Cortical representation of persistent visual stimuli

Edden M. Gerber [a,*], Tal Golan [a], Robert T. Knight [b,c], Leon Y. Deouell [a,d,**]

[a] Edmond and Lily Safra Center for Brain Sciences, The Hebrew University of Jerusalem, Jerusalem, 9190401, Israel
[b] Helen Wills Neuroscience Institute, University of California, Berkeley, Berkeley, CA 94720, USA
[c] Dept. of Psychology, University of California, Berkeley, Berkeley, CA 94720, USA
[d] Dept. of Psychology, The Hebrew University of Jerusalem, Jerusalem, 9190501, Israel

ARTICLE INFO

ABSTRACT

Research into visual neural activity has focused almost exclusively on onset- or change-driven responses and little is known about how information is encoded in the brain during sustained periods of visual perception. We used intracranial recordings in humans to determine the degree to which the presence of a visual stimulus is persistently encoded by neural activity. The correspondence between stimulus duration and neural response duration was strongest in early visual cortex and gradually diminished along the visual hierarchy, such that is was weakest in inferior-temporal category-selective regions. A similar posterior-anterior gradient was found within inferior temporal face-selective regions, with posterior but not anterior sites showing persistent face-selective activity. The results suggest that regions that appear uniform in terms of their category selectivity are dissociated by how they temporally represent a stimulus in support of ongoing visual perception, and delineate a large-scale organizing principle of the ventral visual stream.

1. Introduction

Although visual perception of scenes or objects typically occurs over an extended time window, what we know of human high level visual perception is largely based on onset responses, that is, responses to change (e.g. the EEG/intracranial face-selective N170/N200; Bentin et al., 1996; Allison et al., 1999). While onset responses hold a wealth of information on the functional properties of the respective neural population, they cannot disentangle the transient response to change, from neural activity driven by the ongoing presence of a percept. Thus, how real-time experience persists beyond stimulus onset is unknown, leaving the neural basis for the brunt of the time course of perception unaccounted for. Put simply, our question is how do we know, in real-time and beyond the initial detection, whether a stimulus we gaze at is still present 500 or 1500 ms beyond its onset, and how do we know its content, e.g. whether it is (still) a face or an object? This question is distinct from that of explicit duration estimation tasks (e.g. was a probe stimulus shorter or longer than a reference stimulus; Buhusi and Meck, 2005; Lewis and Miall, 2003) in which judgments are performed post-hoc, that is, after stimulus offset. Here, we are concerned with how the visual system codes the continued presence of an object in real time. Some evidence of sustained responses to visual stimuli can be gleaned

from single unit (Ikeda and Wright, 1974; Kulikowski et al., 1979; Petersen et al., 1988) or fMRI studies (Gilaie-Dotan et al., 2008) which used long duration visual stimuli, and from variable duration EEG/MEG responses in an explicit duration estimation task (N'Diaye et al., 2004; Pouthas et al., 2000). However these studies did not directly examine the neural basis of sustained, real-time perception.

We recorded from ten human subjects implanted with subdural electrodes over visual cortices using electrocorticography (ECoG) while subjects were engaged in a novel target detection paradigm, with faces and objects presented for variable durations. Since even a brief stimulus can produce a sustained response (Fisch et al., 2009), using variable stimulus durations was crucial for isolating the part of the response driven specifically by the stimulus ongoing presence. Notably, no duration estimation was required.

We found that in early visual cortex (EVC) the high-frequency broadband response (HFB, >30 Hz) closely tracked the time course of the stimulus, but this precision decreased along the visual hierarchy, i.e. from V1/V2 to V3/V4, from early visual cortex to inferior temporal (IT) cortex, and from posterior to anterior IT. Only the posterior part of IT robustly encoded the presence of the stimulus over time. This posterior-anterior gradient could not be explained by signal-to-noise ratio, nor by differential influences of attention.

---

* Corresponding author. Edmond and Lily Safra Center for Brain Sciences, The Hebrew University, The Edmond J. Safra Campus, Jerusalem, 9190401, Israel.
** Corresponding author. Edmond and Lily Safra Center for Brain Sciences, The Hebrew University, The Edmond J. Safra Campus, Jerusalem, 9190401, Israel.
E-mail addresses: edden.gerber@mail.huji.ac.il (E.M. Gerber), leon.deouell@mail.huji.ac.il (L.Y. Deouell).

2. Materials and methods

2.1. Subjects

Ten patients (3 female, 7 male, mean age 41.5 years, range 19–65, two left-handed) undergoing surgical treatment for intractable epilepsy and implanted with chronic subdural electrodes participated in the experiment (Table 1). Seven patients were recorded in the Stanford School of Medicine, 2 in the California Pacific Medical Center (CPMC), and 1 in the UCSF Medical Center.

Electrode arrays were implanted on the right hemisphere for 8 of the patients, and on the left hemisphere for 2, with electrode location determined solely by clinical needs. All subjects gave informed consent approved by the UC Berkeley Committee on Human Research and corresponding IRBs at the clinical recording sites.

2.2. Stimuli and tasks

Recording was conducted in the epilepsy ICU. Stimuli were presented on a laptop screen and responses captured on the laptop keyboard. Stimuli were grayscale images of either frontal human faces, man-made objects (round watch-faces, tools, vehicles, furniture, clothing items, musical instruments and household objects), or animals, presented on a square uniform gray background in the center of the screen and extending across approximately 5° of the visual field in each direction. Only the face and object categories were used in the main analyses. Additional categories including body parts and houses were excluded from analysis due to the paucity of exemplars. Face and object images did not significantly differ in terms of luminance (206.2 and 207 mean pixel intensity for faces and objects, respectively) or contrast defined as the standard deviation of pixel intensity (41.8 and 41.1 respectively). Each stimulus was presented for either 300, 600, 900, 1200 or 1500 ms for the first 3 subjects (Fig. 1A), or 300, 900 or 1500 ms for the last 7 subjects (Fig. 1B. the number of durations was reduced to allow the additional control task described below without excessively prolonging the experiment). All categories had the same probability of appearing with each duration. Inter-stimulus interval varied from 600 to 1200 ms with 150 ms steps, during which a fixation cross was presented. In the main experimental condition, subjects were instructed to fixate on the center of the screen, and respond with a button press whenever a clothing item was
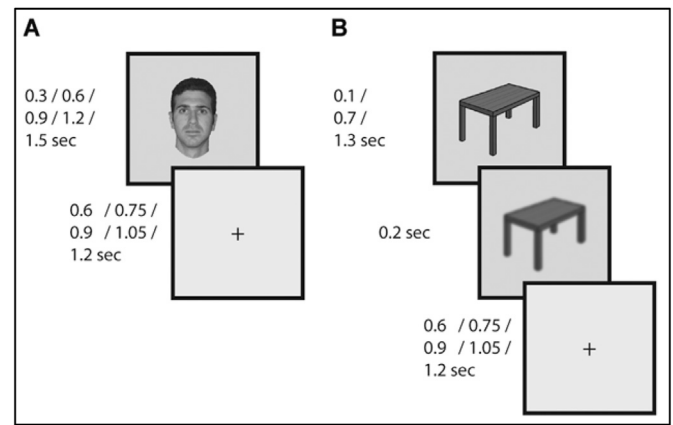

Fig. 1. **Experimental paradigm.** (A) Images were presented for 300, 900, or 1500 ms (for 3 subjects also 600 or 1200 ms) with variable inter-stimulus interval (ISI) during which a fixation cross was presented. Subjects responded with a button press to presentation of targets (clothing images; 10% of trials). (B) Dual-task control. This task was identical to the first except that subjects also had to respond to rare blurring of the image in the last 200 ms of its presentation.

presented (10% of trials). The last 7 subjects performed an additional experimental task (dual attention-control task) where in addition to responding to images of clothing items, they were also required to respond to rare instances when an image of any category became blurry for the last 200 ms of its presentation (clothing targets and blurred image targets each accounted for 5% of the trials). The onset time for each image was registered alongside the ECOG data from a photodiode placed on the laptop screen, recording a white rectangle displayed at the same time as the image at the corner of the screen.

2.3. ECoG acquisition and data processing

Each subject was implanted with subdural arrays containing 53–128 contact electrodes (AdTech Inc.). In total, 1067 electrodes were examined. Each electrode was 2.3 mm in diameter, with 5 or 10 mm spacing between electrodes within an array, arranged in 1-dimensional strips or 2-dimensional grids. Recordings were sampled at 1000 Hz (CPMC), 3051.76 Hz (Stanford, UCSF) or 1535.88 Hz (Stanford) and resampled to 1000 Hz offline (with the exception that the original 3051.76 Hz sampling rate was used for the results reported in Fig. 2). A high-pass filter was applied online to the signal at either 0.1 Hz or 0.5 Hz in different subjects. 159 electrodes manifesting ictal spikes or persistent noise were visually identified and removed from analysis, as were time intervals with excessive noise or ictal activity as determined by one of the authors (RTK). All remaining electrodes were re-referenced offline to the average potential of all non-rejected electrodes, separately for each subject. Unless otherwise indicated, all data processing and analysis were done using custom Matlab code (Mathworks, Natick, MA).

2.4. Electrode localization

Electrode locations were identified manually using BioImageSuite (www.bioimagesuite.org) on a post-operative Computed Tomography (CT) scan co-registered to a pre-operative MR scan using the FSL software package (Jenkinson et al., 2002; Jenkinson and Smith, 2001). Individual subjects' brain images were skull-stripped and segmented using Free-Surfer (http://surfer.nmr.mgh.harvard.edu). Localization errors driven by both co-registration error and anatomical mismatch between pre- and post-operative images were reduced using a custom procedure which uses a gradient descent algorithm to jointly minimize the squared distance between all electrodes within a single electrode array/strip and the cortical pial surface (see Dykstra et al., 2012 for a similar procedure). In contrast to methods which only attempt to correct individual electrodes'

Table 1
ECoG patients.

| Patient ID	Gender	Age	Handedness	Electrode Coverage	Number of Electrodes	Responsive Electrodes
S1	F	38	R	RH: Tem, Occ, Fro	64	32
S2	M	46	R	RH: Tem, Occ, Par	112	28
S3	M	41	L	RH: Tem, Occ, Par	118	44
S4	M	19	R	RH: Tem, Occ, Par, Fro	123	15
S5	F	42	R	RH: Tem, Occ, Par	107	25
S6	M	29	R	RH: Tem, Occ, Par, Fro	113	16
S7	M	47	L	LH: Tem, Occ, Fro	64	9
S8	F	65	R	RH: Tem, Par, Fro	116	39
S9	M	47	R	LH: Tem, Par	128	52
S10	F	36	R	RH: Tem, Occ, Par, Fro	122	32
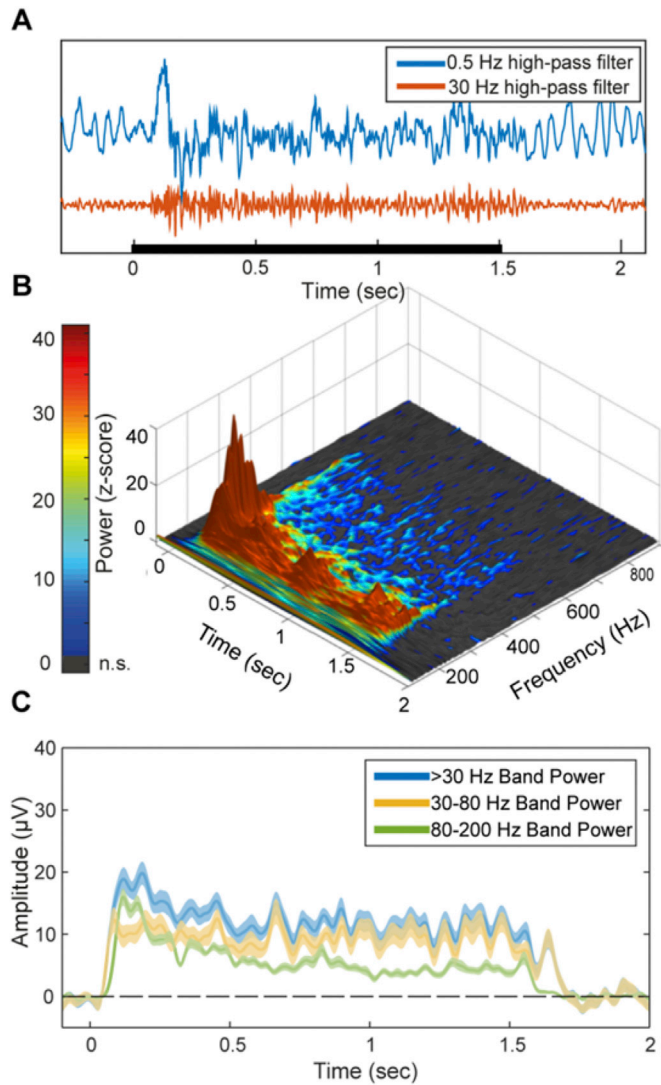
Fig. 2. Representative spectral response profile. (A) The high-frequency response is evident in raw single trials. Blue and Red traces correspond to 0.5 Hz and 30 Hz high-pass filtered traces respectively of a single trial from an early visual cortex (V1/V2) electrode. Black bar indicates the stimulus duration. (B) Time-frequency response to 1500 ms visual stimuli averaged across both categories, in the same electrode. Z-axis and color scale represent z-score of the power at each time-frequency data point as normalized by the baseline for the same frequency band. Since both the baseline and response spectra are dominated by a 1/f power law in terms of mean power and variance, this normalization makes it possible to observe relative modulations in high-frequency bands which are otherwise too small in comparison to lower frequencies. Gray pixels represent non-significant z-score values (p < 0.05, uncorrected). Note that while the early response components contain higher power at high frequencies, the upper bound of significantly above-baseline activity is roughly fixed at 400–500 Hz throughout the activity time course, implying a general attenuation of the response over time rather than a narrowing of the response bandwidth. (C) Amplitude of band-limited activity in three frequency bands for the same electrode. Shading corresponds to standard error. While including lower frequencies increases the response SNR, all subsets of the high-frequency range produce a qualitatively similar response pattern, suggesting that the underlying activity is indeed broadband.

position in relation to the pial surface, this method preserves the original array topography, thus providing a more reliable estimate of actual electrode positions.

Individual patients' brains and electrode coordinates were next co-registered to a common brain template (FreeSurfer's *fsaverage* template) using surface-based registration (Fischl et al., 1999). This registration optimally preserves the mapping of electrode locations to anatomical features between the native and common brain space despite individual variability in cortical folding patterns, in contrast to volume-based registration which preserves the 3D spatial configuration of electrode arrays but does not account for differences in cortical folding patterns. To identify visual areas, each cortical surface was resampled to a standard mesh with a fixed number of co-registered nodes using SUMA (Argall et al., 2006), and visual area labels were imported from a probabilistic map of visual areas topography (Wang et al., 2014). Visualization of electrode data in Fig. 3 was based on surface registration to an MNI152 standard-space T1-weighted average structural template image.

2.5. High frequency broadband power signal

To extract the high-frequency broadband (HFB) signal, the recorded signal was high-pass filtered using a 3rd order Butterworth filter with a 30 Hz cutoff, and the power along time in each electrode was computed using the square of the Hilbert transform of the filtered signal. Following the observation that stimulus-driven responses in most visually-responsive electrodes (averaged across durations and categories) manifests as an increase in a broad frequency band starting at 20–40 Hz, we selected the >30 Hz band as our target for analysis. The 30 Hz lower frequency bound is commonly used to define the high-frequency (or "gamma") band (e.g. Fisch et al., 2009; Engell et al., 2012; Jacques et al., 2016). High frequency broadband power was shown to correlate with local neural firing rate (Manning et al., 2009; Miller et al., 2014, 2009; Ray and Maunsell, 2011). Morlet wavelet analysis was used solely for visualization purposes (Figs. 2, 7 and 8), with a wavelet constant of 12 and center frequency bins ranging from 2 to 800 Hz.

2.6. Identification of visually-responsive electrodes

Visually-responsive electrodes were defined as those exhibiting a significant (p < 0.01) HFB power increase in response to either the face or object stimulus categories across trials, as assessed by a cluster-based permutation test compared to a 500 ms pre-stimulus baseline (Maris and Oostenveld, 2007). Subsequent analyses were performed on the set of electrodes showing a significant response to at least one category. In addition, the response onset latency of each visually responsive electrode was derived as the onset latency of its earliest significant activity cluster.

2.7. Category-selectivity

Category selectivity was defined as a significant difference between the HFB onset response to face vs. object stimuli (note that this type of selectivity does not necessarily imply specificity to the preferred category, and could arise out of a less specific selectivity, e.g. animate vs. non-animate stimuli (Grill-Spector and Weiner, 2014)). The onset response magnitude for each category was computed by taking the root mean square (RMS) of all data points from the first 300 ms of the trial-average HFB response to the category. A selectivity index (SI) was computed as:

$$SI = \frac{R_f - R_o}{R_f + R_o}$$

where R_f and R_o are the RMS of the mean response to faces and objects, respectively. Positive values correspond therefore to face-preference, and negative values to object-preference. Statistical significance was derived with a randomization test whereby in each of 10,000 iterations the category labels were shuffled and the SI recomputed, producing a null-hypothesis distribution. P-values were derived from the rank of the original SI within this distribution (in absolute value), corrected for multiple comparisons across all electrodes using the False Detection Rate (FDR) method (Benjamini and Yosef, 1995). In this analysis and whenever FDR was used in this study, the FDR q value was 0.05.

2.8. Single-trial duration tracking

The purpose of the single-trial duration tracking analysis was to
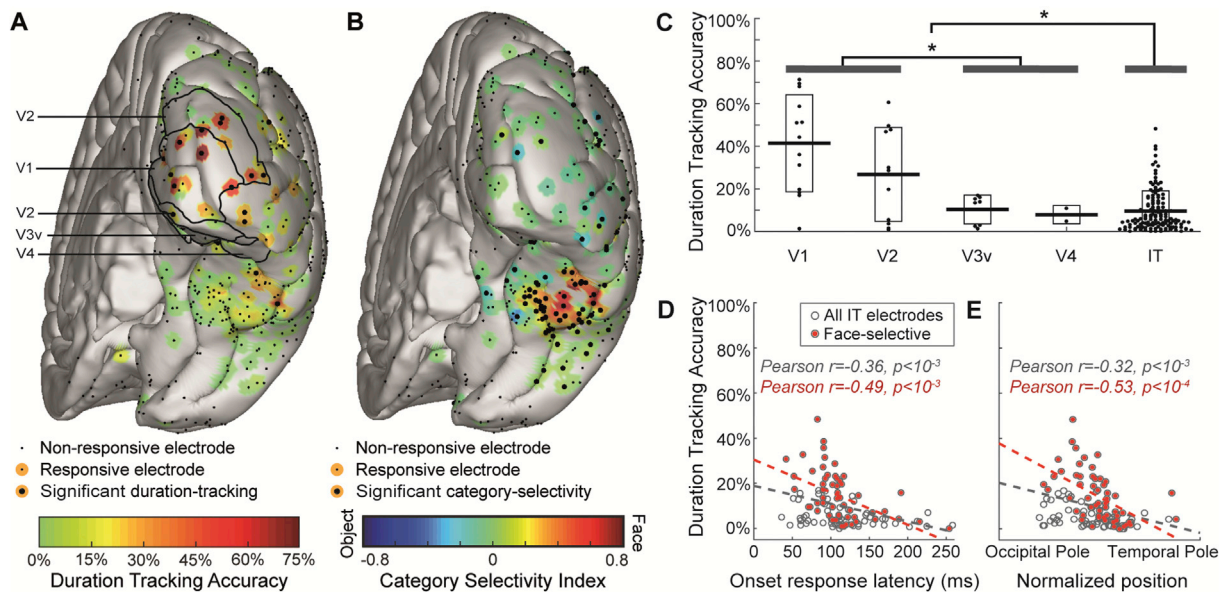
Fig. 3. Duration-tracking accuracy decreases gradually along the ventral visual stream. (A) Duration-tracking accuracy for all subjects' electrodes projected onto a common brain template. Visually responsive electrodes are surrounded by a color patch representing duration-tracking accuracy. Significant duration-tracking (FDR corrected, q = 0.05) is denoted by a thicker black dot. Anatomical regions marked with black contour lines within EVC are based on surface registration to a probabilistic atlas (Wang et al., 2014). (B) Same as (A), but for category selectivity index. Negative values (blue) correspond to object-preference. (C) Relation between duration-tracking and hierarchical position along the ventral stream (EVC areas defined based on the probabilistic atlas). Each dot corresponds to a single electrode, with horizontal dispersion based on data point density. Boxes correspond to standard deviation. Asterisks mark significant difference (p < 0.05) between EVC areas and IT and between V1/V2 and V3v/V4. Note that V3v/V4 sites are not necessarily earlier than all IT sites in terms of response latency. (D) duration-tracking within IT plotted against onset response latency, as a proxy for hierarchical position along the ventral stream. Face selective electrodes are marked red. (E) Same as (D), with hierarchical position measured as the electrode's coordinate along the occipital-temporal axis.

identify electrodes with single-trial HFB responses that are reliably (across trials) sustained above baseline for a duration corresponding to the stimulus duration (i.e. duration-tracking responses). The rationale of this approach is to emulate an online process where the persistence of the stimulus percept can be decoded from the region's current HFB signal, by testing at each consecutive time point whether the signal is still above the decision threshold ("activation") or not ("baseline").

The HFB signal was first smoothed using a 300 ms window moving average. To derive the activation threshold, an "activation" sample was generated by taking the means of the 1200–1500 ms post-stimulus window across 1500-ms-duration trials, and a "baseline" sample was generated by taking the means of the 300 ms pre-stimulus window from the same trials. The late 1200–1500 ms window was selected to allow the test to be sensitive to the relatively low-magnitude activation of the late sustained response as opposed to the high-magnitude early onset response component. The threshold was then set as the median value of the two equally-sized samples taken together. For each trial, the threshold was generated based on all trials excluding the current one. In summary, the threshold indicates whether a power value averaged across 300 ms is more likely to reflect a baseline or active epoch. The decoded duration of the active percept was then defined as the latency (relative to the electrode's onset response latency) of the first time point where the smoothed HFB signal drops below the threshold, within a 1800 ms post-stimulus period. Only trials which included a minimal 1800 ms window from stimulus onset to the next trial stimulus onset were included in this analysis, thus excluding all 300 ms duration trials, and some of the 600 and 900 ms duration trials. Using a shorter analysis window of 1200 ms and thus including all 300, 600 and 900 ms trials, while excluding 1200 and 1500 ms trials, led to very similar results. Finally, the duration-tracking accuracy was computed as the percentage of "correct" estimated durations which were within a ±150 ms error margin around the true stimulus duration. The results of this analysis were not qualitatively changed when selecting a different error margin (i.e. ±100 or ±200 ms), or smoothing window size (i.e. 200 or 400 ms). Statistical significance was determined using a randomization test such that at each of 10,000 iterations, the duration-tracking accuracy was re-computed based on

shuffled stimulus duration labels to produce a null-hypothesis distribution. P-values were derived from the rank of the original accuracy within this distribution, corrected for multiple comparisons using the FDR method. Duration-tracking was assessed across both face and object stimulus trials for non-category-selective electrodes, and for the preferred category for category-selective electrodes.

2.9. Duration-dependence

Duration-tracking accuracy is a strict index of sustained activity as it requires that HFB power does not drop below baseline throughout the stimulus duration in single trials. As such, it may be insensitive to information about stimulus duration that may be present in sustained responses across trials without being reliable on a single-trial level. In order to characterize this type of information, we derived a second index which was aimed at being maximally sensitive to the relation between stimulus duration and sustained activity across trials. This weaker index of duration-tracking, which we will hereby refer to as *duration dependence*, was computed by correlating trials' duration with the number of above-baseline HFB data points in a post-stimulus window. This resulted in a correlation coefficient for each electrode, where 0 indicates no statistical relation between stimulus duration and the number of above-baseline HFB time points and 1 indicates a perfect linear relation. To determine the number of above-baseline points for each trial, we smoothed the HFB signal using a 3rd order Butterworth 10 Hz low-pass filter, and counted the number of time points with amplitude larger than the mean of the 500 ms pre-stimulus baseline period. As in the duration-tracking analysis, trials which were too short to include an 1800 ms post-stimulus onset-to-onset window were excluded. The number of above-baseline data points was counted for each trial within a window of 600–1800 ms post-stimulus (minimizing the influence of the onset responses, which do not contain relevant variance). The use of a count metric was selected from among several possible approaches to quantifying duration-dependence due to its relative insensitivity to amplitude differences, emphasizing instead differences in duration. To rule out the possibility that duration-dependent *offset* responses drive the correlation (e.g.,

longer offset response following longer stimulus duration), the number of above-baseline points in the 600 ms post-offset window for each trial was measured as well, and if it was positively correlated with stimulus duration its explained variance was regressed out prior to the main correlation analysis. Statistical significance of the correlations was derived parametrically from the correlation analysis, and corrected for multiple comparisons using FDR. Duration-dependence was assessed across both face and object stimulus trials for non-category-selective electrodes, and for the preferred category for category-selective electrodes.

2.10. Comparison of responses across single and dual-tasks

As a sensitive test for differences in the duration of the responses between the two tasks, we compared the number of above-baseline time points between the two tasks for the 1500 ms stimulus-duration trials. For each trial, we counted the number of time points with amplitude larger than the mean of the 500 ms pre-stimulus period, in the 0–1800 ms post-stimulus window. The total number of above-baseline time points across trials was compared between the two tasks using a two-sample *t*-test. Results were corrected for multiple comparisons using FDR.

2.11. Saccade detection

Saccadic spike potentials detected from two temporal-pole electrodes were used to indicate the timing of eye movements in a single subject where spike potentials could be detected from electrodes just behind the eyes. Detection of the spike potentials was performed by high-pass filtering the signal from one of the two electrodes at 30 Hz, convolving it with a "canonical" saccadic spike model, and marking supra-threshold time points as saccadic spikes (Keren et al., 2010). To create the saccadic spike model, three healthy scalp EEG control subjects (female, age 26–31) participated in a scalp-EEG version of the same experiment with simultaneous eye-tracking (Eyelink 1000/2K, SR Research, Canada; binocular tracking at 500 Hz). Saccade rates were approximately 1.5/sec as predicted by the literature (Engbert, 2006). Our spike potential model (Fig. 8B) was generated by high-pass filtering the radial electro-ocular channel of each EEG subject at 30 Hz, and averaging it locked to saccade events detected by video eye-tracking and across subjects. After the convolution of the intracranial signal with the model, a 3 std. deviation threshold was set and any above-threshold activity was interpreted as a saccadic spike (Keren et al., 2010).

2.12. Simulation of saccade-driven sustained activity

In addition to the analysis based on the saccadic spike potential, we used a simulation to investigate the hypothesis that sustained activity in the early visual cortex could be driven by successive transient visual-fixation-locked activity (that is, fixed-duration neural responses to the retinal update following a saccade, while the image is presented). We simulated a "best-case" scenario for such a hypothesis, consisting of a fixed-duration high-SNR response contingent on a saccade occurring while the stimulus is presented. We used saccade and micro-saccade detection data from the three healthy subjects (female, age 26–31) who participated in a scalp-EEG version of the same experiment with simultaneous eye-tracking. A high-SNR saccade-driven activity time course was simulated for each of the three subjects by convolving each occurrence of a saccade coinciding with stimulus presentation with a fixed-duration square response window. To avoid making any assumptions about the duration of the saccade-locked response, it was varied in separate simulations between 50 and 1000 ms in 50 ms steps. The first 500 ms of each trial were set as active by default to simulate a fixed stimulus-onset response. For each subject and simulated saccade-locked response duration, the overall simulated response durations were estimated across trials in the same way as in the duration-tracking analysis

of the empirical data, and the mean absolute estimation error was recorded.

3. Results

Subjects viewed grayscale images of faces, objects, or other miscellaneous images presented for a duration of 300, 600, 900, 1200 or 1500 ms (Fig. 1). To maintain attention subjects responded via button press to a rare target category (clothes). All subjects detected the targets successfully (mean ± SD hit rate 90.4% ± 8.6%, false alarm rate 1.9% ± 3.2%). Responses to the rare targets were not included in the following analyses. Stimulus-induced modulation of the ECoG signal was measured as an increase in the power of high-frequency broadband signal (HFB, >30 Hz). Fig. 2 shows characteristic single-trial and average responses. Out of 1067 electrodes, subsequent analysis was performed on 292 electrodes designated as visually-responsive, based on a significant modulation of HFB power following the onset of least one stimulus category.

3.1. Duration-tracking decreases from EVC to category-selective IT cortex

To determine the degree to which the real-time, moment-by-moment presence of the visual percept is analogically reflected in the HFB signal on a single-trial basis, we quantified for each trial and electrode the duration for which the temporally-smoothed HFB signal was continuously sustained above baseline following the response onset. A *duration-tracking accuracy* was then computed as the percentage of trials where the duration of the sustained response was within ±150 ms of the true stimulus duration (see Materials and Methods; the results were similar for margins of ±100 or ±200 ms). The rationale of this approach is to emulate a real-time process where the persistent presence of the stimulus percept can be decoded from the region's HFB signal, by testing, at each consecutive time window, whether the signal is still above the decision threshold ("activation") or not ("baseline"). Of the 292 visually-responsive electrodes, 21 electrodes in eight subjects showed significant duration-tracking accuracy based on a permutation test. Seventeen of these electrodes were located in the posterior occipital lobe (50% of responsive early visual cortex electrodes) and 4 located on IT cortex (1.5% of responsive IT electrodes; see Figs. 3A and 6).

To identify category-selective electrodes, a selectivity index (SI) metric was computed corresponding to the response amplitude difference between faces and objects in the 300 ms post-onset window, normalized by their sum (positive SI corresponds to face-preference; see Materials and Methods). One hundred and twenty seven electrodes across all subjects showed category-selectivity (Fig. 3B), 96 of which were on IT (64% of IT electrodes), and the rest on the occipital, parietal, or inferior frontal lobes. In six category-selective electrodes that were also duration-tracking, we additionally tested whether category selectivity was sustained beyond the onset response, by calculating the SI in the 1200–1500 ms post-stimulus window for 1500-ms-duration stimuli. Significant category selectivity in sustained activity was found in only two duration-tracking electrodes, both in posterior IT.

As expected, category selectivity was low over EVC and maximal over the fusiform gyrus. In contrast, duration-tracking accuracy was maximal over EVC and followed a diminishing gradient along the ventral stream toward anterior IT – such that the majority of responses in regions highly informative about the category of a stimulus showed low sensitivity to stimulus duration, and thus to its presence beyond its onset (Fig. 3).

EVC electrodes showed significantly higher duration-tracking accuracy than IT electrodes (Fig. 3C, t(166) = 7.01, p < 10^{-5}). Within EVC, V1-V2 electrodes showed higher accuracy than downstream V3v-V4 electrodes (t(32) = 3.31, p < 10^{-3}). There was a trend for higher accuracy in V1 than in V2 (t(22) = 1.59, p = 0.06). Within IT, we tested for a posterior-anterior gradient using two converging indexes for position along the ventral stream: onset response latency and anatomical position. Anatomical position was quantified by projecting the coordinates of each

electrode to a line extending from the occipital pole to the temporal pole of each individual subject's right hemisphere, and normalizing the result to the range 0–1 (1 is closer to the temporal pole; electrodes over the left hemisphere were omitted from this analysis due to their paucity). Duration-tracking accuracy was negatively correlated with response onset latency (Fig. 3D, r = −0.36, p < 10⁻³) as well as anatomical position (Fig. 3E, r = −0.32, p < 10⁻³). The same correlations were found when limiting the analysis to face-selective electrodes (r = 0-0.49, p < 10⁻³ for onset latency and r = −0.53, p < 10⁻⁴ for position). Object-selective electrodes showed a similar negative correlation but were too few for robust statistical results. Using a mixed-model regression approach, we verified that these correlations were not driven by differences between subjects, and observed that the anatomical gradient can be seen in individual subjects (Fig. 4). Similar results were obtained when controlling for signal-to-noise ratio across electrodes, as expressed by onset response peak magnitude and baseline noise level (Inline Supplementary Fig. S2), when the mean error of the response estimation (i.e. difference from true stimulus durations) was used as dependent variable instead of the duration-tracking accuracy metric (Inline Supplementary Fig. S3), or with a more liberal index of *duration-dependence*, based on the correlation between the stimulus duration and the number of above-threshold post-stimulus time points across trials (see Materials and Methods and Inline Supplementary Fig. S4).

Sites with low duration-tracking accuracy did not necessarily show only transient onset responses. Rather, many non-duration-tracking electrodes were characterized by prolonged HFB activity that was uninformative about single-trial stimulus duration. Fig. 5 shows representative examples of an EVC electrode with high duration-tracking accuracy (left), a face-selective IT electrode with a prolonged response that is not informative regarding the duration of the stimulus (right), and one of the two rare posterior IT electrode with a category-selective, duration-tracking response (middle). We next describe this rare response profile, encoding both the content and the ongoing presence of a stimulus over time.

3.2. Stimulus presence and category jointly encoded in posterior IT cortex

As noted, two face-selective electrodes from two subjects showed both duration-tracking and sustained category selectivity (Fig. 6). Both were similarly located laterally from the posterior part of the mid-fusiform sulcus (Weiner et al., 2014). Two additional posterior fusiform electrodes, located medial to the mid-fusiform sulcus, were also object-selective and duration-tracking, however their selectivity was limited to the onset response, that is, the response to faces and objects did not differ in the late (1200–1500 ms) time window. The locations of the two face-selective electrodes correspond to the posterior fusiform face-selective region pFus-faces (or FFA-1, Grill-Spector and Weiner, 2014), and to cytoarchitectonic region FG2 (Caspers et al., 2013), suggesting a functional distinction between this region and the more anterior mFus-faces region (FFA-2, related to cytoarchitectonic area FG4, Lorenz et al., 2015), which showed high category selectivity but little duration-tracking.

3.3. Differences in duration-tracking are not explained by task attentional demands

The diminished sustained activity in IT could be explained by the possibility that subjects stopped attending to the stimulus or averted their gaze after its appearance (eye tracking was not feasible at the ICU bedside). If IT is more sensitive than EVC to the level of attention, this could drive the dissociation between these regions in sustained responses (Hasson et al., 2002; Levy et al., 2001). To test this, a second version of the experiment was performed by 7 of the 10 subjects in addition to the original version (Fig. 1B). In this difficult dual-task control, they had to detect a rare brief blurring of any image just before its disappearance, in addition to the detection of clothing items. The dual detection task required gaze and attention to be sustained throughout image presentation as the timing of stimulus offset was unpredictable. Detecting clothing items in the dual-task control
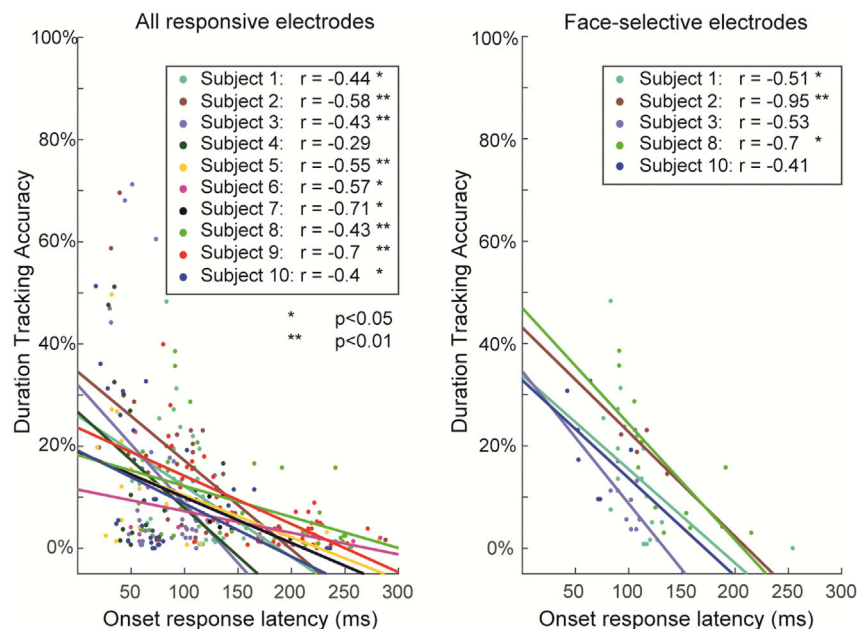

Fig. 4. Single subject correlations. To verify that the negative gradient of duration-tracking accuracy is not driven by differences between subjects (i.e. to verify that tracking accuracy and position along the ventral stream are correlated within subjects, rather than that subjects with more posterior coverage have generally higher duration-tracking), we fitted a mixed-effect linear regression model to the single-trial duration-tracking data. The binomial result of each trial in each electrode (accurate/inaccurate response duration) was used as the dependent variable, and either the electrodes' onset response latency, or its posterior-anterior position was used as an independent variable, with random intercepts and slopes for each subject and random intercepts for individual trials. The continuous measures of latency and anatomical position were used as the independent variables for the analysis of electrodes across all visual areas rather than region labels, because there were not sufficient data for the model to converge with an ordinal independent variable. The results indicated a significant fixed effect for both latency and position for all responsive electrodes, all right-hemisphere IT electrodes, all right-hemisphere face-selective electrodes, and all right-hemisphere object selective electrodes. Left panel: data points for all responsive electrodes divided by subject, and subject-specific regression lines. Right panel: same for right-hemisphere IT face-selective electrodes. Mixed-effects analysis was performed using R (www.r-project.org) with the lme4 software package.
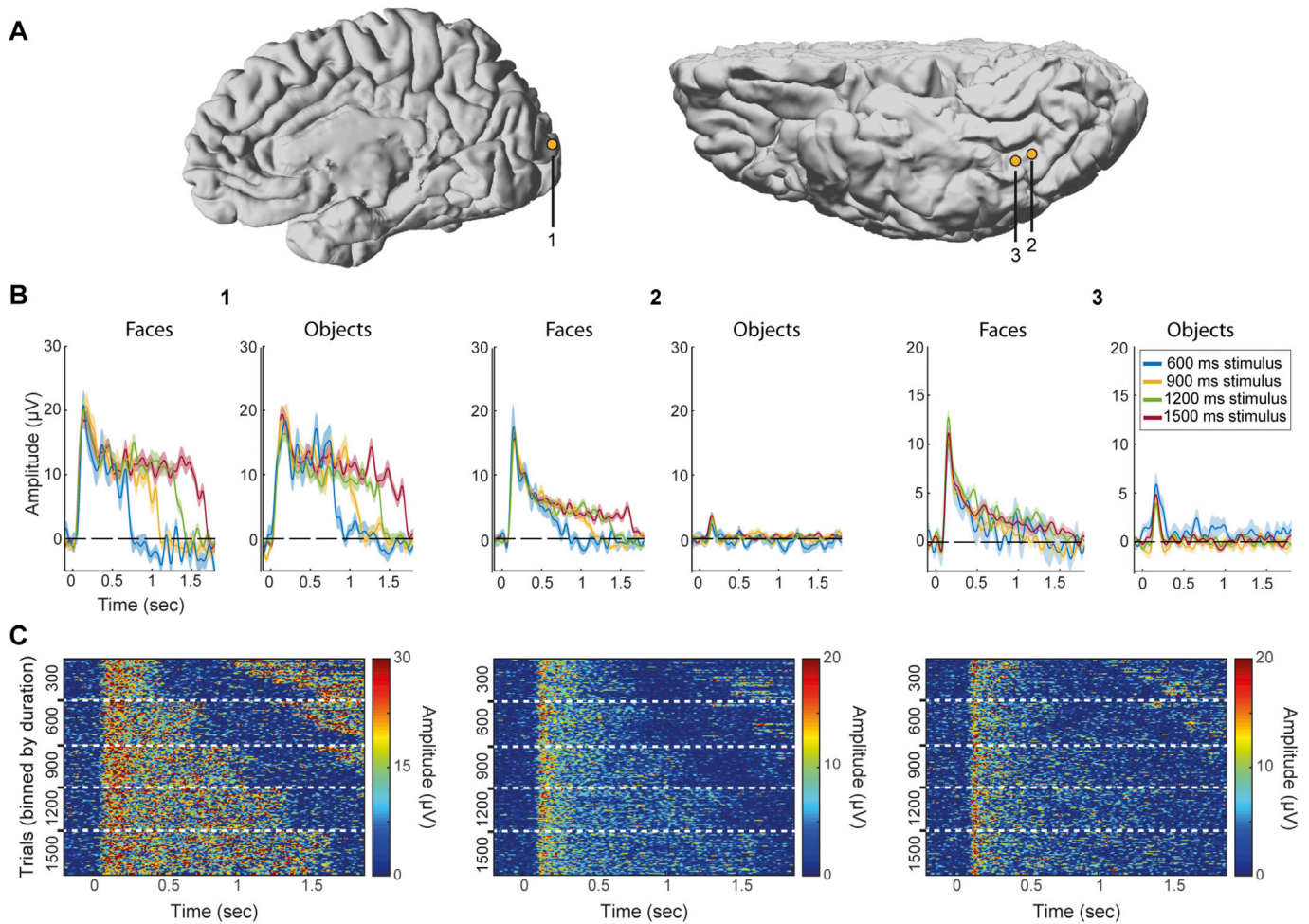
Fig. 5. Three representative electrodes. Electrode 1 - Early visual cortex electrode (dorsal V2/V3 according to fMRI retinotopy performed on this subject for another study, see Parvizi et al., 2012), duration-tracking but not category-selective (trials pooled from both categories). Electrode 2 - Lateral fusiform gyrus electrode with sustained activity that is both duration-tracking and category-selective. Electrode 3 - Slightly anterior to electrode 2 on the lateral fusiform gyrus, category-selective but not duration-tracking. MNI coordinates for the three electrode are (9.4, −94.8, 12.7), (37.6, −29.5, −11.1), (39.1, −25, −11.2). (**A**) Electrode locations on native brains (electrodes 2, 3 are from the same subject). (**B**) Mean response for each stimulus duration and category. For clarity only trials lasting at least 1800 ms (onset to onset) are shown, thereby eliminating 300 ms stimulus duration trials. Shading corresponds to standard error across trials. Note different y-axis scales. (**C**) Single-trial images (stacks) binned by stimulus duration (300, 600, 900, 1200, 1500 ms from top to bottom). Each row in the images represents a single face trial. Increased activation at the end of short-duration trials (top right corner of each image) corresponds to the subsequent trial when the ISI is short.

experiment provided similar results to the main single-task experiment for all subjects. Two subjects with relevant category selective IT electrodes (8 and 21 electrodes over the left IT cortex) adequately performed the second, blurring detection task in the dual-task control experiment (hit rates 75% and 94%; hit rates for detection of clothing items in these subjects: 88% and 94%; overall false alarm rates 4% and 1% respectively). The same relation between HFB and non-target stimulus duration was found with the dual-task control as with the single-task original experiment in the IT electrodes from these two subjects, as well as in all other electrodes for all seven subjects who performed the dual-task control (Fig. 7; the other subjects either missed the majority of the blurring targets while continuing to respond to the clothing items, or did not have suitable electrode coverage for this analysis). This argues against the possibility that the absence of sustained activity in IT electrodes was due to lack of sustained spatial attention, or shifts of gaze away from the stimulus. Indeed, while it could be expected that the introduction of the secondary control task will increase activity along the visual pathway because of the higher attentional demands (Brefczynski and DeYoe, 1999), subjects likely paid sufficient attention to the stimuli already in the main experiment (creating a ceiling effect for attention). Given this insensitivity to other experimental conditions, data from both experiments was pooled in all other analyses.

3.4. Duration-tracking is not confounded with offset responses

Because offset responses are also related to the duration of the stimulus, our analysis was aimed to reduce the effect of such responses over duration-tracking measures. As noted above, duration-tracking required single trial sustained response to remain above baseline for any 300 ms window throughout the duration of the stimulus, for all durations. An electrode exhibiting only onset- and offset-responses would not pass this requirement.

Furthermore, we repeated the analysis of the duration-tracking gradient limited to electrodes not displaying offset responses and obtained the same results. Fourteen electrodes in four subjects (35 prior to multiple comparisons correction) showed a significant HFB offset response (Inline Supplementary Fig. S1). Only 2 of the 14 were also duration-tracking. Eliminating all electrodes with offset responses yielded a duration-tracking gradient which was similar to the gradient found including these electrodes.

3.5. Is duration-tracking activity driven by eye-movements?

The finding of duration-tracking responses could be potentially explained as the temporally summed activity of transient (onset) responses to retinal shifts driven by eye movements (EM) while the
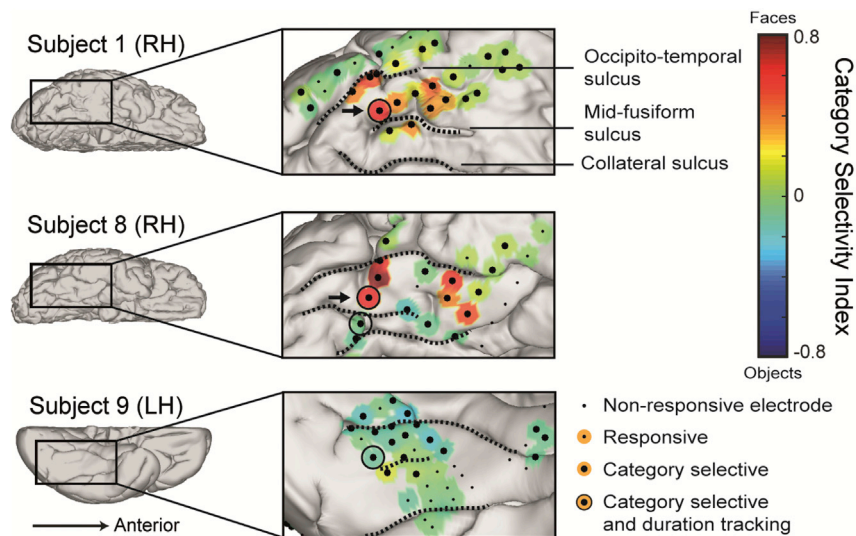
Fig. 6. Electrodes on the posterior fusiform gyrus encode stimulus persistent presence and its category. Four significantly duration-tracking electrodes in three subjects were found within category-selective IT cortex, all located adjacently to the mid-fusiform sulcus on the posterior fusiform gyrus, as opposed to more anterior category-selective sites which do not track ongoing stimulus presence. Of these, the two face-selective electrodes (from two subjects) marked by arrows also maintained category-selectivity persistently throughout the duration of the response (marked with arrows) whereas the two others did not. The electrodes are presented on the subjects' native brains.

stimulus is presented (Nagasawa et al., 2011). However, there are good reasons to suggest that duration-tracking responses are not merely driven by EM. First, eye movements occur about 1–2 (max. 4) per second, and thus EM-driven activity would have taken the form of random sparse bursts in single trials, but these were not consistently observed. Second, although eye movements could not be tracked with fidelity in the clinical ICU setting, saccadic spike potentials elicited by the contraction of the ocular muscles at the onset of saccades (Yuval-Greenberg et al., 2008), can be detected intracranially in electrodes proximal to the eyes (Jerbi et al., 2009; Kovach et al., 2011). We were able to identify reliable saccadic events in one of our patients through the detection of intracranial saccadic spike potentials in two temporal pole electrodes, just posterior to the eye (Fig. 8A). We detected saccadic potentials using a custom filter, based on a canonical spike potential template derived from independent recordings of saccadic spike potentials from three EEG control subjects who performed the same experiment with concurrent eye-tracking (see Materials and Methods). The claim that the intracranial spikes indeed corresponded to saccadic potentials was supported by producing a peri-stimulus saccade rate modulation curve, which matched the pattern observed in the eye-tracked control subjects and was consistent with expected peri-stimulus pattern for saccades (i.e. initial suppression followed by increased saccade rate, Fig. 8C), as well as by the theoretically expected mean saccade rate of about 1.5 saccades/sec (Engbert, 2006). Finally, to test for the effect of saccades on sustained HFB visual responses, we examined data from a posterior face-selective duration-tracking inferior-temporal electrode, in the same subject in which the intracranial spike potentials were recorded. The data was divided into two trial groups: one group of 11 trials having zero detected saccades within the 1500 ms post-stimulus window, and a second group of the same number of trials having the most detected saccades (3–6 per second, mean 4.2). The HFB response to 1500-ms-duration face stimuli was then averaged for each group and compared. The results show no difference between the trial groups (Fig. 8D,E). Similar results were obtained by comparing the trials with more saccades to those with fewer saccades by a median split based on the number of saccades per trial. Thus, the face-selective duration-tracking in a posterior IT electrode was not contingent on the presence of saccades during stimulus presentation. Finally, to further test whether the sustained activity in early visual areas could be explained by saccade-locked activity, we simulated a best-case scenario for this hypothesis and tested it against the empirical results (Fig. 8F). Whereas these results do not account for finer fixational eye

movements like tremor or drift (Martinez-Conde et al., 2004), the simulation results show that saccade-driven responses are unlikely to account for the robust duration-tracking activity observed in early visual cortex. We return to the potential role of eye movements in the discussion.

3.6. Duration-tracking in low-frequency activity

In addition to the analysis of the HFB signal, duration-dependence was also assessed in the time domain (i.e., event related potential) using the same method as for HFB, except that original rather than bootstrapped trials were used and the signal was inverted in polarity if the mean of the 1000–1500 ms window of the response to 1500 ms duration stimuli was negative, to allow for negative sustained potentials. We found positive and negative *duration dependent sustained potentials* (DDSPs), that is, deflections which remained distinct from baseline for the duration of the stimulus (cf. N'Diaye et al., 2004 in a duration-estimation task). These were found in four non-category-selective electrodes on the inferior bank of the calcarine sulcus in three subjects (Fig. 9), all mapped to ventral V1 based on a probabilistic atlas (Wang et al., 2014). Of the four electrodes with significant DDSPs, three showed also significant HFB duration-tracking and one only a non-significant but visually discernible sustained HFB response. In contrast, other sites with strong HFB duration-tracking in the same subjects did not show any type of DDSP. We note that in some of the subjects the 0.5 Hz high pass filter applied during the recording might have reduced the sensitivity to DDSPs.

Duration-dependence could also characterize the power modulation of low-frequency bands (e.g. in the alpha/beta range). However, because of the strong low-frequency power of the onset and offset responses, and the lower temporal resolution of time-frequency analysis in these bands (temporal smearing), resolving low frequency duration-tracking necessitates longer stimulation durations than were used in our study.

4. Discussion

4.1. Duration-tracking as an organizing principle in the ventral stream

We show that while activity in early visual cortex remains above threshold for as long as the stimulus is presented, this correspondence diminishes along the visual hierarchy and is minimal in category-
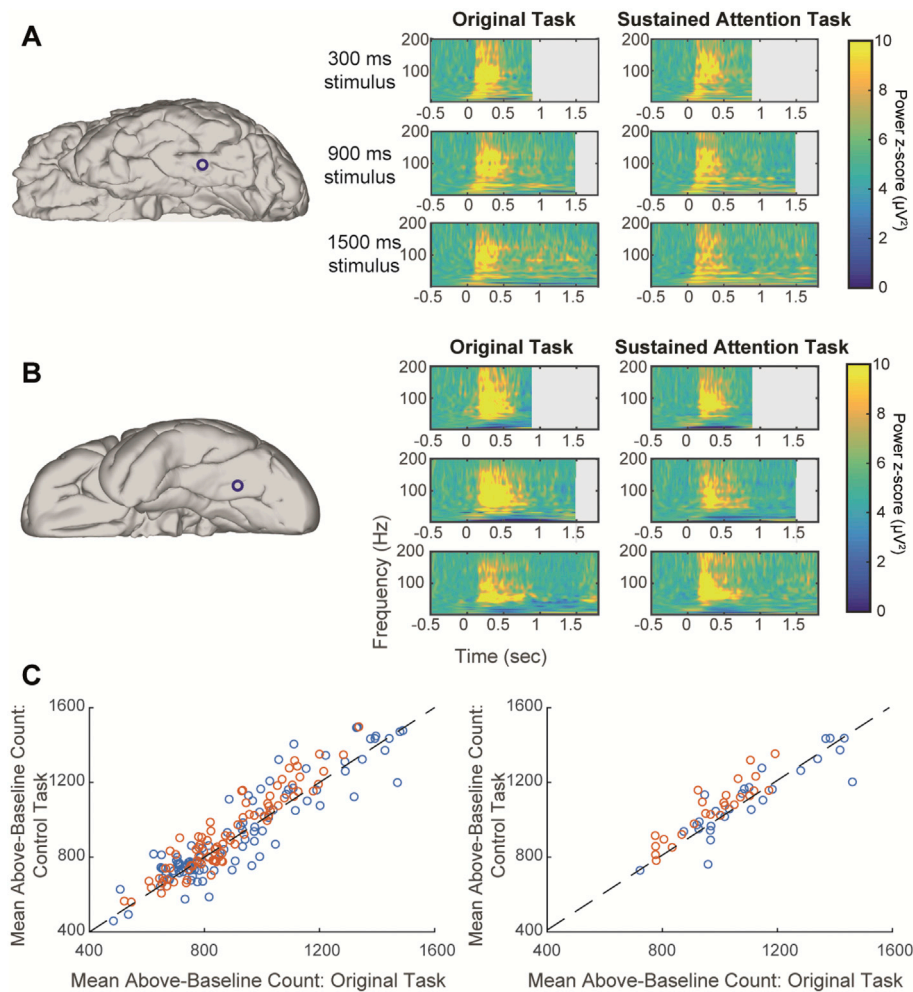
Fig. 7. No differences between single task and dual-task attention control. Panels A, B show time-frequency plots from two representative electrodes in two subjects who performed well for both the clothing items and the blurring events detection task (attention-control). Left column: electrode location on individual subject's brain. Right column: time-frequency plots for a single electrode corresponding to three stimulus durations (top to bottom) and the two experimental conditions. Power is shown as z-score compared to baseline. (**A**) Face-selective, non-duration-tracking IT electrode. (**B**) Object-selective, non-duration-tracking IT electrode. (**C**) Left: each visually-responsive electrode is plotted as one data point comparing the mean number of above-baseline data points (across trials) between the two tasks ($r = 0.92$). The results are clustered around the diagonal demonstrating similar mean response durations across conditions. Orange points are from the three subjects who could perform the control task well. Right: the same relation is shown specifically for category-selective inferior temporal electrodes ($r = 0.88$). There was no significant difference between the two experiments either on the single electrode basis (across trials) or in the mean across electrodes.

selective sites in anterior IT cortex. This anterior-posterior gradient was evident through a direct comparison of duration-tracking accuracy between anatomically-defined visual regions, as well as its correlation with onset response latency and with electrode coordinates on a posterior-anterior vector.

The functional organization of visual IT cortex obeys a number of well-characterized functional-anatomical regularities. These include an increasingly abstract, high-level representation (Kanwisher, 2010; Lerner et al., 2001), increasing spatial receptive field (Wilson and Wilkinson, 2015) and temporal receptive window size (Honey et al., 2012) along the posterior-anterior axis, as well as variable preference for stimulus eccentricity (Hasson et al., 2002), animacy, and stimulus size (Weiner et al., 2014) along the medial-lateral axis. Our results add temporal dynamics as an additional organizing principle, manifested as diminishing real-time perceptual duration tracking along the visual hierarchy. This gradient is comparable in nature to the other posterior-anterior gradients in that it entails a growing invariance to low-level properties of the visual signal.

The differences in the fidelity of sustained activity across the visual hierarchy could be seen as expressing differences in the functional role of different cortical areas. For example, cortical activity in EVC and posterior FFA showing reliable (single-trial) duration-tracking may maintain

an ongoing visual percept, while activity in anterior areas dominated by onset responses can be a signature of stimulus-onset-related processes such as categorization or identification. Activity in non-duration-tracking downstream areas could also reflect high-level cognitive processing, which may not be strictly contingent on stimulus presence (fluctuations in attention to stimulus features, imagery etc.).

Why do different regions show different degrees of duration-tracking? One possible explanation could be intrinsically different temporal dynamics (phasic vs. tonic) of the neurons or local circuits in different regions. An alternative account of the duration-tracking gradient is that it is driven by the interaction of small eye movement with increasing receptive field size along the visual stream. We could record saccades based on the finding of the saccadic spike potentials in one of the patients, and showed that sustained responses did not differ in trials with or without saccades. Additionally, modelling the potential effect of saccades on visual responses could not explain the observed results. As corroborating evidence we also note that in a separate set of EEG experiments with simultaneous eye-tracking, duration-tracking HFB (>50 Hz) activity recorded in posterior scalp electrodes was found to be independent of saccade rate (Gerber and Deouell, 2016). Nevertheless, we cannot completely rule out the option that small jitter in the visual field in the form of drift, smaller saccades than could be detected, or even
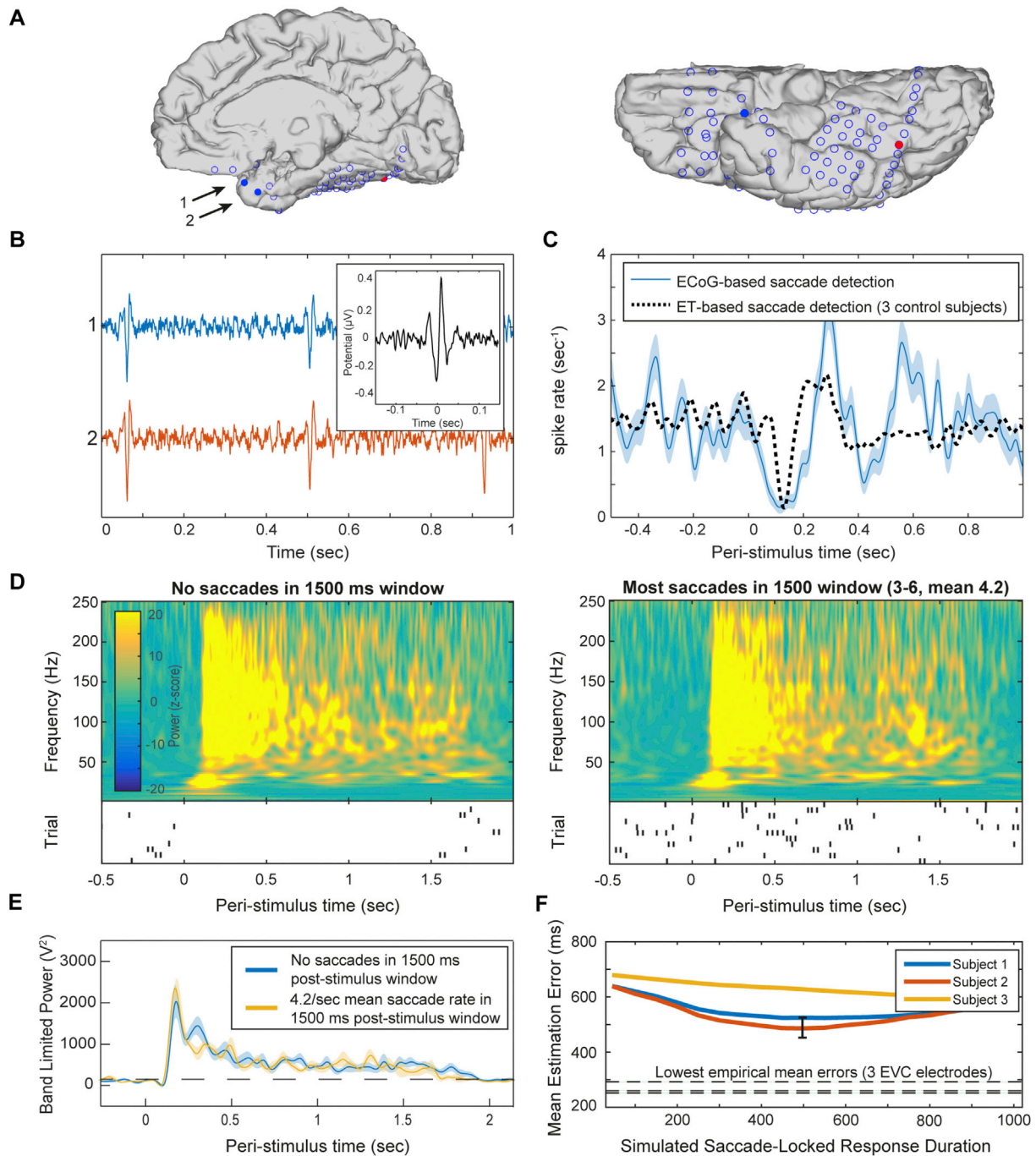
Fig. 8. No effect of saccades on sustained visual activity. Saccadic events were identified in a single subject based on detection of saccadic spike potentials in two electrodes near the temporal pole. (**A**) Anatomical location of electrodes used to detect saccadic spikes (blue filled circles), and an electrode in the same subject exhibiting face-selective duration-tracking (red). (**B**) Representative un-averaged traces (>30 Hz high-pass-filtered) of spike events in the two temporal-pole electrodes. Inset: the averaged saccadic spike model from the 3 EEG control subjects which served as a model and in which saccade onsets were detected by eye tracking. (**C**) Peri-stimulus saccade rate. The ECoG-based saccades follow a typical post-stimulus suppression and rebound (blue), comparable to saccades detected using eye-tracking for the same task in control subjects (black dotted). (**D**) Average time-frequency plots of the responses to 1500 ms face stimuli in the posterior IT electrode showing face-selective duration-tracking, for trials with no saccades during stimulus presentation (left), vs. trials with many saccades (right). Shown below each time frequency plot is a raster plot of saccade events during each trial included in the group. Each row represents a single trial and each tick mark represents a detected saccade. (**E**) Direct comparison of high-frequency band-limited power for the two groups of trials. Cluster-based permutation testing did not detect differences in power at any time point. Dotted line marks mean baseline power. (**F**) Mean error for stimulus duration estimation from simulated HFB power constructed based on individual saccades, as function of the simulated duration of the saccade evoked activity. Each line represents simulation based on saccades from one of three EEG subjects. The lowest mean error achieved in the simulation was 485 ms, which is larger than that of 15 of the most strongly duration-dependent intracranial electrodes in the early visual cortex. The black error bar indicates 95% confidence interval across trials for the best-performing (lowest error) simulation (the three lowest errors from intracranial electrodes are shown as horizontal broken lines). We conclude that it is unlikely that the robust duration-tracking responses in early visual cortex were the result of summed transient responses driven by eye movements, regardless of the duration of these saccade-related bursts.

tremor, created more variable input to neural populations with smaller receptive fields, which could lead to less adaptation over time in early than high-level visual areas. This possible explanation does not undermine the finding that duration-tracking diminishes along the ventral visual pathway. Rather, it could show that a spatial constraint may translate to a critical constraint on temporal dynamics of a neuronal
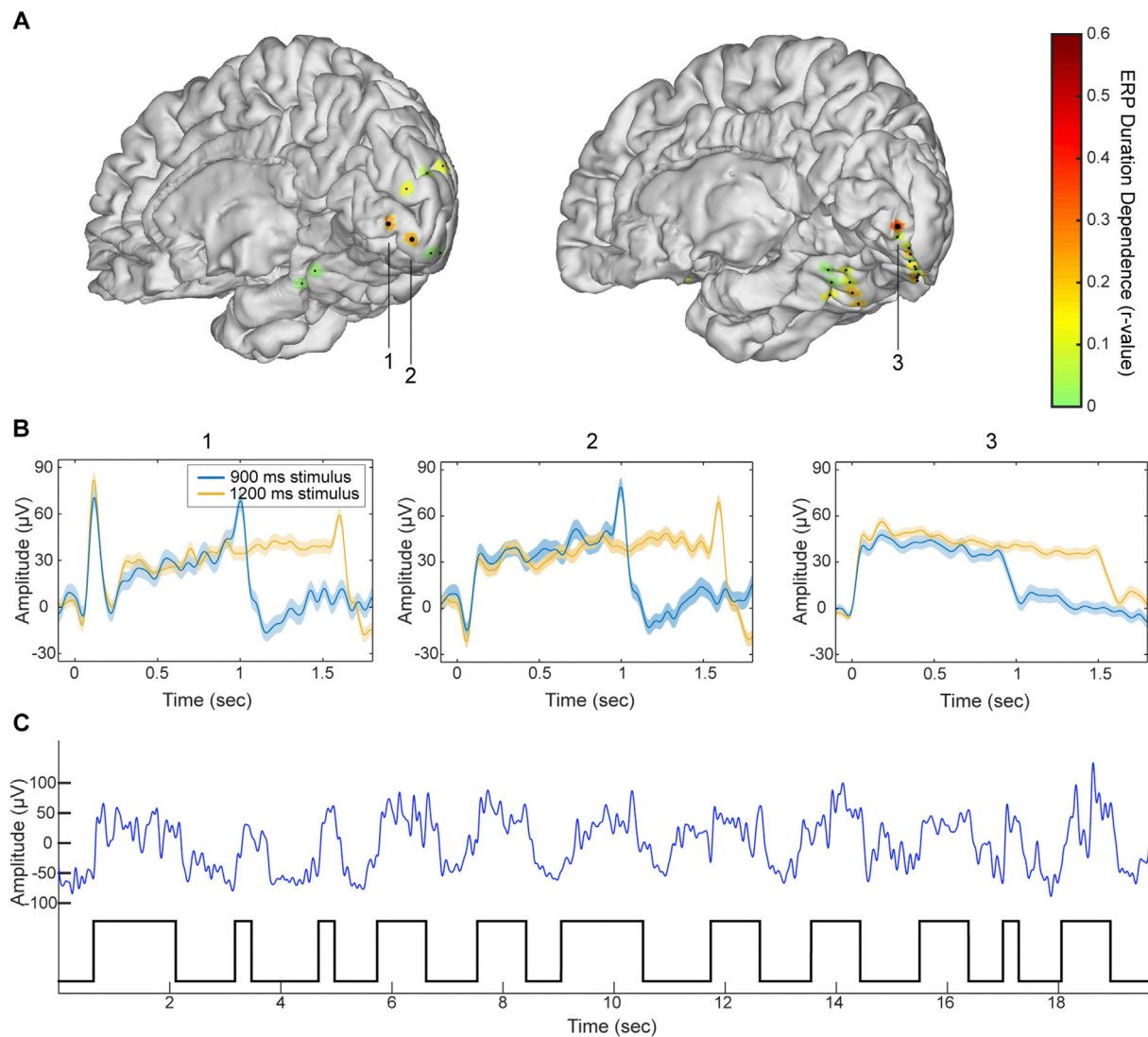
Fig. 9. Duration dependent sustained potentials. (A) Duration-dependence of evoked responses in two individual subject brains (subjects 5 and 10). Colored patches indicate degree of duration-dependence according to the color scale on the right. Large dots indicate statistically significant duration-dependence. (B) Mean evoked responses for the three electrodes identified in panel (A). (C) High-SNR segment of the raw data trace (blue line, low-pass filtered at 10 Hz) taken from electrode 3 and on-off visual stimulation epochs (black trace on bottom). The robust DDSP evidenced without averaging, indicates that DDSP corresponds to sustained single trials deflections rather than to a summation of stimulus-contingent transient deflections in single trials.

pool, which may shape its functional role. Moreover, even if diminishing duration tracking is a consequence of increasing receptive field sizes, the question still stands of how the visual percept is maintained despite adaptation and decreased response fidelity in high-level areas.

Finally, this gradual attenuation of information about stimulus presence may also reflect a basic property of information transfer along a serial processing stream. As formally stated by the data processing inequality theorem (Cover and Thomas, 1991), in any system the amount of information about the initial input cannot be increased as a result of additional processing. Thus, any information about the real-time presence of a visual percept in a cortical site must be present also in the input to that site. Given imperfect synaptic transmission, this information can be expected to become less reliable along the processing stream – and likely more so for sustained than onset-driven activity, as the former is typically of lesser magnitude and therefore more likely to fall below neural detection thresholds. The fact that category selectivity seems to emerge rather than decrease in downstream sites does not contradict this principle, as category information is fully present in the retina, and therefore it is not gained, but rather made explicit by the integration of low-level visual features as well as stored templates into representations

which are increasingly invariant to low-level detail, e.g. to stimulus contrast (Avidan et al., 2002). Persistent stimulus-driven neural activity corresponding to the presence of a visual feature at any processing stage can thus be expected to be only as reliable as the persistence of its input in the previous stage, and as such is bound to show some degree of decay with subsequent processing steps. Functional differentiation in visual processing along the ventral stream could in part arise from this type of fundamental property of the cortical system.

4.2. Differentiation within category-selective cortex

We found that within face-selective regions of the fusiform gyrus, significant duration-tracking activity was found only at some posterior recording sites, corresponding to the FFA-1 region, while more anterior (FFA-2) sites showed non-reliable duration-tracking activity. These results provide evidence that regions which appear homogenous in terms of their category-selectivity (Tsao et al., 2006), but are anatomically and cytoarchitectonically distinct (Caspers et al., 2013; Grill-Spector and Weiner, 2014; Lorenz et al., 2015), differ in how they respond to a stimulus over time and in their potential contribution to ongoing visual

perception.

The temporal response profile difference found between EVC and IT regions and between posterior and anterior category selective regions could be related to previous fMRI results (Weiner et al., 2010), although the paradigm and especially the time scales are quite different. In Weiner et al.'s study, IT regions showed larger fMRI adaptation effects (reduced BOLD response to repeated vs. non-repeated stimuli) than EVC regions, and anterior face- and limb-selective regions showed stronger adaptation than their more posterior counterparts. The latter difference was pronounced in a condition with long lags between repetitions (many seconds and intervening stimuli) and not as much with short lags (half a second). The present study, using direct neural measures and high temporal resolution, may have been more sensitive to differences between regions at the sub-second time scale. Moreover, it suggests that a gradient of temporal response profiles is present even within the response to a single stimulus, and not only in the adaptation across stimuli.

4.3. Measuring neural activity under sustained cognitive conditions

The challenge of measuring neural activity contingent on a sustained cognitive process, while dissociating it from other temporal activity profiles such as transient onset- or offset-driven activity, has not been widely addressed in the literature. A common approach in functional imaging studies of sustained cognitive processes (e.g. working memory, sustained attention or imagery) is to use separate GLM predictors for onset-, sustained- and offset-/response-related activity (Curtis and D'Esposito, 2003; Eriksson et al., 2004; Offen et al., 2009; Portas et al., 2000). However, this does not produce reliable evidence for sustained activity that is independent from onset and offset responses, because any neural activity extending beyond the duration of the modeled onset response will be interpreted as sustained activity, regardless of whether its duration is in fact related to the duration of the process. The current study suggests two general guidelines for measuring duration dependent sustained activity: first, the experimental paradigm should allow for variable response durations; and second, the analysis should directly assess the dependence of the response duration on the experimental duration (see Vilberg and Rugg, 2012 for an example of such analysis in the context of memory retention). Direct examination of duration-tracking in neural activity associated with sustained cognitive processes may benefit a wide range of studies.

5. Conclusions

Onset responses reflect major changes in the environment, yet much of perception occurs between changes in the visual stream. Understanding the neural correlates of perception beyond onset responses has major implications for deciphering the biological underpinning of experience. We addressed the neural basis of sustained visual perception and report dissociations between semantic category ("What") and ongoing presence ("Whether") of a visual event evident along the ventral visual pathway. Specifically, the real-time, single-trials encoding of the presence of a visual stimulus is robust in early visual cortex but decreasingly reliable toward anterior, category-selective areas. These results raise fundamental questions about representation of visual information beyond moments of change, and should constrain models about where and how the experience of a protracted presence of a categorically recognized object is encoded moment-by-moment. The ongoing representation could rely on the intermediate zone in posterior IT showing category selectivity as well as duration-tracking; however, these responses were rare despite extensive coverage of the ventral stream. Alternatively, ongoing representation could be based on a combination of the onset response in category selective regions raising a 'recognition flag' upon the initial recognition of an object, while the sustained activity in EVC provides a presence ('still here') index. However, this option still leaves the question of whether these signals are merged somewhere, or whether the distributed representation is

sufficient for subjective experience. Finally, the subjective experience of a non-discrete, truly continuous perception of a stimulus, may be a post-hoc reconstruction, not requiring high fidelity moment-to-moment neural activity. Resolving these options requires continued efforts, using multiple imaging modalities, to study perception beyond the dramatic moments of change.

Appendix A. Supplementary data

Supplementary data related to this article can be found at http://dx.doi.org/10.1016/j.neuroimage.2017.08.028.

References

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

Argall, B.D., Saad, Z.S., Beauchamp, M.S., 2006. Simplified intersubject averaging on the cortical surface using SUMA. Hum. Brain Mapp. 27, 14–27. http://dx.doi.org/10.1002/hbm.20158.

Avidan, G., Harel, M., Hendler, T., Ben-bashat, D., Zohary, E., Papanikolaou, A., Keliris, G. a, Papageorgiou, T.D., Shao, Y., Krapp, E., Papageorgiou, E., Stingl, K., Bruckmann, A., Schiefer, U., Logothetis, N.K., Smirnakis, S.M., 2002. Contrast sensitivity in human visual areas and its relationship to object recognition. J. Neurophysiol. 87, 3102–3116.

Benjamini, Y., Yosef, H., 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. J. R. Stat. Soc. 57, 289–300.

Bentin, S., Allison, T., Puce, A., Perez, E., McCarthy, G., 1996. Electrophysiological studies of face perception in humans. J. Cogn. Neurosci. 8, 551–565. http://dx.doi.org/10.1162/jocn.1996.8.6.551.

Brefczynski, J.A., DeYoe, E.A., 1999. A physiological correlate of the "spotlight" of visual attention. Nat. Neurosci. 2, 370–374. http://dx.doi.org/10.1038/7280.

Buhusi, C.V., Meck, W.H., 2005. What makes us tick? Functional and neural mechanisms of interval timing. Nat. Rev. Neurosci. 6, 755–765. http://dx.doi.org/10.1038/nrn1764.

Caspers, J., Zilles, K., Eickhoff, S.B., Schleicher, A., Mohlberg, H., Amunts, K., 2013. Cytoarchitectonical analysis and probabilistic mapping of two extrastriate areas of the human posterior fusiform gyrus. Brain Struct. Funct. 218, 511–526. http://dx.doi.org/10.1007/s00429-012-0411-8.

Cover, T.M., Thomas, J.A., 1991. Elements of Information Theory. Wiley, New York.

Curtis, C.E., D'Esposito, M., 2003. Persistent activity in the prefrontal cortex during working memory. Trends Cogn. Sci. 7, 415–423. http://dx.doi.org/10.1016/S1364-6613(03)00197-9.

Dykstra, A.R., Chan, A.M., Quinn, B.T., Zepeda, R., Keller, C.J., Cormier, J., Madsen, J.R., Eskandar, E.N., Cash, S.S., 2012. Individualized localization and cortical surface-based registration of intracranial electrodes. Neuroimage 59, 3563–3570. http://dx.doi.org/10.1016/j.neuroimage.2011.11.046.

Engbert, R., 2006. Microsaccades: a microcosm for research on oculomotor control, attention, and visual perception. Prog. Brain Res. 154, 177–192. http://dx.doi.org/10.1016/S0079-6123(06)54009-9.

Engell, A.D., Huettel, S., Mccarthy, G., 2012. The fMRI BOLD signal tracks electrophysiological spectral perturbations, not event-related potentials. Neuroimage 59, 2600–2606. http://dx.doi.org/10.1016/j.neuroimage.2011.08.079.

Eriksson, J., Larsson, A., Riklund Ahlström, K., Nyberg, L., 2004. Visual consciousness: dissociating the neural correlates of perceptual transitions from sustained perception with fMRI. Conscious. Cogn. 13, 61–72. http://dx.doi.org/10.1016/S1053-8100(03)00050-3.

Fisch, L., Privman, E., Ramot, M., Harel, M., Nir, Y., Kipervasser, S., Andelman, F., Neufeld, M.Y., Kramer, U., Fried, I., Malach, R., 2009. Neural "ignition": enhanced activation linked to perceptual awareness in human ventral stream visual cortex. Neuron 64, 562–574. http://dx.doi.org/10.1016/j.neuron.2009.11.001.

Fischl, B., Sereno, M.I., Tootell, R.B.H., Dale, A.M., 1999. High-resolution inter-subject averaging and a surface-based coordinate system. Hum. Brain Mapp. 8, 272–284.

Gerber, E.M., Deouell, L.Y., 2016. Evoked and high-frequency markers of sustained visual perception in scalp EEG. In: 798.11/Y15. Neuroscience Meeting Planner. Society for Neuroscience, San Diego, CA, p. 2016. Online).

Gilaie-Dotan, S., Nir, Y., Malach, R., 2008. Regionally-specific adaptation dynamics in human object areas. Neuroimage 39, 1926–1937. http://dx.doi.org/10.1016/j.neuroimage.2007.10.010.

Grill-Spector, K., Weiner, K.S., 2014. The functional architecture of the ventral temporal cortex and its role in categorization. Nat. Rev. Neurosci. 15, 536–548. http://dx.doi.org/10.1038/nrn3747.

Hasson, U., Levy, I., Behrmann, M., Hendler, T., Malach, R., 2002. Eccentricity bias as an organizing principle for human high-order object areas. Neuron 34, 479–490. http://dx.doi.org/10.1016/S0896-6273(02)00662-1.

Honey, C.J., Thesen, T., Donner, T.H., Silbert, L.J., Carlson, C.E., Devinsky, O., Doyle, W.K., Rubin, N., Heeger, D.J., Hasson, U., 2012. Slow cortical dynamics and the accumulation of information over long timescales. Neuron 76, 423–434. http://dx.doi.org/10.1016/j.neuron.2012.08.011.

Ikeda, H., Wright, M., 1974. Evidence for" sustained" and" transient" neurones in the cat's visual cortex. Vis. Res. 14, 133–136.

Jacques, C., Witthoft, N., Weiner, K.S., Foster, B.L., Rangarajan, V., Hermes, D., Miller, K.J., Parvizi, J., Grill-Spector, K., 2016. Corresponding ECoG and fMRI category-selective signals in human ventral temporal cortex. Neuropsychologia 83, 14–28. http://dx.doi.org/10.1016/j.neuropsychologia.2015.07.024.

Jenkinson, M., Bannister, P., Brady, M., Smith, S., 2002. Improved optimization for the robust and accurate linear registration and motion correction of brain images. Neuroimage 17, 825–841. http://dx.doi.org/10.1016/S1053-8119(02)91132-8.

Jenkinson, M., Smith, S., 2001. A global optimisation method for robust affine registration of brain images. Med. Image Anal. 5, 143–156. http://dx.doi.org/10.1016/S1361-8415(01)00036-6.

Jerbi, K., Freyermuth, S., Dalal, S., Kahane, P., Bertrand, O., Berthoz, A., Lachaux, J.-P., 2009. Saccade related gamma-band activity in intracerebral EEG: dissociating neural from ocular muscle activity. Brain Topogr. 22, 18–23. http://dx.doi.org/10.1007/s10548-009-0078-5.

Kanwisher, N., 2010. Functional specificity in the human brain: a window into the functional architecture of the mind. Proc. Natl. Acad. Sci. 107, 11163–11170. http://dx.doi.org/10.1073/pnas.1005062107.

Keren, A.S., Yuval-Greenberg, S., Deouell, L.Y., 2010. Saccadic spike potentials in gamma-band EEG: characterization, detection and suppression. Neuroimage 49, 2248–2263. http://dx.doi.org/10.1016/j.neuroimage.2009.10.057.

Kovach, C.K., Tsuchiya, N., Kawasaki, H., Oya, H., Howard, M. a, Adolphs, R., 2011. Manifestation of ocular-muscle EMG contamination in human intracranial recordings. Neuroimage 54, 213–233. http://dx.doi.org/10.1016/j.neuroimage.2010.08.002.

Kulikowski, J.J., Bishop, P.O., Kato, H., 1979. Sustained and transient responses by cat striate cells to stationary flashing light and dark bars. Brain Res. 70, 362–367.

Lerner, Y., Hendler, T., Ben-Bashat, D., Harel, M., Malach, R., 2001. A hierarchical axis of object processing stages in the human visual cortex. Cereb. Cortex 11, 287–297. http://dx.doi.org/10.1093/cercor/11.4.287.

Levy, I., Hasson, U., Avidan, G., Hendler, T., Malach, R., 2001. Center-periphery organization of human object areas. Nat. Neurosci. 4, 533–539. http://dx.doi.org/10.1038/87490.

Lewis, P. a., Miall, R.C., 2003. Brain activation patterns during measurement of sub- and supra-second intervals. Neuropsychologia 41, 1583–1592. http://dx.doi.org/10.1016/S0028-3932(03)00118-0.

Lorenz, S., Weiner, K.S., Caspers, J., Mohlberg, H., Schleicher, A., Bludau, S., Eickhoff, S.B., Grill-Spector, K., Zilles, K., Amunts, K., 2015. Two new cytoarchitectonic areas on the human mid-fusiform gyrus. Cereb. Cortex bhv225. http://dx.doi.org/10.1093/cercor/bhv225.

Manning, J.R., Jacobs, J., Fried, I., Kahana, M.J., 2009. Broadband shifts in local field potential power spectra are correlated with single-neuron spiking in humans. J. Neurosci. 29, 13613–13620. http://dx.doi.org/10.1523/JNEUROSCI.2041-09.2009.

Maris, E., Oostenveld, R., 2007. Nonparametric statistical testing of EEG-and MEG-data. J. Neurosci. Methods 164, 177–190. http://dx.doi.org/10.1016/j.jneumeth.2007.03.024.

Martinez-Conde, S., Macknik, S.L., Hubel, D.H., 2004. The role of fixational eye movements in visual perception. Nat. Rev. Neurosci. 5, 229–240. http://dx.doi.org/10.1038/nrn1348.

Miller, K.J., Honey, C.J., Hermes, D., Rao, R.P.N., DenNijs, M., Ojemann, J.G., 2014. Broadband changes in the cortical surface potential track activation of functionally diverse neuronal populations. Neuroimage 85, 711–720. http://dx.doi.org/10.1016/j.neuroimage.2013.08.070.

Miller, K.J., Sorensen, L.B., Ojemann, J.G., den Nijs, M., 2009. Power-law scaling in the brain surface electric potential. PLoS Comput. Biol. 5, e1000609. http://dx.doi.org/10.1371/journal.pcbi.1000609.

N'Diaye, K., Ragot, R., Garnero, L., Pouthas, V., 2004. What is common to brain activity evoked by the perception of visual and auditory filled durations? A study with MEG and EEG co-recordings. Brain Res. Cogn. Brain Res. 21, 250–268. http://dx.doi.org/10.1016/j.cogbrainres.2004.04.006.

Nagasawa, T., Matsuzaki, N., Juhász, C., Hanazawa, A., Shah, A., Mittal, S., Sood, S., Asano, E., 2011. Occipital gamma-oscillations modulated during eye movement tasks: simultaneous eye tracking and electrocorticography recording in epileptic patients. Neuroimage 58, 1101–1109. http://dx.doi.org/10.1016/j.neuroimage.2011.07.043.

Offen, S., Schluppeck, D., Heeger, D.J., 2009. The role of early visual cortex in visual short-term memory and visual attention. Vis. Res. 49, 1352–1362. http://dx.doi.org/10.1016/j.visres.2007.12.022.

Parvizi, J., Jacques, C., Foster, B.L., Withoft, N., Rangarajan, V., Weiner, K.S., Grill-Spector, K., 2012. Electrical stimulation of human fusiform face-selective regions distorts face perception. J. Neurosci. 32, 14915–14920. http://dx.doi.org/10.1523/JNEUROSCI.2609-12.2012.

Petersen, S.E., Miezin, P., Allman, J., 1988. Transient and sustained responses in four extrastriate visual areas of the owl monkey. Exp. Brain Res. 70, 55–60.

Portas, C.M., Strange, B. a, Friston, K.J., Dolan, R.J., Frith, C.D., 2000. How does the brain sustain a visual percept? Proc. Biol. Sci. 267, 845–850. http://dx.doi.org/10.1098/rspb.2000.1080.

Pouthas, V., Garnero, L., Ferrandez, A.M., Renault, B., 2000. ERPs and PET analysis of time perception: spatial and temporal brain mapping during visual discrimination tasks. Hum. Brain Mapp. 10, 49–60. http://dx.doi.org/10.1002/(SICI)1097-0193(200006)10, 2<49::AID-HBM10>3.0.CO;2–8.

Ray, S., Maunsell, J.H.R., 2011. Different origins of gamma rhythm and high-gamma activity in Macaque visual cortex. PLoS Biol. 9, e1000610. http://dx.doi.org/10.1371/journal.pbio.1000610.

Tsao, D.Y., Freiwald, W.A., Tootell, R.B.H., Livingstone, M.S., 2006. A cortical region consisting entirely of face-selective cells. Science 311, 670–674. http://dx.doi.org/10.1126/science.1119983 (80-.).

Vilberg, K.L., Rugg, M.D., 2012. The neural correlates of recollection: transient versus sustained FMRI effects. J. Neurosci. 32, 15679–15687. http://dx.doi.org/10.1523/JNEUROSCI.3065-12.2012.

Wang, L., Mruczek, R., Arcaro, M., Kastner, S., 2014. Probabilistic maps of visual topography in human cortex. Cereb. Cortex 1–21. http://dx.doi.org/10.1093/cercor/bhu277.

Weiner, K.S., Golarai, G., Caspers, J., Chuapoco, M.R., Mohlberg, H., Zilles, K., Amunts, K., Grill-Spector, K., 2014. The mid-fusiform sulcus: a landmark identifying both cytoarchitectonic and functional divisions of human ventral temporal cortex. Neuroimage 84, 453–465. http://dx.doi.org/10.1016/j.neuroimage.2013.08.068.

Weiner, K.S., Sayres, R., Vinberg, J., Grill-Spector, K., 2010. fMRI-adaptation and category selectivity in human ventral temporal cortex: regional differences across time scales. J. Neurophysiol. 103, 3349–3365. http://dx.doi.org/10.1152/jn.01108.2009.

Wilson, H.R., Wilkinson, F., 2015. From orientations to objects: configural processing in the ventral stream. J. Vis. 15, 1–10. http://dx.doi.org/10.1167/15.7.4.doi.

Yuval-Greenberg, S., Tomer, O., Keren, A.S., Nelken, I., Deouell, L.Y., 2008. Transient induced gamma-band response in EEG as a manifestation of miniature saccades. Neuron 58, 429–441. http://dx.doi.org/10.1016/j.neuron.2008.03.027.