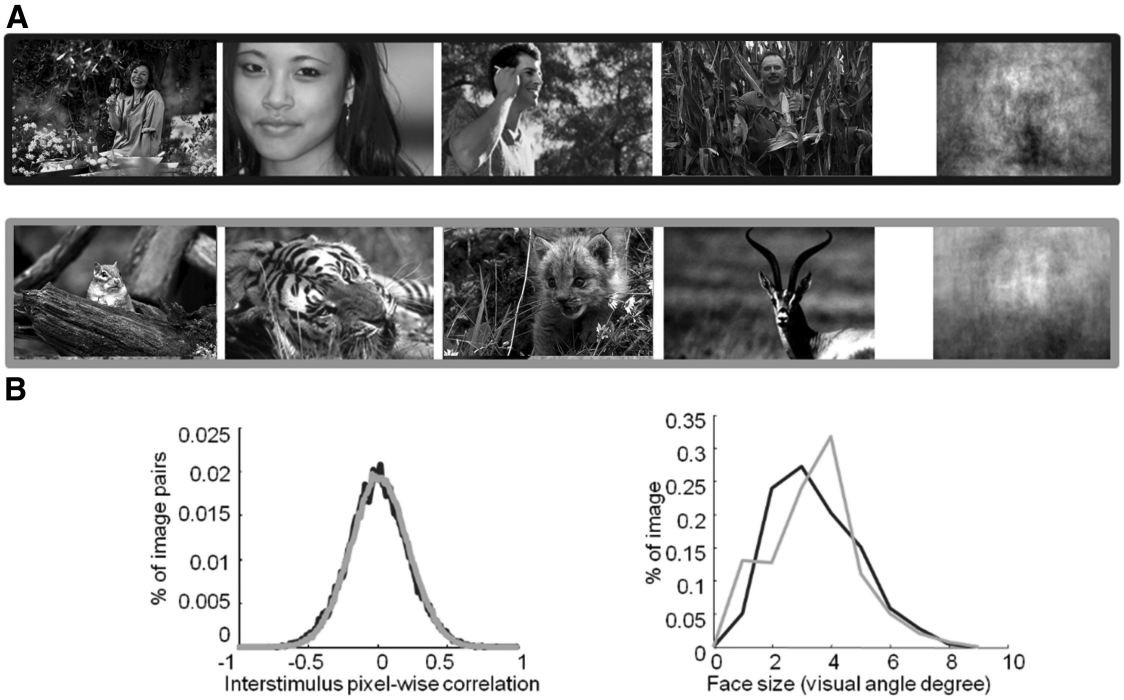

Figure 1. Stimuli used in the study. ***A***, Representative stimuli used in the study and pixel-based sample mean average over the entire stimulus set computed for targets (human faces) and distractors (animal faces). ***B***, Size and eccentricity for targets (black) and distractors (gray).

and Pernet, 2011), we subsequently considered a (whole-brain) multivariate pattern analysis (MVPA) technique to investigate the dynamics of face processing. MVPA techniques, along with other approaches attempting to perform multivariate ERP analyses (Parra et al., 2005; Philiastides and Sajda, 2006), have become increasingly popular in the imaging literature because they enable the detection of subtle effects otherwise undetectable with classical analyses (Kamitani and Tong, 2005). By pooling information across electrodes, whole-brain MVPA may thus increase the statistical power and enable the detection of category information earlier than predicted from single or subsets of EEG electrodes (Cauchoix et al., 2012). This method also allows for an image-based decoding analysis whereby decoding accuracy can be correlated with image properties and/or subject behavioral measures to relate EEG to behavior.

Materials and Methods

Participants. Fifteen females and 13 males ($n = 28$, median age: 24 years, range: 19–37, 25 right-handed) signed informed consent to participate in the experiment. All subjects reported that they had normal or corrected-to-normal visual acuity.

Stimulus set. Target images consisted of grayscale photographs of human faces (270 images) presented in their natural contexts (i.e., the images included some background clutter and no face was artificially pasted). We selected face exemplars that exhibited significant intraclass variations such as viewpoint, gender and race, eccentricity, and size. Sample faces are shown in Figures 1*A* (top row) and 5. Distractor stimuli consisted of (nonhuman) animal faces (270 images), which included different species (mammals, birds, reptiles, etc). The stimulus set was previously used by Rousselet et al. (2004), (Fig. 1*A*, bottom row). Images were 320 × 480 pixels in size. Confidence intervals (CIs; 95%) were computed and are reported in square brackets throughout the paper. The global luminance and root-mean-square contrast were similar between the two groups (mean luminance: 105.8 [102.5 109.1] vs 104.3 [100.1 107.6] for animal and human images, respectively; $t_{(538)} = 0.77$, $p = 0.44$; mean contrast: 53.4 [52.0 54.8] and 54.8 [53.4 55.2] for animal and human images, respectively; $t_{(538)} = 1.60$, $p = 0.1$).

To characterize this intraclass variability, for each image we computed and compared size (approximated by the diameter of a circle containing the same number of pixels as the cropped face) and eccentricity (measured as the distance between the fixation cross and the center of a square manually drawn on the face) for human and animal faces (Fig. 1*B*).

To further verify that the set of target and distractor images did not differ in low-level visual differences, we used a computer vision approach similar to the "tiny images" approach by Torralba (2009). The approach consists of downsampling stimuli to very low-resolution (32 × 48 pixels) grayscale images. A linear Support Vector Machine (SVM) classifier is then fed the corresponding pixel intensities using a classification procedure identical to the one used for neural decoding (see below). Classification was not significantly different from chance level (mean: 52%, $p = 0.24$). Overall, this analysis suggests that low-level visual cues provide insufficient information to perform the task.

Experimental setup. Participants sat in a dimly lit room ~90 cm away from a 19 inch CRT computer screen (resolution: 1024 × 768; vertical refresh rate: 100 Hz) controlled by a PC computer. Photographs were displayed on a black background and subtended a visual angle of ~7 × 11° using the E-prime software. The experiment consisted of a go/no-go paradigm, which was divided in three blocks of 180 photographs each (90 targets and 90 distractors). Participants were familiarized with the experiment using a small set of stimuli (30 targets, 30 distractors) not used for the actual experiment.

Participants were instructed to respond as quickly and accurately as possible by raising their finger from an infrared response pad when a target stimulus was presented (human face target/go response). They were asked to keep their finger on the response pad if a distractor stimulus (animal face) was presented (no-go response). At the beginning of each trial, a fixation cross appeared for a random time interval to prevent anticipatory responses (300–600 ms). This was followed by the presentation of the stimulus (100 ms) and a blank screen (1000 ms; Fig. 2*A*). The order of the stimuli was randomized across blocks and participants.

EEG recording. EEG activity was recorded from 32 electrodes mounted on an elastic cap based on the 10–20 system (Oxford Instruments) with the addition of extra occipital electrodes using a SynAmps amplifier system (Neuroscan). The ground electrode was placed along the midline, in front of Fz, and impedances were kept <5 kΩ. Signals were digitized at a

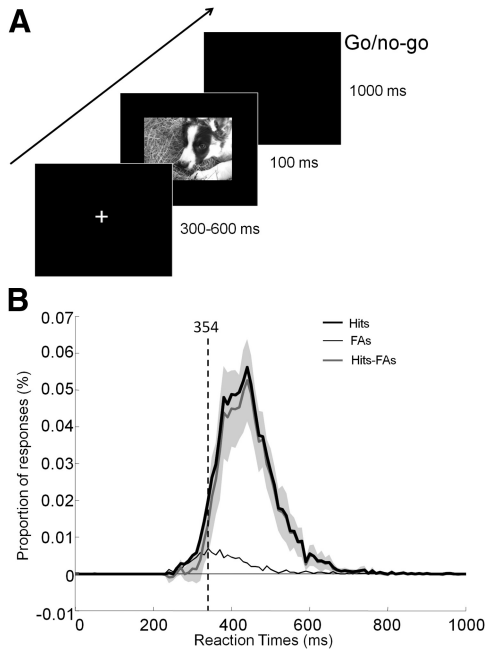

Figure 2. Experimental design and RT distribution. ***A***, After the presentation of a fixation cross for a random time interval (300 – 600 ms), the stimulus was flashed (100 ms) followed by a blank screen. Participants had 1 s to respond using an infrared pad if they perceived a human face in the picture (go response). Otherwise, they had to withhold their response (no-go response). ***B***, Distribution of RTs for hits (black curve), false alarms (FAs; thin black gray curve), and difference across subjects (Hits-FAs) with 95% CIs plotted in light gray. The vertical dashed line indicates the minimal RT at which target and distractor start to be reliably classified (see Materials and Methods).

sampling rate of 1000 Hz and low-pass filtered at 100 Hz. Potentials were referenced on-line to the Cz electrode and average referenced off-line. EEG data analysis was performed using EEGLAB (Delorme and Makeig, 2004), a freely available open source toolbox (http://www.sccn.ucsd.edu/eeglab) running under MATLAB (The Mathworks).

First, EEG data were downsampled to 256 Hz and then digitally filtered using a bidirectional linear filter (EEGLAB FIR filter) that preserves the phase information (pass-band 0.1–40 Hz). For two of the participants, one of the channels also had to be excluded from analysis because of the presence of significant permanent artifacts. Continuous data were then manually pruned from nonstereotypical artifacts such as high amplitude and high-frequency noise (muscle) as well as from electrical artifacts resulting from poor electrode contacts. All remaining data were then submitted to Infomax Independent Component Analysis (Infomax ICA) using the runica algorithm (Makeig et al., 1997) from the EEGLAB toolbox. For each subject, we visually identified and rejected one to three well characterized ICA components for eye blink and lateral eye movements (Delorme et al., 2007). Scalp maps, power spectrum, and raw activity of each component were visually inspected to select and reject these artifactual ICA components.

A total of 540 epochs for each individual participant (15,120 epochs) were extracted (−100–700 ms) and baseline corrected (−100–0 ms). Only correct trials were considered for EEG analyses (14,101 epochs) and further inspected visually. Epochs containing artifacts were excluded from further analysis.

Following this entire procedure, the mean percentage of rejected epochs across all participants was 19.9% ([17.2 22.6]; range: 8.2–37.1%). Thus, further analysis was performed on 11,087 epochs (mean per subject: 396; range: 264–456).

ERP analyses. ERPs were computed separately for correct human face target trials and correct animal face nontarget trials. We report results for the P100 at four bilateral occipital electrodes (O1, O2, PO3, and PO4) and for the N170 at four right hemisphere occipitotemporal electrodes (PO10, PO8, P8, and TP8), where amplitude was maximal or was classically associated with face processing. Amplitudes were quantified for each condition as the mean voltage measured within 30 ms windows centered on the grand average peak latencies of the component's maximum amplitude. Peak latency was extracted automatically at the minimum value between 60 and 140 ms for the P100 and 110 and 190 ms for the N170 (Rossion and Caharel, 2011).

To estimate reliable differences in peak amplitude or latency while limiting possible confounding issues due to multiple comparisons, we ran a paired two-tailed permutation test based on the t_{max} statistic (Blair and Karniski, 1993; Maris and Oostenveld, 2007) using a familywise α-level of 0.05 (32 comparisons) for each component (P100 an N170). All statistic analyses were performed using the Mass Univariate ERP toolbox (Groppe et al., 2011) written in MATLAB.

To precisely track the time course of face information, the same statistical analysis was used for comparing ERPs evoked by human versus animal faces. For this analysis, we considered all time points between −50 and 700 ms (192 time points) across all 32 electrodes (i.e., 6144 comparisons total).

Behavioral performance analysis. To estimate the minimal processing time required to detect target images, we computed the shortest latency (minimal reaction times; RT) at which correct go-responses started to significantly outnumber incorrect go-responses (Rousselet et al., 2003). Minimum RTs across trials were computed using 10 ms sliding time bins (χ^2 test, $p < 0.01$; Rousselet et al., 2003). Across participants, to allow for lower statistical power than with across-trial data since there were fewer trials, we used 30 ms time bins and a Fisher's exact test ($p < 0.01$; Barragan-Jason et al., 2012, 2013; Besson et al., 2012). Minimum RTs were estimated by considering the onset of the first significant bin followed by at least 60 ms of significance (Barragan-Jason et al., 2012; Besson et al., 2012).

MVPA. MVPA was conducted on single-trial ERPs. A linear classifier was trained to decode the presence of a target versus distractor in single trials from individual time bins of the EEG signal across all electrodes. We derived an accuracy measure by averaging the performance of the classifier over multiple random splits of the data (see below). Such decoding analysis characterizes the temporal evolution of the category signal across the whole brain. Each input feature (electrode potential) was normalized (using a *Z*-score) across trials, and a linear SVM was used as classifier.

The classification procedure ran as follows: (1) For each subject, the stimulus set was split equally into a training and a test set that contained an equal proportion of target (correct go responses) and distractor images (correct no-go responses); (2) an optimal cost parameter C was determined through line search optimization using eightfold cross-validation on the training set; and (3) an SVM classifier was trained and tested on each set. For each subject, this procedure was repeated over 100 times where different training and test sets were selected each time at random. A single measure of accuracy was obtained by averaging the classification performance over all repetitions. A measure of chance level was obtained by performing the same analysis on permuted labels. This allowed us to estimate the latency of category information across all participants via a paired, two-tailed permutation test (accuracy measured on permuted vs nonpermuted labels; $p < 0.01$) based on the t_{max} statistic (Blair and Karniski, 1993) using a familywise α-level of 0.05 (i.e., 192 comparisons). Reported decoding latencies correspond to the earliest significant bin. To characterize the contribution of individual electrodes to the overall decoding accuracy, we computed the average weights obtained for each electrode during the cross-validation procedure.

We further considered a classifier confidence for individual images and each participant by averaging out the decoding accuracy for a specific image over 100 cross-validations (imAcc). To evaluate the contribution of various image properties (Weibull, face size) and subject behavior (median RT for individual images calculated across participants) to the neural signal, we fitted a regression model to *z*-scored variable values at each time point: imAcc(t) = b0 + b1**Weibull* + b2* *size* + b3*median RT; using MATLAB glmfit function (Hauk et al., 2006; Clarke et al., 2013). To estimate the contribution of each variable in time, we report the time course of the slopes (b1, b2, and b3) and associated *p* values (corrected for multiple comparisons using false discovery rate methods, $p < 0.01$; Lage-Castellanos et al., 2010). Because each variable was nor-

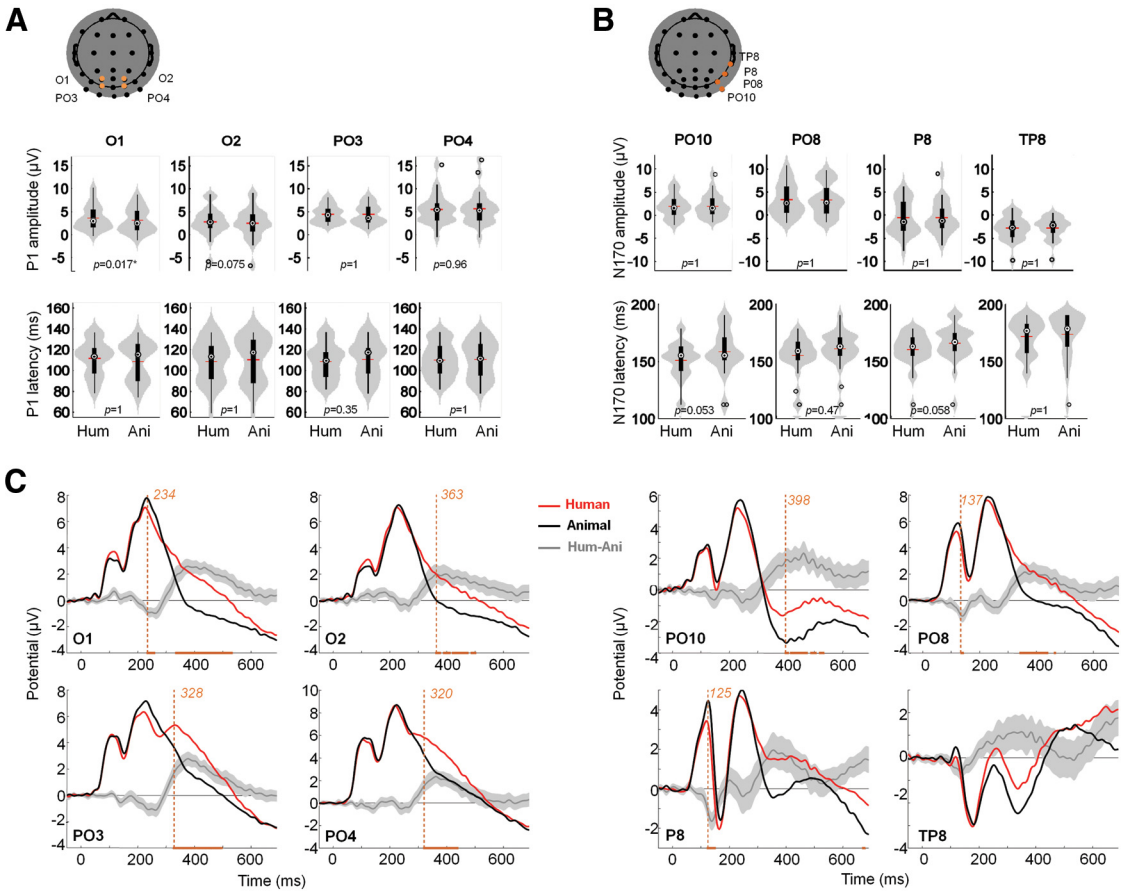

Figure 3. Peak and time course ERP analysis. ***A***, Peak analysis for the P100. Reported electrodes (O1, O2, PO3, and PO4) are indicated in orange overlaid on the scalp topography. Top line plots show peak amplitude distributions, bottom line plots show amplitude distributions for the two conditions (human vs animal faces) using violin plots in gray (Allen et al., 2012) and box plots in black. The small red horizontal bar indicates the mean; *p* values estimated from a permutation paired *t* test using the t_{max} method are shown at the bottom of each plot. Hum, human; Ani, animal. ***B***, Same as ***A*** for the P170. Reported electrodes (PO10, PO8, P8, and TP8) are indicated in orange overlaid on the scalp topography. ***C***, ERPs for targets (black) and distractors (gray). Orange curves correspond to the mean differential activity between the two conditions (±95% CI across participants shown as lighter orange shaded area). Time points for which a significant difference between the two conditions was found (paired permutation *t* test using t_{max} method, $p < 0.05$) are indicated on the *x*-axis. Earliest significant time bin is shown with a vertical dotted line.

malized to zero mean and unit SD, the regression coefficients can be interpreted as "microvolts per SD" of the corresponding variable.

To assess how similar the decoding was across all image stimuli, we ran a clustering algorithm to identify possible image subgroups with similar patterns of decoding accuracy. We ran *k*-means ($k = 2$–5) directly on the temporal decoding curves obtained from individual images. The optimal number of clusters *k* was selected by visual inspection of the cluster centers.

Image feature computation. We obtained a low-level estimate of the contrast of individual images by fitting a Weibull function for individual images (Scholte et al., 2009). The analysis was done using both β and γ parameters estimated from the Weibull function. As both parameters gave highly similar results, figures and statistics are only presented for the β parameter. As an additional low-level image property we considered the size of the face in individual images. The underlying assumption is that as tolerance to position and scale increases along the visual hierarchy (Riesenhuber and Poggio, 1999), one would expect the stimulus scale to correlate with low-level processes and less so with higher level processes.

Results

Participants performed the categorization task (human vs animal face) with a very high level of accuracy (mean: 94.8% [93.6 96.0], range: 89.1%–99.6%) and fast RTs (mean RT: 445 ms [428.7 461.3]). The mean minimum RT across trials was 354 ms (Fig. 2*B*).

To study the face selectivity of the EEG signal, we first computed standard ERPs for each condition and performed a peak analysis on occipitotemporal electrodes. The classical P1-N1-P2 complex can be readily observed (Figs. 3*C*, 4*A*). At the same time, the N170 component appears small compared with components obtained in previous studies using cropped homogenous stimuli (e.g., Rousselet et al., 2004; Thierry et al., 2007a, b; Dering et al., 2011; Rossion and Caharel, 2011). The specific topography of classical face components seems dramatically different compared with what has been previously reported with most occipital electrodes remaining mainly positive during the N170 time widows (Fig. 4*A*).

The maximal P100 amplitude (mean = 5.6 μV [4.5 6.7]) on human face stimuli was recorded on PO4 right temporal electrode at 105 ms poststimuli onset (Fig. 3). Using a paired two-tailed permutation test based on the t_{max} statistic, we found just one electrode significantly modulated in amplitude (O1: $t_{max} = 3.17$, $t_{orig} = 3.56$, df = 27, $p = 0.02$) and no significant modulation in latency ($t_{max} = 2.87$, df = 27, $p > 0.05$) for the P100 (Fig. 3*A*).

The maximal N170 amplitude (mean = −3.6 μV [−4.5 −2.7]) on human face stimuli was recorded on TP8 right temporal electrodes at 180 ms poststimuli onset (Fig. 3). ERPs averaged

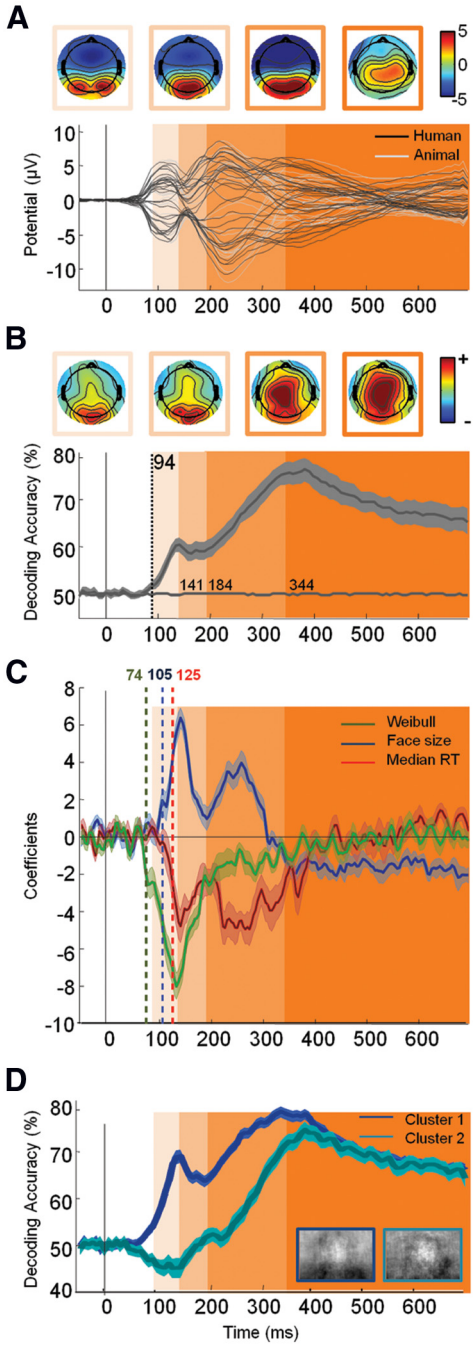

Figure 4. Neural timing of face detection in natural scene. ***A***, Top, Potential topographies (μv) averaged during the four periods defined in ***B*** using shaded areas. Bottom, Average ERP ($n = 28$) for each of the $k = 32$ electrodes for target (black) versus distractor (gray) stimuli. ***B***, Top, Topographies of (normalized) decoding weights for face stimuli averaged during the four periods defined below by shaded areas. Bottom, Average temporal decoding accuracy (±95% CI) across all participants. The decoding accuracy estimated from permuted labels was used to assess chance level (shown in gray on the *x*-axis). The vertical dotted line indicates the latency of the first significant decoded bin (estimated using an aired permutation *t* test based on t_{max} method, $p < 0.05$). Based on the latency of this first bin and changes in the sign of the derivative of the temporal decoding curve, we isolated four distinct temporal widows (shown in shades of orange: 94–141 ms; 141–184 ms; 184–344 ms; >344 ms). These four windows are also shown on ***A*** and ***C*** for improved readability. ***C***, Coefficient derived from the regression analysis on single image decoding (see Materials and Methods) with Weibull α (green), face size (blue), and subject median RTs (red) (±95% bootstrapped CIs). Latency of significance (corrected for multiple comparison using FDR, $p < 0.01$) is indicated by a vertical dotted bar of the corresponding color. The *y*-axis corresponds to the change in accuracy for one SD change of the considered variable. ***D***, Temporal decoding (±95% CI) of the two clusters computed using the *k*-means algorithm on the entire temporal decoding curve computed for individual images.

over classically reported electrode locations from the right hemisphere (equivalent to P8 and PO8; Rossion and Jacques, 2008) showed an even smaller N170 amplitude (P8: mean: −2.7 μV [−4.0 −1.4]; PO10: mean: −0.6 μV [−1.5 0.3]), peaking, respectively, at 162 and 155 ms poststimulus onset (Fig. 3*B*). Using a paired two-tailed permutation test, based on the t_{max} statistic, we found no significant modulation in amplitude ($t_{max} = 3.30$, df = 27, $p > 0.05$) and no significant modulation in latency ($t_{max} = 2.77$, df = 27, $p > 0.05$) for the N170 (Fig. 3*B*).

We systematically tested an amplitude modulation (target vs distractor) for individual time points, using a paired two-tailed permutation test based on the t_{max} statistic (6144 comparisons; Fig. 3*C*). P08 and P8 exhibited a significant early amplitude modulation, respectively, at 137 ms and from 125 to 145 ms poststimulus ($t_{max} = 4.83$, df = 27, $p < 0.05$), while other significant modulations occur rather late (>230 ms). Thus, point-by-point analysis reveals significant modulation happening only in between or after the P100 (105 ms) or the N170 (180 ms). Overall, no or weak modulation of the early ERP components (P100 and N170) was found. Given that the use of faces in natural scenes may have disrupted the classic topography of the electrodes traditionally associated with face processing, we complement this analysis using MVPA, which, by pooling information across all electrodes, may more easily capture the dynamics of face processing.

Figure 4*B* shows the temporal decoding accuracy resulting from the MVPA averaged across all participants. This analysis reveals that significant (192 comparisons, $t_{max} = 3.54$, df = 27, $p < 0.05$) face category information can be readout at very short latencies, as early as 94 ms poststimulus onset. Interestingly, the EEG decoding accuracy does not seem to increase monotonically. Instead, the amount of face information available seems to fluctuate in time, suggesting the possible existence of discrete processing time windows.

To further characterize these time windows, we estimated a temporal derivative of the accuracy curve shown in Figure 4*B*. During an initial phase (~95–140 ms), the decoding accuracy increases monotonically (derivative constantly positive) until a plateau (derivative oscillates between positive and negative values) is reached (~60% accuracy around 140 ms after stimuli onset). During a third time window (~200–350 ms), a monotonic increase can be observed again (reaching ~80% of decoding accuracy around 350 ms), possibly reflecting the accrual of further face or motor information. After ~350 ms, decoding accuracy stabilizes and decreases slowly until 700 ms.

Decoding weight topographies (Fig. 4*B*) suggests that during the first 200 ms of visual processing, most of the information originates from occipitolateral electrodes, while for longer latencies, parietal electrodes seem to contribute more to the overall decoding. It is possible that these discrete time windows may reflect different levels of visual processing.

We thus fitted our classifier confidence for individual stimuli with a number of experimental variables. Here we consider low-level image statistics obtained by fitting the distribution of pixel intensities to a *Weibull* function as done by Scholte et al. (2009). Such low-level image statistics was shown to account for a significant fraction of the variance across single-image evoked potentials. We also considered face size as an additional low-level image property. Building up tolerance to 2D transformations is a hallmark of object processing in the ventral stream (Riesenhuber and Poggio, 1999). It is thus expected that face size should modulate low-level visual processing and less so higher level processes. Last, we considered median reaction times

Figure 5. Easiest and most difficult stimuli to decode organized by time periods indicated into square brackets (ms). In each group, the top/bottom 10 images with the highest/lowest decoding accuracy (ordered left to right) are shown. The average decoding accuracy by stimulus is printed below each image. The 11th images correspond to the average image of the 50 best/worst (top/bottom) stimuli.

across participants (*RT*) for individual stimuli as a marker of higher level decision processes.

Figure 4*C* shows the estimated regression coefficients over time between the classifier confidence derived from single-images (see Materials and Methods) and the three variables described above (*Weibull*, *size*, and *RT*). The analysis suggests a significant contribution ($p < 0.01$, uncorrected for multiple comparison) of the *Weibull* starting very early, ~70–80 ms poststimulus onset, followed by face *size* at ~105 ms and *RT* at ~125 ms.

We found these three variables (*Weibull*, *size*, and *RT*) to be only weakly correlated with one another (*Weibull* vs *size*: $r^2 = 0.01$, $p = 0.032$; *Weibull* vs *RT*: $r^2 = 0.01$, $p = 0.022$; *size* vs *RT*: $r^2 = 0.02$, $p = 0.002$). It is thus unlikely that correlation between these three variables would explain the observed correlations with the classifier confidence.

Two coefficient peaks can be observed—approximately corresponding to the processing windows described above. Before 200 ms, *Weibull*, *size*, and *RT* contribute to decoding accuracy peaking at ~140 ms. Beyond 200 ms, the contribution of the *Weibull* disappears and the contribution of the face size, although significant, is largely reduced. We conducted a clustering analysis directly on the decoding curve obtained for individual stimuli (see Materials and Methods). As shown on Figure 4*D*, one cluster accounting for 62% of the stimulus set seems to reflect a rapid decoding while a second cluster (38% of the stimuli) seems to reflect later decoding.

The 10 easiest and most difficult images to decode for each time window are shown in Figure 5. From these images, it seems that the complexity of the surrounding background clutter may influence the decoding accuracy. Shown on the right are composite averages computed over the top 50 easiest and most difficult images to decode. Stimuli that are well decoded during earlier phases appear more stereotypical and less variable than those that are difficult to decode. This trend seems less pronounced for later phases.

Discussion

The current study investigated the neural dynamics of face processing in natural scenes using EEG recordings. The underlying neural activity was correlated with both image properties and participants' RTs.

Consistent with previous studies on face processing, we observed two ERP components, namely the P1 and N170. However, differential activity for ERPs associated with go and no-go trials was modest and, contrary to numerous studies (Jeffreys, 1989; Bötzel et al., 1995; Bentin et al., 1996; George et al., 1996; Joyce and Rossion, 2005), no amplitude modulation on the N170 component was found. This could be due to the set of distractor stimuli used in the present study (animal faces) or to the fact that participants performed a go/no-go task rather than a yes/no task as in previous studies. Notwithstanding, these results are consistent with previous go/no-go studies that have shown no significant amplitude modulation (Thierry et al., 2007a, b; Dering et al., 2009, 2011) and small but significant latency effect of the N170 using human and animal faces embedded in natural scenes (Rousselet et al., 2004). The presence of background clutter in natural scenes could also have disrupted the classic topography of electrodes associated with face processing, while reducing the

ERP amplitudes (Rousselet et al., 2007b; Thierry et al., 2007a, b; Dering et al., 2009, 2011). This is consistent with both monkey electrophysiology (Desimone and Duncan, 1995; Zhang et al., 2011) and human imaging studies (Reddy and Kanwisher, 2007) that have shown that patterns of brain activity associated with object categories are disrupted by clutter. This would also be consistent with behavioral studies that have shown that background clutter hinders detection (Serre et al., 2007). In addition, the high variability of the face stimuli used in the present study compared with previous studies might have increased the intertrial jitter in the latency of the component, artificially decreasing its average amplitude (Rousselet et al., 2005; Thierry et al., 2007a, b; Dering et al., 2009, 2011). This hypothesis could be tested in future studies by looking at phase coherence and realigning single-trial ERPs (Navajas et al., 2013).

We therefore ran complementary analyses based on MVPA. In the context of the present EEG analysis, this procedure has the great advantage to summarize and quantify the neural information available for the task at hand (here, human face detection) across all electrodes for each time point.

The first important result provided by MVPA is that face category information could be readout very early, starting ~95 ms following stimulus onset. This latency is comparable to onsets of ~90–100 ms obtained by contrasting faces to noise patches (Bieniek et al., 2012; Rousselet, 2012) and is thus remarkable given the complexity and variable nature of the stimuli used in the present study. This very fast category-selective activity supports the claim that the visual system is highly optimized for the processing of natural scenes (Vinje and Gallant, 2000; Simoncelli and Olshausen, 2001). Our estimate is also consistent with previous studies that have reported that categorization information (faces vs objects) can be detected in <100 ms in humans (Liu et al., 2009; Dering et al., 2011).

Additionally, the EEG decoding accuracy did not seem to increase monotonically as would be expected from a pure decision process. We found three distinct phases, which overlap with the ERP component latencies described above: an initial phase starting ~95–140 ms poststimulus onset (P1 time window) followed by a plateau ~140–195 (N170 time window) and a later phase ~185–350 ms poststimulus onset.

Our analyses suggest that the earliest phase reflects low-level processes possibly implemented via an initial feedforward sweep of activity from V1 to occipitotemporal areas (Riesenhuber and Poggio, 1999). Consistent with this idea, we found that the decoding activity correlated well with low-level visual properties of the images (*Weibull* statistics ~75 ms followed by face *size* ~115 ms).

This result is consistent with psychophysics studies that have demonstrated a role played by low-level image statistics such as contrast (Scholte et al., 2009), power spectrum (Rossion and Caharel, 2011) or phase (Bieniek et al., 2012) during rapid object detection tasks. Our estimated latency of decoding is also consistent with the earliest behavioral responses observed in the saccadic choice paradigm, during which participants are asked to saccade toward faces (Kirchner and Thorpe, 2006; Crouzet et al., 2010). This early visual activity seems to be somehow linked to behavioral responses, since we observe a correlation with median RTs as early as ~125 ms poststimulus onset.

The second phase is characterized by a plateau in decoding activity, perhaps reflecting a transitory stabilization of the face information available. It is well established that the occipital face area (OFA) and the fusiform face area (FFA) are involved in face processing (Haxby et al., 2000, 2001; Gobbini and Haxby, 2006). Based on lesion studies, it has been proposed that these regions do not rely on feedforward processing (from posterior occipital areas to the OFA to the FFA), but on re-entrant signals from posterior areas to OFA via the FFA (Rossion, 2008). It is during that period that a high-level individual representation of the face is built (Rossion and Caharel, 2011). Hence, the plateau observed during the second phase could be due to the time needed for this re-entrant processing to take place and switch from a purely externally to an internally driven information processing stage as suggested by the decrease of correlation with low-level statistics and behavior. Numerous studies have reported a similar category-selective activity between 140 and 180 ms leading to the hypothesis that this activity could reflect the build-up of an internal representation of the stimulus independent of low-level visual properties (Schyns et al., 2007; van Rijsbergen and Schyns, 2009). This phase would be necessary to drive behavioral go/no-go responses (VanRullen and Thorpe, 2001) and decision making (Philiastides and Sajda, 2006), as shown by the second phase of correlation with behavioral responses.

The third phase, starting at ~185 ms, is associated with a very significant increase in decoding accuracy. Using a memory task in epileptic patients, it has been previously shown that the coherence of face-selective activity increases in a widespread network of regions including the temporal, parietal, and frontal lobes during a similar time window (from ~160 to 230 ms poststimulus onset; Klopp et al., 2000). Similarly, a period of massively parallel processing has been identified in the entire visual ventral stream starting at ~180 ms and peaking at 240 ms during a face recognition task using intracerebral recordings (Barbeau et al., 2008). Hence, this third phase could reflect the involvement of a distributed network of brain areas in contrast with previous stages related to the activation of relatively posterior visual areas (first stage) or a local network involving posterior areas as well as the OFA and FFA (second stage). This third phase could be associated with conscious access to the face representation ("I know that it is a face"; Sergent et al., 2005; Railo et al., 2011).

Overall, the current study shows a promising application of MVPA techniques to surface electrophysiological signals with an unknown topography and a focus on the temporal dynamics of processing. While MVPA has been extensively used in the context of functional magnetic resonance imaging studies, the use of decoding techniques for M/EEG analysis has been mainly limited to the field of brain computer interface (P300 speller; Farwell and Donchin, 1988). Very few studies have investigated the possibility to read out visual category information from noninvasive human electrophysiological signals. Among them, an MEG study has demonstrated the possible readout of basic object category but with late latencies (incompatible with behavioral results such as those reported in saccadic choice tasks; Crouzet et al., 2010) despite the fact that isolated and cropped stimuli of faces, houses, and other textures were used (Carlson et al., 2011). Another EEG study reached higher decoding accuracy for line drawings of animals versus tools (Simanova et al., 2010). However, the study used a small number of stimuli with many repetitions and evident low-level visual differences between categories. Here, instead, we used a large database of variable natural scenes without any repetition. In any case, future work will be needed to compare more directly and formally the usefulness of MVPA to other recent univariate or multivariate EEG analyses.

In conclusion, we extend previous results and verify that the dynamics of face processing identified using ERPs also applies to faces seen in complex, naturalistic scenes.

References

Allen EA, Erhardt EB, Calhoun VD (2012) Data visualization in the neurosciences: overcoming the curse of dimensionality. Neuron 74:603–608. CrossRef Medline

Barbeau EJ, Taylor MJ, Regis J, Marquis P, Chauvel P, Liégeois-Chauvel C (2008) Spatio temporal dynamics of face recognition. Cereb Cortex 18: 997–1009. CrossRef Medline

Barragan-Jason G, Lachat F, Barbeau EJ (2012) How fast is famous face recognition? Front Psychol 3:454. CrossRef Medline

Barragan-Jason G, Besson G, Ceccaldi M, Barbeau EJ (2013) Fast and famous: looking for the fastest speed at which a face can be recognized. Front Psychol 4:100. CrossRef Medline

Bentin S, Allison T, Puce A, Perez E, McCarthy G (1996) Electrophysiological studies of face perception in humans. J Cogn Neurosci 8:551–565. CrossRef Medline

Besson G, Ceccaldi M, Didic M, Barbeau EJ (2012) The speed of visual recognition memory. Vis Cogn 20:1131–1152. CrossRef

Bieniek MM, Pernet CR, Rousselet GA (2012) Early ERPs to faces and objects are driven by phase, not amplitude spectrum information: evidence from parametric, test-retest, single-subject analyses. J Vis 12(13):12. CrossRef Medline

Blair RC, Karniski W (1993) An alternative method for significance testing of waveform difference potentials. Psychophysiology 30:518–524. CrossRef Medline

Bötzel K, Schulze S, Stodieck SR (1995) Scalp topography and analysis of intracranial sources of face-evoked potentials. Exp Brain Res 104:135–143. Medline

Carlson TA, Hogendoorn H, Kanai R, Mesik J, Turret J (2011) High temporal resolution decoding of object position and category. J Vis 11(10). pii:9. CrossRef Medline

Cauchoix M, Arslan AB, Fize D, Serre T (2012) The neural dynamics of visual processing in monkey extrastriate cortex: a comparison between univariate and multivariate techniques. In: Machine learning and interpretation in neuroimaging, pp. 164–171. New York: Springer.

Clarke A, Taylor KI, Devereux B, Randall B, Tyler LK (2013) From perception to conception: how meaningful objects are processed over time. Cereb Cortex 23:187–197. CrossRef Medline

Crouzet SM, Kirchner H, Thorpe SJ (2010) Fast saccades toward faces: face detection in just 100 ms. J Vis 10(4):16.1–17. CrossRef Medline

Delorme A, Makeig S (2004) EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. J Neurosci Methods 134:9–21. CrossRef Medline

Delorme A, Sejnowski T, Makeig S (2007) Enhanced detection of artifacts in EEG data using higher-order statistics and independent component analysis. Neuroimage 34:1443–1449. CrossRef Medline

Dering B, Martin CD, Thierry G (2009) Is the N170 peak of visual event-related brain potentials car-selective? Neuroreport 20:902–906. CrossRef Medline

Dering B, Martin CD, Moro S, Pegna AJ, Thierry G (2011) Face-sensitive processes one hundred milliseconds after picture onset. Front Hum Neurosci 5:93. CrossRef Medline

Desimone R, Duncan J (1995) Neural mechanisms of selective visual attention. Annu Rev Neurosci 18:193–222. CrossRef Medline

Eimer M, Holmes A (2002) An ERP study on the time course of emotional face processing. Neuroreport 13:427–431. CrossRef Medline

Fabre-Thorpe M (2011) The characteristics and limits of rapid visual categorization. Front Psychol 2:243. Medline

Farwell LA, Donchin E (1988) Talking off the top of your head: toward a mental prosthesis utilizing event-related brain potentials. Electroencephalogr Clin Neurophysiol 70:510–523. CrossRef Medline

George N, Evans J, Fiori N, Davidoff J, Renault B (1996) Brain events related to normal and moderately scrambled faces. Brain Res Cogn Brain Res 4:65–76. CrossRef Medline

Gobbini MI, Haxby JV (2006) Neural response to the visual familiarity of faces. Brain Res Bull 71:76–82. CrossRef Medline

Groppe DM, Urbach TP, Kutas M (2011) Mass univariate analysis of event-related brain potentials/fields I: a critical tutorial review. Psychophysiology 48:1711–1725. CrossRef Medline

Halgren E, Raij T, Marinkovic K, Jousmäki V, Hari R (2000) Cognitive response profile of the human fusiform face area as determined by MEG. Cereb Cortex 10:69–81. CrossRef Medline

Hauk O, Davis MH, Ford M, Pulvermüller F, Marslen-Wilson WD (2006) The time course of visual word recognition as revealed by linear regression analysis of ERP data. Neuroimage 30:1383–1400. CrossRef Medline

Haxby JV, Hoffman EA, Gobbini MI (2000) The distributed human neural system for face perception. Trends Cogn Sci 4:223–233. CrossRef Medline

Haxby JV, Gobbini MI, Furey ML, Ishai A, Schouten JL, Pietrini P (2001) Distributed and overlapping representations of faces and objects in ventral temporal cortex. Science 293:2425–2430. CrossRef Medline

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

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

Joyce C, Rossion B (2005) The face-sensitive N170 and VPP components manifest the same brain processes: the effect of reference electrode site. Clin Neurophysiol 116:2613–2631. CrossRef Medline

Kamitani Y, Tong F (2005) Decoding the visual and subjective contents of the human brain. Nat Neurosci 8:679–685. CrossRef Medline

Kirchner H, Thorpe SJ (2006) Ultra-rapid object detection with saccadic eye movements: visual processing speed revisited. Vision Res 46:1762–1776. CrossRef Medline

Klopp J, Marinkovic K, Chauvel P, Nenov V, Halgren E (2000) Early widespread cortical distribution of coherent fusiform face selective activity. Hum Brain Mapp 11:286–293. CrossRef Medline

Lage-Castellanos A, Martínez-Montes E, Hernández-Cabrera JA, Galán L (2010) False discovery rate and permutation test: an evaluation in ERP data analysis. Stat Med 29:63–74. Medline

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

Liu J, Harris A, Kanwisher N (2002) Stages of processing in face perception: an MEG study. Nat Neurosci 5:910–916. CrossRef Medline

Makeig S, Jung TP, Bell AJ, Ghahremani D, Sejnowski TJ (1997) Blind separation of auditory event-related brain responses into independent components. Proc Natl Acad Sci U S A 94:10979–10984. CrossRef Medline

Maris E, Oostenveld R (2007) Nonparametric statistical testing of EEG-and MEG-data. J Neurosci Methods 164:177–190. CrossRef Medline

Navajas J, Ahmadi M, Quian Quiroga R (2013) Uncovering the mechanisms of conscious face perception: a single-trial study of the N170 responses. J Neurosci 33:1337–1343. CrossRef Medline

Parra LC, Spence CD, Gerson AD, Sajda P (2005) Recipes for the linear analysis of EEG. Neuroimage 28:326–341. CrossRef Medline

Philiastides MG, Sajda P (2006) Temporal characterization of the neural correlates of perceptual decision making in the human brain. Cereb Cortex 16:509–518. Medline

Railo H, Koivisto M, Revonsuo A (2011) Tracking the processes behind conscious perception: a review of event-related potential correlates of visual consciousness. Conscious Cogn 20:972–983. CrossRef Medline

Reddy L, Kanwisher N (2007) Category selectivity in the ventral visual pathway confers robustness to clutter and diverted attention. Curr Biol 17: 2067–2072. CrossRef Medline

Riesenhuber M, Poggio T (1999) Hierarchical models of object recognition in cortex. Nat Neurosci 2:1019–1025. CrossRef Medline

Rossion B (2008) Constraining the cortical face network by neuroimaging studies of acquired prosopagnosia. Neuroimage 40:423–426. CrossRef Medline

Rossion B, Caharel S (2011) ERP evidence for the speed of face categorization in the human brain: disentangling the contribution of low-level visual cues from face perception. Vision Res 51:1297–1311. CrossRef Medline

Rossion B, Jacques C (2008) Does physical interstimulus variance account for early electrophysiological face sensitive responses in the human brain? Ten lessons on the N170. Neuroimage 39:1959–1979. CrossRef Medline

Rossion B, Delvenne JF, Debatisse D, Goffaux V, Bruyer R, Crommelinck M, Guérit JM (1999) Spatio-temporal localization of the face inversion effect: an event-related potentials study. Biol Psychol 50:173–189. CrossRef Medline

Rousselet GA (2012) Does filtering preclude us from studying ERP time-courses? Front Psychol 3:131. Medline

Rousselet GA, Pernet CR (2011) Quantifying the time course of visual object processing using ERPs: it's time to up the game. Front Psychol 2:107. Medline

Rousselet GA, Macé MJ, 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. Medline

Rousselet GA, Macé MJ-M, Fabre-Thorpe M (2004) Animal and human faces in natural scenes: how specific to human faces is the N170 ERP component? J Vis 4(1):13–21. Medline

Rousselet GA, Husk JS, Bennett PJ, Sekuler AB (2005) Spatial scaling factors explain eccentricity effects on face ERPs. J Vis 5(10):755–763. Medline

Rousselet GA, Husk JS, Bennett PJ, Sekuler AB (2007a) Single-trial EEG dynamics of object and face visual processing. Neuroimage 36:843–862. CrossRef Medline

Rousselet GA, Macé MJM, Thorpe SJ, Fabre-Thorpe M (2007b) Limits of event-related potential differences in tracking object processing speed. J Cogn Neurosci 19:1241–1258. CrossRef Medline

Scholte HS, Ghebreab S, Waldorp L, Smeulders AWM, Lamme VAF (2009) Brain responses strongly correlate with Weibull image statistics when processing natural images. J Vis 9(4):29.1–15. CrossRef Medline

Schyns PG, Petro LS, Smith ML (2007) Dynamics of visual information integration in the brain for categorizing facial expressions. Curr Biol 17: 1580–1585. CrossRef Medline

Sergent C, Baillet S, Dehaene S (2005) Timing of the brain events underlying access to consciousness during the attentional blink. Nat Neurosci 8:1391–1400. CrossRef Medline

Serre T, Oliva A, Poggio T (2007) A feedforward architecture accounts for rapid categorization. Proc Natl Acad Sci U S A 104:6424–6429. CrossRef Medline

Simanova I, van Gerven M, Oostenveld R, Hagoort P (2010) Identifying object categories from event-related EEG: toward decoding of conceptual representations. PLoS One 5:e14465. CrossRef Medline

Simoncelli EP, Olshausen BA (2001) Natural image statistics and neural representation. Annu Rev Neurosci 24:1193–1216. CrossRef Medline

Thierry G, Martin CD, Downing P, Pegna AJ (2007a) Controlling for interstimulus perceptual variance abolishes N170 face selectivity. Nat Neurosci 10:505–511. Medline

Thierry G, Martin CD, Downing PE, Pegna AJ (2007b) Is the N170 sensitive to the human face or to several intertwined perceptual and conceptual factors? Nat Neurosci 10:802–803. CrossRef

Torralba A (2009) How many pixels make an image. Vis Neurosci 26:123–131. CrossRef Medline

van Rijsbergen NJ, Schyns PG (2009) Dynamics of trimming the content of face representations for categorization in the brain. PLoS Comput Biol 5:e1000561. CrossRef Medline

VanRullen R, Thorpe SJ (2001) The time course of visual processing: from early perception to decision-making. J Cogn Neurosci 13:454–461. CrossRef Medline

Vinje WE, Gallant JL (2000) Sparse coding and decorrelation in primary visual cortex during natural vision. Science 287:1273–1276. CrossRef Medline

Zhang Y, Meyers EM, Bichot NP, Serre T, Poggio TA, Desimone R (2011) Object decoding with attention in inferior temporal cortex. Proc Natl Acad Sci U S A 108:8850–8855. CrossRef Medline