Sensitivity to timing and order in human visual cortex

Jedediah M. Singer, Joseph R. Madsen, William S. Anderson and Gabriel Kreiman

J Neurophysiol 113:1656-1669, 2015. First published 26 November 2014; doi:10.1152/jn.00556.2014

J Neurophysiol 113: 1656–1669, 2015.
First published November 26, 2014; doi:10.1152/jn.00556.2014.

Sensitivity to timing and order in human visual cortex

Jedediah M. Singer,[1] Joseph R. Madsen,[2] William S. Anderson,[3] and Gabriel Kreiman[1,4,5]

[1]*Department of Ophthalmology, Boston Children's Hospital, Harvard Medical School, Boston, Massachusetts;* [2]*Department of Neurosurgery, Boston Children's Hospital, Harvard Medical School, Boston, Massachusetts;* [3]*Department of Neurosurgery, Johns Hopkins Hospital, Johns Hopkins University School of Medicine, Baltimore, Maryland;* [4]*Center for Brain Science, Harvard University, Cambridge, Massachusetts; and* [5]*Swartz Center for Theoretical Neuroscience, Harvard University, Cambridge, Massachusetts*

Submitted 28 July 2014; accepted in final form 18 November 2014

Singer JM, Madsen JR, Anderson WS, Kreiman G. Sensitivity to timing and order in human visual cortex. *J Neurophysiol* 113: 1656–1669, 2015. First published November 26, 2014; doi:10.1152/jn.00556.2014.— Visual recognition takes a small fraction of a second and relies on the cascade of signals along the ventral visual stream. Given the rapid path through multiple processing steps between photoreceptors and higher visual areas, information must progress from stage to stage very quickly. This rapid progression of information suggests that fine temporal details of the neural response may be important to the brain's encoding of visual signals. We investigated how changes in the relative timing of incoming visual stimulation affect the representation of object information by recording intracranial field potentials along the human ventral visual stream while subjects recognized objects whose parts were presented with varying asynchrony. Visual responses along the ventral stream were sensitive to timing differences as small as 17 ms between parts. In particular, there was a strong dependency on the temporal order of stimulus presentation, even at short asynchronies. From these observations we infer that the neural representation of complex information in visual cortex can be modulated by rapid dynamics on scales of tens of milliseconds.

visual object recognition; intracranial field potentials; temporal coding; electrocorticography; human neurophysiology; ventral visual cortex; computational models

SHAPE RECOGNITION IS ESSENTIAL for most visual tasks and depends on continuous integration of visual cues over space and time. Shape recognition relies on the semihierarchical cascade of linear and nonlinear steps along the ventral visual stream (Haxby et al. 1991; Rolls 1991; Logothetis and Sheinberg 1996; Tanaka 1996; Connor et al. 2007). Several studies have documented the spatial integration properties of neurons along the ventral stream, showing that receptive field sizes increase from early visual cortex all the way to inferior temporal cortex (ITC) (Gattass et al. 1981, 1988; Kobatake and Tanaka 1994; DiCarlo and Maunsell 2000; Yoshor et al. 2007; Dumoulin and Wandell 2008; Agam et al. 2010). Furthermore, the presence of multiple objects can significantly influence the physiological responses within the receptive field (Missal et al. 1999; Gawne and Martin 2002; Zoccolan et al. 2007; Agam et al. 2010; Baeck et al. 2013).

Comparatively less is known about the dynamics underlying temporal integration of visual information, particularly in the highest echelons of ventral cortex. Mean response latencies progressively increase along the ventral stream by ~15 ms at each stage (Schmolesky et al. 1998), and selective responses to complex shapes have been reported in ITC at 100–150 ms poststimulus onset both in macaque monkeys (Richmond et al. 1990; Keysers et al. 2001; Hung et al. 2005) and humans (Thorpe et al. 1996; Liu et al. 2009). This rapid progression in information transmission suggests that fine temporal details of the neural response may be important to the brain's encoding of visual signals and has led to theories describing visual recognition via bottom-up and hierarchical concatenation of linear and nonlinear processing steps (Fukushima 1980; Wallis and Rolls 1997; Riesenhuber and Poggio 1999; Deco and Rolls 2004; Serre et al. 2007). At the same time, neurons often show response durations that span several tens to hundreds of milliseconds (Richmond et al. 1990; Ringach et al. 1997; Keysers and Perrett 2002; De Baene et al. 2007), which may endow them with the potential to integrate visual inputs over time. Behavioral studies have suggested windows of temporal integration that range from several tens to hundreds of milliseconds (Clifford et al. 2004; Singer and Sheinberg 2006; Anaki et al. 2007; Schyns et al. 2007; Singer and Kreiman 2014).

To further our understanding of how sensory stimuli are integrated over space and time in the human ventral visual stream, here we investigated whether changes in the relative timing of incoming visual stimulation affect the representation of object information. We recorded intracranial field potentials while subjects recognized objects whose parts were presented with varying asynchrony. We proposed and evaluated quantitative models based on responses to the individual parts and on responses at different asynchrony values to describe the extent of spatiotemporal integration by the ensuing physiological signals. Temporal asynchrony as short as 17 ms led to significant differences in the neural responses. Furthermore, distinct responses were evoked by altering the relative order of object part presentation. These results demonstrate that the human ventral visual stream is sensitive to relative timing on scales of ~17 ms.

MATERIALS AND METHODS

Subjects

Subjects were five patients (4 female) at Boston Children's Hospital and Johns Hopkins Medical Center, with subdural and/or depth electrodes (Ad-Tech, Racine, WI) implanted for clinical purposes as part of epilepsy treatment. The number of electrodes as well as their location was exclusively dictated by clinical considerations. The number of electrodes per patient ranged from 84 to 186 (total = 628); the electrode locations are described in Table 1. Sixteen healthy subjects (10 female) performed a psychophysics experiment described below. All procedures were performed with informed consent and

Address for reprint requests and other correspondence: G. Kreiman, 3 Blackfan Circle, Rm. 13075, Boston, MA 02215 (e-mail: gabriel.kreiman@tch.harvard.edu).

Table 1. *Electrode/response properties and counts by brain region*

Region	No. Electrodes	No. Responses to Whole	No. Responses to Both Halves	No. Order-Sensitive Responses	No. *Eq. 1*	No. *Eq. 2*
Visual cortex						
Anterior transverse collateral sulcus	3	7 (58%)	4 (57%)	0 (0%)	0 (0%)	0 (0%)
Fusiform gyrus	24	54 (56%)	44 (81%)	21 (48%)	11 (25%)	8 (18%)
Inferior occipital gyrus/sulcus	25	55 (55%)	41 (75%)	18 (44%)	10 (24%)	10 (24%)
Occipital pole	35	70 (50%)	48 (69%)	29 (60%)	17 (35%)	9 (19%)
Lateral occipitotemporal sulcus	2	4 (50%)	4 (100%)	2 (50%)	0 (0%)	1 (25%)
Middle occipital/lunate sulcus	5	8 (40%)	6 (75%)	4 (67%)	0 (0%)	0 (0%)
Middle occipital gyrus	40	63 (39%)	55 (87%)	25 (45%)	14 (25%)	15 (27%)
Inferior temporal gyrus	39	31 (20%)	19 (61%)	12 (63%)	10 (53%)	10 (53%)
Visual cortex totals	173	292	221	111	62	53
Intraparietal sulcus	6	3	0	0		
Parahippocampal gyrus	22	6	1	0		
Superior temporal sulcus	11	2	2	0		
Cuneus	23	4	2	0		
Superior occipital gyrus	7	1	0	0		
Precentral gyrus	9	1	0	0		
Middle temporal gyrus	82	7	3	0		
Supramarginal gyrus	47	4	1	0		
Lingual gyrus	39	3	1	0		
Depth electrodes	26	2	0	0		
Subcentral gyrus/sulcus	14	1	0	0		
Angular gyrus	28	2	1	0		
Lateral superior temporal gyrus	49	3	0	0		
Planum temporale	5	0	0	0		
Postcentral gyrus	14	0	0	0		
Postcentral sulcus	3	0	0	0		
Temporal pole	18	0	0	0		
Short insular gyri	1	0	0	0		
Operculum	6	0	0	0		
Posterior ventral cingulate	1	0	0	0		
Collateral/lingual sulcus	4	0	0	0		
Inferior temporal sulcus	5	0	0	0		
Precuneus	8	0	0	0		
Superior/transverse occipital sulcus	2	0	0	0		
Unclassified	1	0	0	0		
Superior parietal gyrus	4	0	0	0		
Calcarine sulcus	1	0	0	0		
Triangular inferior frontal gyrus	4	0	0	0		
Parieto-occipital sulcus	1	0	0	0		
Middle frontal gyrus	2	0	0	0		
Hippocampus	2	0	0	0		
Totals	618	331	232	111		

Summary of electrode locations and response properties. We parceled the brain into 75 regions (see MATERIALS AND METHODS) and report the number of electrodes in each region, the number of responses to whole images (percentage with respect to total responses in parentheses), the number of responses to both image halves (percentage with respect to responses to whole images in parentheses), the number of order-sensitive responses at 17 ms (percentage with respect to responses to both halves in parentheses), the number of responses described by the simple linear model (*Eq. 1*, percentage with respect to responses to both halves in parentheses), and the short stimulus onset asynchrony (SOA) model with order (*Eq. 2*, percentage with respect to responses to both image halves in parentheses).

approved by the Boston Children's Hospital and Johns Hopkins Medical Center Institutional Review Boards.

Task

Subjects were shown asynchronously presented image parts and were asked to identify the images (Fig. 1*A*). Each subject saw grayscale images with flattened intensity histograms from one of two sets of four stimuli; each stimulus was constructed of two parts (a top part and a bottom part), and each part was present in two images (Fig. 1*B*). Subjects were familiarized with the images and their names before the experiment. It was made explicit that sometimes the two parts of the image would be shown at different times but that the subject should still respond according to the image whose parts were presented. A fixation cross persisted throughout the trial. Each image subtended ~5–6° of visual angle vertically and 4–6° horizontally. Each trial began with 500 ms of low-contrast phase-scrambled noise at 60 Hz. One of the two image parts then appeared on the screen for one screen refresh (17 ms). The second image part was presented, also for 17 ms, with a stimulus onset asynchrony (SOA) of 0 (both parts appeared simultaneously), 1, 2, 3, 6, 15, or 42 screen refreshes (0, 17, 33, 50, 100, 250, or 700 ms, respectively). The flickering noise continued behind and between the two parts, and for 500 ms after the onset of the second image part. The subject was then presented with a screen giving the four image choices and the corresponding buttons

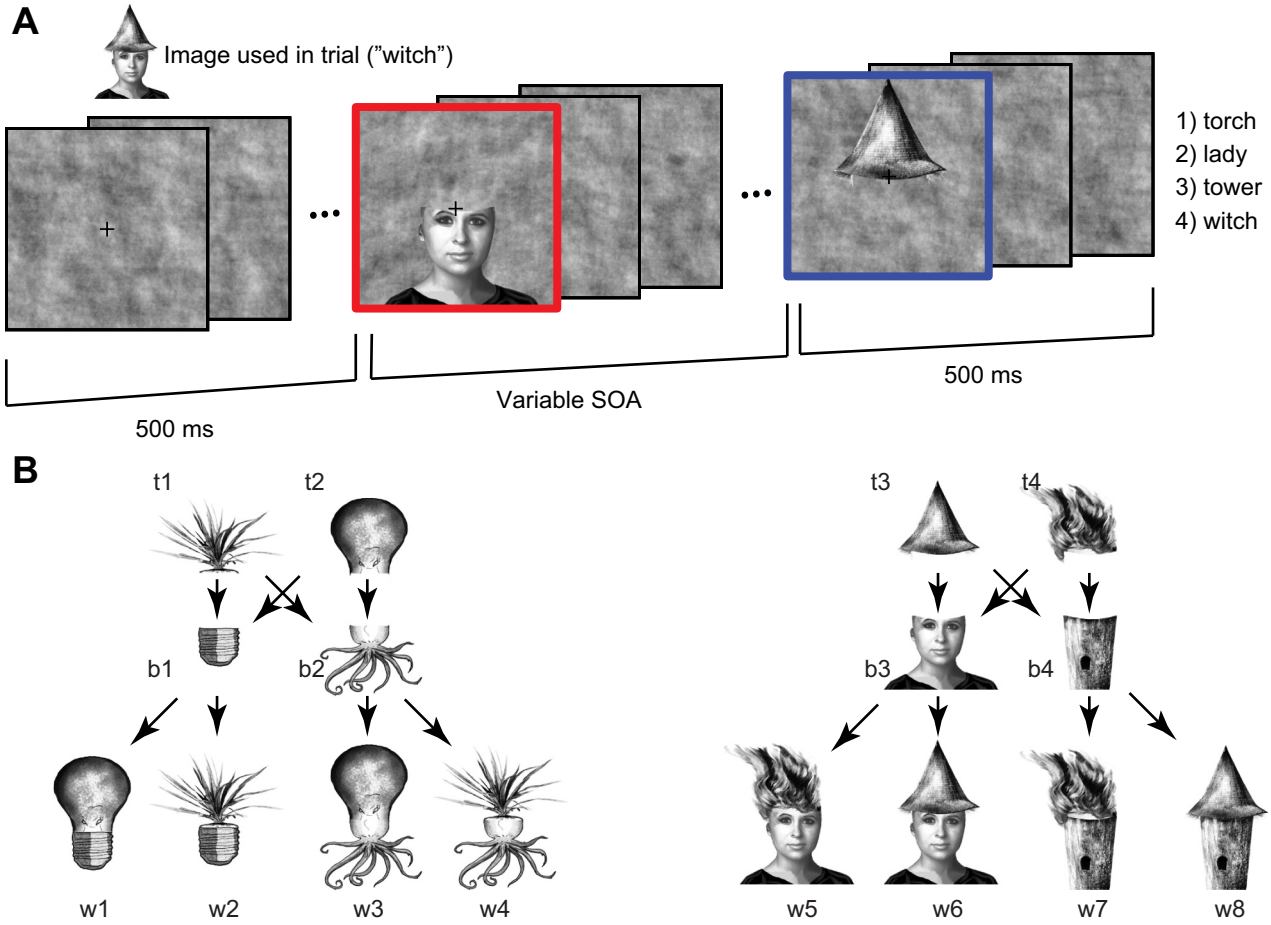

Fig. 1. Trial structure and images used. *A*: subjects identified which of 4 two-part images was shown in a 4-alternative forced-choice task (Supplemental Movie S1; see MATERIALS AND METHODS). Each part was flashed for 17 ms in the midst of low-contrast noise. Parts were separated by a stimulus onset asynchrony (SOA) ranging from 0 to 250 ms. The order of presentation of SOAs, images, and parts was randomized across trials. Frames did not have colored borders in the actual experiment (shown here in relation to Fig. 2). *B*: 2 sets of images used. Each set contained 2 bottom part and 2 top parts that were combined to form 4 possible whole images. Each part was used in 2 images. w1 ("light bulb") = b1 + t2; w2 ("houseplant") = b1 + t1; w3 ("octopus") = b2 + t2; w4 ("turnip") = b2 + t1; w5 ("lady") = b3 + t4; w6 ("witch") = b3 + t3; w7 ("torch") = b4 + t4; and w8 ("tower") = b4 + t3.

to press (4-alternative forced choice). Mapping between images and button presses remained fixed throughout the experiment. No correct/incorrect feedback was provided, except for an overall score at the end of each block of 40 trials. The order of presentation of the different images and SOA values was randomized.

Physiological Recordings

Electrophysiological data were recorded and digitized at 500, 1,000, 1,024, or 2,000 Hz (depending on the subject) using either an XLTEK or Nihon Kohden clinical system. All analyses of electrophysiological data were performed with MATLAB software (Mathworks, Natick, MA). We subtracted the mean across all electrodes from each channel to reduce externally induced artifacts, for example, due to mechanical or electrical noise. We bandpass filtered the data between 1 and 100 Hz, with a notch filter at 60 Hz to remove line noise. To reduce artifacts, we excluded trials in which any sample was more than 4 SDs (over all trials) from the mean response (over trials with the same image and SOA). This excluded 4.5, 4.3, 2.0, 5.9, and 3.7% of trials, respectively, for the five subjects. We considered broadband responses (1–100 Hz) and also the high-gamma-power time course (65–100 Hz). To evaluate the high gamma-power time course, we bandpass filtered the data between 65 and 100 Hz and then calculated the absolute value of the analytic signal generated by the Hilbert transform of the filtered data (Freeman et al. 2003).

Electrodes were localized by coregistering a preoperative structural MRI scan with a postimplantation CT scan. We used Freesurfer software to compute a three-dimensional representation of the cortical surface from the structural MRI and manually located each electrode shown in the CT scan on this surface. Brain regions and Talairach coordinates were also calculated using Freesurfer (Fischl et al. 2004; Destrieux et al. 2010). Freesurfer coregisters each patient's brain with a standardized average brain (Montreal Neurological Institute coordinates in the case of regional parcellation and the Talairach brain in the case of Talairach coordinates). We excluded five electrodes because they appeared to be shorted.

Data Analyses

Visual responsiveness. To assess visual responsiveness, we computed the mean across all whole image trials for each electrode. An electrode was considered to be visually responsive to an image if the range (max-min) of the mean response between 50 and 350 ms after image onset was larger than chance, as determined by a permutation test in which responses to individual trials were randomly multiplied by either 1 or −1 (10,000 iterations; $P < 0.0001$).

We also used a permutation test to partition visual responses into those that showed order sensitivity and those that did not. Trials at 17-ms SOA in both orders were combined and randomly partitioned into two groups; the root-mean-square deviation

(RMSD) between the mean time courses of the two groups was calculated (over the window between 50 and 350 ms after image onset), and this process was repeated 5,000 times. The actual RMSD between the mean responses for each order was then compared with this distribution. We used the same kind of permutation test to evaluate differences between responses to whole images and different SOA values (in both orders).

In some cases, responses to two-part stimuli were dominated by the response to one of the parts; while this sometimes reflected a winner-take-all interaction between two part responses, in other cases the weaker part did not elicit a response even in isolation. These latter responses likely reflected limited receptive fields (spatial or object-related) and thus were not pertinent to this study of interactions between parts. We therefore excluded all responses in which asynchronous presentations could be described by the response to one of the parts (see below) and one of the responses to individual parts failed to reach a threshold ($P < 0.001$) in the same type of permutation test used to ascertain responsiveness to whole images. While these patients were suffering from epilepsy, none of the electrodes that we considered were over seizure foci or areas of early seizure spread. They were affected only when seizures generalized to the entire brain, suggesting that they were not pathological tissue.

Quantitative models to describe the interactions of image parts. We evaluated whether the response to the asynchronously presented objects could be described by linear combination of the responses to the isolated constituent parts. Let t be the time from onset of the first image part. Let $r_B(t)$ denote the response (averaged across trials) to the bottom image part (calculated from the SOA = 700 ms trials), and similarly let $r_T(t)$ be the response to the top part. We denote by $\hat{r}(t,\mathrm{SOA})$ the model's prediction at a given time t and SOA (SOA > 0 represents trials where the top part was presented first and SOA < 0 represents trials where the bottom part was presented first). The linear model can be expressed as:

$$\hat{r}(t,\mathrm{SOA}) = \begin{cases} c_B(\mathrm{SOA})r_B(t) + c_T(\mathrm{SOA})r_T(t-\mathrm{SOA}) & \mathrm{SOA} \le 0 \\ c_B(\mathrm{SOA})r_B(t+\mathrm{SOA}) + C_T(\mathrm{SOA})r_T(t) & \mathrm{SOA} > 0 \end{cases}. \quad (1)$$

where the parameters $c_B(\mathrm{SOA})$ and $c_T(\mathrm{SOA})$ govern the weights of the linear combination of single-part responses at each SOA. In its canonical form, c_B and c_T do not change with SOA. In the general form, we allowed for one coefficient at each SOA. To focus on the most critical time window for visual responses and interactions between responses to image parts, the model was fit and evaluated over the range $50 \le t \le 350$ ms. Parameters were optimized using the method of least squares. We restricted the evaluation of the model to $17 \le |\mathrm{SOA}| \le 50$ ms because independent responses to the two parts were apparent at SOA $\ge$ 100 ms.

We also considered a second model that, rather than predicting responses as linear combinations of responses to parts, evaluated the similarity of responses at longer SOA values to the responses at the shortest SOAs. With this model, we aimed to capture invariances across SOA values regardless of how the responses were generated, a complementary goal to the generative predictions of the first model. Since this model was based upon responses at ±17-ms SOA, it was defined only at SOAs of ± 33 and ± 50 ms. In all cases where we compared this model and the linear model, the linear model was recalculated based only on SOAs of ±33 ms and ±50 ms. Let $r_{TB}(t,\mathrm{SOA})$ denote the response at time t (averaged across trials) to presentation of both parts, with the top part appearing SOA ms before the bottom part. Then, maintaining the other conventions of the previous model, this model can be described as:

$$\hat{r}(t,\mathrm{SOA}) = \begin{cases} k(\mathrm{SOA})r_{TB}(t,-17) & \mathrm{SOA} < -17 \\ k(\mathrm{SOA})r_{TB}(t,17) & \mathrm{SOA} > 17 \end{cases}. \quad (2)$$

This model has a single parameter at each SOA, k(SOA), that governs how the 17-ms SOA response is scaled to fit the response at longer SOAs. In its canonical form k depends only on the sign of the SOA, yielding a model with two parameters. In its general form, there is one coefficient k for each SOA.

Model evaluation. To evaluate how well a model described the data for a given SOA, we computed the RMSD between the model and the mean of the actual responses:

$$\mathrm{RMSD}(\mathrm{SOA}) = \sqrt{\frac{\sum_{t=50}^{350}\left(r_{TB}(t,\mathrm{SOA}) - \hat{r}(t,\mathrm{SOA})\right)^2}{(350-50)s/1,000}}, \quad (3)$$

where s is the sampling rate in samples per second and the sum in the numerator includes all values of t from 50 to 350 ms. RMSD(SOA) takes the value of 0 for a perfect fit and is bounded above by the variation in r_{TB} over time.

The criteria for rejecting a model were determined by comparing the difference between the model fit and the data (evaluated by RMSD) against the trial-to-trial variability observed at the electrode in question. To make this comparison, we estimated the distribution of RMSDs between the averages of complementary subsets of responses to the whole image. Let $r_{i,TB}(t,0)$ indicate the response at time t in presentation i (i = 1,...,N) of the whole image; note that $r_{i,TB}(t,0) = \frac{1}{N}\sum_{i=1}^{N} r_{i,TB}(t,0)$. We consider two nonoverlapping equal sized random partitions of the N trials: π_1 and π_2 ($\pi_1 \cup \pi_2 = \{1,...,N\}, \pi_1 \cap \pi_2 = \varnothing$ and $|\pi_1| = |\pi_2|$). We define the average response over each partition ${}_1r_{i,TB}(t,0) = \frac{1}{|\pi_1|}\sum_{i\in\pi_1} r_{i,TB}(t,0)$ and ${}_2r_{i,TB}(t,0) = \frac{1}{|\pi_2|}\sum_{i\in\pi_2} r_{i,TB}(t,0)$ and the RMSD between those two mean responses, $\mathrm{RMSD}_{whole} = \sqrt{\frac{\sum_{t=50}^{350}({}_1r_{TB}(t,0) - {}_2r_{TB}(t,0))^2}{(350-50)s/1000}}$. Because there were twice as many trials in the whole condition as in other conditions, after partitioning the whole trials into two groups, the number of trials was comparable to those in the asynchronous conditions; therefore, the reliability of RMSD_{whole} is comparable to that of RMSD(SOA).

This procedure was repeated 5,000 times (indexed by j) to generate a sampling distribution D from $\{\mathrm{RMSD}_{whole,j}\}$. We used D as a noise distribution, to indicate what sort of RMSD values one might expect when making comparisons between means of identically generated individual waveforms. We compared the value of RMSD(SOA) for a given model against the distribution D. Let p_{SOA} be the percentile of RMSD(SOA) with respect to D:

$$p_{SOA} = \frac{\left|\left\{\mathrm{RMSD}_{whole,j} | \mathrm{RMSD}_{whole,j} \le \mathrm{RMSD}(\mathrm{SOA})\right\}\right|}{5,000} \quad (4)$$

where $|\,|$ denotes the cardinality of the set and $0 \le p_{SOA} \le 1$. A large value of p_{SOA} indicates a poor model fit. For example, $p_{SOA} = 0.95$ indicates that the model in question would generate data as different (or more different) than those observed only 5% of the time. Similarly, $p_{SOA} = 0.5$ indicates that the difference between the model and the actual data is as small as the median difference between different partitions of the responses to the whole image; in other words, the model fits the data approximately as well as can be expected given trial-to-trial variability. We then obtained the probability p that the model in question could have generated data as bad as, or worse than, those seen at even the worst-case SOA:

$$p = \min_{\mathrm{SOA}}\left(1 - p_{SOA}\right). \quad (5)$$

A model was rejected if this value of p was <0.05 and accepted otherwise. In other words, a model's effectiveness was evaluated based on its worst performance across all SOAs. To prevent bias,

these probabilities were only compared for models fit and evaluated on the same set of SOA values. Note also that parameters were fit based on minimization of overall squared difference across all considered SOA values but evaluated based on the difference at the worst-fit SOA value.

We used a similar formulation to ensure that neither individual part alone could describe the responses to the combined image parts. We evaluated a version of the linear model in which c_B and c_T were fixed at 0 and 1, respectively (to test for dominance of the top part), or vice versa (to test for dominance of the bottom part).

The presentation of different parts in asynchronous fashion generates spatiotemporal features that are absent when the parts are presented simultaneously. These features may lead to electrophysiological responses and percepts that differ between the two presentation orders irrespective of the part shapes. To control for such shape-independent spatiotemporal features, we also evaluated the differences between a given response and the response where the two other parts were presented in the same order and with the same SOA. For example, the response to whole image w6 (Fig. 1*B*) obtained when part t3 was followed 17 ms later by part b3 was compared with the response to whole image w7 obtained when part t4 was followed 17 ms later by part b4. By comparing these responses with respect to the sampling distribution, we evaluated the degree to which temporal order alone was sufficient to drive a response irrespective of shape information.

Akaike Information Criterion. In cases where we compared models with different numbers of parameters, we calculated the corrected Akaike Information Criterion (AICc) (Akaike 1974; Burnham and Anderson 2002). Differences in this value describe how much more likely one model is than another to minimize the information lost due to replacing the data with the models. When calculating AICc, we estimated a model's likelihood as $\prod_{SOA}(1 - p_{SOA})$. We set a threshold of 4 for deeming one model to be significantly better than another, which corresponds to a likelihood ratio of ~7.4.

Order tuning index. Often, responses appeared to change when part order changed, but remained relatively constant in the face of other manipulations of SOA (see text). To quantify this observation, we computed a symmetric comparison matrix (see Fig. 4) that described the differences in the responses for any two SOA values, SOA_1 and SOA_2:

$$\mathrm{RMSD}(\mathrm{SOA}_1, \mathrm{SOA}_2) = \sqrt{\frac{\sum_{t=50}^{350} \left(r_{TB}(t, \mathrm{SOA}_1) - r_{TB}(t, \mathrm{SOA}_2)\right)^2}{(350 - 50)s/1000}}. \quad (6)$$

Note that when (and only when) averaging across responses to plot data (see Fig. 4, *D* and *F*), we normalized these values for each response by dividing them by the median of the distribution *D* of RMSD_{whole} (see above) for that response. We then defined the across-order average difference,

$$\mathrm{RMSD}_{across} = \frac{1}{4} \sum_{\substack{|\mathrm{SOA}_1| \le 33 \\ \mathrm{SOA}_1 \neq 0}} \sum_{\substack{\mathrm{sign}(\mathrm{SOA}_1) \neq \mathrm{sign}(\mathrm{SOA}_2) \\ |\mathrm{SOA}_2 \le 33| \\ \mathrm{SOA}_2 \neq 0 \\ \mathrm{SOA}_1 \neq \mathrm{SOA}_2}} \mathrm{RMSD}(\mathrm{SOA}_1, \mathrm{SOA}_2), \quad (7)$$

the within-order average difference,

$$\mathrm{RMSD}_{within} = \frac{1}{12} \sum_{\substack{|\mathrm{SOA}_1| \le 100 \\ \mathrm{SOA}_1 \neq 0}} \sum_{\substack{\mathrm{sign}(\mathrm{SOA}_1) = \mathrm{sign}(\mathrm{SOA}_2) \\ |\mathrm{SOA}_2 \le 100| \\ \mathrm{SOA}_2 \neq 0 \\ \mathrm{SOA}_1 \neq \mathrm{SOA}_2}} \mathrm{RMSD}(\mathrm{SOA}_1, \mathrm{SOA}_2), \quad (8)$$

and defined the order tuning index (OTI):

$$\mathrm{OTI} = \frac{\mathrm{RMSD}_{across} - \mathrm{RMSD}_{within}}{\mathrm{RMSD}_{across} + \mathrm{RMSD}_{within}}. \quad (9)$$

OTI takes a value >0 when the differences across presentation orders are larger than the differences within the same order and is negative when the differences within orders are larger (OTI ranges from −1 to 1). We used a smaller range of SOAs for the across conditions because the absolute difference between SOA values are larger at a given SOA than they are for the corresponding within condition. This does not actually reflect a larger SOA but does reflect that the temporal positions of the top and bottom parts have been reversed. Using this type of asymmetrical limit ensures that any bias in OTI is towards a negative value (i.e., against differences between orders). Using the whole range of SOAs would lead to larger OTI values (i.e., stronger order effects).

We also repeated all calculations leading up to OTI, including determinations of visual responsiveness and order sensitivity, using a temporal window from 50 to 200 ms after stimulus onset.

Response Latency

We calculated the latency of response to each whole image, as a proxy for hierarchical position in the visual processing pathway. We considered each local peak or trough to be a response if the data in a 20-ms window around the local extremum were significantly different from zero (*t*-test, $P < 0.01$). The latency to response was then taken as the last time before the first such extremum at which the mean signal reached half the value at the extremum.

Psychophysics Tasks

We conducted a variation of the main task to evaluate whether it is possible to detect stimulus asynchrony under the same stimulus presentation conditions. In this psychophysics experiment, 16 healthy subjects (10 female) were asked to indicate whether the two parts were presented simultaneously or asynchronously (2-alternative forced-choice task). Eye position was tracked using an Eyelink 1000 Remote system, and trials were initiated by a 500-ms fixation within 3° of the fixation point at the center of the screen. Each subject completed 5 blocks of 48 trials. In half of the trials, the two parts were presented simultaneously, and in the rest the two parts were presented with 17, 33, 50, or 100 ms SOA (we left out longer SOAs because pilot testing indicated performance was at ceiling).

While subjects might tend to see very short asynchronies as simultaneous, it might nevertheless be possible to determine in which order the parts were presented. We therefore ran a second psychophysics test with the same subjects and stimuli; the task was to indicate whether the top part or the bottom part came first. Only asynchronous stimuli were used, with SOAs of 17, 33, 50, or 100 ms, and there were 5 blocks with 50 trials each, counterbalanced between conditions. Note that both of these behavioral experiments differed from the task performed by subjects during the electrophysiological recordings, which was to identify the image.

RESULTS

Subjects identified which of four possible images was shown in a task where the two constituent image parts were briefly flashed (17 ms) with a SOA ranging from 0 ms (whole images) to 700 ms (Fig. 1; see Supplemental Movie S1; Supplemental Material for this article is available online at the Journal website). Subjects could identify the images in spite of the rapid presentation and intervening frames of visual noise (mean performance = 76% correct; chance = 25%); performance did not depend on SOA for any subject (χ^2-test, $P > 0.1$ uncorrected). We further broke down the error types to exam-

ine whether subjects systematically made errors with respect to part order or location more often than chance. Population-wide, there was a slight tendency to give incorrect answers that matched the bottom part (but not the top) more than the reverse (error rates of 7 vs. 8%, respectively; $P = 0.01$, χ^2-test). No such differences between first and second part were observed ($P = 0.66$, χ^2-test). No subject showed any difference in error types with respect to order or position as a function of SOA (χ^2-test, all $P > 0.07$ uncorrected).

We recorded physiological responses from 628 subdural and depth electrodes in 5 subjects during the task. The responses at an example electrode located in the left fusiform gyrus are shown in Fig. 2. Consistent with previous neuroimaging work (Haxby et al. 2001; Grill-Spector and Malach 2004) and macaque (Richmond et al. 1990; Rolls 1991; Logothetis and Sheinberg 1996; Keysers et al. 2001; Hung et al. 2005; Kiani et al. 2005; Connor et al. 2007) and human (Allison et al. 1999; Privman et al. 2011; Liu et al. 2009) neurophysiological recordings, presentation of each image part (Fig. 2*B*) or the whole image (Fig. 2*C*) elicited a strong selective response commencing ~100 ms after stimulus onset. For long SOAs (e.g., 250 ms; Fig. 2, *D5* and *E5*), the sequential parts elicited distinct responses that were largely independent. As the SOA became shorter, these responses to the two parts overlapped and revealed evidence of interactions (Fig. 2, *D1–4* and *E1–4*). Responses to asynchronous presentations were different from those to the whole two-part image (permutation test, $P = 0.0001$), suggesting that both spatial and temporal features in the stimulus influence the electrode's responses. Moreover, there was a striking dependency on the temporal order with which the two parts were presented (compare Fig. 2, *D* vs. *E*). At long SOAs, dependence on temporal order is trivial given the independence of the responses to each image part. Yet, such temporal order dependencies were evident even at the shortest SOA that we tested (17 ms, permutation test, $P = 0.002$, Fig. 2, *D1* vs. *E1*).

We repeated the analyses by examining the responses in the high-gamma-frequency band (65 to 100 Hz; see MATERIALS AND METHODS). Figure 3 demonstrates the responses in the high-gamma band of an example electrode located in left fusiform gyrus. These high-gamma-band responses also showed the key characteristics demonstrated for the broadband responses in Fig. 2. All in all, the high-gamma-band responses tended to be noisier and less strongly driven by visual inputs (compare Table 1 with Table 2), and we therefore focus on the broadband signals in the following analyses.

The influence of stimulus presentation dynamics on the physiological responses could reflect spatiotemporal features sensitive to input timing and to presentation order. To evaluate the relative influence of SOA and presentation order, we constructed in Fig. 4*A* a matrix of pairwise RMSD (*Eq. 6*) between all the responses shown in Fig. 2, *C–E*. This comparison matrix showed smaller differences between conditions in which the two parts were presented in the same order than when they were presented in opposite orders (see Fig. 4, *B* and *C*, for other examples). To quantify this observation, we calculated an OTI (*Eq. 9*). The OTI ranged from −1 (differences between orders negligible compared with SOA-dependent differences within orders) to 0 (differences within orders as large as differences between orders) to 1 (differences within orders negligible compared with differences between orders). The OTI for the example electrode in Fig. 1 was 0.37. There were 692 responses from 173 electrodes located in ventral visual regions (692 = 173 electrodes × 4 two-part images;

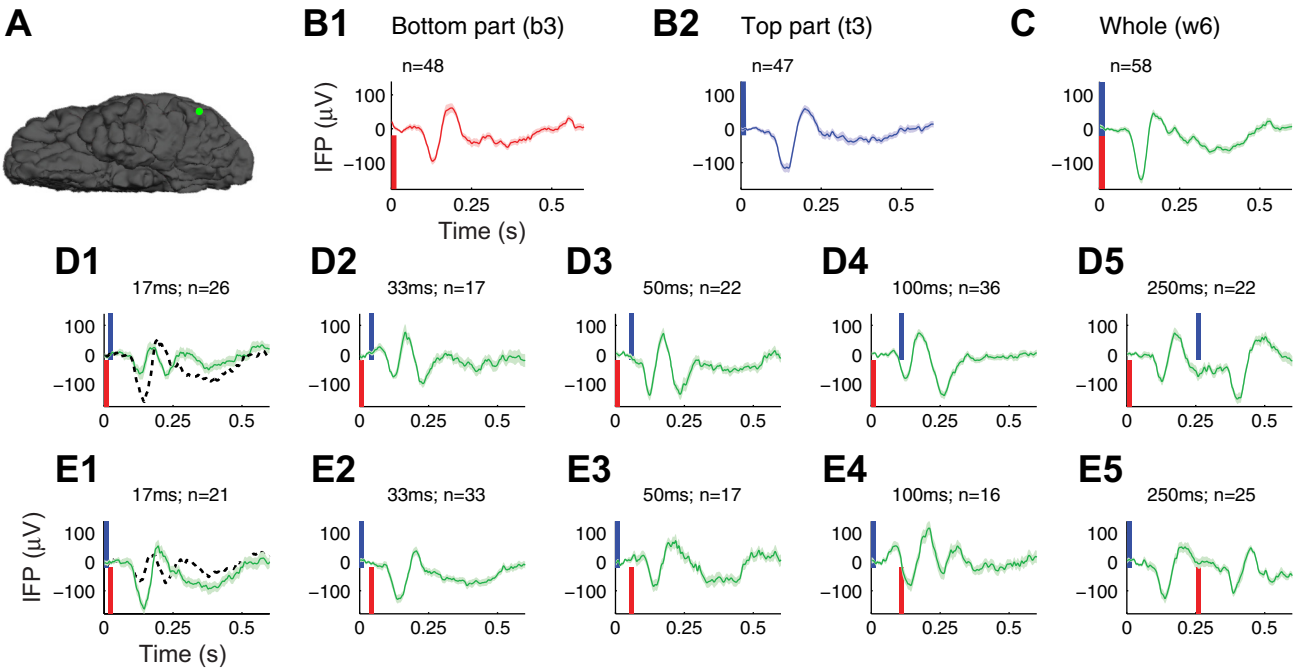

Fig. 2. Example physiological responses, broadband signals. *A*: example electrode located in the left fusiform gyrus. *B*: intracranial field potential (IFP) responses to the 2 parts from Fig. 1*A* (b3, t3) when presented independently (red = b3, blue = t3), aligned to image part onset and averaged over 48 and 47 repetitions respectively. Shaded areas denote SE. *C*: responses to the whole image (w6). *D* and *E*: responses at increasing SOA values, with the bottom part shown first (*D*) or last (*E*). Responses are aligned to the first part onset. The red (blue) rectangle denotes the onset and duration of the bottom (top) part. The dotted line in *D1* shows the response in the opposite-order condition (*E1*) for comparison purposes, and the dotted line in *E1* shows the response from *D1*. Even at the shortest nonzero SOA (17 ms), responses differed depending on part order (cf. *D1* vs. *E1*, $P = 0.002$, permutation test).

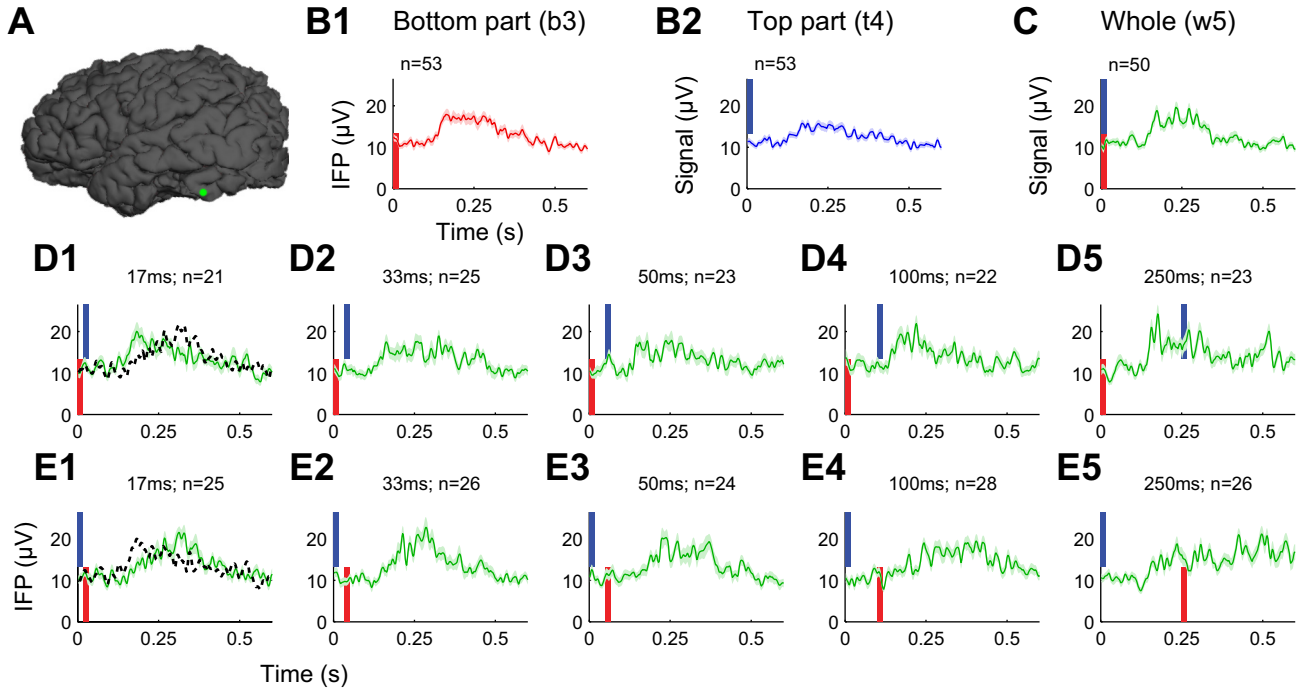

Fig. 3. Example physiological responses, high-gamma band. *A*: example electrode located in the left fusiform gyrus. *B*: high-gamma (65–100 Hz)-power time course in responses to 2 parts when presented independently (red = part b3, blue = part t4; Fig. 1*B*), aligned to image part onset and averaged over 53 repetitions. Shaded areas denote SE. *C*: high-gamma-band responses to the whole image (w5). *D* and *E*: high-gamma-band responses at increasing SOA values, with the bottom part shown first (*D*) or last (*E*). Responses are aligned to the first part onset. The red (blue) rectangle denotes the onset and duration of the bottom (top) part. The dotted line in *D1* shows the response in the opposite-order condition (*E1*) for comparison purposes, and the dotted line in *E1* shows the response from *D1*. Even at the shortest nonzero SOA (17 ms), responses differed depending on part order (cf. *D1* vs. *E1*, $P = 0.001$, permutation test).

Table 1). We calculated OTI values for the 221 of these 692 responses that were visually responsive to the whole image (e.g., Fig. 2*C*) and to both constituent halves (e.g., Fig. 2*B*; see MATERIALS AND METHODS). The average comparison matrix revealed a clear asymmetry depending on whether differences were computed across vs. within orders (Fig. 4*D*). Almost all OTI values were positive (means ± SD = 0.27 ± 0.12; Fig. 4*E*). To minimize the possibility that nonvisual signals could influence the results, we repeated the OTI computations considering only data between 50 and 200 ms after stimulus onset (Fig. 4, *F*–*G*). Within this window, the OTI values were even more strongly positive (OTI = 0.39 ± 0.14). The positive order tuning indexes demonstrate that these visual responses were more sensitive to temporal disruptions that reversed the order of part presentations than to disruptions that preserved order.

We sought to quantitatively describe the dynamic interactions between asynchronous halves that gave rise to this order sensitivity by considering whether a linear model based on the responses to individual parts could account for the observed responses (see MATERIALS AND METHODS). To avoid cases where such a model could trivially explain the data, we restricted the analyses to those SOAs that revealed the strongest interactions between parts (17, 33, and 50 ms) and we focused on 111 responses at 54 electrodes showing significant temporal order sensitivity at 17-ms SOA (permutation test, $P < 0.05$; e.g., Fig. 2, *D1* and *E1*). The responses that did not show such temporal order sensitivity are described in Fig. 5. The model showed even better performance in explaining the order-insensitive responses (Fig. 5) compared with the order sensitive responses (discussed below, see Fig. 7). We focus the rest of the article on describing the responses that are more challenging to explain, namely those that showed order sensitivity.

We considered a weighted linear combination of both halves' responses (Baker et al. 2002; Zoccolan et al. 2007; Agam et al. 2010; Baeck et al. 2013), shifted to reflect SOA (*Eq. 1*, with one coefficient for each part, the "simple linear model"). Figure 6, *A*–*D*, shows an example of an electrode located in the left fusiform gyrus where the data were well fit by this model ($P = 0.21$; compare orange traces vs. green traces) but not by either of the individual image halves.

The simple linear model provided a good description of the data for 44% of the responses (see MATERIALS AND METHODS). There were electrodes where the interactions between image halves could not be explained by this model (e.g., Fig. 6, *E*–*H*). Inspired by the temporal sensitivity documented in Figs. 2 and 4, we considered a general linear model endowed with the flexibility to reflect relative timing by taking into account the SOA and presentation order (*Eq. 1*, with different coefficients at each SOA). Incorporating relative timing allowed the general linear model to describe 62% of the responses, an increase of 41% over the simple linear model. The increased explanatory power was not merely due to the addition of parameters: the general linear model was found to be a significantly better description of the data for 72% of the responses, and significantly worse for none, according to the AICc (ΔAICc > 4; see MATERIALS AND METHODS).

Further evidence for the importance of relative timing was provided by a second model that evaluated the similarity of waveforms as SOA increased beyond 17 ms (*Eq. 2*, with a single scaling parameter for each order). By fixing coefficients within a temporal ordering, this model described 48% of

Table 2. *Electrode/response properties and counts by brain region, high gamma*

Region	No. Electrodes	No. Responses to Whole	No. Responses to Both Halves	No. Order-Sensitive Responses	No. *Eq. 1*	No. *Eq. 2*
Visual cortex						
Anterior transverse collateral sulcus	3	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)
Fusiform gyrus	24	10 (10%)	10 (100%)	3 (3%)	0 (0%)	1 (10%)
Inferior occipital gyrus/sulcus	25	13 (13%)	13 (100%)	2 (2%)	0 (0%)	2 (15%)
Occipital pole	35	17 (12%)	16 (94%)	8 (6%)	2 (13%)	4 (25%)
Lateral occipitotemporal sulcus	2	2 (25%)	2 (100%)	0 (0%)	0 (0%)	0 (0%)
Middle occipital/lunate sulcus	5	1 (5%)	1 (100%)	0 (0%)	0 (0%)	0 (0%)
Middle occipital gyrus	40	12 (8%)	12 (100%)	4 (3%)	2 (17%)	4 (33%)
Inferior temporal gyrus	39	4 (3%)	4 (100%)	3 (2%)	0 (0%)	1 (25%)
Visual cortex totals	173	59	58	20	4	12
Intraparietal sulcus	6	0	0	0		
Parahippocampal gyrus	22	0	0	0		
Superior temporal sulcus	11	0	0	0		
Cuneus	23	0	0	0		
Superior occipital gyrus	7	0	0	0		
Precentral gyrus	9	0	0	0		
Middle temporal gyrus	82	0	0	0		
Supramarginal gyrus	47	0	0	0		
Lingual gyrus	39	2	2	0		
Depth electrodes	26	0	0	0		
Subcentral gyrus/sulcus	14	0	0	0		
Angular gyrus	28	0	0	0		
Lateral superior temporal gyrus	49	0	0	0		
Planum temporale	5	0	0	0		
Postcentral gyrus	14	0	0	0		
Postcentral sulcus	3	0	0	0		
Temporal pole	18	0	0	0		
Short insular gyri	1	0	0	0		
Operculum	6	0	0	0		
Posterior ventral cingulate	1	0	0	0		
Collateral/lingual sulcus	4	0	0	0		
Inferior temporal sulcus	5	0	0	0		
Precuneus	8	0	0	0		
Superior/transverse occipital sulcus	2	0	0	0		
Unclassified	1	0	0	0		
Superior parietal gyrus	4	0	0	0		
Calcarine sulcus	1	0	0	0		
Triangular inferior frontal gyrus	4	0	0	0		
Parieto-occipital sulcus	1	0	0	0		
Middle frontal gyrus	2	0	0	0		
Hippocampus	2	0	0	0		
Totals	618	61	60	20		

Summary of electrode locations and response properties, considering the high-gamma (65–100 Hz)-frequency band. Format and conventions are the same as those in Table 1.

responses (e.g., Fig. 6, *E–H*). Allowing coefficients to vary as a function of SOA (*Eq. 2*, with different scaling parameters at each SOA) did not describe any additional responses. The performance of the short SOA model agrees with the positive OTI values, showing that responses changed relatively little with increasing SOA within the same order. The linear and short SOA models provide different perspectives on the data. The former quantifies our ability to predict responses to asynchronous stimulus presentation based on responses to the constituent parts. The latter quantifies our ability to predict responses by assuming that they change only as a function of temporal order.

Inspired by the performance of the short-SOA model (*Eq. 2*), we considered a variant of the linear model (*Eq. 1*) to evaluate the hypothesis that coefficients might vary as a function of order only. In the most extreme case, this might manifest as the first part entirely dictating the response. We fit a version of *Eq. 1* with one coefficient for the first part and one coefficient for the second part. This model accounted for 52% of the responses. The mean first-part coefficient was 0.80 vs. 0.42 for the second part. These coefficients provide evidence that the first part tends to play a more prominent role in shaping the overall response, but the second part is also important.

The presentation of different parts in asynchronous fashion generates spatiotemporal features that are absent when the parts are presented simultaneously. These features may lead to electrophysiological responses that differ between the two

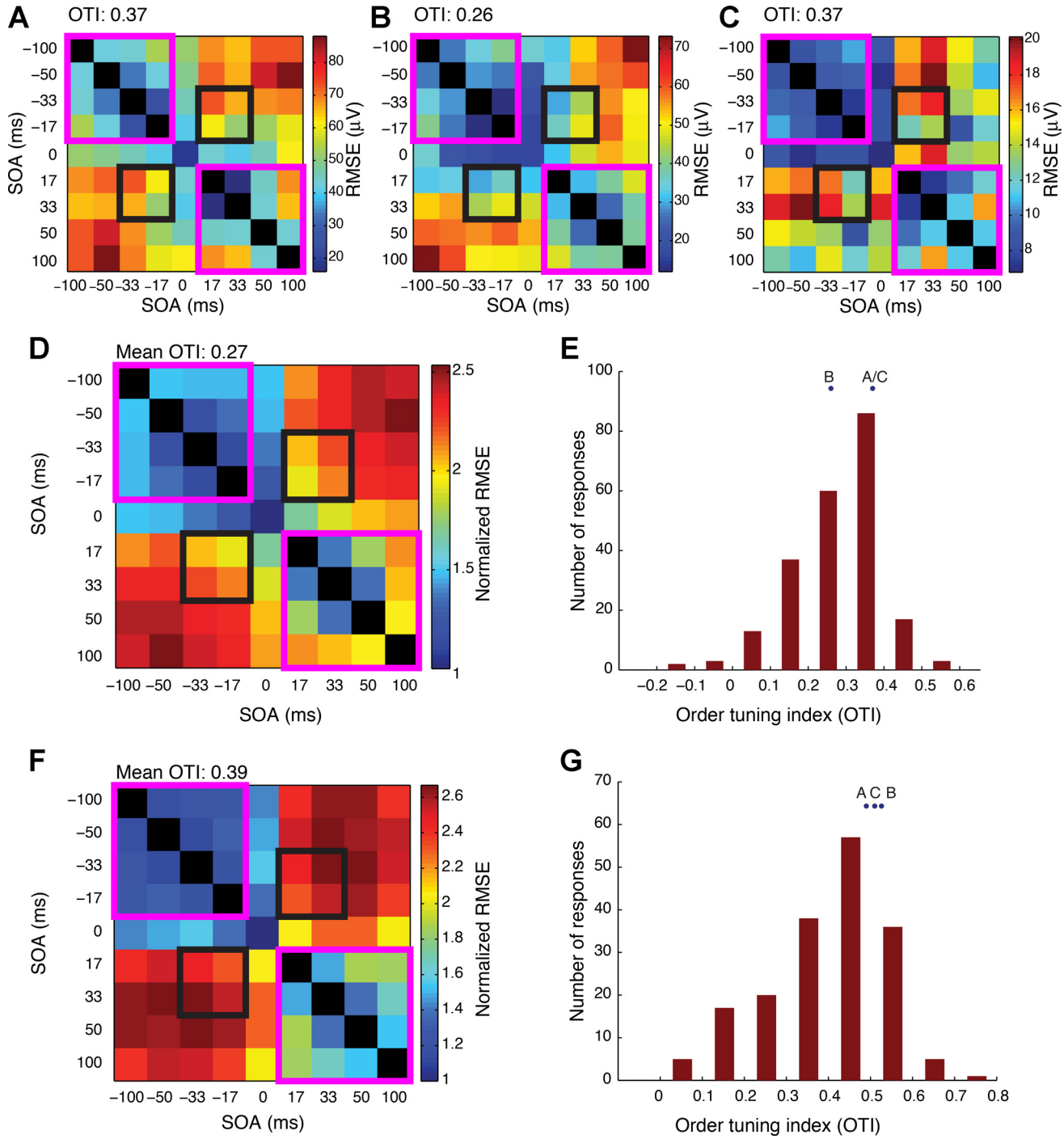

Fig. 4. Comparison of responses across orders and SOAs. *A–C*: for each of 3 example electrodes, we computed a comparison matrix contrasting responses at different SOAs and presentation orders. Entry *i*, *j* in this matrix represents the root-mean-square deviation (RMSE; *Eq. 6*) between mean responses elicited by trials with SOAs given by *i* and *j* (see color scale on the *right*). The root-mean-square deviation (RMSD) was not computed along the diagonal, which is shown as black squares. The order tuning index (OTI) was calculated from the mean difference between responses with different temporal orders (solid black windows) minus the mean difference between responses with the same temporal order (pink windows, *Eq. 9*). Positive OTIs indicate that differences between SOA conditions that preserve order are smaller than those between SOA conditions that do not. *D*: summary comparison matrix contrasting responses at different SOAs and presentation orders ($n = 221$ responses from 78 electrodes, see MATERIALS AND METHODS). Here the RMSD is normalized for each electrode before averaging. *E*: distribution of OTI values. Bin size = 0.1. OTI values for the 3 example electrodes are indicated with circles. *F* and *G*: same analysis as in *D* and *E* except using the interval between 50 and 200 ms after stimulus onset ($n = 179$ visual responses). Mean OTI = 0.39 ± 0.14.

presentation orders, irrespective of the parts' shapes. For example, inhomogeneities between the top and bottom parts of an electrode's receptive field or apparent motion signals could lead to distinct neural signals depending on which part is presented first. To evaluate whether such spatiotemporal features could explain the signals described above, we compared responses to two distinct sets of parts presented under the same order and SOA (for example, comparing part t1 followed by

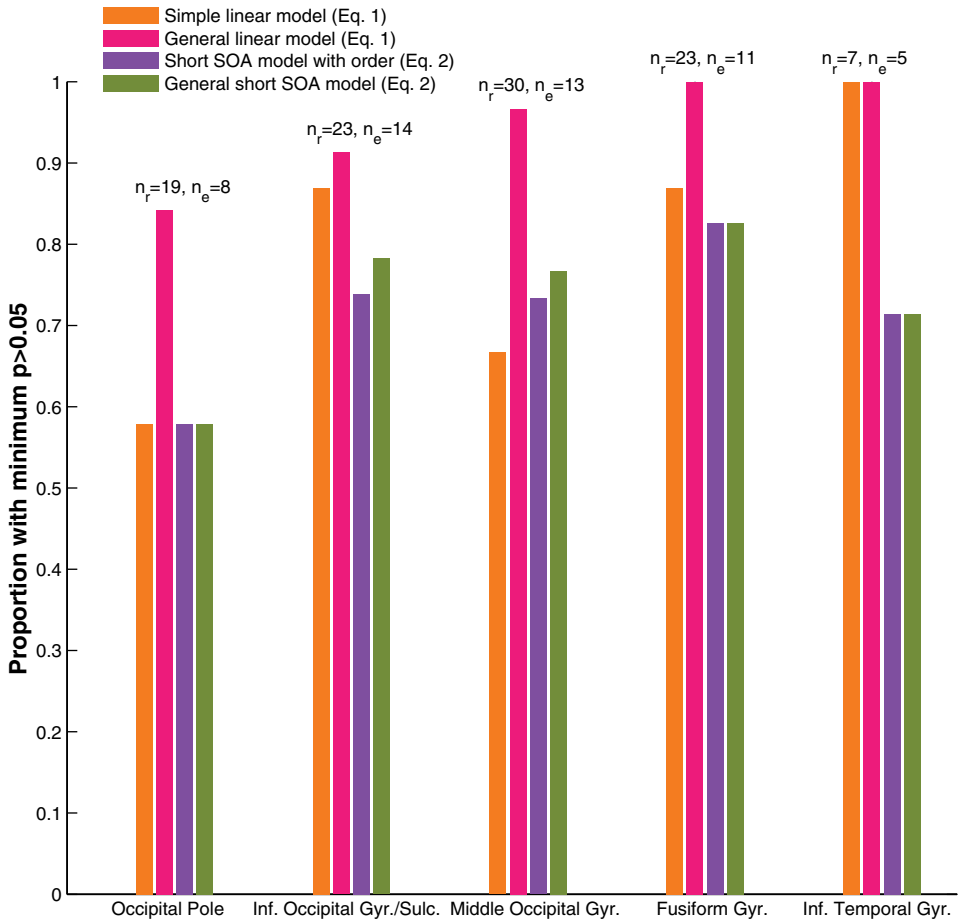

Fig. 5. Comparison among quantitative models for order-insensitive responses. Performance of the 2 different models (*Eqs. 1* and *2*), with 2 variants of each model (see MATERIALS AND METHODS), at the 5 visual surface regions with >5 responses. Here we consider only order-insensitive responses (i.e., responses that did not show a significant difference between the 2 presentation orders at 17 ms). For each region, n_r indicates the number of responses and n_e the number of electrodes considered. For each model, in each region, the proportion of responses characterized with probability $P > 0.05$ is shown. Here the models are compared at SOA= ± 33 and ± 50 ms (and not at SOA= ± 17 ms, because *Eq. 2* is not defined at this SOA). The models are the linear model in *Eq. 1* with 2 free parameters, 1 for each image part (orange), the linear model with 8 free parameters, 1 for each image part at each SOA (magenta), the short SOA model with 2 free parameters, 1 for each order (purple), and the short SOA model with 4 free parameters, 1 for each order and SOA (green).

part b1 vs. part t2 followed by part b2). In all, only 8% of the order-sensitive responses were well matched by these complementary responses. While spatiotemporal features that depend on presentation order may contribute to the responses documented here, the low similarity between responses to different parts with the same spatiotemporal sequence suggests that the responses cannot be purely explained in terms of features that are independent of the parts' shapes.

We parceled each subject's brain into 75 anatomical regions based on preoperative MR and postoperative CT images (see MATERIALS AND METHODS). We focused on the five regions along the ventral visual stream with more than five order-sensitive visual responses: occipital pole, inferior occipital gyrus, inferior temporal gyrus, middle occipital gyrus, and fusiform gyrus. We compared the performance of the simple linear model (*Eq. 1*) vs. the short SOA model (*Eq. 2*), evaluated at 33- and 50-ms SOA. The short SOA model performed significantly worse at the occipital pole (Fig. 7; χ^2-test, $P = 0.02$), but no such differences were observed in the four higher visual areas (χ^2-test, $P > 0.2$). Performance of the short SOA model (*Eq. 2*) varied significantly by region (χ^2-test, $P = 0.02$), explaining 31% of responses at the occipital pole and 83% of responses in inferior temporal gyrus. By contrast, performance of the part-based linear models (*Eq. 1*) remained relatively consistent across regions (χ^2-test, $P > 0.65$).

We also examined whether order sensitivity was uniformly distributed throughout ventral visual cortex. We first performed an ANOVA of OTI across the five regions of ventral visual cortex at which we had more than five electrodes. The distribution deviated significantly from uniform ($P = 0.003$), increasing from 0.31 at the occipital pole to 0.4 in the inferior temporal gyrus. Additionally, there was a small but significant correlation between response latency (see MATERIALS AND METHODS) and OTI ($\rho = 0.24$, $P = 0.01$). The regional analyses of OTI and response latencies are consistent with the notion that order sensitivity increases as information advances through the visual system.

During the electrophysiology experiments, subjects focused on recognizing the two-part images regardless of the presentation asynchrony. To evaluate whether it is possible to detect asynchronous presentation, we conducted a separate behavioral experiment (without physiological recordings) under identical stimulus presentation conditions but asking subjects to determine whether or not the parts were presented synchronously. The behavioral data suggested that many subjects tended to perceive SOAs of 17 ms indistinguishably from simultaneous trials (Fig. 8*A*), although the population mean discriminability (d') was slightly but significantly greater than zero (mean = 0.44; $P = 0.004$, *t*-test). Extrapolating from these results, it seems plausible that the SOA = 17-ms condition was perceptually quite close to the SOA = 0-ms condition during the physiological recordings. To evaluate whether subjects were able to discriminate the relative order in which the two parts were presented, we conducted a second behavioral task (without physiological recordings) requiring subjects to indicate whether the top part was presented first or second (Fig. 8*B*). While there was considerable variability across subjects, the

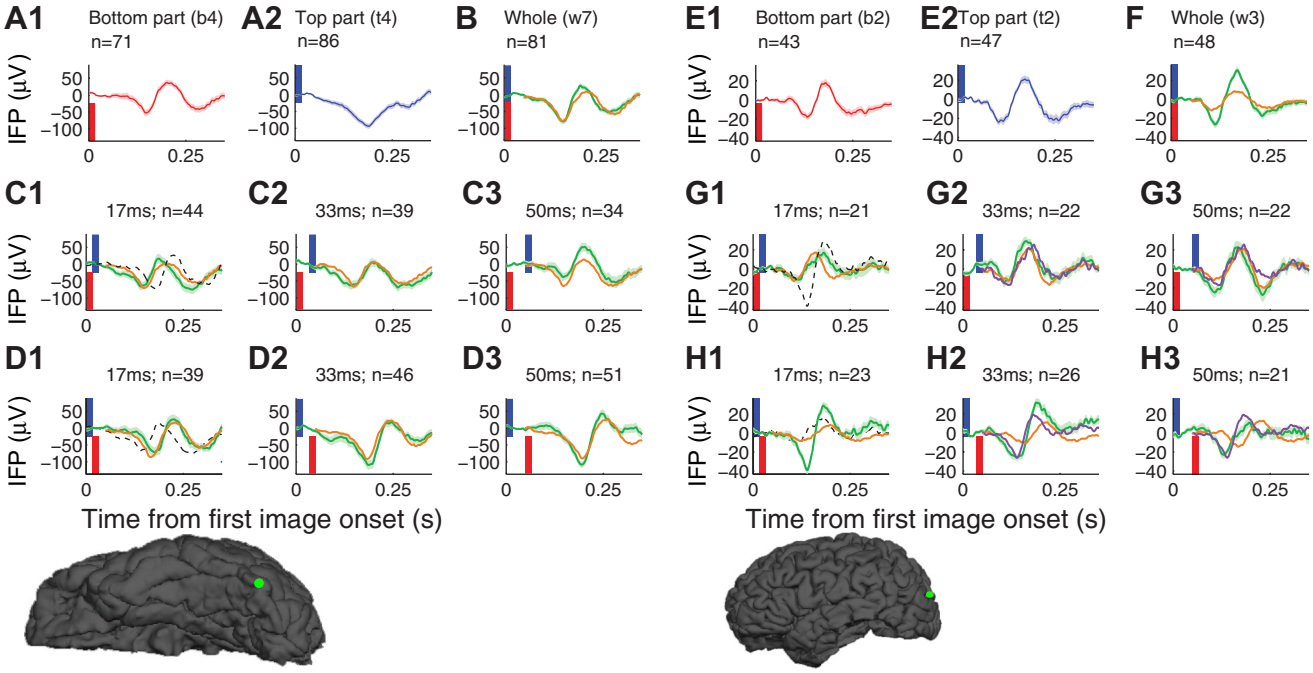

Fig. 6. Examples of quantitative modeling of the physiological responses. *A–D*: responses recorded from an electrode in the left fusiform gyrus upon presentation of the w7 stimulus. Individual panels use the same conventions as Fig. 2, *B–E*. *E–H*: responses recorded from an electrode in the left middle occipital gyrus upon presentation of the w3 stimulus. The orange lines in *C* and *D* and *G* and *H* show fits from a linear model that aims to predict the responses to the images at different SOA levels from the constituent parts (simple linear model, *Eq. 1*; see MATERIALS AND METHODS). This model provided a good fit to the data in *C* and *D* ($P = 0.21$; see MATERIALS AND METHODS) but not in *G* and *H* ($P = 0.007$). Purple lines in *G* and *H* show the fits of a model that compares responses at 33- and 50-ms SOA to responses at 17-ms SOA, with a coefficient that varies depending on order (short-SOA model with order, *Eq. 2*; see MATERIALS AND METHODS). This model described the data well ($P = 0.20$).

population mean d' was greater than zero (means = 0.61, $P = 0.0005$, *t*-test).

DISCUSSION

The current study investigated the neural representation of image parts that were asynchronously presented with intervals ranging from 17 to 250 ms in a task where subjects had to put together the two components to form a whole. Examining intracranial field potentials recorded along the human ventral visual stream, long SOAs of 100–250 ms led to largely independent responses to each separate image part (e.g., Fig. 2, *D5* and *E5*). By contrast, the responses revealed interactions between the parts at shorter asynchrony values. The neural signals reflecting the integration of the two parts were sensitive to the asynchrony of stimulus presentation, even at an SOA as short as 17 ms (e.g., compare Fig. 2, *C* vs. *D1*). Moreover, which part was presented first strongly influenced the ensuing response (e.g., compare Fig. 2, *D1* vs. *E1*). Similar observations were made when considering only the high-gamma-power time course (Wittingstall and Logothetis 2009; Privmann et al. 2011).

Spatial context is known to modulate responses throughout visual cortex from early visual areas (Zipser et al. 1996; Bair et al. 2003; Angelucci and Bressloff 2006; Allman et al. 1985) all the way to intermediate and higher visual areas (Missal et al. 1999; Zoccolan et al. 2007; Reynolds et al. 1999; Chelazzi et al. 1993; Agam et al. 2010; Baeck et al. 2013). In early visual cortex, adding visual stimulation in the surround of a neuron's receptive field typically inhibits the responses to the center stimulus (Sceniak et al. 1999; Allman et al. 1985). Additionally, recordings in visual areas V4 and IT have shown that presenting two or more stimuli (sometimes referred to as "clutter" in the context of visual search tasks) within a neuron's receptive field typically leads to a reduction in the firing rate to the preferred stimulus (Missal et al. 1999; Zoccolan et al. 2007; Reynolds et al. 1999; Chelazzi et al. 1993). In the absence of attention, the responses to object pairs have been successfully described by linear weighted models similar to *Eq. 3* from the current article (with SOA = 0) (Reynolds et al. 1999; Baeck et al. 2013; Zoccolan et al. 2007). The success of these linear weighted models in describing many (but not all, Fig. 5) of the responses in the current study extends the notion of linear combinations to the time domain.

A recent study from our laboratory showed that the weighted sums were biased towards the response to the preferred objects in situations where two random objects were disconnected, independent, simultaneously presented and there was no behavioral integration required during the task (Agam et al. 2010), similar to the studies of Zoccolan et al. (2007) and Gawne and Martin (2002) and one of the conditions in Baeck et al. (2013). Here we report activity from many electrodes along the ventral visual stream in which the response to a pair of asynchronously presented object parts could not be described by the response to either part alone. These observations suggest that the distance between parts and/or the task's requirement to incorporate information from the two parts may play an important role in shaping the responses to spatial context along the ventral visual stream.

In addition to spatial context, temporal context can also influence physiological responses in early visual areas (Nelson

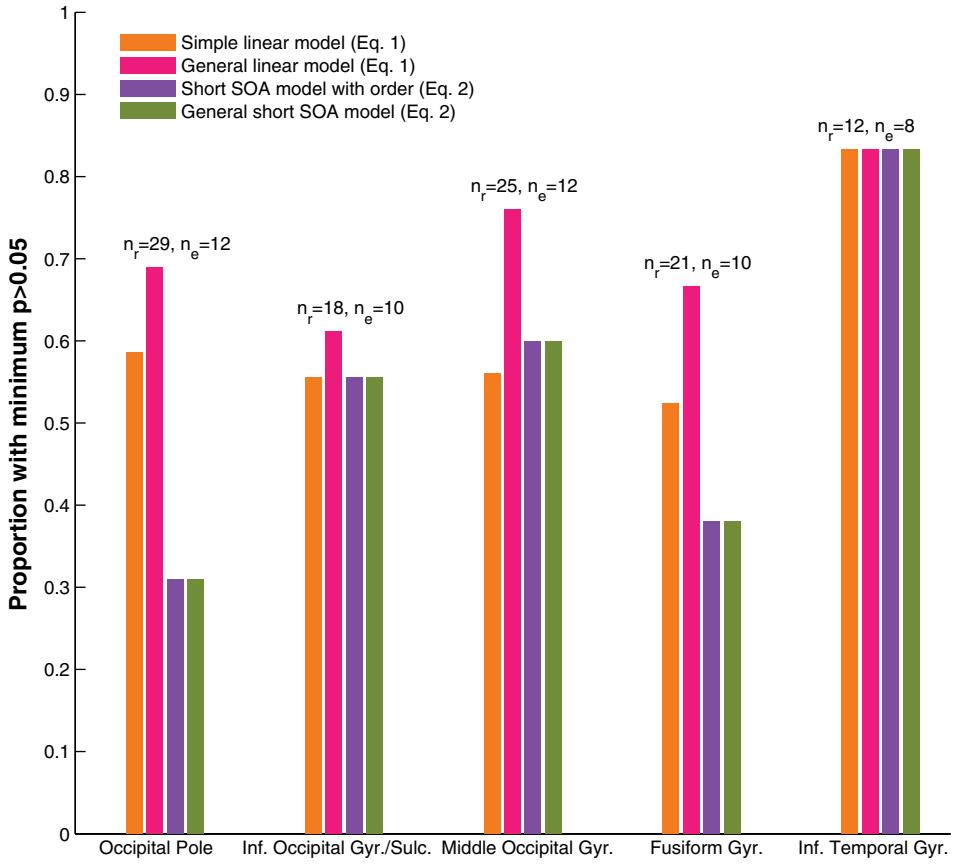

Fig. 7. Comparison of all models for order sensitive responses. Performance of the 2 models, both variants, at the 5 visual regions with >5 responses, considering only order-sensitive responses (see MATERIALS AND METHODS for description of each model). Conventions are otherwise the same as in Fig. 5.

1991; Vinje and Gallant 2002; Bair et al. 2003; Benucci et al. 2009). Less is known about how temporal context modulates neurophysiological signals in higher visual areas. Neuroimaging signals have shown dependence on stimulus temporal order over scales of hundreds of milliseconds to seconds in higher visual areas (Hasson et al. 2008), and several studies have documented how recognition at the behavioral level can be influenced by asynchronous presentation of visual stimuli (Clifford et al. 2004; Eagleman et al. 2004; Singer and Sheinberg 2006; Anaki et al. 2007; Singer and Kreiman 2014). The current study demonstrates that temporal context can significantly affect physiological responses throughout the human ventral visual stream, even within as little as 17 ms. In some cases, the effects of temporal context could be accounted for by weighted linear models that incorporated timing differences (*Eq. 1*; Figs. 5 and 6, *A–D*). However, many other responses

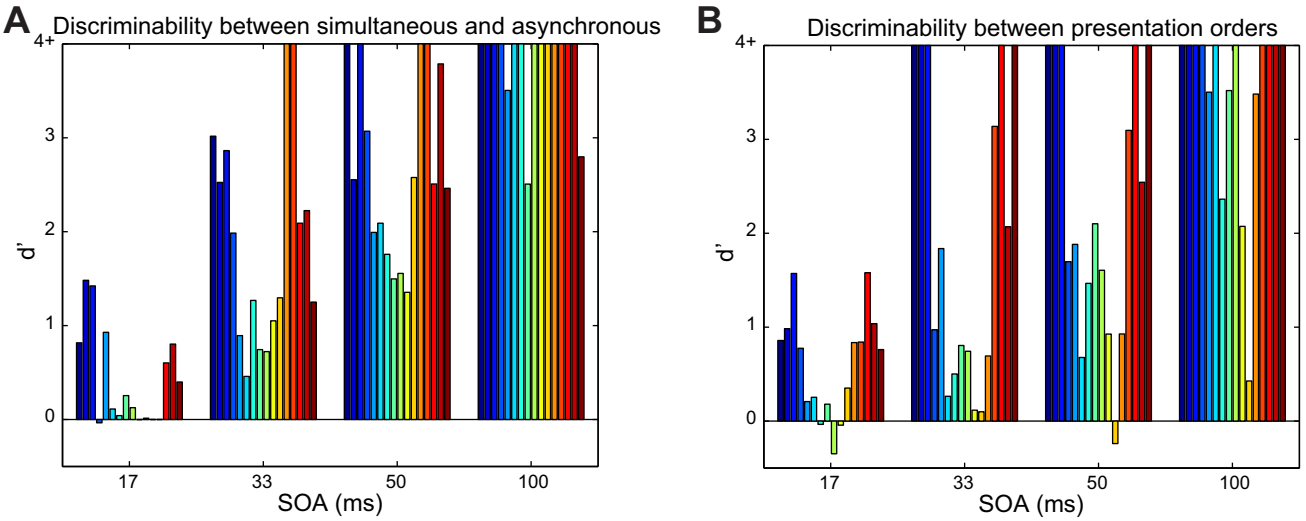

Fig. 8. Results of 2 psychophysics experiments. *A*: stimuli and presentation parameters were identical to those used in the main experiment, but subjects were asked to indicate whether or not the 2 parts were presented simultaneously. Each color denotes the discriminability (d') between simultaneous trials and trials of the indicated SOA for a different subject. *B*: in this task, subjects were asked to indicate which of the 2 asynchronously presented parts was presented first. There were no physiological recordings during these variants of the experiment.

manifested a strong sensitivity to the order in which the stimuli were presented (Figs. 2, 5, and 6, *E–H*) that could not be accounted for by the time-shifted linear models.

During the physiological recordings, subjects were asked to identify the two-part objects. We conducted two additional psychophysics experiments to evaluate whether it is possible to discern whether the parts were presented synchronously or not (Fig. 8*A*) and which part was presented first (Fig. 8*B*). At long SOAs, subjects could reliably report that the stimuli were asynchronous and in which order they appeared. This is consistent with the large differences and independent part responses observed electrophysiologically at long SOA. At the shortest and most difficult SOA value 17 ms, subjects were quite poor at discerning simultaneity and order, although population-wide performance was above chance in both questions. Yet, we emphasize that during the physiological recordings, subjects were performing a recognition task and probably were not deliberately evaluating synchrony or order.

We considered whether the current results could be explained in terms of differential eye movements, adaptation effects, or masking effects. Given that stimuli and asynchrony values were presented in random order and with short asynchrony, it seems difficult to explain the results based on differential eye movements between the whole condition and the 17-ms SOA conditions or between the two 17-ms asynchrony conditions with different orders. Even if subjects were to make distinct saccades triggered and dictated by the stimulus order, it is unclear whether such saccades could be rapid enough to explain differences in the physiological responses during the initial 200 ms. The sensitivity to small asynchrony intervals also argues against an interpretation of the data based on adaptation over long time scales. Short intervals are typical in visual masking studies. A mask presented within tens of milliseconds of a visual stimulus (either before or after) can strongly influence neurophysiological responses as well as visual perception (Kovacs et al. 1995; Macknik 2006; Felsten and Wasserman 1980; Rolls et al. 1999). It seems unlikely that the effects documented here can be ascribed to masking given that *1*) there is no clear masking effect at the behavioral level (Fig. 8); *2*) in most visual masking studies, the mask spatially overlaps the primary stimulus; and *3*) the compatible content and adjacent peri-foveal organization of the images argue against paracontrast or metacontrast masking (Alpern 1953).

The origin of the sensitivity to relative timing reported here is not clear. Distinct spatiotemporal features depending on stimulus order elicited distinct responses even as early as the occipital pole. Consistent with potential early origins, prior work has found that neurons in primary visual cortex show modulation in their responses when stimuli inside or outside their receptive fields are presented in temporal proximity (Nelson 1991; Bair et al. 2003; Benucci et al. 2009). The dependence on relative stimulus timing described in the current study is particularly intriguing in light of theoretical and experimental studies proposing that a robust and efficient representation of information can be encoded in the temporal order with which neurons fire action potentials (Hopfield 1995; vanRullen and Thorpe 2002; Gollisch and Meister 2008) The current observations suggest that spatiotemporal interactions persist even at the highest levels of visual processing within the ventral pathway and with dynamics on the scale of tens of milliseconds.

ACKNOWLEDGMENTS

We thank Roozbeh Kiani, Hanlin Tang, Radhika Madhavan, Kendra Burbank, John Maunsell, and Chou Hung for comments on the manuscript; Morgan King for the stimuli; the patients for participation in this study; and Sheryl Manganaro, Lixia Gao, and Hanlin Tang for facilitating data collection.

GRANTS

Financial support for this study was provided by the National Institutes of Health and National Science Foundation.

DISCLOSURES

No conflicts of interest, financial or otherwise, are declared by the author(s).

AUTHOR CONTRIBUTIONS

Author contributions: J.M.S. and G.K. conception and design of research; J.M.S., J.R.M., and W.S.A. performed experiments; J.M.S. analyzed data; J.M.S. and G.K. interpreted results of experiments; J.M.S. prepared figures; J.M.S. drafted manuscript; J.M.S. and G.K. edited and revised manuscript; J.M.S., J.R.M., W.S.A., and G.K. approved final version of manuscript.

REFERENCES

Agam Y, Liu H, Pappanastassiou A, Buia C, Golby AJ, Madsen JR, Kreiman G. Robust selectivity to two-object images in human visual cortex. *Curr Biol* 20: 872–879, 2010.

Akaike H. A new look at the statistical model identification. *IEEE Trans Automat Contr* 19: 716–723, 1974.

Allison T, Puce A, Spencer D, McCarthy G. Electrophysiological studies of human face perception. I: Potentials generated in occipitotemporal cortex by face and non-face stimuli. *Cereb Cortex* 9: 415–430, 1999.

Allman J, Miezin F, McGuinness E. Stimulus specific responses from beyond the classical receptive field: neurophysiological mechanisms for local-global comparisons in visual neurons. *Annu Rev Neurosci* 8: 407–430, 1985.

Alpern M. Metacontrast. *J Opt Soc Am* 43: 648–657, 1953.

Anaki D, Boyd J, Moscovitch M. Temporal integration in face perception: evidence of configural processing of temporally separated face parts. *J Exp Psychol Hum Percept Perform* 33: 1, 2007.

Angelucci A, Bressloff PC. Contribution of feedforward, lateral and feedback connections to the classical receptive field center and extra-classical receptive field surround of primate V1 neurons. *Prog Brain Res* 154: 93–120, 2006.

Baeck A, Wagemans J, Op de Beeck HP. The distributed representation of random and meaningful object pairs in human occipitotemporal cortex: the weighted average as a general rule. *Neuroimage* 70: 37–47, 2013.

Bair W, Cavanaugh JR, Movshon JA. Time course and time-distance relationships for surround suppression in macaque V1 neurons. *J Neurosci* 23: 7690–7701, 2003.

Baker CI, Behrmann M, Olson CR. Impact of learning on representation of parts and wholes in monkey inferotemporal cortex. *Nat Neurosci* 5: 1210–1216, 2002.

Benucci A, Ringach DL, Carandini M. Coding of stimulus sequences by population responses in visual cortex. *Nat Neurosci* 12: 1317–1324, 2009.

Burnham KP, Anderson D. *Model Selection and Multi-Model Inference* (2nd ed.). New York: Springer, 2002.

Chelazzi L, Miller EK, Duncan J, Desimone R. A neural basis for visual search in inferior temporal cortex. *Nature* 363: 345–347, 1993.

Clifford CW, Holcombe AO, Pearson J. Rapid global form binding with loss of associated colors. *J Vis* 4: 1090–1101, 2004.

Connor CE, Brincat SL, Pasupathy A. Transformation of shape information in the ventral pathway. *Curr Opin Neurobiol* 17: 140–147, 2007.

De Baene W, Premereur E, Vogels R. Properties of shape tuning of macaque inferior temporal neurons examined using rapid serial visual presentation. *J Neurophysiol* 97: 2900–2916, 2007.

Deco G, Rolls ET. A neurodynamical cortical model of visual attention and invariant object recognition. *Vision Res* 44: 621–642, 2004.

Destrieux C, Fischl B, Dale A, Halgren E. Automatic parcellation of human cortical gyri and sulci using standard anatomical nomenclature. *Neuroimage* 53: 1–15, 2010.

DiCarlo J, Maunsell H. Form representation in monkey inferotemporal cortex is virtually unaltered by free viewing. *Nat Neurosci* 3: 814–821, 2000.

Dumoulin SO, Wandell BA. Population receptive field estimates in human visual cortex. *Neuroimage* 39: 647–660, 2008.

Eagleman DM, Jacobson JE, Sejnowski TJ. Perceived luminance depends on temporal context. *Nature* 428: 854–856, 2004.

Felsten G, Wasserman GS. Visual masking: mechanisms and theories. *Psychol Bull* 88: 329–353, 1980.

Fischl B, van der Kouwe A, Destrieux C, Halgren E, Segonne F, Salat DH, Busa E, Seidman LJ, Goldstein J, Kennedy D, Caviness V, Makris N, Rosen B, Dale AM. Automatically parcellating the human cerebral cortex. *Cereb Cortex* 14: 11–22, 2004.

Freeman WJ, Burke BC, Holmes MD. Aperiodic phase re-setting in scalp EEG of beta-gamma oscillations by state transitions at alpha-theta rates. *Hum Brain Mapp* 19: 248–272, 2003.

Fukushima K. Neocognitron: a self organizing neural network model for a mechanism of pattern recognition unaffected by shift in position. *Biol Cybern* 36: 193–202, 1980.

Gattass R, Gross CG, Sandell JH. Visual topography of V2 in the macaque. *J Comp Neurol* 201: 519–539, 1981.

Gattass R, Sousa AP, Gross CG. Visuotopic organization and extent of V3 and V4 of the macaque. *J Neurosci* 8: 1831–1845, 1988.

Gawne TJ, Martin JM. Responses of primate visual cortical V4 neurons to simultaneously presented stimuli. *J Neurophysiol* 88: 1128–1135, 2002.

Gollisch T, Meister M. Rapid neural coding in the retina with relative spike latencies. *Science* 319: 1108–1111, 2008.

Grill-Spector K, Malach R. The human visual cortex. *Annu Rev Neurosci* 27: 649–677, 2004.

Hasson U, Yang E, Vallines I, Heeger DJ, Rubin N. A hierarchy of temporal receptive windows in human cortex. *J Neurosci* 28: 2539–2550, 2008.

Haxby J, Grady C, Horwitz B, Ungerleider L, Mishkin M, Carson R, Herscovitch P, Schapiro M, Rapoport S. Dissociation of object and spatial visual processing pathways in human extrastriate cortex. *Proc Natl Acad Sci USA* 88: 1621–1625, 1991.

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

Hopfield JJ. Pattern recognition computation using action potential timing for stimulus representation. *Nature* 376: 33–36, 1995.

Hung C, Kreiman G, Poggio T, DiCarlo J. Fast Read-out of Object Identity from Macaque Inferior Temporal Cortex. *Science* 310: 863–866, 2005.

Keysers C, Perrett DI. Visual masking and RSVP reveal neural competition. *Trends Cogn Sci* 6: 120–125, 2002.

Keysers C, Xiao DK, Foldiak P, Perrett DI. The speed of sight. *J Cogn Neurosci* 13: 90–101, 2001.

Kiani R, Esteky H, Tanaka K. Differences in onset latency of macaque inferotemporal neural responses to primate and non-primate faces. *J Neurophysiol* 94: 1587–1596, 2005.

Kobatake E, Tanaka K. Neuronal selectivities to complex object features in the ventral visual pathway of the macaque cerebral cortex. *J Neurophysiol* 71: 856–867, 1994.

Kovacs G, Vogels R, Orban GA. Cortical correlate of pattern backward masking. *Proc Natl Acad Sci USA* 92: 5587–5591, 1995.

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

Logothetis NK, Sheinberg DL. Visual object recognition. *Annu Rev Neurosci* 19: 577–621, 1996.

Macknik S. Visual masking approaches to visual awareness. *Prog Brain Res* 155: 177–215, 2006.

Missal M, Vogels R, Li CY, Orban GA. Shape interactions in macaque inferior temporal neurons. *J Neurophysiol* 82: 131–142, 1999.

Nelson SB. Temporal interactions in the cat visual system. I. Orientation-selective suppression in the visual cortex. *J Neurosci* 11: 344–356, 1991.

Privman E, Fisch L, Neufeld MY, Kramer U, Kipervasser S, Andelman F, Yeshurun Y, Fried Itzhak Malach R. Antagonistic relationship between gamma power and visual evoked potentials revealed in human visual cortex. *Cereb Cortex* 21: 616–624, 2011.

Reynolds JH, Chelazzi L, Desimone R. Competitive mechanisms subserve attention in macaque areas V2 and V4. *J Neurosci* 19: 1736–1753, 1999.

Richmond BJ, Optican LM, Spitzer H. Temporal encoding of two-dimensional patterns by single units in primate primary visual cortex. I. Stimulus-response relations. *J Neurophysiol* 64: 351–369, 1990.

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

Ringach DL, Hawken MJ, Shapley R. Dynamics of orientation tuning in macaque primary visual cortex. *Nature* 387: 281–284, 1997.

Rolls E. Neural organization of higher visual functions. *Curr Opin Neurobiol* 1: 274–278, 1991.

Rolls ET, Tovee MJ, Panzeri S. The neurophysiology of backward visual masking: information analysis. *J Cogn Neurosci* 11: 300–311, 1999.

Sceniak MP, Ringach DL, Hawken MJ, Shapley R. Contrast's effect on spatial summation by macaque V1 neurons. *Nat Neurosci* 2: 733–739, 1999.

Schmolesky M, Wang Y, Hanes D, Thompson K, Leutgeb S, Schall J, Leventhal A. Signal timing across the macaque visual system. *J Neurophysiol* 79: 3272–3278, 1998.

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

Serre T, Kreiman G, Kouh M, Cadieu C, Knoblich U, Poggio T. A quantitative theory of immediate visual recognition. *Prog Brain Res* 165C: 33–56, 2007.

Singer J, Sheinberg D. Holistic processing unites face parts across time. *Vision Res* 46: 1838–1847, 2006.

Singer JM, Kreiman G. Short temporal asynchrony disrupts visual object recognition. *J Vis* 14: 1–14, 2014.

Tanaka K. Inferotemporal cortex and object vision. *Annu Rev Neurosci* 19: 109–139, 1996.

Thorpe S, Fize D, Marlot C. Speed of processing in the human visual system. *Nature* 381: 520–522, 1996.

vanRullen R, Thorpe S. Surfing a spike wave down the ventral stream. *Vision Res* 42: 2593–2615, 2002.

Vinje WE, Gallant JL. Natural stimulation of the nonclassical receptive field increases information transmission efficiency in V1. *J Neurosci* 22: 2904–2915, 2002.

Wallis G, Rolls ET. Invariant face and object recognition in the visual system. *Prog Neurobiol* 51: 167–194, 1997.

Whittingstall K, Logothetis NK. Frequency-band coupling in surface EEG reflects spiking activity in monkey visual cortex. *Neuron* 64: 281–289, 2009.

Yoshor D, Bosking WH, Ghose GM, Maunsell JH. Receptive fields in human visual cortex mapped with surface electrodes. *Cereb Cortex* 17: 2293–2302, 2007.

Zipser K, Lamme VA, Schiller PH. Contextual modulation in primary visual cortex. *J Neurosci* 16: 7376–7389, 1996.

Zoccolan D, Kouh M, Poggio T, DiCarlo JJ. Trade-off between object selectivity and tolerance in monkey inferotemporal cortex. *J Neurosci* 27: 12292–12307, 2007.