VISION
RESEARCH

# Dissociable effects of inter-stimulus interval and presentation duration on rapid face categorization

Talia L. Retter[a,b,*], Fang Jiang[b], Michael A. Webster[b], Bruno Rossion[a,c]

[a] Psychological Sciences Research Institute, Institute of Neuroscience, University of Louvain, Belgium
[b] Department of Psychology and Graduate Program in Integrative Neuroscience, University of Nevada, Reno, USA
[c] Neurology Unit, Centre Hospitalier Regional Universitaire (CHRU) de Nancy, F-54000 Nancy, France

## ARTICLE INFO

## ABSTRACT

Fast periodic visual stimulation combined with electroencephalography (FPVS-EEG) has unique sensitivity and objectivity in measuring rapid visual categorization processes. It constrains image processing time by presenting stimuli rapidly through brief stimulus presentation durations and short inter-stimulus intervals. However, the selective impact of these temporal parameters on visual categorization is largely unknown. Here, we presented natural images of objects at a rate of 10 or 20 per second (10 or 20 Hz), with faces appearing once per second (1 Hz), leading to two distinct frequency-tagged EEG responses. Twelve observers were tested with three squarewave image presentation conditions: 1) with an ISI, a traditional 50% duty cycle at 10 Hz (50-ms stimulus duration separated by a 50-ms ISI); 2) removing the ISI and matching the rate, a 100% duty cycle at 10 Hz (100-ms duration with 0-ms ISI); 3) removing the ISI and matching the stimulus presentation duration, a 100% duty cycle at 20 Hz (50-ms duration with 0-ms ISI). The face categorization response was significantly decreased in the 20 Hz 100% condition. The conditions at 10 Hz showed similar face-categorization responses, peaking maximally over the right occipito-temporal (ROT) cortex. However, the onset of the 10 Hz 100% response was delayed by about 20 ms over the ROT region relative to the 10 Hz 50% condition, likely due to immediate forward-masking by preceding images. Taken together, these results help to interpret how the FPVS-EEG paradigm sets temporal constraints on visual image categorization.

## 1. Introduction

Fast periodic visual stimulation with electroencephalography (FPVS-EEG) has advantages in sensitivity and objectivity for measuring rapid visual categorization processes through the insertion of within-category (face) images as a proportion of rapidly presented across-category (object) images (e.g., Rossion, Torfs, Jacques, & Liu-Shuang, 2015; Jacques, Retter, & Rossion, 2016; Jonas et al., 2016; Retter & Rossion, 2016). In the example illustrated in Fig. 1A, natural images are presented at a fixed periodic rate, i.e., 10 images per second with faces appearing as every 1 out of 10 images, leading to frequency-tagged EEG responses (also known as "steady-state visual evoked potentials", SSVEPs; for a general review, see Norcia, Appelbaum, Ales, Cottereau, & Rossion, 2015) at 10 Hz for image presentation and 1 Hz for face-selective responses. By using highly variable natural face images, as well as a wide variety of natural object images, the paradigm is able to capture both generalization across within-category face exemplars and discrimination of face-selective vs. generic visual responses (Rossion et al., 2015).

FPVS-EEG is not only defined by the periodic presentation of visual stimuli: it is also defined by its relatively *fast* rate of stimulation (i.e., compared to standard EEG or behavioral studies). A relatively fast rate, e.g., 4–12 Hz, is valuable because it constrains neural responses to limited time windows. This borrows from the logic of rapid serial visual presentation (RSVP), in which stimuli are serially presented, sometimes in the range of eye fixation rates or faster, with participants subsequently providing responses to the images they perceived or remembered (Potter, 2012; Potter & Levy, 1969; Potter, Wyble, Hagmann, & McCourt, 2014). The brief image duration, as well as perceptual backward-masking from sequential stimuli (typically shown without any inter-stimulus interval) used in RSVP temporally limits the availability of visual information about a stimulus category (Potter, 2012). Consequently, RSVP has been used to probe the contributions of attention and memory to visual perception at a behavioral level. RSVP has also been applied to single neuron recordings to investigate the effects of forward and backward masking, as well as the temporal limits of image processing in the brain (Keysers & Perrett, 2002; Keysers, Xiao, Foldiak, & Perrett, 2005; Keysers, Xiao, Földiák, & Perrett, 2001).

* Corresponding author at: 1664 N Virginia St., Mail Stop 296, Reno, Nevada 89557, USA.
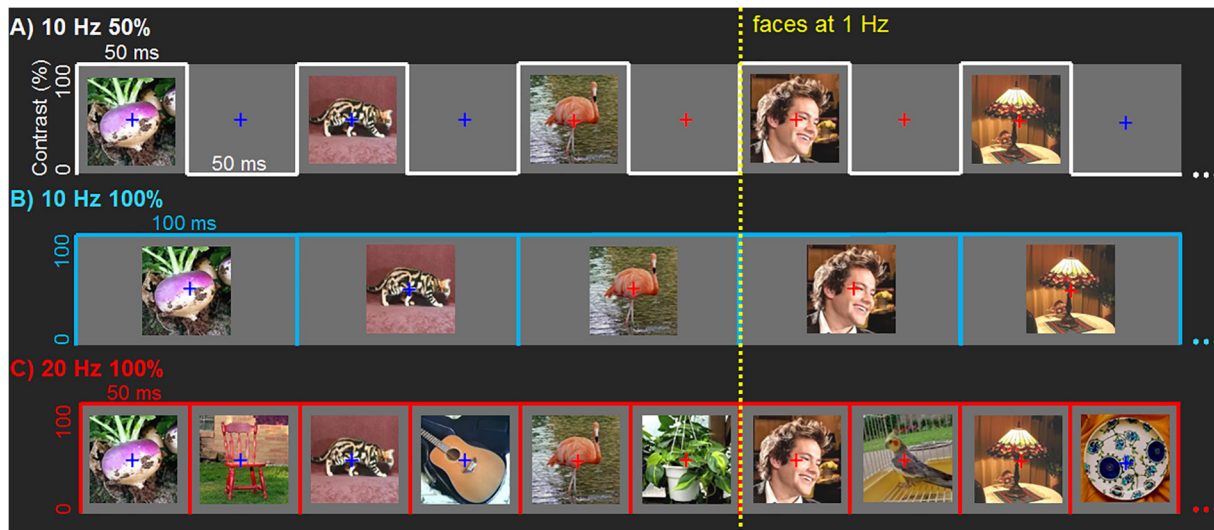E-mail address: tlretter@nevada.unr.edu (T.L. Retter).

**Fig. 1.** Experimental paradigm and conditions. In all stimulus presentation modes, images of natural objects are presented in a random order without repetition, and face stimuli appear every 1 s, i.e., at a rate of 1 Hz. An excerpt of only 500 ms is depicted here for each testing sequence, including an example of the brief fixation cross color change, providing cues for an orthogonal behavioral task. A) 10 Hz 50%: Stimuli are presented at 10 Hz with a 50% duty cycle, i.e., "on" for 50 ms at full luminance contrast and "off" for an ISI of 50 ms at 0% contrast. B) 10 Hz 100%: Stimuli are presented at 10 Hz again, but with a 100% duty cycle, staying on for 100 ms with no ISI. C) 20 Hz 100%: Stimuli are presented at 20 Hz with a 100% duty cycle, appearing on for 50 ms with no ISI, matching the stimulus presentation duration of the first condition but the 0-ms ISI of the second condition.

Building on these techniques, FPVS-EEG presents longer streams of visual images with embedded frequency-tagged categories, allowing for a category-specific neural response to be characterized and quantified (Retter & Rossion, 2016).

The rate of image presentation can be controlled with two factors: 1) the duration of stimulus presentation and 2) the inter-stimulus interval (ISI). For different experimental purposes, stimuli may be presented with or without an ISI, e.g., with an ISI to prevent image after-effects or without an ISI to maximize forward and backward masking, so as to limit visual information. Indeed, removing the ISI may have implications on the degree of backward and forward masking of each stimulus (Crawford, 1947; Macknik & Livingstone, 1998). With relatively short ISIs (below about 90 or 100 ms), visual short term memory may serve to "fill-in" the gaps between stimuli (Potter, 2012), leading to no differences reported in human image detection performance or in monkey superior-temporal sulcus neuron firing rates (Keysers & Perrett, 2002; see also Rolls & Tovée, 1994). Subtle differences may be observed, however, in the onset of neural responses to stimuli presented with or without an ISI. Specifically, there was a small, uninvestigated delay of responses without an ISI in a previous study (see Fig. 3A of Keysers & Perrett, 2002; also in Fig. 3 of Keysers et al., 2005). This might be due to transient off and/or sustained responses produced from forward masking (i.e., from previously presented images) overlapping with and suppressing subsequent image responses (e.g., Breitmeyer, Levi, & Harwerth, 1981; Duysens, Orban, Cremieux, & Maes, 1985; Keysers et al., 2005; Macknik & Livingstone, 1998; Ogmen, Breitmeyer, & Melvin, 2003).

In this study we asked how the neural responses captured by the FPVS face categorization paradigm depend on the temporal properties of the stimulus sequence. To separate the effect of ISI from stimulus duration on rapid face categorization, we compared three squarewave[1]

stimulation modes. The first of these was chosen to closely replicate parameters used in previous studies (e.g., Dzhelyova et al., 2016; Retter & Rossion, 2016), in which stimuli were presented at 10 times a second (10 Hz) with a 50% duty cycle (each image shown for 50 ms and followed by a 50 ms ISI) (Fig. 1A). In the second condition, we removed the ISI but used the same presentation rate. That is, we presented stimuli at 10 Hz but with a 100% duty cycle, presenting images for 100 ms with no ISI (Fig. 1B). In the third condition, we also removed the ISI but used the original presentation duration, so that the images were shown at 20 Hz with a 100% duty cycle, displaying images for 50 ms with no ISI (Fig. 1C). In all conditions, faces appeared at 1 Hz, i.e., with a face stimulus shown every 1 s, and we asked how the 1 Hz response to faces (reflecting face categorization) depended on the temporal properties of the image stream.

Specifically, by comparing the first two conditions, the effect of the ISI (50 vs. 0 ms) may be investigated while controlling the presentation rate at 10 Hz. By comparing the first and third conditions, the effect of ISI (again 50 vs. 0 ms) may be investigated while controlling the image presentation duration at 50 ms. Finally, by comparing the second and third conditions, the effect of stimulus presentation duration (100 vs. 50 ms) may be investigated when the ISI is not present. Practically, it is important to understand the effects of different presentation modes when designing experiments. For example, an ISI may be desired for limiting after-effects (e.g., color or face-related) or apparent motion across sequential stimulus presentations; however, an ISI may be undesired for testing the temporal constraints that forward and backward masking impose on (e.g., the speed of) perception. Thus, one goal was to provide information about the appropriateness of different presentation modes for future studies, as well as to provide baseline data for comparing across past and future studies using different image presentation modes. Theoretically, we also sought to compare the impacts of an ISI on human population-level neural responses to the effects of ISI previously observed in single unit responses.

---

[1] Note that a squarewave presentation was used here in order to identify exactly the presentation duration, whereas the majority of previous FPVS studies applied a sinusoidal modulation of image contrast (e.g., as in Dzhelyova, Jacques, & Rossion, 2016; Jacques, Retter, & Rossion, 2016; Regan, 1966; Retter & Rossion, 2016; Rossion et al., 2015; Silberstein, Schier, Pipingas, et al., 1990; van der Tweel & Verduyn Lunel, 1965; Victor & Zemon, 1985). In sinusoidal presentation, the image display appears smooth and relatively continuous but the contrast level at which each image is identifiable, defining the effective presentation duration, is not always known.

## 2. Materials and Methods

### 2.1. Participants

Participants included 12 students or employees at the University of

Nevada, Reno (20–33 years old, 4 male, 11 right handed). All reported normal or corrected-to-normal visual acuity. Participation was with informed consent and all procedures followed protocols approved by the university's Institutional Review Board and were conducted in accordance with the Code of Ethics of the World Medical Association (Declaration of Helsinki).

### 2.2. Stimuli

Stimuli were 294 natural images (46 faces and 248 objects) taken from an image set used in previous studies, e.g., by Rossion and colleagues (2015) and Retter and Rossion (2016). Images containing faces included different ages, races, sex, ethnicity, etc., and stimuli of objects included diverse categories such as plants, animals, lamps, chairs, fruits, and houses. Across face and object categories, images varied widely in viewpoint, background, lighting conditions, etc. For standardization of experimental testing, they were cropped to the same size (a square of 200 by 200 pixels) and equated for mean luminance. The stimuli are available online at: http://face-categorization-lab.webnode.com/resources, and representative examples are shown in Fig. 1. Viewed from 80 cm on an 800 by 600 pixel resolution monitor, the images subtended approximately 6.5 degrees in visual angle.

### 2.3. Procedure

Participants were fitted with an EEG headcap and then viewed stimuli presented on a cathode ray tube (CRT) monitor with a refresh rate of 120 Hz. The room was otherwise darkened. Observers were presented with three image presentation conditions as described in the Introduction and Fig. 1: 1) 10 Hz 50%: images are presented for 50 ms, separated by a 50-ms ISI; 2) 10 Hz 100%: images are presented for 100 ms with a 0 ms ISI; 3) 20 Hz 100%: images are presented for 50 ms with a 0 ms ISI. Thus, during the testing sequence, 10 or 20 images were presented per second (10 or 20 Hz), with a consistent periodicity of 9 or 19 non-face objects followed by one face, which therefore appeared at a rate of 1 Hz. As mentioned in the Introduction, images were presented with a squarewave stimulus presentation mode, i.e., appearing at either full or zero contrast, in order to precisely control the temporal parameters. This differs from most earlier FPVS studies, in which images were presented with a sinusoidal modulation of luminance contrast (e.g., Alonso-Prieto, Van Belle, Liu-Shuang, Norcia, & Rossion, 2013; Gentile & Rossion, 2014; Rossion et al., 2015; Jacques et al., 2016; Jonas et al., 2016). Stimuli were presented with a script running over Java SE Version 8.

The experimental trials were composed of the following segments: 1) 2–5 s of a central fixation cross on a gray background (corresponding to a gray level of 112 on the 8-bit or 255-level range of the monitor); 2) 3 s of gradually increasing stimulus contrast ("fade-in"); 3) the 90 s testing sequence; 4) 3 s of gradually decreasing stimulus contrast ("fade-out"); 5) 2 s of the fixation cross on the background. To encourage constant attention throughout the test sequences, participants were given the task to respond by pressing on the keyboard's space bar to randomly timed color changes of the fixation cross (blue to red for 250 ms), which occurred 12 times within each sequence. Participants viewed two repetitions of 90 s for each of the three experimental conditions; the order of trials was fully randomized for each participant. This led to a total recording time of about 10 min. Since testing time was short, participants also took part in other experiments during the same recording session.

### 2.4. EEG acquisition

EEG signals were acquired with a BioSemi ActiveTwo EEG system containing 128 Ag-AgCl Active-electrodes. The default BioSemi head cap electrode configuration was used, centered around nine standard 10/20 locations on the primary axes (BioSemi B.V., Amsterdam, Netherlands; for exact position coordinates, see http://www.biosemi.com/headcap.htm). Default BioSemi electrode labels (e.g., A1, A2, etc.) were changed to closely match the more conventional 10/5 system (Oostenveld & Praamstra, 2001; for exact relabeling, see Rossion et al., 2015, Fig. S2). Offsets were kept below 40 mV, referenced through the common mode sense (CMS) and driven right leg (DRL) electrodes situated in the head cap. Vertical and horizontal electrooculograms (EOG) were also recorded with four additional flat-type Active-electrodes: two electrodes above and below the participant's right eye and two lateral to the external canthi. The EEG and EOG were digitized at a sampling rate of 2048 Hz.

### 2.5. Analysis

Analyses were performed with Letswave 5, an open source toolbox (http://nocions.webnode.com/letswave), running over MATLAB R2013b (MathWorks, USA),

#### 2.5.1. Preprocessing

Two filtering steps were applied to the continuously recorded individual participant data: 1) a Butterworth band-pass filter with cutoff values of 0.1–120 Hz (second-order, zero-phase); 2) a multi-notch frequency filter at 60 and 120 Hz (width of 0.5 Hz) to remove electrical line noise. To reduce file size, data were down-sampled to 512 Hz. Data were then segmented by condition, including fade-in and fade-out time for each testing sequence. Independent component analysis was applied on the data of only three participants who blinked more than 0.26 times/s (the mean blink rate across all subjects was 0.20 times/s, SD = 0.35) on average during the sequences; only a single component, the first in all cases, was removed. High-artifact channels, exhibiting several deflections greater than 100 μV, were linearly interpolated with several neighboring channels (less than 1.2% of channels on average). All 128 EEG channels were then re-referenced to a common average. Finally, data were precisely segmented from the time of stimulus onset to 89.03 s, an interval containing an integer number of cycles at 0.99965 Hz (the actual frequency of the nominal 1 Hz stimulus, since the monitor had a refresh rate slightly below 120 Hz).

#### 2.5.2. Frequency domain analysis of periodic responses

Sequences of each condition were averaged in the time domain, reducing activity that was not phase-locked across trials (i.e., not driven by visual stimulation). Then a fast Fourier transform (FFT) was applied to transform the data into a normalized amplitude spectrum (μV) covering 0–256 Hz in the frequency domain. The resolution of this spectrum is 0.011 Hz, i.e., 1/89.03 s. In an attempt to isolate the signal from the baseline noise level of the amplitude spectrum, a baseline-subtraction was applied at each frequency bin (e.g., as in Mouraux et al., 2011; Retter & Rossion, 2016), using a baseline of twenty neighboring frequency bins not including the immediately adjacent, maximum, and minimum frequency bins. The grand-average amplitude and baseline-subtracted amplitude spectra were computed for display at the group-level.

Responses to the face-presentation frequency and the image-presentation frequency occur at their respective frequency values (e.g., $F$) and their harmonics (e.g., $2F$, $3F$, etc.). The total response to each tagged frequency was reconstructed by summing its unique harmonics (Retter & Rossion, 2016). The frequency range over which we summed the harmonics was determined based on a previous FPVS-EEG face categorization study: up to 19 Hz for face-selective responses, and up to 40 Hz for image-presentation responses (Retter & Rossion, 2016). Significant harmonics calculated according to previously applied protocols based on Z-scores (e.g., Rossion et al., 2015; Jacques, et al., 2016; Retter & Rossion, 2016) would instead be based on 14 Hz for faces and image-presentation harmonics up to 40 Hz. However, the inclusion of these extra harmonics is not expected to change the results, since the baseline-subtracted amplitude is zero on average where a stimulus-
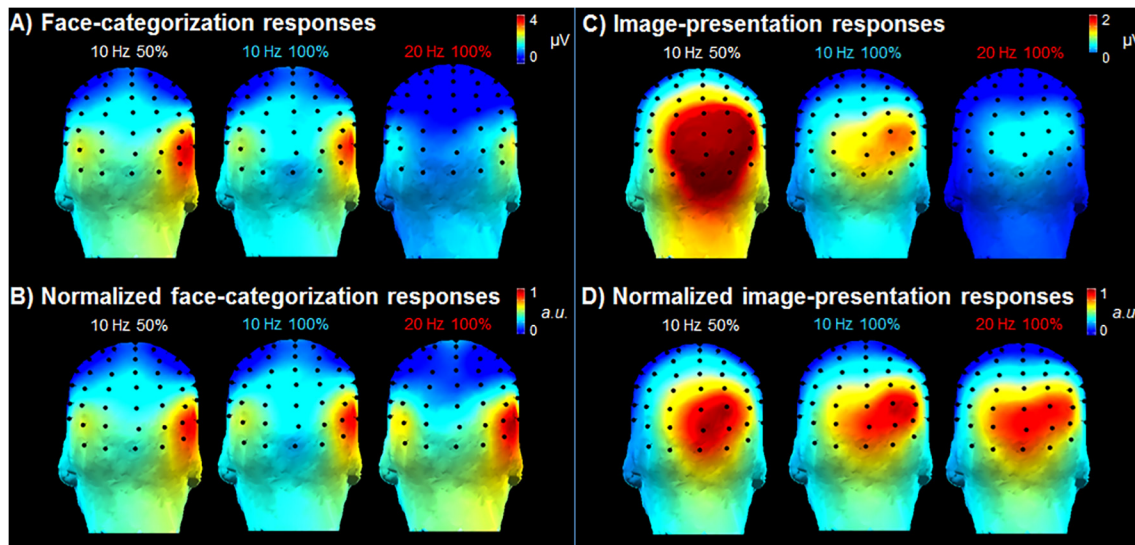
**Fig. 2.** Grand-averaged response topographies across the back of the head for each of the three experimental conditions. A) Face-categorization responses, comprising the sum of baseline-subtracted responses at the face-presentation frequency of 1 Hz and its specific harmonic responses up to 19 Hz. B) Normalized topographies of A. C) Image-presentation responses, constituting the sum of the baseline-subtracted responses to the image-presentation rate of 10 Hz (or 20 Hz in the third condition) and its harmonics up to 40 Hz. D) Normalized topographies of C. (a.u. = arbitrary unit).

driven response is not present. Topographical head maps of individual harmonic and summed-harmonic responses were plotted to display the data across all channels; additionally, the summed-harmonic topographies were normalized to remove general amplitude differences for better comparison of scalp distributions across conditions, according to the method of McCarthy and Wood (1985).

The amplitude of summed-harmonic face-selective responses was evaluated over three regions-of-interest (ROIs), defined a priori according to previous studies using this paradigm (e.g., Dzhelyova & Rossion, 2014; Retter & Rossion, 2016), as well as verified post hoc by examining the maximal channels at the group level (see Results). A right occipito-temporal (ROT) ROI averages channels PO8, PO10, PO12, P10, and P8, a left-occipito temporal (LOT) ROI was defined homologously, and a medial occipital (MO) ROI encompasses channels POOz, POz, Oz, Oiz, and Iz. Given differing distributions of image-presentation responses across the scalp for the 0 ms ISI conditions here, these responses were quantified over a novel medial, occipito-parietal (OP) ROI, defined by channels O2, POI2, OIz, Iz, O1, and POI1, identified post hoc (see Results). Statistical comparisons of conditions were executed separately for face-selective and image-presentation responses. For testing face-categorization responses, repeated measures ANOVAs, with factors of Condition (10 Hz 50%, 10 Hz 100%, 20 Hz 100%) and *Region* (ROT, LOT, and MO) were applied. For image-presentation responses, a one-way ANOVA was tested over the OP region, with three levels of *Condition* (as above). In the case where Mauchly's test of sphericity was significant, a Greenhouse-Geisser correction was applied.

### 2.5.3. Time domain analysis

Re-referenced data segmented by condition were filtered more conservatively with a second-order, zero-phase Butterworth low-pass filter, with a cutoff value of 30 Hz, as commonly used in time-domain analyses of event-related potentials. A frequency-domain notch-filter was applied to selectively remove the response to image-presentation at 10 Hz (or 20 Hz in the third condition) and its harmonics up to 40 Hz (width 0.05 Hz), in order to isolate face-selective responses (e.g., Rossion et al., 2015; Jacques et al., 2016; Retter & Rossion, 2016). Data were then segmented by each face presentation within a condition, i.e., in intervals of 2 s starting 1 s before and continuing 1 s after each face presentation onset. The extra time was taken for initial display purposes only; data were analyzed and ultimately displayed over non-

overlapping time windows, i.e., a baseline of 100 ms prior to face stimulus onset up to 900 ms after. Epochs were then averaged, and a baseline correction was applied to each channel, with the baseline defined over the time period 100 ms before stimuli onset. Channels within the ROT, LOT, and MO ROIs were averaged. Finally, data were grand-averaged across participants by condition for display at the group level. To statistically analyze when a deflection in the time domain differed from zero, two-tailed t-tests with a significance threshold of p < .01 were run on each bin (512 bins/s) from −100 to 900 ms relative to face stimulus onset. To reduce the probability of false positives due to multiple comparisons, only groups of at least 5 consecutive time bins (about 10 ms) were considered for each ROI (e.g., Dzhelyova & Rossion, 2014; Dzhelyova et al., 2016; Jacques et al., 2016).

## 3. Results

### 3.1. EEG responses to the face presentation frequency and image presentation frequency for the different temporal conditions

As noted in the Methods, periodic presentation of faces or images produces signals in the EEG at the presentation frequencies and their higher harmonics, and were characterized firstly by summing the specific baseline-corrected harmonics. In the 10 Hz 50% condition, we expected the generic face categorization response to be maximal across occipito-temporal regions, with a right hemispheric dominance, in accordance with the literature including previous FPVS-EEG face-categorization studies (e.g., Rossion et al., 2015; Jacques et al., 2016; Retter & Rossion, 2016; Jonas et al., 2016). While expecting roughly the same pattern of responses across the other two conditions, we used the frequency-domain analysis to explore whether the temporal restrictions imposed by varying presentation duration and ISI would impact the distribution of the response across the scalp and/or its maximal amplitude.

Examination of the distribution of these responses for faces confirmed that the responses were strongest over the occipito-temporal cortex: in all three conditions, the response peaked at right occipito-temporal channel P10, followed by three adjacent channels, PO10, PO12, and P8 (Fig. 2A). In order to more clearly visualize the distribution of the face-selective response across the scalp channels topographies were normalized for each condition (see Methods; Fig. 2B). This normalization hinted at a less right-lateralized response at 20 Hz
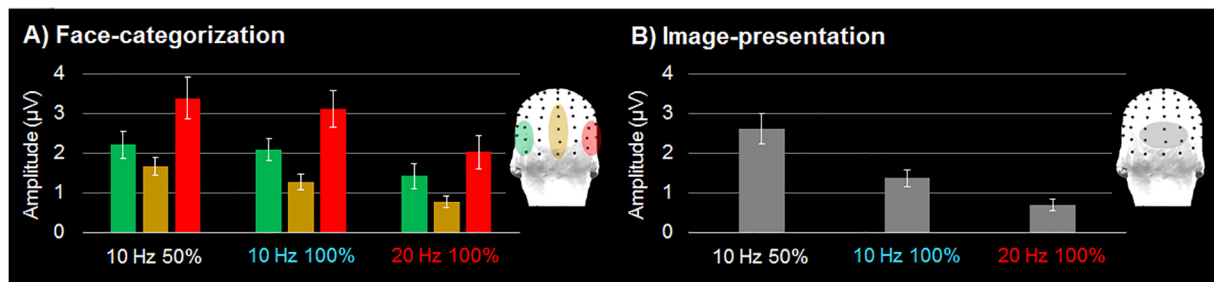
**Fig. 3.** Summed-harmonic, frequency-domain responses. Error bars show + 1/− one standard error from the mean. A) Face-selective response amplitudes for the three experimental conditions across each of the three five-channel ROIs (left occipito-temporal (green), medial-occipital (brown), and right occipito-temporal (red)). B) Image-presentation response amplitudes across a single medial, occipito-parietal ROI (gray). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

100%, and less medial activation over medial occipital channels in the 10 Hz 100% relative to 50% condition.

In order to quantitatively compare the amplitude and distribution of face-categorization responses across conditions, the amplitude was evaluated over three regions of interest (ROIs), corresponding to the right occipito-temporal, the homologous left occipito-temporal, and the medial-occipital, for each of the three conditions (Fig. 3A). Over the right occipito-temporal ROI, the amplitude of the response was similar between the two conditions at 10 Hz, but reduced by about 40% in the 20 Hz condition: *10 Hz 50%* (M = 3.38 μV; SE = 0.53 μV), *10 Hz 100%* (M = 3.11 μV; SE = 0.46 μV), and *20 Hz 100%* (M = 2.02 μV; SE = 0.41 μV).

Statistically, there were significant main effects of *Condition*, $F_{2,22} = 16.9$, $p < .001$, $\eta_p^2 = 0.61$, and *Region*, $F_{1.34,14.8} = 8.11$, $p = .008$, $\eta_p^2 = 0.42$, but no interaction between these factors, $F_{4,44} = 2.10$, $p = .097$, $\eta_p^2 = 0.16$. Post-hoc two-tailed pairwise comparisons of marginal means for *Condition*, with Bonferroni-corrected alpha values for three comparisons, found that the response to the *20 Hz 100%* condition (M = 1.40 μV; SE = 0.22 μV) was significantly lower than that to both *10 Hz 50%* (M = 2.41 μV; SE = 0.27 μV), $p = .003$, and *10 Hz 100%* (M = 2.15 μV; SE = 0.23 μV), $p = .002$. The two conditions at 10 Hz were not significantly different in amplitude across ROIs ($p = .29$).

Additionally, corresponding one-tailed comparisons for *Region* showed that, as predicted, the medial-occipital ROI (M = 1.23 μV; SE = 0.16 μV) gave a significantly lower response than both the right occipito-temporal ROI (M = 2.84 μV; SE = 0.43 μV), $p = .003$, and the left occipito-temporal ROI (M = 1.90 μV; SE = 0.29 μV), $p = .037$; however, the response of the right occipito-temporal ROI was not significantly larger than of the left occipito-temporal ROI ($p = .14$). As suggested in Fig. 2B and 3A, this lack of significant right-lateralization may have been due to the slight decrease of right-lateralization in the *20 Hz 100%* condition. Indeed, when removing the *20 Hz 100%* condition from the comparison, the right (M = 3.25 μV; SE = 0.48 μV) had a significantly larger response than the left occipito-temporal ROI (M = 2.15 μV; SE = 0.30 μV), $t_{11} = 2.10$, $p = .030$, $d = 0.94$.

A decreased medial vs. occipito-temporal activation in the *10 Hz 100%* condition relative to *10 Hz 50%* was hinted at in the data, as can be seen in Fig. 2B. Additionally, comparison of the responses across ROIs (Fig. 3A) showed that the selective response to faces in the *10 Hz 100%* condition was reduced by 23% (0.38 μV) at the medial-occipital ROI relative to the *10 Hz 50%* condition, as compared to being reduced by only 6–8% (0.13–0.27 μV) in the left and right occipito-temporal ROIs, respectively. A relative reduction in the response at medial occipital region could suggest that the face-categorization response was less impacted by associated low-level visual information. However, comparing the medial-occipital ROI to an average of the right and left occipito-temporal ROIs for the two conditions at 10 Hz revealed only a main effect of *Region*, $F_{1,11} = 23.0$, $p = .001$, $\eta_p^2 = 0.68$, with the occipito-temporal ROI (M = 2.70 μV; SE = 0.22 μV) being larger than the

medial-occipital ROI (M = 1.46 μV; SE = 0.15 μV); there was neither a main effect of *Condition*, $F_{1,11} = 3.87$, $p = .075$, $\eta_p^2 = 0.26$, nor a significant interaction between these two factors, $F_{1,11} = 1.24$, $p = .29$, $\eta_p^2 = 0.10$. In sum, the face-categorization results showed similar response amplitudes and scalp topographies for the two conditions at 10 Hz, indicating that removing the ISI does not necessarily impact these response features. However, a significantly decreased 20 Hz 100% face-categorization response, and lack of pronounced right lateralization, suggest that presenting masked images too rapidly may impair face-selective processing.

The response to image presentation at 10 Hz or 20 Hz (again based on the summed harmonics), is reported to give a baseline of a general (i.e., not face-specific) response to rapidly presented object stimuli. This response had a different spatial distribution across the scalp than the face-selective response, peaking maximally over occipito-parietal channel O2 in all three conditions (Fig. 2C and D). However, this image-presentation response showed even more pronounced amplitude differences across conditions (Fig. 3B). Over the occipito-parietal ROI, the *10 Hz 50%* condition gave the largest response (M = 2.60 μV; SE = 0.38 μV), followed by *10 Hz 100%* (M = 1.36 μV; SE = 0.21 μV), and finally *20 Hz 100%* (M = 0.694 μV; SE = 0.14 μV). Statistically, these differences produced a main effect of *Condition*, $F_{1.2,22.5} = 25.5$, $p < .001$, $\eta_p^2 = 0.70$, with post hoc pairwise comparisons of marginal means revealing significant differences between all three conditions, all p's ≤ .002. Interestingly, the amplitude differences across face- and image-presentation responses were dissociated: regarding the two conditions at 10 Hz, the amplitude was decreased in the image presentation responses without an ISI, while the amplitude did not differ across these conditions in the face presentation responses.

### 3.2. Face-selective harmonic responses are maximal up to 6 Hz

The responses to faces were also analyzed by comparing the separate harmonics in the response (Fig. 4). The distribution of amplitudes across these harmonic frequencies reveals something about the slower aspects of the response at lower frequencies and faster aspects at higher frequencies, although this interpretation is limited since different harmonic responses are not independent of one another (see Section 4.1 in the Discussion of Retter & Rossion, 2016). Still, differences in the harmonic distributions could suggest differences in the temporal dynamics of face categorization response, inspiring investigation in a complimentary time domain analysis (reported in the following section). Across all conditions, harmonic response amplitudes as evident in the spectra appeared maximal from 1 to 6 Hz: the averaged amplitude across conditions at each of these harmonics is at least 0.25 μV; at 7 Hz, the amplitude is 0.22 μV; at 8 Hz, 0.15 μV; the amplitude decreased progressively across higher harmonic frequencies.

As in the summed-harmonic comparisons, the response to faces in the *20 Hz 100%* condition was lower than the response to the conditions at 10 Hz across all harmonic response frequencies, while the *10 Hz 50%*
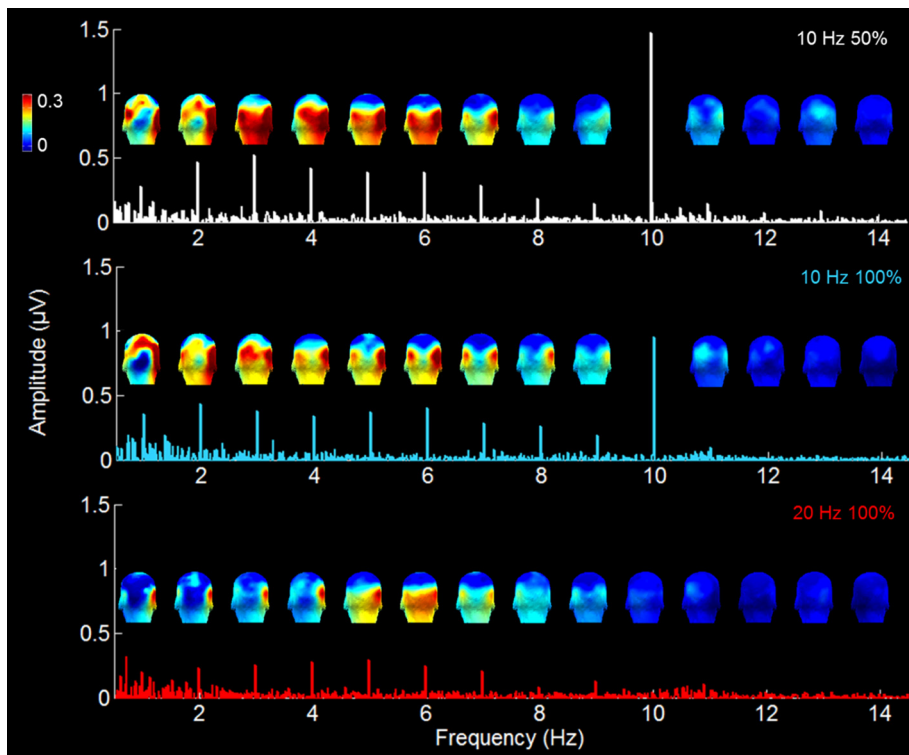
**Fig. 4.** The baseline-subtracted frequency-domain amplitude spectrum for a single right occipito-temporal channel, PO10, for each of the three experimental conditions. The back-of-the-head topographies are plotted for each face-selective harmonic response from 1 to 14 Hz, at a common scale (indicated in the top left), for each of the three conditions.

condition generally appeared similar in magnitude to that of *10 Hz 100%* condition. The apparent summed-harmonic reduction of the medial-occipital ROI (as identified in Fig. 3) in the *10 Hz 100%* relative to *10 Hz 50%* condition was shown to appear over a restricted range of harmonic frequencies, being most pronounced from about 4–9 Hz, which may have contributed to its lack of significance overall. In this range, the averaged occipito-temporal ROIs were greater than the MO ROI by about 0.05–0.1 μV for the *10 Hz 100%* condition at each harmonic response in this frequency range. The *20 Hz 100%* condition produced some changes in the face categorization response, e.g., less right-lateralization at 2 Hz and at 6–10 Hz (see the topographies in Fig. 4) and decreased amplitude at the higher harmonics. However, the responses remained distributed across a similar frequency range as in the two 10 Hz conditions. Note that these harmonic responses are displayed after a baseline-subtraction of the noise, in order to roughly equalize the noise level to zero across the spectrum, e.g., removing relatively increased noise in the alpha range or generally at lower frequencies; the original uncorrected amplitude spectrum nevertheless showed similar trends in harmonic distribution (Supplemental Fig. 1).

### 3.3. Generic face categorization response deflections in the time domain

The first aim of the temporal analysis was to test whether the delay hinted at by single unit recordings for image presentations without an ISI would be present at the population level in human face categorization responses. Additionally, we aimed to explore if the temporal dynamics of the face categorization responses might differ qualitatively in other ways across conditions. Generally, the two conditions at 10 Hz showed similar response deflections to each other, as well as to previously reported face-selective "components" from FPVS-EEG face categorization studies (Fig. 5A; Rossion et al., 2015; Retter & Rossion, 2016). Here, each condition had significant deflections at the time of the previously defined "P1-face", "N1-face", and "P2-face" which continued to be significantly above zero until the time of the "P3-face" peak. The peak times over the ROT ROI for each of these components are given in Fig. 5B.

Interestingly, there was a somewhat consistent delay in the peak

timing of the 10 Hz 100% relative to the 10 Hz 50% condition. This delay over the ROT was on average 13.8 ms (SD = 4.71 ms), being largest (about 18 ms) across the first two peaks. The significant ROT onset of the first response component, the "P1-face," also showed a similar delay (19.5 ms). Note, however, that a similar delay was evident when examining responses to image presentation over the same ROT region (Supplemental Fig. 2), suggesting that this effect was not specific to face categorization.

The 20 Hz 100% condition was more differentiated in its temporal responses, but its comparisons with the 10 Hz conditions are complicated by differences in magnitude, as well as decreased right-lateralization. The temporal response showed a similar P1-face peak latency to the 100% at 10 Hz over the ROT ROI (both 143 ms), although this component did not reach significance (at the threshold of p < .01) due to its decreased magnitude. However, the ROT response then became more in line with the 50% condition at 10 Hz (N1-face peak at 186 ms), proceeded to show the fastest P2-face peak (272 ms), and finally did not show a clear deflection near the time of the P3-face. The offset of its significant response time over the ROT was at 333 ms, at least 70 ms earlier than in the 10 Hz conditions.

The spatial distribution of these response components across the scalp appeared largely consistent across conditions (Fig. 5B). In the two conditions at 10 Hz, there was again the appearance of less medial activation at 100% vs. 50%, particularly in the first two response components. Finally, the ROT response (see Fig. 5A) was compared with the response over more low-level MO region (Supplemental Fig. 3). The peak magnitudes of the MO ROI response in the 100% condition was reduced relative to the 50% condition by over 10% at the P1-face and over 20% at the N1-face. In contrast, over the ROT ROI the 100% condition magnitude did not considerably differ from the 50% condition (a slight decrease at the P1-face, and a slight increase at the N1-face, both by less than 10%). Such MO differences in these two components may be related to similar effects described in the harmonic frequency responses about 4–9 Hz, but again this reflects an effect too focal to have produced significant summed-harmonic differences.

It may be noted that the MO ROI response time course (as shown in Supplemental Fig. 3) did not show all the same differences as the ROT
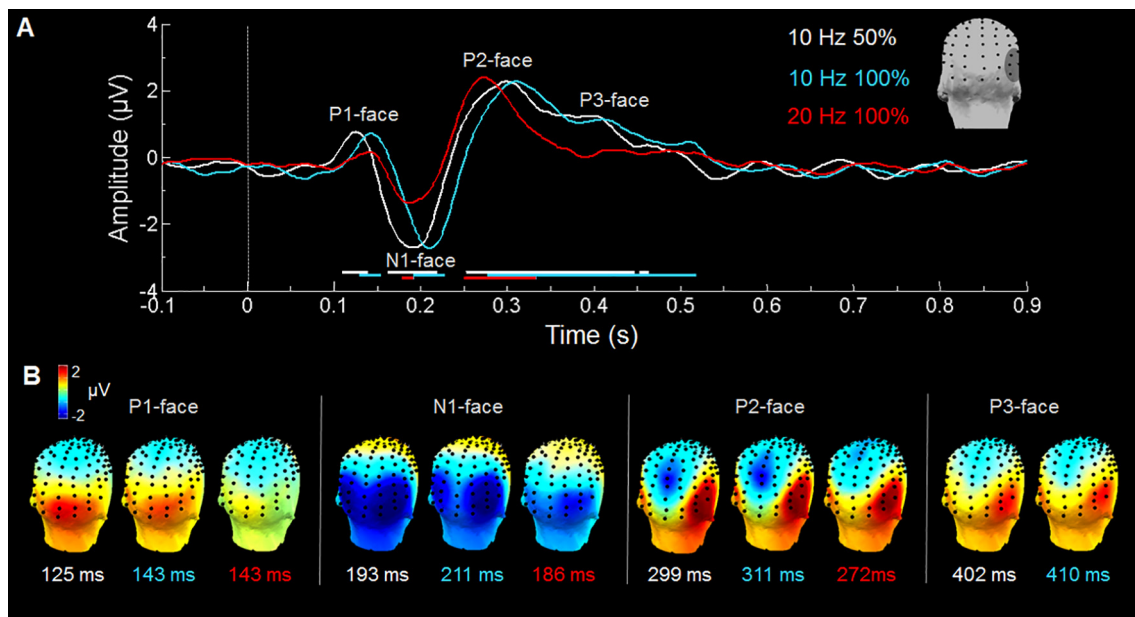
**Fig. 5.** A) Time-domain responses over the right occipito-temporal ROI for each of the experimental conditions. The vertical line at 0 s indicates face onset time; horizontal bars below the waveform indicate when deflections are significantly different from zero. B) Scalp topographies at the time of the peak magnitude of the response deflections labeled on the waveform in part A for each condition. The exact peak time for each condition at every deflection is indicated below each topography, color-coded to indicate which condition is plotted, in the same schema and order as part A.

ROI across conditions, although its weaker overall magnitude limited comparisons. At the two 10 Hz conditions, the MO ROI significant onset difference was less than 4 ms, but the peak latency was 7.8 ms later in the 100% condition. At the N1-face, the MO response was similarly delayed to the ROT response in the 100% relative to 50% 10 Hz condition (21 ms later in significant onset latency; 22 ms in peak latency). In neither 10 Hz condition did the MO ROI show any significant deflections by the time of the P2-face (the last significant response time in any condition was 260 ms). In the 20 Hz condition, the MO ROI response peaked around the same time as the 10 Hz 50% condition at the P1-face (difference < 1 ms). Similarly as with the ROT ROI, the MO 20-Hz response peaked earlier than the other two conditions by the N1-face (5–23 ms). Still, overall, the temporal dynamics of the face-categorization responses across conditions were marked by similarities than differences, apart from the delay produced without an ISI.

## 4. Discussion

To summarize, we investigated the effects of removing the ISI on FPVS-EEG face categorization and image presentation responses while controlling for image presentation duration within a squarewave presentation mode. The results showed similar face-categorization response amplitudes across the scalp for the two conditions at the rate of 10 Hz, with and without an ISI, but a decreased and less right-lateralized response at 20 Hz without an ISI (matching the presentation duration of images in the first 10 Hz condition with an ISI). However, the conditions without an ISI showed a delayed response onset, by about 18 ms at the peak of the first, P1-face, component. Thus, there was a dissociation in the effects of ISI and stimulus presentation duration on the rapid face categorization response. In comparison, the responses to image presentation showed parallel effects on these two parameters: a decreased response amplitude when removing the ISI at 10 Hz as well as a delay in response onset.

### 4.1. Removing the ISI does not reduce the face categorization response amplitude

Taken together, these results suggest that removal of the ISI at 10 Hz

did not impact the amplitude and topography of selective face categorization responses. There was an effect only when the removal of the ISI coincided with an increased image presentation rate, i.e., decreasing the SOA from 100 ms to 50 ms, in the 20 Hz condition. This is line with the results of behavioral studies using RSVP, which have shown that ISI is not important for picture memory when images are already presented for relatively long durations, i.e., about 110–180 ms (Intraub, 1980; Potter, Staub, & O'Connor, 2004). It also corresponds with the results from single-cell recordings in monkey superior-temporal sulcus, in which the firing rate in response to monkey head images is not considerably reduced when the ISI (51–93 ms) is removed but the stimulus presentation duration is extended, such that the SOA is unaffected. However, when the ISI is removed but the presentation rate is consequently increased, the response is significantly decreased (Keysers & Perrett 2002; Keysers et al., 2005). In these studies by Keysers and colleagues, a complementary face detection task was also performed by human participants, with similar patterns of results in accuracy, suggesting that the neural effects corresponded with perception. Here, we show that the effects like those measured in single monkey neurons are also evident at the population-level in the EEG responses of the human brain, lending further support to this link between neural activity and behavior.

A lack of a response reduction when removing the ISI and maintaining a 10 Hz presentation rate is also in line with the results from Retter and Rossion (2016), in which a 12.5 Hz 50% squarewave presentation mode showed no amplitude, scalp topography, or harmonic frequency response distribution differences as compared to a sinusoidal presentation mode at the same frequency (see Fig. 7 of Retter & Rossion, 2016). Note that, as mentioned in the Introduction, in sinusoidal stimulus presentation stimuli are modulated in contrast at each display frame according to a sinusoidal function. Thus, while sinusoidal presentation employs an ISI of at least a single frame, given the varying contrast levels within each stimulus presentation it is difficult to identify the exact stimulus presentation and ISI durations in that presentation mode. This was the reason that a squarewave presentation mode was employed in the present experiment.

While the face categorization response amplitude was not affected by removal of the ISI at 10 Hz, the image presentation response was

significantly reduced over a medial, occipito-parietal ROI. This may be attributable to the lack of both stimulus onset and stimulus offset against a uniform field, each occurring at 10 Hz in the 10 Hz 50% condition. Between the two conditions without an ISI, the distribution of the response is less medially-centered at 10 than 20 Hz (see Fig. 2D), perhaps hinting at relatively less reliance on low-level image cues for processing object images at 10 Hz.

This dissociation suggests that a reduction of amplitude in lower-level visual areas does not necessarily propagate to a high-level face-selective response. An alternative explanation is that a raw face categorization response would be reduced without an ISI, but in proportion to the reduced general response to image presentation, such that the differential response at 1 Hz and its harmonics is equal to that with an ISI. However, the image presentation and face categorization responses mainly originate from differing areas of the brain (Jonas et al., 2016), and project with different distributions across the scalp. Moreover, there is no apparent increased occipito-temporal relative to medial-occipital activation in the image presentation response of the 10 Hz condition with an ISI; if anything, it is the opposite (see Fig. 2). Thus, these data provide an illustration of the independence of the face categorization and image presentation responses with scalp EEG.

### 4.2. Presenting images at 20 Hz (with no ISI) decreases the face categorization response amplitude

A decreased face categorization response amplitude was found here only for the 20 Hz 100% condition. Such an effect was also reported and is evident in the previously referenced figures from Keysers and colleagues for stimuli presented without an ISI below about 90 ms (Keysers & Perrett, 2002; Keysers et al., 2005). Note that here, while faces are presented as every one out of 20 images at 20 Hz, and as every 1 out of 10 images for the two conditions at 10 Hz, this difference in proportion is not expected to be consequential (Retter & Rossion, 2016).

On the other hand, many previous studies have focused on a different question, asking what is the minimal amount of time for perceiving stimuli with RSVP and other masking paradigms. For example, in a single cell study in monkeys, neurons in the temporal cortex showed some RSVP image discrimination up to a rate of 72 Hz without any ISI (Keysers et al., 2001). In behavioral RSVP studies, meaning may be identified from rapid image sequences at up to about 77 Hz (Potter et al., 2014). However, in the present study we sought to estimate the complete face-categorization processing that takes place under natural conditions (possible from 40 to 50 ms of image duration across this experiment and that of Retter & Rossion, 2016), instead of the minimal amount of time for face detection. In these terms, decreased responses have been reported for stimuli presented with sinusoidal contrast modulation above 6 Hz for individual identity discrimination (EEG: Alonso-Prieto et al., 2013; fMRI: Gentile & Rossion, 2014). The corresponding limits of full temporal responses for generic face categorization have yet to be explored in detail, but the present evidence shows that responses are reduced between 10 and 20 Hz.

This reduction could be explained by both forward and backward interference of the responses to temporally adjacent images, and thus may point to limits in the finite temporal window over which the categorization process occurs. Such neural interference, termed "competition" by Keysers and Perrett in 2002, has been shown previously in FPVS-EEG responses to face stimuli (Retter & Rossion, 2016). These results agree with other studies pointing to interrupted visual processing from masking at about 20 Hz and above (e.g., Liu, Agam, Madsen, & Kreiman, 2009; Perrett, Hietanen, Oram, & Benson, 1992; Thorpe & Fabre-Thorpe, 2001).

### 4.3. Removing the ISI increases response latency

In the study of Retter and Rossion (2016), differences were reported in terms of a 20-ms delay of response onset when presenting stimuli

with a sinewave, although this was likely affected by the delayed, sinusoidally-increasing stimulus onset (i.e., progressing from 0 to 15 to 50% contrast over the first three monitor refresh frames at 100 Hz, with a response likely not originating until the third frame onset, 20 ms after stimulus presentation onset was recorded). Here, we report a delay from the removal of the ISI of about 18–20 ms for the response onset of the ROT region, evident in both 10 Hz and 20 Hz 100% conditions relative to 10 Hz 50% (and for the image presentation as well as the face categorization responses). Since a squarewave presentation mode is used, stimuli are always presented at full contrast, thus this delay is not attributable to an onset delay from sinusoidally increasing stimulus contrast.

As noted in the introduction, while such differences in response onset were not targeted by Keysers and Perrett (2002), a delayed response onset from the removal of the ISI was hinted at in the reported data (Fig. 3A; also in Keysers et al., 2005: Fig. 3). Such a delay could be explained by transient off and/or sustained responses generated from forward masking suppressing the (onset) responses to subsequent images (Breitmeyer et al., 1981; Keysers et al., 2005; Macknik & Livingstone, 1998; Ogmen et al., 2003). While forward masking may affect stimuli presented even with a brief ISI, this effect may be particularly salient without an ISI because image offset and image onset occur at exactly the same time, possibly producing the maximal amount of interference. Such masking could be attributed to low-level interference in early visual areas and/or to interference in higher-level areas, with effects depending on the overlap of processing between the masking and test stimuli (e.g., Kim & Mullen, 2015; but see also Rolls & Tovée, 1994).

While neural interference may be sufficient to account for our results, this interpretation is speculative and so we will continue to address the data in light of additional factors, such as the potential neural effects of forward masking. (Note that we only discuss forward masking at present, since backward masking may interrupt processing of the previous image but occurs too late to affect its onset time). Forward masking without an ISI may effectively decrease the stimulus energy at onset, leading to a slower accumulation of information in the visual system, since replacing a full-contrast image with a full-contrast image is a more subtle change than replacing a zero-contrast image with a full-contrast image. Stimulus onset is also less coherent, since at a local level luminance value changes across pixels are less synchronized across a wider range (i.e., compared to the zero-contrast background, the difference at each stimulus pixel ranges from the mean luminance; while compared to another stimulus, the difference at each stimulus pixel ranges across the full luminance range). This could contribute to less synchronized transient onset responses in early visual areas, delaying the activation of further visual processes (Samonds & Bonds, 2005).

Additionally, the sustained level of contrast across stimuli presented without an ISI may also contribute to increased contrast adaptation, i.e., perceptually lowered contrast sensitivity (e.g., Bonds, 1991). Images, including both upright and inverted faces as well as grating stimuli, presented at lower contrast have been shown to elicit greater response onset latencies (Albrecht, 1994; Schneider, DeLong, & Busey, 2007; Hu et al., 2011; Sani, Santandrea, Golzar, Morrone, & Chelazzi, 2013) [2]. Moreover, it has been proposed that neurons in early visual areas have faster temporal responses to stimuli presented with ISIs than without as an effect of temporal context (Hu et al., 2011). As a reminder, we do not think this delay is specific to face categorization because comparable delays were also evident in the responses to rapidly presented object stimuli (Supplemental Fig. 2).

---

[2] Schneider & DeLong, 2007 also report decreased component magnitudes with lower stimulus contrast. Here, a reduction in amplitude may not be revealed over the right occipito-temporal region in the 10 Hz 100% relative to 10 Hz 50% condition due to the amount of contrast invariance in higher-level visual areas (Avidan et al., 2002) and because the amplitude of the face categorization response is measured in comparison to the image presentation response (Rossion et al., 2015).

#### 4.4. Independence of face-categorization and image-presentation responses

As mentioned in Section 1, the consistency of face-categorization response amplitudes across 10 Hz conditions is paired with inconsistency in terms of the image-presentation response, for which there was a reduction over the occipito-parietal ROI of over 50% in the 100% condition relative to the 50% condition. This is in striking contrast to the effects reported for latency in Section 3 of the two 10 Hz conditions, wherein face categorization and image presentation responses were similarly delayed when no ISI was present. It appears that latency differences, likely present in early visual areas contributing to more low-level image presentation responses (see Supplemental Fig. 3), propagate to higher-level visual areas (e.g., Samonds & Bonds, 2005). However, when present, this latency difference did not change the qualitative evolution of response components, amplitudes, or scalp topographies across the 10 Hz condition, implying that it does not have strong functional implications.

The dissociation of the face and image response amplitudes across these conditions was even more pronounced. This may be because high-level processes must retain some level of invariance to facilitate consistent recognition across a wide range of low-level changes in the natural environment, e.g., due to lighting, viewpoint, etc. In support of this possible explanation, varying levels of image coherence produced a categorical face detection response with FPVS-EEG for normal faces, processed with high-level expertise, but not inverted or contrast-reversed faces (Liu-Shuang, Ales, Rossion, & Norcia, 2015).

#### 4.5. Practical implications

To summarize again, we have shown that removal of an ISI reliably delays face categorization responses by about 18–20 ms, but does not affect this response amplitude or scalp topography. However, when the removal of the ISI concurs with an increased stimulus presentation rate, the response amplitude and right-lateralization is shown to be decreased at 20 relative to 10 Hz. The practical implications of these results regard the design of future experiments, as well as interpretations across experiments using different stimulus presentation modes.

Considering the design of future experiments, our results suggest that presenting stimuli without an ISI is not detrimental to measuring an EEG response, beyond inducing its general delay. This could be useful, because presenting stimuli without an ISI enables testing perceptual aspects (e.g., the minimal stimulus presentation duration required for a response) with the tightly controlled limits imposed by forward- and backward-masking (e.g., Crawford, 1947; Macknik & Livingstone, 1998; Potter, 2012). Practically, this 100% duty-cycle presentation mode may also be desired because it allows for higher and more flexible presentation rates, given the rate of available monitor refresh frames. This is because an ISI requires a minimum of two frames per image presentation cycle (one "on" and one "off"). For example, given a 120 Hz monitor, stimuli may be presented for one frame without an ISI at each 120, 60, 40, 30, 24 Hz, etc., while if presented with an ISI the available rates with a 50% duty cycle are reduced to 60, 30, 20, 15, 12 Hz, etc.

On the other hand, our results show that a 50% duty cycle squarewave presentation mode also gives a full face-categorization response, at least at 10 Hz. As proposed in the Introduction, including an ISI may be useful in studies where it is important to reduce after-effects and/or the appearance of motion across fast serial presentations of images. After-effects may be particularly salient for structurally homogeneous stimuli, such as those isolating color or orientation, but may also affect higher-level aspects, e.g., the sex or attractiveness of faces. The appearance of motion across sequential images may emerge when testing at high rates, particularly with complex images (additionally, the appearance of viewing distance changes may also emerge when image size changes at each cycle). An ISI may thus be desired to temporally separate subsequent images; its optimal duration may vary with the stimuli and experimental questions, but should be sufficiently long to reduce noticeable perceptual interactions across stimuli.

Comparing across experiments using different stimulus presentation modes (e.g., different squarewave duty cycles or sinewave presentation), given reasonably low stimulus presentation rates (at least 10 Hz or below), the most important factor to remember is the expected shift in response latencies. That is, when studies use a squarewave presentation with a 50% duty cycle (e.g., Dzhelyova et al., 2016; Retter & Rossion, 2016), response onset is as immediate as possible, i.e., about 100 ms for generic face categorization or expression change over the right occipito-temporal cortex in those studies. Studies using no ISI will be relatively delayed in time, by about 20 ms here for both face categorization and common object responses over the right occipito-temporal cortex. We predict that studies using any duty cycle considerably below 100% (e.g., greater than a 10-ms ISI) will also have an immediate onset, i.e., that the delay due to removal of the ISI reported here is all-or-none. However, this remains to our knowledge an open question: it is possible that parametrically reducing ISI duration, particularly below about 50 ms as tested here, may produce incremental increases in response onset latency. This could be investigated in future experiments.

Finally, studies using a sinewave contrast presentation mode reporting time-domain and/or phase data (e.g., Dzhelyova et al., 2016; Jacques et al., 2016; Retter & Rossion, 2016; Rossion et al., 2015) will have an onset delayed by the gradual onset of stimulus presentation, which will be dependent on the rate of presentation and the monitor refresh rate, as well as the diagnostic characteristics of luminance contrast in the stimuli. Such a delay thus may be around only 20 ms with about a 12 Hz stimulus presentation rate (Dzhelyova et al., 2016; Retter & Rossion, 2016), but may be longer with lower presentation rates, e.g., in the range of 40–60 ms around 6 Hz (Rossion et al., 2015). These practical considerations will be important for reconciling findings across studies on the spatio-temporal dynamics of face categorization, and visual perception in general, in the human brain.

### Acknowledgments

### Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.visres.2018.02.009.

### References

Albrecht, D. G. (1994). Visual cortex neurons in monkey and cat: Effect of contrast on the spatial and temporal phase transfer functions. *Visual Neuroscience, 12,* 1191–1210.

Alonso-Prieto, E. A., Van Belle, G., Liu-Shuang, J., Norcia, A. M., & Rossion, B. (2013). The 6 Hz fundamental frequency rate for individual face discrimination in the right occipito-temporal cortex. *Neuropsychologia, 51,* 2863–2975.

Avidan, G., Harel, M., Hendler, T., Ben-Bashat, D., Zohary, E., & Malach, R. (2002). Contrast sensitivity in human visual areas and its relationship to object recognition. *Journal of Neurophysiology, 87*(6), 3102–3116.

Bonds, A. B. (1991). Temporal dynamics of contrast gain in single cells of the cat striate cortex. *Visual Neuroscience, 6,* 239–255.

Breitmeyer, B., Levi, D. M., & Harwerth, R. S. (1981). Flicker masking in spatial vision. *Vision Research, 21*(9), 1377–1385.

Crawford, B. H. (1947). Visual adaptation in relation to brief conditioning stimuli. *Proceedings of the Royal Society of London Series B, 134*(875), 283–302.

Duysens, J., Orban, G. A., Cremieux, J., & Maes, H. (1985). Visual cortical correlates of

visible persistence. *Vision Research, 25*(2), 171–178.

Dzhelyova, M., Jacques, C., & Rossion, B. (2016). At a single glance: Fast periodic visual stimulation uncovers the spatio-temporal dynamics of brief facial expression changes in the human brain. *Cerebral Cortex, 27*(8), 4106–4123.

Dzhelyova, M., & Rossion, B. (2014). The effect of parametric stimulus size variation on individual face discrimination indexed by fast periodic visual stimulation. *BMC Neuroscience, 15*(87), 1–12.

Gentile, F., & Rossion, B. (2014). Temporal frequency tuning of cortical face-sensitive areas for individual face perception. *NeuroImage, 90*, 256–265.

Hu, M., Wang, Y., & Wang, Y. (2011). Rapid dynamics of contrast responses in the cat primary visual cortex. *PLoS One, 6*(10), e25410.

Intraub, H. (1980). Presentation rate and the representation of briefly glimpsed pictures in memory. *Journal of Experimental Psychology. Human Learning, 6*, 1–12.

Jacques, C.∗, Retter, T. L.∗, & Rossion, B. (2016). A single glance at a face generates larger and qualitatively different category-selective spatio-temporal signatures than other ecologically-relevant categories in the human brain. *NeuroImage, 137*, 21–33.

Jonas, J.∗, Jacques, C.∗, Liu-Shuang, J., Brissart, H., Colnat-Coulbois, S., Maillard, L., et al. (2016). A face-selective ventral occipito-temporal map of the human brain with intracerebral potentials. *Proceedings of the National academy of Sciences of the United States of America, 113*, E4088–E4097.

Keysers, C., & Perrett, D. I. (2002). Visual masking and RSVP reveal neural competition. *Trends in Cognitive Science, 6*, 120–125.

Keysers, C., Xiao, D. K., Földiák, P., & Perrett, D. I. (2001). The speed of sight. *Journal of Cognitive Neuroscience, 13*, 90–101.

Keysers, C., Xiao, D.-K., Foldiak, P., & Perrett, D. I. (2005). Out of sight but not out of mind: The neurophysiology of iconic memory in the superior temporal sulcus. *Cognitive Neuropsychology, 22*(3–4), 316–332.

Kim, Y. J., & Mullen, K. T. (2015). The dynamics of cross-orientation masking at monocular and interocular sites. *Vision Research, 116A*, 80–91.

Liu, H., Agam, Y., Madsen, J. R., & Kreiman, G. (2009). Timing, timing, timing: Fast decoding of object information from intracranial field potentials in human visual cortex. *Neuron, 62*, 281–290.

Liu-Shuang, J., Ales, J. M., Rossion, B., & Norcia, A. M. (2015). Separable effects of inversion and contrast reversal on face detection thresholds and response functions: A sweep VEP study. *Journal of Vision, 10;15*(2).

Macknik, S. L., & Livingstone, M. S. (1998). Neuronal correlates of visibility and invisibility in the primate visual system. *Nature Neuroscience, 1*, 144–149.

McCarthy, G., & Wood, C. C. (1985). Scalp distributions of event-related potentials: An ambiguity associated with analysis of variance models. *Electroencephalography and Clinical Neurophysiology, Evoked Potentials Section, 62*, 203–208.

Mouraux, A., Iannetti, G. D., Colon, E., Nozaradan, S., Legrain, V., & Plaghki, L. (2011). Nociceptive steady-state evoked potentials elicited by rapid periodic thermal stimulation of cutaneous nociceptors. *Journal of Neuroscience, 3116*, 6079–6087.

Norcia, A. M., Appelbaum, L. G., Ales, J. M., Cottereau, B., & Rossion, B. (2015). The steady-state visual evoked potential in vision research: A review. *Journal of Vision,*

*15*(6) 4, 1–46.

Ogmen, H., Breitmeyer, B. G., & Melvin, R. (2003). The what and where in visual masking. *Vision Research, 12*, 1227–1350.

Oostenveld, R., & Praamstra, P. (2001). The five percent electrode system for high-resolution EEG and ERP measurements. *Clinical Neurophysiology, 112*(4), 713–719.

Perrett, D., Hietanen, J., Oram, M., & Benson, P. (1992). Organization and functions of cells responsive to faces in the temporal cortex. *Philosophical Transactions of the Royal Society B, 335*, 23–30.

Potter, M. C. (2012). Recognition and memory for briefly presented scenes. *Frontiers in Psychology, 3*, 32.

Potter, M. C., & Levy, E. I. (1969). Recognition memory for a rapid sequence of pictures. *The Journal of Experimental Psychology, 81*, 10–15.

Potter, M. C., Staub, A., & O'Connor, D. H. (2004). Pictorial and conceptual representation of glimpsed pictures. *Journal of Experimental Psychology: Human Perception and Performance, 30*, 478–489.

Potter, M. C., Wyble, B., Hagmann, C. E., & McCourt, E. S. (2014). Detecting meaning in RSVP at 13 ms per picture. *Attention, Perception, & Psychophysics, 76*(2), 270–279.

Regan, D. (1966). An effect of stimulus colour on average steady-state potentials evoked in man. *Nature, 210*, 1056–1057.

Retter, T. L., & Rossion, B. (2016). Uncovering the neural magnitude and spatio-temporal dynamics of natural image categorization in a fast visual stream. *Neuropsychologia, 91*, 9–28.

Rolls, E. T., & Tovée, M. J. (1994). Processing speed in the cerebral cortex and the neurophysiology of visual masking. *Proceedings of the Royal Society B, 257*, 9–15.

Rossion, B., Torfs, K., Jacques, C., & Liu-Shuang, J. (2015). Fast periodic presentation of natural face images reveals a robust face-selective electrophysiological response in the human brain. *Journal of Vision, 15*(1), 1–18.

Samonds, J. M., & Bonds, A. B. (2005). Gamma oscillation maintains stimulus structure-dependent synchronization in cat visual cortex. *Journal of Neurophysiology, 93*(1), 223–236.

Sani, I., Santandrea, E., Golzar, A., Morrone, M. C., & Chelazzi, L. (2013). Selectively tuning for contrast in macaque area V4. *Journal of Neuroscience, 33*(47), 18583–18596.

Schneider, B. L., DeLong, J. E., & Busey, T. A. (2007). Added noise affects the neural correlates of upright and inverted faces differently. *Journal of Vision, 7*(4), 1–24.

Silberstein, R. B., Schier, M. A., Pipingas, A., et al. (1990). *Brain Topography, 3*, 337–347.

Thorpe, S., & Fabre-Thorpe, M. (2001). Seeking categories in the brain. *Science, 291*, 260–263.

van der Tweel, L. H., & Verduyn Lunel, H. F. E. (1965). Human visual responses to sinusoidally modulated light. *Electroencephalography and Clinical Neurophysiology, 18*(6), 587–598.

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