An objective signature for visual binding of face parts in the human brain

Adriano Boremanse
Institute of Research in Psychology and Institute of Neuroscience, University of Louvain, Louvain la Neuve, Belgium

Anthony M. Norcia
Department of Psychology, Stanford University, Stanford, CA, USA

Bruno Rossion
Institute of Research in Psychology and Institute of Neuroscience, University of Louvain, Louvain la Neuve, Belgium

Whether and how the parts of a visual object are grouped together to form an integrated (“holistic”) representation is a central question in cognitive neuroscience. Although the face is considered to be the quintessential example of holistic representation, this issue has been the subject of much debate in face perception research. The implication of holistic processing is that the response to the whole cannot be predicted from the sum of responses to the parts. Here we apply techniques from nonlinear systems analysis to provide an objective measure of the nonlinear integration of parts into a whole, using the left and right halves of a face stimulus as the parts. High-density electroencephalogram (EEG) was recorded in 15 human participants presented with two halves of a face stimulus, flickering at different frequencies (5.88 vs. 7.14 Hz). Besides specific responses at these fundamental frequencies, reflecting part-based responses, we found intermodulation components (e.g., 7.14 – 5.88 = 1.26 Hz) over the right occipito-temporal hemisphere, reflecting nonlinear integration of the face halves. Part-based responses did not depend on the relative alignment of the two face halves, their spatial separation, or whether the face was presented upright or inverted. By contrast, intermodulations were virtually absent when the two halves were spatially misaligned and separated. Inversion of the whole face configuration also reduced specifically the intermodulation components over the right occipito-temporal cortex. These observations indicate that the intermodulation components constitute an objective, configuration-specific signature of an emergent neural representation of the whole face that is distinct from that generated by the parts themselves.

Introduction

How are the parts of a visual object combined into a holistic representation? This fundamental question of visual neuroscience is one of the most difficult to resolve, and to date the answer remains largely a mystery (Spillmann & Ehrenstein, 1996; Ungerleider & Bell, 2011). Gestalt psychologists claimed that a whole visual stimulus is qualitatively different than the sum of its parts and defined a set of principles of part organization that govern the perception of a stimulus as a whole (Wertheimer, 1923; see Wagemans et al., 2012 for a review). Starting in the second half of the 20th century, neurophysiological recordings in the monkey brain have begun to suggest how holistic processing might emerge over a succession of processing stages. Early studies demonstrated that neurons in visual areas such as the primary visual cortex (V1) and early extrastriate areas (V2, V3, V4, MT) have small receptive fields and are sensitive to simple features such as contrast, orientation of line segments, spatial frequency, color, or motion (Spillmann & Ehrenstein, 1996; Werner & Chalupa, 2004; Ungerleider & Bell, 2011). By contrast, higher order areas of the inferior temporal cortex (IT) contain neurons with much larger receptive fields and are sensitive to simple features such as contrast, orientation of line segments, spatial frequency, color, or motion (Spillmann & Ehrenstein, 1996; Werner & Chalupa, 2004; Ungerleider & Bell, 2011). By contrast, higher order areas of the inferior temporal cortex (IT) contain neurons with much larger receptive fields and are sensitive to simple features such as contrast, orientation of line segments, spatial frequency, color, or motion (Spillmann & Ehrenstein, 1996; Werner & Chalupa, 2004; Ungerleider & Bell, 2011). By contrast, higher order areas of the inferior temporal cortex (IT) contain neurons with much larger receptive fields and are sensitive to simple features such as contrast, orientation of line segments, spatial frequency, color, or motion (Spillmann & Ehrenstein, 1996; Werner & Chalupa, 2004; Ungerleider & Bell, 2011). By contrast, higher order areas of the inferior temporal cortex (IT) contain neurons with much larger receptive fields and are sensitive to simple features such as contrast, orientation of line segments, spatial frequency, color, or motion (Spillmann & Ehrenstein, 1996; Werner & Chalupa, 2004; Ungerleider & Bell, 2011). By contrast, higher order areas of the inferior temporal cortex (IT) contain neurons with much larger receptive fields and are sensitive to simple features such as contrast, orientation of line segments, spatial frequency, color, or motion (Spillmann & Ehrenstein, 1996; Werner & Chalupa, 2004; Ungerleider & Bell, 2011).
a review and historical perspective). Further single-cell studies at intermediate stages of the visual pathway suggested a hierarchy of processes along the ventral visual pathway through which parts, then simple configurations of parts, are combined into whole objects (e.g., Brincat & Connor, 2004). The process culminates in the most anterior region of IT, the center of object recognition (Ungerleider & Haxby, 2011).

In humans, functional magnetic resonance imaging (fMRI) studies have found that the ventral visual pathway, extending from the primary visual cortex to higher visual areas, is increasingly sensitive to the whole of an object (Lerner, Hendler, Ben-Bashat, Harel, & Malach, 2001; Schiltz & Rossion, 2006; Kubilius, Wagemans, & Op de Beeck, 2011).

However, whether the representation of a whole object is different from the summed representation of its parts even in high-level visual areas remains largely unclear. This is because integration of visual parts is commonly determined by comparing the summed response to each of the parts presented in isolation to the response to the whole visual stimulus. The rationale is that under no integration, the response to the whole stimulus should be a linear combination (i.e., sum) of the response to the two input parts. However, this approach is indirect, and its findings can be difficult to interpret. For instance, some IT neurons do not respond to any part of a moderately complex shape presented in isolation but discharge only to the combination of the parts (Fujita et al., 1992; Tsunoda, Yamane, Nishizaki & Tanifuji, 2001; Baker, Behrmann, & Olson, 2002; Brincat & Connor, 2004; Yamane, Tsunoda, Matsumoto, Phillips, & Tanifuji, 2006). Although such observations suggest nonlinear mechanisms, the neurons may respond to simple juxtapositional features created where the parts abut (Sripati & Olson, 2010). Secondly, because the part and whole responses are measured at separate times, it can be difficult to control the state of arousal or attention under the two conditions and variation in these factors can spuriously suggest that nonlinear combination of the parts has occurred.

These problems present particular challenges for understanding how faces are perceived. The face is considered by many as the quintessential whole, or Gestalt (Pomerantz & Kubovy, 1986; Biederman & Kalocsai, 1997; Palmer, 1999). Numerous behavioral studies have shown that performance at recognizing a facial part is influenced by the insertion of this part into a whole face stimulus, and these results have been interpreted as indicative of a nonlinear combination of face parts into a whole. For instance, face parts are better recognized if they are presented in a whole face than if they are presented in isolation (Tanaka & Farah, 1993). In the same vein, performance in recognizing the top half of a face (i.e., a part) is lower if it is aligned than if it is misaligned with the bottom half of another face (Young, Hellawell, & Hay, 1987; Rossion, 2013 for a review). This approach has been extended to the neural level, using neuroimaging. For instance, the blood oxygen level-dependent (BOLD) response to a face half in the right fusiform gyrus differs if the face half is aligned (forming a whole face) versus misaligned with the other half of a face (Schiltz & Rossion, 2006; see also Harris & Aguirre, 2008; Andrews, Davies-Thompson, Kingstone, & Young, 2010; Schiltz, Dricot, Goebel, & Rossion, 2010; Goffaux, Schiltz, Mur & Goebel, 2012).

Collectively, these findings have been generally taken as evidence for holistic face perception—namely the perception of a face in which the parts are glued into an integrated representation. In this view, the representation of a whole face is somehow different than the mere sum of independent face parts. However, studies that have quantified the response to each face part and compared the summed response to the response to the whole have reported discrepant results. Behavioral performance at recognizing a whole face may not be better than for a single diagnostic isolated face part, leading to the conclusion that a face is no more than the sum of its parts (Gold, Mundy, & Tjan, 2012; but see Sergent, 1984). In contrast, an fMRI study that compared responses to various combinations of face parts showed that responses to configurations of parts do not equal the sum of responses to the parts separately. In a posterior face-selective area (occipital face area, OFA), the response to an image with both eyes present was equal to that from an image with only the right eye. In the mid fusiform face-selective area (FFA), responses to combinations of parts and single parts all produced the same activation (Arcurio, Gold, & James, 2012). Each of these effects indicates that the responses to part combinations is subadditive. In the same vein, the responses of single neurons of the monkey IT cortex to isolated parts or to whole faces, also showing additivity failures, remain ambiguous (Kobatake & Tanaka, 1994; Freiwald, Tsao, & Livingstone, 2009).

One reason for this failure is that additivity experiments are sufficient to demonstrate the presence of nonlinear combination, but they give only a limited characterization of it. When additivity fails (i.e., the sum of the separate part responses is not equal to the measured response to the parts presented together), it is impossible to determine how much of the measured nonlinearity comes from a change in the part responses, the addition of a new nonlinear interaction process, or both, as these are lumped together in the difference signal formed by the additivity test. Another important issue is more specific to faces. Unlike objects that can be formed of simple meaningful elements (e.g., a triangle formed of three connected lines), a facial part...
such as a single eye is meaningful by itself and may be sufficient to activate an entire representation by completion. This is particularly true for half of a face, which is sufficient to activate representation of the other half. In this condition, the response to the whole face may be lower than the summed responses to the face halves presented separately, suggesting nonlinearity. For this reason, paradigms that are used successfully with simple nonface stimuli to determine the "emergent features" of a Gestalt (e.g., Pomerantz & Portillo, 2011) may give unexpected outcomes when applied to faces (Pomerantz et al., 2003).

In order to provide a direct test of the hypothesis that a whole face stimulus is qualitatively different than the sum of its parts, it is necessary to use an approach that allows the experimenter to separate the responses arising from the representation of each of the parts presented simultaneously from those generated by the representation of the whole visual stimulus that they form. Ideally, the approach should address the part/whole issue at the global scale of visual system organization. These conditions are met by the "frequency-tagging" approach obtained from the electroencephalogram (EEG) recorded on the human scalp (Regan & Heron, 1969; Regan & Cartwright, 1970). The approach takes advantage of the fact that presenting a periodic visual stimulus to the human brain leads to periodic responses directly related to the frequency of stimulation (the "steady state visual evoked potential," ssVEP, Regan, 1966). This property is very useful in the design of highly selective filters that can define the signal objectively (i.e., the response to the stimulus at the precise, experimentally defined stimulation frequency). Most importantly, the method can be used to record responses from multiple simultaneously presented stimuli "tagged" at different frequencies (e.g., \( f_1 \) and \( f_2 \)), and disentangle objectively their contribution to the brain’s response ("frequency-tagging," Regan & Heron, 1969; Regan & Cartwright, 1970; Regan & Regan, 1988; Morgan, Hansen, & Hillyard, 1996; Appelbaum, Wade, Vildavski, Pettet, & Norcia, 2006; Andersen, Muller, & Hillyard, 2009).

In the two-frequency input method, part responses are still accessible when the whole configuration is present and it is thus possible to track what happens to the part responses (their suppression or facilitation) as well as getting a second very specific indication of the strength and temporal dynamics of the nonlinear interaction. This second indicator of nonlinear interaction manifests itself in the form of responses at “intermodulation” (IM) frequencies (e.g., \( f_1 \pm f_2 \)). Responses at these frequencies can only be generated if there is a population of neurons that receives inputs from the two stimuli, and only if those stimuli interact nonlinearly within that population (Zemon & Ratliff, 1984; Victor & Conte, 2000; Appelbaum, Wade, Pettet, Vildavski, & Norcia, 2008; Giani et al., 2012).

IM components have been used previously to study nonlinear interactions between simple inputs such as those arriving from the two eyes, these interactions being lost when the participant is stereo blind (Baitch & Levi, 1988; Norcia, Harrad, & Brown, 2000). The strength of nonlinear interactions between contrast-modulated stimuli depends strongly on spatial separation (Zemon & Ratliff, 1984; Norcia, Wesemann, & Manny, 1999; Victor & Conte, 2000; Appelbaum et al., 2006; Fuchs, Andersen, Gruber, & Müller, 2008) relative alignment (Norcia et al., 1999; Victor and Conte, 2000) or relative contrast (Tsai, Wade, & Norcia, 2012) of the different stimulus regions. Increased IM components have also been observed when two hemifield images flickering at different frequencies and presented to each eye are perceived as combined (“interocular grouping”) rather than with one suppressing the other (“binocular rivalry,” Sutoyo & Srinivasan, 2009).

Here we used this approach to objectively identify a signature of the representation of the simultaneously presented parts and of the whole of a face. Each of the left and right halves of a face were tagged with different frequencies (i.e., \( f_1 \): 5.88 Hz and \( f_2 \): 7.14 Hz) while recording high-density EEG in 15 human observers. Objective representation of parts was defined as the EEG response at each of the two fundamental frequencies and their harmonics (e.g., \( 2f_1 \): 11.76 Hz). Beyond this part-based representation, we expected to observe responses at intermodulation (IM) frequencies (e.g., \( f_1 - f_2 \): 1.26 Hz; \( f_1 + f_2 \): 13.01 Hz), which could be used as an objective signature of a holistic (i.e., integrated) face representation. Moreover, we predicted that when introducing a physical gap between the parts, or spatially misaligning them, the response at intermodulation components—but not the responses to parts—should be largely reduced due to a failure to integrate the parts under these unnatural conditions. Finally, because inversion is thought to impair holistic perception of faces (Young et al., 1987; Tanaka & Farah, 1993; Rossion, 2008 for a review), we also ran a second experiment that included inverted faces, making the prediction that the IM should be reduced for these inverted faces, either specifically or relatively more than the part-based responses.

**Methods**

**Participants**

Fifteen healthy adults (two left-handed; four males; age range: 18–37, mean age 23.6) with normal or
corrected-to-normal vision participated in the study for payment. Eleven different adults took part in second experiment that included the exact same conditions, with the addition of inverted faces (all right handed; two males; age 19–28, mean age 23). Written informed consent was obtained from all participants prior to the experiments, which were approved by the Biomedical ethical committee of the University of Louvain.

**Stimuli**

Fifteen full-front color pictures of faces from a well-known database (The Max Planck Institute, Tübingen, Germany) were used. To make the stimuli, each face was vertically cut in half (Figure 1A). Each half face subtended a visual angle of approximately 6.04° in height and 4.3° in width. The face stimuli were unfamiliar to the participants.

**Procedure**

After electrode-cap placement, participants were seated in a light- and sound-attenuated room at a viewing distance of 100 cm (there was no chin rest, but instructions to maintain constant distance were provided to the participants and controlled regularly). Stimuli were displayed on a light grey background (38 cd/m²) using an in-house application (SinStim) written in MATLAB (MathWorks Inc., Natick, MA). In the main condition, a whole face divided into a left and a right half that flickered at different frequencies was presented for 70 s (Movie 1). The two half faces were sinusoidally contrast-modulated, one at 5.88 Hz and the other one at 7.14 Hz (Figure 1). Each participant was presented with a single face throughout the entire set of recordings. Each of the 15 participants saw one out of 15 different faces. Although presenting the same face might lead to repetition suppression (“neural adaptation”) effects on the face-related EEG periodic response (Rossion & Boremanse, 2011; Rossion, Prieto, Boremanse, Kuefner, & Van Belle, 2012), face identity was held constant for a given participant in order to facilitate perceptual binding of the two halves.

Two control conditions were used in the main experiment. In the first control condition (gap), the half faces were spatially separated by a 0.3° gap that matched the luminance of the background. In the second control condition (misaligned), the half faces were spatially misaligned along the vertical axis (Figure 1C). The two control conditions, gap and misaligned, were designed to disrupt local interactions and whole object integration related processes, respectively. The gap condition was inspired by previous EEG frequency-tagging experiments investigating local interactions for low-level visual stimuli. For instance, a decrease of figure/background interaction has been reported following the introduction of small gaps between the two regions (Appelbaum et al., 2008). Misaligning the face

---

**Figure 1. Schematic illustration of the stimulation.** (A) and (B) The left and right face halves are contrast-modulated at different frequency rates. Contrast modulation was between 0% (background screen) and 100% (full half face picture). Here the left part (thick red line) flickers at 7.14 Hz and the right part (dashed blue line) at 5.88 Hz. (C) The three conditions of the study: whole face, left and right halves spatially separated (gap condition), left and right halves vertically misaligned (misaligned condition).
halves was done to disrupt the integration of object parts while keeping the physical border between the two parts of equal length. Thus, in the second control, the face halves were also elongated (7° in height) so that they shared the same length of common border as in the normal whole-face condition. The width of the face was kept identical to the other two conditions in order to minimize any potential amplitude increase of the ssVEP due to an increase of stimulus size (Regan, 1989). This slight distortion of the stimulus was unnoticed by the participants, who were asked specifically about their impression of the elongated faces after the experiment.

Each participant performed each condition four times, for a total of 12 blocks of 70 seconds each. Such long duration windows, which provide a high-frequency resolution, have been used in a number of previous ssVEP studies, including recent studies with contrast-modulated faces (Rossion & Boremanse, 2011; Rossion et al., 2012). For half of the blocks, the right side of the face (left visual field) was contrast-modulated at 5.88 Hz and left aside (right visual field) at 7.14 Hz, and for the remaining blocks the modulation was reversed. Close temporal frequencies were used because the ssVEP response characteristics depend on the temporal frequency of the visual input and may have an influence on the emergence of IM terms (Regan, 1989; Srinivasan, Russell, Edelman, & Tononi, 1999). The specific stimulation frequencies were constrained by the refresh rate of the monitor (i.e., integers of 100 Hz, 100/5.88 = 17; 100/7.14 = 14) and were chosen based on the consideration of several factors. First, stimulation frequencies between 4 Hz and 9 Hz (theta-band) appear to elicit maximal responses to contrast-modulated face stimulation as used here (Prieto, Van Belle, Liu-Shuang, Norcia, & Rossion, in press). In addition, both the fundamental frequencies of the evoked response, as well as at the second-order sum intermodulation (IM) component (i.e., 5.88 + 7.14 = 13.02 Hz) were located outside of the typical alpha frequency band (8–12 Hz), in order to maximize signal-to-noise ratio of these responses (Regan, 1989). In the first six blocks, each participant performed the full set of conditions counterbalanced for the side of higher frequency stimulation. The order of presentation of the six blocks was randomized across participants. The next six blocks were presented in reversed order for each participant, repeating the
experiment. Thus, every participant was tested in the three conditions (whole face, gap, misaligned) for a total of 12 blocks. The total time of the experiment did not exceed 20 min (with pauses). During each block, participants were instructed to fixate a small black cross located in the center of the face, below the eye line. They had to detect brief color changes (200 ms) (black to red) of the fixation cross (length: 0.29 line. They had to detect brief color changes (200 ms) of the fixation cross (length: 0.29° and width: 0.06°; six to eight changes per block) by pressing a response key. The goal of this task was to ensure that participants maintain a consistent level of attention throughout the duration of the block for all conditions (Rossion & Boremanse, 2011). There were no differences between conditions in terms of hit rates and correct reaction times (RTs) in this simple task (95% accuracy, mean RTs of about 600 ms). Triggers coding for each face half presentation were sent from the parallel port of the stimulation computer to the EEG recording computer at the beginning and end of the sequences and at each minimal level of visual stimulation for each half face (the frame where only the uniform background was present, Figure 1, see also Rossion & Boremanse, 2011; Rossion et al., 2012).

EEG recording

The EEG was recorded using 128 Ag/AgCl electrodes placed on the scalp using a cap system (Waveguard128 cap, Cephalon A/S, Denmark). Electrode positions included the standard 10–20 system locations and additional intermediate sites (for a 2-D mapping of electrode labels and positions, see http://www.ant-neuro.com/products/caps/waveguard/layouts/128/). The ground electrode was positioned on the forehead. Vertical and horizontal eye movements were monitored using four additional electrodes placed on the outer canthus of each eye and over the inferior and superior areas of the right orbit. During EEG recording, all electrodes were referenced to a centro-frontal channel (AFZ) and electrode impedances were kept below 10 kΩ. EEG was digitized at a 1000 Hz (128 channel high-speed amplifier, Advanced Neuro Technology, The Netherlands) and a digital anti-aliasing filter of 0.27 × sampling rate was applied at recording (at 1000 Hz sampling rate, the usable bandwidth is 0 to ~ 270 Hz).

Data analysis

EEG analyses were carried out using Analyzer 2.0 (Brain Products, Germany), Letswave (http://nocions.webnode.com/letswave) (Mouraux & Iannetti, 2008), MATLAB and EEGLAB (http://sccn.ucsd.edu). Blocks of EEG records (12 blocks for each participant) were cropped into 66.67 s windows containing exactly 392 and 476 cycles, respectively, of the two input frequencies \( f_1 \) (5.88 Hz) and \( f_2 \) (7.14 Hz). The first 2 s were removed to avoid contamination from transient responses triggered by the onset of visual stimulation train and to allow time for the system to be entrained by the stimulation. The EEG recording blocks were band-pass filtered between 0.1 and 100 Hz (Butterworth zero-phase filter, 24 db/octave – fourth order) to remove slow drifts in the recorded signals and very high frequencies. The signals were then downsampled to 250 Hz in order to reduce file size and increase the speed of data processing. Next the segmented EEG recordings were averaged together for each condition and each participant separately (four blocks averaged by condition for each participant). Averaging in the time domain maintains the complex phase of the response while canceling out EEG activity that is not phase locked with the stimulus. The resulting average waveforms were then transformed into the frequency domain using a discrete Fourier transform (FFT) yielding an amplitude spectrum (\( \mu \)V) at values ranging from 0 to 100 Hz with a very high-frequency resolution (1/66.67 = 0.015 Hz). An advantage of such a high-frequency resolution provided by the analysis of a long recording window is that all of the responses of interest, and thus all of the potential differences between conditions, can be concentrated in a discrete frequency band around the frequency of interest (Regan, 1989; Rossion et al., 2012). This frequency band occupies a very small fraction of the total EEG bandwidth. By contrast, biological noise is distributed throughout the EEG spectrum, resulting in a potentially very high signal-to-noise ratio (SNR) at the bandwidth of interest (Regan, 1989). At each electrode, SNR spectra were computed by taking the value at each frequency bin and dividing it by the average value of the 20 neighboring bins (Srinivasan et al., 1999; Rossion & Boremanse, 2011). Finally, the spectra were averaged across participants for each condition separately, for the display of grand-averaged spectra and topographical maps.

Topographical maps as well as previous studies of periodic visual stimulation with faces at these frequency rates (Rossion & Boremanse, 2011; Ales, Farzin, Rossion, & Norcia, 2012; Rossion et al., 2012) were used to define regions of interest (ROIs) for statistical tests. For the main analyses (parts and difference IM components), these two ROIs included eight right occipito-temporal channels (P06, P8, P06h, P6, PPO10h, TPP8h, P08, P010) as well as the left homologous ROI (i.e., P05, P7, P05h, P5, PPO9h, TPP7h, P07, PO9). Separate statistical analyses on these ROIs were performed based on individual participants’ values using repeated measures analyses of variance (ANOVAs), or using \( t \) tests for simple comparisons between the whole-face condition and
Results

When the two halves formed a fully integrated face (whole-face condition), there were clear responses at the two fundamental stimulation frequencies $f_1$ (5.88 Hz) and $f_2$ (7.14 Hz) and at their harmonics (Figure 2), mainly at occipital and occipito-temporal sites. Distinct responses were also observed at a number of frequencies corresponding exactly to the intermodulation components, namely the differences ($f_2 - f_1 = 1.26$ Hz, $2f_2 - 2f_1 = 2.52$ Hz, etc.), as well as the sums ($f_1 + f_2 = 13.02$ Hz, etc.) of the frequencies of interest (Figure 2; Supplementary Figure 1).

Response to parts

Fundamental frequency responses were strongly lateralized according to the half of the face that was frequency-modulated (e.g., 5.88 Hz on the left/right side yielded a large 5.88 Hz right/left posterior response; Figures 2 and 3; Supplementary Figures 1, 2, and 3). The contralateral dominant response was evident for both the left (Figure 3) and the right side of the face, for all conditions, giving a significant two-way interaction between hemisphere and side of stimulation ($F(1, 14) = 19.87, p = 0.00054$). Overall, the signal-to-noise ratio (SNR, see experimental procedures and also Rossion & Boremanse, 2011) was larger at 5.88 Hz (mean SNR = 5.29, 95% confidence interval (CI): 0.46, 10.11) than 7.14 Hz (mean SNR = 4.37, CI [1.63, 7.12]; $F(1, 14) = 6.78, p = 0.021$). These effects were qualified by a three-way interaction between frequency, hemisphere, and side of stimulation ($F(1, 14) = 14.64, p = 0.002$) reflecting an advantage of the 5.88 Hz stimulation frequency over the 7.14 Hz frequency only for contralateral responses (left hemisphere: mean SNR at 5.88 Hz = 5.91 [1.93, 9.9], mean SNR at 7.14 Hz = 4.61 [2.23, 6.99]; $F(1, 14) = 5.69, p = 0.032$; right hemisphere: mean SNR at 5.88 Hz = 7.46 [3.79, 11.13], mean SNR at 7.14 Hz = 5.89 [3.28, 8.51], $F(1, 14) = 10.43, p = 0.006$; NS for ipsilateral: left hemisphere: $F(1, 14) = 3.14, p = 0.098$; right hemisphere: $F(1, 14) = 0.07, p = 0.791$). Importantly, there were no main effects or interactions involving the factor condition (main effect: $F(2, 28) = 0.32, p = 0.725$; all other $p > 0.1$) for the fundamental response components: Relative to the whole-face condition their amplitude did not differ when the two face halves were separated by a physical gap (gap condition: $F(1, 14) = 0.004, p = 0.951$) or were laterally misaligned (misaligned condition: $F(1, 14) = 0.70, p = 0.415$) (Figure 3).

The same effects were observed for the second harmonic (Figure 2; Supplementary Figures 1, 2, and 3), although responses were larger at 14.28 Hz (2 × 7.14 Hz, mean SNR = 6.09 [0.39, 11.78]) than 11.74 Hz (2 × 5.87 Hz, mean SNR = 4.08 [0.37, 7.79]; $F(1, 14) = 15.70$,

---

**Figure 2.** Grand-averaged amplitude spectrum (0-15 Hz) at a right occipito-temporal channel (PO8), for the whole-face condition (right face stimulation, left visual field, at 5.88 Hz, left face 7.14 Hz). The responses at fundamental frequencies (5.88 Hz, 7.14 Hz) and the second harmonics (11.76 Hz, 14.28 Hz) are contralateralized. A number of intermodulation components (e.g., 7.14 – 5.88 = 1.26 Hz; 7.14 + 5.88 = 13.02 Hz) are also present. The responses are comparable when the pattern of stimulation is reversed (left visual field at 7.14 Hz and right visual field at 5.88 Hz; Supplementary Figure 1).
There were no differences between conditions, either for the left visual field stimulation at right lateralized occipito-temporal channels ($F(2, 28) = 0.16, p = 0.853$) or for the right visual field stimulation at left lateralized occipito-temporal channels ($F(2, 28) = 1.94, p = 0.162$; all other $ps > 0.1$). While stimulation of the left visual field was associated with highly lateralized responses, the channels showing the largest responses for the right visual field stimulation were located more medially. An additional analysis on medial occipital channels showed only a main effect of frequency, responses again larger at 14.28 Hz than at 11.76 Hz ($F(1, 14) = 27.93, p = 0.0001$) and no other effects (all $ps > 0.09$). Responses at the third harmonics were much smaller and were no longer lateralized. Again, no main effects or interactions involving the factor condition were found for any of the harmonics (all $ps > 0.1$).

In summary, there were large EEG responses at the fundamental stimulation frequencies and second har-
monics corresponding to the side of the stimulated face for all three conditions. These responses were strongly contralateralized. Importantly, if anything, left visual stimulation yielded fundamental and second harmonic frequency responses of slightly larger amplitude for the gap and misaligned conditions than for the whole face condition (Figure 3C), although the differences were not significant.

**Responses to whole: IM components**

**Difference IM components**

On grand-averaged data, all four IM components at frequencies equal to low-order differences between $f_1$ and $f_2$ ($f_2 - f_1$: 1.26 Hz; $2f_2 - 2f_1$: 2.52 Hz; $3f_2 - 3f_1$: 3.78 Hz; $4f_2 - 4f_1$: 5.04 Hz) were prominent for the whole face condition only (Figure 4). The average SNR of these components shows that the low-term difference IM responses were mainly localized in two posterior regions: the right occipito-temporal cortex (centered on PO8) and, to a lesser extent, the left homologous region (Figure 5). The largest and most consistent differences between the whole-face condition and the two control conditions were also found on these same electrode sites (Figure 5).

On the same regions of interests (ROIs) used to define the part-based responses, there was a much larger response in the whole-face condition as compared to the other two conditions (mean SNR whole face = 2.03 [1.46, 2.60]; gap = 1.22 [1.06, 1.38]; misaligned = 1.31 [1.09, 1.53], $F(2,28) = 16.14$, $p = 0.00002$; left hemisphere: mean SNR whole face = 1.75 [1.33, 2.17]; gap = 1.18 [1.04, 1.32]; misaligned = 1.29 [1.11, 1.48], $F(2,28) = 7.07$, $p = 0.003$; and right hemisphere: mean SNR whole face = 2.30 [1.82, 2.79]; gap = 1.27 [1.1, 1.44]; misaligned = 1.33 [1.16, 1.5], $F(2,28) = 18.24$, $p = 0.0001$). The gap and misaligned conditions did not differ from each other in any of the two hemispheres ($F(1.14) = 1.02$, $p = 0.330$). There was also a significant effect of the factor hemisphere (left hemisphere: 1.41 [1.08, 1.74]; right hemisphere: 1.64 [1.29, 1.99], $F(1, 14) = 6.45$, $p = 0.024$) and a significant interaction between condition and hemisphere ($F(1, 14) = 5.38$, $p = 0.011$) because only the whole-face condition showed a significant difference in favor of the right hemisphere ($F(1, 14) = 8.38$, $p = 0.00177$; NS for the two control conditions, gap: $F(1,14) = 0.84$, $p = 0.375$; and misaligned: $F(1,14) = 0.20$, $p = 0.660$). In addition to these low-term IM components, which reflect an equal contribution of both face parts, there were also smaller responses at other difference components (e.g., $2f_2 - f_1$: 8.40 Hz, Figure 2; and $2f_1 - f_2$: 4.62 Hz), showing a more occipital medial distribution but also larger responses to whole faces than the other two conditions over the right occipito-temporal region (Supplementary Figure 4).

At the individual level, when considering the average of all four low-term difference components, only two participants failed to present IM significant responses for the whole-face condition over the right occipito-temporal cortex (Z-score range of 13 participants: 3.12–12.96; $p < 0.001$) (Figure 5). For the gap and misaligned condition however, only four (Z-score range: 2.04–5.98) and five (Z-score range: 2.45–5.00) participants presented significant responses, respectively. For the 13 participants who showed significant IM components in the whole-face condition, SNR ranged between 1.64 (64% signal increase) and 3.96. The two participants who did not present with IM component responses above noise in the whole-face condition topographical maps are depicted at the right occipito-temporal site where these responses were the most significantly above noise for the average of the four IM low-term difference components (PO8). There were very few if any intermodulation responses in the gap and misaligned conditions.
condition nevertheless presented with significant responses to parts, which were in the low normal range. For the whole face condition, the Pearson correlation across individuals between the sum of the SNR values for the fundamental responses to parts (left + right side of the face) and the first-order difference IM component (1.26 Hz) was significant (electrode PO8; $r = 0.69$, $r^2 = 0.48$, $p = 0.004$; for the mean of the four IM low-term difference components: $r = 0.68$, $r^2 = 0.46$, $p = 0.006$; Supplementary Figure 5). This result indicates that the magnitude of the response to parts across participants is only weakly related to the magnitude of the nonlinear response to the whole face.

**Sum IM components**

Sum IM components (i.e., $f1 + f2 = 13.02$ Hz) were observed almost exclusively in the whole-face and the misaligned conditions, and these components were found on occipital medial channels (Supplementary Figure 6). There was a significant effect of condition (whole face: 4.67 [3.26, 6.07]; gap: 2.55 [2.06, 3.03]; misaligned: 5.07 [3.78, 6.36], $F(2, 28) = 13.38$, $p = 0.00008$) on a occipital medial ROI, reflecting the reduced response in the gap condition. The whole-face and misaligned conditions did not differ from each other on any of the ROIs (central ROI: $F(1, 14) = 0.83$, $p = 0.376$; left ROI: $F(1, 14) = 0.082$, $p = 0.778$; right ROI: $F(1, 14) = 1.57$, $p = 0.231$).

**Replication and inverted faces**

To strengthen these observations, the whole experiment was replicated in a new group of participants ($N = 11$, not tested in the first experiment) and included a condition with whole faces presented upside down (the fixation cross...
remained in between the eyes of the face, Figure 6). The rationale for including inverted faces is that the visual stimulus remains exactly identical, although there is behavioral evidence that perceptual integration of facial parts is disrupted for inverted faces (Sergent, 1984; Young et al., 1987; Tanaka & Farah, 1993; Biederman & Kalocsai, 1997; for a review see Rossion, 2008). All IM difference components (n/2 – n/2) were observed primarily at right occipito-temporal sites in the whole-face condition and were almost absent in the gap and misaligned conditions. There was a main effect of condition (whole face: 1.81 [1.33, 2.29]; gap: 1.22 [1.07, 1.38]; misaligned: 1.22 [1.11, 1.34]; inverted: 1.62 [1.38, 1.86], F(3, 30) = 15.24, p < 0.000001; whole face > all other conditions together: t(1, 10) = 13.21, p = 0.002). This effect was qualified by a significant interaction between condition and hemisphere, (F(1, 10) = 2.69, p = 0.064) due to a larger IM response in the whole-face condition than in the gap and misaligned conditions in both hemispheres (versus gap: t(1, 10) = 1.82, p = 0.049; versus misaligned: t(1, 10) = 2.52, p = 0.015, for the left hemisphere; versus gap: t(1, 10) = 5.61, p = 0.0001; versus misaligned: t(1, 10) = 4.09, p = 0.001, for the right hemisphere), while a significant difference with inverted faces was observed only in the right hemisphere (t(1, 10) = 2.27, p = 0.020; left hemisphere: t(1, 10) = −0.28, p = 0.394; Figure 6).

Thus, inversion was not sufficient to abolish perceptual integration between the two face halves in the left hemisphere. IM components were larger in the inverted condition as compared to the other control conditions, versus gap: F(1, 10) = 24.28, p = 0.0003; versus misaligned: F(1, 10) = 27.89, p = 0.0002. As in the main experiment, the response to parts did not differ between conditions at lateral ROIs (F(3, 30) = 1.37, p = 0.272).

At the individual level, only two participants failed to present IM significant responses for the whole-face condition over the right occipito-temporal cortex (Z-score range of nine participants: 2.52–5.68; p < 0.01). In the inverted condition, eight participants presented significant responses (Z-score range: 2.23–4.70; p < 0.01) but all of them showed larger responses in the upright condition (Figure 6).

**Discussion**

By taking advantage of the frequency-tagging approach in EEG, we measured the response to each simultaneously presented half of a complex visual stimulus. Our most important observation was the specificity of the intermodulation responses to the intact whole face stimulus. These IM responses can only be generated by populations of neurons that not only receive inputs from both face halves but also combine them in a nonlinear fashion. The presence of IM components in the EEG when the whole face was presented indicates unambiguously that the response to the whole face is different than the sum of the responses to its parts.

**Intermodulation responses as a signature of holistic face representation**

As mentioned in the Introduction, nonlinear interactions reflected in IM response components have been observed in response to simple visual stimuli, such as collinear bars, the border between a figure and its background, or the same checkerboard flickering in size and luminance (e.g., Zemon & Ratliff, 1984; Norcia et al., 1999; Victor & Conte, 2000; Appelbaum et al., 2008; Giani et al., 2012). Two recent studies have also reported an increase in intermodulation components when simple visual stimuli form an illusory object (Aissani, Cottereau, Dumas, Paradis, & Lorenceau, 2011; Gündlach & Müller, in press). These previous studies offer a basis for the interpretation of our observations, which are nevertheless original at several levels.

First and foremost, unlike these previous studies, which used low-level stimuli only or in which a shape emerges from the interaction between low-level visual stimuli, here each half of a face is a complex stimulus in itself. Since each face half is sufficient to recognize the stimulus as a face, and to possibly activate a whole face representation, demonstrating unambiguously that the whole is different than the sum of its parts is particularly challenging with such stimuli. This makes the present paradigm and observations highly valuable for understanding how facial parts are integrated into a meaningful whole.

Second, the studies that reported an increase in IMs when simple visual stimuli form an illusory object reported differences between conditions for sum components only (Aissani et al., 2011: f1+f2; Gündlach & Müller, in press: f1+f2 and 2f1+2f2). These sum components were either found at unexpected electrode sites (right centro-frontal) or shared the same topography as the flickering frequencies on medial occipital sites (Aissani et al., 2011; Gündlach & Müller, in press, respectively). Here the sum component (13.02 Hz) decreased substantially only in the gap condition when there was a physical separation between the two parts and no contiguous border. This observation, together with the medial occipital topography of the low-term sum component at 13.02 Hz, suggests that it reflects local spatial interactions in early retinotopic areas that are more
or less preserved by the misalignment manipulation. This may occur because there are still adjacent contrast regions across the border between the face parts that could interact in a fashion that does not depend on strict local or global alignment. By contrast, the multiple high SNR difference components (1.26 Hz, 2.52 Hz, etc.) observed here were the most interesting for our purpose. These components depend on the global alignment of face parts in that the simple local juxtaposition of the parts is not sufficient to generate them. Importantly, all of these relevant IM components are associated with a consistent scalp topography, pointing to cortical sources in high-level visual areas of the right occipito-temporal cortex.

A possible interpretation of the dissociation between the difference and sum components comes from a set of prior two-input ssVEP experiments, which were originally interpreted within the context of a cascade of linear and nonlinear stages of processing (Tsai, Wade, & Norcia, 2012). Within this model, the two mechanisms in the present study reflected by the sum and the difference components may be the result of parallel processes with different filtering properties after their separate nonlinearities. The process generating the sum terms may be temporally band pass or high pass and may reflect more local spatial interactions in early visual cortex based on its scalp distribution. The process generating the difference terms appears to involve considerable temporal integration as it only passes low-

Figure 6. Signal-to-noise ratio (SNR) averaged for the four intermodulation components for upright (whole) and inverted faces in the second study (11 participants). The histograms show the group mean ($\pm SE$) and the SNR values for every individual tested of the eight right occipito-temporal channels selected in the right hemispheric region of interest. (B). Topographical maps (back of the head) showing the right occipito-temporal focus of the intermodulation response for upright and inverted faces, and the difference between the two.
frequency interaction terms at its output, which would be consistent with a part integration process in (right) occipito-temporal cortex where this component has large amplitude.

**Binding by convergence**

The IM responses constitute a signature of an integrated representation, indicating that populations of neurons see the two parts of the object. Yet, these two parts oscillate at different low frequencies in our paradigm. Thus, these observations cannot provide support to the binding-by-synchrony hypothesis, according to which different populations of neurons code for different parts of an object by oscillating in phase, at the same high-frequency rate ("gamma band oscillations," Singer & Gray, 1995; Tallon-Baudry & Bertrand, 1999). However, they may support a binding by convergence view, whereby a representation of the whole emerges in higher order areas from combining the processing of the (two) parts in lower level areas (Barlow, 1972). An interesting issue is whether the whole is processed after the parts have been fully analyzed or if a whole representation can be constructed early on from a very coarse definition of its parts, and then refined through top-down processing (Mumford, 1992; Hochstein & Ahissar, 2002). Our study cannot address this issue because the phases of the IM components and of the fundamental responses cannot be directly compared. Thus, our data cannot be used to make direct inferences about the spatio-temporal hierarchy between the representation of the wholes and the parts.

**Right hemispheric dominance of part-based responses**

Despite using relatively small centrally presented stimuli, the response at the fundamental frequency to each face half was strongly contralateralized, in line with previous observations of dominant theta/alpha (4–12 Hz) responses on the left and right occipito-temporal electrodes processing the contralateral eye of a face (Smith, Gosselin, & Schyns, 2007). Here this contralateralization was especially true for the face half presented in the left visual field, which led to a strongly right lateralized occipito-temporal response. This finding is in agreement with the general dominance of the right hemisphere in face perception (Hecaen & Angelergues, 1962; Sergent, Ohta, & MacDonald, 1992; Bentin, Allison, Puce, Perez, & McCarthy, 1996). The increased right lateralization also fits with the long standing evidence from chimeric faces that the right half of the face looks more like the whole face than the left half (Gilbert & Bakan, 1973), except in patients with right temporal lobe damage (Kolb, Milner, & Taylor, 1983).

**Whole-part relationships in face perception**

Our approach and findings offer a fresh look to the long-standing issue of part/whole relationships in face perception. Numerous behavioral studies have shown that the processing of a facial part is influenced by the other parts, and by how these parts are organized in a whole configuration (e.g., Sergent, 1984; Young et al., 1987; Tanaka & Farah, 1993; Rossion, 2013). Evidence for the holistic representation of faces at the neural level is also typically based on the interpretation of how the presence of an intact whole face configuration modifies the response to a part (Schiltz & Rossion, 2006; Jacques & Rossion, 2009). These observations have generally been taken as evidence for the interactivity of processing between parts and for a whole face being different than the sum of its parts. However, as outlined in the Introduction, the evidence remains indirect and has been questioned by studies that attempt to quantify the behavioral or neural response to a part independently of the response to a whole face and compare these responses offline (Freiwald et al., 2009; Arcurio et al., 2012; Gold et al., 2012). In these conditions, a “nonlinearity” could be due to a part being sufficient to evoke another part or an entire face representation. In contrast, here the evidence for holistic face processing is direct in the sense that a nonlinear response can only be due to an integration of the two parts. This response is identified objectively in the EEG spectrum because it is confined to small frequency bins corresponding exactly to the difference between the part-related frequencies that are known by the experimenter (e.g., 7.14 Hz – 5.88 Hz = 1.26 Hz). This whole-based response can also be objectively separated in the EEG spectrum from the response to the parts and directly quantified in terms of signal-to-noise ratio.

Importantly, the substantial reduction of the intermodulation difference response in the gap and misaligned control conditions, coupled with the lack of decrease of the part response, provides unambiguous evidence that a perceptually integrated whole face stimulus is different than the sum of its parts in the human brain. This observation prompts two further remarks. First, the absence of intermodulation responses, that is, of interaction between face halves when they were separated by a physical gap, is in line with single-cell recordings of IT neurons showing that nonlinearities are more difficult to observe when the parts of an object are physically separated (Sripati &
Specificity to faces

Our decision to use faces as the image category in this frequency-tagging approach applied to high-level vision was motivated by several factors. Faces are visually homogenous and almost symmetrical, making them ideal to study the interaction between their left and right halves. They are highly salient and familiar stimuli of our visual environment, and their perception is associated with large and at least partly specific neural responses (e.g., Desimone et al., 1984; Sergent et al., 1992; Allison, Puce, Spencer, & McCarthy, 1999; Haxby, Hoffman, & Gobbini, 2000; Tsao et al., 2006; Weiner & Grill-Spector, 2013 for a review). As explained in the Introduction, because their parts are meaningful and can activate other parts through completion, faces also pose particularly difficult challenges to the study of part/whole relationships. Given this, one may wonder whether our approach and findings are specific to faces or could be extended to the perceptual integration of other classes of natural images. On the one hand, there are no reasons why the approach could not be extended to study perceptual integration of other classes of images. However, since we did not test other object stimuli, we cannot and do not claim that our findings are specific to faces. On the other hand, there are at least two aspects of our data that point to face-specific mechanisms. First, our IM responses were strongly right lateralized and observed specifically over the same electrode sites where face-sensitive electrophysiological responses are recorded (the “N170” component, Bentin et al., 1996; Rossion & Jacques, 2011 for a review). This right occipitotemporal activation has been specifically associated with holistic perception of faces (e.g., Sergent, 1988; Hillger & Koenig, 1991; Rossion, Dricot, Goebel, & Busigny, 2011; Caharel, Leleu, Bernard, Lolande, & Rebai, in press). Second, if the stimulus was merely a 2-D circle, introducing a gap between its left and right halves might not reduce IM responses significantly, because the visual system could “fill the gap,” thanks to contour closure (Wertheimer, 1923; Wagemans et al., 2012). However, the introduction of a gap between face halves in our study does only break a face in two halves; it also changes substantially the width/height ratio of the face, so that it may not fit a holistic face template anymore. Moreover, the introduction of a vertical gap breaks the face-specific local configuration of internal diagnostic parts (nose, mouth, combination of the two eyes and eyebrows, etc.), possibly contributing to the reduction of the IM responses in this condition. Finally, inversion of the whole face reduced significantly the IM components while leaving the part-based responses intact, an observation offering direct support to the view that inversion reduces holistic/configural face perception (Sergent, 1984; Young et al., 1987; Rhodes, Brake, & Atkinson, 1993; Tanaka & Farah, 1993; Rossion, 2008; Sekunova & Barton, 2008; Van Belle, de Graef, Verfaillie, Rossion, & Lefevre, 2010; Rossion, 2013). Given that the decrease of behavioral performance for inverted relative to upright stimuli is a well-known marker of face-specific processes (Yin, 1969), future studies may also find that this electrophysiological effect is specific to faces.

Keywords: face perception, Gestalt, ssVEP, holistic perception, inversion, intermodulation, nonlinearity

Acknowledgments

This research was supported by an ERC grant (faccozep 284025) to BR. AB and BR are supported by the Belgian National Fund for Scientific Research (FNRS). The authors thank Jeremy Balder, Sylvie Nozaradan, and three anonymous reviewers for their helpful comments on a previous version of this paper.

Commercial relationships: none.
Corresponding author: Bruno Rossion.
Email: bruno.rossion@uclouvain.be.
Address: Institute of Research in Psychology & Institute of Neuroscience, Université catholique de Louvain, Louvain la Neuve, Belgium.
References


Gold, J. M., Mundy, P. J., & Tjan, B. S. (2012). The perception of a face is no more than the sum of its parts. *Psychological Science, 23,* 427–434.


consisting entirely of face-selective cells. Science, 311, 670–674.


